

Tiago Gräf

CARACTERIZAÇÃO MOLECULAR DA EPIDEMIA DO HIV-1 EM CIDADES DO INTERIOR DE SANTA CATARINA E RIO GRANDE DO SUL E FILOGEOGRAFIA DO SUBTIPO C NO BRASIL

Tese submetida ao Programa de Pós-Graduação em Biotecnologia e Biociências da Universidade Federal de Santa Catarina, para obtenção de título de Doutor.

Orientadores:
Prof. Dr. Aguinaldo R. Pinto
Prof^a. Dr^a Sabrina E. M. De Almeida

Florianópolis
2015

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

Gräf, Tiago

CARACTERIZAÇÃO MOLECULAR DA EPIDEMIA DO HIV-1 EM
CIDADES DO INTERIOR DE SANTA CATARINA E RIO GRANDE DO SUL
E FILOGEOGRAFIA DO SUBTIPO C NO BRASIL / Tiago Gräf ;
orientador, Aguinaldo Roberto Pinto ; coorientadora,
Sabrina Esteves de Matos Almeida. - Florianópolis, SC, 2015.
181 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro de Ciências Biológicas. Programa de Pós-
Graduação em Biotecnologia e Biociências.

Inclui referências

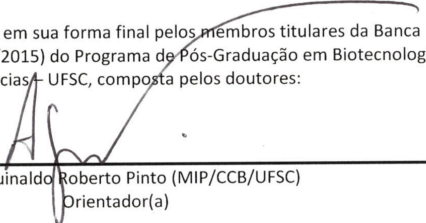
1. Biotecnologia e Biociências. 2. Epidemiologia
Molecular. 3. HIV-1. 4. Filodinâmica. 5. aids. I. Pinto,
Aguinaldo Roberto. II. Almeida, Sabrina Esteves de Matos.
III. Universidade Federal de Santa Catarina. Programa de
Pós-Graduação em Biotecnologia e Biociências. IV. Título.

“Caracterização molecular da epidemia do HIV-1 em cidades do interior de Santa Catarina e Rio Grande do Sul e filogeografia do subtipo C no Brasil”

Por

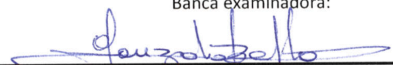
Tiago Gräf

Tese julgada e aprovada em sua forma final pelos membros titulares da Banca Examinadora (11/PPGBTC/2015) do Programa de Pós-Graduação em Biotecnologia e Biociências - UFSC, composta pelos doutores:




Dr(a) Aguinaldo Roberto Pinto (MIP/CCB/UFSC)
Orientador(a)

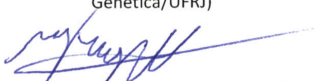
Banca examinadora:



Dr(a) Gonzalo Bello Bentancor (Laboratório de AIDS e Imunologia Molecular/Departamento de Imunologia/Instituto Oswaldo Cruz)



Dr(a) Marcelo Alves Soares (Laboratório de Viologia Humana/Departamento de Genética/UFRJ)



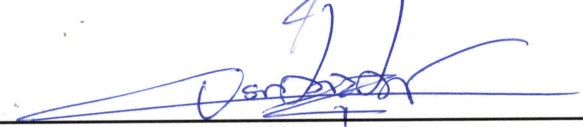
Dr(a) Emil Kupek (Departamento de Saúde Pública/CCS/UFSC)



Dr(a) Carlos Roberto Zanetti (MIP/CCB/UFSC)



Dr(a) Edmundo Carlos Grisard (MIP/CCB/UFSC)



Prof(a). Dr(a). Oscar Bruna Romero
Coordenador(a) do Programa de Pós Graduação em Biotecnologia e Biociências

Florianópolis, 10 de julho de 2015.

AGRADECIMENTOS

Em primeiro lugar, eu preciso agradecer ao meu orientador, professor Aguinaldo Pinto. Trabalhar com ele foi uma experiência muito enriquecedora e certamente fez de mim uma pessoa mais madura e um cientista mais crítico. Agradeço especialmente a confiança e a liberdade que recebi para executar meu projeto de doutorado. Durante os mais de seis anos de convivência, além de um grande mestre, eu tive um bom amigo para me aconselhar.

Eu também sou muito grato à minha coorientadora, professora Sabrina Almeida. As conversas durante o café e as reuniões com o grupo sempre atçaram minha curiosidade, fazendo-me querer aprender mais sobre o HIV e sua epidemia. Obrigado pela confiança e acolhida ao grupo.

Muito obrigado aos meus colegas de laboratório Dennis e Rúbia. Certamente, sem eles esta tese teria sido mais pobre, menor e/ou levaria mais um ano para ser finalizada. Obrigado por compartilharem comigo tantos ensinamentos, tantos projetos (um dia concluiremos aqueles engavetados) e também os momentos de frustração. Vocês são demais!

Obrigado ao professor Philippe Lemey e seu grupo de pesquisa do Laboratório de Epidemiologia e Virologia Clínica da KU Leuven. Eu sou imensamente grato pela acolhida, paciência e amizade no meu período como pesquisador visitante em terras belgas. A experiência que tive lá foi certamente decisiva no direcionamento da minha carreira acadêmica. Dank je wel!!

A todo o pessoal do CDCT/FEPPS, funcionários e estudantes, agradeço a amizade e o companheirismo. Obrigado especialmente à Hegger, que foi importantíssima no momento em que eu precisei de um par de braços extra para terminar as amplificações e os sequenciamentos.

A este programa de pós-graduação, sou grato por permitir que eu realizasse meu doutorado afastado da UFSC.

Muito obrigado a todas as unidades de saúde das cidades participantes deste projeto. Sem a parceria delas e a cooperação dos coordenadores e funcionários, este projeto nunca teria saído do papel.

Eu preciso agradecer também aos meus amigos que foram pacientes neste período da minha vida. Especialmente aos meus companheiros de composições, Victório e Tonho, que, apesar de nem

tão pacientes, não desistiram de mim. Caras, quem sabe agora lançamos um disco!

Obrigado à Karine Pardon, uma grande amiga que foi praticamente uma segunda mãe no período do doutorado sanduíche.

Muito obrigado ao meu tio Paulo e sua família, que me receberam muito carinhosamente toda a vez que eu precisei ir a Florianópolis. Du e Letícia, desculpem-me por roubar seus quartos às vezes, era por insistência da mãe de vocês. Berna, obrigado por sempre se preocupar comigo. Paulo, obrigado por todas as caminhadas na Beira-Mar, todas as caronas, tantas histórias e todas as conversas. Você é uma pessoa especial que eu tenho muita sorte de ter na minha vida.

Obrigado à minha irmã, Anna, que é um exemplo pra mim. Mana, obrigado pela inspiração, obrigado pela preocupação, obrigado pela amizade e por último, mas não menos importante, obrigado por colocar no mundo o ser humano mais fofinho que eu já conheci.

Muito obrigado a meus pais, Inácio e Nila. Pais, vocês construíram a base desta tese, obrigado por me ensinarem a ser correto, a questionar e a buscar o conhecimento sempre. Obrigado por me apoiarem nos meus sonhos, mesmo quando internamente vocês desejavam que eu voltasse para casa. Pai e mãe, amo vocês.

Por último, eu quero agradecer à pessoa que esteve sempre comigo, em toda essa trajetória. Obrigado, Lívia, por largar tudo e ir viajar comigo, obrigado por me fazer sorrir quando eu estou mal-humorado, por planejar os passeios mais incríveis e por me fazer pensar em um modelo evolutivo para a minha vida, e não apenas naquele que se adapta melhor ao HIV. Muito obrigado por fazer eu me empolgar com cada pequena conquista. Te amo!

Existem muitas hipóteses em ciência que estão erradas. Isso é perfeitamente aceitável, elas são a abertura para achar as que estão certas.

Carl Sagan

Este trabalho foi realizado em colaboração com o Centro de Desenvolvimento Científico e Tecnológico (CDCT) da Fundação Estadual de Produção e Pesquisa em Saúde (FEPPS) do Rio Grande do Sul. Recebeu apoio da CAPES, através da concessão da bolsa de doutorado, e recursos financeiros da FAPESC, FAPERGS e CNPq.

APRESENTAÇÃO DA TESE

A presente tese de doutorado foi redigida em um formato que inclui uma introdução geral, objetivos, três capítulos de resultados, uma discussão global dos resultados, conclusões e perspectivas. A introdução geral aborda os conceitos mais relevantes para a interpretação dos resultados apresentados aqui e é dividida em quatro seções: 1) conhecimentos básicos sobre o vírus e a doença; 2) uma revisão sobre métodos em filogenia e filogeografia de maior aplicação no estudo da epidemia do HIV; 3) a história e situação atual da epidemia do HIV-1; 4) o cenário epidemiológico do Brasil e as peculiaridades da região Sul que motivaram a realização deste estudo. Os capítulos de resultados estão apresentados no formato exigido por revistas científicas, apresentando um resumo, breve introdução, materiais e métodos específicos, resultados e discussão.

RESUMO

A epidemia de HIV/aids na região Sul do Brasil é distinta da observada em outras regiões do país. Principalmente nos estados de Santa Catarina (SC) e do Rio Grande do Sul (RS) as taxas de incidência anual de aids e de mortalidade pela doença são bem maiores que a média nacional. Além disso, na mesma região é encontrada grande prevalência de subtipo C enquanto nos outros estados brasileiros há predominância do subtipo B. Uma minuciosa revisão da bibliografia revelou que, apesar de vários estudos já terem descrito a epidemia molecular do subtipo C nas capitais do Sul do Brasil, uma investigação que abrangesse de forma representativa a população dessa região ainda fazia-se necessária. Em vista disso, o trabalho apresentado aqui estudou a epidemia do HIV em 13 municípios do interior dos estados de SC e do RS. Material biológico e dados epidemiológicos de aproximadamente 350 pacientes foram coletados e utilizados para montar um banco de dados com informações moleculares, clínicas e demográficas. Em praticamente todas as cidades amostradas, mas principalmente em SC, o HIV-1 subtipo C mostrou-se como forma predominante. No RS, um número maior de formas recombinantes foi observado com três possíveis novas formas circulantes recombinantes (CRF) identificadas. Análises filodinâmicas revelaram um crescimento mais rápido da epidemia do HIV-1 subtipo C em relação ao subtipo B, sendo esta diferença ainda mais pronunciada em SC, onde também observou-se uma clara segregação dos subtipos virais entre categorias de exposição. Os dados coletados aqui ainda foram utilizados em um estudo filogeográfico para descrever a expansão da epidemia do subtipo C pelo Brasil. Através da análise integrada de dados ecológicos e moleculares foi identificado que a origem dessa epidemia ocorreu em Porto Alegre e a partir desta cidade o subtipo C foi sendo progressivamente introduzido em cidades mais distantes e/ou isoladas. A prevalência de HIV-1 e o número de pessoas infectadas pelo subtipo C foram identificados como preditores da dispersão viral do Sul ao Norte do país. Em conclusão, os resultados apresentados aqui revelam o rápido crescimento que a epidemia do HIV-1 subtipo C teve em SC e RS; discute como a separação em diferentes grupos de exposição pode impactar na velocidade de expansão e também na formação de novas recombinantes; e ainda descreve as rotas de dispersão do subtipo C rumo ao norte do Brasil, alertando para o potencial

expansivo dessa epidemia em outras regiões, principalmente no Centro-Oeste.

Palavras-chave: Filogenia. Filogeografia. Filodinâmica. Epidemiologia Molecular. Aids. HIV. Subtipo C. Região Sul do Brasil.

ABSTRACT

The HIV/AIDS epidemic in Brazilian Southern region is distinct than that observed in other parts of the country. Mainly in the states of Santa Catarina (SC) and Rio Grande do Sul (RS) the annual aids incidence and mortality rates are far higher than the national average. Furthermore, in the same states HIV-1 epidemic is driven by subtype C, while in other Brazilian regions subtype B is more prevalent. A detailed review of the bibliography revealed that studies describing the molecular epidemiology in south Brazil were mainly performed in the capital cities and that a more comprehensive characterization of the epidemic was necessary. In view of this, the study presented here investigated the HIV epidemic in 13 countryside municipalities from SC and RS states. Blood samples and epidemiological data were collected and assembled in a databank with molecular, clinical and demographic information from around 350 HIV sero-positive individuals. HIV-1 subtype C was observed to be the most prevalent in virtually all sampled locations, with higher frequencies in SC. HIV recombinant forms were observed in higher frequencies in RS, where some possible new CRFs were also identified. Phylodynamic analyses revealed a faster epidemic growth rate for subtype C in comparison with subtype B. This difference was more pronounced in SC where was also observed subtype segregation in categories of exposure. In addition, data collected here was also used in a phylogeographical investigation of the subtype C dispersion throughout Brazil. By using a new approach which integrates ecological and molecular data in the phylogeographical reconstructions, Porto Alegre was found to be the origin of subtype C epidemic and the central hub of dispersion towards north Brazil. HIV prevalence and subtype C population size were identified as predictors of the viral diffusion. In conclusion, the results presented here reveal the fast epidemic growth that subtype C went through in RS and SC; discuss how viral segregation in separated exposure categories can impact in the epidemic growth rates and in the emergence of recombinant forms; and, finally, describe the northward dispersion of subtype C, highlighting the potential to increase in prevalence mainly in Central-West region.

Keywords: Phylogenetics. Phylogeography. Phylodynamics. AIDS. HIV. Subtype C. Brazilian Southern Region.

LISTA DE ABREVIATURAS E SIGLAS

Inglês

AI	<i>Association index</i>
aids	<i>Acquired immunodeficiency syndrome</i>
aLRT	<i>Approximate Likelihood ratio test</i>
BSSVS	<i>Bayesian stochastic search variable selection</i>
CRF	<i>Circulating recombinant form</i>
DBS	<i>Dried blood spot</i>
ESS	<i>Effective sample size</i>
GLM	<i>Generalized linear model</i>
GTR	<i>General time reversible</i>
HIV	<i>Human immunodeficiency virus</i>
HIV-1C_BR	<i>HIV-1 subtype C Brazilian sequences</i>
HPD	<i>High posterior density</i>
HSB	<i>Homens que fazem sexo com homens</i>
IDU	<i>Injection drug users</i>
MC	<i>Monophyletic clade</i>
MCC	<i>Maximum clade credibility</i>
MCMC	<i>Markov chain Monte Carlo</i>
MEP	<i>Measurable evolving population</i>
ML	<i>Maximum likelihood</i>
MP	<i>Maximum parsimony</i>
MSM	<i>Men who have sex with men</i>
N_e	<i>Effective population size</i>
NJ	<i>Neighbor joining</i>
PP	<i>Probabilidade posterior</i>
PS	<i>Parsimony score</i>
r	<i>Growth rate</i>
RT	<i>Reverse transcriptase ou transcriptase reversa</i>
SDRM	<i>Surveillance Drug Resistance Mutations</i>
SGR	<i>Second-generation recombinants</i>
SH-aLRT	<i>Shimodaira-Hasegawa aLRT</i>
SIV	<i>Simian immunodeficiency virus</i>
URF	<i>Unique recombinant form</i>
WHO	<i>World health organization</i>
λ	<i>Epidemic doubling time</i>

Português

ARV	Antirretroviral
HET	Heterossexual
PR	Paraná
RS	Rio Grande do Sul
SC	Santa Catarina
UDI	Usuários de drogas injetáveis

SUMÁRIO

CAPÍTULO 1 - Introdução geral	23
1.1 – <i>O vírus e a doença</i>	24
1.1.2 – A taxa evolutiva do HIV	27
1.2 – <i>Os métodos de inferência filogenética e filogeográfica</i>	28
1.2.1 – A importância da filogenia no estudo das epidemias virais	28
1.2.2 – A árvore filogenética	28
1.2.3 – Modelos evolutivos	30
1.2.4 – Reconstrução filogenética pelos métodos tradicionais	35
1.2.5 – Estatística Bayesiana aplicada à inferência filogenética	39
1.2.6 – Detecção de eventos de recombinação	41
1.2.7 – Inferência filogeográfica	43
1.3 – <i>A origem do HIV e epidemiologia molecular</i>	46
1.3.1 – Dos primatas ao ser humano	46
1.3.2 – A epidemiologia molecular do HIV-1 no mundo	49
1.4 – <i>A epidemia do HIV no Brasil e as peculiaridades da região Sul</i>	51
1.5 – <i>Objetivos</i>	55
1.6 – <i>Referências</i>	56
CAPÍTULO 2 - The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent	65
2.1 – <i>Abstract</i>	66
2.2 – <i>Introduction</i>	66
2.3 – <i>Origin of Brazilian HIV-1 C clade</i>	69
2.4 – <i>Recognition of the HIV-1 C epidemic in Rio Grande do Sul</i>	71
2.4.1 – Identification of a new CRF_BC in Rio Grande do Sul	72
2.5 – <i>The HIV-1 C epidemic in Santa Catarina and Paraná</i>	73
2.6 – <i>Temporal trends and spatial dynamics of HIV-1 C expansion and CRF31_BC epidemics in Southern Brazil</i>	74
2.7 – <i>HIV-1C expansion throughout the country</i>	76
2.8 – <i>HIV-1C in South American countries</i>	77
2.9 – <i>The subtype C epidemic and its relevance</i>	79
2.10 – <i>Concluding remarks</i>	82
2.11 – <i>Acknowledgments</i>	83
2.12 – <i>References</i>	83

CAPÍTULO 3 - Comprehensive characterization of the HIV-1 molecular epidemiology and demographic history in the Brazilian region most heavily affected by AIDS	97
3.1 - <i>Abstract</i>	98
3.2 - <i>Importance</i>	98
3.3 - <i>Introduction</i>	99
3.4 - <i>Materials and Methods</i>	100
3.4.1 - Patients and samples	100
3.4.2 - Subtype classification	101
3.4.3 - Bayesian phylodynamics analyses	101
3.4.4 - Statistical analysis	102
3.5 - <i>Results</i>	102
3.6 - Discussion	109
3.7 - <i>Acknowledgements</i>	112
3.8 - <i>References</i>	112
3.9 - <i>Supplementary Information</i>	118

CAPÍTULO 4 - The contribution of epidemiological predictors in unravelling the phylogeographic history of HIV-1 subtype C in Brazil	119
4.1 - <i>Abstract</i>	120
4.2 - <i>Introduction</i>	120
4.3 - <i>Methods</i>	122
4.3.1 - Patients, samples and new sequences	122
4.3.2 - Sequence dataset compilation	122
4.3.3 - Phylogenetic divergence time estimation and population dynamics inference	123
4.3.4 - Phylogeny-trait association	123
4.3.5 - Phylogeographic inference with epidemiological predictors	124
4.4 - <i>Results</i>	125
4.4.1 - Sequence dataset compilation	125
4.4.2 - Phylogeny-trait association	127
4.4.3 - Inconsistencies in root state estimates	127
4.4.4 - Predictors of viral spread	128
4.4.5 - Porto Alegre as a central hub of the HIV-1C epidemic	130
4.5 - <i>Discussion</i>	132
4.6 - <i>Acknowledgements</i>	134

<i>4.7 - References</i>	135
<i>4.8 - Supplementary Material</i>	139
CAPÍTULO 5 - Discussão Geral	147
<i>5.1 - Referências</i>	155
CAPÍTULO 6 - Conclusões	159
CAPÍTULO 7 - Perspectivas	161
Apêndices	163
1 - Outros trabalhos publicados:	163
<i>2 - Termo de consentimento Livre e Esclarecido</i>	164
<i>3 - Questionário utilizado na entrevista de recrutamento dos voluntários</i>	169
<i>4 - Relatório Sucinto</i>	171

CAPÍTULO 1.

Introdução geral

A síndrome da imunodeficiência adquirida (aids) é uma das doenças infecciosas de maior importância a afetar a espécie humana em toda sua história. Atualmente, as mortes causadas pela aids são mais numerosas que doenças como tuberculose ou malária (WHO, 2010). Os primeiros relatos de indivíduos com aids foram registrados no ano de 1981 em pessoas homossexuais que apresentavam doenças raras, como pneumonia pneumocística e candidíase muco-cutânea disseminada (Gottlieb et al., 1981). A causa de tais infecções era uma severa imunodeficiência de etiologia ainda desconhecida. Todavia em 1983, apenas dois anos após o registro dos primeiros casos desta síndrome inexplicável, foi isolado o agente causador da doença: o vírus da imunodeficiência humana (HIV) (Barre-Sinoussi et al., 1983; Gallo et al., 1984).

1.1 – O vírus e a doença

O HIV é um vírus de RNA que sintetiza um genoma intermediário de DNA em seu ciclo replicativo, sendo desta forma membro da família *Retroviridae*. Além disso, o HIV pertence ao gênero *Lentivirus* devido a seu longo período de incubação e lento desenvolvimento dos sintomas que compõem a doença que ele causa (Hoffmann et al., 2007). Existem dois tipos de HIV: HIV-1 e HIV-2. Ambos são morfologicamente semelhantes, mas o HIV-2 é menos patogênico e causa infecções restritas ao oeste da África, o HIV-1, por sua vez, é o grande causador da pandemia mundial de aids (Rambaut et al., 2004).

O genoma do HIV-1 é composto de duas fitas simples de RNA idênticas, cada uma contendo aproximadamente 10 mil nucleotídeos e nove genes. Os três genes principais, presentes em todos os retrovírus, são *env*, *pol* e *gag*. Esses genes codificam respectivamente as proteínas gp120 e gp41 presentes no envelope viral; as enzimas transcriptase reversa (RT), integrase e protease; e as proteínas estruturais que compõem o capsídeo viral (Hoffmann et al., 2007). Os outros seis genes são considerados acessórios e são denominados *vif*, *vpu*, *vpr*, *tat*, *rev* e *nef*. Suas funções variam desde a promoção da transcrição do DNA proviral e transporte do RNAm para o citoplasma, até evasão de mecanismos de resposta imune inata e adaptativa (Miller e Sarver, 1997; Cullen, 1998; Peter, 1998).

Os principais eventos do ciclo replicativo do HIV são resumidos pela Figura 1.1. Inicialmente as proteínas do envelope viral interagem com a molécula CD4 encontrada na superfície de células como linfócitos T, monócitos, macrófagos, células dendríticas e da micróglia. Após uma segunda interação com o coreceptor CXCR4 ou CCR5 ocorre a entrada do vírus na célula. Então, as enzimas virais promovem a transcrição reversa e a integração do DNA viral ao genoma celular. Sob a forma de um componente integral do genoma da célula hospedeira, o DNA proviral é transcrito utilizando a maquinaria celular. A partícula viral é montada contendo duas cópias do RNA viral e as enzimas RT, integrase e protease. Essa última age na clivagem das poliproteínas traduzidas a partir do genoma viral e na maturação do virion após o brotamento da célula, processo necessário para tornar o HIV infeccioso (Hoffmann et al., 2007).

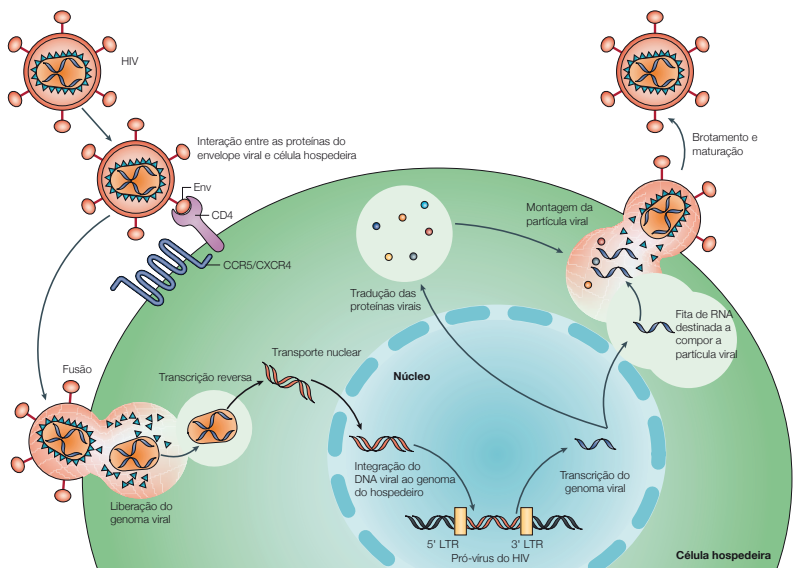


Figura 1.1 – Figura esquemática destacando os principais eventos do ciclo replicativo do HIV. Adaptado de Rambaut et al. (2004)

As principais células alvo de infecção pelo HIV são os linfócitos T CD4. Essas células são atacadas e destruídas principalmente nos

linfonodos, devido a sua grande concentração neste tecido e a um ambiente pró-inflamatório que promove a replicação do vírus (Peterlin e Trono, 2003). A infecção causada pelo HIV pode ser dividida em três fases: a fase aguda, que ocorre logo após a infecção, a fase crônica e a fase de aids (Figura 1.2) (Hoffmann et al., 2007). Durante a fase aguda há uma alta carga viral e sintomas semelhantes a gripe podem ser observados. Após algumas semanas, linfócitos T citotóxicos HIV-específicos são capazes de controlar a viremia (Mellors et al., 1995). A fase aguda termina quando ocorre a soroconversão, momento em que os níveis de anticorpos anti-HIV passam a ser detectáveis. O momento antes da soroconversão também é chamada janela imunológica.

A fase crônica é caracterizada por um período de equilíbrio entre replicação viral e resposta imune do hospedeiro, podendo durar muitos anos sem manifestações clínicas da doença. Nessa fase ocorre uma progressiva morte das células T CD4, culminando em uma acentuada imunodepressão do organismo. A fase de aids é então atingida quando o sistema imune não é mais capaz de proteger o indivíduo de infecções por patógenos oportunistas. Níveis de linfócitos T CD4 abaixo de 200 células/ μ l de sangue são um indicativo desse estado (Hoffmann et al., 2007).

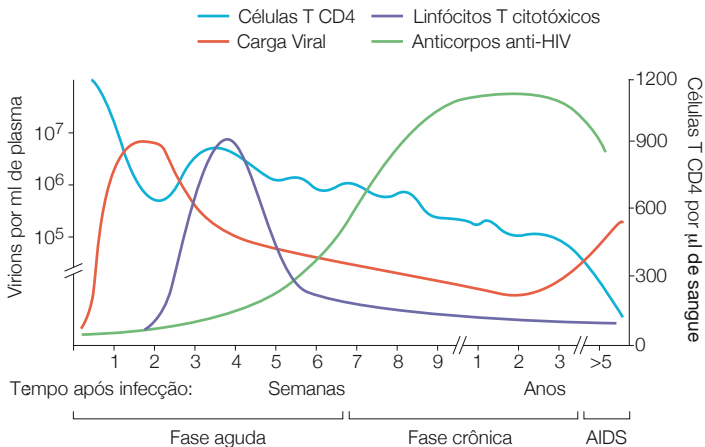


Figura 1.2 – História natural da infecção pelo HIV. Adaptado de Goulder et al. (2004).

1.1.2 - A taxa evolutiva do HIV

Uma característica marcante do HIV é sua grande variabilidade genética. Tal característica é devida a uma combinação de fatores que tornam o HIV um dos organismos conhecidos de mais rápida evolução (Rambaut et al., 2004). Primeiramente a ausência da atividade de correção de erros da enzima RT é responsável por uma alta taxa mutacional. A etapa de transcrição reversa do RNA viral gera em torno de 0,2 mutações por genoma por ciclo replicativo (Preston et al., 1988). Esta alta taxa mutacional do HIV, por sua vez, é potencializada por uma taxa replicativa de 10^{10} a 10^{12} novas partículas virais por dia (Perelson et al., 1996), levando o HIV a uma taxa evolutiva estimada em 2 a 4×10^{-3} substituições por sítio por ano (s/s/y) (Abecasis et al., 2009). Para se ter uma ideia da velocidade evolutiva do HIV, sua taxa é cerca de 2 milhões de vezes mais rápida que a taxa evolutiva do ser humano, que é aproximadamente 1×10^{-9} s/s/y (Scally e Durbin, 2012). Além disto, altas taxas de recombinação entre diferentes variantes do HIV também contribuem para aumentar a diversidade de formas virais. Eventos de recombinação ocorrem quando uma célula é coinfetada por duas partículas virais e no momento do empacotamento do material genético, fitas de RNA de origens distintas são inseridas no mesmo capsídeo, formado uma partícula viral híbrida (Vuilleumier et al., 2015). A recombinação ocorre porque a enzima RT pode trocar de fita de RNA molde durante a síntese do DNA. Se estas duas fitas são diferentes o DNA transcrito será um recombinante de ambas.

A rápida taxa evolutiva do HIV faz com que o sistema imune do hospedeiro seja incapaz de combater tantas variantes virais, torna o desenvolvimento de vacinas um desafio e permite o vírus adquirir resistência aos fármacos antirretrovirais. Esta mesma característica faz com que a evolução do HIV, assim como de outros vírus de RNA, seja mensurável em uma escala de tempo observável pelo ser humano, o que facilita em muito a reconstrução de sua história evolutiva (Grenfell et al., 2004). Isto, somado a riqueza de sequências nucleotídicas disponíveis em bancos de dados públicos na internet, tornam o HIV um excelente modelo de estudos filogenéticos, auxiliando também no desenvolvimento de novas ferramentas de bioinformática para análise de sequências genômicas.

1.2 – Os métodos de inferência filogenética e filogeográfica

1.2.1 – A importância da filogenia no estudo das epidemias virais

Filogenia é o estudo das relações evolutivas entre genes e organismos. O método clássico de estimar tais relações entre as espécies é através da comparação de estruturas morfológicas, permitindo também o estudo de registros fósseis de espécies já extintas (Freeman e Herron, 2009). Por sua vez, a filogenia molecular utiliza sequências nucleotídicas, aminoacídicas ou outros dados moleculares que tornaram-se extremamente abundantes a partir do advento da biologia molecular. O estudo da filogenia de vírus é possível apenas através da análise de dados moleculares, pois estes organismos não deixam registro fóssil. Desta forma, a única maneira de se estudar a história evolutiva dos vírus é através da análise filogenética das populações existentes, ou, em alguns casos, através do estudo de genomas virais antigos inseridos em espécies atuais (Lemey et al., 2009a).

A filogenia molecular se tornou uma ferramenta essencial na investigação de novos vírus, bem como em estudos de epidemiologia viral. Quando um novo vírus é descrito o estudo de suas relações filogenéticas com outras populações virais é aplicado para descrever sua origem e história de disseminação na espécie hospedeira. Esta abordagem filogenética da história evolutiva de vírus ou outros patógenos é também conhecida como epidemiologia molecular, disciplina com o objetivo de descrever os padrões e taxas de transmissão de patógenos através da análise comparativa de suas sequências nucleotídicas (Holmes, 1998). Ademais, a inferência filogenética também é útil no campo da virologia e epidemiologia através de sua aplicação no monitoramento do surgimento de formas resistentes a fármacos antivirais, levantamento da diversidade genética para desenvolvimento de vacinas e monitoramento do combate de epidemias virais por programas vacinais (McCormack e Clewley, 2002).

1.2.2 – A árvore filogenética

A principal tarefa de uma análise filogenética é reconstruir a árvore filogenética, um gráfico que descreve a relação descendente-ancestral dos organismos em investigação. Basicamente o método

filogenético considera que as sequências em análise são homólogas, ou seja, compartilham um ancestral comum. Embora, segundo Darwin, todos os organismos compartilhem um ancestral comum, homologia não é encontrada entre as sequências de qualquer organismo. Com o passar do tempo modificações, sejam elas mutações, inserções ou deleções, vão se acumulando no genoma até o ponto de duas sequências não possuírem mais informação suficiente para se inferir a relação filogenética entre elas. O termo homologia é usado apenas quando o ancestral comum é recente o suficiente para que suas sequências tenham uma taxa de similaridade que permita a inferência filogenética. Em geral, se as sequências nucleotídicas apresentarem menos de 60% de similaridade, a árvore filogenética construída a partir delas é de pouca valia (Lemey et al., 2009a).

Cada sequência nucleotídica representada em uma árvore filogenética é denominada táxon, sendo estes representados na parte mais externa dela e também chamados de nodos externos, terminais ou folhas da árvore (Figura 1.3). Os nodos internos representam os hipotéticos ancestrais dos táxons, que é uma informação não observada nos dados (as sequências sob análise) mas estimada a partir deles. Cada nodo da árvore é ligado por um ramo, sendo este de tamanho variado que pode representar o número de substituições nucleotídicas/aminoácídicas por sítio da região analisada ou uma escala de tempo. Árvores filogenéticas podem ser desenhadas de forma não enraizada, representando simplesmente a relação entre os táxons, ou de forma enraizada onde a história evolutiva é representada de forma direcionada a partir de um ancestral comum a todos os táxons (a raiz) para os nodos externos (Freeman e Herron, 2009).

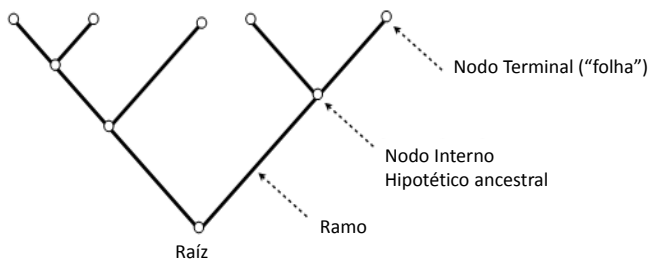


Figura 1.3 – Terminologia de uma árvore filogenética.

Uma etapa essencial no processo da inferência filogenética é o alinhamento múltiplo das sequências, onde o objetivo é posicionar na mesma coluna (alinhar) os caracteres homólogos das sequências. Este processo torna-se complexo ao considerarmos que inserções ou deleções (*indels*) possam ter ocorrido na história evolutiva dos táxons, sendo necessário então a inserção de *gaps* (regiões vazias no alinhamento) para algumas sequências (Lemey et al., 2009). Atualmente vários algoritmos de alinhamento estão disponíveis, sendo o Clustal W e o Muscle dois dos mais utilizados (Thompson et al., 1994; Edgar, 2004). Entretanto, estes programas não são infalíveis e uma inspeção manual do alinhamento é altamente recomendada.

1.2.3 – Modelos evolutivos

Uma infinidade de fatores pode interferir na forma com que os organismos evoluem. Uma inferência filogenética perfeita deveria refletir a ação destas forças evolutivas, entretanto, com os métodos disponíveis atualmente, podemos modelar apenas uma parte delas. Os modelos evolutivos, desta forma, visam promover a reconstrução de árvores filogenéticas da forma mais realística possível. A escolha do modelo mais apropriado para a análise é o passo seguinte ao alinhamento das sequências e é de grande importância para a acurácia da árvore filogenética reconstruída. Atualmente um modelo evolutivo pode ser composto de um modelo de substituição nucleotídica ou aminoacídica, um modelo de relógio molecular, um modelo de coalescência e um modelo filogeográfico de difusão espacial (descrito em mais detalhes na seção 1.2.7).

Os modelos de substituição nucleotídica descrevem o processo de mudança dos nucleotídeos de uma sequência no tempo (Whelan, 2001). O número de diferenças observadas entre duas sequências nucleotídicas é geralmente menor que o número de substituições que ocorreram durante o processo evolutivo. Múltiplas substituições no mesmo sítio e mutações de reversão podem fazer com que a distância genética calculada diretamente pelo número de diferenças entre duas sequências subestime a real distância genética entre elas. Os modelos de substituição nucleotídica tem por objetivo corrigir, ou pelo menos diminuir, o efeito das mudanças não observáveis. Diferentes modelos de substituição fazem diferentes suposições sobre a probabilidade de um

nucleotídeo ser substituído por outro e sobre a frequência com que ocorrem. Por exemplo, alguns modelos consideram distintas probabilidades de ocorrerem transições - mudanças entre nucleotídeos de mesma classe - e transversões - trocas entre nucleotídeos de classes diferentes. Apesar do grande número de modelos de substituição disponíveis, análises filogenéticas do HIV utilizam frequentemente o modelo GTR (Tavaré, 1986). Esse considera que cada nucleotídeo pode ter uma frequência e probabilidade de mudança próprio, sendo essa última simétrica, ou seja, a probabilidade de troca $A \rightarrow G$ é igual a troca $G \rightarrow A$ (Figura 1.4). Além disso, os modelos podem ainda considerar que a taxa de mudança é heterogênea ao longo dos sítios da sequência e também que uma porção dos sítios é invariável (Yang, 1993).

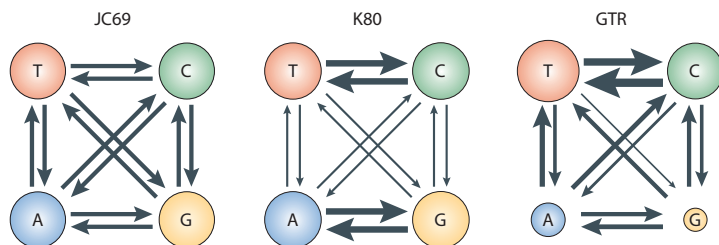


Figura 1.4 - Desenho esquemático de três diferentes modelos de substituição nucleotídica. A espessura das setas indica as taxas de troca entre os quatro nucleotídeos e o tamanho dos círculos representa suas frequências. O modelo JC69 considera a mesma probabilidade de substituição para todos nucleotídeos e frequências iguais. O modelo K80 considera maior probabilidade de ocorrerem transições e também considera frequências iguais. O modelo GTR assume que cada par de nucleotídeos possui uma probabilidade própria de trocar entre si e que as frequências podem ser diferentes, o que por sua vez também afeta a taxa de mudança entre os nucleotídeos. Adaptado de Yang (2012).

A hipótese do relógio molecular foi postulada em meados da década de 60 após estudos demonstrarem que a distância genética entre genes ortólogos era proporcional ao tempo de divergência entre as espécies (Zuckermandl e Pauling, 1962 e 1965). Este padrão foi observado em vários genes codificantes de proteínas em diversas espécies, levando à conclusão que a taxa evolutiva de qualquer proteína é

constante através do tempo. Tais estudos, alguns anos mais tarde, serviram de base para Kimura desenvolver a Teoria Neutralista da Evolução Molecular, propondo que a grande maioria das substituições nucleotídicas fixadas nas populações são neutras e não positivamente selecionadas (Kimura, 1968 e 1983). Seguindo esta linha de raciocínio a evolução molecular seguiria um relógio constante, também chamado de restrito, e desvios desse padrão aconteceriam apenas em eventos de adaptação, mudanças rápidas no tamanho efetivo da população ou relaxamento de barreiras funcionais em proteínas. Segundo Kimura, tais eventos seriam raros na história evolutiva das espécies.

Considerando-se a existência de um relógio molecular, a taxa evolutiva de qualquer gene pode ser calculada e isto pode ser usado para datar o tempo de divergência de duas espécies, ou seja, o tempo do ancestral comum mais recente delas. Da mesma forma, informações temporais de registros fósseis podem ser usados na inferência da taxa evolutiva dos genes, ajudando a “calibrar” o relógio molecular. Alternativamente, também pode-se aumentar a precisão da inferência da taxa evolutiva de um gene através da utilização de sequências isoladas em diferentes momentos no tempo, compondo um banco de sequências chamado heterocrônico. Quando há significativo acúmulo de mudanças genéticas entre sequências de uma mesma espécie amostradas em um dado período de tempo, tal população é considerada como de evolução mensurável no tempo (MEP – *Measurable evolving population*) (Drummond et al., 2003). Devido a suas altas taxas mutacionais, os vírus de RNA estão entre as principais MEPs estudadas atualmente.

O conceito de um relógio molecular é central no estudo de MEPs, pois ele conecta a informação temporal contida na amostragem heterocrônica à distância genética inferida a partir das sequências. Com base nestes conceitos, foi possível demonstrar que o modelo de um relógio molecular restrito é uma simplificação que não pode ser aplicada a todos os organismos. A análise de bancos de sequências heterocrônicas de 50 diferentes vírus de RNA demonstrou que apenas sete deles não rejeitam o modelo de relógio restrito (Jenkins et al., 2002). O modelo de relógio molecular relaxado foi então desenvolvido para acomodar taxas evolutivas variáveis no tempo (Drummond et al., 2006). Este modelo assume que não existe nenhuma correlação entre as taxas evolutivas dos ramos adjacentes de uma árvore filogenética, o que permite a mudança deste parâmetro ao longo da história evolutiva do organismo.

