UNIVERSIDADE FEDERAL DE SANTA CATARINA
Centro de Ciências Biológicas
Departamento de Microbiologia, Imunologia e Parasitologia
Laboratório de Imunologia Aplicada

IDENTIFICAÇÃO E CARACTERIZAÇÃO FILOGENÉTICA DE UMA NOVA FORMA RECOMBINANTE DE HIV-1 PRESENTE NO SUL DO BRASIL

Vinicius Provenzi Coltro

Florianópolis
2016

UNIVERSIDADE FEDERAL DE SANTA CATARINA
Centro de Ciências Biológicas
Departamento de Microbiologia, Imunologia e Parasitologia
Laboratório de Imunologia Aplicada

IDENTIFICAÇÃO E CARACTERIZAÇÃO FILOGENÉTICA DE UMA NOVA FORMA RECOMBINANTE DE HIV-1 PRESENTE NO SUL DO BRASIL

Vinicius Provenzi Coltro

Trabalho de conclusão de curso apresentado ao Curso de Ciências Biológicas, Centro de Ciências Biológicas da Universidade Federal de Santa Catarina, como requisito parcial para obtenção do título de Bacharel em Ciências Biológicas.

Orientador: Profº. Drº. Aguinaldo R. Pinto

Co-orientador: Drº. Tiago Gräf.

Florianópolis
2016

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

Coltro, Vinicius
IDENTIFICAÇÃO E CARACTERIZAÇÃO FILOGENÉTICA DE UMA NOVA FORMA RECOMBINANTE DE HIV-1 PRESENTE NO SUL DO BRASIL
/ Vinicius Coltro ; orientador, Aguinaldo Roberto Pinto ; co-orientador, Tiago Gräf. - Florianópolis, SC, 2016.
43 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro de Ciências Biológicas. Graduação em Ciências Biológicas.

Inclui referências

1. Ciências Biológicas. 2. HIV-1. 3. Recombinação. 4. Análises Filogenéticas. I. Roberto Pinto, Aguinaldo. II. Gräf, Tiago. III. Universidade Federal de Santa Catarina. Graduação em Ciências Biológicas. IV. Título.

VINICIUS PROVENZI COLTRO

IDENTIFICAÇÃO E CARACTERIZAÇÃO FILOGENÉTICA DE UMA NOVA FORMA RECOMBINANTE DE HIV-1 PRESENTE NO SUL DO BRASIL

Florianópolis, 05 de julho de 2016

_____

Profª. Drª. Maria Risoleta Freire Marques
Coordenadora de Graduação do Curso de Ciências Biológicas

Banca Examinadora:

_____

Profº. Drº. Aguinaldo Roberto Pinto
Presidente

_____

Profª. Drª. Andrea Rita Marrero
Membro Titular

_____

Drº. Guilherme Toledo
Membro Titular

_____

Profº. Drº. Glauber Wagner
Membro Suplente

"*A ciência não é a atividade séria que muitos pensam. Impregnados de preconceitos de toda ordem, nós, os cientistas, atrasamos, inúmeras vezes, o progresso da própria ciência simplesmente porque, em geral, não dispomos do brilho intelectual, da imparcialidade, da visão crítica, da liberdade de pensamento, da capacidade de bem julgar etc., características que vivemos alardeando como nossas principais qualidades. Não somos melhores que os comerciantes, os operários, os sacerdotes, os poetas. Damos muita importância à posição exercida pelo cientista que faz declarações, à língua em que o trabalho é redigido, à nacionalidade de quem criou a hipótese, à revista em que o trabalho é publicado, ao fato de estar a nova verificação de acordo com verificações antigas etc. A ciência não vive de verdades, o cientista não é um construtor de verdades. No reino das grandes teorias e mesmo, às vezes, das pequenas hipóteses, a ciência se alimenta, muitas vezes, de crenças baseadas em dados insuficientes.*"

Newton Freire-Maia

AGRADECIMENTOS

Agradeço primeiramente (os demais não são menos importantes) aos meus pais, Marivete Provenzi e Paulo Nadir Coltro, por uma excelente criação, desprovida de decoro (porém não de tentativas), –fato não obstante observado– mas rica em amor. Por ~~tentar insistentemente~~ me ensinar o valor do trabalho e da dedicação, por me incentivarem sempre a buscar meus objetivos, a dar sempre o meu melhor em tudo que faço, honrando meus compromissos. Por mostrarem que posso manter os pés no chão seguindo cético e vigiante, sem perder o encanto e o amor pelo que faço e pelos que caminham ao meu lado. Por respeitarem e apoiarem as minhas decisões sendo, para sempre, os maiores exemplos da minha vida.

OBS: Mãe, teu nome é tão estranho que nem o auto corretor do Word identifica essa naba.

Agradeço a Maria Terezinha Alves pelo fornecimento de ~~deliciosos~~ macronutrientes (s*trogonoff*, baião de dois, feijoada, carne assada, saladas das mais diversas, tainha, entre outros), apoio psicológico (por aguentar meus desabafos sobre o TCC, graduação e vida em geral), material didático e espaço físico adequado (por perderem um quarto para eu poder transforma-lo em um escritório) e transporte privado (caronas em grande estilo) tudo isso para o bom desenvolvimento deste trabalho. Você é como uma segunda mãe.

A Beatriz Alves Bentes de Sá por todo seu companheirismo e apoio durante está fase da minha vida, obrigado por compreender ~~as vezes com certa resistência~~ minhas ausências em corpo presente durante praticamente um ano todo de trabalho. Por todo amor e carinho de que precisei para melhor suportar as longas jornadas de estudo. Além do fornecimento de material adequado para o acondicionamento e transporte do aparelhamento tecnológico empregado neste estudo (uma mochila da hora com bolso estofado pra *notebook*).

Ao meu co-orientador, Tiago Gräf, por toda a paciência, ensinamentos, paciência, incentivo, paciência, tutela com o trabalho e os resultados, paciência, orientações a longa distância (o Skype trabalho nesse projeto) e mais paciência. Desculpe-me por toda a mediocridade que assola o meu intelecto (a tal da burrice).

Ao meu orientador, Aguinaldo Roberto Pinto, por esses 3 anos (+-0,5) de iniciações cientificas ~~não tão bem sucedidas~~ e TCC (agora vai!). Por todos os ensinamentos, correções de relatórios/banners/apresentações, conselhos e pela boa convivência e amizade no tempo em que estive no LIA. Obrigado por ter insistido e me

apoiado, e não ter me expulsado (o que eu sempre pensei que aconteceria em algum momento) em todo o tempo em que estive no LIA. Você é um cara realmente corajoso e insistente!

Agradeço a minha turma (em especial ao vadio's team – tanto aos remanescentes como aos que ficaram pelo caminho) 11.2, por toda a falta de comprometimento de todos com as atividades/festas/rodas de mobilidade urbana/oficinas de macramê/espiculas/reuniões no CA oferecidas pelo pessoal da biologia, graças a isso somos os únicos tetracampeões do troféu de turma lipídio da história deste curso. Sério, eu tenho muito orgulho disso.

A todas as pessoas que dividiram apê comigo nestes anos de faculdade (PV, Ezequiel, Verees, Boca, Alisson, Bia). Obrigado pelos bons momentos de conversa, jantas e terapias psicossomáticas a base de álcool etílico. Agradeço especialmente ao: Boca (vulgo Gustavo Rocha) pelo fornecimento de equipamentos de áudio e vídeo (uma tv de LED HD de 27'); Sr. Nariz (vulgo Ezequiel Alievi) pelo fornecimento de trajes adequados (uma camisa massa bagarai); Vereés (vulgo Felipe Ramon Provenzi Coltro) por ter importado (͡° ͜ʖ ͡°) um excelente *notebook;* todos indispensáveis para o bom desenvolvimento e apresentação deste trabalho.

A todos os outros amigos que fiz aqui em Floripa durante a graduação (Isa, Pedro, Lucas, Mara, Leos, Guis, Jhonny, Larissa, Anas, Rafa, Joãos, Arielle, Renan, Luiz, Otavio – se eu esqueci alguém peço perdão pelo vacilo). Obrigado pelas festas, bares, discussões filosóficas ~~sob efeito do alcool~~, karaokês, praias, carnavais, viradas de ano, bebedeiras e por muitas risadas.

Aos LIAnos sempre solícitos e prestativos, ajudando-me nos momentos de dificuldade (empréstimo de jaleco pra aulas práticas e caneca pro cafezinho da tarde). Obrigado por todo o companheirismo nesses 3 (+- 0,5) anos de lab.

Aos meus amigos de Chapecó (Ade, Salame, Cadu, Ezequiel, Luana, Babi, Ronald, Japa, Duda, Jack, Johan, Zé, Gobbi, Taina, Tavão, Leo, Adrew – se eu esqueci alguém peço perdão pelo vacilo) pelos ótimos momentos proporcionadas durante os meus períodos de férias. Obrigado por fazer parecer que nunca estive longe de vocês mesmo depois de tanto tempo.

Por último, mas não menos importante, aos membros da banca (Andrea, Guilherme e Glauber) por disponibilizarem seu tempo em prol da avaliação deste trabalho.

RESUMO

A aids, causada pelo HIV-1, é uma doença caracterizada pela falência do sistema imune e consequente predisposição do organismo a doenças infecciosas. O HIV-1 possui um genoma viral formado por duas cópias lineares de RNA simples fita associado a enzimas necessárias à sua replicação. Entre estas enzimas está a transcriptase reversa (RT), responsável pela síntese do cDNA a partir do RNA viral e que, por não possuir um sistema de reparo, pode gerar 0,2 mutações por genoma por ciclo de replicação. Além disto, a RT possui a capacidade de mudar de substrato durante a transcrição, gerando assim um genoma recombinante entre as duas fitas, que atrelada a esta alta taxa replicativa contribui para o surgimento de uma ampla gama de formas recombinantes circulantes (CRFs). No sul do Brasil a epidemia de HIV/aids apresenta um caráter distinto do panorama nacional, havendo um predomínio dos subtipos C e B (56% e 22% respectivamente) que co-circulam nesta região há aproximadamente 40 anos. Até o momento somente uma CRF havia sido descrita no sul do país (CRF31_BC), refletindo a falta de estudos e consequentemente mascarando o cenário real das formas virais circulantes nesta região do Brasil. No presente estudo foram caracterizadas oito sequências de 1125 pb correspondente ao fragmento que vai da posição 2265 a 3290 do genoma do HIV-1, com um padrão inédito de recombinação entre os subtipos B e C. O primeiro fragmento, que inicia na posição 2265 e termina na posição 2688, corresponde ao subtipo C. O segundo fragmento corresponde ao subtipo B e seu início encontra-se entre as posições 2689 e 2715 (2702, +-13), terminando entre os pontos 3075 e 3087 (3081, +-6). O terceiro e último fragmento corresponde ao subtipo C, iniciando na posição 3088 e terminando na posição 3290. As análises filogenéticas aqui realizadas indicam que este evento de recombinação se deu no Brasil em meados de 1982 (95% de probabilidade: 1979 – 1987). Apesar dos fortes indícios da ocorrência do evento de recombinação aqui apresentados, neste trabalho foi avaliado apenas um pouco mais de 10% de toda a extensão gênica do HIV-1. Assim, as demais regiões gênicas não avaliadas por este estudo podem conter outros pontos de recombinação. Isso torna necessário a realização de mais estudos envolvendo diferentes regiões gênicas do HIV-1 com este padrão de recombinação, incluindo uma caracterização de genomas completos, para a determinação desta como uma nova CRF entre os subtipos B e C do HIV-1.

Palavras-chave: HIV; subtipos; recombinação; análises filogenéticas.

ABSTRACT

AIDS, caused by HIV-1, is a disease characterized by the failure of the immune system, and consequent predisposition of the organism to infectious diseases. HIV-1 genome is composed by two linear copies of single stranded RNA, combined with enzymes required for replication. One of these enzymes, reverse transcriptase (RT) is responsible for the synthesis of cDNA from the viral RNA. As RT does not have a repair system, 0.2 mutations per genome are generated by replication cycle. Moreover, RT has the ability to change substrate during transcription, being able to generate a recombinant genome among two or more strains. This feature, along with high replicative rate, contributes for the appearance of a wide variety of circulating recombinant forms (CRFs). In southern Brazil epidemic of HIV/aids has a distinct character of national scenery, with the prevalence of subtype C and B (56% and 22% respectively) that co-circulate on this region approximately for 40 years. To date, only one CRF had been described in southern Brazil (CRF31_BC), reflecting the absence of studies and therefore masking the real scenery of the circulating viral forms at the region of Brazil. On the present study eight sequences, of 1125 pb were characterized, spanning from position 2265 to 3290 of the HIV-1 genome, with an unprecedented recombination between subtypes B and C. First fragment corresponds to subtype C and starts at position 2265 and ends at position 2688. The second fragment corresponds to subtype B and begins at 2689 and 2715 (2702 + -13), ending between 3075 and 3087 (3081 + -6). Third and last fragment corresponds to subtype C, starting at position 3088 and ending at position 3290. Additionally, phylogenetic analysis presented here show that this recombination event occurred in Brazil in 1982 (95% probability: 1979-1987). Although the strong evidences of the occurrence of recombination presented in this study, only slightly more than 10% of the whole genome of HIV-1 was evaluated here. Thus, other genic regions, not evaluated here, can have other recombination points. This makes necessary to perform further studies involving different genetic regions of HIV-1 with this standard recombination, including a characterization of complete genomes, for the determination such a new CRF between subtypes B and C HIV-1.

Key word: HIV; subtypes; recombination; phylogenetic analysis.

SUMÁRIO

1 – INTRODUÇÃO

1.1 – HIV, ciclo viral e patogênese

A síndrome da imunodeficiência adquirida (aids), causada pelo vírus da imunodeficiência humana (HIV-1), é uma doença caracterizada pela falência do sistema imune e consequente predisposição do organismo a doenças infecciosas. Em 1981 foram registrados os primeiros casos de aids em indivíduos homossexuais que apresentavam doenças raras como Sarcoma de Kaposi e pneumonia pneumocística [1]. O HIV-1 pode ser transmitido de diferentes maneiras, sendo que atualmente o contato sexual desprotegido é a maneira mais comum de contágio, bem como através do compartilhamento de agulhas por usuários de drogas injetáveis (UDI), transmissão vertical durante a gestação ou através da amamentação e pela transfusão de hemoderivados contaminados com o vírus [1].

