

VICTOR FIGUEIREDO PIMENTEL

**ESTUDO DA DIVERSIDADE GENÉTICA DO HIV INTER E INTRA-HOSPEDEIRO
EM PACIENTES SOROCONVERTORES RECENTES**

Tese apresentada ao programa de pós-graduação em Microbiologia do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do Título de Doutor em Ciências.

São Paulo
2016

VICTOR FIGUEIREDO PIMENTEL

**ESTUDO DA DIVERSIDADE GENÉTICA DO HIV INTER E INTRA-HOSPEDEIRO
EM PACIENTES SOROCONVERTORES RECENTES**

Tese apresentada ao programa de pós-graduação em Microbiologia do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do Título de Doutor em Ciências.

Área de concentração: Microbiologia

Orientador: Prof^a. Dr^a. Charlotte Marianna Hársi Domotor

Co-orientador: Dr. Luís Fernando de Macedo Brígido

Versão original

São Paulo
2016

CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e informação Biomédica
do Instituto de Ciências Biomédicas da Universidade de São Paulo

Ficha Catalográfica elaborada pelo(a) autor(a)

Figueiredo Pimentel, Victor
Estudo da diversidade genética do HIV inter e
intra hospedeiro em pacientes sorosconvertores
recentes / Victor Figueiredo Pimentel; orientador
Charlotte Marianna Hársi Domotor; coorientador Luís
Fernando de Macedo Brígido. -- São Paulo, 2016.
118 p.

Tese (Doutorado) -- Universidade de São Paulo,
Instituto de Ciências Biomédicas.

1. HIV-1. 2. Redes de transmissão. 3.
Diversidade genética. I. Marianna Hársi Domotor,
Charlotte , orientador. II. Fernando de Macedo
Brígido, Luís, coorientador. III. Título.

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Candidato(a): Victor Figueiredo Pimentel

Título da Tese: Estudo da diversidade genética do HIV inter e intra-hospedeiro em pacientes soroconvertedores recentes

Orientador(a): Prof. Dra. Chalotte Marianna Hársi Domotor

A Comissão Julgadora dos trabalhos de Defesa da Tese de Doutorado, em sessão pública realizada a/...../....., considerou

Aprovado(a)

Reprovado(a)

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Examinador(a): Assinatura:
Nome:
Instituição:

Presidente: Assinatura:
Nome:
Instituição:



UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Cidade Universitária "Armando de Salles Oliveira"
Av. Prof. Lineu Prestes, 2415 - cep. 05508-000 São Paulo, SP - Brasil
Telefone : (55) (11) 3091.7733 telefax : (55) (11) 3091-8405
e-mail: cep@icb.usp.br

São Paulo, 15 de dezembro de 2014.

PARECER 1208/CEPSH

A Comissão de *Ética em Pesquisas em Seres Humanos* do ICB, nesta data, **APROVOU** o projeto intitulado: "Estudo da diversidade genética do HIV inter e intra-hospedeiro em pacientes soroconvertores recentes" dos pesquisadores Profa. Dra. Charlotte Marianna Hársi e Victor Figueiredo Pimentel.

Cabe aos pesquisadores elaborar e apresentar a este Comitê, relatórios anuais (parciais e final), de acordo com a Resolução nº 466/12, item II, II.19 e II.20, do Conselho Nacional de Saúde, conforme modelo constante no site: icb.usp.br.

Aos pesquisadores cabe também finalizar o processo junto à Plataforma Brasil quando do encerramento deste.

O primeiro relatório deverá ser encaminhado à Secretaria deste CEP em **15.12.2015**.

Atenciosamente,

Prof. Dr. PAOLO M.A. ZANOTTO
Coordenador da Comissão de Ética em
Pesquisas com Seres Humanos - ICB/USP

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Estudo da diversidade genética do HIV inter e intra-hospedeiro em pacientes sorosconvertidos recentes.

Pesquisador: Charlotte Marianna Hársi Domotor

Área Temática:

Versão: 1

CAAE: 39709314.4.0000.5467

Instituição Proponente: Universidade de São Paulo

Patrocinador Principal: Universidade de São Paulo

DADOS DO PARECER

Número do Parecer: 913.113

Data da Relatoria: 10/12/2014

Apresentação do Projeto:

Título da Pesquisa: Estudo da diversidade genética do HIV inter e intra-hospedeiro em pacientes sorosconvertidos recentes.

Pesquisador: Charlotte Marianna Hársi Domotor

Instituição Proponente: Universidade de São Paulo

Objetivo da Pesquisa:

Caracterizar a diversidade genética do HIV utilizando técnicas de sequenciamento de DNA, de forma a compreender os processos de diversificação do vírus mediados por recombinação e mutação.

Avaliação dos Riscos e Benefícios:

Não envolve nenhum risco para os indivíduos. Os benefícios são claros quanto a uma melhor compreensão dos mecanismos evolutivos de vírus e sua correlação com o desenvolvimento da doença e resistência aos antirretrovirais.

Comentários e Considerações sobre a Pesquisa:

O projeto tem como objetivo analisar a diversidade genética do HIV, utilizando diversas técnicas de sequenciamento de DNA. A amostra consiste de plasma de 60 pacientes, sorosconvertidos recentes, armazenados no biorrepositório do serviço de virologia do Instituto Adolfo Lutz. Cabe

Endereço: Av. Profº Lineu Prestes, 2415

Bairro: Cidade Universitária

CEP: 05.508-000

UF: SP

Município: SAO PAULO

Telefone: (11)3091-7733

Fax: (11)3091-8405

E-mail: cep@icb.usp.br

For my treasures: my mother and sister.