Entre os fatores que podem alterar a taxa evolutiva de um organismo estão o tamanho efetivo da sua população (N_e) e as alterações desta no tempo. Por definição, N_e é proporcional às alterações na diversidade genética da espécie, podendo ser estimada a partir de informações moleculares (Lemey et al., 2009a). A história demográfica de uma população pode ser reconstruída através dos modelos de coalescência, introduzido por Kingman em 1982. A teoria da coalescência descreve a relação entre a história demográfica de uma população e a ancestralidade compartilhada entre os indivíduos amostrados desta população e representados em uma árvore filogenética (Figura 1.5). O tempo de coalescência dos ramos da árvore, conforme regressa-se no tempo, traz informações sobre o tamanho da população e se este sofreu alterações (Griffiths e Tavaré, 1994). Os modelos de coalescência atuais são aplicáveis em MEPs (Rodrigo e Felsenstein, 1999) e podem inferir a história demográfica utilizando um parâmetro de crescimento demográfico (constante, exponencial, logístico, etc) ou através de modelos não paramétricos que aceitam mudanças no padrão de crescimento através do tempo (skyline, skyride ou skygrid) (Ho e Shapiro, 2011; Gill et al., 2013). Os modelos de coalescência são de grande valia na epidemiologia molecular, pois a partir deles pode-se descrever a história demográfica de epidemias passadas, estimando também o número de indivíduos afetados, ou inferir a taxa de expansão de epidemias enquanto elas acontecem (no caso de vírus de RNA), possibilitando que medidas de contenção sejam tomadas (Rambaut et al., 2008; Pybus et al., 2001; Robbins et al., 2003).

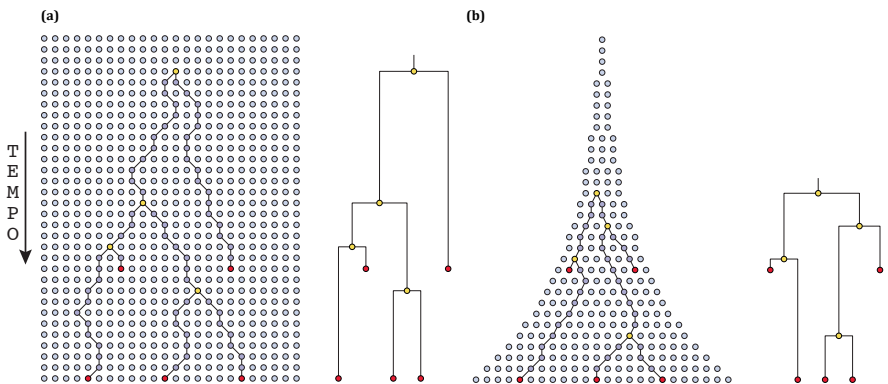


Figura 1.5 – Desenho esquemático representando o processo de coalescência em duas populações. Os pontos vermelhos representam indivíduos amostrados em dois momentos de tempo (heterocrônicos), os pontos amarelos representam o ancestral comum entre eles e os pontos cinzas representam indivíduos da população não amostrados. Os dois quadros representam as relações filogenéticas entre indivíduos amostrados de uma população de tamanho constante (a) e de crescimento exponencial (b). Regressando-se no tempo a partir das amostras mais recentes é possível observar que o tempo de um evento de coalescência ocorrer é inversamente proporcional ao tamanho da população naquele momento. Esta relação é refletida no tamanho dos ramos da árvore filogenética e pode ser usada para estimar a história demográfica da população. Adaptado de Drummond et al. (2003).

A escolha dos modelos de substituição nucleotídica, de relógio molecular e coalescência podem ter efeitos dramáticos na topologia da árvore filogenética, afetando o tamanho dos ramos e também seu arranjo. Estes três modelos compõem a base das suposições assumidas sobre o processo evolutivo e tornam o complexo problema computacional da reconstrução filogenética em uma tarefa exequível. Desta forma, procedimentos que comparem modelos alternativos e forneçam um critério objetivo de escolha do melhor modelo devem ser aplicados antes ou durante a análise filogenética (Sullivan e Joyce, 2005). Testes como a Razão de Verossimilhança e Critério Informativo Bayesiano ou de Akaike podem ser aplicados na abordagem de reconstrução filogenética por Máxima Verossimilhança (ML – *Maximum Likelihood*) e analisam o quanto o modelo se adapta aos dados (o alinhamento no caso da inferência filogenética) (Posada, 2008). Alternativamente, na abordagem Bayesiana a adaptação dos modelos aos dados é analisada através do cálculo da Verossimilhança Marginal integrada sobre todos os valores dos parâmetros e ponderado sobre a distribuição dos *priors* dos modelos (Baele et al., 2012). A subtração da Verossimilhança Marginal de dois modelos gera o Fator Bayes, uma medida de suporte estatístico sobre a melhor adequação de um modelo em comparação a outro (Suchard et al., 2001). É importante ressaltar que o melhor modelo não é aquele que descreve com perfeição os dados, mas sim o que consegue explicar os dados apresentando um balanço entre simplicidade e flexibilidade.

1.2.4 – Reconstrução filogenética pelos métodos tradicionais

A tarefa de encontrar a árvore filogenética que melhor descreve os dados é árdua. O número de possíveis topologias cresce exponencialmente com o número de táxons incluídos, sendo três quando quatro táxons são analisados e chegando a 2×10^6 topologias em uma análise de 10 táxons (Felsenstein, 1978). Isto torna a busca exaustiva, onde todas as possíveis árvores são investigadas, uma tarefa praticamente impossível de ser realizada para análises com algumas dezenas de táxons. Nestes casos são utilizados métodos heurísticos de procura. Estes algoritmos geralmente geram uma árvore inicial aleatória e realizam pequenos rearranjos locais dos clados na tentativa de melhorar a árvore na perspectiva do método de inferência utilizado (Lemey et al., 2009a). Os métodos de busca heurística não dão garantia de encontrar a melhor árvore, entretanto, eles tornam possível análises filogenéticas para um grande número de táxons.

Os métodos de inferência filogenética podem ser classificados pelo tipo de dados que eles utilizam: caracteres discretos (geralmente nucleotídeos ou aminoácidos) ou uma matriz contendo a distância genética das sequências (Tabela 1). Os métodos de distância utilizam o alinhamento de sequências para calcular a distância genética entre todas elas par a par, sendo apenas estes valores considerados na reconstrução da árvore. Por sua vez, os métodos baseados em caracteres comparam simultaneamente todas as sequências do alinhamento considerando um sítio - uma coluna no alinhamento - por vez e atribuindo um escore para aquela árvore. Esse escore é conhecido também como critério-ótimo e é utilizado para comparar todas as árvores e encontrar a melhor. Com exceção da Máxima Parcimônia, todos os métodos utilizam um modelo de substituição nucleotídica, seja para o cálculo da distância genética ou para estimar as probabilidades de troca entre os caracteres (Yang e Rannala, 2012).

Tabela 1: Principais métodos de inferência filogenética e o software correspondente para sua utilização

Tipo de dado	Método	Software
Matriz de Distância	Neighbor-Joining	PAUP*, MEGA, PHYLIP, Seaview
	Minimum Evolution	PAUP*, MEGA, PHYLIP
	UPGMA	PAUP*, MEGA, PHYLIP
Caracteres	Máxima Parcimônia	PAUP*, MEGA, PHYLIP, Seaview
	Máxima Verossimilhança	PAUP*, MEGA, PHYLIP, Seaview, PAML
	Estatística Bayesiana	BEAST, MrBayes

A metodologia mais simples de inferência filogenética é a Máxima Parcimônia (MP). Ela foi originalmente desenvolvida para ser usada em caracteres morfológicos sendo posteriormente aplicada na filogenia molecular (Holder e Lewis, 2003). A premissa da MP é simples: a melhor árvore é aquela que conta a história evolutiva dos táxons através do mínimo número de substituições de nucleotídeos. Para tanto, caracteres são designados aos nodos internos e o número de mudanças para atingir aquele estado a partir dos nodos externos é contado por um algoritmo (Fitch, 1971). Acoplado a isso, um algoritmo heurístico vasculha o espaço das possíveis árvores à procura daquela que minimiza o número de mudanças nucleotídicas, ou seja, a árvore de máxima parcimônia. Apesar de sua simplicidade e rapidez, este método possui sérios problemas uma vez que ele não considera nenhum modelo de substituição nucleotídica. Isso leva a uma subestimativa do tamanho dos ramos da árvore pois a distância genética entre os táxons não é corrigida para possíveis múltiplas substituições no mesmo sítio (ver seção 1.2.3). Essa deficiência, somada a incapacidade da MP de lidar com medidas de incerteza filogenética e de incorporar informações sobre a data da amostragem, tem levando ao desuso deste método.

Os métodos de distância são muito utilizados pela sua rapidez e resultados de boa acurácia, sendo o mais popular dentre eles o método de *Neighbor Joining* (NJ) (Saitou e Nei, 1987). A inferência por NJ parte de uma árvore inicial que contém todos os táxons mas nenhuma estrutura, ou seja, uma árvore em forma de estrela onde todos os ramos possuem um mesmo ancestral comum. Então um algoritmo de aglomeração junta no mesmo clado as sequências com menor distância genética de modo a minimizar a distância estimada da árvore. No passo

seguinte os dois táxons juntados em um clado são substituídos pelo ramo de seu ancestral comum e a mesma avaliação é realizada para juntar os próximos dois táxons com menor distância. Esse procedimento é mantido até que toda a árvore esteja resolvida, ou seja, cada nodo interno dá origem a apenas dois ramos. A principal desvantagem do NJ é na análise de sequências muito divergentes, geralmente subestimando sua distância. Isto ocorre pois métodos de distância perdem muitas informações ao transformar o alinhamento em uma simples matriz numérica. Além disso, métodos de distância são sensíveis a *gaps* no alinhamento (Yang e Rannala, 2012).

O conceito de verossimilhança está associado à probabilidade de uma hipótese ser verdadeira (verossímil). Na filogenia, a verossimilhança é a medida de quanto os dados concordam com a hipótese de uma árvore construída sob as predições de um modelo evolutivo. O método da Máxima Verossimilhança (ML) é uma forma de encontrar qual topologia da árvore, tamanhos dos ramos e parâmetros do modelo evolutivo (ex.: taxa de transição e transversão, frequência dos nucleotídeos, heterogeneidade da taxa de evolutiva entre os sítios, etc.) são capazes de maximizar a verossimilhança das sequências do alinhamento (Lemey et al., 2009a). Em outras palavras, a função de verossimilhança é a probabilidade condicional dos dados (alinhamento) dada uma hipótese (o modelo evolutivo e seus parâmetros θ , e a árvore τ com os específicos tamanhos de ramos):

$$L(\theta, \tau) = P(\text{Dados} \mid \theta, \tau)$$

Onde L é a verossimilhança da hipótese e P é a probabilidade condicional dos dados dada a hipótese.

A busca pela árvore de ML pode iniciar-se com uma árvore gerada por outro método (NJ por exemplo) e envolve a otimização do tamanho dos ramos e da topologia por métodos heurísticos. A cada otimização um novo valor de verossimilhança é designado à árvore e se este for melhor que o da árvore antiga ela mantém-se. Esta busca estende-se até que nenhuma modificação seja capaz de melhorar a verossimilhança da árvore.

Uma das principais vantagens do método de ML é o grande repertório de modelos evolutivos que podem ser utilizados e testados,

incluindo modelos demográficos e de relógio molecular. Através do Teste de Razão de Verossimilhança também é possível identificar qual o modelo que melhor se encaixa nas sequências amostradas, levando a um entendimento mais detalhado do processo evolutivo. O custo de tamanha versatilidade é um grande peso computacional, sendo esta a principal desvantagem da ML. Entretanto, desde o primeiro algoritmo para análise de ML em um contexto filogenético, desenvolvido por Felsenstein em 1981, avanços nos métodos de busca heurístico (Guindon et al., 2010) e do poder computacional das máquinas modernas tem diminuído o tempo de análise.

Nos métodos de inferência filogenética discutidos até aqui, a melhor árvore é estimada mas não há nenhum suporte estatístico que mostre o quão forte é a relação dos táxons agrupados nos clados. Esta informação é geralmente inferida através de um método chamado *bootstrapping* (Felsenstein, 1985). Neste método colunas do alinhamento de sequências são escolhidas randomicamente e sucessivamente adicionadas a um novo alinhamento, chamado de pseudorréplica, até o mesmo ter o mesmo tamanho do alinhamento original. Então, o método de inferência filogenética é aplicado aos novos alinhamentos (geralmente 1000) e a frequência com que os ramos agrupam-se em clados é anotada à árvore estimada a partir do alinhamento original. Para cada clado é designado um valor de *bootstrap* indicando em quais partes da árvore pode-se ter mais confiança quanto as relações evolutivas reconstruídas. O método de *bootstrapping*, porém, demanda um grande poder computacional, pois a melhor árvore filogenética terá de ser inferida para cada pseudorréplica. Nos métodos de distância ou MP isso pode não ser um grande problema, mas para uma análise de ML com um número razoável de táxons, calcular o *bootstrap* dos clados pode ser proibitivo. Novos métodos propostos recentemente são capazes de avaliar com rapidez o suporte dos clados em árvores de ML, permitindo a análise de várias centenas de sequências. Entre eles estão o Teste Aproximado da Razão de Verossimilhança (aLRT) (Anisimova e Gascuel, 2006) e o Shimodaira-Hasegawa aLRT (SH-aLRT) (Guindon et al., 2010).

1.2.5 – Estatística Bayesiana aplicada à inferência filogenética

Apesar da inferência Bayesiana preceder a estatística frequentista em 150 anos, apenas no final da década de 1990 ela foi introduzida na filogenia molecular (Rannala e Yang, 1997; Mau e Newton, 1997; Li et al., 2000). Em 1763 foi publicado postumamente o trabalho mais importante de Thomas Bayes: a formalização de um método para inferir a probabilidade de um evento incorporando-se evidências prévias (probabilidades a priori ou simplesmente *prior*) sobre este evento. Isso implica que para cada evento (hipótese) que se quer inferir a probabilidade de ocorrência, um *prior* deve estar associado (Brown e Kilmer, 2003). Na estatística Bayesiana a probabilidade dos eventos ocorrerem após analisar-se os dados e considerar os *priors* é chamada de probabilidade posterior (PP). A inferência Bayesiana é baseada no Teorema de Bayes:

$$P(\theta, \tau | \text{Dados}) = \frac{P(\theta, \tau) \times P(\text{Dados} | \theta, \tau)}{P(\text{Dados})}$$

Onde $P(\theta, \tau)$ é probabilidade a priori dos parâmetros e da árvore, $P(\text{Dados} | \theta, \tau)$ é a verossimilhança dos dados condicionado aos parâmetros e a árvore e $P(\theta, \tau | \text{Dados})$ é a PP. O denominador $P(\text{Dados})$ (também chamado de verossimilhança marginal) é a probabilidade de todas as hipóteses alternativas ocorrerem, ou seja, de todas as possíveis árvores e valores de parâmetros, o que torna impraticável o cálculo direto da PP de uma árvore filogenética.

Para tornar a inferência filogenética Bayesiana possível, um procedimento estocástico de amostragem é utilizado: o algoritmo de Markov Chain Monte Carlo (MCMC). Este algoritmo propõe mudanças aleatórias na topologia τ e valores dos parâmetros θ , sendo que cada proposta é um novo estado da cadeia representado por uma θ e τ específicos. A cada nova proposta é calculada a razão da PP do novo estado sobre o antigo. Se PP do novo estado é maior, ele é aceito imediatamente; e se ele for menor é aceito com uma probabilidade proporcional a razão das probabilidades posteriores. O cálculo da razão da PP faz com que não seja necessário o cálculo de $P(\text{Dados})$ pois são duas constantes que se cancelam na equação (Yang e Rannala, 2012).

Outra propriedade importante da utilização da MCMC é que regiões de menor probabilidade também são “visitadas”, o que pode ser um passo importante para atingir-se regiões de alta probabilidade que localizam-se após um vale no espaço da distribuição dos parâmetros (Figura 1.6).

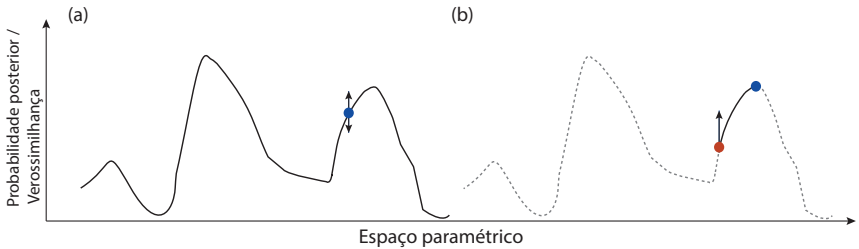


Figura 1.6 – Desenho esquemático contrastando a estratégia de estimativa dos parâmetros da inferência Bayesiana e de Máxima Verossimilhança (ML). As paisagens desenhadas representam o espaço paramétrico, destacando vales de baixa probabilidade e picos de alta probabilidade. (a) Na inferência Bayesiana a MCMC (Markov Chain Monte Carlo) - círculo azul - possui liberdade para se mover por regiões de menor probabilidade posterior (PP), conseguindo assim atravessar vales de baixa PP. Seu objetivo é descrever o espaço paramétrico buscando as regiões de mais alta PP. (b) Na inferência de ML é buscado apenas parâmetros de maior verossimilhança, assim uma pequena porção do espaço paramétrico é visitado e não é possível atravessar vales de baixa probabilidade. O círculo vermelho representa a região de início da busca e o círculo azul o conjunto de parâmetros de máxima verossimilhança encontrado.

Uma grande vantagem da inferência Bayesiana é a acomodação da incerteza filogenética em suas análises. Isto é possível pois o objetivo final da inferência Bayesiana com MCMC não é encontrar os valores dos parâmetros que maximizem a verossimilhança dos dados, mas sim descrever a distribuição das PP no espaço paramétrico (Brown e Kilmer, 2003). A MCMC pode ser mantida por um tempo indeterminado, mas ela é geralmente interrompida quando atinge um platô de alta PP e passa a vagar por essa região, neste momento a cadeia é classificada como estacionária. As árvores e valores dos parâmetros coletados pela MCMC na fase estacionária são então usados para representar a distribuição de mais alta probabilidade posterior (HPD – *High posterior*

density) sendo possível a partir disso calcular a média e intervalos de confiança para cada parâmetro inferido.

Os métodos Bayesianos são hoje o estado da arte da inferência filogenética. Seu principal software de aplicação, BEAST, permite a utilização de uma enorme gama de modelos evolutivos, como modelos de substituição nucleotídica, demográficos, relógios moleculares e até mesmo modelos de difusão filogeográfica (Drummond et al., 2012). O grande potencial desta ferramenta é muito bem utilizado no estudo de MEPs e inúmeros estudos recentes descrevendo a epidemiologia molecular, filodinâmica e filogeografia de patógenos virais tem aplicado a inferência filogenética Bayesianiana através do BEAST.

1.2.6 – Detecção de eventos de recombinação

A evolução dos organismos não é apenas impulsionada pelo acúmulo de mutações randômicas através das gerações (evolução vertical). A transferência de material genômico e sua mistura entre indivíduos de linhagens diferentes (evolução horizontal) possui um papel importante na progressiva diferenciação entre os indivíduos, especialmente entre os vírus. Apesar dos avanços das técnicas de inferência filogenética nos últimos anos, a representação da evolução em forma de árvore não é um bom método de se capturar sinais de evolução horizontal, fazendo-se necessária a utilização de métodos alternativos para a identificação de formas recombinantes.

Entretanto para a detecção de recombinantes ser possível as sequências parentais que originaram a forma recombinante precisam ser divergentes o suficiente para poder-se identificar quem é o ancestral de cada fragmento recombinante atuais (LEMEY et al., 2009a). Esta condição é facilmente atingida por eventos de recombinação entre diferentes subtipos do HIV-1 grupo M (recombinação intersubtipo). Por outro lado, quando recombinações intrasubtipo ocorrem, a divergência das formas parentais é tão baixa que é muito difícil sua detecção. Um segundo obstáculo para a detecção de eventos de recombinação são as mutações de ponto que podem se acumular na região recombinada. Se o evento tiver ocorrido há muito tempo, pode não ser mais possível identificar-se a linhagem parental.

De forma geral os diferentes métodos de detecção de sinal de recombinação envolvem passos similares. Todos eles particionam o

alinhamento em porções sobrepostas chamadas janelas e a relação evolutiva entre as sequências é avaliada separadamente para cada janela de uma forma progressiva até cobrir todo o alinhamento. As diferenças entre os métodos encontra-se principalmente na forma de medir a relação entre as sequências e o suporte estatístico que se aplica a essa medida.

O método mais popular de detecção de recombinação é o *bootscan*. Este método constrói árvores de NJ para cada janela do alinhamento para avaliar as relações evolutivas da sequência suspeita de ser recombinante com sequências referências. O suporte dos clados é calculado por *bootstrap* e quando um evento de recombinação está presente, o padrão do *bootstrap* muda, significando que a sequência alvo pulou de um clado agrupado com uma sequência parental para outro (Figura 1.7). O método de *bootscan* é rápido e apresenta resultados confiáveis, permitindo a realização de varreduras em grandes alinhamentos. Atualmente ele está implementado nos programas Simplot (Lole et al., 1999) e RDP3 (Martin et al., 2010).

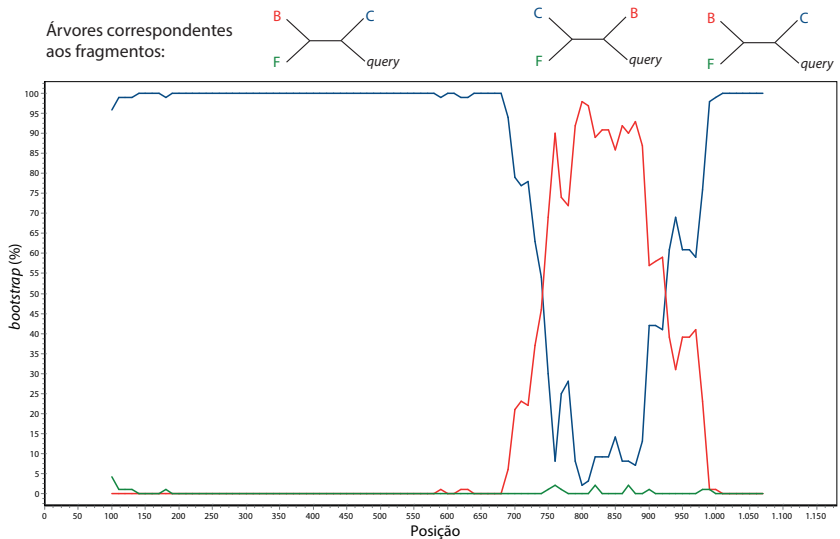


Figura 1.7 – Análise de *bootscan* de uma sequência recombinante. O gráfico mostra o suporte de *bootstrap* para o agrupamento da sequência em investigação (*query*) com uma das três possíveis sequências parentais

ao longo do alinhamento. Na parte superior do gráfico estão resumidas as árvores que representam com maior suporte a relação filogenética das sequências nas diferentes regiões do alinhamento.

1.2.7 – Inferência filogeográfica

Interpretado literalmente, o termo filogeografia significa análise filogenética de organismos considerando o contexto geográfico de sua distribuição. Embebida nesta definição está a possibilidade de se testar hipóteses sobre como os fenômenos geográficos - ou a compartimentalização em diferentes tecidos do hospedeiro no caso dos vírus - podem impactar na separação dos organismos e no processo de especiação. De uma forma mais geral, a filogeografia pode então ser definida como a extrapolação de processos microevolutivos ocorrendo dentro dos organismos para explicar diferenças macroevolutivas observadas (Avice et al., 1987).

O processo de expansão das epidemias, disseminação por diferentes hospedeiros e adaptação a novos sistemas imunes acaba moldando as filogenias virais, deixando “pistas” que podem ser mensuradas. O impacto das dinâmicas de transmissão e seleção sobre a variação genética dos vírus pode, dessa forma, ser usada para estudar-se a dinâmica de dispersão das epidemias, a dinâmica espaço-temporal das populações virais, transmissões zoonóticas, tropismo celular e deriva antigênica. À investigação de todos estes processos dá-se o nome de filodinâmica viral e dentro dela, os métodos filogeográficos podem ser aplicados na investigação do processo de difusão das epidemias (Grenfell et al., 2004).

Um vez obtida a árvore filogenética que melhor representa as relações evolutivas entre os táxons, a estruturação geográfica pode ser investigada como se fosse um traço herdado do ancestral (Figura 1.8). Pode-se então estimar quais são os traços - localizações geográficas neste caso - presentes nos nodos ancestrais da árvore condicionados às observações das sequências amostradas. A estruturação geográfica está presente na árvore quando o número de transições entre as localidades é menor que em uma população panmítica. Esta hipótese pode ser testada através da randomização dos traços observados nos nodos externos e comparando-se o número de transições com a árvore original. Esta

randomização pode ser repetida múltiplas vezes para obter-se uma distribuição nula e a partir disso é possível calcular um valor de p para a diferença entre as contagens (Slatkin e Maddison, 1989). A inferência de traços ancestrais pode ser feita através de parcimônia (Fitch, 1971), Máxima Verossimilhança (Cunningham et al., 1998) ou estatística Bayesiana (Lemey et al., 2009b). As vantagens e desvantagens de cada método são semelhantes às observadas para a inferência filogenética, portanto será aqui abordada apenas o método Bayesiano.

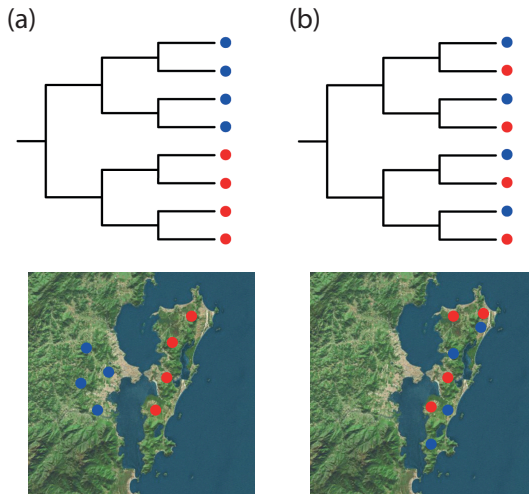


Figura 1.8 – Efeito da estruturação geográfica sobre a filogenia de uma população. a) Árvore filogenética de duas populações isoladas geograficamente, sendo perceptível o impacto do isolamento na estruturação da árvore. b) Árvore filogenética de duas populações cocirculando na mesma região geográfica.

Na inferência filogeográfica Bayesiana, a localização geográfica de cada táxon é tratada como um caractere, da mesma forma que os nucleotídeos. Um modelo de difusão espacial guia o processo de substituição desses caracteres, o que neste caso significa um evento de migração entre uma localidade e outra. Ao fim da reconstrução filogeográfica será inferido a cada localidade uma taxa de transição, relativo ao número de migrações das linhagens entre cada localidade ao longo da árvore. Os modelos de difusão podem especificar taxas de

transição simétricas ($A \rightarrow B = B \rightarrow A$) ou assimétricas ($A \rightarrow B \neq B \rightarrow A$) e ainda aplicar um método estocástico de procura e escolha de variáveis (BSSVS) que mantem no modelo apenas as transições de maior relevância (Lemey et al., 2009b). Através do BSSVS é possível calcular o Fator de Bayes das transições e identificar quais possuem maior suporte estatístico.

Os modelos de difusão descritos acima compõem a filogeografia em espaço discreto, onde as localidades ancestrais estimadas pertencem necessariamente ao grupo de localidades anotadas aos táxons. Alternativamente existem os modelos de difusão contínua, capazes de reconstruir a dispersão geográfica de forma mais realística, sobretudo quando as amostras foram coletadas de maneira difusa no espaço e anotadas com dados de longitude e latitude (Lemey et al., 2010). A difusão no espaço é guiada por diversos modelos, podendo ser randômico como o Movimento Browniano ou obedecendo a padrões de movimento animal (Bloomquist et al., 2010). Os modelos de difusão contínua são mais aplicáveis à filogeografia de epidemias em populações animais não-humanas ou de plantas, enquanto que para patógenos humanos os modelos discretos funcionam melhor (Lemey et al., 2010).

O rápido desenvolvimento de novas técnicas de análise filogenética e filogeográfica tem contribuído muito para a caracterização das epidemias virais que afetam a população humana nos dias atuais, sejam elas um problema antigo como o influenza, ou vírus emergentes como o chikungunya (Bedford et al., 2010; Lo Presti et al., 2012). As novas técnicas de sequenciamento também têm fomentado os avanços no campo da filodinâmica, atuando em conjunto com a bioinformática para fornecer respostas rápidas a surtos epidêmicos correntes, como no caso da epidemia de Ebola na África ocidental em 2014 (Dudas e Rambaut, 2014). Além disso, métodos recentes tem permitido a utilização de dados externos à filogenia (ecológicos, geográficos ou demográficos) para serem avaliados como preditores do processo de difusão viral pelo espaço. Usando esta abordagem foi possível demonstrar a distância de estradas como preditor da dispersão do vírus rábico em cães no norte da África (Talbi et al., 2010) e o fluxo aéreo de passageiros determinando a dispersão global do influenza H3N2 (Lemey et al., 2014).

Dentre os vírus, o HIV talvez seja o patógeno mais estudado. A seção 1.3 apresenta em detalhes a história evolutiva do HIV como contada pelos métodos descritos até aqui.

1.3 – A origem do HIV e epidemiologia molecular

1.3.1 – Dos primatas ao ser humano

A descoberta do vírus da imunodeficiência símia (SIV) permitiu aos pesquisadores entenderem as origens do HIV. Assim como o HIV, o SIV é um lentivirus da família *Retroviridae*. Sua epidemiologia molecular foi detalhadamente descrita através de métodos não invasivos capazes de recuperar o material genômico viral presente em fezes e urina de primatas (Santiago et al., 2003). Mais de 40 espécies de primatas africanos foram identificadas como portadoras de alguma variante do SIV, sendo a grande maioria destas linhagens virais espécie-específicas (Klatt et al., 2012). Dentre os vários lentivirus de primatas, o SIV encontrado em chimpanzés *Pan troglodytes* (SIVcpz) foi observado como sendo a linhagem mais proximamente relacionada ao HIV-1, enquanto o SIV dos macacos verde africanos *Cercocebus atys* (SIVsmm) é mais semelhante ao HIV-2 (Sharp et al., 2011). A infecção de seres humanos pelo SIV provavelmente ocorreu em sucessivos eventos de exposição ao sangue de primatas não-humanos durante a caça e manipulação da carne para consumo (Peeters et al., 2002).

O HIV-1 é composto por quatro diferentes grupos, denominados grupo M, N, O e P, cada um deles originado de um evento separado de transmissão do macaco para o ser humano (Hemelaar, 2013). Estudos filogenéticos demonstraram que o grupo M e N foram originados a partir de uma passagem do SIVcpz para o ser humano (Keele et al., 2006), enquanto que o grupo P e O teriam se originado de uma linhagem do SIV encontrado em gorilas (SIVgor) (Plantier et al., 2009; D'arc et al., 2015) (Figura 1.9). É interessante observar que o SIVgor também é originado do SIVcpz, evidenciando que o ancestral comum a todos os grupos do HIV-1 é o vírus circulante nos chimpanzés. Pesquisas realizadas em diferentes comunidades selvagens de chimpanzés foram capazes de demonstrar que a subespécie *Pan troglodytes troglodytes*, que habita as florestas do sudeste de Camarões, é portadora da linhagem de SIV que originou as linhagens virais

encontradas hoje no ser humano e também no gorila (Keele et al., 2006; Van Heuverswyn et al., 2007).

Enquanto que infecções pelo HIV-2 são quase que restritas a alguns países da África ocidental, como Guiné-Bissau e Senegal (De Silva et al., 2008), o HIV-1 está disperso mundialmente. Entretanto, apenas o HIV-1 grupo M teve real sucesso em disseminar-se pela espécie humana, sendo hoje o grande responsável pela pandemia mundial que afeta mais de 30 milhões de pessoas (Hamelaar, 2013). Por sua vez, infecções por linhagens virais dos grupos N, O e P representam menos de 1% das infecções pelo HIV-1 no mundo e foram identificadas em Camarões, Gabão ou em imigrantes vindos desses países (Sharp et al., 2011). Evidências tem demonstrado que o ponto inicial de ignição da pandemia do HIV-1 grupo M foi na cidade de Kinshasa, atual República Democrática do Congo, na década de 1920 (Faria et al., 2014). Na época esta era a maior cidade do centro-oeste da África e ricamente ligada, por via férrea e fluvial, a várias outras cidades. Provavelmente as linhagens ancestrais do HIV-1 grupo M chegaram em Kinshasa através de balsa, em indivíduos vindos do sudeste de Camarões, e encontraram nesta cidade condições ideais para sua adaptação e disseminação na espécie humana. Entre essas condições estão os altos índices de prostituição em Kinshasa, consequência das mudanças que a exploração colonialista europeia provocou na estrutura sócio-demográfica africana, e políticas de saúde colonialista (Pepin, 2011). Essas últimas visavam combater doenças infecciosas como sífilis, buba, lepra e tripanossomíases através de fármacos que eram inoculados utilizando-se seringas de vidro não esterilizadas entre as aplicações, o que pode ter promovido a infecção por outros patógenos presentes no sangue dos pacientes.

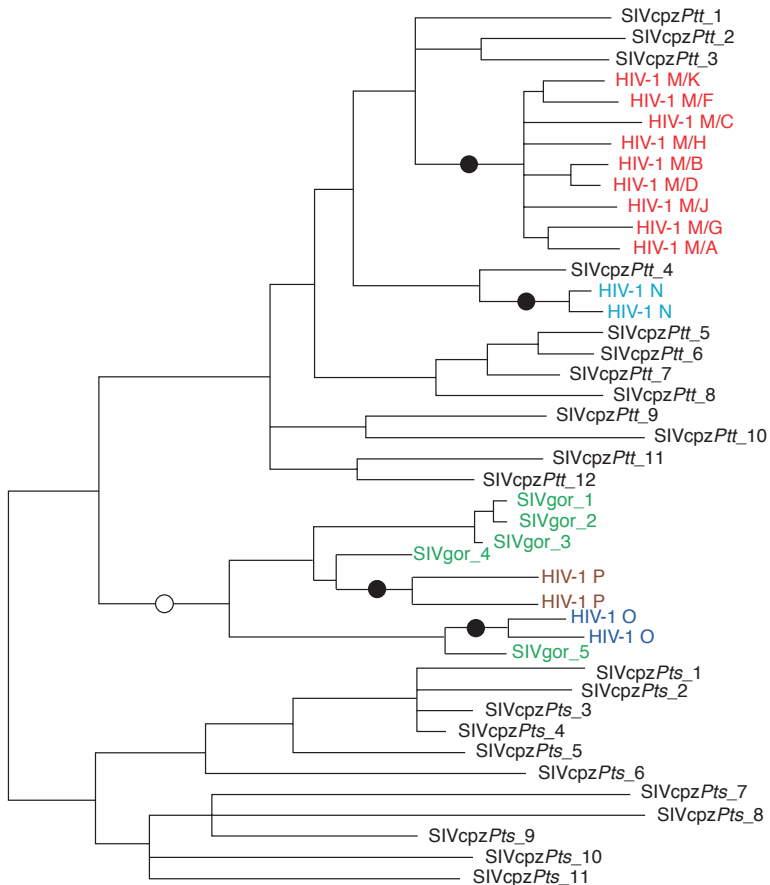


Figura 1.9 – As origens do HIV-1. Árvore filogenética demonstrando as relações evolutivas entre os grupos do HIV-1, SIVcpz e SIVgor. Sequências de SIVcpz estão em preto e correspondem a duas subespécies do chimpanzé: *Pan troglodytes troglodytes* (SIVcpzPtt) e *Pan troglodytes schweinfurthii* (SIVcpzPts). Sequências do SIVgor estão coloridas em verde e cada grupo do HIV-1 está pintado de uma cor diferente. Os círculos em preto representam os quatro eventos de transmissão para o ser humano e o círculo em branco representam o possível momento de transmissão do chimpanzé para o gorila. Adaptado de Sharp et al., 2011 e atualizado com informações de D’arc et al., 2015.

1.3.2 - A epidemiologia molecular do HIV-1 no mundo

O sucesso em disseminar-se por um grande número de indivíduos permitiu ao HIV-1 grupo M vasta diversificação genética. Atualmente este grupo é dividido nos subtipos A, B, C, D, F, G, H, J e K e nos sub-subtipos A1, A2, F1 e F2 (Hemelaar, 2013). Esta classificação é baseada na similaridade de seus genomas, podendo haver uma divergência de até 30% entre os diferentes subtipos. A alta diversidade do grupo M do HIV-1 é ainda mais pronunciada ao consideramos as formas recombinantes entre os subtipos. Quando a ocorrência de uma forma recombinante está restrita a menos de três indivíduos não relacionados ela é classificada como forma recombinante única (URF), do contrário ela é denominada forma recombinante circulante (CRF) (Vuilleumier, et al., 2015). Todos os anos novas CRFs são identificadas em diferentes regiões do mundo e até o momento 70 delas já foram descritas

(<http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>), representando mais de 20% das infecções mundiais pelo HIV-1.

A disseminação global dos subtipos do grupo M do HIV-1 envolveu eventos fundadores e gargalos populacionais que promoveram uma distribuição heterogênea dos subtipos no mundo (Tebit e Arts, 2011). O HIV-1 subtipo B (HIV-1B) foi o primeiro a infectar indivíduos fora do continente africano. Esta forma viral foi a responsável pelos primeiros casos de aids detectados no início da década de 80 nos Estados Unidos (Gilbert et al., 2007). Estima-se que o HIV-1B foi introduzido no Haiti em meados da década de 60, quando vários haitianos retornaram ao seu país de origem após anos de trabalho voluntário no Congo. A partir do Haiti o vírus se espalhou para os Estados Unidos e dali para o mundo. Atualmente esta forma do HIV é predominante na Europa, Américas, Austrália e Japão (Tebit e Arts, 2011). Por ser o grande responsável pela epidemia nos países ricos, este é o subtipo viral mais estudado e todas as inovações tecnológicas que visam o combate ao HIV são desenvolvidas a partir e testadas contra o subtipo B.

Apesar do HIV-1B ser um subtipo altamente disseminado pelo mundo ele corresponde a apenas 10% das infecções mundiais (Figura 1.10) (Hemelaar, 2013). Atualmente a forma do HIV-1 mais prevalente no mundo é o subtipo C (HIV-1C), responsável por mais da metade das

infecções. O HIV-1C é predominantemente encontrado no sul e nordeste da África, na Índia e vem apresentando crescente incidência na China e no Brasil. Outras formas do HIV-1 de importância mundial são o subtipo A1, de alta prevalência nos países centro-africanos, leste europeu e norte da Ásia; CRF02_AG, muito frequente no oeste da África; e CRF01_AE, dominante na epidemia do HIV-1 nos países do sudeste da Ásia (Hemelaar, 2013). É importante ressaltar ainda que nos países desenvolvidos mais de 40% das infecções recentes são de subtipos não-B, devido às altas taxas migratórias com origem de países em desenvolvimento (Ariën et al., 2007; Tebit e Arts, 2011).