Uma partícula viral do HIV-1 possui um diâmetro aproximado de 100nm de diâmetro e apresenta um envelope lipoproteico proveniente da célula hospedeira (Figura 1) [1, 2]. Estão ancoradas neste envelope as proteínas virais gp120 e gp41, além de proteínas do hospedeiro, como adesinas e moléculas de MHC que, durante o brotamento viral, são carregadas junto com a membrana da célula hospedeira e podem facilitar na adesão a outras células alvo. Ancorado do lado interno do envelope viral está a proteína p17, formadora da matriz viral, estrutura que cerca o capsídeo [1, 2]. O capsídeo é composto pela proteína p24 e no seu interior encontram-se as enzimas necessárias à replicação viral: transcriptase reversa (RT), integrase (INT) e protease (PR), assim como o genoma viral formado por duas cópias lineares de RNA (Figura 1) [1, 2].

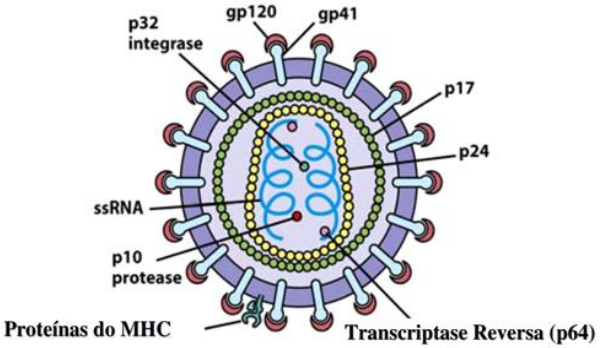

Figura 1 - Representação esquemática do HIV-1, evidenciando sua organização estrutural. Fonte: Retirado de Kuby e colaboradores (2008) [2].

Cada uma das duas fitas de RNA genômico contem aproximadamente 10 mil nucleotídeos (Figura 2). As proteínas do capsídeo, as enzimas virais e as proteínas presentes no envelope são, respectivamente, codificadas pelos três principais genes virais *gag*, *pol* e *env*, presentes em todos os retrovírus e compondo a maior parte da estrutura gênica do HIV [1, 2]. Existem ainda outros seis genes denominados *vif*, *vpu*, *vpr*, *tat*, *rev* e *nef*. Estes genes, ditos acessórios, apresentam funções regulatórias e são responsáveis por permitir toda a cascata de eventos necessária para o processo de replicação viral. Os nove genes do HIV são ladeados por longas regiões terminais repetitivos (LTR). Tais regiões não codificam proteínas, mas são importantes na regulação da expressão gênica e integram o genoma viral ao DNA da célula hospedeira durante o processo de incorporação (Figura 2) [1].

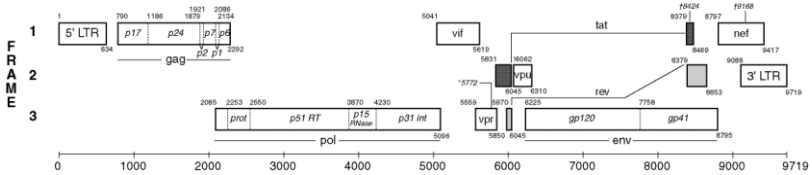

Figura 2 – Mapa genômica do HIV-1 referente à cepa HXB2 (K03455). Os quadros de leitura estão representados pelos retângulos. Os números no canto superior esquerdo dos retângulos e seus pontilhados indicam a posição de início dos respectivos genes (códon ATG). Os números nos cantos inferiores direitos indicam a posição do códon de parada. Os retângulos sombreados indicam os exons dos genes *tat* e *ver*. Fonte: Retirado de *Los Alamos HIV database* (www.hiv.lanl.gov).

O início da interação entre o HIV e as células do hospedeiro ocorre pela interação entre a molécula CD4 da membrana celular e a gp120 viral. O CD4 encontra-se ancorado na superfície de células como linfócitos T, monócitos, macrófagos, células dendríticas e da micróglia [3]. A proteína gp120 do envelope viral é capaz de ligar-se à molécula CD4 e iniciar o processo de infecção. Depois desta ligação, o sítio da gp120 interage com um co-receptor de quimiocinas, CCR5 ou CXCR4, sendo que o tropismo por um destes co-receptores dependerá da célula alvo, da cepa viral infectante e do estágio da doença [3]. A ligação da gp120 com a molécula de CD4 e em seguida com o co-receptor de quimiocinas conduz a mudanças conformacionais da gp120 e como resultado à dissociação da gp41. Finalmente, sequências peptídicas

hidrofóbicas da gp41 promovem a fusão entre a membrana da célula e o envelope viral [3].

Uma vez livre na célula, o capsídeo viral libera o material genético do vírus e as enzimas que ele carrega consigo. Em seguida a RT produz uma fita de DNA complementar (cDNA) por meio do RNA viral utilizando nucleotídeos presentes na célula hospedeira. O cDNA viral então é conduzido ao núcleo onde a enzima viral integrase liga-o covalentemente ao genoma da célula hospedeira, tornando-se assim um provírus. Sob a forma de um componente integral do genoma da célula hospedeira, o DNA proviral é transcrito utilizando a maquinaria celular [1, 3].

Geralmente as primeiras células infectadas pelo HIV são células dendríticas. Estas células encontram-se em diversos tecidos humanos, entre eles a mucosa anal e o trato genital, principais vias de contagio do HIV [1, 3]. Após serem infectadas, as células dendríticas dirigem-se aos linfonodos, onde ocorre a infecção de linfócitos T CD4, principal alvo do HIV. Os tecidos linfóides compõem o ambiente ideal para o vírus invadir novas células e se replicar rapidamente [1, 3]. Nos tecidos linfoides existe uma enorme concentração de células T CD4 e abundante produção de citocinas pró-inflamatórias, levando à ativação de linfócitos infectados e consequentemente a replicação do vírus [1, 2]. A presença do provírus no genoma de células T em linfonodos é até 10 vezes mais comum do que em células mononucleares sanguíneas, fazendo destes tecidos os principais depósitos virais e o sitio de maior depleção de linfócitos T CD4 [1, 2, 3].

A doença ocasionada pelo HIV pode ser dividida em três fases: a fase aguda, que ocorre logo depois da infecção, a fase de latência (ou crônica) e a fase de aids [1]. No decorrer da fase aguda da infecção, podem aparecer alguns sintomas como febre e linfoadenopatia em decorrência dos elevados níveis de replicação viral [1]. Na fase aguda da infecção não há resposta imune adaptativa e a carga viral plasmática comumente é superior a 10^7 partículas/mm^3 de sangue [1, 4]. Passadas algumas semanas, a resposta imune mediada por linfócitos T citotóxicos HIV-específicos reduz a viremia e a carga viral estabiliza-se em níveis inferiores [1]. Devido à produção de anticorpos também acontece a soroconversão. A resposta imune é capaz de controlar momentaneamente a replicação viral e o número de células T CD4 circulantes aumenta consideravelmente, porém nunca voltam aos valores prévios anteriores a infecção [1].

Tanto fatores genéticos do hospedeiro quanto a cepa do HIV podem intervir na dinâmica de replicação viral na infecção aguda, fazendo com que o nível de partículas virais circulantes seja bastante variável [1, 4]. Tal acontecimento é imprescindível para se delimitar a velocidade de evolução da infecção e o quão rápido os sinais clínicos da doença aparecerão. Por definição, a infecção aguda estende-se do primeiro episódio de exposição ao HIV até o desenvolvimento da resposta imune humoral [1]. Em razão da falta de anticorpos contra HIV e da incapacidade de diagnóstico da infecção por meio de testes sorológicos esta etapa também pode ser nomeada de janela imunológica [1].

Ao fim da fase aguda da infecção um período de estabilidade entre replicação viral e resposta imune do hospedeiro é alcançado, dando início a fase de latência. A fase de latência pode perdurar por vários anos sem haver sinais clínicas da doença. Porém, mesmo aparentemente saudáveis, as pessoas infectadas têm uma enorme quantidade de vírus, o que leva continuamente à perda de células T CD4 ocasionada tanto pela lise devido à replicação e brotamento de partículas virais quanto pela eliminação via apoptose e por linfócitos T CD8 das células infectadas [1, 2]. Por estas razões, durante a fase de latência ocorre a maior parte das transmissões de HIV-1 através de contato sexual ou por compartilhamento de seringas [1].

A fase de aids é marcada pelo término da fase de latência e por uma imunossupressão bastante acentuada, tornando o sistema imunológico incapaz de proteger o indivíduo soropositivo de infecções por patógenos oportunistas [1, 2]. Quando o indivíduo soropositivo apresentar alguma infecção por um patógeno oportunista característico ou a contagem de linfócitos T CD4 for abaixo de 200 células/mm3 de sangue ele é denominado como portador de aids [1].

1.2 – Epidemiologia do HIV

Atualmente estima-se que aproximadamente 36,7 milhões de pessoas vivam com HIV em todo mundo, caracterizando-a como uma das maiores pandemias da atualidade [5]. Há uma enorme disparidade na distribuição geográfica das pessoas vivendo com HIV/aids. Na África subsaariana encontram-se aproximadamente 69% de todos os casos de infectados pelo vírus no mundo, sendo que a prevalência de indivíduos infectados em determinados países africanos pode chegar a mais de 25%. Além do mais, 90% das crianças soropositivas de todo o mundo vivem na África subsaariana [5]. Outros continentes também apresentam números

preocupantes, apesar de não terem a magnitude da epidemia na África. Na Ásia, há aproximadamente 5,1 milhões de pessoas infectadas, 2 milhões na América do Norte, 2 milhões na América Latina e 1,5 milhões no Leste Europeu [5]. Na América Latina surgem cerca de 100 mil novos casos e mais de 50 mil mortes causadas pela aids a cada ano. A região latino-americana lidera as atividades de prevenção do HIV-1 pelo sistema de saúde pública, tendo 55% dos indivíduos soropositivos em tratamento antirretroviral, sendo uma das três regiões com maior cobertura de tratamento da doença em todo o mundo [5].

No Brasil estima-se que aproximadamente 798 mil indivíduos vivam com o HIV/aids, sendo que em 2015 foram notificados 40 mil novos casos e 12.449 óbitos em decorrência da aids [6]. A taxa de detecção de HIV no Brasil tem apresentado estabilização nos últimos dez anos, com uma média de 20,5 casos para cada 100 mil habitantes. Apesar disto, em 2014 registrou-se 47% do total de novos casos contabilizados na América Latina. Desta forma, o Brasil é a nação com maior número de casos de infecção pelo HIV de toda a América Latina [6].

Nas regiões sul e sudeste concentram-se mais de 73% de todos os casos brasileiros de HIV/aids, com 53,8% dos indivíduos soropositivos vivendo no sudeste e outros 20% residindo na região sul [6]. Porém, com cerca de 31,1 novos casos a cada 100 mil habitantes, a região sul apresentou a maior taxa de detecção observada entre as demais regiões do pais em 2014, seguida pela região norte (25,7/100.000 hab.), região sudeste (18,6/100.000 hab.), região centro-oeste (18,4/100.000 hab.) e região nordeste (15,2/100.000 hab.) [6]. Dos três estados que compõem a região sul do Brasil, dois deles mostraram taxas de detecção para o ano de 2014 maiores que a média nacional (20,5/100.000 hab.): Rio Grande do Sul (38,3/100.000 hab.) e Santa Catarina (29,6/100.000 hab.) [6].

O estado de Santa Catarina (SC) possui aproximadamente 6 milhões de habitantes e desde 1984, ano do primeiro caso de HIV/aids registrado no estado, 40.307 casos da doença foram notificados e mais de 10 mil óbitos em decorrência da aids [6]. Dentre os municípios de SC, os que apresentam maiores números de casos acumulados da doença são Florianópolis com 6,646 casos (prevalência de 1,14%), Joinville com 4.170 casos (prevalência de 0,6%), Itajaí com 3,756 casos (prevalência de 1,3%), São José com 2,638 casos (prevalência de 0,9%) e Blumenau com 2,393 casos (prevalência de 0,5%). Em 2014, Florianópolis apresentou 53,2 novos casos a cada 100 mil habitantes, sendo esta a quinta maior incidência de casos de aids dentre as capitais brasileiras [7].

1.3 – Filogenética e epidemiologia molecular do HIV-1

Por meio de estudos sorológicos, foram encontrados dois tipos antigênicos do HIV: o HIV-1 e o HIV-2. O HIV-1 é o tipo mais virulento e é amplamente disseminado em todo o mundo. Já o HIV-2, que apresenta distribuição quase que exclusiva no oeste africano, é caracterizado por apresentar uma progressão da doença mais lenta, assim como menor transmissibilidade quando comparado ao HIV-1 [8].

Segundo análises filogenéticas de amostras do HIV-1 de diferentes regiões geográficas, este vírus pode ser classificado em grupos, subtipos, sub-subtipos, bem como em formas recombinantes circulantes. Dentre os grupos filogeneticamente distintos, quatro já foram descritos: M (*major*), O (*outlier*), N (*non M-O*) e P (Seguindo a ordem alfabética) (Figura 3) [8, 9, 10, 11]. O grupo M abarca a grande maioria das linhagens do HIV-1, sendo também o responsável por mais de 90% das infecções por HIV em todo o mundo [8, 9]. O grupo O é composto por linhagens amplamente divergentes, endêmicas em Camarões e em alguns países vizinhos da África Central Ocidental, além de casos isolados na Europa e nos Estados Unidos [11]. O grupo N é representado por um pequeno número de linhagens, com casos restritos à República de Camarões [8]. O grupo P foi o último a ser descrito, tendo sido isolado e sequenciado a partir de uma mulher nascida nos Camarões [10].