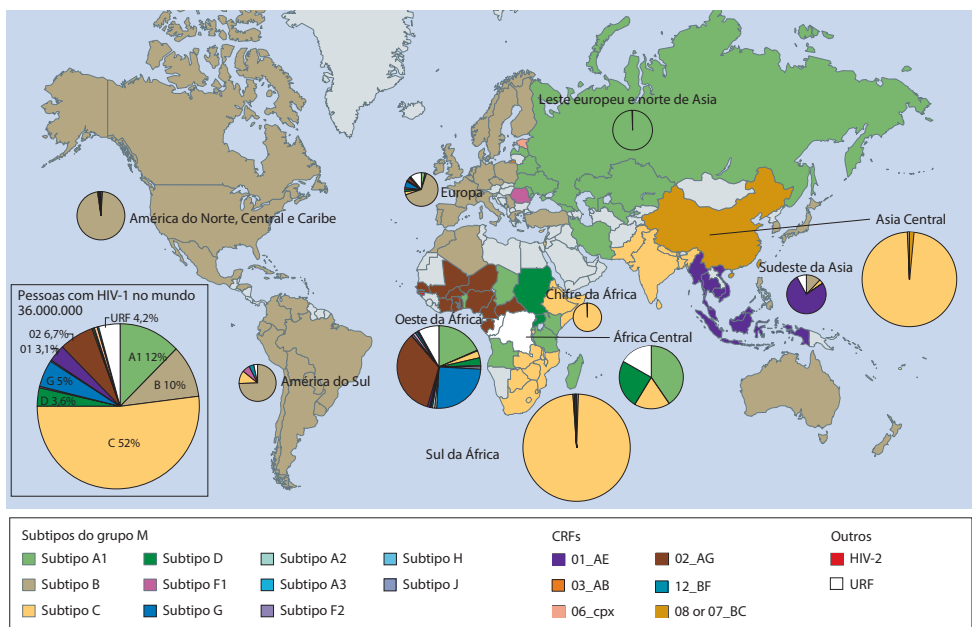


Figura 1.10 – Distribuição mundial dos subtipos do HIV-1 e CRFs. Países estão coloridos de acordo com a variante predominante e os gráficos detalham as frequências em grandes regiões geográficas. Países pintados de cinza não dispõem de dados suficientes. Adaptado de Ariën et al., 2007.

1.4 – A epidemia do HIV no Brasil e as peculiaridades da região Sul

Por ser um país de dimensões continentais, os números acerca da epidemia de HIV/aids no Brasil variam conforme a região geográfica analisada. Estimativas do Ministério da Saúde apontam que atualmente 734 mil pessoas estejam vivendo com HIV no Brasil (Departamento de DST, aids e Hepatites Virais, 2014). A prevalência entre 15 e 49 anos é de 0,6%, número que se mantém estável desde 2004. A razão de sexo é um índice que apresenta diferenças regionais importantes. De uma forma geral existe uma tendência ao equilíbrio, sendo que a razão de casos de aids entre homens e mulheres passou de 15:1 em 1986 para 1,8:1 em 2013. Entretanto a partir de meados dos anos 2000 a parcela dos homens infectados vem aumentando muito no Centro-Oeste e Sudeste, alcançando uma razão de 2:1 em 2013. Esse aumento recente da porção masculina de indivíduos infectados com HIV deve-se principalmente a crescente incidência de infecções em homens que fazem sexo com homens (HSH), sendo este aumento mais acentuado nas regiões Sudeste e Centro-Oeste, onde cerca de 50% dos homens infectados em 2013 eram HSH. As demais regiões do Brasil apresentaram comportamento semelhante, embora menos acentuado, destacando-se o Sul do Brasil com a menor proporção de HSH entre homens infectados – 30% em 2013. Além disso, o Sul do Brasil, é a região com maior proporção de mulheres infectadas, em torno de 10 para cada 14 homens, ou seja, uma razão homem:mulher de 1,4:1. Tais dados revelam uma importância maior da transmissão heterossexual na região Sul (Departamento de DST, aids e Hepatites Virais, 2014).

Uma segunda característica contrastante entre as regiões é a taxa de incidência de aids. No Brasil essa encontra-se estável desde o ano de 2002, em torno de 21 novos casos por 100 mil habitantes (Departamento de DST, aids e Hepatites Virais, 2014). Quase todas as regiões apresentam números semelhantes a este, com exceção da região Sul com aproximadamente 30 novos casos anuais por 100 mil hab. Essa taxa incide, por sua vez, na mortalidade por aids, que no Sul do Brasil gira em torno de 8,5 mortes por 100 mil hab., a maior do Brasil onde a média é de 6 mortes. Analisando-se com mais atenção os estados que compõem a região Sul, percebe-se que entre eles também há diferenças consideráveis nos indicadores da epidemia de aids. Nos últimos 10 anos, os estados de Santa Catarina (SC) e do Rio Grande do Sul (RS) vem

apresentando taxas de incidência sempre acima de 30 novos casos por 100 mil hab., enquanto no Paraná (PR) esse índice mantém-se abaixo de 20. A mesma diferenciação é observada na taxa de mortalidade, maior em SC e RS e similar à média brasileira no PR.

Os fatores que podem estar causando estas acentuadas diferenças entre a epidemia de aids no Sul e nas outras regiões do Brasil ainda são desconhecidos, mas um fato muito peculiar é marcante na distinção entre elas: a circulação do HIV-1 subtipo C no Sul do Brasil. Nas regiões Sudeste, Norte, Nordeste e Centro-Oeste do país o subtipo B é responsável por 70-90% das infecções. Nestas regiões os subtipos F1 e recombinantes BF1 apresentam prevalências variando de 3-17% e 2-14%, respectivamente, enquanto que a frequência de HIV-1C registrada na maioria dos estudos não passa de 3% (Teixeira et al., 2004; Sa Filho et al., 2005; Pedroso et al., 2007; Gadelha et al., 2003; Cardoso et al., 2009; Machado et al., 2009). Na região Sul, entretanto, o HIV-1C é encontrado com prevalências variando de 20-80%, sendo menor no PR e maior em SC e RS (Brígido et al., 2007; Medeiros et al., 2011; Silveira et al., 2012; Rodrigues et al., 2011; Gräf et al., 2011; Raboni et al., 2010; Silva et al., 2010).

Análises filodinâmicas da epidemia do subtipo B sugerem diferentes introduções, entre meados da década de 1960 e início da década de 1970, com origem provável nos Estados Unidos (Bello et al., 2007; Junqueira et al., 2011). Recentemente, demonstrou-se ainda que a epidemia brasileira inclui uma baixa prevalência de vírus introduzidos diretamente do Caribe (Cabello et al., 2015). Inicialmente, o HIV deve ter circulado predominantemente entre HSH e usuários de drogas injetáveis, sendo aos poucos introduzido na população heterossexual, tendo sido esse padrão também observado na Europa e Estados Unidos. O subtipo F1 por sua vez, chegou ao Brasil no final da década de 1970, com evidências apontando a República Democrática do Congo como origem (Bello et al., 2012). Ambas as formas do vírus apresentaram uma taxa de expansão crescente da epidemia até o final da década de 1980, e seu declínio é concordante com o histórico de novos casos de aids registrados anualmente no Brasil.

As características filodinâmicas da epidemia do HIV-1 subtipo C no Brasil já foram investigadas por diversos autores. No Capítulo 2 desta tese é apresentado uma detalhada revisão sobre todos estes estudos. Em grande parte, estas investigações foram realizadas nas capitais dos

estados do Sul e nenhuma delas era capaz de representar de uma forma mais ampla a população da região. Em vista disso e das peculiaridades da epidemia do HIV na região Sul do Brasil, mais precisamente nos estados de SC e do RS, nesta tese foi montado um banco de dados contendo dados epidemiológicos e sequências nucleotídicas do HIV de pacientes de 13 municípios do interior do RS e SC. No Capítulo 3 são apresentados os resultados de um trabalho de epidemiologia molecular acrescido de uma análise comparativa da história demográfica das epidemias do HIV-1 subtipo B e C na região amostrada. Por fim, no Capítulo 4 a história filogeográfica do HIV-1C no Brasil é reconstruída a partir de uma grande amostragem, utilizando-se técnicas avançadas de inferência filogenética e um modelo de difusão capaz de integrar dados epidemiológicos e genéticos na análise da dispersão viral.

1.5 - Objetivos

Esta tese possui dois objetivos principais. O primeiro é descrever de uma forma abrangente a epidemia do HIV-1 nos estados de Santa Catarina e Rio Grande do Sul e o segundo é entender a dinâmica de dispersão da epidemia do HIV-1 subtipo C entre diferentes cidades da região Sul e do Brasil. Especificamente, esta tese buscou atingir os seguintes objetivos:

- Revisar a bibliografia disponível sobre a epidemia do HIV-1 subtipo C no Brasil;
- Caracterizar molecularmente a epidemia do HIV-1 quanto aos subtipos e formas virais recombinantes circulantes em municípios distantes das principais zonas urbanas nos estados de Santa Catarina e Rio Grande do Sul;
- Avaliar o perfil da população infectada pelo HIV-1 quanto a características demográficas e epidemiológicas na região amostrada;
- Avaliar comparativamente a taxa de expansão das principais formas do HIV-1 circulantes na região amostrada;
- Caracterizar filogeograficamente a dispersão do HIV-1 subtipo C pelo Brasil;
- Avaliar a presença de preditores da dispersão do HIV-1 subtipo C.

1.6 – Referências

- Abecasis AB, Vandamme AM, Lemey P. Quantifying differences in the tempo of human immunodeficiency virus type 1 subtype evolution. *J Virol.* 2009;83(24):12917-24.
- Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol.* 2006;55(4):539-52.
- Ariën KK, Vanham G, Arts EJ. Is HIV-1 evolving to a less virulent form in humans? *Nat Rev Microbiol.* 2007;5(2):141-51.
- Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, et al. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Ann Rev Ecol Syst.* 1987;18(11):489–522.
- Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol.* 2012;29(9):2157-67.
- Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). 1983. *Rev Invest Clin.* 2004;56(2):126-9.
- Bedford T, Cobey S, Beerli P, Pascual M. Global migration dynamics underlie evolution and persistence of human influenza A (H3N2). *PLoS Pathog.* 2010;6(5):e1000918.
- Bello G, Afonso JM, Morgado MG. Phylodynamics of HIV-1 subtype F1 in Angola, Brazil and Romania. *Infect Genet Evol.* 2012;12(5):1079-86.
- Bello G, Eyer-Silva WA, Couto-Fernandez JC, Guimarães ML, Chequer-Fernandez SL, Teixeira SL, et al. Demographic history of HIV-1 subtypes B and F in Brazil. *Infect Genet Evol.* 2007;7(2):263-70.
- Bloomquist EW, Lemey P, Suchard MA. Three roads diverged? Routes to phylogeographic inference. *Trends Ecol Evol.* 2010;25(11):626-32.
- Brigido LFM, Nunes CC, Oliveira CM, Knoll RK, Ferreira JLP, Freitas CA, 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*. 2007; 23(12):1579–1585.
- Brown JW, Kilmer AJ. *The state of bayesian phylogenetics: Bayes for the uninitiated*. Kingston: Queen's University; 2003.
- Cabello M, Junqueira DM, Bello G. Dissemination of nonpandemic Caribbean HIV-1 subtype B clades in Latin America. *AIDS*. 2015;29(4):483-92.
- Cardoso LP, Queiroz BB, Stefani MM. HIV-1 pol phylogenetic diversity and antiretroviral resistance mutations in treatment naïve patients from Central West Brazil. *J Clin Virol*. 2009;46(2):134-9.
- Departamento de DST, aids e Hepatites Virais; Ministério da Saúde. Brasil. Boletim Epidemiológico AIDS. Ano III, nº. 1, até semana epidemiológica 52^a – Dezembro de 2013. Brasília; 2014.
- Cullen BR. HIV-1 auxiliary proteins: making connections in a dying cell. *Cell*. 1998;93(5):685-92.
- Cunningham CW, Omland KE, Oakley TH. Reconstructing ancestral character states: a critical reappraisal. *Trends Ecol Evol*. 1998;13(9):361-6.
- D'arc M, Ayouba A, Esteban A, Learn GH, Boué V, Liegeois F, et al. Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proc Natl Acad Sci U S A*. 2015;112(11):E1343-52.
- De Silva TI, Cotten M, Rowland-Jones SL. HIV-2: the forgotten AIDS virus. *Trends Microbiol*. 2008;16(12):588-95.
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006;4(5):e88.
- Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. Measurably evolving populations. *Trends Ecol Evol*. 2003;18(9):481–8.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012;29(8):1969-73.
- Dudas G, Rambaut A. Phylogenetic Analysis of Guinea 2014 EBOV Ebolavirus Outbreak. *PLoS Curr*. 2014;6.
- Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004;5:113.

- Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science*. 2014;346(6205):56-61.
- Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985;39(4):783-791.
- Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol*. 1981;17(6):368-76.
- Felsenstein J. The number of evolutionary trees. *Syst Zool*. 1978; 27(1): 27-33.
- Fitch WM. Toward defining the course of evolution: minimum change for a specific tree topology. *Syst Zool*. 1971;20(4):406-416.
- Freeman S, Herron JC. *Análise evolutiva*. 4. ed. Porto Alegre: Artmed; 2009.
- Gadelha SR, Shindo N, Cruz JN, Morgado MG, Galvão-Castro B. Molecular epidemiology of human immunodeficiency virus-1 in the state of Ceará, Northeast, Brazil. *Mem Inst Oswaldo Cruz*. 2003;98(4):461-3
- Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science*. 1984;224(4648):500-3.
- Ganser-Pornillos BK, Yeager M, Sundquist WI. The structural biology of HIV assembly. *Curr Opin Struct Biol*. 2008;18(2):203-17.
- Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A*. 2007;104(47):18566-70.
- Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol Biol Evol*. 2013;30(3):713-24.
- Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, et al. *Pneumocystis carinii* pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med*. 1981;305(24):1425-31.
- Goulder PJ, Watkins DI. HIV and SIV CTL escape: implications for vaccine design. *Nat Rev Immunol*. 2004;4(8):630-40.
- Gräf T, Passaes CPB, Ferreira LGE, Grisard EC, Morgado MG, Bello G, 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.* 2011; 51(3):186–191.
- Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science.* 2004;303(5656):327–32.
- Griffiths RC, Tavaré S. Sampling theory for neutral alleles in a varying environment. *Philos Trans R Soc Lond B Biol Sci.* 1994;344(1310):403–10.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307–21.
- Hemelaar J. Implications of HIV diversity for the HIV-1 pandemic. *J Infect.* 2013;66(5):391–400.
- Ho SY, Shapiro B. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol Ecol Resour.* 2011;11(3):423–34.
- Hoffmann C, Rockstroh J, Kamps B. *HIV Medicine.* 15. ed. Paris: Flying Publisher; 2007.
- Holder M, Lewis PO. Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet.* 2003;4(4):275–84.
- Holmes EC. Molecular epidemiology and evolution of emerging infectious diseases. *Br Med Bull.* 1998;54(3):533–43.
- Jenkins GM, Rambaut A, Pybus OG, Holmes EC. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *J Mol Evol.* 2002;54(2):156–65.
- Junqueira DM, de Medeiros RM, Matte MC, Araújo LA, Chies JA, Ashton-Prolla P, et al. Reviewing the history of HIV-1: spread of subtype B in the Americas. *PLoS One.* 2011;6(11):e27489.
- Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science.* 2006;313(5786):523–6.
- Kimura M. Evolutionary rate at the molecular level. *Nature.* 1968;217(5129):624–6.
- Kimura M. *The neutral theory of molecular evolution.* Cambridge: Cambridge University Press; 1983.
- Kindt T, Goldsby RA, Osborne B, Kuby J. *Kuby immunology.* 6th. ed. New York: WH Freeman and Company; 2006.

- Kingman JFC. The coalescent. *Stoch Proc Appl.* 1982(1);13:235–248.
- Klatt NR, Silvestri G, Hirsch V. Nonpathogenic simian immunodeficiency virus infections. *Cold Spring Harb Perspect Med.* 2012;2(1):a007153.
- Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, et al. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog.* 2014;10(2):e1003932.
- Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. *PLoS Comput Biol.* 2009b;5(9):e1000520.
- Lemey P, Rambaut A, Welch JJ, Suchard MA. Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol.* 2010;27(8):1877–85.
- Lemey P, Salemi M, Vandamme A. *The Phylogenetic Handbook: a practical approach to phylogenetic analysis and hypothesis testing.* Cambridge: Cambridge University Press; 2009a.
- Li S, Pearl D, Doss H. Phylogenetic tree reconstruction using Markov chain Monte Carlo. *J Am Stat Assoc.* 2000;95(1):493–508.
- Lo Presti A, Ciccozzi M, Cella E, Lai A, Simonetti FR, Galli M, et al. Origin, evolution, and phylogeography of recent epidemic CHIKV strains. *Infect Genet Evol.* 2012;12(2):392–8.
- Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol.* 1999;73(1):152–60.
- Machado LF, Ishak MO, Vallinoto AC, Lemos JA, Azevedo VN, Moreira MR, 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.* 2009;25(10):961–6.
- Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics.* 2010;26(19):2462–3.
- Mau B, Newton MA. Phylogenetic inference for binary data on dendrograms using Markov chain Monte Carlo. *J Comput Graph Stat.* 1997;6(1):122–131.

- McCormack GP, Clewley JP. The application of molecular phylogenetics to the analysis of viral genome diversity and evolution. *Rev Med Virol.* 2002;12(4):221-38.
- Medeiros RM de, Junqueira DM, Matte MCC, Barcellos NT, Chies JAB, Matos Almeida SE. 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.* 2011; 83(10):1682–1688.
- Mellors JW, Kingsley LA, Rinaldo CR, Todd JA, Hoo BS, Kokka RP, et al. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. *Ann Intern Med.* 1995;122(8):573-9.
- Miller RH, Sarver N. HIV accessory proteins as therapeutic targets. *Nat Med.* 1997;3(4):389-94.
- Pedroso C, Queiroz AT, Alcântara LC, Drexler JF, Diaz RS, Weyll N, 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.* 2007;45(2):251-3.
- Peeters M, Courgnaud V, Abela B, Auzel P, Pourrut X, Bibollet-Ruche F, et al. Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. *Emerg Infect Dis.* 2002;8(5):451-7.
- Pepin J. *The origins of AIDS.* Cambridge: Cambridge University Press; 2011.
- Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science.* 1996;271(5255):1582-6.
- Peter F. HIV nef: the mother of all evil? *Immunity.* 1998;9(4):433-7.
- Peterlin BM, Trono D. Hide, shield and strike back: how HIV-infected cells avoid immune eradication. *Nat Rev Immunol.* 2003;3(2):97-107.
- Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemée V, et al. A new human immunodeficiency virus derived from gorillas. *Nat Med.* 2009;15(8):871-2.
- Posada D. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* 2008;25(7):1253-6.
- Preston BD, Poiesz BJ, Loeb LA. Fidelity of HIV-1 reverse transcriptase. *Science.* 1988;242(4882):1168-71.

- Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. The epidemic behavior of the hepatitis C virus. *Science*. 2001;292(5525):2323-5.
- Qadir MI, Malik SA. HIV fusion inhibitors. *Rev Med Virol*. 2010;20(1):23-33.
- Raboni, SM, Almeida SM, Rotta I, Ribeiro CE, Rosario D, Vidal LR, Nogueira MB, Riedel M, Winhescki MAG, Ferreira KA, Ellis R. Molecular epidemiology of HIV-1 clades in Southern Brazil. *Mem. Inst. Oswaldo Cruz*. 2010; 105, 1044–1049.
- Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. *Nat Rev Genet*. 2004;5(1):52-61.
- Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. The genomic and epidemiological dynamics of human influenza A virus. *Nature*. 2008;453(7195):615-9.
- Rannala B, Yang Z. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J Mol Evol*. 1996;43(3):304–311.
- Robbins KE, Lemey P, Pybus OG, Jaffe HW, Youngpairaj AS, Brown TM, et al. U.S. Human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. *J Virol*. 2003;77(11):6359-66.
- Rodrigues R, Manenti S, Romao PRT, Paula Ferreira JL de, Batista JPG, Siqueira AFAC, et al. Young Pregnant Women Living with HIV/AIDS in Criciúma, Southern Brazil, Are Infected Almost Exclusively with HIV Type 1 Clade C. *AIDS Res Hum Retroviruses*. 2010; 26(3):351–357.
- Rodrigo AG, Felsenstein J. Coalescent approaches to HIV-1 population genetics. In: Crandall K. *The Evolution of HIV*. Baltimore: Johns Hopkins University Press; 1999. p. 233–272.
- Sa Filho DJ, Sanabani S, Diaz RS, Munerato P, Brunstein A, Fusuma E, 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*. 2005;21(2):145-51.
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987;4(4):406-25.

- Santiago ML, Lukasik M, Kamenya S, Li Y, Bibollet-Ruche F, Bailes E, et al. Foci of endemic simian immunodeficiency virus infection in wild-living eastern chimpanzees (*Pan troglodytes schweinfurthii*). *J Virol*. 2003;77(13):7545-62.
- Scally A, Durbin R. Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet*. 2012;13(10):745-53.
- Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med*. 2011;1(1):a006841.
- Silva MM, Telles FQ, da Cunha CA, Rhame FS. HIV subtype, epidemiological and mutational correlations in patients from Paraná, Brazil. *Braz. J. Infect. Dis*. 2010; 14, 495–501.
- Silveira J, Santos AF, Martínez AMB, Góes LR, Mendoza-Sassi R, Muniz CP, et al. Heterosexual transmission of human immunodeficiency virus type 1 subtype C in southern Brazil. *J Clin Virol*. 2012; 54(1):36–41.
- Slatkin M, Maddison WP. A cladistic measure of gene flow measured from the phylogenies of alleles. *Genetics*. 1989;123(3):603–613.
- Sleasman JW, Goodenow MM. 13. HIV-1 infection. *J Allergy Clin Immunol*. 2003;111(2 Suppl):S582-92.
- Suchard MA, Weiss RE, Sinsheimer JS. Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol*. 2001;18(6):1001-13.
- Sullivan J, Joyce P. Model selection in phylogenetics. *Annu Rev Ecol Syst*. 2005;36(1):445-466.
- Talbi C, Lemey P, Suchard MA, Abdelatif E, Elharrak M, Nourlil J, et al. Phylogenomics and human-mediated dispersal of a zoonotic virus. *PLoS Pathog*. 2010;6(10):e1001166.
- Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*. 1986;17(1):57-86.
- Tebit DM, Arts EJ. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis*. 2011;11(1):45-56.
- Teixeira SL, Bastos FI, Telles PR, Hacker MA, Brigido LF, de F Oliveira CA, 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.* 2004;31(3):221-6.
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22(22):4673-80.
- Van Heuverswyn F, Li Y, Bailes E, Neel C, Lafay B, Keele BF, et al. Genetic diversity and phylogeographic clustering of SIVcpzPtt in wild chimpanzees in Cameroon. *Virology.* 2007;368(1):155-71.
- Vuilleumier S, Bonhoeffer S. Contribution of recombination to the evolutionary history of HIV. *Curr Opin HIV AIDS.* 2015;10(2):84-9.
- Whelan S, Liò P, Goldman N. Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends Genet.* 2001;17(5):262-72.
- WHO library cataloguing in publication data, Joint United Nations Programme on HIV/AIDS. Global report: UNAIDS report on the global AIDS epidemic; 2010.
- Wu L, Kewalramani VN. Dendritic-cell interactions with HIV: infection and viral dissemination. *Nat Rev Immunol.* 2006;6(11):859-68.
- Yang Z, Rannala B. Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method. *Mol Biol Evol.* 1997;14(7):717-24.
- Yang Z, Rannala B. Molecular phylogenetics: principles and practice. *Nat Rev Genet.* 2012;13(5):303-14.
- Yang Z. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol Biol Evol.* 1993;10(6):1396-401.
- Zuckerlandl E, Pauling L. Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ. *Evolving genes and proteins.* New York: Academic press; 1965. p. 97-166.
- Zuckerlandl E, Pauling L. Molecular disease, evolution and genetic heterogeneity. In: Kasha M, Pullman B. *Horizons of biochemistry.* New York: Academic press; 1962. p. 185-225.

CAPÍTULO 2.

The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent

Publicado em:

Gräf T, Pinto AR. The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. *Virology*. 2013;435(1):170-8

2.1 – Abstract

The HIV-1 has evolved swiftly and the scenario of HIV-1 genetic diversity is constantly changing. In South America, recombinant forms of subtypes B, F1, and BF1 have historically driven the HIV-1 epidemic. In recent years, however, infection with subtype C has gained prominence as its prevalence increased in Southern Brazil as well as neighboring countries. Current studies point to a single introduction of closely related strains as the beginning of the Brazilian subtype C epidemic. However, the place of origin of these strains, date, and route of introduction are under continuous debate as well as the clinical outcomes of the emergence of subtype C. Therefore, this paper reviews the history of the HIV-1 subtype C in Brazil, particularly in the Southern region, covering its demographic and evolutionary history and the possible implications to the Brazilian AIDS epidemic as well as to neighboring countries.

2.2 – Introduction

The first AIDS case was diagnosed in Brazil in 1982 and since then 608,230 cases had been reported throughout the country (Brazilian Ministry of Health, 2011). The prevalence of AIDS in the adult population is 0.6% and the incidence of new cases was 17.9/100,000 inhabitants in 2010. With a population exceeding 192 million, Brazil has the largest number of HIV infected people among countries from Central and South America, and according to UNAIDS accounts for one third of all people living with HIV in this region (Brazilian Institute of Geography and Statistics [IBGE], 2012 and World Health Organization [WHO], 2010).

Brazil is a country of continental dimensions, spread over an area of 8,514,877 km² which occupies roughly half of South America, divided into 26 states and a Federal District, which are grouped into five geopolitical regions. The states of Rio Grande do Sul (RS), Santa Catarina (SC), and Paraná (PR) compose the Southern region, which is the smallest region of the country with a population of 27 million, 14% of its total population (Brazilian Institute of Geography and Statistics [IBGE], 2012). This region shares borders with Uruguay, Argentina,

and Paraguay, as well as with the Brazilian states of São Paulo and Mato Grosso do Sul (Fig. 2.1).

The epidemiological situation in Southern Brazil regarding AIDS is concerning. The AIDS-related mortality is steadily increasing and ever since the beginning of the 2000s is highest among Brazilian regions (Brazilian Ministry of Health, 2011). The number of newly diagnosed cases is also the highest in the country, and in 2010 the incidence rates in the states of Rio Grande do Sul and Santa Catarina were 37.6 and 30.2 new cases per 100,000 persons, respectively, whereas in the state of Paraná the incidence rate was 19 new cases per 100,000 persons. Moreover, the top ten Brazilian cities with the highest incidence of AIDS in 2010 were from the Southern region, and Porto Alegre, the capital city of Rio Grande do Sul, had an incidence of 99,8 new cases per 100,000 persons, the highest incidence of AIDS in the country (Brazilian Ministry of Health, 2011).

High molecular variability is a striking feature of HIV-1 in the global AIDS pandemic. Among four HIV-1 groups (M, N, O, and P), HIV-1 group M is globally distributed, being subdivided into nine subtypes (A–D, F–H, J, and K) and, 51 circulating recombinant forms (CRFs) identified so far (Ng et al., 2012, Plantier et al., 2009 and Tebit and Arts, 2011). The AIDS pandemic in Brazil is mostly caused by the HIV-1 subtype B, while subtypes C, F1, and BF1 recombinants are observed in lower frequencies (Gadelha et al., 2003, Machado et al., 2009, Pedroso et al., 2007, Sa Filho et al., 2005, Stefani et al., 2007 and Teixeira et al., 2004). However, the distribution of HIV-1 variants is heterogeneous in the inner parts of the country, while subtype C is highly prevalent in Southern Brazil (Gräf et al., 2011, Medeiros et al., 2011 and Silva et al., 2010).

Currently, subtype C and C-containing forms are responsible for more than 51% of all worldwide HIV-1 infections (Tebit and Arts, 2011). Dominant in India, Southern African countries, and with increasing presence in China and Eastern Africa, subtype C and C-containing recombinant forms appear to spread more rapidly than other group M subtypes (Ariën et al., 2007 and Tebit and Arts, 2011). Moreover, increasing subtype C prevalence has been reported in some countries from Western Europe, the Americas, and Australia (Abecasis et al., 2011, Chibo and Birch, 2012, Dougan et al., 2005, Tatt et al., 2004 and Tebit and Arts, 2011). Due to the high prevalence of HIV-1

subtype B in developed countries, research on the natural history and health technology improvements have largely focused on this variant. However, the actual global epidemiological profile requires further studies evaluating the impact of HIV-1 diversity, especially the dominant subtype C on disease progression and treatment outcomes. Thus, molecular epidemiology studies are pivotal for evaluating phenotypic relevance and providing underlying data concerning viral diversity and evolution.

In Brazil, studies published over a decade have investigated the molecular features of the local HIV epidemic, producing valuable data regarding the importance of subtype C and C-containing recombinant forms in the Southern region and its implications to neighboring countries as well other Brazilian states. Given the extensive bibliography available and the actual global relevance of HIV-1 subtype C, this article aims to review the history of subtype C in Brazil, from the early cases in the South to the increasing prevalence in the North.

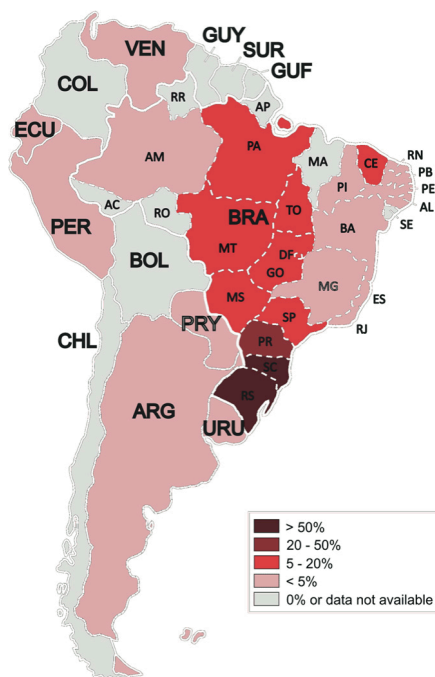


Figure 2.1. Prevalence of HIV-1 subtype C and C-containing sequences in Brazilian states and South American countries. Acronyms for Brazilian states: RS (Rio Grande do Sul), SC (Santa Catarina), PR (Paraná), SP (São Paulo), RJ (Rio de Janeiro), MG (Minas Gerais), ES (Espírito Santo), BA (Bahia), SE (Sergipe), AL (Alagoas), PE (Pernambuco), PB (Paraíba), RN (Rio Grande do Norte), CE (Ceará), PI (Piauí), MA (Maranhão), PA (Pará), AP (Amapá), AM (Amazonas), RR (Roraima), AC (Acre), RO (Rondônia), MT (Mato Grosso), TO (Tocantins), GO (Goiás), MS (Mato Grosso do Sul), and DF (Distrito Federal).

2.3 - Origin of Brazilian HIV-1 C clade

The history of how HIV-1 subtype C was introduced in Brazil is being revealed as new epidemiological data are published and more Brazilian subtype C sequences become available. The first phylogenetic analyses indicated that subtype C may have entered Brazil with a single introduction of closely related strains, since Brazilian HIV-1 subtype C sequences (HIV-1C_BR) formed a monophyletic cluster when compared to sequences of the same subtype from India and some African countries (Monteiro et al., 2007, Sanabani et al., 2006 and Soares et al., 2003). Furthermore, amino acid signatures absent in the subtype C world consensus sequence were identified in HIV-1C_BR, reinforcing a monophyletic introduction hypothesis (M.A. Soares et al., 2003). Early studies also reported a smaller genetic distance between sequences of HIV-1C_BR compared with Brazilian subtype B sequences, suggesting a more recent introduction of HIV-1 C than subtype B in Brazil (Monteiro et al., 2007, Sanabani et al., 2006 and Soares et al., 2005).

The strong evidence supporting a monophyletic origin of the Brazilian HIV-1C epidemic prompted questions regarding the geographical origin and migration route of the founder strains. In 2008, two studies began elucidating this issue, reporting that HIV-1C_BR sequences were closely related to subtype C sequences from Middle East African countries (Bello et al., 2008 and Fontella et al., 2008). More specifically, sequences from Burundi at the root of the Burundi-Brazilian clade showed that HIV-1C_BR originated from a strain related to this country (Bello et al., 2008). Due to scarce social, cultural,

and economic relationships between Brazil and Burundi, a direct introduction of HIV-1C from Burundi or neighboring countries (Democratic Republic of Congo, Rwanda, or Tanzania) into Brazil was further evaluated (Brígido, 2009). Regarding cultural, social, and historical determinants, Mozambique was a putative source for the Brazilian HIV-1C epidemic; however, phylogeny estimations lacked evidence for a Mozambican ancestry (Brígido, 2009 and Fontella et al., 2009). A new hypothesis for the introduction of HIV-1C in Brazil emerged recently. Phylogenetic analysis revealed that subtype C sequences from the United Kingdom (UK), Brazil, and East Africa grouped together in a monophyletic cluster, where some UK sequences appeared at the base of the Brazilian HIV-1C clade (Oliveira et al., 2010). This suggests that the UK may have played a key role in the introduction of HIV-1C into Brazil. A more recent phylogeographic analysis, however, found no evidence of a direct viral flow from the UK to Brazil, only from both East Africa and Brazil to the UK (Véras et al., 2011), which may be expected since the UK is home to large Brazilian and East African immigrant populations. Thus, the precise migration route of HIV-1C from East Africa to Brazil remains unclear.

The exact time when the HIV-1C clade was introduced in Brazil is also controversial. The onset date of subtype C epidemic in Brazil was first estimated by Salemi et al. (2005), which reported the early 1990s as the most likely period for HIV-1C introduction, about 30 years after subtype B was introduced. Some years later, Bello et al., 2008 reestimated the onset year for the HIV-1 subtype C epidemic in Brazil, citing the early 1980s as the probable time of introduction. Subtype C sequences from patients infected in the 1980s were included in this study, supporting the hypothesis of an introduction prior to 1990. More recently, Véras et al. (2011) reevaluated the dates, suggesting an introduction between the 1960s and the 1970s. The latter study included sequences from the three states that compose Southern Brazil plus the state of São Paulo, while Bello et al. only included sequences from the state of Rio Grande do Sul, which may explain differences between the two estimates.

2.4 - Recognition of the HIV-1 C epidemic in Rio Grande do Sul

Despite the idea that HIV-1C_BR clade probably circulated in Brazil before the early 1980s, the existence of an established subtype C epidemic in the country was only recognized in the 1990s. The WHO Network for HIV Isolation and Characterization (1994) reported for the first time the existence of HIV-1C in Latin America. This subtype was found in 14 clinical samples obtained in Porto Alegre (capital city of Rio Grande do Sul) between 1992 and 1993. Shortly after that, the same group identified subtype C in three other Brazilian patients, two from Porto Alegre and one from São Paulo (capital city of São Paulo) (Osmanov et al., 1994). At the time only subtypes B and F1 had been identified in South America, and the possibility of a subtype C epidemic in Southern Brazil motivated several other studies focusing on the molecular epidemiology of HIV-1.

The real scenario of the HIV epidemic in Southern Brazil began to be unraveled in 2002 when Martínez et al. found 22% of subtype C among samples collected between 1994 and 1997 in Rio Grande, a port city in the state of Rio Grande do Sul. In 2003 Soares M.A. et al. corroborated these findings reporting a 37% prevalence of HIV-1C in the state of Rio Grande do Sul, reaching as high as 44% when recombinant viruses containing subtype-C related sequences were considered. Also in 2003, the Brazilian Network for HIV Drug Resistance Surveillance reported a high frequency of BC recombinant strains in samples from Rio Grande do Sul, showing the intermixing of subtype B and C epidemics in Southern Brazil (Brindeiro et al., 2003). Table 2.1 summarizes all studies performed in the state of Rio Grande do Sul before the publication of this review. The co-circulation of multiple HIV-1 clades in a restricted geographic region favors the emergence of recombinant forms, as observed frequently in some sub-Saharan countries (Tebit and Arts, 2011). In Southern Brazil, the epidemiological scenario also proved favorable to the emergence of HIV recombinant forms.

Table 2.1. Summary of studies in HIV molecular epidemiology performed in the state of Rio Grande do Sul (RS), Southern Brazil.

Reference	City	Sampling n (ARV exposure)	Sampling year	HIV-1 diversity	Gene analyzed
Martinez et al. 2002	Rio Grande	69 (naïve and treated)	1994-1997	22% C, 75% B, 3% F	<i>env</i>
Soares, M.A. et al. 2003	Porto Alegre	57 (naïve)	2001	37% C, 7% CB, 56% B+F1	<i>pol</i> (PR/RT)
Brindeiro et al. 2003	Not available	139 (naïve)	2001	45% C, 30% B, 3% F1, 22% recombinants	<i>pol</i> (PR/RT)
Soares, E.A. et al. 2003	Porto Alegre	77 (naïve and treated)	2002	31% C, 45% B, 12% F1, 9% CRF31*, 1,5% BF1, 1,5% BD	<i>pol</i> (PR/RT)
Soares, E.A. et al. 2005	Rio Grande	85 (naïve and treated)	2002	41% C, 42% B, 5% F1, 4% CRF31*, 2% D, 6% URF	<i>pol</i> (PR/RT)
Rodrigues et al. 2006	Porto Alegre	108 (naïve)	2004	33% C, 32% B, 3% F1, 25% CRF31*, 7% others	<i>pol</i> (PR/RT)
Santos et al. 2006	Porto Alegre and Rio Grande	152 (unknown)	2002-2003	31% C, 46% B, 4% F1, 10% CRF31, 9% others	<i>pol</i> (PR/RT)
Monteiro et al. 2007	Porto Alegre	22 (unknown)	2003	27% C, 37% B, 4,5% F1, 27% BC, 4,5% BF1	<i>env</i> and <i>gag</i>
Brigido et al. 2007	Porto Alegre	122 (naïve)	2004-2006	27% C, 30% B, 10% F1, 21% CRF31, 8% URF BC, 4% URF	<i>env</i> and <i>pol</i> (PR/RT)
Dias et al. 2009	Porto Alegre	128 (naïve)	2004 - 2005	32% C, 25% B, 11% F1, 26% CRF31, 6% URF	<i>pol</i> (PR/RT)
Simon et al. 2010	Canoas	80 (naïve and treated)	2008-2009	44% C, 19% B, 2% F1, 35% CRF31	<i>pol</i> (PR/RT)
Medeiros et al. 2011	Porto Alegre	99 (naïve)	2006-2007	40% C, 26% B, 1% F1, 19% CRF31, 12% URF BC, 2% URF BF1	<i>pol</i> (PR/RT)
Silveira et al. 2012	Rio Grande	245 (naïve)	2005-2008	56% C, 21% B, 5% F1, 10,5% URF_BC, 0,5% CRF31, 6% URF BF1, 1% others	<i>pol</i> (PR/RT, CN and RH)

PR/RT to protease/reverse transcriptase, CN to connection, RH to RNase H. *Values reported by Santos et al. (2007).

2.4.1 - Identification of a new CRF_BC in Rio Grande do Sul

The high variability of HIV is also a consequence of recombination events. Inter-subtype recombination occurs when an individual is co-infected with two different HIV strains at once, or when a previously infected individual acquires a different strain of HIV, which is called superinfection. These two phenomena lead to the emergence of unique recombinant forms (URFs) or circulating recombinant forms (CRFs). In Brazil recombination events were reported between subtypes B and F1, with the BF1 mosaic as the most prevalent HIV genetic form after subtype B in most of the country (Guimarães et al., 2008 and Sa Filho et al., 2006). The introduction of subtype C in the Southern region, however, gave rise to several BC mosaic strains and a new CRF.

Most of the early studies revealing the high prevalence of subtype C in the state of Rio Grande do Sul also reported frequencies of recombinant BC forms ranging between 6% and 22% (Brindeiro et al., 2003, Soares et al., 2003 and Soares et al., 2005). Rodrigues et al. reported in 2006 a remarkable 25% of BC recombinants in samples from Porto Alegre. Detailed analyses of these BC sequences showed

high similar recombinant structure, lower intra-cluster distances, and common amino acid signatures, pointing strongly to a new circulating recombinant variant. Shortly afterward, Santos et al. (2006) characterized this relevant BC recombination form as a new CRF named CRF31_BC. This HIV-1 variant features a predominantly subtype C genome, with a 240bp subtype B fragment inserted in the reverse transcriptase gene.

Previous studies indicate that the CRF31_BC clade likely originated in the late 1980s and rapidly expanded in the city of Porto Alegre (Passaes et al., 2009). The prevalence of the CRF31_BC in this city was estimated to be 10% in samples collected between 2002 and 2003 and 25% in samples from 2004 (Santos et al., 2007) (Table 2.1). Despite occasional detection of CRF31_BC clade beyond Porto Alegre, the prevalence of this recombinant variant in other cities in Rio Grande do Sul, Santa Catarina, and Paraná seems to be much lower (<5%) (Brígido et al., 2007, Gräf et al., 2011, Raboni et al., 2010, Rodrigues et al., 2010, Silva et al., 2010 and Silveira et al., 2012).

2.5 - The HIV-1 C epidemic in Santa Catarina and Paraná

The unique features of the HIV/AIDS epidemic in the state of Rio Grande do Sul galvanized researchers to better understand HIV-1 molecular epidemiology in the Southern Brazilian region. In 2007, the first cases of subtype C infection were reported in Santa Catarina, a southern state without previous HIV-1 molecular data, in a small group of three female sex workers from the city of Imbituba (Schuelter-Trevisol et al., 2007). Shortly after that, Locateli et al. (2007) reported a 48% frequency of subtype C in samples obtained in 2004 from Florianópolis (capital city of Santa Catarina). Frequencies of subtype C over 60% were reported a few years later in Florianópolis, Camboriú, Itajaí, and Criciúma (Brígido et al., 2007, Gräf et al., 2011, Rodrigues et al., 2010 and Pinto et al., 2012), making Santa Catarina the state with the highest prevalence of HIV-1C in Brazil. High prevalence (16%) of URF_BC in Santa Catarina was also reported, reflecting the co-existence of subtypes C and B in the same population (Gräf et al., 2011).

The first molecular epidemiology study in the state of Paraná, conducted by Brindeiro et al. in 2003, reported a subtype C prevalence

of 30%, the same frequency subsequently reported by Ferreira et al., 2008 in the capital city (Curitiba). More recently, Toledo et al. (2010), Raboni et al. (2010) and Silva et al. (2010) evaluated patient samples from several cities in Paraná, reporting 21–27% of subtype C in patients with therapeutic failure. Raboni et al. (2010) also observed higher proportion of subtype C (31%) in cities located in the southern region of the state of Paraná, bordering Santa Catarina, while in the northern region of the state, which borders São Paulo state, only 13% of subtype C was observed. Table 2.2 summarizes the studies performed in the states of Santa Catarina and Paraná.

Table 2.2. Summary of studies in HIV molecular epidemiology performed in the states of Santa Catarina (SC) and Paraná (PR), Southern Brazil.

Reference	City/State	Sampling n (ARV exposure)	Sampling year	HIV-1 diversity	Gene analyzed
Shuelter-Trevisol et al. 2007	Imbituba/SC	4 (naïve)	2003-2004	75% C, 25% B	<i>pol</i> (PR/RT)
Locateli et al. 2007	Florianópolis/SC	80 (naïve and treated)	2004	49% C, 22.5% B, 22.5% BC, 6% BF1	<i>env</i> and <i>gag</i>
Brígido et al. 2007	Itajai and Camboriú/ SC	83 (naïve)	2004-2006	64% C, 26.5% B, 6% F1, 3.5% CRF31	<i>env</i> and <i>pol</i> (PR/RT)
Rodrigues et al. 2010	Criciúma/SC	42 (naïve and treated)	2007	79% C, 7% B, 9.5% F1, 2% CRF31, 2.5% CF1	<i>pol</i> (PR/RT)
Gräf et al. 2011	Florianópolis/SC	82 (naïve)	2008-2009	66% C, 13.5% B, 1% F1, 2.5% CRF31, 16% URF_BC, 1% URF_BCf	<i>env</i> and <i>pol</i> (PR/RT/INT)
Brindeiro et al. 2003	Not available/PR	147 (naïve)	2001	30% C, 64% B, 4% F1, 2% rec.	<i>pol</i> (PR/RT)
Ferreira, J.L. et al. 2008	Curitiba/PR	38 (naïve)	2005-2006	30% C, 52% B, 2% F1, 14% BC, 2% CF1	<i>env</i> and <i>pol</i> (PR/RT)
Toledo et al. 2009	several cities/PR	389 (virological failure)	2003-2006	21% C, 61% B, 5% F1, 2.5% BC, 10.5% BF1	<i>pol</i> (PR/RT)
Raboni et al. 2010	several cities/PR	239 (virological failure)	1999-2007	26.5% C, 52% B, 10.5% F1, 8.5% URF	<i>pol</i> (PR/RT)
Silva et al. 2010	Curitiba/PR	191 (virological failure)	1987-2008	21.5% C, 69% B, 5% F1, 4% URF_BC, 0.5% URF_CF	<i>pol</i> (PR/RT)

PR/RT to protease/reverse transcriptase, CN to connection, RH to RNase H

2.6 - Temporal trends and spatial dynamics of HIV-1 C expansion and CRF31_BC epidemics in Southern Brazil

Reconstruction of the demographic history of subtype C and CRF31_BC epidemics in Porto Alegre, performed using a Bayesian coalescent-based approach, indicates that both HIV-1 clades experienced a period of fast exponential spread in this Brazilian city during the 1980s and 1990s (Bello et al., 2009). Consistent with this finding, E.A. Soares et al. (2003) observed an increasing frequency of subtype C in the city of Porto Alegre from 0% among patients

diagnosed before 1990 to 43% in patients diagnosed between 2001 and 2002. The following studies confirmed the continued increase in HIV-1 subtype C infections in Porto Alegre: 27% reported by Brígido et al. (2007), 32% observed by Dias et al. (2009), and 40% reported by Medeiros et al. (2011). A study conducted in the city of Canoas (Porto Alegre metropolitan area) found 44% of subtype C in 2010 (Simon et al., 2010). In the same reports, the frequency of subtype B sharply decreased since the early 2000s.

In general, the proportion of CRF31_BC also increased over time in Porto Alegre, though not linearly. A wide variation in the frequency of CRF31_BC between published studies is observed and may reflect the difficulties in identifying the recombination pattern, featured by a short subtype B insertion within a whole subtype C genome. In addition, minimal variations in the point of recombination of CRF31_BC get in the way of the classification, which can reclassify the strain into a second generation recombinant form or a URF_BC (Passaes et al., 2009). Therefore, the differences between CRF31_BC and unique recombination forms may be minimal, explaining why a lower prevalence of CRF31_BC is observed along with a high prevalence of URFs, mostly URF_BC in some studies.

A temporal increase of subtype C was also observed in other cities from Rio Grande do Sul and Santa Catarina, including Rio Grande and Florianópolis, respectively. A subtype C prevalence of 41% was reported in the city of Rio Grande in 2005 and 56% in 2012, over twice as high as the first work performed in the same city in 2002 (Martínez et al., 2002, Silveira et al., 2012 and Soares et al., 2005). Evaluation of the HIV-1 genotype distribution at the env gene from two studies performed in Florianópolis also revealed a significant increase in the proportion of subtype C in the heterosexual population over time: from 56% in 2004 to 83% in 2008–2009 (Gräf et al., 2011), while the proportion of subtype B in that population decreased from 38% to just 15% in the same period.

No similar temporal increase in the prevalence of subtype C has been detected in Paraná. The prevalence of this HIV-1 subtype has remained relatively stable (20–30%) across different studies performed between 2003 and 2010 (Brindeiro et al., 2003, Ferreira et al., 2008, Raboni et al., 2010, Silva et al., 2010 and Toledo et al., 2010). A significant increase in the frequency of URFs_BC over time, however,

was reported in Paraná by Ferreira et al. (2008), probably reflecting the frequent and ongoing co-infections of patients living in this region with HIV-1 subtypes C and B clades.

Regarding the spatial dynamics of the HIV-1C epidemic in Brazil, two recent reports using phylogeography inference point to the state of Paraná as the most likely source of subtype C in the southern region (Bello et al., 2012a and Vêras et al., 2011). The study of Vêras et al. (2011) indicates that subtype C expanded through Southern Brazil with an asymmetrical net viral flow following a north to south axis, from Paraná to both Santa Catarina and Rio Grande do Sul. A recent study performed by our group with a large number of sequences collected from the three southern capitals reveals a much more complex scenario (Bello et al., 2012a). According to our analyses, Florianópolis (capital of Santa Catarina) acted as an important staging post between Curitiba (capital of Paraná) and Porto Alegre (capital of Rio Grande do Sul), both receiving and sending viral lineages to the neighboring cities. Meanwhile, the direct viral migration flow between Curitiba and Porto Alegre seems to be much lower. Such estimated subtype C migration flows could be partially explained by the spatial distance and routine migration of people between Southern Brazilian states.

2.7 - HIV-1C expansion throughout the country

In general, subtype C seems to be slowly expanding north in the Brazilian territory as can be seen in Fig. 2.1. In 2010 a slight increase in the prevalence of HIV-1C was observed in the state of São Paulo, ranging from 0.3% in samples from 2005 to 2.9% in 2008, reaching 8.3% if only newly diagnosed patients from 2008 are considered (Brígido et al., 2011). These findings were recently confirmed by a study reporting 7.3% of subtype C and C-containing sequences in the city of São Paulo (Alcalde et al., 2012). No evidence of fast expansion of subtype C has been reported in other Southeast states, including Rio de Janeiro and Minas Gerais (Castro et al., 2010 and Westin et al., 2011). Thus, despite the close geographical proximity and the frequent routine movement of people from the Southern region, subtype B continues to be the widely dominant HIV-1 clade in Southeast Brazil.

Nonetheless, a relatively high prevalence of HIV-1C infections has been reported recently in some states from the Central-West, North

and Northeast regions. In 2009, a study from the Brazilian Network for HIV Drug Resistance Surveillance found 6% of subtype C in the state of Pará and in the national capital Brasília (Inocencio et al., 2009). In 2011, two studies reported, respectively, 11.7% and 9.6% of subtype C and C-containing forms in the inland cities of Cuiabá, capital of Mato Grosso (MT), and Palmas, capital of Tocantins (TO) (Ferreira et al., 2011 and Carvalho et al., 2011) (Fig. 2.2). Also in 2011, a frequency of subtype C and C-containing forms of close to 20% was reported by a study performed in inmates from Central-West Brazil (Cardoso et al., 2011). Due to job opportunities in agriculture and livestock, Brazilian Central-West region attracts a large number of immigrants, mainly from the Southern region. Therefore, it is possible that this migration route is responsible for the dissemination of subtype C to this region in recent years. Moreover, a prevalence of 5.4% HIV-1C was recently reported in the state of Ceará (CE), the highest prevalence ever recorded in that state (Arruda et al., 2011).

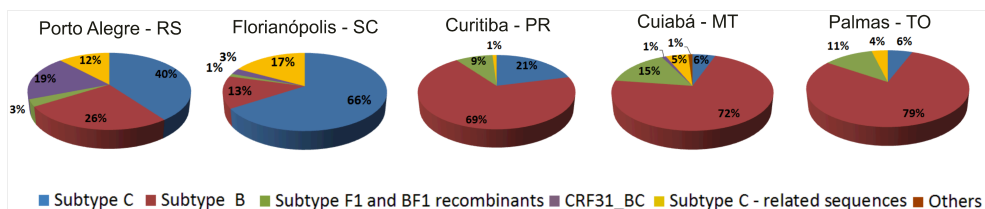


Figure 2.2. HIV-1 diversity in the Brazilian state capitals mainly affected by subtype C epidemic according to the most recent studies. Data according to Medeiros et al. (2011), Gräf et al. (2011), Silva et al. (2010), Ferreira et al. (2011), and Carvalho et al. (2011)

2.8 - HIV-1C in South American countries

The HIV epidemic in South American countries is characterized by a predominant circulation of subtypes B and BF recombinants, with the latter being observed with higher prevalence in Southern Cone countries (Dilernia et al., 2007, Montano et al., 2005 and Ruchansky et al., 2009). Carrion et al., 2004 were the first to report HIV-1 subtype C infections in South America after analyzing 1629 samples from Ecuador, Colombia, Bolivia, Venezuela, Peru, Chile, Argentina,

Paraguay, and Uruguay. Seven subtype C and one BC strains were identified in Argentina, Paraguay, and Uruguay, countries that share borders with the Brazilian Southern region and compose the so-called Southern Cone. Corroborating these findings, subsequent reports also observed prevalence of less than 1% in subtype C or C-containing forms in patients from Ecuador, Peru, and Southern Cone countries (Aulicino et al., 2005a, Aulicino et al., 2005b, Montano et al., 2005 and Pando et al., 2006). In these reports, patients infected with HIV-1C or C-containing forms frequently reported some epidemiologic link to Brazil, such as living in border towns or having had sexual partners from Brazil.

Lately several studies have detected the circulation of subtype C in Paraguay, Uruguay, and Argentina, although still at low frequencies, accounting for less than 5% of the HIV infections in those countries (Aguayo et al., 2008, Pando et al., 2011 and Ruchansky et al., 2009). In addition, Jones et al. described in 2009 two CRF31_BC sequences in Argentina. These studies point to a pattern of very slow spread of HIV-1C and related BC sequences from Brazil to neighboring countries from the Southern Cone. In fact, HIV-1C sequences from Argentina and Uruguay were found to be closely related to Brazilian strains in several phylogenetic studies, supporting the hypothesis of a single introduction of subtype C in Southern Brazil which later spread to other South American countries (Dilernia et al., 2007, Fontella et al., 2008, Jones et al., 2009, Monteiro et al., 2007 and V eras et al., 2011). Nevertheless, studies performed in Venezuela in 2005 and 2012 found subtype C sequences that grouped outside the Brazilian HIV-1C cluster (Castro et al., 2005 and Rangel et al., 2012). Epidemiologic data showed that two of these sequences were originated from patients most likely infected in South Africa, while the other case was identified in an Amerindian woman who lacked epidemiological information. Though these studies raised the possibility that new HIV-1C lineages may have been introduced in South America, more specifically in Venezuela, a more comprehensive study performed in 2009 did not detect subtype C infections in 425 Venezuelan patients (Rangel et al., 2009). The HIV-1 epidemic in Venezuela is predominantly driven by subtype B and the few reported cases of HIV-1C infections might not mean that subtype C is circulating in the Venezuelan population. In Fig. 2.1 the overall prevalence of HIV-1C and C-containing sequences in South American countries is shown.

2.9 - The subtype C epidemic and its relevance

The rapid emergence of HIV-1 subtype C globally warrants special attention to the pathogen itself. However, little is known about genotypic and phenotypic differences that could explain how or why HIV-1 subtype C is associated with the disastrous epidemic in Southern Africa and its rapid emergence in other countries. Salemi et al. (2005) investigated differences in epidemic behavior between subtypes B and C in Brazil. Using a Maximum Likelihood coalescent-based approach, Salemi et al. estimated that the HIV-1C_BR epidemic expanded twice as fast as Brazilian subtype B or even South African subtype C. Subsequent studies using Bayesian coalescent-based methods also indicate that the mean growth rate of the HIV-1C_BR epidemic (0.70–0.87 year⁻¹) was higher than that estimated for the BBR (0.46–0.56 year⁻¹) and F1BR (0.59–0.69 year⁻¹) epidemics (Bello et al., 2008, Bello et al., 2009 and Bello et al., 2012b). Despite such differences, the confidence intervals of those estimates displayed a great overlap making it unclear whether the initial rate of dissemination of different Brazilian HIV-1 subtypes was or in fact significantly different (Bello et al., 2011).

As reviewed by Ariën et al. (2007), the successful worldwide dissemination of subtype C may be due to this strain being less virulent in comparison with other HIV-1 group M subtypes while maintaining the same transmission efficiency. Subtype C strains predominantly use CCR5 co-receptor and rarely switch to CXCR4 or dual tropic use (Vasan et al., 2006). A switch to CXCR4 usage has been associated with acceleration of disease progression. Furthermore, Abraha et al. (2009) showed in ex vivo experiments that subtype C HIV-1 isolates might have a reduced replicative fitness in peripheral blood mononuclear cells. This HIV-1C feature may reflect in a decrease in CD4 cell count, which supports the findings reported by Silveira et al. (2012) with higher CD4 cell counts in subtype C infected patients compared with patients infected with other HIV-1 subtypes. A lower pathogenicity of HIV-1C would lead to a long asymptomatic period, which would increase opportunity for transmission and favor the epidemic expansion. Although the lower pathogenicity of HIV-1C may have contributed to a rapid dissemination of this subtype in Southern Brazil, the relatively slow rate of expansion of subtype C in other Brazilian regions and South

American countries that share borders with Southern Brazil suggests that factors other than intrinsic viral transmissibility have also shaped the dissemination pattern of HIV subtypes in the Southern Cone.

Among these factors, the viral transmission networks and the availability of host individuals may have had a central role in the forging of the current epidemiological scenario. Several epidemiological studies in Southern Brazil observed a relevant association between HIV-1C and heterosexual individuals of both genders, while HIV-1B was found to be more frequent in men who have sex with men (MSM) (Brindeiro et al., 2003, Gräf et al., 2011, Raboni et al., 2010, Silva et al., 2010, Silveira et al., 2012, Soares et al., 2005 and Toledo et al., 2010). The number of HIV infected people in the heterosexual exposure category has been increasing rapidly in Brazil due to a process of feminization of the AIDS epidemic: in 1990 the man/woman new AIDS cases ratio was 5.4 while in 2009 it had decreased to 1.6 (Brazilian Ministry of Health, 2011). It is conceivable that an efficient introduction of subtype C in the heterosexual group allowed for a rapid spread through a large and expanding network of host individuals. Therefore, segregation of exposure categories and HIV-1 variants may be due to differences in the transmission networks where subtypes were introduced and disseminated at early times, rather than to differences in viral transmissibility. A study assessing the temporal dynamics of HIV-1 subtypes in different exposure categories in Porto Alegre observed a significant increase in subtype C in both heterosexual and MSM individuals over time, resulting in a progressive loss of the association between HIV subtypes and exposure categories (Almeida et al., 2011). A progressive intermixing of the epidemics and the consequent loss of associations between HIV variants and exposure categories appears to be a natural tendency as seen in other countries (Tovanabutra et al., 2004). On the other hand, a study performed in Porto Alegre by Dias et al. (2009) found evidence of a different viral transmission pattern. This study, although based on a small sample size, linked anal sex to clade B transmission in both genders – and not only in the MSM exposure category. These findings suggest that though transmission networks might justify most clade segregation, clade C could be less dependent on high-risk anal sex behavior, a factor that could favor its heterosexual transmission. However, very little is known about differences in

pathogenic potential and transmissibility of HIV-1 subtypes, and further studies are needed to explain this issue.

Another important issue that needs to be addressed is whether HIV-1 subtypes may differ in susceptibility to antiretroviral (ARV) treatment and the rate of fixation of mutations conferring drug resistance. Again, little data is available on the impact of antiretroviral therapy on subtype C and other non-subtype B strains, since HIV-1B are dominant in developed countries and most ARVs were designed for subtype B. Nevertheless, some studies have reported that under drug selective pressure subtype C may acquire drug resistance faster than other strains. K65R mutation confers partial resistance to most nucleoside reverse transcriptase inhibitors (NRTIs) and is selected by NRTIs such as tenofovir (TDF), abacavir (ABC), and didanosine (ddI) (Brenner and Coutinho, 2009). The overall incidence of K65R is very low due to high genetic barriers and fitness constraints, but K65R was observed in 23–30% of subtype C patients failing therapy in Africa who were genotyped (Doualla-Bell et al., 2006 and Hosseinipour et al., 2009). *In vitro* studies, thereafter, demonstrated that two subtype C polymorphisms in 64/65 RT codons accelerate K65R selection leading to low susceptibility to NRTIs (Brenner et al., 2006 and Invernizzi et al., 2009). The lower genetic barrier of subtype C, combined with suboptimal treatment regimens in most poor resource countries, may favor the acquisition of K65R, leading to limitations in treatment.

In contrast, another report suggested a lower rate of accumulation of drug resistance mutations in subtype C (Soares et al., 2007). In this study, patient samples from Porto Alegre were collected and grouped by HIV-1 clade, ARV class, and time of exposure. The results showed a significantly higher frequency of resistance mutations in subtype B infected patients compared with subtype C. These findings were observed for protease inhibitors (PI) and NRTI, but not for non-NRTI, and were later supported by Munerato et al. (2010) who found less resistance mutations to PI and NRTI in subtype C infected patients. Such observations were somehow unexpected, undermining concerns that non-subtype B infected individuals are at a disadvantage because ARV drugs have been designed primarily for subtype B infections. Reinforcing this point, Scherrer et al. (2011) observed an improved virological outcome in non-subtype B infected patients on ARV therapy from the Swiss HIV Cohort Study. Although subtype C infected

patients only accounted for 12% of the non-B subtype individuals in the Swiss Cohort, these findings are consistent with a lower acquisition of drug resistance mutations in subtype C. However, follow-up studies with larger number of individuals are required to elucidate the real impact of ARV therapy on non-subtype B strains (Soares et al., 2012).

Naturally occurring mutations and polymorphism that may impair the effectiveness of entry inhibitors, a new class of ARV drugs, have also been investigated (Araújo et al., 2012 and Gonzalez et al., 2010). These studies indicated that, compared with subtype B, HIV-1C has a higher number of natural accumulated polymorphisms that may cause resistance to entry inhibitors. So far, the Food and Drug Administration have approved only two entry inhibitors: Maraviroc, a negative allosteric modulator of the CCR5 cell receptor, and Efavirtide, which blocks gp41 protein conformational changes. In Brazil, entry inhibitors are still seldom used except in cases of multidrug virological failure, even though the genetic background of HIV-1C may reduce susceptibility to this class of antiretroviral drug. In vitro and in vivo studies are still required to confirm such findings, and careful planning is needed before employing this class of ARV as rescue therapy.

2.10 - Concluding remarks

The global HIV pandemic is characterized by dynamism and complexity, and the Brazilian experience with the introduction of a new variant to a pre-defined epidemic can yield valuable data regarding HIV expansion and interactions between different strains. Presently, subtype C is responsible for more than half of all global HIV infections, mainly in South and East Africa, and India. Therefore, Southern Brazil represents an important site for epidemiological studies, where HIV-1C (the most prevalent) and HIV-1B (the most studied) co-exist in similar frequencies and infected individuals who have access to antiretroviral treatment and receive follow-up in a reasonable health care system from an HIV care stand point. Despite some obscure issues, studies have shown that HIV-1 subtype C was introduced in Brazil between 1960 and 1980 somewhere in the Southern region, with evidence pointing to the state of Paraná, and originated from an East African country. This strain was very successful in spreading among heterosexual individuals, but not limited to them, becoming highly prevalent in the Brazilian

region that is most heavily affected by AIDS. Whether HIV-1C is the cause of the most disturbing scenario of AIDS epidemic in Brazil, or whether it just benefited from the conditions present in the site where it was introduced, is still up for debate. The recent emergence of subtype C infections in other Brazilian regions and South American countries signals the possibility of continental expansion, although it seems that HIV-1C is not spreading as quickly in these regions as in Southern Brazil. The understanding of the dynamics and diversity of the HIV-1 epidemic in Southern Brazil is of utmost importance for designing new public health policies, affecting prevention, treatment, and vaccine development. Therefore, continued epidemiologic surveillance should be maintained in these regions as well as further evolutionary studies and a close clinical monitoring of infected patients to ensure a better understanding of dissemination within a population, differences in virulence, and transmissibility of the different HIV circulating strains.

2.11 - Acknowledgments

The authors wish to thank Dr. Gonzalo Bello for critically reading and improving the manuscript, and the CNPq, the FAPESC, and the CAPES for the financial support.

2.12 – References

Abecasis, A.B., Martins, A., Costa, I., Carvalho, A.P., Diogo, I., Gomes, P., Camacho, R.J., Group, P.H.R.S., 2011. Molecular epidemiological analysis of paired pol/env sequences from Portuguese HIV type 1 patients. *AIDS Res. Hum. Retroviruses* 27, 803–805.

Abraha, A., Nankya, I.L., Gibson, R., Demers, K., Tebit, D.M., Johnston, E., Katzenstein, D., Siddiqui, A., Herrera, C., Fischetti, L., Shattock, R.J., Arts, E.J., 2009. 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, 5592–5605.

Aguayo, N., Laguna-Torres, V.A., Villafane, M., Barboza, A., Sosa, L., Chauca, G., Carrion, G., Coenca, B., Perez, J., Galeano, A., Bautista, C.T., Sanchez, J.L., Carr, J.K., Kochel, T., 2008. 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 Paraguary. *Rev. Soc. Bras. Med. Trop.* 41, 225–231.

Alcalde, R., Guimaraes, M.L., Duarte, A.J., Casseb, J., 2012. Clinical, epidemiological and molecular features of the HIV-1 subtype C and recombinant forms that are circulating in the city of São Paulo, Brazil. *Viol. J.* 9, 156.

Almeida, S.E.M., Medeiros, R.M., Junqueira, D.M., Gräf, T., Bello, G., Passaes, C.P.B., Morgado, M., Guimarães, M.L., 2011. Temporal dynamics of HIV-1 circulating subtypes in different exposure categories in the city of Porto Alegre, Brazil. In: IX Brazilian Symposium of HIV/AIDS research [SIMPAIDS]. Salvador, Brazil, 1–3 September 2011.

Araújo, L.A., Junqueira, D.M., de Medeiros, R.M., Matte, M.C., Almeida, S.E., 2012. Naturally occurring resistance mutations to HIV-1 entry inhibitors in subtypes B, C, and CRF31_BC. *J. Clin. Virol.* 54, 6–10.

Ariën, K.K., Vanham, G., Arts, E.J., 2007. Is HIV-1 evolving to a less virulent form in humans? *Nat. Rev. Microbiol.* 5, 141–151.

Arruda, E., Simões, L., Sucupira, C., Medeiros, M., Diaz, R.S., Lima, A., 2011. Short communication: intermediate prevalence of HIV type 1 primary antiretroviral resistance in Ceará State, Northeast Brazil. *AIDS Res. Hum. Retroviruses* 27, 153–156.

Aulicino, P.C., Kopka, J., Mangano, A.M., Rocco, C., Iacono, M., Bologna, R., Sen, L., 2005a. Circulation of novel HIV type 1A, B/C, and F subtypes in Argentina. *AIDS Res. Hum. Retroviruses* 21, 158–164.

Aulicino, P.C., Kopka, J., Rocco, C., Mangano, A., Sen, L., 2005b. Sequence analysis of a South American HIV type 1 BC recombinant. *AIDS Res. Hum. Retroviruses* 21, 894–896.

Bello, G., Afonso, J.M., Morgado, M.G., 2012b. Phylodynamics of HIV-1 subtype F1 in Angola, Brazil and Romania. *Infect. Genet. Evol.* 12, 1079–1086.

- Bello, G., Guimarães, M.L., Passaes, C.P., Matos Almeida, S.E., Veloso, V.G., Morgado, M.G., 2009. Short communication: Evidences of recent decline in the expansion rate of the HIV type 1 subtype C and CRF31_BC epidemics in southern Brazil. *AIDS Res. Hum. Retroviruses* 25, 1065–1069.
- Bello, G., Passaes, C.P., Guimarães, M.L., Lorete, R.S., Matos Almeida, S.E., Medeiros, R.M., Alencastro, P.R., Morgado, M.G., 2008. Origin and evolutionary history of HIV-1 subtype C in Brazil. *AIDS* 22, 1993–2000.
- Bello, G., Soares, M.A., Schrago, C.G., 2011. The use of bioinformatics for studying HIV evolutionary and epidemiological history in South America. *AIDS Res. Treat* 154945, 1–13.
- Bello, G., Zanotto, P.M., Iamarino, A., Gräf, T., Pinto, A.R., Couto-Fernandez, J.C., Morgado, M.G., 2012a. Phylogeographic analysis of HIV-1 subtype C dissemination in Southern Brazil. *PLoS One* 7, e35649.
- Brazilian Institute of Geography and Statistics [IBGE], 2012. Census—2010 [in Portuguese]. Available from: [/http://www.censo2010.ibge.gov.br/S](http://www.censo2010.ibge.gov.br/S). (accessed 04.15.2012).
- Brazilian Ministry of Health, 2011. AIDS Epidemiological Bulletin January–June 2011; Year VIII, nu 01 [in Portuguese]. Available from: [/http://www.aids.gov.br/publicacao/2011/boletim_epidemiologico_2011S](http://www.aids.gov.br/publicacao/2011/boletim_epidemiologico_2011S). (accessed 04.15.2012).
- Brenner, B.G., Coutsinos, D., 2009. The K65R mutation in HIV-1 reverse transcriptase: genetic barriers, resistance profile and clinical implications. *HIV Ther.* 3, 583–594.
- Brenner, B.G., Oliveira, M., Doualla-Bell, F., Moisi, D.D., Ntemgwa, M., Frankel, F., Essex, M., Wainberg, M.A., 2006. HIV-1 subtype C viruses rapidly develop K65R resistance to tenofovir in cell culture. *AIDS* 20, F9–13.
- Brígido, L.F., 2009. On the origin of South America HIV-1 C epidemic. *AIDS* 23, 543–544.

Brígido, L.F., Ferreira, J.L., Almeida, V.C., Rocha, S.Q., Ragazzo, T.G., Estevam, D.L., Rodrigues, R., Workgroup, S.P.H.S., 2011. Southern Brazil HIV type 1 C expansion into the state of São Paulo, Brazil. *AIDS Res. Hum. Retroviruses* 27, 339–344.

Brígido, L.F., Nunes, C.C., Oliveira, C.M., Knoll, R.K., Ferreira, J.L., Freitas, C.A., Alves, M.A., Dias, C., Rodrigues, R., Program, R.C., 2007. 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* 23, 1579–1586.

Brindeiro, R.M., Diaz, R.S., Sabino, E.C., Morgado, M.G., Pires, I.L., Brígido, L., Dantas, M.C., Barreira, D., Teixeira, P.R., Tanuri, A., Surveillance, B.N.f.D.R., 2003. Brazilian network for HIV drug resistance surveillance (HIV-BResNet): A survey of chronically infected individuals. *AIDS* 17, 1063–1069.

Cardoso, L.P., da Silveira, A.A., Francisco, R.B., da Guarda Reis, M.N., Stefani, M.M., 2011. Molecular characteristics of HIV type 1 infection among prisoners from Central Western Brazil. *AIDS Res. Hum. Retroviruses* 27, 1349–1353.

Carrion, G., Eyzaguirre, L., Montano, S.M., Laguna-Torres, V., Serra, M., Aguayo, N., Avila, M.M., Ruchansky, D., Pando, M.A., Vinales, J., Perez, J., Barboza, A., Chauca, G., Romero, A., Galeano, A., Blair, P.J., Weissenbacher, M., Birx, D.L., Sanchez, J.L., Olson, J.G., Carr, J.K., 2004. Documentation of subtype C HIV Type 1 strains in Argentina, Paraguay, and Uruguay. *AIDS Res. Hum. Retroviruses* 20, 1022–1025.

Carvalho, B.C., Cardoso, L.P., Damasceno, S., Stefani, M.M., 2011. Moderate prevalence of transmitted drug resistance and interiorization of HIV type 1 subtype C in the inland North State of Tocantins, Brazil. *AIDS Res. Hum. Retroviruses* 27, 1081–1087.

Castro, C.A., Grinsztejn, B., Veloso, V.G., Bastos, F.I., Pilotto, J.H., Morgado, M.G., 2010. Prevalence, estimated HIV-1 incidence and viral diversity among people seeking voluntary counseling and testing services in Rio de Janeiro, Brazil. *BMC Infect. Dis.* 10, 224.

Castro, E., Moreno, M., Deibis, L., de Pérez, G., Salmen, S., Berrueta, L., 2005. Trends of HIV-1 molecular epidemiology in Venezuela: introduction of subtype C and identification of a novel B/C mosaic genome. *J. Clin. Virol.* 32, 257–258.

Chibo, D., Birch, C., 2012. Increasing diversity of human immunodeficiency virus Type 1 subtypes circulating in Australia. *AIDS Res. Hum. Retroviruses* 28, 578–583.

Dias, C.F., Nunes, C.C., Freitas, I.O., Lamego, I.S., Oliveira, I.M., Gilli, S., Rodrigues, R., Brigido, L.F., 2009. 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, 191–196.

Dilernia, D.A., Gomez, A.M., Lourtau, L., Marone, R., Losso, M.H., Salomón, H., Gómez-Carrillo, M., 2007. HIV type 1 genetic diversity surveillance among newly diagnosed individuals from 2003 to 2005 in Buenos Aires, Argentina. *AIDS Res. Hum. Retroviruses* 23, 1201–1207.

Doualla-Bell, F., Avalos, A., Brenner, B., Gaolathe, T., Mine, M., Gaseitsiwe, S., Oliveira, M., Moisi, D., Ndwapi, N., Moffat, H., Essex, M., Wainberg, M.A., 2006. High prevalence of the K65R mutation in human immunodeficiency virus type 1 subtype C isolates from infected patients in Botswana treated with didanosine-based regimens. *Antimicrob. Agents Chemother.* 50, 4182–4185.

Dougan, S., Gilbert, V.L., Sinka, K., Evans, B.G., 2005. HIV infections acquired through heterosexual intercourse in the United Kingdom: findings from national surveillance. *BMJ* 330, 1303–1304.

Ferreira, A.S., Cardoso, L.P., Stefani, M.M., 2011. Moderate prevalence of transmitted drug resistance and high HIV-1 genetic diversity in patients from Mato Grosso State, Central Western Brazil. *J. Med. Virol.* 83, 1301–1307.

Ferreira, J.L., Thomaz, M., Rodrigues, R., Harrad, D., Oliveira, C.M., Oliveira, C.A., Batista, J.P., Ito, T.S., Brigido, L.F., 2008. 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, 800–808.
- Fontella, R., Soares, M.A., Schrago, C.G., 2008. On the origin of HIV-1 subtype C in South America. *AIDS* 22, 2001–2011.
- Fontella, R., Soares, M.A., Schrago, C.G., 2009. The origin of South American HIV-1 subtype C: lack of evidence for a Mozambican ancestry. *AIDS* 23, 1926–1928.
- Gadelha, S.R., Shindo, N., Cruz, J.N., Morgado, M.G., Galvão-Castro, B., 2003. Molecular epidemiology of human immunodeficiency virus-1 in the state of Ceará, Northeast, Brazil. *Mem. Inst. Oswaldo Cruz* 98, 461–463.
- Gonzalez, S., Gondwe, C., Tully, D.C., Minhas, V., Shea, D., Kankasa, C., M'soka, T., Wood, C., 2010. Short communication: antiretroviral therapy resistance mutations present in the HIV type 1 subtype C pol and env regions from therapy-naïve patients in Zambia. *AIDS Res. Hum. Retroviruses* 26, 795–803.
- Gräf, T., Passaes, C.P., Ferreira, L.G., Grisard, E.C., Morgado, M.G., Bello, G., Pinto, A.R., 2011. 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.* 51, 186–191.
- Guimarães, M.L., Eyer-Silva, W.A., Couto-Fernandez, J.C., Morgado, M.G., 2008. Identification of two new CRF_BF in Rio de Janeiro State, Brazil. *AIDS* 22, 433–435.
- Hosseini-pour, M.C., van Oosterhout, J.J., Weigel, R., Phiri, S., Kamwendo, D., Parkin, N., Fiscus, S.A., Nelson, J.A., Eron, J.J., Kumwenda, J., 2009. The public health approach to identify antiretroviral therapy failure: high-level nucleoside reverse transcriptase inhibitor resistance among Malawians failing first-line antiretroviral therapy. *AIDS* 23, 1127–1134.
- Inocencio, L.A., Pereira, A.A., Sucupira, M.C., Fernandez, J.C., Jorge, C.P., Souza, D.F., Fink, H.T., Diaz, R.S., Becker, I.M., Suffert, T.A., Arruda, M.B., Macedo, O., Simão, M.B., Tanuri, A., 2009. Brazilian network for HIV drug resistance surveillance: a survey of individuals recently diagnosed with HIV. *J. Int. AIDS Soc.* 12, 20.

- Invernizzi, C.F., Coutsinos, D., Oliveira, M., Moisi, D., Brenner, B.G., Wainberg, M.A., 2009. Signature nucleotide polymorphisms at positions 64 and 65 in reverse transcriptase favor the selection of the K65R resistance mutation in HIV-1 subtype C. *J. Infect. Dis.* 200, 1202–1206.
- Jones, L.R., Dileria, D.A., Manrique, J.M., Moretti, F., Salomón, H., Gomez-Carrillo, M., 2009. In-depth analysis of the origins of HIV type 1 subtype C in South America. *AIDS Res. Hum. Retroviruses* 25, 951–959.
- Locateli, D., Stoco, P.H., de Queiroz, A.T., Alcântara, L.C., Ferreira, L.G., Zanetti, C.R., Rodrigues, R., Grisard, E.C., Pinto, A.R., 2007. Molecular epidemiology of HIV-1 in Santa Catarina State confirms increases of subtype C in Southern Brazil. *J. Med. Virol.* 79, 1455–1463.
- Machado, L.F., Ishak, M.O., Vallinoto, A.C., Lemos, J.A., Azevedo, V.N., Moreira, M.R., Souza, M.I., Fernandes, L.M., Souza, L.L., Ishak, R., 2009. 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* 25, 961–966.
- Martínez, A.M., Barbosa, E.F., Ferreira, P.C., Cardoso, F.A., Silveira, J., Sassi, G., da Silva, C.M., Mendonça-Signorini, V., Antunes, C.M., 2002. Molecular epidemiology of HIV-1 in Rio Grande, RS, Brazil. *Rev. Soc. Bras. Med. Trop.* 35, 471–476.
- Medeiros, R.M., Junqueira, D.M., Matte, M.C., Barcellos, N.T., Chies, J.A., Matos Almeida, S.E., 2011. 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, 1682–1688.
- Montano, S.M., Sanchez, J.L., Laguna-Torres, A., Cuchi, P., Avila, M.M., Weissenbacher, M., Serra, M., Vinholes, J., Russi, J.C., Aguayo, N., Galeano, A.H., Gianella, A., Andrade, R., Arredondo, A., Ramirez, E., Acosta, M.E., Alava, A., Montoya, O., Guevara, A., Manrique, H., Lama, J.R., de la Hoz, F., Sanchez, G.I., Ayala, C., Pacheco, M.E., Carrion, G., Chauca, G., Perez, J.J., Negrete, M., Russell, K.L.,

Bautista, C.T., Olson, J.G., Watts, D.M., Birx, D.L., Carr, J.K., Group, f.t.S.A.H.M.S.W., 2005. Prevalences, genotypes, and risk factors for HIV transmission in South America. *J. Acquir. Immune. Defic. Syndr.* 40, 57–64.

Monteiro, J.P., Ferraro, G.A., Oliveira, T., Goldani, L.Z., Kashima, S., Alcantara, L.C., Morgado, M.G., Bou-Habib, D.C., Galvão-Castro, B., 2007. Genetic and biologic characterization of HIV type 1 subtype C isolates from south Brazil. *AIDS Res. Hum. Retroviruses* 23, 135–143.

Munerato, P., Sucupira, M.C., Oliveros, M.P., Janini, L.M., de Souza, D.F., Pereira, A.A., Inocencio, L.A., Diaz, R.S., 2010. HIV type 1 antiretroviral resistance mutations in subtypes B, C, and F in the City of São Paulo, Brazil. *AIDS Res. Hum. Retroviruses* 26, 265–273.

Ng, O.T., Eyzaguirre, L.M., Carr, J.K., Chew, K.K., Lin, L., Chua, A., Leo, Y.S., Redd, A.D., Quinn, T.C., Laeyendecker, O., 2012. Identification of new CRF51_01B in Singapore using full genome analysis of three HIV type 1 isolates. *AIDS Res. Hum. Retroviruses* 28, 527–530.

Oliveira, T., Pillay, D., Gifford, R.J., Resistance, U.C.G.o.H.D., 2010. The HIV-1 subtype C epidemic in South America is linked to the United Kingdom. *PLoS One* 5, e9311.