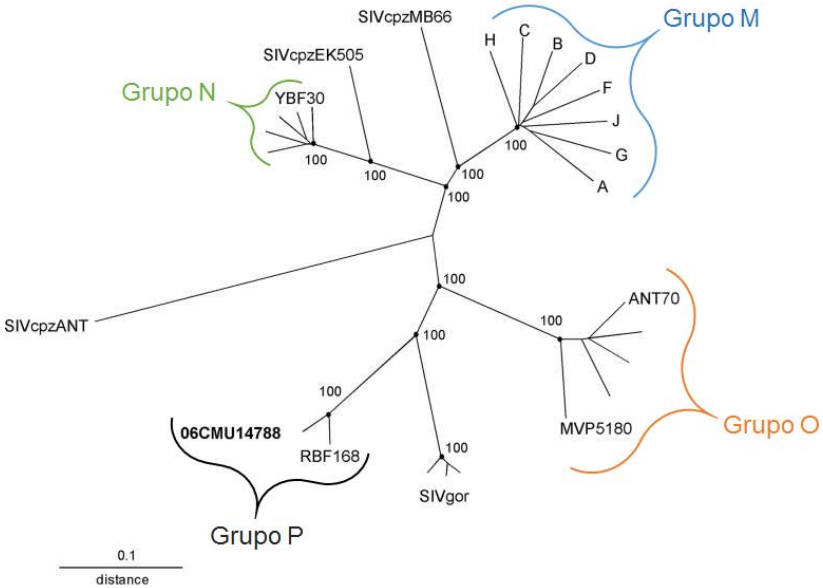

Figura 3 – Árvore filogenética derivada do alinhamento de sequências de nucleotídeos de genomas do HIV-1. O Grupo M é representado por sequências únicas para cada subtipo que o compõem (de A a H). Os grupos N e O são, cada um, representados por cinco sequências junto com as cepas de referência YBF30, ANT70 e MVP5180. O SIV (Vírus da Imunodeficiência Símia) de *G. gorila* (SIVgor) é representado por três sequências (CP684, CP2135 e CP2139). O alinhamento é constituído por sequências de 7.509 nucleotídeos. Reprodutibilidade dos nós principais é mostrada em porcentagem. A cepa de referência SIVcpzANT de chipanzé (*Pan troglodytes*) foi utilizado como grupo externo. Fonte: Modificado de Vallari e colaboradores (2011) [10].

O grupo M ostenta uma estrutura filogenética bastante característica. A grande maioria das sequências inerentes a este grupo pertence a um grande número limitado de clados equidistantes, o que permite classifica-lo em subtipos [8]. Para uma linhagem pertencer a um subtipo ela deve, ao longo de todo o seu genoma, ser aparentada a um determinado subtipo e a nenhum outro existente. Tomando como base tal princípio, nove subtipos compõem a divisão de clados do grupo M: A-D, F-H, J e K [8, 12]. Mais recentemente, os subtipos A e F foram separados em subclados, originando os sub-subtipos A1-A4, F1 e F2. Entre os subtipos do HIV-1 a similaridade entre os nucleotídeos é de 20% para a região do envelope, 16% para a região correspondente ao gene *gag* e 13% para a região da polimerase. Para os sub-subtipos, por sua vez, a

divergência é de 11 a 16% para o gene do envelope e de 7 a 12% para *gag* [12].

Ademais, existem mais de 70 formas recombinantes circulantes (CRFs) descritas (www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html). Tais CRFs são definidas como cepas virais causadoras de infecções em três ou mais indivíduos não relacionados epidemiologicamente (ou seja, que não foram infectados pelo mesmo indivíduo) [12, 13]. As CRFs são responsáveis por mais de 20% dos casos de HIV/aids em todo o mundo. Enquanto em um período de cerca de 8 anos o número de CRFs cresceu 4,5%, houve um acréscimo de 50% no número de infecções causadas por CRFs, sendo que durante um período de 30 anos, de 18 a 21,6% das linhagens de HIV foram submetidas a recombinação inter-subtipo [12, 13].

Na Europa, Américas e Oceania o subtipo predominante é o B [8, 9]. Por ser o principal causador das epidemias nos países mais ricos, este é consequentemente o subtipo viral mais estudado e grande parte das inovações tecnológicas que visam combater o HIV são testadas e desenvolvidas contra o mesmo. No entanto, os subtipos A e C são os agentes causadores da maioria das infecções, uma vez que só no continente Africano correspondem a mais de 80% dos infectados [8, 9]. O subtipo C é comumente encontrado no Sul e Leste da África e na Índia, enquanto que o subtipo predominante no Oeste africano é o A. É importante destacar que em países desenvolvidos, apesar da dominância do subtipo B, atualmente mais de 40% das infecções são causadas por outros subtipos do grupo M, um provável efeito do aumento dos movimentos migratórios em todo o mundo [8, 9].

No Brasil, o subtipo predominante é o B, sucedido pelo subtipo F1, formas recombinantes BF1, subtipo C e formas recombinantes BC. O subtipo B é o responsável por cerca de 70 a 90% das infecções nas regiões sudeste, nordeste, norte e centro-oeste do país. O subtipo F1 e recombinantes BF1 exibem prevalências, nestas mesmas regiões brasileiras, que oscilam entre 3 e 7% e 2 e 14%, respectivamente. A frequência de subtipo C registrada em grande parte dos estudos chega no máximo a 3% [41]. Entretanto, no sul do Brasil, o contexto epidemiológico é diferente do restante do país. Estudos têm retratado uma prevalência de 27 a 47% das infecções pelo subtipo C no Rio Grande do Sul, de 48 a 79% em Santa Catarina e de 20 a 30% no Paraná [14, 15, 16, 17, 18, 19]]. Ademais, uma cepa mosaico denominada de CRF31_BC, originária de uma recombinação entre os subtipos C e B no gene *pol*, tem sido observada prevalentemente no sul do Brasil. No Rio Grande do Sul

sua frequência chega a 26%, sendo encontrada em menor frequência no Paraná e em Santa Catarina [14, 15, 16, 17].

Existem poucos estudos referentes a epidemiologia molecular em cidades interioranas nos três estados do sul do país e especialmente em Santa Catarina [15, 16, 17, 18, 19]. Por esta razão, Gräf (2015) desenvolveram um estudo que teve por objetivos esclarecer aspectos filogeográficos da epidemia de HIV-1 no sul do Brasil. Neste estudo, intitulado "*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*", foram sequenciadas 317 amostras de voluntários recrutados em diferentes cidades dos estados do Rio Grande do Sul e Santa Catarina. Deste total 56% foram classificadas como subtipo C, 22% como subtipo B, 5% como CRF31_BC e 3% como subtipo F1 ou D. Os 14% restantes eram de formas recombinantes únicas, destacando-se um grupo de sequências provenientes de indivíduos soropositivos de Joinville/SC que formaram um *cluster* individualizado com alto suporte, não se agrupando a qualquer subtipo puro. Estes resultados indicam que tais sequências possivelmente pertencem a uma variante ainda não descrita como uma forma recombinante, exigindo uma investigação mais detalhada a respeito destas sequências [14].

1.4 – Mecanismos de geração de variabilidade genética do HIV-1

Uma das características mais marcantes da biologia do HIV-1 é seu potencial de diversificação genética, tanto no paciente quanto na população. Grande parte desta variabilidade é gerada pela enzima RT. Devido à ausência de um sistema de reparo de erros pela RT, a etapa de transcrição reversa do RNA viral pode gerar 0,2 mutações por genoma a cada ciclo de replicação [20]. Outro mecanismo que pode cooperar para a diversificação genética é a alta capacidade replicativa do HIV, que gera em torno de 10^{10} partículas virais por dia. Assim, o HIV dispõe de uma elevada variabilidade genética, devido à sua elevada taxa de geração de novos vírions e de substituições nucleotídicas incorretas durante a transcrição reversa, culminando em uma taxa de 2 a 4 x 10^{-3} mutações por sitio por ano [21].

Além disso, a coinfecção de um indivíduo com duas ou mais linhagens virais pode desencadear um fenômeno de troca de material genético entre elas. Esses eventos de recombinação dão origem a genomas mosaico do HIV, constituídos por subtipos distintos. Tal episódio de recombinação ocorre durante a transcrição reversa, por meio

da mudança de substrato pela RT quando duas fitas distintas de RNA viral são usadas na síntese do cDNA [22]. Por causa da alta heterogeneidade do genoma do HIV, tanto pela grande taxa de erros cometidos pela RT quanto pelo fenômeno de recombinação, a pandemia do HIV é extremamente complexa e apresenta por vez um caráter regional [22].

A similaridade das sequências não parece ser um fator limitante para a recombinação, podendo ser observada em intergrupos (entre os grupos M e O), e inter e intrasubtipos dentro do grupo M do HIV-1 [23]. Assim, novos vírus recombinantes passam a expressar pontos de recombinação discretos e passivos de identificação através de associações filogenéticas distintas. Uma vez que os locais físicos de recombinação distribuem-se de maneira aleatória ao longo do genoma viral, os locais preferências para a ocorrência de recombinações do HIV-1 permanecem desconhecidos [24]. A ocorrência de recombinações passou a ser um evento comum entre os diferentes subtipos do HIV-1, sendo a recombinação intersubtipo a mais observada. Recombinações intrasubtipo podem ocorrer com tanta ou mais frequência que recombinações intersubtipo, entretanto são de difícil identificação [25].

Mutações, recombinações e altas taxas replicativas exercem um papel fundamental na geração de diversidade dos retrovírus. Este potencial de geração de variabilidade genética confere aos retrovírus a enorme capacidade de responder rapidamente a pressões seletivas, sejam de base imunológica ou farmacológica [22, 23, 25]. A elevada taxa mutacional do HIV-1 faz com que o sistema imune seja incapaz de combater estas inúmeras variantes virais de forma eficiente, tornando o desenvolvimento de vacinas uma tarefa difícil e, ainda, proporcionando a habilidade do vírus tornar-se resistente aos fármacos antirretrovirais [1, 22, 26].

1.5 – Métodos de inferência filogenética

Em 1859, Charles Darwin difundiu a ideia de que todos os seres vivos são fruto de um processo evolutivo lento e gradual e que todos as espécies descendem de um único ancestral comum. O postulado de Darwin sobre a origem das espécies cedeu espaço para o surgimento de uma nova ciência, a taxonomia [27, 28]. Em termos práticos, a taxonomia tenta elucidar as relações evolutivas entre os organismos utilizando o máximo de características observáveis possíveis. Tradicionalmente a reconstrução da história evolutiva é realizada por meio da comparação de

estruturas morfológicas, possível tanto pelo estudo de organismos atuais, como pelo estudo de registros fósseis de espécies já extintas [27, 28].

Comumente, a representação de uma análise evolutiva é organizada na forma de uma filogenia linear [28]. As filogenias lineares consistem em representações gráficas ramificadas semelhante a uma árvore. Os nós terminais mais externos (ou folhas) na filogenia identificam os indivíduos, genes ou proteínas que foram amostrados e incluídos na análise filogenética (táxons). Os nós internos representam indivíduos não amostrados que correspondem aos prováveis parentais dos nós adjacentes. O tamanho dos ramos (galhos da árvore) pode ter significados diferentes dependendo do método empregado (Figura 4) [28].

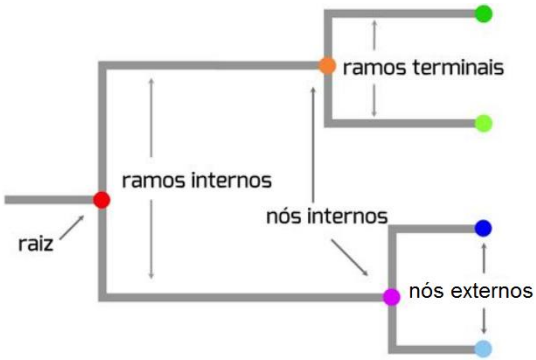

Figura 4 – Representação esquemática de uma árvore filogenética elucidando sua estrutura. Retirado de: Verli (2014) [28].

Com os avanços da biologia molecular, principalmente no desenvolvimento de técnicas para o sequenciamento gênico e proteico, surge a filogenia molecular [27, 28]. Além de uma nova abordagem baseada em um conjunto de dados de sequências aminoácidicas e nucleotídicas, a filogenia molecular oferece uma oportunidade para se estudar as relações filogenéticas entre entidades biológicas microscópicas como os vírus, fato antes impossível pelos métodos tradicionais. No entanto, obter a árvore filogenética que descreve melhor a relação evolutiva num determinado conjunto dados é uma tarefa árdua. Do ponto de vista molecular, uma série de fatores alteram a forma como um determinado organismo evolui, resultantes de diferentes forças evolutivas que reorganizam as sequências e a própria estrutura do gene ou proteína ao longo do tempo [29]. De tal modo, a escolha de um modelo evolutivo

eficaz deveria levar em consideração estas alterações evolutivas descrevendo os processos de inserção, substituição, deleção e duplicação, assim como a ocorrência de transposição e recombinações. Porém, os modelos evolutivos concebidos até hoje satisfazem apenas parcialmente estes pré-requisitos [27, 28].

Um importante passo na construção de uma filogenia acurada baseia-se na correta escolha do modelo de substituição nucleotídica. Numerosos modelos de substituição fazem diferentes estimativas acerca da probabilidade de um nucleotídeo ser substituído por outro e sobre a frequência com que estas substituições acontecem. Estes modelos podem considerar distintas probabilidades de ocorrência de transições (mudanças entre nucleotídeos de mesma classe) e transversões (trocas entre nucleotídeos de classes diferentes) [27, 28]. Apesar desta ampla gama de modelos de substituição existentes, o modelo GTR (*general time-reversible*) é o mais utilizado em análises filogenéticas do HIV (Figura 5) [27, 28].

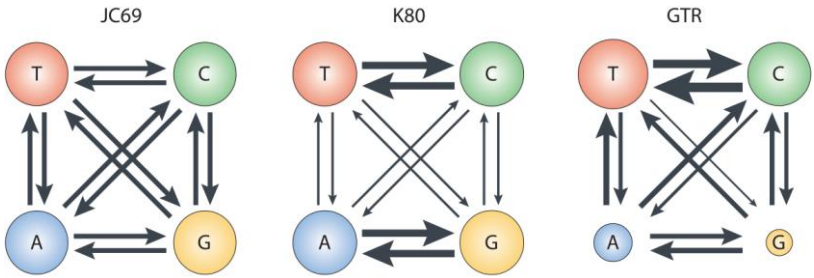

Figura 5 – Desenho esquemático exemplificando três diferentes modelos de substituição nucleotídica. A espessura das setas indica as taxas de troca entre os quatro nucleotídeos e o tamanho dos círculos representa suas frequências. O modelo JC69 (Junken e Cantor, 1969) assume a mesma probabilidade de troca para todos nucleotídeos e frequências idênticas. O modelo K80 (Kimura, 1980) 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. Fonte: Adaptado de Yang e Rannala (2012) [29].