Osmanov, S., Heyward, W.L., Esparza, J., 1994. The World Health Organization network for HIV isolation and characterization: summary of a pilot study. *AIDS Res. Hum. Retroviruses* 10, 1325–1326.

Pando, M.A., Eyzaguirre, L.M., Segura, M., Bautista, C.T., Marone, R., Ceballos, A., Montano, S.M., Sánchez, J.L., Weissenbacher, M., Avila, M.M., Carr, J.K., 2006. First report of an HIV-1 triple recombinant of subtypes B, C and F in Buenos Aires, Argentina. *Retrovirology* 3, 59.

Pando, M.A., Gómez-Carrillo, M., Vignoles, M., Rubio, A.E., dos Ramos Farias, M.S., Vila, M., Rossi, D., Ralón, G., Marone, R., Reynaga, E., Sosa, J., Torres, O., Maestri, M., Avila, M.M., Salomón, H., 2011. 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, 17–23.
- Passaes, C.P., Bello, G., Lorete, R.S., Matos Almeida, S.E., Junqueira, D.M., Veloso, V.G., Morgado, M.G., Guimarães, M.L., 2009. Genetic characterization of HIV-1 BC recombinants and evolutionary history of the CRF31_BC in Southern Brazil. *Infect. Genet. Evol.* 9, 474–482.
- Pedroso, C., Queiroz, A.T., Alcântara, L.C., Drexler, J.F., Diaz, R.S., Weyll, N., Brites, C., 2007. 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.* 45, 251–253.
- Pinto, A.R., Petry, A., Gräf, T., Vandresen, R., Kupek, E., 2012. Case report of a haemovigilance investigation using phylogenetic analysis of HIV-1 in Brazil. *Transfus. Med.* 22, 57–62.
- Plantier, J.C., Leoz, M., Dickerson, J.E., De Oliveira, F., Cordonnier, F., Lemée, V., Damond, F., Robertson, D.L., Simon, F., 2009. A new human immunodeficiency virus derived from gorillas. *Nat. Med.* 15, 871–872.
- Raboni, S.M., Almeida, S.M., Rotta, I., Ribeiro, C.E., Rosario, D., Vidal, L.R., Nogueira, M.B., Riedel, M., Winhescki, M.a.G., Ferreira, K.A., Ellis, R., 2010. Molecular epidemiology of HIV-1 clades in Southern Brazil. *Mem. Inst. Oswaldo Cruz* 105, 1044–1049.
- Rangel, H.R., Garzaro, D., Gutiérrez, C.R., Vásquez, L., Guillen, G., Torres, J.R., Pujol, F.H., 2009. HIV diversity in Venezuela: predominance of HIV type 1 subtype B and genomic characterization of non-B variants. *AIDS Res. Hum. Retroviruses* 25, 347–350.
- Rangel, H.R., Maes, M., Villalba, J., Sulbarán, Y., de Waard, J.H., Bello, G., Pujol, F.H., 2012. Evidence of at least two introductions of HIV-1 in the Amerindian Warao population from Venezuela. *PLoS One* 7, e40626.
- Rodrigues, R., Manenti, S., Romao, P.R., de Paula Ferreira, J.L., Batista, J.P., Siqueira, A.F., de Macedo Brigido, L.F., 2010. Young pregnant women living with HIV/ AIDS in Criciúma, Southern Brazil, are infected almost exclusively with HIV type 1 clade C. *AIDS Res. Hum. Retroviruses* 26, 351–357.

- Rodrigues, R., Scherer, L.C., Oliveira, C.M., Franco, H.M., Sperhacker, R.D., Ferreira, J.L., Castro, S.M., Stella, I.M., Brigido, L.F., 2006. 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.* 116, 201–207.
- Ruchansky, D., Casado, C., Russi, J.C., Arbiza, J.R., Lopez-Galindez, C., 2009. Identification of a new HIV Type 1 circulating recombinant form (CRF38_BF1) in Uruguay. *AIDS Res. Hum. Retroviruses* 25, 351–356.
- Sa Filho, D.J., Sanabani, S., Diaz, R.S., Munerato, P., Brunstein, A., Fusuma, E., Sabino, E.C., Janini, L.M., 2005. 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* 21, 145–151.
- Sa Filho, D.J., Sucupira, M.C., Caseiro, M.M., Casiero, M.M., Sabino, E.C., Diaz, R.S., Janini, L.M., 2006. Identification of two HIV type 1 circulating recombinant forms in Brazil. *AIDS Res. Hum. Retroviruses* 22, 1–13.
- Salemi, M., de Oliveira, T., Soares, M.A., Pybus, O., Dumans, A.T., Vandamme, A.M., Tanuri, A., Cassol, S., Fitch, W.M., 2005. Different epidemic potentials of the HIV-1B and C subtypes. *J. Mol. Evol.* 60, 598–605.
- Sanabani, S., Neto, W.K., de Sa Filho, D.J., Diaz, R.S., Munerato, P., Janini, L.M., Sabino, E.C., 2006. Full-length genome analysis of human immunodeficiency virus type 1 subtype C in Brazil. *AIDS Res. Hum. Retroviruses* 22, 171–176.
- Santos, A.F., Schrago, C.G., Martinez, A.M., Mendoza-Sassi, R., Silveira, J., Sousa, T.M., Lengrubler, R.B., Soares, E.A., Sprinz, E., Soares, M.A., 2007. Epidemiologic and evolutionary trends of HIV-1 CRF31_BC-related strains in southern Brazil. *J. Acquir. Immune Defic. Syndr.* 45, 328–333.
- Santos, A.F., Sousa, T.M., Soares, E.A., Sanabani, S., Martinez, A.M., Sprinz, E., Silveira, J., Sabino, E.C., Tanuri, A., Soares, M.A., 2006.

Characterization of a new circulating recombinant form comprising HIV-1 subtypes C and B in southern Brazil. *AIDS* 20, 2011–2019.

Scherrer, A.U., Ledergerber, B., von Wyl, V., Böni, J., Yerly, S., Klimkait, T., Bürgisser, P., Rauch, A., Hirschel, B., Cavassini, M., Elzi, L., Vernazza, P.L., Bernasconi, E., Held, L., Günthard, H.F., Study, S.H.C., 2011. Improved virological outcome in white patients infected with HIV-1 non-B subtypes compared to subtype B. *Clin. Infect. Dis.* 53, 1143–1152.

Schuelter-Trevisol, F., da Silva, M.V., Oliveira, C.M., Rodrigues, R., 2007. HIV genotyping among female sex workers in the State of Santa Catarina. *Rev. Soc. Bras. Med. Trop.* 40, 259–263.

Silva, M.M., Telles, F.Q., da Cunha, C.A., Rhame, F.S., 2010. HIV subtype, epidemiological and mutational correlations in patients from Paraná, Brazil. *Braz. J. Infect. Dis.* 14, 495–501.

Silveira, J., Santos, A.F., Martínez, A.M., Góes, L.R., Mendoza-Sassi, R., Muniz, C.P., Tupinambás, U., Soares, M.A., Greco, D.B., 2012. Heterosexual transmission of human immunodeficiency virus type 1 subtype C in southern Brazil. *J. Clin. Virol.* 54, 36–41.

Simon, D., Béria, J.U., Tietzmann, D.C., Carli, R., Stein, A.T., Lunge, V.R., 2010. Prevalence of HIV-1 subtypes in patients of an urban center in Southern Brazil. *Rev. Saú de Pú blica* 44, 1094–1101.

Soares, E.A., Martínez, A.M., Souza, T.M., Santos, A.F., Da Hora, V., Silveira, J., Bastos, F.I., Tanuri, A., Soares, M.A., 2005. HIV-1 subtype C dissemination in southern Brazil. *AIDS* 19 (Suppl. 4), S81–S86.

Soares, E.A., Santos, A.F., Sousa, T.M., Sprinz, E., Martinez, A.M., Silveira, J., Tanuri, A., Soares, M.A., 2007. Differential drug resistance acquisition in HIV-1 of subtypes B and C. *PLoS One* 2, e730.

Soares, E.A., Santos, R.P., Pellegrini, J.A., Sprinz, E., Tanuri, A., Soares, M.A., 2003. Epidemiologic and molecular characterization of human immunodeficiency virus type 1 in southern Brazil. *J. Acquir. Immune Defic. Syndr.* 34, 520–526.

- Soares, E.A., Santos, A.F., Soares, M.A., 2012. HIV-1 subtype and virological response to antiretroviral therapy: acquired drug resistance. *Clin. Infect. Dis.* 54, 738–739.
- Soares, M.A., De Oliveira, T., Brindeiro, R.M., Diaz, R.S., Sabino, E.C., Brigido, L., Pires, I.L., Morgado, M.G., Dantas, M.C., Barreira, D., Teixeira, P.R., Cassol, S., Tanuri, A., Surveillance, B.N.f.D.R., 2003. A specific subtype C of human immunodeficiency virus type 1 circulates in Brazil. *AIDS* 17, 11–21.
- Stefani, M.M., Pereira, G.A., Lins, J.A., Alcantara, K.C., Silveira, A.A., Viegas, A.A., Maya, N.C., Mussi, A.H., 2007. Molecular screening shows extensive HIV-1 genetic diversity in Central West Brazil. *J. Clin. Virol.* 39, 205–209.
- Tatt, I.D., Barlow, K.L., Clewley, J.P., Gill, O.N., Parry, J.V., 2004. Surveillance of HIV-1 subtypes among heterosexuals in England and Wales, 1997–2000. *J. Acquir. Immune Defic. Syndr.* 36, 1092–1099.
- Tebit, D.M., Arts, E.J., 2011. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect. Dis.* 11, 45–56.
- Teixeira, S.L., Bastos, F.I., Telles, P.R., Hacker, M.A., Brigido, L.F., de F Oliveira, C.A., Bongertz, V., Morgado, M.G., 2004. HIV-1 infection among injection and ex-injection drug users from Rio de Janeiro, Brazil: prevalence, estimated incidence and genetic diversity. *J. Clin. Virol.* 31, 221–226.
- Toledo, P.V., Carvalho, D.S., Rossi, S.G., Brindeiro, R., de Queiroz-Telles, F., 2010. Genetic diversity of human immunodeficiency virus-1 isolates in Paraná, Brazil. *Braz. J. Infect. Dis.* 14, 230–236.
- Tovanabutra, S., Beyrer, C., Sakkhachornphop, S., Razak, M.H., Ramos, G.L., Vongchak, T., Rungruengthanakit, K., Saokhieo, P., Tejafong, K., Kim, B., De Souza, M., Robb, M.L., Birx, D.L., Jittiwutikarn, J., Suriyanon, V., Celentano, D.D., McCutchan, F.E., 2004. The changing molecular epidemiology of HIV type 1 among northern Thai drug users, 1999 to 2002. *AIDS Res. Hum. Retroviruses* 20, 465–475.

Vasan, A., Renjifo, B., Hertzmark, E., Chaplin, B., Msamanga, G., Essex, M., Fawzi, W., Hunter, D., 2006. Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. *Clin. Infect. Dis.* 42, 843–852.

Véras, N.M., Gray, R.R., Brígido, L.F., Rodrigues, R., Salemi, M., 2011. High-resolution phylogenetics and phylogeography of human immunodeficiency virus type 1 subtype C epidemic in South America. *J. Gen. Virol.* 92, 1698–1709.

Westin, M.R., Biscione, F.M., Fonseca, M., Ordonez, M., Rodrigues, M., Greco, D.B., Tupinambas, U., 2011. Resistance-associated mutation prevalence according to subtypes B and non-B of HIV type 1 in antiretroviral-experienced patients in Minas Gerais, Brazil. *AIDS Res. Hum. Retroviruses* 27, 981–987.

WHO Network for HIV Isolation and Characterization, 1994. HIV type 1 variation in World Health Organization-sponsored vaccine evaluation sites: genetic screening, sequence analysis, and preliminary biological characterization of selected viral strains. *AIDS Res. Hum. Retroviruses* 10, 1327–1343.

World Health Organization [WHO], 2010. Global Report: UNAIDS Report on the Global AIDS Epidemic 2010. Available from: http://www.unaids.org/globalreport/Global_report.htm. (accessed 03.17.2012).

CAPÍTULO 3.

Comprehensive characterization of the HIV-1 molecular epidemiology and demographic history in the Brazilian region most heavily affected by AIDS

Artigo submetido ao *Journal of Virology*

Tiago Gräf, Hegger Machado Fritsch, Rúbia Marília de Medeiros,
Dennis Maletich Junqueira, Sabrina Esteves de Matos Almeida e
Aguinaldo Roberto Pinto

3.1 - Abstract

High incidence of AIDS cases and dominance of HIV-1 subtype C infections are two features that distinguish HIV-1 epidemic in the two southernmost Brazilian states (RS and SC) from the epidemic in other parts of the country. Nevertheless, studies on HIV molecular epidemiology were mainly conducted in capital cities and more comprehensive understanding of factors driven this unique epidemic in Brazil is necessary. Blood samples were collected from 13 municipalities in Brazilian southern region. HIV-1 *env* and *pol* genes were submitted to phylogenetic analyses for assignment of subtype and viral population phylogenetics were reconstructed applying skyride and logistic coalescent models in a Bayesian analysis. High prevalence of subtype C was observed in all sampled locations, however increased frequency of recombinant strains was found in RS, with evidences for new CRFs. In SC state subtype B and C epidemics were associated to distinct transmission chains. In both locations subtype C epidemic was showed to be in a faster expanding rate than subtype B epidemic. Our results highlight the fast expansion rate of HIV-1C in south Brazil and also discuss how the heterosexual and MSM transmission chains might have impacted in the current prevalence of HIV-1 subtypes in the region.

3.2 - Importance

AIDS epidemic in south Brazil is expanding fast, but the circumstances driven this condition are not well known. High prevalence of HIV-1 subtype C was reported in the capital cities of the region, contrasting with the subtype B dominance in the rest of the country. This study sought to comparatively investigate HIV-1 subtype B and C epidemic by sampling several cities in the two states with the highest AIDS incidence in Brazil. Our analyses showed distinct expansion rates for the two epidemics and we also found evidences suggesting that separate transmission chains can be impacting in the viral phylogenetic and in the emergence of new recombinant forms. These findings highlight the fast growing subtype C epidemic in the region and also provide insights about the potential of expanding throughout Brazil and in other parts of the globe where HIV-1 subtype C was introduced.

3.3 - Introduction

High molecular diversity is one of the main features of HIV-1. Due to the very fast replication rate and to the lack of a proofreading activity of the viral reverse transcriptase (RT) enzyme, HIV-1 group M – the pandemic strain - has evolved into a diverse number of variants (1). Among 9 HIV-1 group M subtypes, subtype C (HIV-1C) is responsible for more than 50% of the worldwide infections, occurring mostly in Southern and East African countries and India. Subtype B (HIV-1B), in turn, is widely dispersed in the world, being the dominant strain in Europe and in the Americas, but with a world prevalence of only 10% (1, 2). HIV-1 high variability is also consequence of genetic recombination process which can occur in unique forms (URF), when less than three individuals are infected, or in circulating forms (CRF) when three individuals not epidemiologically related are infected with the same mosaic strain (3). Currently more than 60 CRFs have been described, encompassing around 20% of the HIV-1 infections around the world (<http://www.hiv.lanl.gov> - accessed in Feb, 2015).

The HIV-1 molecular epidemic in Brazil is mainly driven by subtype B, with minor circulation of subtype F1, BF1 recombinant forms and subtype C (4). However, in the two southernmost states - Rio Grande do Sul (RS) and Santa Catarina (SC) - a large prevalence of HIV-1C is observed (40-80%), followed by BC mosaic forms (20-30%, including the frequency of CRF31_BC) and subtype B (15-30%) (5). HIV-1C prevalence decreases in northward direction in Brazil, representing up to 30% in Paraná State and between 5-20% in other nearby states. Besides the high prevalence of HIV-1C, RS and SC states registered the highest AIDS incidence rates among the Brazilian states in the last 10 years, notifying around 35 new cases per 100k habitants per year (6). AIDS incidence was also observed to have increased more than two times in countryside cities from RS and SC over 10 years (7). The same process is also observed in other Brazilian states, but in much more slower rate.

A decade of epidemiological studies has described the HIV-1C epidemic in south Brazil, investigating the origin and highlighting its dispersion throughout the country and continent (5, 8-10). However, these studies mainly included samples from the capital cities and very little molecular information is available from countryside locations in

south region. In face of this interesting epidemiological scenario, where high AIDS incidence is observed along with unique molecular features, a better characterization of the HIV-1 epidemic in a broad sampling area is necessary. The present study aims to describe the HIV-1 molecular epidemiology in countryside cities from RS and SC states. To do so we sampled several cities, including border towns, in an effort to construct a comprehensive databank with viral sequences and patients' information. In addition, we analyzed the phylodynamics of HIV-1C and B in the region and showed how the demographic history reconstruction can be informative to explain the current differences in the epidemic observed between the two sampled states.

3.4 - Materials and Methods

3.4.1 - Patients and samples

HIV-1 seropositive patients were invited to enroll this study throughout 13 cities of SC and RS Brazilian states (Figure 3.1). The inclusion criteria were positive diagnosis for HIV infection and no previous exposure to antiretroviral (ARV) treatment. The Universidade Federal de Santa Catarina (UFSC) and the Fundação Estadual de Pesquisa e Produção em Saúde (FEPPS-RS) ethic committees have assessed and approved this study. Informed consent was obtained from all enrolled volunteers.

Blood samples were collected in FTA cards (Whatman Inc.) and dried at room temperature. Following manufacturer instructions, three to five 0.5mm dried blood spot (DBS) punches were washed and dried. HIV-1 provirus DNA was amplified by Nested-PCR directly from the washed punches and primers and reaction conditions were as follows (11). HIV-1 envelope (*env*) (HXB2 6846-7360 bp) and polymerase (*pol*) (HXB2 2274-3545 bp) were the genome target regions. Purified PCR products were sequenced using BigDye Terminator v3.1 Cycle Sequencing (Applied Biosystems) and read by ABI 3100 Genetic Analyzer (Applied Biosystems). The generated sequences were then assembled and visually inspected by using SeqMan software in LaserGene package (DNASTAR). The sequences generated in the present study were deposited in GenBank under accession numbers: KR065788 - KR066336 and KP224476-KP224501.

3.4.2 - Subtype classification

Sequences generated here were submitted to the REGA, RIP and SCUEAL online subtyping tools (12, 13, 14). The subtype assigned by these automated tools was confirmed by phylogenetic tree reconstruction. Subtype reference sequences were downloaded from Los Alamos HIV Sequence Database (www.hiv.lanl.gov) and aligned with the new sequences using Muscle algorithm (15). Maximum Likelihood (ML) tree reconstruction was performed using PhyML with GTR+I+ Γ 4 nucleotide substitution model, as implemented in Seaview program (16). Aiming to identify the recombination breakpoints, *pol* sequences classified as recombinant were submitted to bootscanning analysis using Simplot 3.5.1 software (17).

3.4.3 - Bayesian phylodynamics analyses

In order to better understand the expansion of HIV-1 epidemic in south Brazil, time-scaled phylogenetic tree reconstruction was performed using BEAST/BEAGLE software for HIV-1 subtype B and C *pol* sequences (18). Additional sequences from SC and RS states were downloaded from Los Alamos HIV Database and aligned with the countryside sequences dataset. Aiming to identify Surveillance Drug Resistance Mutations (SDRM), sequences obtained from the public databanks were submitted to CPR tool (19). Codon sites with a SDRM detected in at least two different sequences were deleted in the alignment and sequences that did not pass in the CPR Quality Assessment tool and/or had five or more SDRM detected were excluded. Bayesian skyride coalescent model with an uncorrelated lognormal relaxed clock were used to estimate the changes in viral effective population size (N_e) through time (20). Nucleotide substitutions were modeled with GTR+I+ Γ 4 and informative priors were used on the tMRCA of the tree root for HIV-1C and HIV-1B as previously published (8, 21). Sequences' sampling dates were integrated out over the known sampling time interval when the exact sampling date was not known (22). To estimate the growth rate (r) of the epidemic of HIV-1B and HIV-1C the logistic growth coalescent model (23) was used as the best-fit demographic model, after visually

diagnosed by the skyride reconstruction, in a Bayesian phylogenetic analysis as described above. Epidemic doubling time (λ) was calculated applying growth rate (r) to the formula $\lambda = \ln(2)/r$. Monte Carlo Markov Chains (MCMC) were run sufficiently long to ensure stationary and adequate effective sample size (ESS) for the main parameters. Tracer software (available at: <http://beastbioedacuk/Tracer>) was used to diagnose MCMC, adjust initial burn-in and to perform skyride demographic reconstruction. Due to large differences in HIV-1C and B sample size, Bayesian analyses were repeated with a similar number of sequences for both subtypes. To do so, HIV-1C datasets were down-sampled to around 60 sequences each using CD-HIT online tool (24).

3.4.4 - Statistical analysis

Patients were invited to answer a questionnaire in the moment of enrollment. Information about gender, age, exposition group, ethnicity, education level and marital status were included in the questionnaire, while clinical data was retrieved from patient medical records. Association among discrete variables was evaluated using Chi-square test or Fisher's exact test when appropriate. Statistical significance was achieved when p values were ≤ 0.05 .

3.5 - Results

An amount of 359 samples were collected between October 2009 and February 2014; 163 (45%) samples were from RS state and 196 (55%) samples were from SC state. HIV-1 sequences from *pol* region were obtained for 253 (71%) samples and from *env* region for 284 (79%) samples, totaling 317 (88%) patients with molecular data for *pol*, *env* or both.

Among 317 patients with subtype information, 45% were woman and 55% were male. Mean diagnostic year was May 2010 and mean patients' age was 40 years. Most frequent related HIV-1 exposure group was heterosexual (71%), followed by men who have sex with men (MSM) (20%). Seventy two percent of the enrolled patients were self defined white, while 17% defined themselves as Afro-descendants or Amerindians. See supplementary data (Table S3.1) for the summary of

the main demographic and epidemiological information obtained through patient interview.

Phylogenetic analysis revealed 56% of HIV-1 subtype C infections, 22% of subtype B, 10% of URF_BC, 5% of CRF31_BC, 4% of URF_BF1 and 3% of other viral forms (subtype D, F1, URF_CF1 and complex unique recombinants). In SC we observed higher prevalence of HIV-1C (66% against 44% in RS), while BC recombinant forms (CRF31_BC and URF_BC) are more frequently found in RS state (23% against 10% in SC) (Figure 3.1). HIV-1B has roughly the same prevalence in the two states.

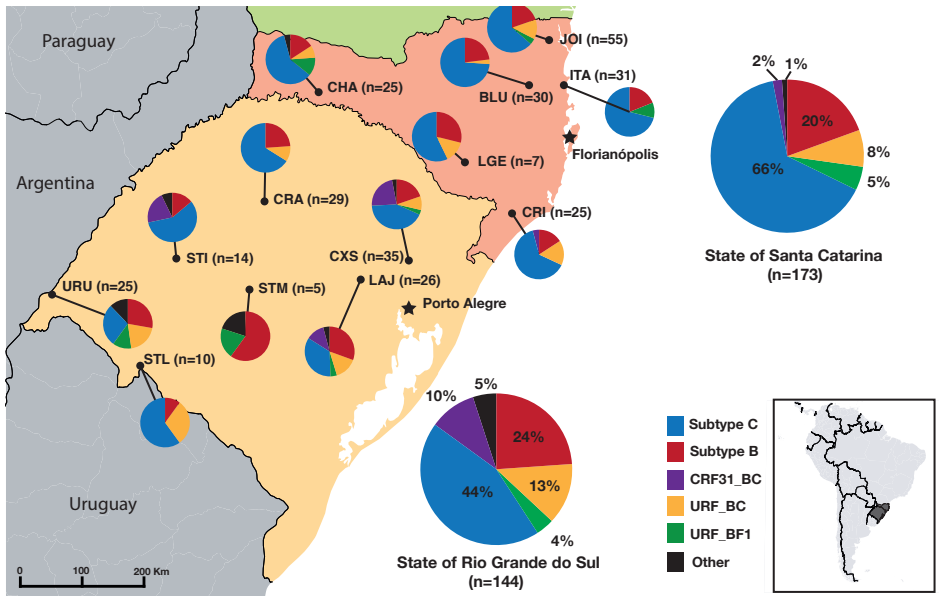


Figure 3.1. HIV-1 molecular diversity in countryside cities from states of Santa Catarina and Rio Grande do Sul, Brazil. The location of sampled cities is showed by dots in the map and the location of capital cities are marked with a black star. Pie chart colors are according to the legend and an inset map shows the location of the sampled region in South America. **Acronyms for cities:** BLU: Blumenau; CHA: Chapecó; CRA: Cruz Alta CRI: Criciúma; CXS: Caxias do Sul; ITA: Itajaí; JOI: Joinville; LAJ: Lajeado; LGE: Lages; STI: Santiago; STL: Santana do Livramento; STM: Santa Maria; URU: Uruguaiana.

Recombinant strains were observed in 21% (66/317) of the patients studied here, being that 77% of them (51/66) were infected by unique forms (URF). Among the URF_BC (n=33), 30% were intergenic recombinants whose recombination breakpoints were not possible to be observed in the *pol* and *env* fragments sequenced here. For the URF_BF1 (n=14), intergenic mosaic strains represented 57%. Some of the URF identified here were close related in phylogenetic trees and also showed similar recombination breakpoints in the bootscanning analysis. Two sequences from Chapecó city were intergenic URF_BF1. They were very close related in the trees, but direct transmission is unlikely once the samples were obtained from two women reporting heterosexual exposure to HIV. Among the intragenic recombinant mosaic forms, three distinct groups of URF_BC strains were identified with similar breakpoints and clustering together in trees (Figure 3.2). They were the sequences LAJ03 and LAJ29 from Lajeado city – named BC_cluster1; URU09 and URU11 from Uruguaiana city – named BC_cluster2; and CRA96, URU05 and URU18 from Cruz Alta and Uruguaiana cities – named BC_cluster3. Heterosexual men and women composed each of these clusters.

A CRF31_BC related recombination pattern was observed for BC_cluster1 and BC_cluster2 sequences (Figure 3.2). The sequences in these clusters share the first recombination breakpoint, between positions 2965 and 2984 relative to the HXB2 strain, with the CRF31_BC. However they have a smaller subtype B insertion in a *pol* subtype C backbone when compared to CRF31_BC. In the BC_cluster2 was still possible to detect another recombination breakpoint, although caution is necessary when interpreting bootscanning analysis for the extremities of the sequences. The BC_cluster3, in turn, shows a distinct mosaic display with a subtype B insertion in the positions close to 2850 and 3060 (HXB2). It is also possible to observe an additional recombination breakpoint close to the beginning of the analyzed fragment.

Statistically significant association was observed among HIV-1 subtypes (constrained to comparisons between HIV-1B and HIV-1C) and epidemiological data (Table 3.1). In SC HIV-1B was associated to MSM exposure group and also to the white ethnical group, HIV-1C in turn was statistically more present among heterosexuals. Exposure group and ethnicity were not related to each other in this location (data not

show). In RS state the only association observed was between HIV-1B infected individuals and ≥ 12 years of education. Interesting to note that in both states MSM group was associated with more years of study while patients from the heterosexual group more frequently declared to have studied up to eight years (data not shown). No statistical association was observed among other HIV-1 forms and demographic/epidemiological patient's data.

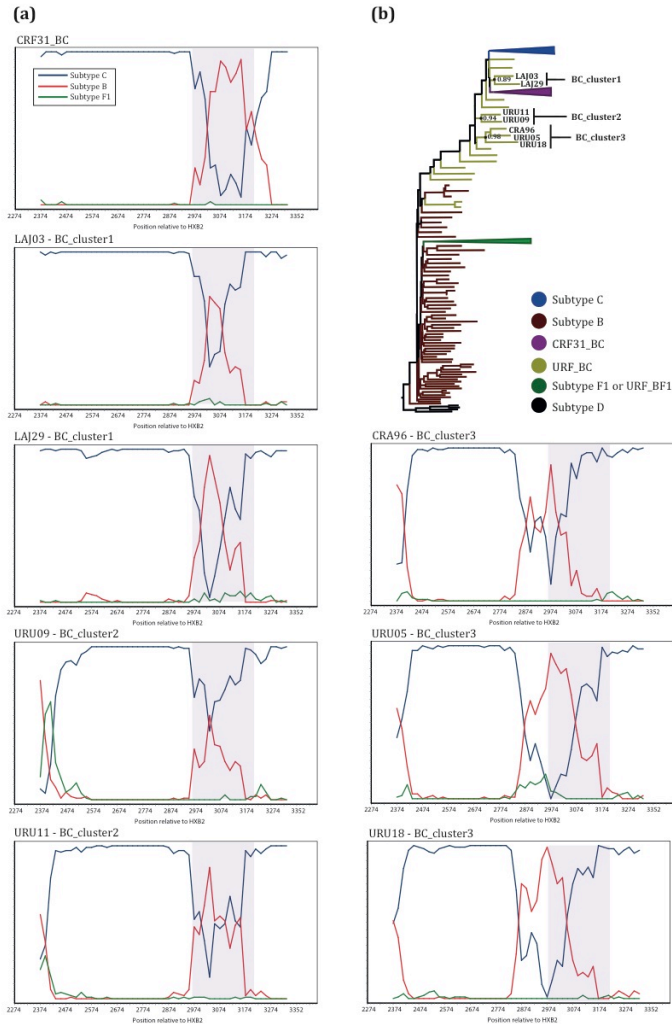


Figure 3.2. Related BC mosaic strains identified in bootscanning and phylogenetic analyzes. a) Bootscanning plots of the sequences compounding BC_cluster1-3 and a CRF31_BC reference sequence (top-left). Purple shade in plots indicates CRF31_BC recombination region. Line colors are according to the legend in the first plot. b) Maximum likelihood tree indicating the position of the sequences compounding BC_cluster1-3 and their respective SH-aLRT node support. Branches are colored according to the legend and some of the subtype clades are condensed.

Table 3.1. Demographic and epidemiological data of the population infected by HIV-1B and HIV-1C in Rio Grande do Sul (RS) and Santa Catarina (SC) states, Brazil .

Characteristic	HIV-1 subtype - RS			HIV-1 subtype - SC		
	B	C	Total	B	C	Total
Gender						
Female	15 (30%)	35 (70%)	50	11 (19%)	46 (81%)	57
Male	20 (41%)	29 (59%)	49	23 (25%)	68 (75%)	91
Exposure Group						
Heterosexual	30 (39%)	46 (61%)	76	16 (17%)	79 (83%)*	95
IDU	0 (0%)	4 (100%)	4	0 (0%)	2 (100%)	2
MSM	5 (42%)	7 (58%)	12	14 (34%)*	27 (66%)	41
Other	0 (0%)	2 (100%)	2	2 (67%)	1 (33%)	3
NA	0 (0%)	5 (100%)	5	2 (29%)	5 (71%)	7
Ethnicity						
White	21 (34%)	41 (66%)	62	33 (27%)*	91 (73%)*	124
Afro-descendants/ Amerindians/Other	7 (35%)	13 (65%)	20	0 (0%)	21 (100%)*	21
NA	7 (41%)	10 (59%)	17	1 (33%)	2 (67%)	3
Education level						
≤ 8 years	19 (33%)	39 (67%)	58	12 (20%)	49 (80%)	61
9 - 11 years	5 (21%)	19 (79%)	24	14 (26%)	40 (74%)	54
≥ 12 years	4 (80%)*	1 (20%)	5	8 (27%)	22 (73%)	30
NA	7 (58%)	5 (42%)	12	0 (0%)	3 (100%)	3

* P value < 0.05; NA – data not available; MSM – men who have sex with men; IDU – injection drug users; values between brackets are relative percentages of the total of the line.

Differences between HIV-1 subtype B and C were further investigated regarding the epidemic expansion rate. Skyride coalescent analysis indicated a logistic growth curve for both HIV-1C and HIV-1B (Figure 3.3). A difference in time of around 10 years was observed for the onset of the subtype B (1960, 1949-1972 95%HPD) and subtype C (1972, 1964-1978 95% HPD) epidemics. Visually inspecting the Skyride plot, HIV-1C seems to have expanded faster than HIV-1B, overcoming the latter's effective population size in the end 1990's, when both subtypes slowed down the growth rate. The number of yearly reported AIDS cases compiled for the RS and SC states follows the same trend. Analyzing the HIV-1 phylodynamics separately for RS and SC states reveals a more pronounced growth curve for HIV-1C in SC. In only 10 years after the introduction, HIV-1C epidemic reached a bigger N_e than HIV-1B, while in RS it took 30 years to HIV-1C reach the same N_e than HIV-1B (Figure 3.3).

The logistic growth coalescent model was used to quantify the epidemic growth rate in RS and SC. Despite some overlap of the HPD intervals, our analysis revealed that subtype C expanded in a faster rate than subtype B in all datasets (Table 3.2). As visually observed in skyride plots, differences in growth rate for subtype C and B were more pronounced in SC (0.73 year⁻¹ and 0.45 year⁻¹, respectively) than in RS (0.54 year⁻¹ and 0.38 year⁻¹, respectively). Estimated r for HIV-1B among the different datasets was more similar than those for HIV-1C.

Table 3.2. Logistic growth (r) and population doubling time (λ) estimated for HIV-1C and HIV-1B in Rio Grande do Sul and Santa Catarina states, Brazil.

Subtype	Dataset	Sample size	r , year ⁻¹ (95% HPD)	λ , year (95% HPD)
C	RS&SC	322	0.63 (0.50-0.78)	1.11 (0.89-1.39)
	RS	134	0.54 (0.36-0.75)	1.29 (0.92-1.90)
	SC	188	0.73 (0.54-0.96)	0.95 (0.72-1.28)
B	RS&SC	108	0.48 (0.34-0.64)	1.46 (1.09-2.03)
	RS	66	0.38 (0.25-0.54)	1.83 (1.28-2.81)
	SC	42	0.45 (0.28-0.66)	1.53 (1.05-2.47)

Abbreviations: HPD – highest posterior density; RS – Rio Grande do Sul; SC – Santa Catarina.

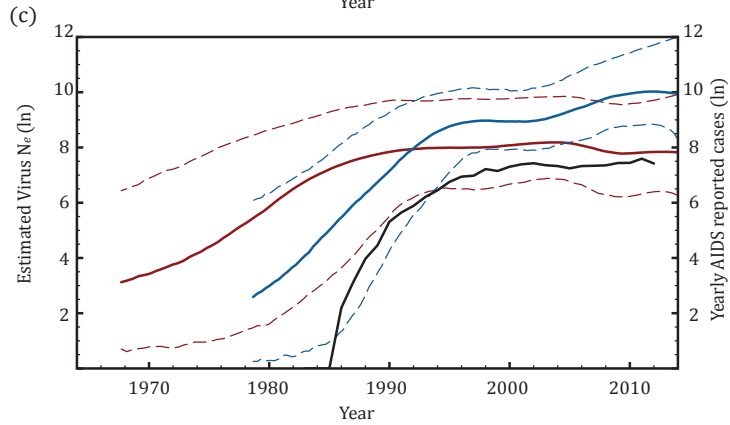
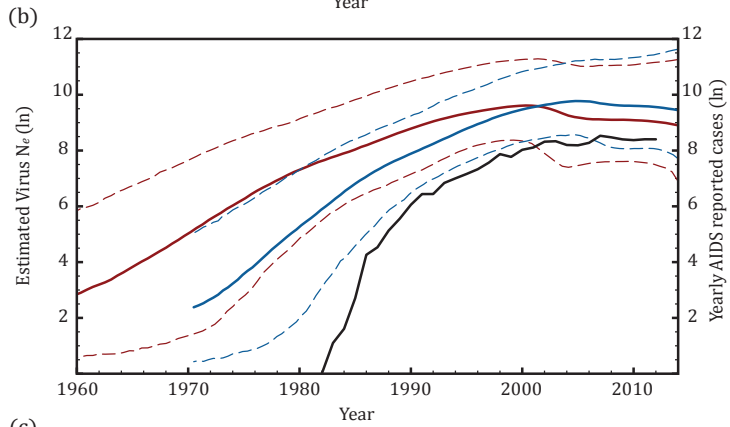
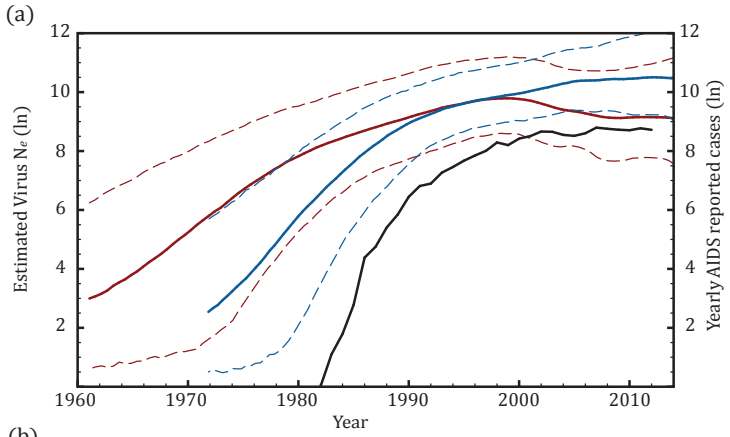


Figure 3.3. Bayesian skyride plots representing changes in virus effective population size (N_e) through time. Solid lines represent the median N_e and dashed lines the 95% HPD. Subtype B is colored in red and subtype C in blue. The solid black line shows the number of new cases of AIDS yearly reported, as indicated in the second y-axis. a) Santa Catarina (SC) and Rio Grande do Sul (RS) dataset; b) RS dataset; c) SC dataset.

3.6 - Discussion

The study presented here describes the HIV-1 molecular epidemiology in a large area, comprehending 13 municipalities from the two southernmost Brazilian states. In terms of molecular diversity our results corroborate previous studies that also reported a high dominance of subtype C (~65%) in SC state (25-28) and a more diverse epidemic in RS state mainly because of high prevalence of BC mosaics (~25%) (29-31). In both states low prevalence of subtype F1 and BF1 recombinants (~3%) was observed, while this frequency can reach up to 20% in northern regions of Brazil (32,33) and 40% in Argentina (34). Interesting to note that the locations with higher subtype F1 or BF1 prevalence were cities close to the border with Argentina (i.e. Uruguiana and Chapecó) and the port city of Itajaí, where high international population flux can foster the contact with new HIV-1 strains (Figure 3.1).

The CRF31_BC strain was almost exclusively found in RS state, more precisely in cities close to Porto Alegre (i.e. Lajeado and Caxias do Sul) (Figure 3.1), where the CRF31_BC frequency is around 22% and may reach to 35% in the metropolitan region (29, 35). Despite a more pronounced occurrence in Porto Alegre and nearby cities, CRF31_BC is not limited to this region once it was also observed in more distant cities in the current study. Among the BC mosaic forms, some displayed a CRF31_BC-related recombination pattern, likely originated from a second recombination event within the CRF31_BC *pol* fragment as previously reported in other samples from Porto Alegre (36). In our sampling this pattern was observed in two clusters of sequences (Figure 3.2) and in some other isolated sequences classified here as URF_BC (data not shown). Although very similar, only a thorough analysis of the

recombination region can confirm if these sequences are indeed second-generation recombinants (SGR) derived from CRF31_BC.

Three possible new CRFs were identified here: BC_cluster1 and 2, that might be SGRs from CRF31_BC; and BC_cluster3 with a different recombination display. Further studies aiming to sequence the full-length genome of these strains are needed to confirm this hypothesis but our findings point to a higher diversity of circulating BC mosaic strains than previously reported (36). Ten distinct BC CRF have been described worldwide so far, eight of them in the Chinese province of Yunnan, a region close to India where subtypes C and B co-circulates (37). High prevalence of AIDS in injection drug users (IDU) makes Yunnan province a hot spot for HIV recombination. Very small prevalence of AIDS cases among IDU in Brazil (6) could explain the low diversity of BC CRFs observed in south region. Nevertheless, due to high prevalence of HIV-1C and B, it is likely that CRFs other than CRF31_BC are circulating in the region.