Além da adoção de um modelo de substituição nucleotídica acurado, os métodos de inferência filogenética devem idealmente explorar o máximo possível de informações contidas em um conjunto especifico de sequências, procurando elucidar a verdadeira história evolutiva dos organismos [27, 28]. Dentre os métodos largamente

empregados na reconstrução de filogenias estão a Máxima Verossimilhança (MV) e a filogenia bayesiana [27, 28, 30].

A MV reconstrói a história evolutiva mais consistente com relação aos dados fornecidos pelo conjunto de sequências. Nele a hipótese (topologia da árvore) é avaliada pela capacidade de explicar a relação entre os dados observados (alinhamento de sequências). Basicamente, a hipótese (árvore) com maior valor de verossimilhança é a que mais provavelmente pode ser explicada pelos dados (alinhamento) [27, 28, 29]. Devido ao grande espaço amostral (2×10^6 possíveis topologias em uma análise com 10 taxons), encontrar a verdadeira árvore que melhor explica os dados é uma tarefa praticamente impossível [27, 28, 29]. Este fato implica na adoção de métodos de busca heurística (uma busca realizada com base em pressupostos previamente estabelecidos pelo programa) que tornam computacionalmente possível a realização da análise. Porém, com uso de heurística, apenas uma dezena de possíveis topologias é amostrada [27, 28, 29]. Apesar de ser considerado um dos melhores métodos de inferência filogenética, a MV deixa de explorar um conjunto muito grande de possíveis topologias de um alinhamento qualquer. Além disso, estes métodos implicam em encontrar a árvore com o valor máximo de verossimilhança entre todas as árvores amostradas, o resultado final sempre fornecerá apenas uma filogenia, ao contrário dos métodos bayesianos [28].

Em 1753 o método formal para inferência de probabilidade de um evento ocorrer atrelando evidencias prévias a este evento foi então publicado pelo reverendo Thomas Bayes [27, 28, 30]. O método desenvolvido por Bayes permite que qualquer informação útil, fornecida pelo pesquisador previamente a própria reconstrução da filogenia, poderá ser convertida em uma probabilidade anterior (ou *prior*) para ser inserida na análise filogenética [28, 30]. Entre os principais parâmetros que podem ser conhecidos antes da reconstrução filogenética pode-se destacar a taxa evolutiva, parâmetros do modelo de substituição, datas de coleta das amostras, datação proveniente de registro fóssil ou relatos históricos e organização monofilética de um grupo de indivíduos [28]. Porém, se mal estimados, os parâmetros previamente fornecidos pelo usuário podem representar um empecilho na reconstrução da filogenia que melhor represente o caminho trilhado por um dado conjunto de táxons ao longo do tempo, conduzindo o pesquisador a conclusões falhas a respeito da história evolutiva destes.

Outro ponto positivo da inferência filogenética bayesiana é que esta pode explorar um número muito maior de possíveis topologias para

um determinado conjunto de dados (Figura 6) [28, 30]. Isso é possível graças ao emprego do algoritmo de Cadeias de Markov Monte Carlo (MCMC). A ideia central do MCMC é causar pequenas alterações aleatórias em uma filogenia (topologia, tamanho dos ramos, parâmetros do modelo de substituição, etc.) e posteriormente aceitar ou rejeitar a nova hipótese de acordo com o cálculo de razão das probabilidades (ou probabilidade posterior – PP) [28, 30]. O número de vezes que uma MCMC é configurada para rodar (exemplo 10.000.000) representa o número de possíveis topologias avaliadas pelos métodos de estatística bayesiana [28, 30].

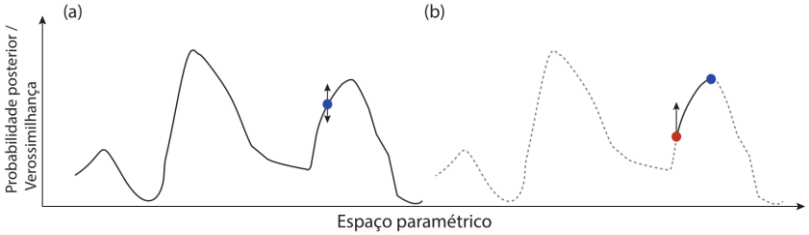

Figura 6 – Desenho esquemático mostrando a diferença entre a estimativa dos parâmetros da inferência bayesiana e de MV. (a) Na inferência bayesiana a MCMC (círculo azul) possui flexibilidade para transitar por regiões de baixa PP, atravessando os vales de baixa PP e caracterizando o espaço paramétrico na busca das regiões de mais alta PP. (b) Na inferência por MV apenas os parâmetros de maior verossimilhança são buscados. Deste modo uma pequena fração do espaço paramétrico é visitado e não sendo possível atravessar vales de baixa probabilidade. O círculo azul corresponde ao conjunto de parâmetros de máxima verossimilhança encontrado, já o círculo em vermelho a região de início da busca. Retirado de: Gräf e colaboradores (2015) [14].

Outra grande vantagem da inferência bayesiana é o emprego de uma incerteza filogenética associada a análises [28, 30] que é possível graças ao grande número de topologias amostradas capazes de explicar a distribuição dos dados. Assim, o intuito final da inferência bayesiana com MCMC não é encontrar um valor para os parâmetros que maximizem a verossimilhança dos dados, mas sim caracterizar a distribuição das topologias com alta probabilidade no espaço paramétrico [28].

1.6 – Métodos de detecção de recombinação

A evolução de um organismo é guiada não só pelo acúmulo de mutações ao longo do tempo (evolução vertical). A ocorrência de

transferência de material genético entre indivíduos pertencentes a diferentes linhagens também possui papel importante na geração de variabilidade entre os indivíduos (evolução horizontal), principalmente entre os vírus [31]. Isso faz com que a inferência filogenética representada na forma de uma arvore lineares não seja o modo mais acurado de se identificar os sinais de evolução horizontal [27]. Por consequência, métodos distintos foram e continuam sendo desenvolvidos para o estudo de recombinação em alinhamentos de nucleotídeos. Cada algoritmo foi desenvolvido para avaliar aspectos diferentes da recombinação, como por exemplo, identificar pontos de quebra, possíveis parentais recombinantes e determinar taxas de recombinação [32]. Estes algoritmos utilizados para a detecção de eventos de recombinação podem ser subdivididos de acordo com a abordagens metodológicas empregada [32].

Os métodos de distância buscam por alterações nos padrões de distância genética entre sequências. Geralmente estes métodos utilizam algum tipo de medida de distância genética bruta, por exemplo, similaridade ou dissimilaridade entre os diferentes pares de sequência presentes no alinhamento [27, 32]. Estas medidas são calculadas em segmentos adjacentes do alinhamento movendo-se por exemplo, em segmentos de 200nt a cada 10nt (*sliding window*) e então comparando a sequência estudada com os possíveis parentais [32]. Seu funcionamento é bastante simples, e os resultados são organizados em um gráfico com as medidas de distância representadas na ordenada e a posição no alinhamento na abscissa, sendo os eventos de recombinação identificados por inversões no padrão de distância [31, 32]. Estes métodos foram implementados em programas como o *SimPlot*, *Recscan*, *RAT* e *RIP* [32].

Os métodos baseados na distribuição de substituições são métodos bastante eficientes na determinação dos pontos de quebra da sequência recombinante. Estes comparam os possíveis parentais com a sequência alvo e avaliam se as substituições são uniformemente distribuídas [27, 32]. A não uniformidade na distribuição das substituições é assumida como evidência de recombinação. Estes métodos foram implementados em programas como o *Chimaera*, *RDP3*, *MaxChi*, *3seq* e *SisScan* [27, 32].

Métodos baseados em análises filogenéticas utilizam reconstrução filogenética para avaliar a existência de incongruência entre topologias de árvores inferidas a partir de diferentes regiões do alinhamento de sequências. Tais métodos são bastante eficientes na determinação das sequências parentais de um dado recombinante [32]. Dependendo do programa empregado, estes podem utilizar uma série de

modelos evolutivos (como os já citados GTR, K80 e JC69 – Figura 5), além de diferentes tipos de dados, e formas de estimar as filogenias (como análises bayesianas ou MV). As medidas são estimadas em segmentos adjacentes do alinhamento, movendo-se por exemplo em segmentos de 200nt a cada 10nt (*sliding window*) e então construindo uma árvore filogenética do alinhamento para cada fragmento criado ao longo da análise. A existência de incongruência entre topologias é evidência de recombinação entre as sequências amostradas. Estes métodos foram implementados em programas como, *SimPlot*, *Topal*, *RDP*, *JPHmm* e *JAMBE* [27, 32].

Entre os métodos supracitados, os embasados em distância como o *RIP* e o *SimPlot* e inferência filogenética, (principalmente o *SimPlot*) são os mais empregados para estudar eventos de recombinação em HIV-1 [32]. Os programas *RIP* e *SimPlot* calculam a similaridade dos segmentos adjacentes do alinhamento comparando a sequência alvo com um conjunto de sequências de referência, sendo que alterações neste padrão de distância sugerem possíveis eventos de recombinação [32]. Além do método de distância, o programa *SimPlot* também implementa o método de *BOOTSCAN*, que ao invés de estimar a similaridade entre as sequências, usa um método de inferência filogenética por uma matriz de distância obtida a partir do alinhamento (atualmente o único método disponível é o *Neighbor-Joining*) [32]. Ademais, os valores de suporte para cada agrupamento (*bootstrap*) são organizados em um gráfico semelhante ao apresentado na Figura 7. Este método também foi empregado nos programas *REGA* e *RDP V3*.5.1 [27, 32].

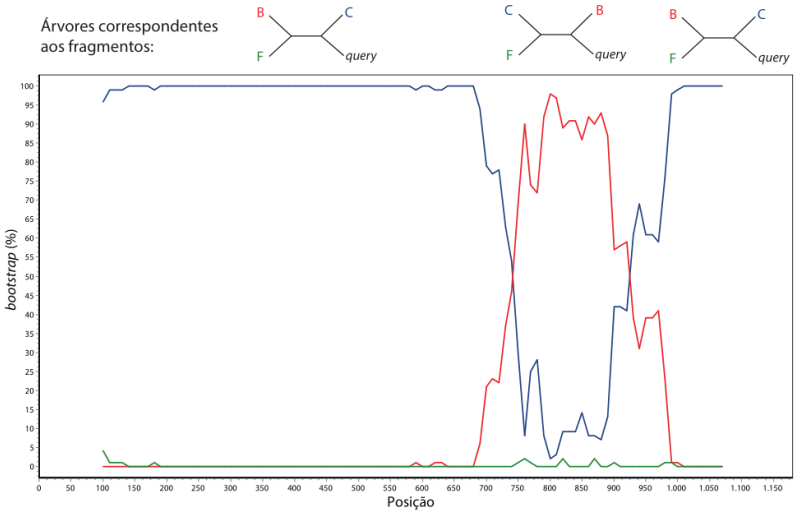

Figura 7 – Análise de *BOOTSCAN* de uma sequência recombinante. O gráfico mostra o suporte de *bootstrap* para o agrupamento da sequência em investigação com uma das três possíveis sequências parentais ao longo do alinhamento. Na parte superior do gráfico estão resumidas as árvores que representam com maior suporte a relação filogenética das sequências nas diferentes regiões do alinhamento. Fonte: Retirado de Gräf (2015) [37].

Implementado no programa *RDP3.5.1*, a abordagem de consenso metodológico é um dos mais completos meios de detecção de eventos de recombinação. Esta abordagem é a junção de todos os métodos supracitados, incorporados em um programa e utilizados conjuntamente na detecção de um evento de recombinação com base em um esquema de votação ou *score* [32]. Tal abordagem permite ao usuário utilizar métodos específicos para detecção dos pontos de quebra como o *MaxChi* e o *3seq* juntamente com métodos específicos para a detecção de sequências parentais por meio de filogenias, como o *BOOTSCAN* e o *SisScan* [32]. Assim, o programa pode estimar com mais precisão quem são os parentais e quais os pontos de quebra com base nestes dados parentais, fornecendo ao usuário um suporte estatístico para o evento de recombinação encontrado. Tais características do *RDP V3.5.1* fazem deste uma das melhores escolhas para análise de recombinação [32].

2 – JUSTIFICATIVA

Embora diversas novas CRFs surjam a cada ano, até o presente momento somente uma foi identifica no sul do Brasil, a CRF_31BC. Com uma alta frequência de indivíduos infectados pelo HIV-1 e diferentes subtipos circulantes, este dado provavelmente é fruto de uma subestimativa, refletindo a falta de estudos e consequentemente mascarando a real diversidade molecular da epidemia de aids no sul do país. Recentemente foi identificado na cidade de Joinville/SC um *cluster* de sequências do gene *pol* do HIV-1 que podem apresentar um padrão de recombinação inédito a qualquer outro subtipo ou CRF descrita. A caracterização e o estudo destas possíveis novas CRFs é importante para o melhor conhecimento dos padrões de disseminação do HIV-1. Novas sequências recombinantes, como as aqui estudadas, podem abarcar mutações de resistência a fármacos empregados no tratamento antirretroviral, ter maior aptidão replicativa e maior capacidade de infectar novas células, representando assim um obstáculo em potencial para as ações adotadas no controle da epidemia no Brasil. A vigilância epidemiológica sobre as novas variantes do HIV-1 pode ter importância no desenvolvimento de estratégias públicas de combate as novas infecções, desenvolvimento de novos fármacos e vacinas.

3 – OBJETIVOS

3.1 – Objetivo geral

Caracterizar filogeneticamente prováveis novas formas recombinantes do HIV-1 previamente identificadas em indivíduos soropositivos residentes na cidade de Joinville/SC.