Indeed, separated transmission chains for subtype B and C can be acting in the compartmentalization of the epidemic in south Brazil. In SC we observed an association between MSM group and HIV-1B, the same association was previously reported in the state capital city Florianópolis (25). The lower prevalence of BC mosaic strains in SC might be consequence of the segregation of HIV-1C and B in separated transmission chains, scenario that is not observed in RS state where higher number of mosaic strains is observed. Moreover, it was reported that IDU could have had an initial role in the dispersion of subtype C in RS, promoting the admixture with the subtype B epidemic (38). Our study only captured an association between HIV-1B and higher levels of education in RS state. This could be interpreted as a consequence from the higher education levels among MSM; however only two (50%) MSM were observed in the group of HIV-1B infected individuals with ≥ 12 years of education in RS. A second demographic factor that can be shaping the viral transmissions in south Brazil is the ethnicity. In SC state all patients self-declared as African-descendants or Amerindians were infected with HIV-1C. Although this demographic feature does not represent an HIV exposure group, factors related to ethnicity or social economical class might be segregating HIV-1 subtypes in separate epidemics. However, only studies better designed to capture

representative samples from different social economical classes can clarify this issue.

Phylogenetic analysis performed here showed a faster epidemic expansion of HIV-1C in SC ($r=0.73$) in comparison to RS ($r=0.54$). This can also be linked to the segregation of subtypes in SC state, once the HIV-1C introduction (1979, 1970-1986 95%HPD) in heterosexual group in SC might have happened when very few HIV-1B infections were present in this group, allowing HIV-1C to expand quickly in a transmission chain with lot of potential hosts. In RS state the two epidemics expanded in more similar rates (Figure 3.3 and Table 3.2), what can be consequence of a co-circulation in the same exposure groups and under the same conditions.

Regardless of sampling location, HIV -1C growth rate estimated here was faster than HIV-1B, despite some overlap of the HPD intervals (Table 3.2). The growth rate estimated in our datasets for HIV-1C epidemic is in agreement with previous studies that reported r values of 0.63 - 0.81 (39, 40). HIV-1B epidemic growth rate was similar to Bello et al. (0.56) (39) but higher than Salemi et al. (~0.30) (40). Skyride plots showed that HIV-1C effective population size increased until the middle 2000's, around ten years more than previous estimates (39). This difference can be the result of more recent samples analyzed here, allowing us to capture the growing trends in the recent years, or because of methodological differences. The Bayesian skyride coalescent method reconstructs smoother plots when compared to Bayesian skyline, therefore sudden population changes are reconstructed in a more evenly way (20). Nevertheless the viral N_e growing curves estimated from our datasets fit well in the history of yearly reported cases of AIDS in RS and SC states, analyzed together or in separate.

In summary our study depicts the molecular epidemic in several cities from Santa Catarina and Rio Grande do Sul states. This is the first time that a broad sampling area was investigated in south Brazil. Although the sample size from some of the cities is small (e.g. Santa Maria and Lages), we were able to capture the diversity of HIV-1 in the region. Our results highlight the fast expansion rate of HIV-1C epidemic, stabilizing only in recent years, and show how separated transmission chains might have impacted the phylogenetic of the viral populations. In addition we also described potential new circulating BC mosaic strains, whose recombination breakpoints can be derived from

CRF31_BC. The region studied here is the most heavily affected by AIDS in Brazil and our findings can be valuable to the development of new public health policies that aim to slow down HIV dissemination in south Brazil and beyond.

3.7 - Acknowledgements

We wish to thank all collaborating municipal health centers from states of Santa Catarina and Rio Grande do Sul, Brazil. We also thank Dr. Philippe Lemey for his helpful suggestions and assistance.

This study was supported by Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

The authors have no conflicts of interest.

3.8 – References

1. Tebit DM, Arts EJ. 2011. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis* 11:45–56.
2. Hemelaar J. 2013. Implications of HIV diversity for the HIV-1 pandemic. *J Infect* 66:391–400.
3. Vuilleumier S, Bonhoeffer S. 2015. Contribution of recombination to the evolutionary history of HIV. *Curr Opin HIV AIDS* 10:84–89.
4. Alencar CS, Sabino EC, Carvalho SMF, Leao SC, Carneiro-proietti AB, Capuani L, Oliveira CL, Carrick D, Birch RJ, Gonçalves TT, Keating S, Swanson PA, Hackett J, Busch MP, *Retrovirus N*. 2013. HIV Genotypes and Primary Drug Resistance Among HIV-Seropositive Blood Donors in Brazil : Role of Infected Blood Donors as Sentinel Populations for Molecular Surveillance of HIV. *JAIDS* 63:387–392.

5. Gräf T, Pinto AR. 2013. The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. *Virology* 435:170-178.
6. Brazilian Ministry of Health. 2014. AIDS Epidemiological Bulletin July 2013-June 2014. Brasília, DF.
7. Brazilian Ministry of Health. 2009. AIDS Epidemiological Bulletin July 2008-June 2009. Brasília, DF.
8. Delatorre E, Couto-Fernandez JC, Guimarães ML, Vaz Cardoso LP, de Alcantara KC, Stefani MMDA, Romero H, Freire CCM, Iamarino A, de A Zannotto PM, Morgado MG, Bello G. 2013. Tracing the origin and northward dissemination dynamics of HIV-1 subtype C in Brazil. *PLoS One* 8:e74072.
9. Jones LR, Dilernia DA, Manrique JM, Moretti F, Salomón H, Gomez-Carrillo M. 2009. In-depth analysis of the origins of HIV type 1 subtype C in South America. *AIDS Res Hum Retroviruses* 25:951-959.
10. Carrion G, Aguayo N, Avila MM, Ruchansky D, Pando MA, Vinales J, Perez J, Barboza A, Chauca G, Romero A, Galeano A, Blair PJ, Weissenbacher M, Birx DL, Sanchez JL, Olson JG, Carr JK. 2004. Documentation of Subtype C HIV Type 1 Strains in Argentina, Paraguay, and Uruguay. *AIDS Res Hum Retroviruses* 20:1022-1025.
11. Bello G, Passaes CP, Guimarães ML, Lorete RS, Matos Almeida SE, Medeiros RM, Alencastro PR, Morgado MG. 2008. Origin and evolutionary history of HIV-1 subtype C in Brazil. *AIDS* 22:1993-2000.
12. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, Snoeck J, van Rensburg EJ, Wensing AMJ, van de Vijver DA, Boucher CA, Camacho R, Vandamme AM. 2005. An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* 21:3797-3800.
13. Pond SLK, Posada D, Stawiski E, Chappay C, Poon AFY, Hughes G, Fearnhill E, Gravenor MB, Brown AJL, Frost SDW. 2009. An evolutionary model-based algorithm for accurate phylogenetic

breakpoint mapping and subtype prediction in HIV-1. *PLoS Comput Biol* 5:e1000581

14. Siepel AC, Halpern AL, Macken C, Korber BT. 1995. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Res Hum Retroviruses* 11:1413–1416.

15. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.

16. Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27:221–4.

17. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 73:152–160.

18. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29:1969–1973.

19. Gifford RJ, Liu TF, Rhee SY, Kiuchi M, Hue S, Pillay D, Shafer RW. 2009. The calibrated population resistance tool: Standardized genotypic estimation of transmitted HIV-1 drug resistance. *Bioinformatics* 25:1197–1198.

20. Minin VN, Bloomquist EW, Suchard MA. 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol* 25:1459–1471.

21. Bello G, Eyer-Silva W a, Couto-Fernandez JC, Guimarães ML, Chequer-Fernandez SL, Teixeira SLM, Morgado MG. 2007. Demographic history of HIV-1 subtypes B and F in Brazil. *Infect Genet Evol* 7:263–70.

22. Shapiro B, Ho SYW, Drummond AJ, Suchard MA, Pybus OG, Rambaut A. 2011. A bayesian phylogenetic method to estimate unknown sequence ages. *Mol Biol Evol* 28:879–887.

23. Griffiths RC, Tavaré S. 1994. Sampling theory for neutral alleles in a varying environment. *Philos Trans R Soc Lond B Biol Sci* 344:403–410.
24. Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics* 26:680–682.
25. Gräf T, Passaes CPB, Ferreira LGE, Grisard EC, Morgado MG, Bello G, Pinto AR. 2011. 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* 51:186–191.
26. Locateli D, Stoco PH, De Queiroz ATL, Alcântara LCJ, Ferreira LGE, Zanetti CR, Rodrigues R, Grisard EC, Pinto AR. 2007. Molecular epidemiology of HIV-1 in Santa Catarina State confirms increases of subtype c in southern Brazil. *J Med Virol* 79:1455–1463.
27. Rodrigues R, Manenti S, Romao PRT, de Paula Ferreira JL, Batista JPG, Siqueira AFAC, de Macedo Brigido LF. 2010. Young pregnant women living with HIV/AIDS in Criciúma, Southern Brazil, are infected almost exclusively with HIV type 1 clade C. *AIDS Res Hum Retroviruses* 26:351–357.
28. Brigido LFM, Nunes CC, Oliveira CM, Knoll RK, Ferreira JLP, Freitas CA, Alves MA, Dias C. 2007. HIV Type 1 Subtype C and CB Pol Recombinants Preval at the Cities with the Highest AIDS Prevalence Rate in Brazil. *AIDS Res Hum Retroviruses* 23:1579–1585.
29. de Medeiros RM, Junqueira DM, Matte MCC, Barcellos NT, Chies JAB, Matos Almeida SE. 2011. 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* 83:1682–1688.
30. Silveira J, Santos AF, Martínez AMB, Góes LR, Mendoza-Sassi R, Muniz CP, Tupinambás U, Soares MA, Greco DB. 2012. Heterosexual transmission of human immunodeficiency virus type 1 subtype C in southern Brazil. *J Clin Virol* 54:36–41.
31. Dias CF, Nunes CC, Freitas IO, Lamego IS, De Oliveira IMR, Gilli S, Rodrigues R, Brigido LF. 2009. 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* 51:191–196.

32. Brindeiro RM, Diaz RS, Sabino EC, Morgado MG, Pires IL, Brigido L, Dantas MC, Barreira D, Teixeira PR, Tanuri A. 2003. Brazilian Network for HIV Drug Resistance Surveillance (HIV-BResNet): a survey of chronically infected individuals. *AIDS* 17:1063–1069.

33. Bongertz V, Bou-Habib DC, Brígido LF, Caseiro M, Chequer PJ, Couto-Fernandez JC, Ferreira PC, Galvão-Castro B, Greco D, Guimarães ML, Linhares de Carvalho MI, Morgado MG, Oliveira CA, Osmanov S, Ramos CA, Rossini M, Sabino E, Tanuri A, Ueda M. 2000. HIV-1 diversity in Brazil: genetic, biologic, and immunologic characterization of HIV-1 strains in three potential HIV vaccine evaluation sites. Brazilian Network for HIV Isolation and Characterization. *Journal of acquired immune deficiency syndromes. JAIDS* 23:184-193.

34. Pando MA, Gómez-Carrillo M, Vignoles M, Rubio AE, dos Ramos Farias MS, Vila M, Rossi D, Ralón G, Marone R, Reynaga E, Sosa J, Torres O, Maestri M, Avila MM, Salomón H. 2011. 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* 27:17–23.

35. Simon D, Béria JU, Tietzmann DC, de Carli R, Stein AT, Lunge VR. 2010. Prevalência de subtipos do HIV-1 em amostra de pacientes de um centro urbano no sul do Brasil. *Rev Saude Publica* 44:1094–1101.

36. Passaes CPB, Bello G, Lorete RS, Matos Almeida SE, Junqueira DM, Veloso VG, Morgado MG, Guimarães ML. 2009. Genetic characterization of HIV-1 BC recombinants and evolutionary history of the CRF31_BC in Southern Brazil. *Infect Genet Evol* 9:474–482.

37. Chen M, Yang L, Ma Y, Su Y, Yang C, Luo H, Chen H, Chen L, Yan W, Shi Y, Jia M, Lu L. 2013. Emerging variability in HIV-1 genetics among recently infected individuals in Yunnan, China. *PLoS One* 8:e60101.

38. Almeida SE, de Medeiros RM, Junqueira DM, Gräf T, Passaes CPB, Bello G, Morgado MG, L Guimarães M. 2012. Temporal dynamics of HIV-1 circulating subtypes in distinct exposure categories in southern Brazil. *Virology* 9:306.
39. Bello G, Guimarães ML, Passaes CPB, Matos Almeida SE, Veloso VG, Morgado MG. 2009. Evidences of recent decline in the expansion rate of the HIV type 1 subtype C and CRF31_BC epidemics in southern Brazil. *AIDS Res Hum Retroviruses* 25:1065–1069.
40. Salemi M, De Oliveira T, Soares MA, Pybus O, Dumans AT, Vandamme AM, Tanuri A, Cassol S, Fitch WM. 2005. Different epidemic potentials of the HIV-1B and C subtypes. *J Mol Evol* 60:598–605.

3.9 – Supplementary Information

Table S3.1. Demographic and Epidemiological characterization of the population enrolled in the study.

Characteristic	State		Row Total (n=317)	
	RS (n=144)	SC (n=173)		
Gender	Female	74 (51%)	69 (40%)	143 (45%)
	Male	70 (49%)	104 (60%)	174 (55%)
Transmission	Transfusion	2 (1%)	4 (2%)	6 (2%)
	Heterosexual	114 (79%)	111 (64%)	225 (71%)
	IDU	4 (3%)	3 (2%)	7 (2%)
	MSM	16 (11%)	47 (27%)	63 (20%)
	Vertical	1 (1%)	-	1 (0%)
	NA	7 (5%)	8 (5%)	15 (5%)
Ethnicity	White	86 (60%)	142 (82%)	228 (72%)
	Afro-descendants/ Amerindians/Other	27 (19%)	28 (16%)	55 (17%)
	NA	31 (22%)	3 (2%)	34 (11%)
Education level	≤ 8 years	74 (51%)	75 (43%)	149 (47%)
	9 - 11 years	39 (27%)	60 (35%)	99 (31%)
	≥ 12 years	7 (5%)	35 (20%)	42 (13%)
	NA	24 (17%)	3 (2%)	27 (9%)
Marital Status	Steady partner	28 (19%)	39 (23%)	67 (21%)
	Married	43 (30%)	44 (25%)	87 (27%)
	Divorced/widower	8 (6%)	6 (3%)	14 (4%)
	Single	59 (41%)	82 (47%)	141 (44%)
	NA	6 (4%)	2 (1%)	8 (3%)
Age	<20	3 (2%)	2 (1%)	5 (2%)
	20-29	36 (25%)	38 (22%)	71 (22%)
	30-39	41 (28%)	50 (29%)	74 (23%)
	40-49	29 (20%)	47 (27%)	91 (29%)
	≥50	35 (24%)	36 (21%)	76 (24%)

NA – data not available; MSM – men who have sex with men; IDU – injection drug users; values between brackets are relative percentages of the total of the line.

CAPÍTULO 4.

The contribution of epidemiological predictors in unravelling the phylogeographic history of HIV-1 subtype C in Brazil

Artigo submetido à Journal of Infectious Diseases

Tiago Gräf, Bram Vrancken, Dennis Maletich Junqueira, Rúbia
Marília De Medeiros, Marc A. Suchard, Philippe Lemey, Sabrina
Esteves De Matos Almeida, Aguinaldo Roberto Pinto

4.1 – Abstract

Background: The phylogeographic history of the Brazilian HIV-1 subtype C (HIV-1C) epidemic is still unclear. Previous studies have mainly focused on Brazilian capital cities and the fact that HIV-1C infections increase at a higher rate than subtype B in Brazil calls for a better understanding of the process of spatial spread.

Methods: A comprehensive sequence dataset representing 22 Brazilian locations was assembled and analyzed. A Bayesian phylogeographic generalized linear model approach was used to reconstruct the spatiotemporal history of HIV-1C in Brazil considering several potential explanatory predictors of the viral diffusion process. Analyses were performed on several subsampled datasets in order to mitigate potential sample biases.

Results: We reveal a central role for Porto Alegre city, the capital of the southernmost state, in the Brazilian HIV-1C epidemic (HIV-1C_BR), and the northwards expansion of HIV-1C_BR could be linked to source populations with higher HIV-1 burden and larger proportions of HIV-1C infections.

Conclusion: The results presented here bring new insights to the continuing discussion about the HIV-1C epidemic in Brazil, and raise an alternative hypothesis for its spatiotemporal history. The current work also highlights the importance of incorporating external information in phylogeographic analyses and demonstrates how sampling bias can confound such analyses.

4.2 – Introduction

Upon emergence into the human population in central Africa around 1920 [1], HIV-1 group M has diversified into nine subtypes and numerous circulating recombinant forms (CRFs) through a series of founder effects and recombination [2,3]. Although HIV-1 subtype B (HIV-1B) is dominant in many countries in Europe and Americas [2], more than 50% of the infections worldwide are caused by HIV-1 subtype C (HIV-1C), which is the most prevalent subtype in southern Africa countries and India, and increases in prevalence in China and South America [2,4].

The epidemic in Brazil is mainly driven by HIV-1B, followed by lower frequencies of HIV-1C, F1 and BF1 recombinants [5]. In the southern regions of Brazil, however, the spread of HIV-1B is matched by HIV-1C, which co-circulates in similar proportions and can even be responsible for up to 80% of the infections [4]. In addition, the two southernmost Brazilian states, Rio Grande do Sul (RS) and Santa Catarina (SC), have the highest AIDS incidence in the country [6]. These patterns have motivated several investigations into the history and dynamics of the Brazilian HIV-1C epidemic (HIV-1C_BR), which revealed that it has started in the late 1960s to middle 1970s in the state of Paraná (PR) [7,8,9]. Because the HIV-1C incidence in more northern states has only recently begun to increase [4,10-14], this suggests viral diffusion would be driven by unknown factors that promote a fast dissemination to the south while constraining spread to the north.

An important goal in molecular epidemiology research is the identification of factors that determine the spread of pathogens. Recently, a model-based Bayesian phylogeographic framework has been extended to integrate molecular sequence data and epidemiological sources of information [15]. The approach to jointly estimate phylogeographic history and identify the contribution of explanatory factors (predictors) to the patterns of spatial spread allows for a bi-directional cross-talk between the genetics and the relevant predictors: the predictors are selected for the ability to explain the location transition history but they can also assist in shaping this history.

In the present study we reconstruct the phylogeographic history of HIV-1C in Brazil incorporating newly obtained sequence data. While previous studies mostly included sequences from the capital cities, we here expanded the spatial sampling by including HIV-1C sequences from 10 rural locations in the southernmost states RS and SC. Our study demonstrates for the first time that augmenting the molecular sequence data with relevant epidemiological information can contribute to the robustness of phylogeographic reconstructions.

4.3 – Methods

4.3.1 - Patients, samples and new sequences

A total of 360 patients from 13 cities from the states of SC and RS (Figure 4.1) were enrolled in this study. Blood samples were collected and HIV-1 envelope (HXB2 6846-7360 bp) and polymerase (HXB2 2274-3545 bp) fragments were amplified from whole cellular DNA by nested-PCR and sequenced as described elsewhere [16]. Sequences were subtyped using the REGA, RIP and SCUEAL online subtyping tools [17,18,19] and by performing maximum likelihood phylogenetic inference incorporating HIV-1 subtype reference sequences available from the Los Alamos HIV sequence database (www.hiv.lanl.gov). Recombinant sequences were identified through bootscanning analysis using Simplot 3.5.1 [20] (see supplemental material, Text S4.1, for methodological details). The sequences generated in the present study were deposited in GenBank under accession numbers: KR065788-KR066336 and KP224476-KP224501.

4.3.2 - Sequence dataset compilation

Briefly, we complemented our new sequence data with all publically available Brazilian HIV-1C sequences (HIV-1C_BR) from pol and env genes ($n = 385$) and the 50 best hits per query in a BLAST [21] search of HIV1C_BR against all non-Brazilian HIV-1 subtype C sequences. After extensive data cleaning this resulted in datasets with 1522 pol and 621 env sequences that were down-sampled to around 500 sequences each to reduce the computational burden in subsequent Bayesian statistical analyses (see Supplementary Text S4.1, for full details of the followed procedure). Six additional sub-samplings containing only Brazilian sequences were made for pol and env, to allow assessing the robustness of the phylogeographic reconstructions (see below). For this we aimed at reducing sampling bias by creating three random down-samples in two groups: a) Rand10 - with a maximum of ten sequences by location and b) Rand20 - with a maximum number of 20 sequences by location (see Supplementary Table S4.2, for HIV-1C_BR sequences in each dataset).

4.3.3 - Phylogenetic divergence time estimation and population dynamics inference

Time-scaled phylogenetic trees were reconstructed using a Bayesian inference method implemented in the BEAST/BEAGLE software [23,24]. All analyses were performed using the GTR+I+ Γ 4 nucleotide substitution model and an uncorrelated lognormal relaxed molecular clock under the Bayesian Skyride coalescent model [25,26]. Due to the low temporal signal of the datasets, the use of an informative prior on the tMRCA of the Brazilian subtype C clade was required. For this purpose, we specified a normal distribution with mean (1976) and standard deviation (5.1) based on previous estimates of the time of introduction of subtype C in Brazil [8]. When exact sampling dates were unknown, the dates were integrated out over a known sampling time interval [27]. Monte Carlo Markov Chains (MCMC) were run sufficiently long to ensure stationarity and adequate effective sample size (ESS >200) as diagnosed by Tracer (<http://beastbioedacuk/Tracer>). Maximum clade credibility (MCC) trees were summarized using the TreeAnnotator tool and visualized in Figtree v1.4.0 (<http://tree.bio.ed.ac.uk/software/figtree/>). A representative sample of 1000 trees was collected and used as an empirical tree distribution for estimating the virus migration processes (see below). To ensure that subsequent phylogeographic analyses are based on histories specific to Brazil we pruned sequences clustering outside the HIV-1C_BR cluster or non-Brazilian sequences clustering inside HIV-1C_BR cluster from these trees using PAUP (<http://paup.csit.fsu.edu>) (see Text S1, for methodological details).

4.3.4 - Phylogeny-trait association

We tested for significant phylogenetic clustering by location in two different ways. First, we calculated the Association Index (AI), Parsimony Score (PS) and Monophyletic Clade (MC) measures using BaTS [28]. For our second approach we introduced the use of the path sampling (PS) and stepping-stone (SS) sampling marginal likelihood estimators as implemented in BEAST [29,30] to evaluate the extent to which the topology is required as a correlation structure to explain the

traits. For this we specified a discrete symmetric (reversible) model of location transitioning and incorporate Bayesian stochastic search variable selection (BSSVS) procedure [31] fitting the trait diffusion process on: 1) a fixed MCC tree summarized from the Bayesian phylogenetic analysis of the complete dataset; 2) a star-like tree with the same trait annotations as in the MCC tree.

4.3.5 - Phylogeographic inference with epidemiological predictors

To assess the impact of potential explanatory variables (predictors) of the viral diffusion process on phylogeographic reconstructions, we made use of the recent generalized linear model (GLM) extension of Bayesian discrete phylogeographic models [15]. This allows reconstructing the spatial diffusion history throughout the tree while simultaneously evaluating the contribution of various potential predictors. Well-supported predictors are identified using a BSSVS procedure, and the contribution of each predictor is quantified as a GLM coefficient.

Using this approach, we tested the following predictors (see Supplementary Text S4.1, for methodological details):

- Geographic distance: the great-circle distance among each pair of cities;
- Passenger air traffic: the number of passengers traveling between each pair of airports;
- HIV population size: the total number of AIDS notifications in a period of 10 years reported in each city;
- HIV prevalence: $(\text{HIV population size} / \text{city population size}) \times 100,000$ habitants;
- HIV-1C population size: HIV population size times the proportion of HIV-1C as reported in the literature [4,10-14];
- HIV-1C prevalence: $(\text{HIV-1C population size} / \text{city population size}) \times 100,000$ habitants;
- Sample size: the number of sequences by location.

Because not all sampling locations have an airport, we specified a different geographic partitioning for evaluating predictor 2 (passenger air traffic). This partitioning is not well suited for the epidemiological predictors, which led us to test predictor 2 in separate analyses including only sample size as an additional potential predictor.

GLM analyses were run in BEAST using previously recommended prior specifications on the set of empirical trees obtained by the Bayesian phylogenetic analysis [15]. Bayes Factors (BF) were calculated to determine the support for the inclusion of each predictor in the model and predictor contributions are reported as effect sizes conditional on the effect being included in the model.

A phylogeographic analysis with BSSVS was performed with asymmetric transition rates being informed by the predictors supported by the GLM analysis.

In other words, for each sub-sampled dataset, we used the rate estimates for prior specification based on the corresponding GLM analysis. SPREAD software was used to identify the well-supported transition rates based on BFs > 3 [32]. We complemented this analysis with Markov jump estimation of the number of location transitions throughout the evolutionary history [33]. RStudio (<http://www.rstudio.org/>) was used to calculate the Bayes factors and effect sizes, and to summarize the posterior densities of the highly supported transitions from the BEAST log files.

4.4 – Results

4.4.1 - Sequence dataset compilation

We sequenced 140 pol and 202 env HIV-1C isolates in 13 locations in SC and RS states, 10 of which have not been sampled before (Figure 4.1). By combining the generated sequence data with publicly available Brazilian and international HIV-1C sequences we were able to compile comprehensive pol and env based datasets for reconstructing the spatiotemporal history of HIV-1C in Brazil. In summary, the complete pol dataset contained 380 Brazilian and 120 international sequences while the env dataset totaled 293 Brazilian and 170 international sequences. The Brazilian pol sequences are distributed over 21 locations and the env sequences represent 17 locations totalizing 22 locations represented with pol or env sequences, most of them in SC and RS (15/21 for pol and 14/17 for env). Considering the complete Brazilian dataset, sequences represent the time period between 2002 and 2014 (see Supplementary Table S4.2).

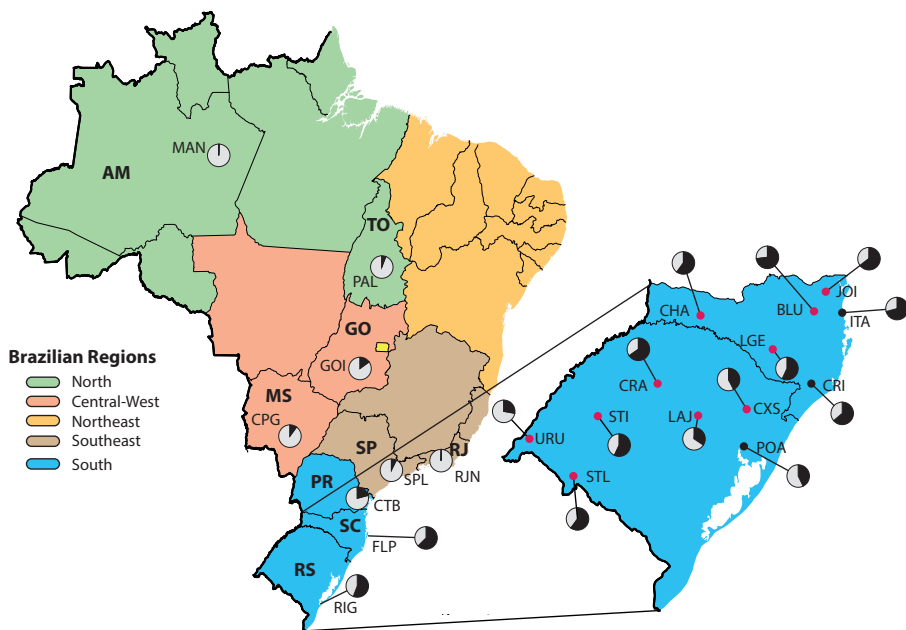


Figure 4.1. Geo-political map of Brazil indicating the locations from where HIV-1C sequences were obtained. Pie charts show the HIV-1C (black) percentage of infections relative to other HIV-1 strains (grey) in all cities with pol or env sequences included in this study. State name abbreviations are shown in bold. The inset shows an enlarged map with the sampling locations (black and red dots) in Santa Catarina and Rio Grande do Sul from which new sequence data were generated. Red dots: cities sampled for the first time. Black dots: sampling locations from where sequence data from other studies were also available. Brazilian regions are colored according to the legend. Acronyms for states: AM: Amazonas; GO: Goiás; MS: Mato Grosso do Sul; PR: Paraná; RJ: Rio de Janeiro; SC: Santa Catarina; SP: São Paulo; TO: Tocantins; RS: Rio Grande do Sul. Acronyms for cities: BLU: Blumenau; CHA: Chapecó; CPG: Campo Grande; CRA: Cruz Alta; CRI: Criciúma; CTB: Curitiba; CXS: Caxias do Sul; FLP: Florianópolis; GOI: Goiania; ITA: Itajaí; JOI: Joinville; LAJ: Lajeado; LGE: Lages; MAN: Manaus; PAL: Palmas; POA: Porto Alegre; RIG: Rio Grande; RJN: Rio de Janeiro; SPL: São Paulo; STI: Santiago; STL: Santana do Livramento; URU: Uruguaiiana.

4.4.2 - Phylogeny-trait association

Because our preliminary analyses suggested a considerable degree of phylogenetic mixing by location, we formally tested whether the datasets containing only Brazilian sequences still supported spatial population structure. The hypothesis of a panmictic population could be rejected for the pol and env datasets by the association index (AI) and parsimony score (PS) statistics ($p < 0.05$), but the monophyletic clade (MC) scores revealed that for 12/21 (57%) of pol locations and 12/17 (71%) of env locations random clustering could not be rejected (see Supplementary Table S4.3, for MC scores). The results of the approach based on model testing also provided strong support against the absence of phylogenetic association by sampling location in the pol and env datasets (Bayes factors of 74 and 39 respectively).

4.4.3 - Inconsistencies in root state estimates

The results of the phylogeographic reconstruction showed, with strong agreement between most datasets and models applied (50/56 analyses), that the epidemic ignited in SC or RS. Its exact location of introduction could, however, not be unambiguously determined using only virus genetic data. Whereas in the complete pol and env datasets Florianópolis (FLP) was consistently estimated as the most likely location at the root, other cities - most notably Porto Alegre (POA) (7/48) and Criciúma (CRI) (7/48) - were implicated in 60% (29/48) of the analyses based on the Rand10 and Rand20 subsampled datasets (Table 4.1).

Table 4.1. Modal root state and posterior probability estimates resulting from different discrete Bayesian phylogeographic analyses applied to different datasets.

		Subsampling						
		All	RAND10A	RAND10B	RAND10C	RAND20A	RAND20B	RAND20C
<i>pol</i>	<i>Symmetric-BSSVS</i>	FLP (1.00)	SPL (0.99)	CTB (0.99)	RJN (0.99)	POA (0.99)	RIG (0.99)	RIG (0.99)
	<i>Symmetric</i>	FLP (0.99)	CRI (0.99)	FLP (0.97)	CRI (0.97)	FLP (0.99)	FLP (0.98)	ITA (0.99)
	<i>Asymmetric-BSSVS</i>	FLP (1.00)	CRI (1.00)	FLP (0.99)	CTB (0.99)	ITA (1.00)	FLP (0.99)	POA (0.99)
	<i>Asymmetric</i>	FLP (0.99)	RJN (0.99)	FLP (0.99)	CTB (0.99)	ITA (1.00)	FLP (0.99)	FLP (0.99)
<i>env</i>	<i>Symmetric-BSSVS</i>	FLP (0.97)	FLP (0.99)	POA (0.99)	POA (0.99)	FLP (0.99)	FLP (0.99)	CXS (0.99)
	<i>Symmetric</i>	FLP (0.97)	FLP (0.99)	POA (0.99)	FLP (0.96)	FLP (0.99)	FLP (0.99)	CRI (0.99)
	<i>Asymmetric-BSSVS</i>	FLP (0.99)	CRI (0.99)	POA (0.99)	LAJ (0.99)	CRI (1.00)	POA (0.99)	FLP (1.00)
	<i>Asymmetric</i>	FLP (0.99)	BLU (0.99)	FLP (0.99)	CRI (0.69)	FLP (0.99)	CRA (0.99)	FLP (1.00)

Acronyms for cities: BLU: Blumenau; CRI: Criciúma; CTB: Curitiba; CXS: Caxias do Sul; FLP: Florianópolis; ITA: Itajaí; LAJ: Lajeado; POA: Porto Alegre; RIG: Rio Grande; RJN: Rio de Janeiro.

4.4.4 - Predictors of viral spread

Using a phylogeographic GLM approach, we evaluated which measures predict the rates of location exchange in the complete and subsampled datasets (Table 4.2). In the *pol* and *env* complete datasets only the origin and destination sample size yielded strong Bayes factor (BF) support, reflecting the fact that only sample sizes and their heterogeneity are needed to explain the number of location transitions. We also found strong support for destination sample size in the model as a predictor in all *pol* and *env* subsampled datasets (Rand10 and Rand20). This indicates that despite the more homogeneous distribution of sequences by sampling locations in these subsampled datasets, the remaining heterogeneity still has an impact on the phylogeographic reconstructions.

Table 4.2. Bayes factor support for an explanatory role in the HIV-1C_BR diffusion process for all tested predictors in all datasets

	Predictor	Dataset						
		Complete	RAND10A	RAND10B	RAND10C	RAND20A	RAND20B	RAND20C
<i>pol</i>	Geographical Distance	0.1	0.0	0.0	0.1	0.0	0.0	0.0
	Origin Sample Size	Inf	0.5	0.4	0.4	0.4	0.4	0.4
	Destination Sample Size	Inf	6583.3	5758.4	1674.5	Inf	Inf	Inf
	Origin HIV pop. size	0.6	1.4	1.1	1.6	1.2	1.1	1.2
	Dest. HIV pop. size	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Origin HIV prevalence	0.2	6.1	10.8	16.1	17.3	13.1	7.9
	Dest. HIV prevalence	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Origin HIV-1C pop. size	0.3	14.4	22.4	9.6	18.1	23.4	38.8
	Dest. HIV-1C pop. size	0.0	0.0	0.1	0.1	0.0	0.0	0.0
	Origin HIV-1C prevalence	0.3	7.2	1.3	4.1	1.1	1.4	1.0
	Dest. HIV-1C prevalence	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	<i>env</i>	Geographical Distance	0.0	0.1	0.1	0.1	0.0	0.0
Origin Sample Size		Inf	0.3	0.4	0.9	0.3	0.3	0.4
Destination Sample Size		Inf	20.6	32.5	29.4	Inf	Inf	Inf
Origin HIV pop. size		0.3	0.8	1.1	2.6	0.8	0.9	0.9
Dest. HIV pop. size		0.0	0.0	0.0	0.0	0.0	0.0	0.0
Origin HIV prevalence		0.3	19.6	18.9	15.3	22.3	26.0	23.4
Dest. HIV prevalence		0.0	0.0	0.0	0.0	0.0	0.0	0.0
Origin HIV-1C pop. size		0.3	13.8	14.0	15.5	13.0	11.3	12.9
Dest. HIV-1C pop. size		0.0	0.1	0.1	0.0	0.0	0.0	0.0
Origin HIV-1C prevalence		0.4	0.4	0.6	1.0	0.7	0.6	0.5
Dest. HIV-1C prevalence		0.0	0.6	0.6	0.6	0.0	0.0	0.0

Epidemiological predictors included in all Rand10 and Rand20 datasets are in bold, as well as $BF \geq 3$; Dest.: destination.

Two predictors, “origin HIV prevalence” and “origin subtype C population size”, were included in all pol and env Rand10 and Rand20 datasets with Bayes factor estimates ranging from moderate ($BF=6$) to strong ($BF=39$) support and with positive mean conditional effect sizes (Figure 4.2). Hence, locations with higher HIV prevalence and larger HIV-1C populations tend to act as sources for onwards spread.

In addition to epidemiological predictors we also tested geographical distance or air transportation data (in a separate analysis, data not shown) as a predictors of HIV-1C diffusion, but these did not result in noticeable support by any of the analyzed datasets.

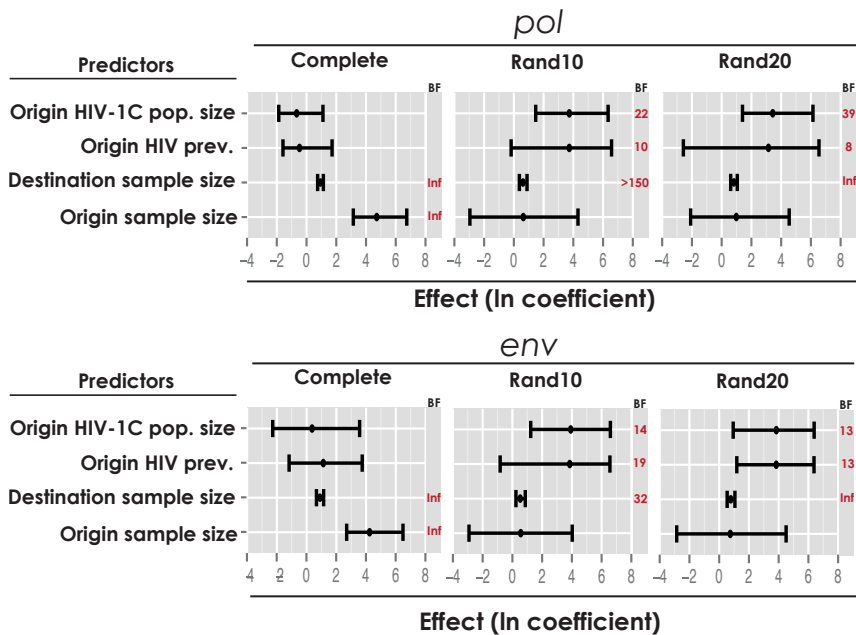


Figure 4.2. Significant predictors of the Brazilian HIV-1C epidemic spread among sub-sampling datasets of *pol* and *env* genes. Inclusion probabilities are represented as Bayes factors (BF) and a BF=3 threshold was used as positive indication of the predictor inclusion. The effect of each predictor, conditional to its inclusion, is represented by the posterior mean (black dot) and 95% HPD of the GLM coefficients in log scale.

Interestingly, incorporating relevant epidemiological information into the phylogeographic reconstructions resulted in consistent root state estimates: using the GLM model we find POA as the modal root state in all (12/12) *pol* and *env* Rand10 and Rand20 datasets. Only in the complete datasets, where the sampling bias is more severe, Florianópolis (FLP) was still estimated as the modal root state.

4.4.5 - Porto Alegre as a central hub of the HIV-1C epidemic

We subsequently estimated the most likely migration patterns using an asymmetrical phylogeographic analysis with BSSVS and priors on the location exchange rates that are based on the GLM rate

estimates. The robustness of the ancestral reconstructions was somewhat lower because in this analysis the predictors can only influence the analysis through the prior specification: POA was found the root state location in 10/12 pol and env Rand10 and Rand20 datasets (data not shown). Nonetheless, POA was strongly linked to all other locations (Bayes factors ≥ 3) while only a few additional well-supported transitions were found. Because this suggests a central role for POA in the Brazilian HIV-1C dissemination, we address its role in more detail.

The arrival of HIV-1C in POA was estimated in 1973 (1966 – 1980, 95% HPD) for pol and 1971 (1963-1978, 95%HPD) for env, and the spread to other cities started around 1980. The timing of these events reveals a consistent pattern. Nearby locations within RS (Rio Grande and Uruguaiiana cities) were initially affected, followed by export to south and southeast state capitals in the early 1980s (e.g. to Florianópolis, Curitiba, Rio de Janeiro and São Paulo). More distant locations were affected at a later stage, first in Central-West region (Campo Grande) in middle 1980s and later in North region (Palmas and Manaus) in the late 1980s and early 1990s. Only two exceptions to this pattern are found (one in the pol and one in the env datasets): the capital city Goiânia, where HIV-1C appears to have been introduced from POA in 1981 (pol), and Rio de Janeiro, where the introduction of HIV-1C has been more recent according to the env datasets (see Supplementary Table S4.4, for time of first transition from Porto Alegre).