3.2 – Objetivos específicos

a) Investigar em bancos de dados a presença de sequências semelhantes;

b) Identificar o padrão de recombinação das sequências a partir de análises do gene *pol*;

c) Identificar o possível local e ano de origem destas variantes;

4 – METODOLOGIA

Este trabalho é derivado de achados da tese de doutorado de Tiago Gräf. Na tese intitulada "*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*" [14], Gräf estudou a epidemia molecular do HIV-1 em várias cidades do interior do Rio Grande do Sul e Santa Catarina.

Em seu trabalho, Gräf (2015) evidenciou um grupo monofilético de cinco sequências obtidas a partir de amostras de sangue periférico de indivíduos soropositivos de Joinville/SC (AN: KR066224, KR066237, KR066211, KR066215 e KR066234). Este *cluster* de sequências agrupou-se na base do clado formado por sequências brasileiras do subtipo C e com as demais sequências também pertencentes ao subtipo C coletadas e identificadas por Gräf, compartilhando um ancestral comum entre elas e indicando assim um possível parentesco longínquo entre as sequências do subtipo C e as cinco sequências deste *cluster*. Com o objetivo de identificar se estas cinco sequências realmente apresentavam o mesmo padrão filogenético foi realizado uma genotipagem por meio do programa *RIP V3* (www.hiv.lanl.gov/content/sequence/RIP/RIP.htm).

Com base na genotipagem, foi observado que destas cinco cepas pertencentes a este *cluster*, apenas três (AN: KR066211, KR066215 e KR066234) aparentemente compartilhavam um padrão comum. A partir desta análise prévia, as etapas subsequentes para a identificação e caracterização destas três sequências foram realizadas (Figura 8). O estudo de Gräf foi aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal de Santa Catarina (UFSC), parecer número 34560.

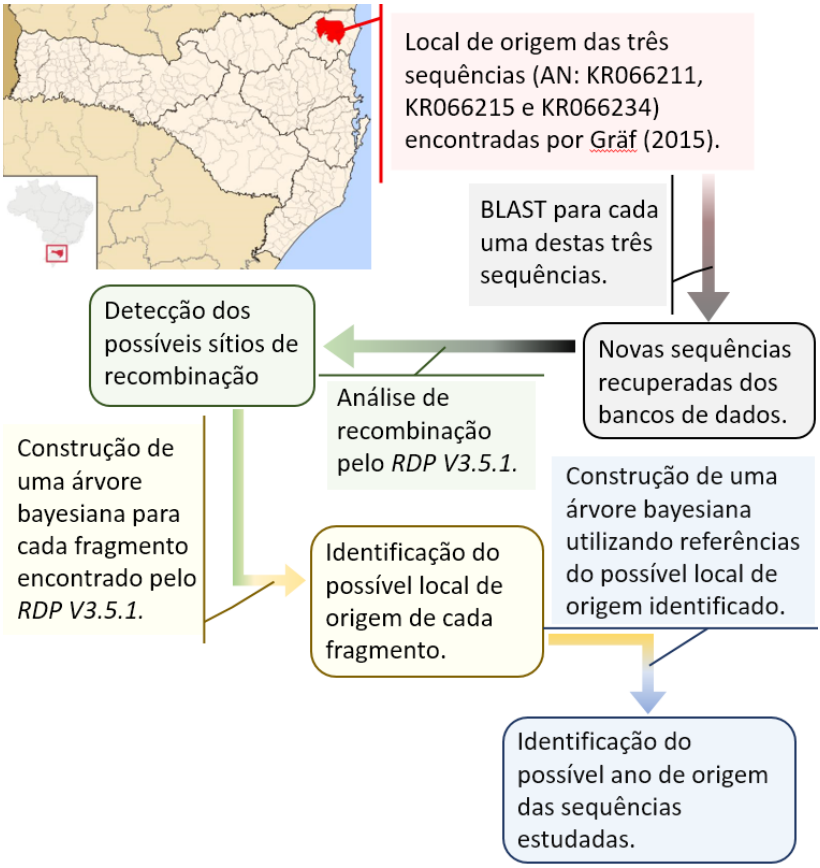

Figura 8 – Fluxograma de trabalho descrevendo cada etapa realizada para se cumprir os objetivos deste estudo.

4.1 – Montagem do *dataset* e busca por sequências de referência

A fim de garantir o bom andamento das análises filogenéticas das sequências aqui estudadas, estas foram alinhadas manualmente com sequências de referência. Para tal, sequências representantes dos subtipos B e C foram empregadas como possíveis parentais (grupo interno) e sequências dos subtipos A1 e F1, selecionadas a partir do banco de referências do programa *RIP V3* (www.hiv.lanl.gov/content/sequence/RIP/RIP.html), foram utilizadas como grupo externo. A ferramenta *Genotyping* (ncbi.nlm.nih.gov/projects/genotyping/formpage.cgi) foi empregada para

confirmar os subtipos das sequências de referência e eliminar sequências possivelmente mal identificadas. Para tanto, o programa *Genotyping* foi configurado com um tamanho de janela de 250 pb com um tamanho de passo 20 pb, sendo que as demais configurações foram mantidas no modo padrão. Para remoção das sequências de referência muito similares entre si e manutenção da heterogeneidade deste *dataset* foi empregado o programa *CD-Hit* (weizhongli-lab.org/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit), tendo sido aplicado um corte de 97% de identidade. Posteriormente, a edição do alinhamento foi executada manualmente com o auxílio do programa *AliView* [33], respeitando a fase de leitura e a igualdade do comprimento das sequências avaliadas.

4.2 - Análise dos pontos de recombinação e identificação do possível local de origem dos fragmentos recombinantes

Para a detecção dos possíveis sinais de recombinação foram utilizados os métodos *RDP*, *BootScanning*, *Chimaera*, *3Seq* e *MaxChi*. A confirmação dos sinais detectados foi realizada a partir dos métodos *SisScan* e *BootScanning*, implementados no programa *RDP V3.5.1* [32]. Foram considerados apenas os sinais de recombinação detectados por dois ou mais métodos e que apresentaram uma pontuação recombinante superior a 0,600 (*recombinant score*) nas análises individuais para cada sequência. Os parâmetros utilizados em cada método foram:

- *RDP*
  - Selecionar apenas sequências de referência internas ou sem referência.
  - Tamanho da janela: 29 pb.
  - Detectar recombinação entre sequências com 0% a 100% de identidade.

- *BootScanning*
  - Tamanho da janela: 250 pb.
  - Tamanho do passo: 10 pb.
  - Número de bootstrap: 100.
  - Usar árvores Neighbor-joining.
  - Modelo: Felsenstein, 1984.

- *Chimaera*
  - Tamanho da janela: Fixo
  - Fração de sítios variáveis por janela: 50.

- *MaxChi*
  - Tamanho da janela: Fixo.
  - Fração de sítios variáveis por janela: 50.
  - Ignorar *Gap*: Sim.
- *SisScan*
  - Tamanho da janela: 250 pb.
  - Tamanho do passo: 10 pb.
  - Número de permutações:100.
  - Ignorar *Gap*: Sim.

Para determinar o intervalo de confiança nas regiões de quebra foi utilizado o cálculo de desvio padrão a partir da média aritmética dos pontos de início e termino das regiões recombinantes para cada sequência. Objetivando estimar o possível local de origem dos fragmentos encontrados pelo *RDP V3.5.1* o mesmo alinhamento utilizado na detecção dos possíveis sinais de recombinação foi dividido em fragmentos selecionados conforme os pontos de recombinação identificados e submetidos a uma análise filogenética bayesiana (como descrito no item 4.4). Os fragmentos resultantes, correspondentes a um mesmo subtipo, foram concatenados para aumentar a região analisada e melhorar a inferência filogenética.

4.3 – Busca por sequências similares em bancos de dados

Para averiguar a existência de mais cepas similares as encontradas por Gräf, disponíveis nas bases de dados do *Los Alamos HIV database* (www.hiv.lanl.gov), cada uma das sequências analisadas foi submetida a um *BLAST*, utilizando a ferramenta *on-line HIV BLAST* (www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html). Foram selecionadas as 100 sequências mais similares a cada recombinante, totalizando ao final desta etapa 300 novas sequências. Logo após, todas as sequências obtidas foram unidas em um único alinhamento, excluindo aquelas repetidas obtidas dos três diferentes *BLAST*. Para identificar quais das sequências restantes realmente apresentavam o mesmo padrão de recombinação aqui estudado, cada sequência foi submetida a uma análise previa de recombinação utilizando o programa *RIP V3* (www.hiv.lanl.gov/content/sequence/RIP/RIP.html). Foram empregadas as sequências de referências disponibilizadas pelo programa,

sendo que os subtipos A1 e F1 foram empregados como grupo externo e C e B como possíveis parentais. Para tanto, o programa foi configurado para um tamanho de janela de 200 pb com um mínimo de 90% de confiança sendo que, as demais opções foram mantidas no modo padrão. Após esta identificação prévia, as sequências que apresentavam o mesmo padrão que as sequências alvo foram submetidas à análise descrita no item 4.2 a fim de identificar a possível presença de fragmentos recombinantes para cada nova sequência.

4.4 – Representações gráficas e construção das escalas dos recombinantes

Para determinar a qual porção gênica do HIV-1 as sequências aqui estudadas pertenciam e as respectivas posições dos pontos de quebra encontrados, foi empregado a ferramenta *HIV-1 Sequence Locator* (www.hiv.lanl.gov/content/sequence/LOCATE/locate.html) utilizando como referência a cepa HXB2 (AN: FB341548). A construção das representações gráficas da estrutura mosaico do HIV-1, obtida a partir das análises de recombinação, foi realizada com o auxílio do programa *Recombinant HIV-1 Drawing Tool V2.1.0* (www.hiv.lanl.gov/content/sequence/draw_crf/recom_mapper.html).

As escalas com as posições nucleotídicas de interesse para a determinação do padrão de recombinação foram construídas manualmente. Para tal, alinhou-se as sequências de interesse juntamente com o principal parental de cada subtipo identificados nas análises de recombinação (C - AY727523 e B - FJ487849). Em seguida, todas as colunas que apresentavam regiões conservadas entre todas as sequências do alinhamento foram removidas mantendo-se apenas as colunas com sítios que apresentavam alguma dissimilaridade. Posteriormente, o alinhamento cortado foi submetido a uma análise de recombinação empregando o método *3seq* no programa *RDP V3.5.1* para a obtenção gráfica da escala [32].

4.5 - Detecção de mutações de resistência aos antirretrovirais

Para determinar a presença de mutações relacionadas à resistência e susceptibilidade a antirretrovirais, as sequências aqui estudadas foram submetidas à análise pela ferramenta *HIVdb program* (sierra2.stanford.edu/sierra/servlet/JSierra?action=sequenceInput).

4.6 – Estimativa temporal do clado de recombinantes e analises filogenéticas

As análises filogenéticas e a data de origem do clado das sequências alvo deste estudo foi estimada através de inferência filogenética bayesiana utilizando-se o pacote *BEAST V1.8.3* [34]. Para tanto, foi empregado o modelo de substituição nucleotídica GTR, incluindo heterogeneidade de sítios modelado por distribuição *Gamma* e sítios invariáveis (GTR+G4+I), juntamente com um relógio molecular relaxado com distribuição lognormal e modelo de coalescência não paramétrico *Bayesian Skyline*. Devido ao pequeno sinal temporal do *dataset* utilizado para esta análise, empregou-se um *prior* normal (média = 39 anos, desvio padrão = ± 5.1 anos) para origem do clado HIV-1 subtipo C brasileiro, conforme estimado anteriormente por Delatorre e colaboradores (2013) [18].

A MCMC foi corrida por 100 milhões de estados. O programa *Tracer v1.6* (beast.bio.ed.ac.uk/Tracer) foi utilizado para checar a convergência dos parâmetros que foram considerados satisfatórios quando o tamanho efetivo de amostragem (ESS) estava acima de 200. A árvore de máxima credibilidade dos clados (MCC) foi selecionada da distribuição posterior de árvores pelo programa *TreeAnnotator V1.8.3*, disponível no pacote de programas *BEAST*. Para tanto, foram descartadas os 10% iniciais (*burn-in*). As árvores foram visualizadas e editadas no programa *FigTree V1.4.2* (tree.bio.ed.ac.uk/software/figtree/).

5 – RESULTADOS E DISCUSSÃO

5.1 – Obtenção e caracterização filogenética de sequências similares

Com o objetivo de obter mais sequências similares às três previamente observadas foi realizado um *BLAST*. Cinco novas variantes com o mesmo padrão filogenético foram obtidas, totalizando oito sequências (AN: KT744389, KT744015, KT747433, FJ591352, AF009376, KR066211, KR066215 e KR066234). Todas foram obtidas de amostras de sangue de indivíduos HIV positivos brasileiros coletadas entre os anos de 1992 e 2013 e apresentam um comprimento que varia entre 1029 e 1302 pb, correspondente ao primeiro terço do gene *pol* do HIV-1. Além disso, estas oito sequências foram empregadas em estudos

anteriores; porém, em nenhum dos trabalhos investigou-se o caráter recombinante das mesmas (Tabela 1).

Tabela 1 – Dados de interesse disponíveis sobre as sequências estudadas.

Número de acesso	País/Data	Posição (HXB2)	Autor	Objetivos do Estudo
KT744389	Br/2009	2262 - 3290	Diaz *et al.* (2015) [35]	Avaliaram o impacto do tratamento antirretroviral no estado imunológico e virológico de pacientes HIV soropositivos ao longo do tempo.
KT744015	Br/2009	2262 - 3290		
KT747433	Br/2008	2262 - 3290		
KR066211	Br/2012	2253 - 3516	Gräf *et al*. (2015) [14]	Descreveram a epidemia do HIV-1 nos estados de Santa Catarina e Rio Grande do Sul, avaliando a dinâmica de dispersão do subtipo C do HIV-1 em diferentes cidades.
KR066215	Br/2012	2254 - 3508		
KR066234	Br/2013	2253 - 3552		
AF009376	Br/1992	2265 - 3440	Cornelissen *et al*. (1997) [36]	Observaram o surgimento natural de mutações de resistência a inibidores de protease em cinco diferentes subtipos do HIV-1 (A, B, C, D e E) e formas recombinantes.
FJ591352	Br/2005	2253 - 3554	Martinez-Cajas *et al*. (2009) [37]	Uma revisão bibliográfica sistemática de estudos publicados entre 1996 e 2008 feitos com amostras de pacientes com infecções causadas por subtipos não B.

Posição (HXB2): Posição de início e término de cada sequência em relação a cepa de referência HXB2 (AN: FB341548).

Além destas, um grupo de sequências originárias da Itália, com um padrão de recombinação bastante semelhante ao aqui estudado, também foi obtido a partir do *BLAST*. Tais sequências são representantes

da CRF60_BC. Descrita por Simonetti e colaboradores em 2013, esta CRF apresenta três fragmentos correspondentes ao subtipo B, intercalados em uma sequência predominantemente subtipo C, com pontos de quebra bem delimitados [38]. O primeiro fragmento encontra-se entre a posição 1 e 463 do HIV-1, em uma região correspondente aos dois primeiros terços da porção 5' LTR. O segundo fragmento está entre as posições 2804 e 3038, na região correspondente ao gene *pol* do HIV-1. O terceiro e último fragmento apresenta-se entre as posições 8662 e 9719, abrangendo uma pequena porção no final do gene *env* até os dois primeiros terços da região 3' LTR (Figura 9).

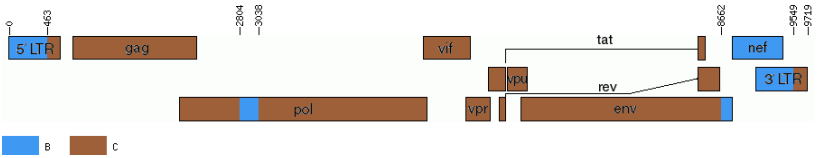

Figura 9 – Representação gráfica da estrutura em mosaico da CRF60_BC. As posições dos pontos de quebra de recombinação foram estimadas em relação a cepa HXB2**.** Pontos de recombinação foram estimados através do programa *SimPlot V3.5.1*. Fonte: Retirado de Simonetti e colaboradores (2013) [38].

Simonetti e colaboradores (2013), com base em análises filogenéticas dos fragmentos correspondentes ao subtipo C, levantaram a hipótese do provável surgimento da CRF60_BC na América do Sul, uma vez que tais fragmentos formaram um *cluster* com uma sequência brasileira do subtipo C. Entretanto, esses autores não apresentaram dados referentes a esta filogenia. Além disto, também discutiram a possibilidade do surgimento da CRF60_BC na Itália, uma vez que um dos indivíduos, o soropositivo diagnosticado há mais tempo incluído no estudo, relatou ter visitado o continente sul-americano. Com base nisso, Simonetti e colaboradores afirmaram que o indivíduo em questão pode ter contraído o subtipo C, que circula comumente no sul do Brasil, durante a sua viagem para o continente sul americano, originando uma coinfecção entre os dois subtipos parentais desta CRF [38].

Com a hipótese de um provável surgimento da CRF60_BC na América do Sul, foi construída uma árvore bayesiana empregando-se seis sequências representantes da CRF60_BC (AN: JQ675734, JQ675738, JQ675741, JQ675749, JQ675750 e JQ675752) e as oito sequências aqui estudadas a fim de se identificar a relação entre elas (Figura 10). Como pode ser observado na figura 10, tanto as sequências aqui estudadas como as cepas representantes da CRF60_BC formam clados monofiléticos

individualizados entre si, porém não apresentando um suporte estatístico significante no nó comum entre elas (PP < 0,9), o que sugere a origem de ambas em diferentes eventos de recombinação. Assim, entende-se que não há necessidade de prosseguir com a análise conjunta da CRF60_BC, uma vez que esse evento distinto de recombinação já foi caracterizado e estas não são representantes de um mesmo clado.

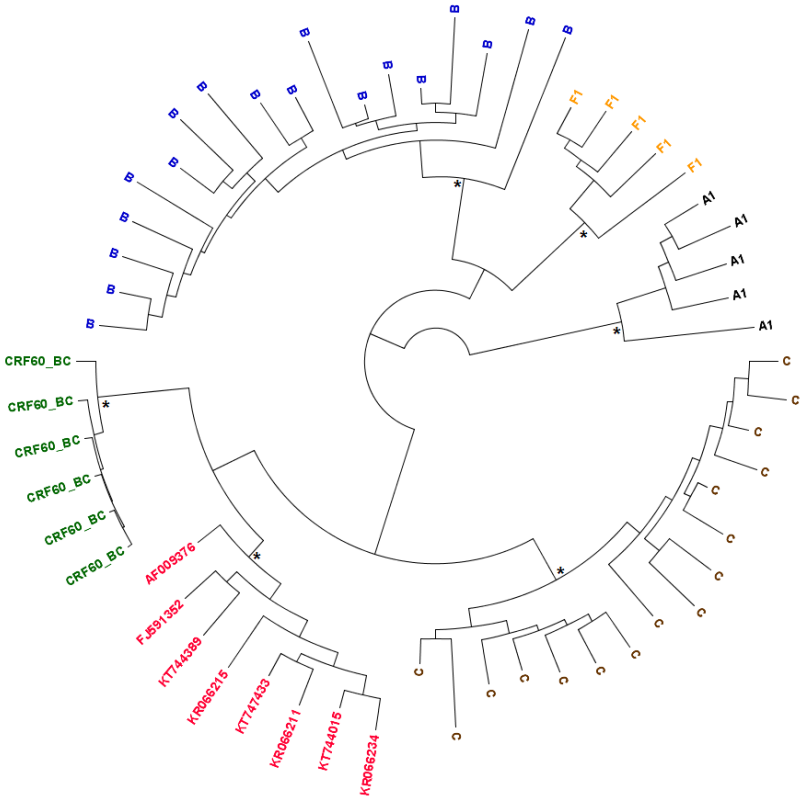

Figura 10 – Árvore bayesiana das CRF60_BC e das oito sequências alvo deste estudo (clado vermelho) com base nos fragmentos do gene *pol* do HIV-1. Os ramos foram coloridos de acordo com o subtipo e forma recombinante. Os asteriscos nos pontos de ramificação correspondem aos valores de PP superiores ou iguais a 0,9. Cepas do subtipo A1 e F1 foram empregadas como grupo externo.

5.2 – Análise do padrão de recombinação

Com o objetivo de identificar um possível evento de recombinação, as oito sequências alvo deste estudo foram submetidas a uma análise de recombinação pelo *RDP V3.5.1*. Com base nas análises individuais realizadas aqui, dois pontos de quebra semelhantes foram encontrados para cada sequência, dividindo-as em três fragmentos recombinantes. O primeiro fragmento corresponde ao subtipo C e inicia-se na posição 2265 e termina na posição 2688. O segundo fragmento corresponde ao subtipo B e seu início encontra-se entre as posições 2689 e 2715 (2702, +-13), terminando entre os pontos 3075 e 3087 (3081, +-6). O terceiro e último fragmento corresponde ao subtipo C, iniciando na posição 3088 e terminando na posição 3290 (Figura 11).

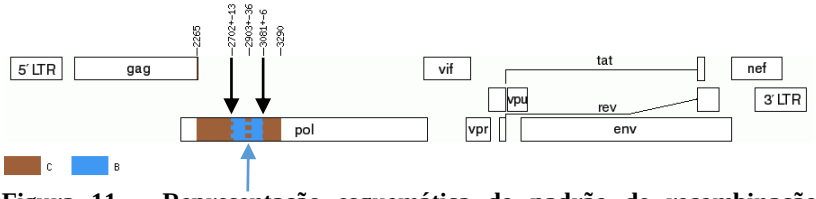

Figura 11 – Representação esquemática do padrão de recombinação identificado no recombinante aqui estudado. As regiões hachuradas (indicada pelas setas pretas) no primeiro (2702, +-13) e segundo (3081, +-6) pontos de quebra, identificam possíveis regiões de ocorrência dos pontos de quebra. A região hachurada (indicada pela seta azul) no centro do fragmento B (2903, +-36) denota uma zona de incerteza característica entre os subtipos B e C.

Além disso, foi evidenciado uma zona de incerteza entre os subtipos parentais no centro do fragmento B (2903, +-36). Esta região apresenta-se mais claramente nas sequências mais derivadas (AN: KT744389, KT744015, KT747433, FJ591352, KR066211, KR066215 e KR066234), sendo que em alguns casos, métodos mais sensíveis para a detecção de pontos de quebra, como o *MaxChi* e o *3seq*, identificam esta região como um fragmento pertencente ao subtipo C (dados não apresentados). Tal zona de incerteza pode ser melhor visualizada em uma escala construída destacando-se os nucleotídeos de importância para a detecção deste evento de recombinação ao longo das sequências. Nesta escala, pode-se observar pontos em comum ao principal parental do subtipo C (AN: AY727523) indicado pelo *RDP3 V3.5.1* para este evento de recombinação (Figura 12).

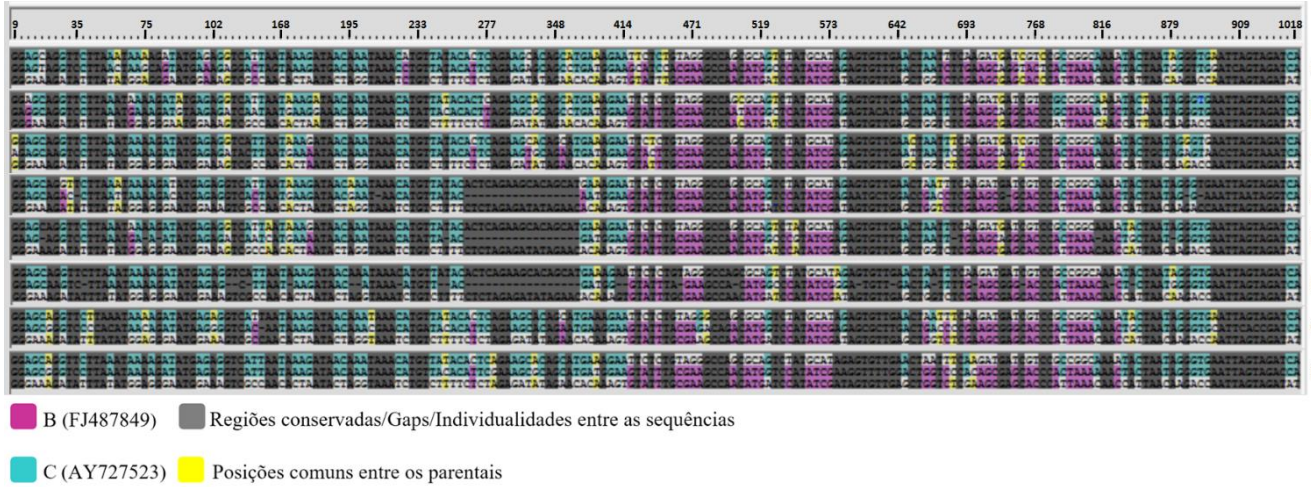

Figura 12 – Representação em escala das sequências recombinantes em relação aos dois parentais B (AN: FJ487849) e C (AN: AY727523) mais recorrentes. As posições nucleotídicas mantidas na escala são aquelas de interesse para a determinação do padrão de recombinação. Os intervalos entre os pontos são dependentes da similaridade com os parentais, portanto não apresenta intervalos regulares. As letras na lateral direita da figura representam os recombinantes (A: KR066211, B: KR066234, C: KR066215, D: KT744389, E: KT744015, F: KT747433, G: FJ591352, H: AF009376). A sequências no centro de cada trinca corresponde ao recombinante. A sequências na base de cada trinca representa o parental B e a sequência no topo de cada trinca representa o parental C.

Aparentemente, tal zona de incerteza não possui qualquer associação com a exposição prévia a antirretrovirais pois, nas análises realizadas aqui, apenas três sequências apresentavam alguma mutação de resistência (AN: FJ591352, KT747433 e KT744015). Entretanto, nenhuma dessas mutações ocorreu na referida região de incerteza. Além disso, a sequência proteica codificada por essa zona de incerteza não apresenta nenhum aminoácido diferente daqueles comumente observados nas sequências utilizadas como referência (dados não apresentados). Isto indica que todas as mutações que aconteceram neste fragmento ambíguo são do tipo silenciosa (mutação que não altera o aminoácido codificado pelo códon onde ela ocorre). Estes resultados sugerem que esta zona de incerteza pode ter surgido gradualmente por simples erro da RT cometidos em diferentes ciclos de infecção, permanecendo por não causar qualquer alteração que prejudique estas cepas.

Independente de como esta região de incerteza surgiu, é fato que ela permanece como algo recorrente mesmo quando outros parentais C (AN: DQ191006, U52953 e KR066289) ou B (AN: EF514711, AY037269, JQ619614, AF042101 e DQ295194) são empregados na detecção deste evento de recombinação (dados não apresentados), consolidando ainda mais a possibilidade de uso desta zona de incerteza como um indicador adicional para a detecção do evento de recombinação elucidado aqui.

A fim de se confirmar as regiões recombinantes delimitadas pelos pontos de quebra encontrados na análise de recombinação, foi construída uma árvore bayesiana para o fragmento do subtipo B e outra para os fragmentos pertencentes ao subtipo C concatenados. Através da árvore da região B, observa-se a formação de um *cluster* monofilético das sequências aqui estudadas (PP = 0,97) agrupando-se à base do clado formado pelas sequências do subtipo B (PP = 0,99), confirmando os achados iniciais deste trabalho (Figura 13). Tal distinção das demais sequências do subtipo B, adotadas como referência, pode ser justificada pela presença da zona de incerteza demonstrada anteriormente (Figura 12).

Ademais, uma sequência pertencente ao subtipo B puro (AN: JF487849) agrupou-se na base do clado formado pelo fragmento B dos recombinantes alvos deste estudo (sequências em vermelho na figura 13). Esta sequência é proveniente de um indivíduo do Rio Grande do Sul, obtida em 2006 [19]. Isto indica, mais uma vez, o surgimento desta variante no Brasil. Além disto, como já apresentado, esta foi a sequência

que mais se repetiu como o principal parental do subtipo B nas análises de recombinação pelo *RDP3 V3.5.1* (Figura 12).