More insights into the temporal pattern of spread were obtained by mapping the density of location transitions from POA to the other state capital cities through time. This reveals a period of higher density of viral influx 25 to 30 years ago to the South region capital cities Florianópolis and Curitiba. Among the sampled capitals in the Southeast region, a similar pattern emerged for Rio de Janeiro but there is a more evenly distributed transition density through time to São Paulo. Such a shift of transition density towards more recent times is slightly noticeable for the capital cities of the Central-West and North regions (see Supplementary Figure S4.5, for transitions by time from Porto Alegre).

4.5 - Discussion

We reconstructed the phylogeographic history of HIV-1C in Brazil using a comprehensive set of pol and env subtype C sequences from 22 different cities, of which 10 were sampled for the first time. Using a new model-testing based approach and by calculating several phylogeny-trait association measures using BaTS [28] we could reject random mixing in both datasets. However, as seen in MC scores, not all locations contributed equally to phylogenetic signal resulting in a considerable degree of uncertainty in the phylogeographic inferences. Nevertheless, after balancing the number of samples per location to mitigate the confounding effects of sampling biases, we were able to identify support for two epidemiological predictors of the viral spatial diffusion process.

Specifically, we found higher migration intensity from cities with larger numbers of HIV-1C infected patients and higher HIV prevalence. Interestingly, this is in agreement with a pattern of HIV-1C spread towards the north of Brazil, where the prevalence of HIV is smaller and only few cases of HIV-1C infection are found [4,6]. An intriguing result illustrating the complexity of modeling human mobility is that neither “geographical distance” nor “passenger air traffic” predicted viral spread. The sample size of source and/or recipient locations, on the other hand, were always included in the model (in isolation or together, Table 4.2). Samples sizes are expected to predict the number of transitions to some extent, and it was not our intention to formally demonstrate this. Rather, we wanted to avoid the circumstance that other predictors would yield support because of some degree of correlation with sample sizes.

To explore how sampling heterogeneity also impacts ancestral reconstructions, we analyzed six random down-sampled datasets in parallel with the complete pol and env datasets. This highlighted a substantial variability in the root state estimates (Table 4.1) and confirms that the sampling scheme can indeed have a profound effect on the inferred location state probabilities at the internal nodes of the tree. The impact of sampling biases was most likely aggravated by the relatively high degree of mixing observed in the pol and env datasets (see Supplementary Table S4.3, for MC scores). The geographical partitioning is also an important factor in discrete phylogeographic

analyses because this determines the level of spatial detail that can be recovered. Whereas previous studies investigating the spread of HIV-1C in Brazil discretized locations according to political states or geopolitical regions [7-9], we opted for a higher-resolution scheme and defined cities as the locations of interest. This allowed us to include more precise predictors in the GLM model.

We were able to largely resolve sampling-bias related inconsistencies by informing the phylogeographical reconstructions with relevant epidemiological information. Our results consistently identify POA, and not the state of Paraná [7-9], as the point of introduction. Several lines of evidence support this hypothesis. The population in the metropolitan area of POA has about 4 million inhabitants, the largest in the South region, and the AIDS incidence rate in POA and its metropolitan area is the highest rate in Brazil [6]. This suggests that the virus found ideal circumstances for onwards transmission and explains why the HIV-1C prevalence in Paraná's capital Curitiba is much lower (~22%) compared to POA (~40% and up to ~60% if the proportion of CRF31_BC - a local circulating form with a small subtype B insertion in a subtype C backbone- is considered) [5,34].

Differences in risk-group associations between the subtype B and C epidemics in Brazil also seem to support our findings. Whereas in POA the association between men-having-sex-with-men (MSM) and HIV-1B disappeared in more recent sampling because of an expansion of HIV-1C in heterosexual (HET) and MSM groups [35], compartmentalized epidemics are still observed in other cities from the South region, including in Paraná, which could be explained by a later introduction of HIV-1C [4,36-38].

Finally, a central role for POA is also reflected in the high support for transitions from POA to all other locations and the reconstructed temporal pattern of dissemination. After its introduction in the early 1970s, HIV-1C started spreading to other cities in the beginning of the 1980s, first to nearby locations and then to locations progressively further away. It is interesting to note that we could recover a noticeably higher fraction of recent jumps from POA to São Paulo when compared to transitions from POA to other South or South-East region capital cities, which points towards a strong longstanding epidemiological link between both cities.

Although our analysis provides support for POA as the central dissemination point of HIV-1C in Brazil, some caution is required when analyzing the number of transitions in star-like trees such as those typically found for HIV-1. The absence of clear phylogenetic structure deeper in the trees also offers little opportunity to capture clear spatial structuring and transitions beyond those out of the location state at the root. In the current work, our sampling strategy focused on broad geographic coverage rather than on a dense sampling, and a small sample from a large and diverse population that has grown exponentially through time, generally results in star-like topologies. Thus, despite the support for a central role of POA, we can recover little detail on viral spread beyond transitions out of this location.

In conclusion, we present a comprehensive reconstruction of the spatial and temporal dynamics of HIV-1C in Brazil based on pol and, for the first time, env sequence data, and included data from 10 newly sampled cities. By augmenting the viral genetic information with epidemiological data, we revealed a central role for POA city in the spread of HIV-1C in Brazil. In addition, we also identified locations with high HIV prevalence and large subtype C population sizes as key in the epidemic expansion towards the north of Brazil.

4.6 - Acknowledgements

We wish to thank all collaborating municipal health centers from Santa Catarina and Rio Grande do Sul.

Contributions: TG, ARP and SEMA conceived and designed the study, as well as coordinated the sampling. TG, RMM and DMJ performed samples amplification and sequencing. TG, BV, MAS and PL designed and performed phylogenetic and phylogeographic analyses. All authors read and critically reviewed the paper.

Sponsorship: This study was supported by Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the European Union Seventh Framework Programme [FP7/2007-2013] under Grant Agreement nr. 278433-PREDEMICS and ERC Grant Agreement nr. 260864.

The authors have no conflicts of interest.

4.7 - References

1. Faria NR, Rambaut A, Suchard MA, et al. The early spread and epidemic ignition of HIV-1 in human populations. *Science* **2014**; 346:56–61.
2. Tebit DM, Arts EJ. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis* **2011**; 11:45–56.
3. Ariën KK, Vanham G, Arts EJ. Is HIV-1 evolving to a less virulent form in humans? *Nat Rev Microbiol* **2007**; 5:141–51.
4. Gräf T, Pinto AR. The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. *Virology* **2013**; 435:179-178.
5. Inocencio LA, Pereira AA, Sucupira MCA, et al. Journal of the International AIDS Brazilian Network for HIV Drug Resistance Surveillance : a survey of individuals recently diagnosed with HIV. *J Int AIDS Soc* **2009**; 6:1–6.
6. Brazilian Ministry of Health. AIDS Epidemiological Bulletin July 2013-June 2014. Brasília; **2014**.
7. Véras NMC, Gray RR, de Macedo Brígido LF, Rodrigues R, Salemi M. High-resolution phylogenetics and phylogeography of human immunodeficiency virus type 1 subtype C epidemic in South America. *J Gen Virol* **2011**; 92:1698–709.
8. Delatorre E, Couto-Fernandez JC, Guimarães ML, et al. Tracing the origin and northward dissemination dynamics of HIV-1 subtype C in Brazil. *PLoS One* **2013**; 8:e74072.
9. Bello G, Zanotto PMA, Iamarino A, et al. Phylogeographic Analysis of HIV-1 Subtype C Dissemination in Southern Brazil. *PLoS One* **2012**; 7:e35649.

10. Brígido LFM, Rodrigues R. Southern Brazil HIV Type 1 C Expansion into the State of São Paulo, Brazil. *AIDS Res Hum Retroviruses* **2010**; 27:339-344.
11. Cardoso LPV, Pereira GAS, Pela LE, Rosado SM. HIV-1 Primary and Secondary Antiretroviral Drug Resistance and Genetic Diversity Among Pregnant Women From Central Brazil. **2010**; 357:351-357.
12. Carvalho BC, Cardoso LPV, Damasceno S, Stefani MM de A. Moderate Prevalence of Transmitted Drug Resistance and Interiorization of HIV Type 1 Subtype C in the Inland North State of Tocantins , Brazil. *AIDS Res Hum Retroviruses* **2011**; 27:1081-1087.
13. Ferreira AS, Cardoso LPV, Stefani MM de A. Moderate Prevalence of Transmitted Drug Resistance and High HIV-1 Genetic Diversity in Patients From Mato Grosso State , Central Western Brazil. *J Med Virol* **2011**; 1307:1301-1307.
14. da Silveira AA, Cardoso LPV, Francisco RBL, de Araújo SMM. HIV type 1 molecular epidemiology in pol and gp41 genes among naive patients from Mato Grosso do Sul State, central western Brazil. *AIDS Res Hum Retroviruses* **2012**; 28:304-307.
15. Lemey P, Rambaut A, Bedford T, et al. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog* **2014**; 10:e1003932.
16. Librelotto CS, Gräf T, Simon D, Almeida SEM de, Lunge VR. HIV-1 epidemiology and circulating subtypes in the countryside of South Brazil. *Rev Soc Bras Med Trop* **2015**; 48:249-257.
17. de Oliveira T, Deforche K, Cassol S, et al. An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* **2005**; 21:3797-3800.
18. Pond SLK, Posada D, Stawiski E, et al. An evolutionary model-based algorithm for accurate phylogenetic breakpoint mapping and subtype prediction in HIV-1. *PLoS Comput Biol* **2009**; 5: e1000581.
19. Siepel AC, Halpern AL, Macken C, Korber BT. A computer program designed to screen rapidly for HIV type 1 intersubtype

- recombinant sequences. *AIDS Res Hum Retroviruses* **1995**; 11:1413–1416.
20. Lole KS, Bollinger RC, Paranjape RS, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **1999**; 73:152–160.
21. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinformatics* **2009**; 10:421.
22. Gifford RJ, Liu TF, Rhee SY, et al. The calibrated population resistance tool: Standardized genotypic estimation of transmitted HIV-1 drug resistance. *Bioinformatics* **2009**; 25:1197–1198.
23. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **2012**; 29:1969–1973.
24. Ayres DL, Darling A, Zwickl DJ, et al. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Syst Biol* **2012**; 61:170–173.
25. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol* **2008**; 25:1459–1471.
26. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* **2006**; 4:699–710.
27. Shapiro B, Ho SYW, Drummond AJ, Suchard MA, Pybus OG, Rambaut A. A bayesian phylogenetic method to estimate unknown sequence ages. *Mol Biol Evol* **2011**; 28:879–887.
28. Parker J, Rambaut A, Pybus OG. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect Genet Evol* **2008**; 8:239–246.
29. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol* **2012**; 29:2157–2167.

30. Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. *Mol Biol Evol* **2013**; 30:239–243.
31. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. *PLoS Comput Biol* **2009**; 5:e1000520.
32. Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* **2011**; 27:2910–2912.
33. O'Brien JD, Minin VN, Suchard MA. Learning to count: Robust estimates for labeled distances between molecular sequences. *Mol Biol Evol* **2009**; 26:801–814.
34. Passaes CPB, Bello G, Lorete RS, et al. Genetic characterization of HIV-1 BC recombinants and evolutionary history of the CRF31_BC in Southern Brazil. *Infect Genet Evol* **2009**; 9:474–482.
35. Almeida SE, de Medeiros RM, Junqueira DM, et al. Temporal dynamics of HIV-1 circulating subtypes in distinct exposure categories in southern brazil. *Virology* **2012**; 9:306.
36. Raboni SM, Almeida SM De, Rotta I, et al. Molecular epidemiology of HIV-1 clades in Southern Brazil. *Memorias do Inst Oswaldo Cruz* **2010**; 105:1044–1049.
37. Gräf T, Passaes CPB, Ferreira LGE, 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* Published **2011**; 5:186-191.
38. Silveira J, Santos AF, Martínez AMB, et al. Heterosexual transmission of human immunodeficiency virus type 1 subtype C in southern Brazil. *J Clin Virol* **2012**; 54:36-41.

4.8 - Supplementary Material

Supplementary Text S4.1 - Additional information on Methods

Patients, samples and new sequences:

From October 2009 to February 2014 HIV-1 seropositive patients from 13 cities in Santa Catarina and Rio Grande do Sul were invited to enroll in this study which was approved by the ethics committees of the Federal University of Santa Catarina (UFSC) and the Foundation of Research and Production in Health of the Rio Grande do Sul state (FEPPS-RS). Informed consent was obtained from all patients enrolled.

Briefly, 360 blood samples were collected in FTA cards (Whatman Inc.) and purified to downstream molecular procedures following the manufacturer's protocol. HIV-1 envelope (HXB2 6846-7360 bp) and polymerase (HXB2 2274-3545 bp) fragments were amplified by nested-PCR and sequenced as described elsewhere [1]. Sequences were assembled and visually inspected using the SeqMan software in the LaserGene package (DNASTAR).

Sequence dataset compilation:

We constructed an as complete as possible dataset of Brazilian *pol* and *env* subtype C sequences (HIV-1C_BR) with sampling time and location annotation by complementing our new sequence data with publically available HIV-1C sequences. To this end, all *pol* and *env* HIV-1C_BR sequences were downloaded from the Los Alamos HIV database ($n = 385$). When sampling location and/or time was not available, we attempted to retrieve this from the study reporting the sequences. We added the 50 best hits per query in a BLAST search of HIV1C_BR against all non-Brazilian HIV-1 subtype C sequences [2], but kept only those assigned as subtype C by REGA, RIP and SCUEAL [3-5] for further analyses. To avoid erroneous clustering due to convergent evolution as a consequence of drug selective pressure [6] we checked all *pol* sequences -including the newly generated ones- for the presence of Surveillance Drug Resistance Mutations (SDRM) by submitting them to the CPR tool available in the Stanford HIV Database (<http://hivdb.stanford.edu>) [7]. Low-quality sequences,

sequences with ≥ 5 SDRMs and sites with a SDRM detected in at least two different sequences were removed.

The sequences were aligned using Muscle [8] and the alignment was manually edited using AliView [9]. Phylogenetic trees were reconstructed using RAxML [10] under a GTR+I+ Γ 4 nucleotide substitution model and Brazilian sequences that clustered outside the general HIV-1C_BR clade were removed. Sequences outliers in a regression of root-to-tip divergence versus sampling time identified using Path-O-Gen software (<http://tree.bio.ed.ac.uk/software/pathogen/>) were also removed. This procedure resulted in datasets with 1522 *pol* and 621 *env* sequences, which were down-sampled to around 500 sequences each. For this we kept all Brazilian subtype C sequences and selected 120 *pol* and 170 *env* international sequences using CD-HIT [11]. Six additional sub-samplings containing only Brazilian sequences were made for *pol* and *env* (see the main text).

Adjusting predictors for phylogeographic inference

All epidemiological information was collected from the DATASUS system, a databank hosted by the Brazilian Ministry of Health (<http://www2.datasus.gov.br/DATASUS/index.php?area=0203>). To extract the GLM predictors, we focused on a 10 years period (2004-2013) as a representative time frame for the current infected population. Population data and geographic coordinates were obtained from the Brazilian Institute of Geography and Statistics (IBGE) (<http://cidades.ibge.gov.br/xtras/home.php>) and air traffic information was collected as previously described [12]. For location-specific measures (3 to 7), we considered separate origin and destination predictors for the rates in the GLM design matrix, which implies that, for origin predictors, the measures for the origin locations are used to predict the pairwise transition rates between the origin and destination location and vice versa.

References:

1. Librelotto CS, Gräf T, Simon D, Almeida SEM de, Lunge VR. HIV-1 epidemiology and circulating subtypes in the countryside of South Brazil. *Rev Soc Bras Med Trop* **2015**; 48:249–257.
2. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinformatics* **2009**; 10:421.
3. de Oliveira T, Deforche K, Cassol S, et al. An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* **2005**; 21:3797–3800.
4. Pond SLK, Posada D, Stawiski E, et al. An evolutionary model-based algorithm for accurate phylogenetic breakpoint mapping and subtype prediction in HIV-1. *PLoS Comput Biol* **2009**; 5:e1000581
5. Siepel AC, Halpern AL, Macken C, Korber BT. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Res Hum Retroviruses* **1995**; 11:1413–1416.
6. 1. Lemey P, Derdelinckx I, Rambaut A, et al. Molecular footprint of drug-selective pressure in a human immunodeficiency virus transmission chain. *J Virol* **2005**; 79:11981–11989.
7. Gifford RJ, Liu TF, Rhee SY, et al. The calibrated population resistance tool: Standardized genotypic estimation of transmitted HIV-1 drug resistance. *Bioinformatics* **2009**; 25:1197–1198.
8. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **2004**; 5:113.
9. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics* **2014**; 30:3276–3278.
10. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**; 30:1312–1313.
11. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics* **2010**; 26:680–682.

12. Lemey P, Rambaut A, Bedford T, et al. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog* **2014**; 10:e1003932

Supplementary Table S4.2. Number of Brazilian sequences according to state and city in the complete dataset and sub-samplings for *pol* and *env* genes.

State	City	Sampling time span	<i>pol</i> datasets			State	City	Sampling time span	<i>env</i> datasets		
			Complete	Rand10	Rand20				Complete	Rand10	Rand20
RS	CRA	2011-2012	17	10	17	RS	CRA	2011-2012	19	10	19
	CXS	2013-2014	11	10	11		CXS	2013-2014	18	10	18
	LAJ	2013	2	6	6		LAJ	2013	15	10	15
	POA	2002-2011	50	10	20		POA	2003-2011	18	10	18
	RIG	2002-2007	20	10	20		STI	2013	11	10	11
	STI	2013	2	2	2		STL	2013-2014	8	8	8
	STL	2013-2014	4	4	4		URU	2009-2010	9	9	9
	URU	2009-2010	6	6	6						
SC	BLU	2012	15	10	15	SC	BLU	2012	19	10	19
	CHA	2012	10	10	10		CHA	2012	16	10	16
	CRI	2007-2013	30	10	20		CRI	2007-2013	20	10	20
	FLP	2008-2011	63	10	20		FLP	2008-2011	64	10	20
	ITA	2004-2014	35	10	20		ITA	2004-2014	16	10	16
	JOI	2012-2013	25	10	20		JOI	2012-2013	34	10	20
LGE	2013	3	3	3	LGE	2013	5	5	5		
GO	GOI	2003-2011	17	10	17	AM	MAN	2006	2	2	2
MS	CPG	2008-2010	4	4	4	RJ	RJN	2008-2010	3	3	3
PR	CTB	2004-2006	15	10	15	SP	SPL	2003-2010	12	10	12
RJ	RJN	2004-2011	19	10	19						
SP	SPL	2003-2010	21	10	20						
TO	PAL	2008	2	2	2						
	TOTAL		371	167	271		TOTAL		289	147	231

Acronyms for states: AM: Amazonas; GO: Goiás; MS: Mato Grosso do Sul; PR: Paraná; RJ: Rio de Janeiro; SC: Santa Catarina; SP: São Paulo; TO: Tocantins; RS: Rio Grande do Sul. Acronyms for cities: BLU: Blumenau; CHA: Chapecó; CPG: Campo Grande; CRA: Cruz Alta; CRI: Criciúma; CTB: Curitiba; CXS: Caxias do Sul; FLP: Florianópolis; GOI: Goiania; ITA: Itajaí; JOI: Joinville; LAJ: Lajeado; LGE: Lages; MAN: Manaus; PAL: Palmas; POA: Porto Alegre; RIG: Rio Grande; RJN: Rio de Janeiro; SPL: São Paulo; STI: Santiago; STL: Santana do Livramento; URU: Uruguiana.

Supplementary Table S4.3. Monophyletic Clade analysis for each location in *pol* and *env* datasets

City	T Sample Number	<i>pol</i> - Monophyletic Clade (MC)						
		observed mean	lower 95% CI	upper 95% CI	null mean	lower 95% CI	upper 95% CI	<i>p</i> value
FLP	63	2.89	2	4	2.30	1.92	3.03	0.091
BLU	15	2.00	2	2	1.17	1.00	2.00	0.048*
CHA	10	2.00	2	2	1.07	1.00	1.38	0.02*
CRA	17	3.03	3	3	1.23	1.00	2.00	0.002*
CRI	30	2.55	2	3	1.59	1.06	2.09	0.008*
CXS	11	1.01	1	1	1.10	1.00	1.85	1.000
ITA	35	2.54	2	3	1.74	1.14	2.18	0.341
JOI	25	2.00	2	2	1.45	1.02	2.02	0.171
LAJ	2	1.00	1	1	1.03	1.00	1.12	1.000
LGE	3	1.00	1	1	1.01	1.00	1.00	1.000
POA	50	6.10	4	10	2.05	1.54	2.89	0.001*
STI	2	1.09	1	2	1.00	1.00	1.00	1.000
STL	4	2.00	2	2	1.01	1.00	1.01	0.002*
URU	6	5.90	6	6	1.03	1.00	1.12	0.001*
GOI	17	2.02	2	2	1.24	1.00	2.00	0.080
CPG	4	1.00	1	1	1.01	1.00	1.01	1.000
PAL	2	1.00	1	1	1.00	1.00	1.00	1.000
SPL	21	3.05	3	3	1.34	1.01	2.00	0.003*
RJN	19	4.00	4	4	1.26	1.00	2.00	0.001*
RIG	20	2.05	2	2	1.31	1.00	2.00	0.118
CTB	15	2.00	2	2	1.18	1.00	2.00	0.057

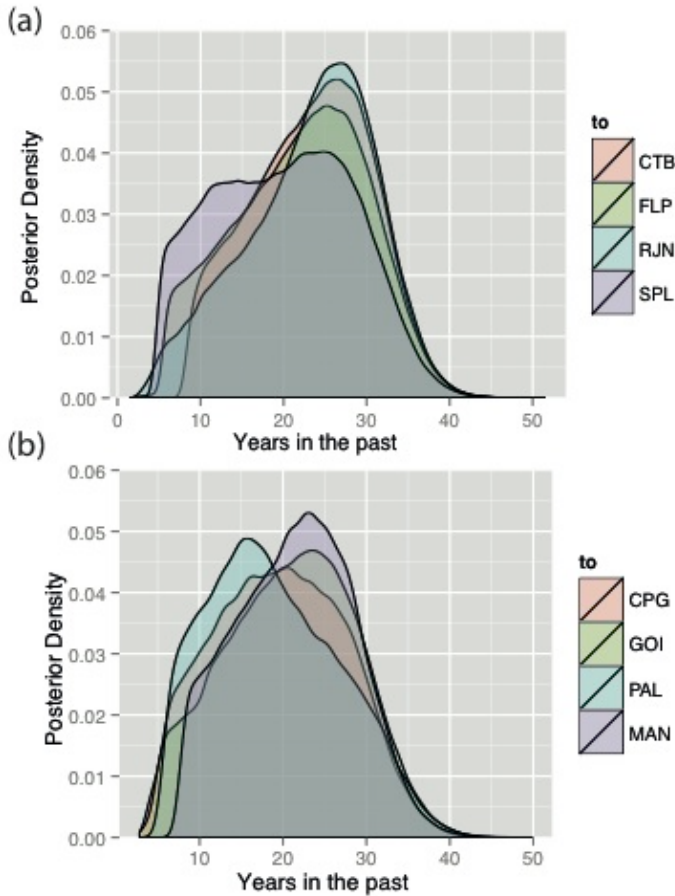
City	Sample Number	<i>env</i> - Monophyletic Clade (MC)						
		observed mean	lower 95% CI	upper 95% CI	null mean	lower 95% CI	upper 95% CI	<i>p</i> value
BLU	19	2.00	2	2	1.36	1.00	2.01	0.137
CHA	16	2.00	2	2	1.26	1.00	2.00	0.093
CRA	19	2.04	2	2	1.36	1.00	2.01	0.144
CRI	20	2.00	2	2	1.39	1.01	2.01	0.159
CXS	18	3.99	4	4	1.33	1.00	2.00	0.001*
FLP	64	3.16	2	4	2.45	2.04	3.21	0.111
ITA	16	2.01	2	2	1.25	1.00	2.00	0.084
JOI	34	2.06	2	3	1.85	1.17	2.33	0.483
LAJ	15	2.00	2	2	1.22	1.00	2.00	0.072
LGE	5	1.00	1	1	1.02	1.00	1.05	1.000
POA	18	1.60	1	2	1.33	1.00	2.00	0.129
STI	11	2.01	2	2	1.13	1.00	1.95	0.038*
STL	8	2.00	2	2	1.06	1.00	1.46	0.02*
URU	9	4.99	2	6	1.08	1.00	1.63	0.001*
SPL	12	2.00	2	2	1.15	1.00	2.00	0.054
RJN	3	2.00	2	2	1.01	1.00	1.01	0.005*
MAN	2	1.00	1	1	1.00	1.00	1.00	1.000

* Significant *p* values ($p < 0.05$); CI: confidential interval. **Acronyms for cities:** BLU: Blumenau; CHA: Chapecó; CPG: Campo Grande; CRA: Cruz Alta; CRI: Criciúma; CTB: Curitiba; CXS: Caxias do Sul; FLP: Florianópolis; GOI: Goiania; ITA: Itajaí; JOI: Joinville; LAJ: Lajeado; LGE: Lages; MAN: Manaus; PAL: Palmas; POA: Porto Alegre; RIG: Rio Grande; RJN: Rio de Janeiro; SPL: São Paulo; STI: Santiago; STL: Santana do Livramento; URU: Uruguaiana.

Supplementary Table S4.4: Time of first transition and 95%HPD from Porto Alegre to all Brazilian sampled cities as estimated for *pol* and *env* Rand20 datasets*

City	<i>pol</i>	<i>env</i>
	Year (95% HPD)	Year (95% HPD)
RIG	1979 (1970-1986)	NA
FLP	1980 (1972-1987)	1979 (1969-1987)
URU	1980 (1972-1987)	1979 (1971-1986)
CRI	1981 (1973-1987)	1980 (1971-1988)
CTB	1981 (1973-1987)	NA
ITA	1981 (1973-1987)	1983 (1974-1991)
RJN	1981 (1973-1987)	1993 (1980-2005)
GOI	1981 (1973-1988)	NA
SPL	1981 (1973-1988)	1983 (1973-1991)
BLU	1982 (1974-1989)	1982 (1972-1989)
CRA	1982 (1974-1989)	1980 (1971-1989)
JOI	1982 (1974-1989)	1980 (1972-1989)
CXS	1982 (1974-1990)	1982 (1972-1989)
LAJ	1985 (1975-1995)	1982 (1972-1990)
CHA	1985 (1976-1994)	1982 (1973-1990)
CPG	1986 (1977-1997)	NA
STL	1987 (1977-1999)	1985 (1976-1995)
LGE	1988 (1976-2003)	1989 (1979-1999)
STI	1988 (1977-2003)	1983 (1973-1991)
MAN	NA	1989 (1978-2001)
PAL	1990 (1977-2003)	NA

*mean from the three Rand20 sub-samplings. **HPD**: highest probability density; **NA**: not available; **Acronyms for cities**: BLU: Blumenau; CHA: Chapecó; CPG: Campo Grande; CRA: Cruz Alta; CRI: Criciúma; CTB: Curitiba; CXS: Caxias do Sul; FLP: Florianópolis; GOI: Goiania; ITA: Itajaí; JOI: Joinville; LAJ: Lajeado; LGE: Lages; MAN: Manaus; PAL: Palmas; POA: Porto Alegre; RIG: Rio Grande; RJN: Rio de Janeiro; SPL: São Paulo; STI: Santiago; STL: Santana do Livramento; URU: Uruguaiana.



Supplementary Figure S4.5. Estimated density of location transitions by time from Porto Alegre to other Brazilian capitals. Smooth density estimates of all transitions counted in *pol* and *env* Rand20 datasets analysis by time, with origin in Porto Alegre. Time zero in Y-axis is equal to 2014. Colors are according to the legend on the right. a) Transitions to South (FLP and CTB) and Southeast (RJN and SPL) regions; b) Transitions to Central-West (CPG and GOI) and to North (MAN and PAL) regions. CTB: Curitiba; FLP: Florianópolis; RJN: Rio de Janeiro; SPL: São Paulo; CPG: Campo Grande; GOI: Goiânia; PAL: Palmas; MAN: Manaus.

CAPÍTULO 5.

Discussão geral

A epidemia de aids no Rio Grande do Sul e em Santa Catarina é preocupante no cenário nacional. As taxas de incidência chegam a ser cinco vezes mais altas do que a média brasileira, impactando a taxa de mortalidade por aids, que também é superior à média do país (Departamento de DST, aids e Hepatites Virais, 2014). A grande prevalência do HIV-1C nesses dois estados acaba por gerar questionamentos sobre o impacto dessa variante viral no cenário epidemiológico local. Seria o HIV-1C o responsável pelo rápido crescimento da epidemia de aids no Sul do país ou estaria ele apenas se beneficiando das condições locais para disseminar-se em grandes proporções nessa região?

Atualmente, o HIV-1C é o subtipo mais prevalente no mundo, principalmente devido a seu protagonismo na epidemia da Índia e dos países do sul da África, locais que possuem uma enorme população e uma grande prevalência de soropositivos para o HIV, respectivamente (Tebit e Arts, 2011). Além dessas regiões, estudos recentes têm demonstrado um crescente número de novas infecções por HIV-1C em países da Ásia e da Europa, bem como na Austrália e no Brasil (ver seção 2.2). Como possível explicação para esse fenômeno expansivo do HIV-1C, podem ser citados alguns estudos que reportaram menor patogenicidade e igual transmissibilidade dessa variante em comparação com outros subtipos, o que aumentaria o período assintomático e possibilitaria um maior número de transmissões (ver seção 2.9).

Entretanto, o efeito desse distinto potencial patogênico dos subtipos do grupo M do HIV-1 não é tão claro ao acompanhar o tempo de progressão à aids. Estudos que avaliaram o histórico de medidas de carga viral e contagens de células CD4 em coortes de indivíduos infectados com diferentes subtipos vêm reportando resultados divergentes sobre o tema. Enquanto o subtipo D tem sido relacionado com progressão rápida em vários estudos, o HIV-1C pode (Touloumi et al., 2013) ou não ser responsável por uma progressão mais lenta (Easterbrook et al., 2010; Nunes et al., 2014; Kiwanuka et al., 2010). Essa falta de consistência dos achados evidencia a necessidade de mais estudos, mas também destaca que maiores prevalências de um subtipo em detrimento de outros que estejam cocirculando na mesma região podem ser simplesmente artefatos de efeitos fundadores em cadeias de transmissão separadas e de dinâmicas de transmissão próprias.

Independentemente de quais fatores estariam agindo na epidemia de aids no Sul do Brasil, os resultados apresentados aqui apontam para uma taxa de expansão maior da epidemia do HIV-1C em comparação com o HIV-1B. Nossas análises foram executadas em uma grande população, amostrada de tal maneira que representasse igualmente todas as regiões dos estados de SC e RS. Além disso, as análises de reconstrução da história demográfica não mostraram estar enviesadas pela disparidade do n amostral entre sequências de HIV-1C e HIV-1B. Na casuística brasileira, taxas diferenciadas de expansão da epidemia podem estar sendo afetadas pela ação de cadeias parcialmente separadas de transmissão heterossexual e HSH. Nossas análises filodinâmicas mostraram uma acentuada explosão populacional do HIV-1C no estado de SC, onde também foram observadas as associações entre essa variante e a transmissão heterossexual e entre o HIV-1B e HSH. No estado do RS, essa separação não foi verificada no nosso grupo amostral, o que também pode explicar taxas de expansão um pouco mais similares entre HIV-1C e HIV-1B e o maior número de recombinantes BC encontrados.

Em regiões do mundo onde o HIV-1C é predominante, como na África do Sul e na Índia, as infecções ocorrem principalmente por via heterossexual (WHO, 2010). No Brasil, essa forma de exposição também é a mais frequente, mas é interessante observar que a razão entre os índices de infecções heterossexuais e HSH varia obedecendo a um padrão semelhante à prevalência do HIV-1C. Atualmente, nos três estados da região Sul, infecções por via heterossexual são de 4 a 6 vezes mais frequentes do que infecções no grupo HSH, enquanto nos três principais estados do Sudeste (São Paulo, Rio de Janeiro e Minas Gerais), essa relação é de 1,5 a 3,5 vezes, coincidindo com a média nacional (Figura 5.1). Já nos estados do Centro-Oeste e no Tocantins, o número de infecções heterossexuais pode ser ainda maior do que na região Sul, com uma frequência entre 4 e 8 vezes maior do que infecções em HSH. A exceção está no estado de Goiás, que apresentou, recentemente, um grande aumento no número de infecções no grupo HSH (Ministério da Saúde do Brasil, 2015).

Em concordância com esses dados, pode-se observar que a prevalência do HIV-1C é muito maior nos estados do Sul do Brasil e vem aumentando consideravelmente nos estados do Centro-Oeste e no Tocantins (Cardoso et al., 2009 e 2011; Ferreira et al., 2011; Carvalho

et al., 2011; Alcântara et al., 2012; Costa et al., 2013). A relação entre infecções por via heterossexual e o HIV-1C fica ainda mais clara no estado de Santa Catarina. Na Figura 5.1, é possível observar um aumento mais rápido do número de heterossexuais infectados em SC em comparação aos outros estados do Sul e ao Sudeste. No início da década de 1990, essa categoria de exposição já era mais numerosa do que o grupo HSH em SC. Nesse estado, o gráfico de crescimento demográfico do HIV-1C (Figura 3.3) reconstrói uma dinâmica muito semelhante a essa, destacando a precisão das análises aqui realizadas e a importância da relação entre o subtipo C e a via de transmissão heterossexual.

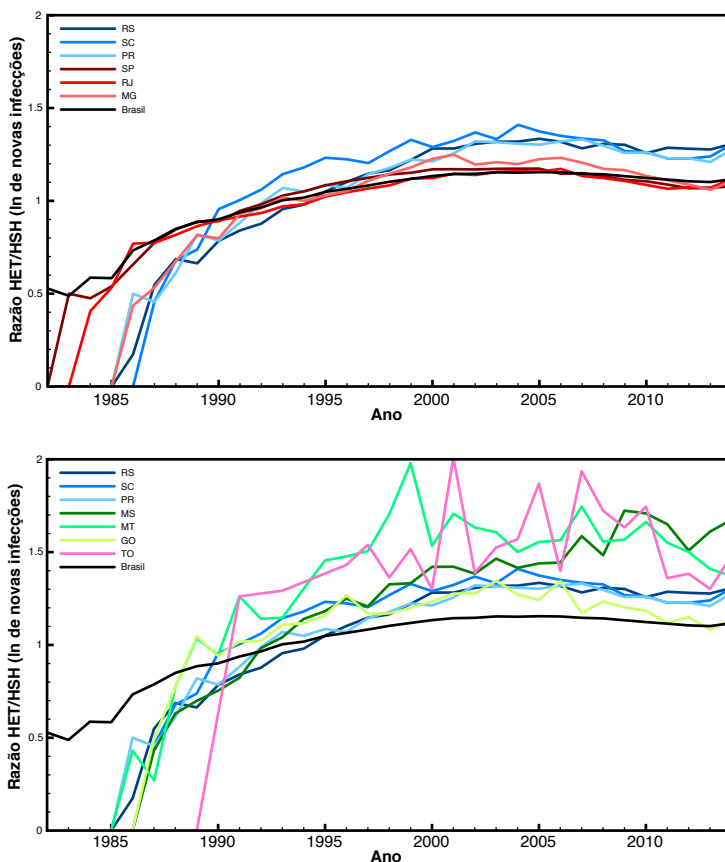


Figura 5.1 – 5.1 – Relação de novas infecções anuais entre HET e HSH por estado da federação. Dados históricos do número de infecções por categoria de exposição nos estados de interesse foram obtidos no sistema DATASUS. Para uma melhor apresentação visual dos gráficos, o logaritmo natural (ln) dos dados brutos foi utilizado para o cálculo da razão. HET – heterossexuais; HSH – homens que fazem sexo com homens.

Torna-se, então, factível a hipótese de que as cadeias de transmissão heterossexual estejam agindo na dispersão do HIV-1C pelo Brasil. Esse pode ser o principal fator que possibilitou o eficiente estabelecimento do HIV-1C na região Sul, ao mesmo tempo que limita sua prevalência no Sudeste e promove sua recente difusão pelo Centro-Oeste. Alternativamente, também podemos considerar que a circulação do HIV-1C talvez esteja levando ao “alargamento” das redes de transmissão heterossexual, aumentando o número de pessoas infectadas por essa via. Novamente, permanece a dúvida sobre o que é causa e o que é efeito na relação entre o HIV-1C e a dinâmica da epidemia de aids brasileira.

As análises filogeográficas apresentadas aqui destacam a dispersão do HIV-1C rumo ao norte do país. A partir da origem da epidemia no Brasil, em Porto Alegre, no início da década de 1970, o HIV-1C tem sido introduzido progressivamente em lugares mais ao norte do país ou em regiões remotas, como algumas cidades do interior do RS e de SC. A introdução nas capitais da região Sul e nas cidades de São Paulo e Rio de Janeiro aconteceu quase ao mesmo tempo (entre o fim da década de 1970 e o início da década de 1980), mas apenas em Curitiba e Florianópolis o HIV-1C efetivamente disseminou-se pela população. As diferentes características da epidemia de aids em São Paulo e no Rio de Janeiro em comparação com as capitais do Sul, discutidas anteriormente, podem ter sido responsáveis por manter o HIV-1C circulando em baixa prevalência. Por outro lado, a introdução no Tocantins e em Mato Grosso do Sul ocorreu a partir da metade da década de 1980, e as condições presentes nesses estados talvez possam favorecer uma expansão mais efetiva. De fato, estudos recentes têm relatado prevalências de mais de 10% de subtipo C ou de formas recombinantes contendo HIV-1C nos estados do Centro-Oeste e no Tocantins (Alcântara et al., 2012; Costa et al., 2013; Carvalho et al., 2011;

Cardoso et al., 2011). A hipótese de que Porto Alegre possa ter sido a origem e o centro de disseminação da epidemia do HIV-1 subtipo C brasileira é originada de resultados que só puderam ser observados por meio da utilização de modernas técnicas estatísticas e de modelagem filogenética. O sinal temporal contido nas sequências aqui analisadas não era suficientemente forte para permitir uma inferência confiável da localização geográfica designada aos nodos ancestrais das árvores. Esses resultados inconsistentes só foram detectados após 12 subamostragens randômicas do banco de sequências completo serem analisadas. O objetivo foi justamente avaliar a robustez dos resultados da análise filogeográfica. A dificuldade de gerar resultados consistentes entre as subamostragens provavelmente decorreu de uma amostragem pouco densa de uma população muito grande e de crescimento rápido, impondo às árvores longos ramos terminais e pequenos ramos internos que coalescem rapidamente. Árvores com esse tipo de topologia são comuns para o HIV e possuem uma estrutura que se assemelha a uma estrela, em que todos os ramos partem de um ponto central comum. Esse tipo de topologia dificulta qualquer tipo de inferência que tente reconstruir o estado ancestral de um traço.

Apesar de os testes de sinal filogenético suportarem que havia uma estruturação geográfica nas árvores reconstruídas (ver seção 4.4.2), apenas após a inclusão de dados externos às filogenias foi possível fazer uma inferência robusta acerca da história filogeográfica do HIV-1C no Brasil. Fatores como o tamanho da população infectada pelo subtipo C e a prevalência de HIV foram escolhidos como preditores da difusão viral, impelindo a epidemia para locais menos afetados por esses fatores, ou seja, do sul para o norte. Os dados moleculares complementados pelos preditores epidemiológicos possibilitaram a reconstrução de um padrão de dispersão consistente com a situação da epidemia do HIV observada atualmente, ressaltando a utilidade de incorporar dados ecológicos às análises filogeográficas. A técnica de GLM integrada à abordagem Bayesiana foi recentemente disponibilizada para ser usada no programa BEAST e já tem mostrado grande potencial em estudos de vigilância epidemiológica e até em modelos de predição de surgimento de epidemias (Lemey et al., 2014; Magee et al., 2015). Nesta tese, demonstramos pela primeira vez a sua utilidade para melhorar a robustez de análises filogeográficas.

No Capítulo 4, foi ainda abordada uma questão técnica muito importante em estudos epidemiológicos e análises estatísticas em geral: o viés de amostragem. Neste estudo, o desenho do n amostral de cada cidade participante foi feito por meio da estratificação pelo tamanho de sua população. Entretanto, nem sempre o número idealizado foi atingido, e, para muitas amostras, não foi possível gerar sequências nucleotídicas. Além disso, utilizamos todas as sequências de HIV-1C do Brasil disponíveis nos bancos públicos, desde que aprovadas no controle de qualidade. Isso tudo acabou gerando uma heterogeneidade muito grande no número de sequências por localidade, tornando necessária a aplicação de métodos para detectar a influência do número de amostras nas inferências. A partir da inclusão do número amostral de cada localidade na análise de seleção de preditores, foi possível detectar o quanto suscetíveis as análises filogeográficas são ao viés de amostragem. Locais mais bem amostrados apresentaram maiores chances de serem estimados como a localização ancestral dos nodos. Isso se deve ao fato de a taxa de transição inferida para cada par de localidades ser influenciada pela frequência com que cada localidade aparece no sítio em análise, da mesma forma que a frequência dos nucleotídeos interfere na taxa de substituição entre eles.

Por meio de um processo de randomização que buscou diminuir a heterogeneidade amostral entre as cidades, conseguiu-se suavizar o efeito do número de amostras na inferência filogeográfica. A partir disso, foi possível detectar o impacto de reais preditores sobre o processo de difusão viral, permitindo a reconstrução filogeográfica de forma consistente, mesmo que alguma interferência do número de amostras ainda estivesse agindo basalmente.

O grande número de localidades amostradas aqui permitiu que, pela primeira vez, a diversidade molecular do HIV-1 fosse descrita em uma extensa região interiorana do Brasil. Especialmente em cidades do RS, um número variável de formas do HIV-1 foi observado. Dentre elas, destaca-se a grande diversidade em Uruguaiiana, apresentando os subtipos B, C e D e as formas recombinantes BC, BF1 e BCF1. Uruguaiiana é uma cidade localizada na fronteira com a Argentina e possui grande importância para as relações comerciais entre esse país e o Brasil. Em seu porto seco, transitam mais de 150 mil caminhões por ano, entre exportações e importações (Nino, 2011). Esse grande fluxo populacional pode estar impulsionando a troca de variantes virais entre

os dois países, exportando o HIV-1C para a Argentina e importando formas BF1, muito prevalentes por lá (Pando et al., 2011). Essa última variante também foi observada em Chapecó e em Itajaí, cidades que podem ter vínculos internacionais devido à relativa proximidade com a fronteira e à presença de um grande porto internacional, respectivamente.

Além da circulação da CRF31_BC, praticamente só observada no RS e mais acentuada em cidades próximas a Porto Alegre, foram identificadas algumas possíveis novas CRFs entre HIV-1C e B. Elas também foram observadas em cidades do RS e reforçam a hipótese de que os subtipos B e C circulam nas mesmas redes de transmissão nesse estado. Por fim, cabe ressaltar que a frequência de formas recombinantes reportada aqui pode ser ainda maior, uma vez que, para algumas amostras, apenas sequências de *env* estavam disponíveis, e recombinações são muito mais prováveis de ocorrer em regiões mais conservadas do genoma do HIV, como, por exemplo, no gene *pol*.

A descentralização dos serviços de saúde e a adaptação às demandas regionais estão entre os principais objetivos da nova Política Nacional de Atenção Básica em Saúde, sendo a pesquisa parte essencial desse processo (Secretaria de Atenção à Saúde, 2012). Os resultados apresentados nesta tese de doutorado trazem uma descrição detalhada sobre a epidemia do HIV em cidades do interior do RS e de SC. Entre os principais resultados, estão a identificação de uma rápida expansão da epidemia do HIV-1C e o papel central da cidade de Porto Alegre na disseminação dessa variante para outros locais do Brasil. Esses achados vão ao encontro da degradante situação da epidemia de aids observada nesses dois estados e destacam a necessidade de medidas de ação de combate ao HIV. Campanhas que enfoquem a transmissão heterossexual, em detrimento daquelas voltadas aos grupos historicamente mais associados à infecção por HIV (ex.: HSH, UDI e profissionais do sexo), podem ser mais eficazes na diminuição da escalada do HIV na região Sul, talvez com um efeito ainda mais pronunciado na contenção do HIV-1C. Tais medidas, aplicadas aos estados do Centro-Oeste, podem servir como uma medida profilática e impedir que um cenário semelhante ao do RS e de SC se instale naquela região.

Entretanto, ainda são necessários estudos que descrevam comportamentos que possam predispor diferentes populações no Brasil à

infecção pelo HIV. Esse tipo de abordagem poderia elucidar, por exemplo, se a grande incidência de aids no RS e em SC é causada por uma maior exposição da população ao vírus, decorrente de comportamentos de risco mais frequentes. Na contramão do que seria esperado, uma das poucas pesquisas sobre o tema em âmbito nacional indica que indivíduos morando na região Sul do Brasil utilizam com mais frequência preservativos na primeira relação sexual (Departamento de DST, aids e hepatites virais, 2011). Se as atitudes e práticas relacionadas à infecção por HIV na população da região Sul não condizem com os altos índices da epidemia, a questão biológica sobre o diferente potencial de disseminação dos subtipos do HIV-1 deve ser considerada. Nesse cenário, a importância do diagnóstico precoce seria ainda mais destacada, uma vez que, com uma fase assintomática mais longa, o HIV-1C promoveria um maior número de transmissões sem o indivíduo saber de seu estado sorológico.

Por fim, permanece a mesma pergunta que abriu este capítulo. Para respondê-la, são necessários estudos comparativos entre o HIV-1 subtipo C e o subtipo B em uma grande população, principalmente para investigar o tempo de progressão à aids e o tempo de transmissão das redes nas duas formas virais. Assim, os resultados apresentados nesta tese trazem valiosas informações que podem guiar futuras pesquisas que visem ao combate da epidemia de aids no Brasil e ao esclarecimento de diferentes potenciais epidêmicos dos subtipos do HIV-1.

5.1 – Referências

- Alcântara KC, Reis MN, Cardoso LP, Bello G, Stefani MM. Increasing heterosexual transmission of HIV-1 subtype C in Inland Central Western Brazil. *J Med Virol.* 2013;85(3):396-404.
- Cardoso LP, da Silveira AA, Francisco RB, da Guarda Reis MN, Stefani MM. Molecular characteristics of HIV type 1 infection among prisoners from Central Western Brazil. *AIDS Res Hum Retroviruses.* 2011;27(12):1349-53.
- Cardoso LP, Queiroz BB, Stefani MM. HIV-1 pol phylogenetic diversity and antiretroviral resistance mutations in treatment naïve patients from Central West Brazil. *J Clin Virol.* 2009;46(2):134-9.

- Carvalho BC, Cardoso LP, Damasceno S, Stefani MM. Moderate prevalence of transmitted drug resistance and interiorization of HIV type 1 subtype C in the inland North State of Tocantins, Brazil. *AIDS Res Hum Retroviruses*. 2011;27(10):1081-7.
- Departamento de DST, aids e Hepatites Virais; Ministério da Saúde. Brasil. Boletim Epidemiológico AIDS. Ano III, nº. 1, até semana epidemiológica 52^a – Dezembro de 2013. Brasília; 2014.
- Costa ZB, Stefani MM, de Lima YA, de Souza WV, de Siqueira Filha NT, Turchi MD, et al. Estimated incidence and genotypes of HIV-1 among pregnant women in central Brazil. *PLoS One*. 2013;8(11):e79189.
- Departamento de DST, aids e hepatites virais; Ministério da Saúde. Brasil. Pesquisa de conhecimentos, atitudes e práticas da população brasileira. Brasília; 2011.
- Easterbrook PJ, Smith M, Mullen J, O'Shea S, Chrystie I, de Ruiter A, et al. Impact of HIV-1 viral subtype on disease progression and response to antiretroviral therapy. *J Int AIDS Soc*. 2010;13:4.
- Ferreira AS, Cardoso LP, Stefani MM. Moderate prevalence of transmitted drug resistance and high HIV-1 genetic diversity in patients from Mato Grosso State, Central Western Brazil. *J Med Virol*. 2011;83(8):1301-7.
- Kiwanuka N, Robb M, Laeyendecker O, Kigozi G, Wabwire-Mangen F, Makumbi FE, et al. HIV-1 viral subtype differences in the rate of CD4+ T-cell decline among HIV seroincident antiretroviral naive persons in Rakai district, Uganda. *J Acquir Immune Defic Syndr*. 2010;54(2):180-4.
- Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, et al. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog*. 2014;10(2):e1003932.
- Magee D, Beard R, Suchard MA, Lemey P, Scotch M. Combining phylogeography and spatial epidemiology to uncover predictors of H5N1 influenza A virus diffusion. *Arch Virol*. 2015;160(1):215-24.

- Ministério da Saúde do Brasil [homepage na internet]. DataSUS [Acesso em Maio de 2015]. Disponível em: <http://www2.datasus.gov.br/DATASUS/index.php?area=0203> .
- Secretaria de Atenção de Saúde; Ministério da Saúde. Brasil. Política Nacional de Atenção Básica. Brasília; 2012
- Nino AAS. A importância do transporte internacional para a cidade de Uruguaiana. Porto Alegre. Monografia [Especialização em Gestão Pública] – UFRGS; 2011.
- Nunes CC, Matte MC, Dias CF, Araújo LA, Guimarães LS, Almeida S, 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*. 2014;56(3):205-11.
- Pando MA, Gómez-Carrillo M, Vignoles M, Rubio AE, dos Ramos Farias MS, Vila M, 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*. 2011;27(1):17-23.
- Tebit DM, Arts EJ. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis*. 2011;11(1):45-56.
- Touloumi G, Pantazis N, Pillay D, Paraskevis D, Chaix ML, Bucher HC, et al. Impact of HIV-1 subtype on CD4 count at HIV seroconversion, rate of decline, and viral load set point in European seroconverter cohorts. *Clin Infect Dis*. 2013;56(6):888-97.
- WHO library cataloguing in publication data, Joint United Nations Programme on HIV/AIDS. Global report: UNAIDS report on the global AIDS epidemic; 2010.

CAPÍTULO 6.

Conclusões

- A revisão bibliográfica mostrou que o HIV-1C disseminou-se rapidamente pelos estados da região Sul do Brasil, mas sua dispersão por estados e países próximos à região Sul tem sido bem mais lenta.
- O subtipo C do HIV-1 foi observado como forma prevalente na grande maioria dos municípios amostrados.
- Nos municípios do Rio Grande do Sul foi observado maior prevalência de formas recombinantes, entre elas a CRF31_BC que possui circulação mais acentuada nos arredores de Porto Alegre, e três possíveis novas CRFs.
- No estado de Santa Catarina foi observado a associação do HIV-1C com transmissões heterossexuais e também com indivíduos autodeclarados afrodescendentes, enquanto o HIV-1B foi associado ao grupo HSH.
- Análises filodinâmicas reconstruíram taxa de expansão mais rápida para a epidemia do HIV-1C em comparação com HIV-1B. A história demográfica das epidemias nos estados de Santa Catarina e Rio Grande do Sul concorda com a associação dos subtipos do HIV com diferentes grupos de transmissão.
- Assim como demonstrado anteriormente, a epidemia do HIV-1C brasileira teve início na primeira metade da década de 1970.
- Porto Alegre foi o local estimado de introdução do HIV-1C e possui um papel central na disseminação desta forma viral pelo país.
- Prevalência de HIV e tamanho da população do HIV-1C foram identificados como preditores da dispersão da epidemia do subtipo C, ajudando na reconstrução robusta de um padrão de dispersão do Sul ao Norte do país.

CAPÍTULO 7.

Perspectivas

- Investigar o tempo de progressão à aids de pacientes infectados com diferentes subtipos do HIV a partir dos dados clínicos coletados retrospectivamente.
- Investigar a taxa de diagnóstico tardio e possíveis associações com forma viral e dados demográficos da população.
- Investigar a prevalência de mutações de resistência a antirretrovirais associando sua presença a dados virais e demográficos e investigar a transmissão de resistência primária entre os indivíduos coletados.
- Investigar a presença de redes de transmissão entre os indivíduos amostrados e descrever suas características.

Apêndices

1 - Outros trabalhos publicados:

Período de 2011/06 - 2015/06

Pinto AR, Petry A, Gräf T, Vandresen R, Kupek E. Case report of a haemovigilance investigation using phylogenetic analysis of HIV-1 in Brazil. *Transfus Med*, 2012;22(1):57–62.

Bello G, Zanutto PMD, Iamarino A, Gräf T, Pinto AR, Couto-Fernandez JC, et al. Phylogeographic Analysis of HIV-1 Subtype C Dissemination in Southern Brazil. Sandstrom P, editor. *PLoS One*, 2012;7(4):e35649.

Almeida SE, de Medeiros RM, Junqueira DM, Gräf T, Passaes CPB, Bello G, et al. Temporal dynamics of HIV-1 circulating subtypes in distinct exposure categories in southern Brazil. *Virology*, 2012;9(1):306.

Junqueira DM, Medeiros RM De, Leite TCNF, Guimarães ML, Gräf T, Pinto AR, et al. Detection of the B"-GWGR variant in the southernmost region of Brazil: unveiling the complexity of the human immunodeficiency virus-1 subtype B epidemic. *Mem Inst Oswaldo Cruz*, 2013;108(6):735–40.

Librelotto CS, Gräf T, Simon D, Esteves S, Almeida M De, Lunge VR. Major Article HIV-1 epidemiology and circulating subtypes in the countryside of South Brazil. *Rev Soc Bras Med Trop*. 2015;48(May):1–9.

2 - Termo de consentimento Livre e Esclarecido:

UNIVERSIDADE FEDERAL DE SANTA CATARINA
Departamento de Microbiologia, Imunologia e Parasitologia
Centro de Ciências Biológicas
88049-900, Campus Universitário, Florianópolis, SC, Brasil

PROJETO DE PESQUISA: Análise da dinâmica de transmissão e disseminação do HIV no Sul do Brasil: uma aplicação da epidemiologia molecular na saúde pública

RESPONSÁVEL: Prof. Dr. Aguinaldo Roberto Pinto

TERMO DE CONSENTIMENTO E INFORMAÇÃO

A AIDS é uma doença causada pelo vírus HIV e causa uma diminuição das defesas do corpo. Este vírus destrói as células responsáveis pela defesa do organismo, e assim o corpo fica sem uma proteção natural. A contaminação pelo HIV pode ocorrer através da relação sexual desprotegida, de mães grávidas para seus filhos, através da troca de seringas entre usuários de drogas e mais raramente em transfusões de sangue. No início da doença os sintomas podem ser parecidos com os de uma gripe comum. Após alguns anos, outros sintomas mais graves aparecem. Quando o exame para HIV é positivo o médico solicita constantemente outros exames, como carga viral e contagem de células CD4/CD8, para saber se a doença esta progredindo e como o vírus está reagindo aos medicamentos. O vírus HIV esta presente em várias partes do mundo e pode se apresentar de diferentes formas, que são chamadas de subtipos. Conhecer o subtipo do vírus é importante para se entender melhor como a doença se espalha na população humana e a partir disto desenvolver campanhas de prevenção mais efetivas e até mesmo novas formas de tratamento ou vacina. Além disso, alguns vírus apresentam mutações que conferem resistência contra certos medicamentos, tornando esses medicamentos ineficazes contra o vírus, então, dependendo do subtipo do vírus o médico pode utilizar um tipo diferente de tratamento para o paciente. Mas para se alcançar estes

resultados é necessário estudar e pesquisar o vírus presente no sangue de indivíduos portadores.

Este projeto pretende pesquisar o tipo de HIV presente em pacientes soropositivos do Estado de Santa Catarina, bem como saber contra quais medicamentos estes vírus são resistentes e como eles se disseminam na população do estado. Este é um estudo inicial e as informações aqui obtidas poderão ajudar estudos futuros a tentarem desenvolver tratamentos mais adequados, vacinas contra AIDS e campanhas de prevenção mais efetivas. A sua participação neste estudo é muito importante e o senhor/a senhora não terá que pagar nada. O senhor/a senhora está sendo convidado a fazer parte deste estudo juntamente com muitas outras pessoas.

PROCEDIMENTO

Se o senhor/a senhora concordar com a participação, deverá autorizar a utilização do seu sangue em alguns testes para fins desta pesquisa. Será coletado apenas uma pequena quantidade de seu sangue através de uma leve picada no dedo e nada mais. Não haverá qualquer outro procedimento que possa ocasionar desconfortos para o participante a não ser o desconforto da picada da lanceta que coletará o sangue. Pedimos também autorização para consultar seu prontuário médico. O benefício deste estudo poderá não ser imediato, porém fornecerá informações a pesquisadores sobre quais os tipos de vírus que os pacientes apresentam e este conhecimento poderá beneficiar outros pacientes futuramente e evitar que um número tão grande de novas pessoas se infecte.

PARTICIPAÇÃO E CONFIDENCIALIDADE

A participação dos pacientes neste estudo é voluntária, confidencial e nenhum nome será divulgado em qualquer tipo de publicação. Caso o senhor/a senhora não queira participar, isto não influenciará o tratamento que você tem recebido. Pense bem, faça a pergunta que desejar e esclareça suas dúvidas. Todos os itens deste termo de consentimento e informação serão cumpridos e respeitados pelos pesquisadores envolvidos.

Se você tiver alguma dúvida ou mesmo depois de assinar este documento não quiser mais participar desta pesquisa, pode entrar em

contato com qualquer um destes pesquisadores abaixo, comunicar sua decisão e você será excluído da pesquisa:

Prof. Dr. Aguinaldo R. Pinto/Universidade Federal de Santa Catarina –
Fone: 3721-5206

CONSENTIMENTO

Eu recebi uma cópia e li (ou leram para mim) as informações acima descritas. Foram explicados todos os procedimentos deste estudo. Sei que posso perguntar o que desejar e compreendo exatamente que o meu sangue será utilizado para realização de pesquisas sobre o HIV. Sei também dos possíveis desconfortos e benefícios com a participação neste estudo. Sou livre para autorizar ou não a minha participação neste estudo.

Pesquisador responsável:

Prof. Dr. Aguinaldo R. Pinto

DECLARAÇÃO

Declaro estar ciente das informações ora prestadas, tendo lido atentamente e concordado com o todo teor.

_____, _____ de _____ de 20__.

Assinatura do Paciente ou Responsável

Nome completo do paciente: _____

Nome Responsável: _____

RG: _____

CPF: _____

3 - Questionário utilizado na entrevista de recrutamento dos voluntários:

Questionário Paciente

1. Identificação do Paciente:

2. Data de Nascimento: ____ / ____ / ____
3. Sexo: () Masculino () Feminino
4. Município de Residência: _____
5. Estado Civil: () Solteiro () Casado () Acompanhado
6. Etnia (auto-declaração): () Branco () Não-branco
7. Profissão: _____
8. Em atividade: () Sim () Não
9. Escolaridade: _____

10. Data da última sorologia negativa para HIV: ____ / ____ / ____
11. Data da primeira sorologia positiva para HIV: ____ / ____ / ____
12. Possível forma de Transmissão:
() Heterossexual () HSH () UDI () Transfusão sanguínea/Transplante
() Transmissão vertical (Materno fetal) () Outro.
Qual? _____
13. Tabagista: () Sim () Não
14. Etilista: () Sim () Não
15. Uso de drogas: () Sim. Qual? _____ () Não
16. Comorbidades e Coinfeccções:
() Diabetes () Cardiopatia () Hemodiálise ()
Hepatite B () Hepatite C () Influenza A ()
() Gripe, resfriado
() HTLV () Tuberculose () Outras DSTs. Qual?

() Outros. Qual? _____
17. Viajou para o exterior: () Sim, onde: _____ () Não

18. Teve contato sexual/drogas injetáveis no exterior: () Sim () Não Onde? _____
19. Teve contato sexual/drogas injetáveis com estrangeiros no Brasil:
 () Sim. Qual a origem do contato?
 _____ () Não
20. Apresenta alguma das doenças auto-imunes abaixo:
 () Diabetes () Lupus (LES)/problemas de pele () Artrite
 () Outra. Qual? _____
 Apresenta histórico familiar de doenças auto-imunes? Se sim, qual? _____
21. Últimos exames de contagem de linfócitos CD4 e Carga Viral:
 CD4: _____ data: _____;
 CD4: _____ data: _____;
 CD4: _____ data: _____; CD4: _____
 data: _____; CD4: _____ data: _____;
 CD4: _____ data: _____;
 Carga Viral _____ data: _____; Carga
 Viral _____ data: _____;
 Carga Viral _____ data: _____; Carga
 Viral _____ data: _____;

4 - Relatório Sucinto

Documento distribuído às Unidades de Saúde participantes deste estudo

Relatório sucinto:

HIV/aids – diversidade viral
e mutações de resistência
em cidades de Santa Catarina e Rio Grande do Sul

Tiago Gräf^{1,2}, Hegger M. Fritsch², Sabrina E. M. Almeida²
& Aguinaldo R. Pinto¹

¹ Laboratório de Imunologia Aplicada, Departamento de Microbiologia, Imunologia e Parasitologia, Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil.

² Grupo de Pesquisa em HIV/aids, Centro de Desenvolvimento Científico e Técnico, Fundação Estadual de Pesquisa e Produção em Saúde, Porto Alegre, RS, Brasil.

Realização:



Apoio:



Instituições parceiras

Estado de Santa Catarina:

- Centro de Testagem e Aconselhamento (CTA) de Joinville
- Serviço de Assistência Especializado (SAE) em DST/HIV/aids de Blumenau
- Centro de Referência de Doenças Infecciosas (CEREDI) de Itajaí
- Hospital Dia de Chapecó
- CTA de Lages
- Programa de Atenção Municipal às DST/HIV/AIDS de Criciúma

Estado do Rio Grande do Sul:

- Centro Especializado de Saúde (CES) de Caxias do Sul
- SAE/CTA de Lajeado
- SAE de Cruz Alta
- SAE/CTA de Uruguaiana
- SAE de Santiago
- SAE de Santa do Livramento
- Hospital Universitário da Universidade Federal de Santa Maria

Apresentação

O cenário da epidemia de HIV/aids nos estados do Rio Grande do Sul (RS) e Santa Catarina (SC) é preocupante. Os dois estados apresentam taxas de incidência anual de aids bem maiores do que a média nacional, o que por sua vez impacta as taxas de mortalidade por essa doença, também muito elevadas no RS e em SC. Além disto, o subtipo do HIV que circula com maior frequência nesses estados é o subtipo C, enquanto no restante do Brasil é o subtipo B. Ainda não se sabe se a circulação do subtipo C nestes locais pode estar ocasionando o crescimento mais rápido da epidemia no RS e em SC. Da mesma forma também é desconhecido se pacientes infectados com o subtipo C apresentam mais resistência primária ao tratamento se comparado àqueles infectados com o subtipo B. Alguns estudos abordando essas questões já foram realizadas nas capitais Porto Alegre e Florianópolis, mas poucos investigaram cidades do interior. Essas peculiaridades da epidemia de HIV/aids em SC e no RS e a escassez de informações para as regiões interioranas motivaram a realização deste estudo.

Neste documento são apresentados sucintamente alguns resultados que podem ser de maior interesse e/ou importância aos centros de saúde e profissionais que atuam na atenção primária aos pacientes HIV-soropositivos. Sem a participação desses, a realização deste estudo não teria sido possível e esta é uma forma de dizer obrigado a todos os envolvidos nas etapas de recrutamento, coleta e entrevista dos pacientes. Este documento também ressalta a importância de parcerias entre a Universidade e o serviço público de saúde, beneficiando assim, em última instância, o cidadão brasileiro.

Introdução

O vírus da imunodeficiência humana (HIV) é um retrovírus cujo genoma é constituído por duas fitas simples de RNA (ácido ribonucleico) de elevada variabilidade, possibilitando a classificação das partículas virais em diferentes grupos e subtipos. A descoberta do HIV ocorreu no início da década de 1980, quando os primeiros casos de aids começaram a ser identificados nos Estados Unidos. Em vista do rápido crescimento dessa nova epidemia, o primeiro medicamento capaz de controlar o ciclo de replicação do HIV, o AZT, foi liberado em 1987, poucos anos após o descobrimento do vírus.

Devido a grande variabilidade do genoma do HIV, por processos de recombinação e mutação, é possível classifica-lo em três grupos: M, N e O. Os grupos N e O encontram-se distribuídos de forma mais discreta, sendo endêmicos apenas em determinadas regiões do oeste africano. O grupo M, por sua vez, apresenta maior relevância epidemiológica e clínica, visto que é o responsável pela pandemia mundial. Esse grupo é subdividido em 9 subtipos diferentes e mais de 60 formas recombinantes circulantes (CRFs). Esses subtipos encontram-se distribuídos de forma bastante heterogênea no mundo. O subtipo C é atualmente o mais prevalente, sendo encontrado em países populosos como a Índia e região sul da África, enquanto o subtipo B, o de maior dispersão mundial, é responsável pela maioria de casos no continente americano, Europa e Oceania. No Brasil, o subtipo B também é o mais frequente, com exceção da região sul, que apresenta elevada incidência de subtipo C. Com relação a efetividade da terapia ou severidade da doença, até o momento não está esclarecido se existe alguma diferença entre os subtipos. Portanto estudos são necessários para investigar essas questões, sendo muito importante manter-se vigilante sobre a distribuição das variantes do HIV.

Atualmente existem seis classes de antirretrovirais utilizados no controle da replicação do HIV. Embora essas medicações sejam utilizadas em tratamentos combinados, atuando em mais de uma etapa da replicação viral, muitas vezes o tratamento pode falhar devido a mutações de resistência para um ou mais medicamentos. As mutações de resistência são originadas devido a elevada taxa de mutação do HIV e podem ser transmitidas a indivíduos que nunca tomaram antirretrovirais (virgens de tratamento). Esses, podem por sua vez apresentar falha terapêutica no momento de início da terapia (resistência primária). Em

um país de acesso universal ao tratamento antirretroviral, o monitoramento da prevalência de mutações de resistência em pacientes virgens de tratamento é muito importante.

Objetivos do estudo

Avaliar a diversidade de subtipos do HIV e a prevalência de mutações de resistência aos antirretrovirais em municípios do interior do Rio Grande do Sul e Santa Catarina. Os municípios incluídos neste estudo foram: Rio Grande do Sul – Uruguaiana, Santana do Livramento, Santiago, Santa Maria, Cruz Alta, Lajeado e Caxias do Sul; Santa Catarina – Criciúma, Chapecó, Lages, Blumenau, Itajaí e Joinville.

Resultados

1. Características da população estudada

Tabela 1. Características gerais da população estudada nos estados do Rio Grande do Sul e Santa Catarina

Característica	Estado		Total (n=317)	
	RS (n=144)	SC (n=173)		
Sexo	Feminino	74 (51%)	69 (40%)	143 (45%)
	Masculino	70 (49%)	104 (60%)	174 (55%)
Forma de transmissão	Transusão	2 (1%)	4 (2%)	6 (2%)
	Heterossexual	114 (79%)	111 (64%)	225 (71%)
	UDI	4 (3%)	3 (2%)	7 (2%)
	HSH	16 (11%)	47 (27%)	63 (20%)
	Vertical	1 (1%)	-	1 (0%)
	ND	7 (5%)	8 (5%)	15 (5%)
Etnia	Branco	86 (60%)	142 (82%)	228 (72%)
	Afro-descendente/ Ameríndio/Outro	27 (19%)	28 (16%)	55 (17%)
	ND	31 (22%)	3 (2%)	34 (11%)
Anos de estudo	≤ 8 anos	74 (51%)	75 (43%)	149 (47%)
	9 - 11 anos	39 (27%)	60 (35%)	99 (31%)
	≥ 12 anos	7 (5%)	35 (20%)	42 (13%)
	ND	24 (17%)	3 (2%)	27 (9%)
Estado civil	Parceiro fixo	28 (19%)	39 (23%)	67 (21%)
	Casado	43 (30%)	44 (25%)	87 (27%)
	Divorciado/viúvo	8 (6%)	6 (3%)	14 (4%)
	Solteiro	59 (41%)	82 (47%)	141 (44%)
	ND	6 (4%)	2 (1%)	8 (3%)
Idade	<20	3 (2%)	2 (1%)	5 (2%)
	20-29	36 (25%)	38 (22%)	74 (23%)
	30-39	41 (28%)	50 (29%)	91 (29%)
	40-49	29 (20%)	47 (27%)	76 (24%)
	≥50	35 (24%)	36 (21%)	71 (22%)

UDI: Usuários de drogas injetáveis; HSH: Homens que fazem sexo com homens; ND: Informação não disponível; n: número de amostras.

Tabela 2. Características gerais da população estudada em cada cidade participante

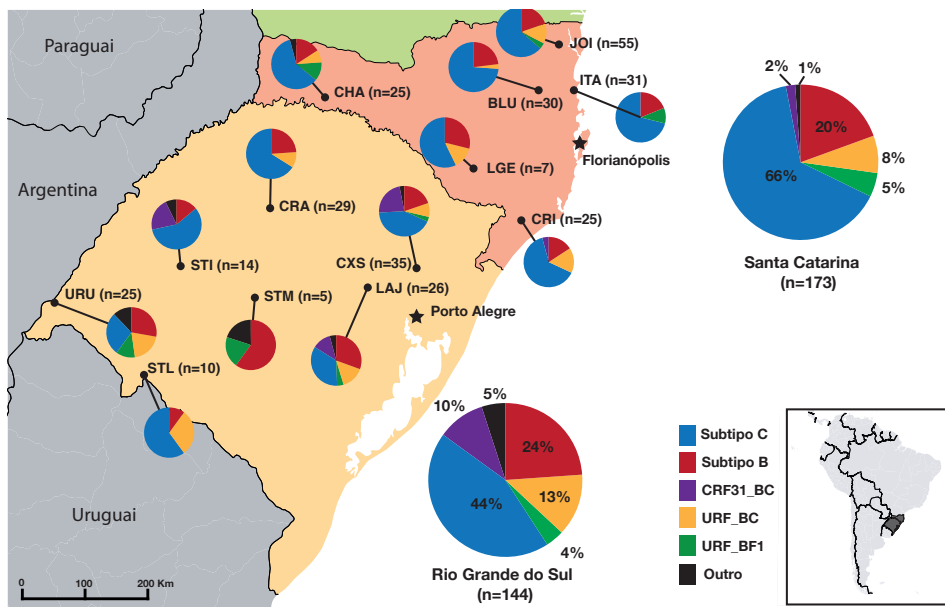
Característica		BLU (n=30)	CHA (n=30)	CRI (n=28)	CXS (n=41)	ITA (n=35)	JOI (n=60)	LAJ (n=29)	STL (n=12)	STI (n=15)	URU	
Sexo	Feminino	9 (30%)	15 (50%)	14 (50%)	20 (49%)	17 (49%)	21 (35%)	13 (45%)	8 (67%)	7 (47%)	13	
	Masculino	21 (70%)	15 (50%)	14 (50%)	21 (51%)	18 (51%)	39 (65%)	16 (55%)	4 (33%)	8 (53%)	16	
Forma de transmissão	Transfusão	-	-	1 (3,6%)	3 (8%)	2 (6%)	1 (2%)	-	-	1 (7%)		
	Heterossexual	14 (47%)	26 (87%)	21 (75%)	30 (73%)	20 (57%)	38 (63%)	25 (86%)	10 (83%)	9 (60%)	20	
	UDI	-	-	1 (3,6%)	1 (2,3%)	1 (3%)	1 (2%)	1 (4%)	-	-	2	
	HSH	16 (53%)	3 (10%)	4 (14%)	5 (12%)	11 (31%)	15 (25%)	3 (10%)	-	3 (20%)	6	
	Vertical	-	-	-	1 (2,3%)	-	-	-	-	-	-	
	ND	-	1 (3%)	1 (3,6%)	1 (2,3%)	1 (3%)	5 (8%)	-	2 (17%)	2 (13%)	1	
Etnia	Branco	23 (77%)	26 (87%)	26 (93%)	25 (61%)	28 (80%)	51 (85%)	22 (76%)	6 (50%)	11 (73%)	2	
	Afro-descendente/ Ameríndio/Outro	7 (23%)	3 (10%)	2 (7%)	11 (27%)	7 (10%)	8 (13%)	7 (24%)	4 (33%)	3 (20%)		
	ND	-	1 (3%)	-	5 (12%)	-	1 (2%)	-	2 (17%)	1 (7%)	27	
Anos de estudo	≤ 8 anos	14 (47%)	12 (40%)	17 (61%)	26 (63%)	14 (40%)	22 (37%)	16 (55%)	12 (100%)	6 (40%)	1	
	9 - 11 anos	6 (20%)	5 (17%)	8 (29%)	10 (24%)	19 (54%)	27 (45%)	10 (34%)	-	7 (47%)	2	
	≥ 12 anos	10 (33%)	11 (37%)	3 (11%)	3 (7%)	2 (6%)	10 (17%)	3 (10%)	-	2 (13%)		
	ND	-	2 (7%)	-	2 (5%)	-	1 (2%)	-	-	-	26	
Estado civil	Parceiro fixo	8 (27%)	4 (13%)	8 (29%)	1 (2%)	8 (23%)	16 (27%)	13 (45%)	1 (8%)	3 (20%)	10	
	Casado	5 (17%)	14 (47%)	5 (18%)	13 (32%)	8 (23%)	15 (25%)	5 (17%)	1 (8%)	5 (33%)	7	
	Divorciado/viuvo	2 (7%)	-	3 (11%)	1 (2%)	-	2 (3%)	2 (7%)	-	-	1	
	Solteiro	15 (50%)	12 (40%)	11 (39%)	25 (61%)	19 (54%)	26 (43%)	9 (31%)	8 (67%)	7 (47%)	6	
	ND	-	-	1 (4%)	1 (2%)	-	1 (2%)	-	2 (17%)	-	5	
Idade	<20	1 (3%)	-	1 (4%)	1 (2%)	1 (3%)	-	1 (3%)	1 (8%)	-		
	20-29	9 (30%)	7 (23%)	4 (14%)	9 (22%)	7 (20%)	13 (22%)	7 (24%)	1 (8%)	4 (27%)	8	
	30-39	4 (13%)	12 (40%)	9 (32%)	13 (32%)	12 (34%)	15 (25%)	5 (17%)	6 (50%)	4 (27%)	5	
	40-49	5 (17%)	6 (20%)	11 (39%)	8 (20%)	10 (29%)	18 (30%)	8 (28%)	-	3 (20%)	7	
	≥50	11 (37%)	5 (17%)	3 (11%)	10 (24%)	5 (14%)	14 (23%)	8 (28%)	4 (33%)	4 (27%)	9	

UDI: Usuários de drogas injetáveis; HSH: Homens que fazem sexo com homens; ND: Informação não disponível; n: número de amostras por cidade. Siglas das cidades: BLU: Blumenau; CHA: Chapecó; CRA: Cruz Alta CRI: Criciúma; CXS: Caxias do Sul; ITA: Itajaí; JOI: Joinville; LAJ: Lajeado; LGE: Lages; STI: Santiago; STL: Santana do Livramento; STM: Santa Maria; URU: Uruguaiana.

As tabelas 1 e 2 demonstram as características demográficas coletadas através de entrevista no momento de recrutamento dos voluntários. Algumas características possuem diferença estatística entre RS e SC. A primeira delas é o número de homens e mulheres participando do estudo: no RS há mais mulheres amostradas que em SC. De uma forma geral, ambos estados apresentaram uma proporção de mulheres infectadas por HIV maior que a média nacional, ressaltando a importância desta parcela da população na epidemia de aids da região. A menor representatividade de homens na amostra do RS pode ser explicada em parte pelo pequeno número de indivíduos HSH participando do estudo neste estado. Neste estudo o número de indivíduos HSH infectados pelo HIV no RS é menor que a média do Brasil, enquanto que em SC a taxa é igual a média nacional. Entretanto não pode-se concluir se estas diferenças são reflexo de algum erro de amostragem, ou se elas realmente refletem a realidade das populações infectadas nestes dois estados.

Outras características não possuem diferença estatística, seja entre RS e SC, bem como entre estes estados e a média nacional. A única exceção é para a autodeclaração de etnia, que apresentou maiores índices de indivíduos brancos no RS e SC se comparado ao Brasil como um todo. Entretanto esta diferença apenas reflete a composição étnica naturalmente encontrada nestes estados e não significa que mais indivíduos brancos estejam sendo infectados pelo HIV em comparação a indivíduos afrodescendentes ou indígenas.

2. Diversidade de variantes do HIV encontradas na região estudada



Siglas das cidades: BLU: Blumenau; CHA: Chapecó; CRA: Cruz Alta; CRI: Criciúma; CXS: Caxias do Sul; ITA: Itajaí; JOI: Joinville; LAJ: Lajeado; LGE: Lages; STI: Santiago; STL: Santana do Livramento; STM: Santa Maria; URU: Uruguaiana.
n: número de amostras por cidade.

De uma forma geral foi encontrada grande prevalência de subtipo C em todas as cidades estudadas. Esta é uma característica da epidemia do HIV no Sul do Brasil, sendo que em todas as outras regiões brasileiras o subtipo B é o mais prevalente e o subtipo C é encontrado em frequências menores de 10%. No estado do Paraná é observada uma prevalência intermediária de subtipo C, em torno de 20%. É possível observar que em média as cidades catarinenses possuem uma maior prevalência de subtipo C, enquanto no Rio Grande do Sul há mais formas recombinantes (CRF31_BC e URF_BC).

A forma viral CRF31_BC é uma cepa recombinante entre o subtipo B e subtipo C que surgiu no Brasil e é muito prevalente em Porto Alegre. Neste estudo ela foi encontrada praticamente só no Rio

Grande do Sul e em cidades próximas a capital deste estado, sugerindo que sua origem provavelmente se deu na região metropolitana de Porto Alegre.

3. Resistência a antirretrovirais

Tabela 3. Prevalência (%) de mutações de resistência a antirretrovirais (MRA) nos estados do Rio Grande do Sul e Santa Catarina

Categoria de resistência	Número de amostras analisadas	Número contendo MRA	%
Amostras com qualquer MRA	248	23	9.3
Amostras com IP MRA	246	9	3.7
Amostras com INTR MRA	247	6	2.4
Amostras com INNTR MRA	247	13	5.3
Amostras com INTR+INNTR MRA	247	2	0.8
Amostras com IP+INTR+INNTR MRA	245	2	0.8

IP: inibidores da protease; INTR: inibidores nucleosídicos da transcriptase reversa; INNTR: inibidores não nucleosídicos da transcriptase reversa

A prevalência de aproximadamente 9% está de acordo com estudos anteriores realizados no Brasil, demonstrando que os índices de resistência no interior do Rio Grande do Sul e Santa Catarina é similar ao de outras regiões do Brasil. Não foram encontradas diferenças estatísticas significantes na prevalência de MRA em indivíduos virgens de tratamento infectados com subtipo C ou subtipo B.

Conclusões

Os resultados apresentados aqui demonstram uma proporção grande de mulheres infectadas pelo HIV. Esta característica da epidemia é diferente do resto do país e merece uma atenção especial das políticas públicas de prevenção à aids. Talvez parte das grandes taxas de incidência de aids no RS e SC possa ser decorrente do aumento do número de infecções entre mulheres. Outro resultado que deve ser ressaltado é a alta prevalência do subtipo C do HIV nas cidades do interior de SC e RS, o que corrobora estudos anteriores que também encontraram alta prevalência desta forma do HIV em Florianópolis e Porto Alegre. Por fim, apesar de apresentar uma forma viral distinta da do resto do país, as taxas de mutações de resistência primária em SC e no RS são similares a estudos anteriores realizados em diferentes locais do Brasil. Esse dado, a princípio, sugere que ao introduzir-se a terapia não há necessidade de adaptação da combinação de antirretrovirais ao subtipo do HIV que o paciente possui.