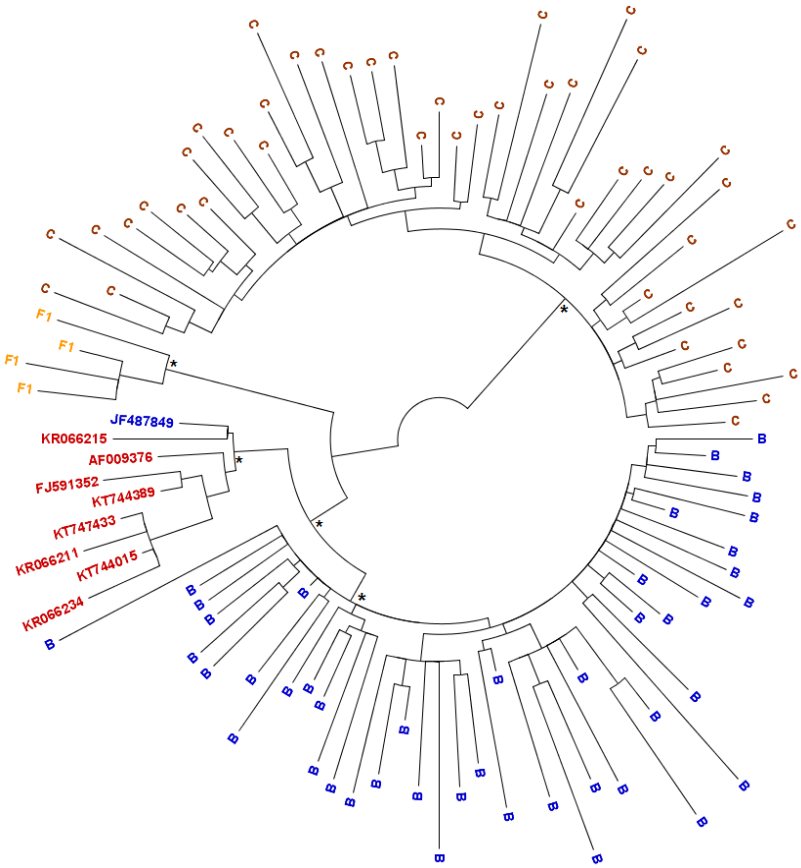

Figura 13 – Árvore bayesiana do fragmento recombinante do subtipo B com base em fragmentos do gene *pol* do HIV-1. Com exceção das sequências alvo deste estudo (em vermelho) e seu principal parental (AN: JF487849 – em azul) os ramos foram coloridos e nomeadas de acordo com os subtipos empregados. Os asteriscos nos pontos de ramificação correspondem aos valores de PP superiores ou iguais a 0,9. Cepas do subtipo F1 foram empregadas como grupo externo.

A árvore construída a partir dos fragmentos do subtipo C concatenados exibiu resultado semelhante, sendo que as sequências aqui estudadas (em vermelho) formaram um grupo monofilético com sequências do subtipo C brasileiras (Figura 14). A principal diferença

desta para a árvore do fragmento B é que as variantes alvo deste estudo não formaram um *cluster* individual, mas distribuíram-se ao longo de todo o grupo de sequências brasileiras. Este resultado pode, ao contrário do que se observa no fragmento do subtipo B, indicar a boa conservação dos fragmentos correspondentes ao subtipo C deste recombinante, uma vez que como já demonstrado por Delatorre e colaboradores em 2013 e Mendonça e colaboradores em 2010, as sequências que compõem o clado do subtipo C brasileiro derivam de um evento único de inserção ocorrido em meados dos anos 70, justificando este padrão de agrupamento [18, 42].

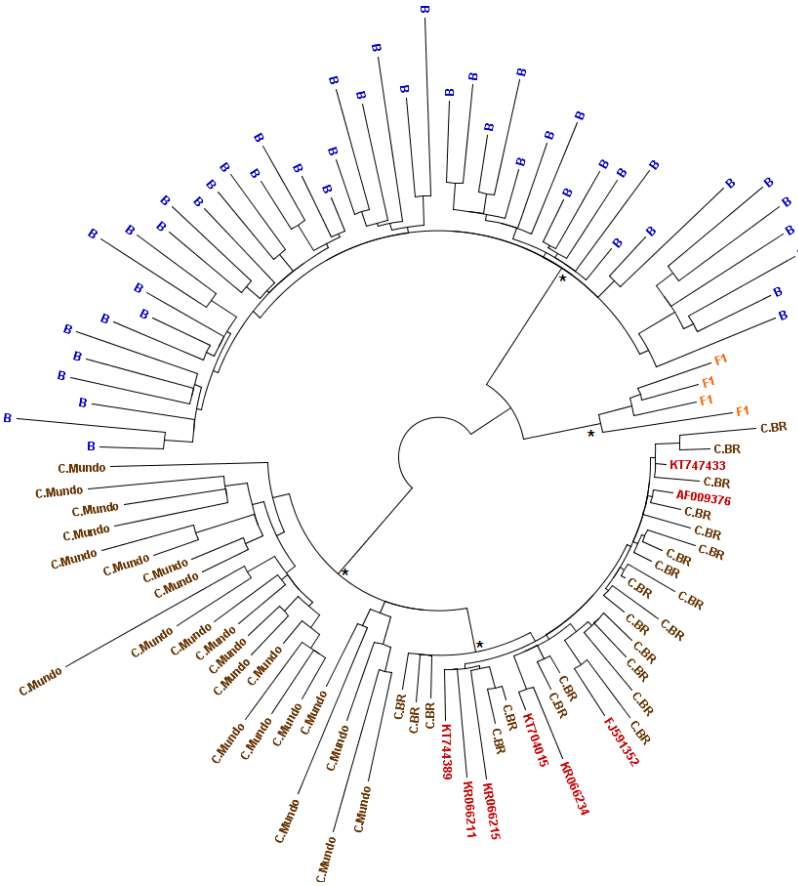

Figura 14 – Árvore bayesiana do fragmento recombinante do subtipo C com base em fragmentos do gene *pol* do HIV-1. Com exceção das sequências alvo deste estudo (em vermelho) os ramos foram coloridos e nomeadas de acordo com os subtipos empregados. Os asteriscos nos pontos de ramificação correspondem aos valores de PP superiores ou iguais a 0,9. Cepas do subtipo F1 foram empregadas como grupo externo.

Diante do exposto até aqui, entende-se que as oito sequências alvos deste estudo possivelmente descendem de um mesmo evento de recombinação, caracterizando-as como novas candidatas a CRF com origem no Brasil. Além dos resultados apresentado, outros recombinantes entre os subtipos B e C, com possível origem no Brasil ou/e com algum dos parentais possivelmente brasileiros, foram descritos em estudos

anteriores [14, 15, 16, 38, 39, 40], indicando o Brasil como um local de ampla cocirculação e possível ocorrência de eventos de recombinação envolvendo estes dois subtipos.

5.3 – Estimativa temporal do clado de recombinantes

Com o objetivo de identificar o ano de origem deste novo candidato a CRF foi construída uma árvore bayesiana com escala de tempo utilizando como referência sequências do subtipo C brasileiras e a forma recombinante CRF31_BC também do Brasil (Figura 15).

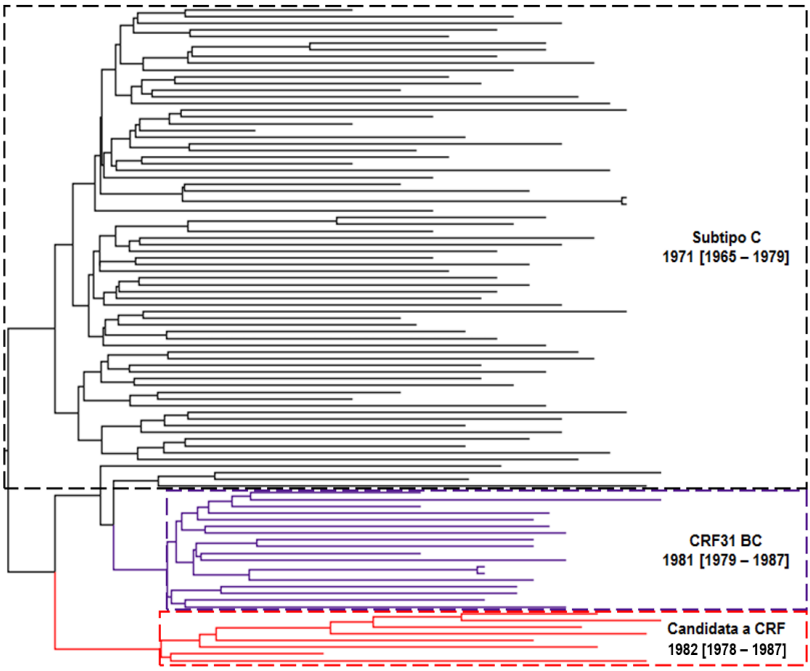

Figura 15 – Árvore bayesiana para determinação do ano de surgimento do ancestral comum da nova candidata a CRF (clado vermelho) com base em fragmentos do gene *pol* do HIV-1. As datas de surgimento do ancestral comum de cada clado estão dentro dos retângulos pontilhados que os delimitam. Os limites para o surgimento de cada clado, assumidos a partir de uma probabilidade posterior de 95% (HPD), estão entre colchetes. A altura dos nodos apresentados na figura é baseada na árvore de máxima credibilidade selecionada da distribuição de árvores estimadas pelo BEAST. As datas de origem e 95% HPD são baseados em todas as árvores estimadas após exclusão do *burn-in*.

As análises bayesianas sugerem que a nova variante aqui apresentada, provém de um evento de recombinação ocorrido em 1982 (95% HPD: 1978 – 1987). Este resultado é condizente com o surgimento da CRF31_BC, nove anos após a introdução do subtipo C no Brasil. Ademais, estudos anteriores demostram que o subtipo C brasileiro é altamente aparentado a sequências comumente isoladas de pacientes soropositivos do Burundi, no leste da África [18]. Provavelmente sua chegada deu-se devido a uma grande onda migratória ocasionada por uma guerra civil, na qual mais de 300 mil refugiados deixaram este pais. Esta migração em massa desempenhou um papel fundamental na disseminação internacional do subtipo C do leste africano. Apesar da rota exata de dispersão deste subtipo do Burundi para o Brasil ser desconhecida, análises filogeográficas indicam que sua introdução ocorreu através da região sul [18]. Tal fato é, também, sustentado pela alta prevalência do subtipo C nos três estados do sul e seu decréscimo desta região para o norte e países vizinhos [41].

Além disso, o subtipo B circula no Brasil desde meados da década de 60, dez anos antes da inserção do subtipo C e dezenove anos antes do surgimento da forma recombinante aqui descrita [42]. O evento de introdução do subtipo B é ainda mais nebuloso. Estudos indicam que este evento pode ter ocorrido a partir da América do Norte, Ásia e outros países da América do Sul (como Venezuela, Argentina e Colômbia), sustentando a hipótese de diferentes eventos de introdução do subtipo B no Brasil [42]. Além disso, Mendonça e colaboradores (2010) demonstraram que os dois principais eventos de introdução do subtipo B, acontecerem nos estados do Rio Grande do Sul e no Rio de Janeiro em meados de 60 [42]. Outro fato importante é que o subtipo B é o segundo mais prevalente no sul do Brasil (aproximadamente 22%) e o mais prevalente no restante dos estados brasileiros e países da América do Sul, sendo que o subtipo C apresenta em média uma baixa prevalência nas demais regiões do país (aproximadamente 10%) [41].

A recente inserção do subtipo C na região sul do Brasil, sua co-circulação em altas prevalências com o subtipo B há pelo menos 40 anos, a recorrente presença de formas recombinantes entre os subtipos B e C somado aos resultados obtidos nas análises realizadas aqui, corroboram a hipótese de surgimento desta candidata a CRF no sul do Brasil.

Ademais, outras CRFs entre os subtipos B e C, com um padrão de recombinação semelhante ao observado neste trabalho, foram anteriormente descritos (Figura 16) [15, 38, 44, 45, 46]. Apesar deste semelhante padrão, tais recombinantes são bastante divergentes, sendo

suficientemente distantes entre si ao ponto de não apresentarem qualquer parentesco próximo (Figura 17). Além destas conhecidas CRFs, existem ainda uma série de formas recombinantes únicas amplamente observadas nos bancos de dados e na literatura com este mesmo padrão [14, 16, 25, 39, 40, 47]. Tais indícios apontam esta região do gene *pol*, correspondente a uma porção codificante da RT, como um possível *hotspot* de recombinação e consequente geração de variabilidade entre ambos.

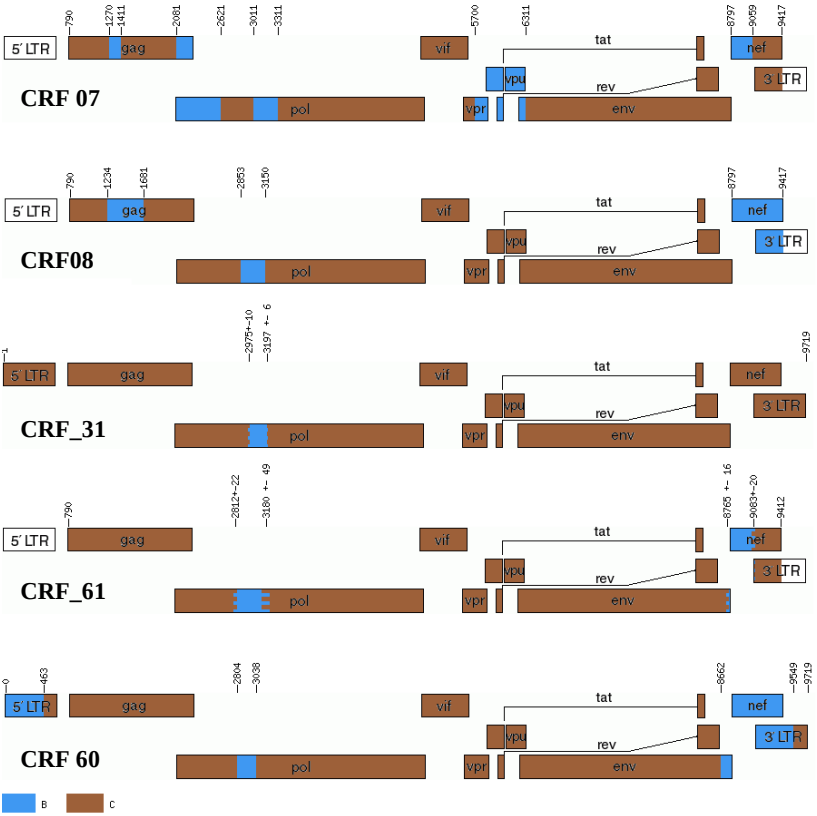

Figura 16 – Representação gráfica da estrutura em mosaico das CRFs entre o subtipo B e C que apresentam pontos de recombinação na porção codificante da RT no gene *pol* do HIV-1. As posições dos pontos de quebra de recombinação foram estimadas em relação a cepa HXB2**.** Fonte: Adaptado de *Los Alamos HIV database* (www.hiv.lanl.gov).

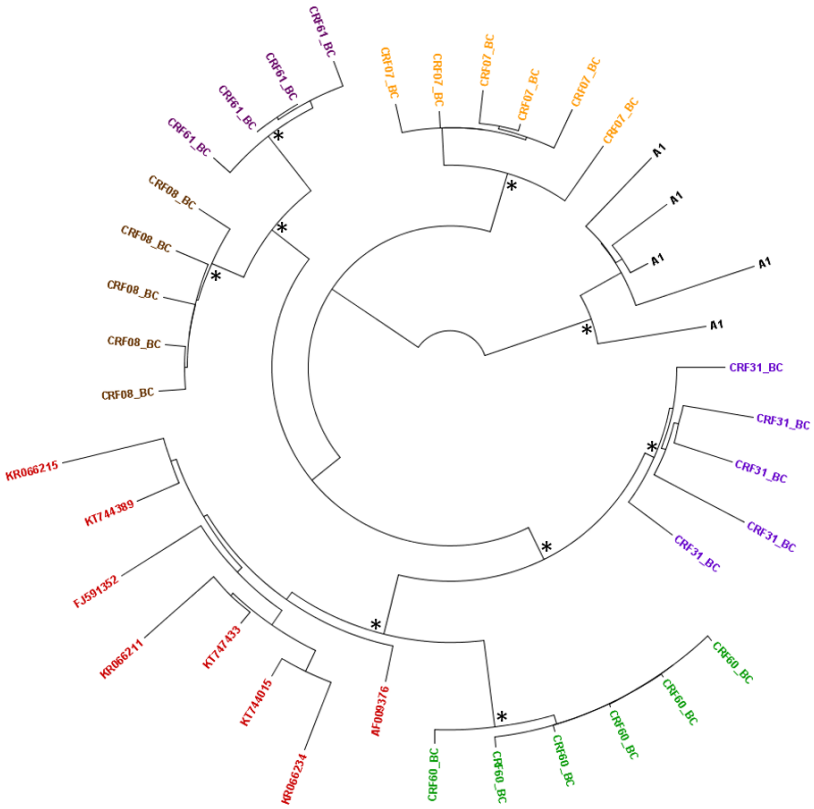

Figura 17 – Arvore bayesianas das CRFs entre os B e C com base em fragmentos do gene *pol* do HIV-1. Com exceção das sequências alvo deste estudo (em vermelho) os ramos foram coloridos e nomeadas de acordo com as CRFs representadas. Os asteriscos nos pontos de ramificação correspondem aos valores de PP superiores ou iguais a 0,9. Cepas do subtipo A1 foram empregadas como grupo externo**.**

6 – CONCLUSÕES

Diante dos resultados obtidos, entende-se que as oito sequências alvo deste estudo possivelmente derivam de um mesmo evento de recombinação entre cepas dos subtipos B e C do HIV-1. Tais recombinantes apresentam dois pontos de quebra, que o divide em 3 fragmentos. O primeiro fragmento corresponde ao subtipo C e inicia-se na posição 2265 e termina na posição 2688. O segundo fragmento

corresponde ao subtipo B e seu início encontra-se entre as posições 2689 e 2715 (2702, +-13), terminando entre os pontos 3075 e 3087 (3081, +-6). O terceiro e último fragmento corresponde ao subtipo C, iniciando na posição 3088 e terminando na posição 3290. Além disso, as análises filogenéticas realizadas, juntamente com as evidências previamente descritas, sugerem que este recombinante provavelmente surgiu no sul do Brasil no início dos anos 80. Apesar dos fortes indícios aqui apresentados, apenas um pouco mais de 10% de toda a extensão gênica do HIV-1 foi avaliada para este recombinante. Isso torna necessário a realização de mais estudos envolvendo diferentes regiões gênicas de cepas do HIV-1 com este padrão de recombinação, incluindo a caracterização de genomas completos para a efetiva determinação desta como uma nova CRF entre os subtipos B e C do HIV-1.

7 – REFERÊNCIAS BIBLIOGRÁFICAS

1. HOFFMANN, C.; ROCKSTROH, J.; KAMPS, B. HIV Medicine. 15ª edição: Paris, **Flying Publisher**, 2007.
2. KINDT, T. J.; GOLDSBY, R. A.; OSBORN, B. A. Kuby Immunology. 6ª edição: New York, **Artmed**, 2008.
3. WU, L.; KEWALRAMANI, V. N. Dendritic-cell interactions with HIV: infection and viral dissemination. **Nature Reviews Immunology**, v. 6, n. 11, p. 859- 68, 2006.
4. MELLORS, J. W. *et al*. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. **Annals of Internal Medicine**, v. 122, n. 8, p. 573-9, 1995.
5. WHO/UNAIDS. Joint United Nations Programme on HIV/AIDS. **Global AIDS update 2016**. 2016.
6. BRASIL. Departamento de Dst, Aids e Hepatites Virais. Ministério da Saúde (Org.). **Boletim Epidemiológico: HIV/AIDS**, p. 64, 2015.
7. BRASIL. Departamento de Dst, Aids e Hepatites Virais. Ministério da Saúde (Org.). **Indicadores e dados básicos do HIV/AIDS dos municípios brasileiros**. Disponível em: http://svs.aids.gov.br/aids/. Acessado em 2016.

8. PINTO, M. E.; STRUCHINER, C. J. Implications of HIV diversity for the HIV-1 pandemic. **Journal of Infection**, v. 3, n. 22, p. 473-484, 2013.

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

10. VALLARI, A. *et al*. Confirmation of putative HIV-1 group P in Cameroon. **Journal of Virology**, v. 85, n. 3, p.1403-1407, 2011.

11. PLANTIER, J. C. *et al*. A new human immunodeficiency virus derived from gorillas. **Nature Medicine**, v. 1, n. 15, p. 871-72, 2009.

12. KIJAK, G. H.; MCCUTCHAN, F. E. HIV diversity, molecular epidemiology, and the role of recombination. **Current Infectious Disease Reports**, v. 7, n. 6, p. 480-488, 2005.

13. AU, K. A.; WONG, J. J. l. Current trends of HIV recombination worldwide. **Infectious Disease Reports**, v. 2, n. 5, p.15-20, 2013.

14. GRÄF, T. **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**. 151 f. Tese (Doutorado) – Programa de Pós-graduação em Biotecnologia e Biociência, UFSC, Florianópolis, 2015.

15. BRIGIDO, L. F. *et al*. HIV type 1 subtype C and CB Pol recombinants prevail at the cities with the highest AIDS prevalence rate in Brazil. **AIDS Research and Human Retroviruses**, v. 23, n. 12, p. 1579-1586, 2007.

16. SANTOS, A. F. *et al*. Epidemiologic and evolutionary trends of HIV-1 CRF31_BC-related strains in southern Brazil. **Journal of Acquired Immune Deficiency Syndrome**, v. 45, n. 3, p. 328-233, 2007.

17. LOCATELI, D. *et al.* Molecular epidemiology of HIV-1 in Santa Catarina State confirms increases of subtype C in Southern Brazil. **Journal of Medical Virology**, v. 79, n. 10, p. 1455-1463, 2007.

18. DELATORRE, E. *at al*. Tracing the origin and northward dissemination dynamics of HIV-1 Subtype C in Brazil. **PlosOne**, v. 8, n. 9, e74072, 2013.

19. MEDEIROS, R. M. *et al*. Co-circulation HIV-1 subtypes B, C, and CRF31_BC in a drug-naive populations from southernmost Brazil: analysis of a primary resistance mutations. **Journal of Medical Virology**, v. 83, n. 10, p. 1-11, 2011.

20. ROBERTS, J. D.; BENBENEK, K.; KUNKEL, T. A. The accuracy of revese transcriptase from HIV-1. **Science**, v. 242, n. 4882, p. 1171-1173, 1988.

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

22. RAMBAUT, A. *et al*. The causes and consequences of HIV evolution. **Nature Reviews**, v. 5, n. 1, p. 52-61, 2004.

23. PEETERS, M. *et al*. Characterization of a highly replicative intergroup M/O human immunodeficiency virus type 1 recombinant isolated from a Cameroonian patient. **Journal of Virology**, v.7, n. 3, p. 7368-7375, 1999.

24. ARCHER, J. *et al*. Identifying the important HIV-1 recombination breakpoints. **PLOS Computational Biology**, v. 4, n. 9, e1000178, 2008.

25. SCHLUB, T. E. *et al*. Fifteen to twenty percentage of HIV substitution mutations are associated with recombination. **Journal of Virology**, v. 88, n. 7, p. 3837-3849, 2014.

26. ZEICHNER, S. L. The molecular biology of HIV. Insights into pathogenesis and targets for therapy. **Clinical Perinatology Journal**, v. 21, p. 39-73, 1994.

27. LEMEY, P.; SALEMI, M.; VANDAMME, A. The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. 2ª edição: Cambrige, **Cambridge University Press**, 2009..

28. VERLI, H. Bioinformática: da Biologia à Flexibilidade Molecular. 1ª edição: São Paulo, **SBBq**, 2014.

29. YANG, Z.; RANNALA, B. Molecular phylogenetics: principles and practice. **Nature Reviews Genetics**, v. 13, n. 5, p. 303-14, 2012.

30. BROWN, J. W.; KILMER, A. J. The state of bayesian phylogenetics: Bayes for the uninitiated. 1ª edição: Kingston, **Queen's University**, 2003.

31. FREEMAN, S.; HERRON J. C. Análise evolutiva. 4 edição: Porto Alegre, **Artmed**, 2009.

32. MARTIN, D. P. *et al*. RDP4: Detection and analysis of recombination patterns in virus genomes. **Virus Evolution**, v. 1, n. 1, p. 1-5, 2015.

33. LARSSON, A. AliView: a fast and lightweight alignment viewer and editor for large data sets. **Bioinformatics**, v. 30, n. 22, p. 3276-3278, 2014.

34. DRUMMOND, A. J.; RAMBAUT, A. BEAST: Bayesian evolutionary analysis by sampling trees. **BMC Evolutionary Biology**, v.7, n. 1 p. 1-8, 2007.

35. DIAZ. R. *et al*. The virological and immunological characteristics of the HIV-1-Infected population in Brazil: from Initial diagnosis to impact of antiretroviral use. **PlosOne**, v. 10, n. 10, e0139677, 2015.

36. CORNELISSEN, M. *et al.* pol gene diversity of five human immunodeficiency virus type 1 subtypes: for naturally occurring mutations that contribute to drug resistance, limited recombination patterns, and common ancestry for subtypes B and D. **Journal of Virology**, v. 71, n. 9, p. 6348-6358, 1997.

37. MARTINEZ-CAJAS, J. *et al*. Differences in resistance mutations among HIV-1 non-subtype B infections: a systematic review of evidence (1996-2008). **Journal of the International AIDS Society**, v. 12, n. 11, p. 1-11, 2009.

38. SIMONETTI, F. R. *et al.* Identification of a new HIV-1 BC circulating recombinant form (CRF60_BC) in italian young men having sex with men. **Infection, Genetics and Evolution**, v. 23, n. 1, p. 176-181, 2014.

39. TEIXEIRA, S. L. *et al.* HIV-1 infection among injection and ex-injection drug users from Rio de Janeiro, Brazil: prevalence, estimated incidence and genetic diversity. **Journal of Clinical Virology**, v. 31, n. 3, p. 221-226, 2004.

40. TOLEDO, P. V. *et al*. Genetic diversity of human immunodeficiency virus-1 isolates in Paraná, Brazil. **Brazilian Journal of Infectious Diseases**, v. 14, n. 3, p. 230-236, 2010.

41. GRÄF, T.; PINTO, A. R. The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. **Virology**, v. 433, n. 1, p. 170-178, 2013.

42. MENDONÇA, N. C. V. **História evolutiva do HIV-1 no Brasil**. 2010. 286 f. Tese (Doutorado) – Programa de Pós-Graduação em Biologia Molecular, UnB, Brasilia, 2010.

43. MACHADO, L. F. *et al*. Molecular epidemiology of HIV type 1 in northern Brazil: identification of subtypes C and D and the introduction of CRF02_AG in the Amazon region of Brazil. **AIDS Research and Human Retroviruses**, v. 25, n. 10, p. 961-966, 2009.

44. MCCUTCHAN, F. E. Understanding the genetic diversity of HIV-1. **AIDS**, v. 14, n. 3, p. 33-44, 2000.

45. LI, X. *et al*. Genome sequences of a novel HIV-1 circulating recombinant form (CRF61_BC) identified among heterosexuals in China. **Genome Announcements**, v. 1, n. 3, p. 1-13, 2013.

46. SU, L. *et al*. Characterization of a virtually full-length human immunodeficiency virus type 1 genome of a prevalent intersubtype (C/B') recombinant strain in China. **Journal of Virology**, v. 74, n. 23, p. 11367-11376, 2000.

47. MONTEIRO, J. P. **Análise da variabilidade genética do vírus da imunodeficiência humana (HIV): epidemiologia molecular no estado da Bahia**. 137 f. Dissertação (Mestrado) - Pós-graduação em Biotecnologia em Saúde e Medicina Investigativa, FIOCRUZ, Salvador, 2009.