For my uncles.

AGRADECIMENTOS

Penso ser esta uma das piores etapas da jornada de um estudante de doutorado quando lhe é atribuído um espaço exíguo para manifestar o seu bem querer e sinceros agradecimentos às pessoas que tiveram papel importante nesta jornada. As relações interpessoais não é colocada em análise de regressão logística e sem dúvidas não é pelo valor de p que identificamos a significância das pessoas na nossa trajetória.

Como de praxi gostaria de agradecer aos meus pais Sr. Jonival Pimentel e Sra. Tânia Pimentel pelo apoio incondicional a mim atribuído, mantendo-se ao meu lado sempre como uma mola propulsora para que eu chegasse ao meu objetivo de maneira satisfatória.

Agradeço aos meus irmãos e sobrinha: Ana Paula Pimentel, Jonas Pimentel e Duda, que sempre me motivaram, entenderam as minhas faltas, as minhas escolhas de migrar por diferentes cidades (cada vez mais distantes), mas sempre me encorajaram e estimularam a estudar cada vez mais. E que ao seu modo sempre se orgulharam de mim e confiaram no meu trabalho. Obrigado pela confiança!

Aos meus tios, tias, primos e primas que vibraram comigo, desde a aprovação na prova de ingresso no doutorado, e sempre fizeram propaganda mais que positiva, diria até que exagerada ao meu respeito.

Aos amigos fora da academia científica (Thiago Farias, Adriano/Stênio, Marcos Alan e Leonardo Alves), que só querem o meu bem e me valorizam como ser humano. Muito, mais muito obrigado pela amizade e existência! Sei que vocês foram capazes de compreender a minha ausência no momento doloroso, porém gratificante da tese.

Agradeço aos meus orientadores: Dra. Charlotte Mariana Hársi Domotor. Aqui não caberiam palavras para agradecer, pelo voto de confiança quando tudo parecia *fuzzy*. Resumí-la a minha orientadora seria pouco!!!!!! Muito obrigado pelas conversas científicas, pelo conhecimento transferido e claro, pelo momento café sagrado das manhãs. Ao Dr. Luís Brígido pelas conversas acadêmicas mais que proveitosas, pelo incentivo e por estímulo na forma bem sutil (rs).

Aos meus orientadores e colaboradores internacionais Dra. Annemiek Vandamme e Dra. Ana Abecasis pelo tratamento impecável, pela forma como me aceitaram e me integraram no seu grupo de pesquisa, além claro do “Lab meeting” quando no restaurante à margem do rio. Não poderia faltar espaço para minha mais nova família portuguesa (Martinha, Paty, Dani, amo vocês!!!).

Aos colegas de Laboratório (IAL) João Leandro, Luana Portes, Jaqueline (Jack), Paula Morena, Gabriela Bastos.

A minha companheira de trabalho incansável Stella Melo que chegou pra ficar no laboratório e acabou fazendo parte da minha vida. Obrigado chuchu.

Aos amigos da ala da virologia que SEMPRE foram singulares comigo, com muita paciência mesmo quando eu pegava pesado nas brincadeiras.

Não poderia deixar de citar os amigos mais próximos da virologia. Em escala de essencialidade vem a mais recente doutora do grupo Danielle Ferreira (Dani) que com seu jeito ímpar me ensinou a ser uma pessoa melhor, sem ela este trabalho não teria acontecido. Muito obrigado por tudo!

À minha querida Susane , esta colombiana que chegou e tomou conta do meu coração, sempre com seu jeito doce a me ensinar o melhor atalho para as minhas análises computacionais. Amo-te chuchu!

À minha querida baby (Suellen), que chegou meio assim num dia de outono e cravou sua semente da amizade. Muito obrigado pela companhia do cafezinho e pelos papos acadêmicos no mais alto nível. Meu pós doutorado versará em REC, com certeza.

Não poderiam faltar agradecimentos aos amigos do laboratório do prof. Dr. Enrique Boccardo (Bruninha, ao amigo Wallason, Van, Aline, Sumara .ops desculpem se faltou alguém).

Last but not least, ao CNPq e à CAPES, que fomentaram a minha bolsa de doutorado no Brasil e durante o meu doutorado sanduíche. Sem estes auxílios não conseguiria desenvolver a minha pesquisa.

*“The greatest enemy of knowledge
is not ignorance, it is the illusion of knowledge”*

Stephen Hawking

RESUMO

Pimentel, VF. Estudo da diversidade genética do HIV inter e intra-hospedeiro em pacientes sorosconvertedores recentes. [tese (Doutorado em Microbiologia)]. São Paulo: Instituto de Ciências Biomédicas, Universidade de São Paulo; 2016.

A epidemia de HIV-1 está mudando em São Paulo, com uma crescente preocupação no número de jovens de homens que fazem sexo com homens (HSH). A epidemia é dominada pelos subtipos B, F1, porém tem aumentado a prevalência do subtipo C. O objetivo do estudo foi identificar e caracterizar as redes de transmissão em São Paulo, comparando entre os subtipos epidêmicos, fatores de risco e exposição à terapia antirretroviral.

Nós compilamos dados clínicos, epidemiológicos e as sequências virais dos indivíduos acompanhados no Instituto Adolfo Lutz de Jan 2004 a fev de 2015. Após o controle de qualidade, 2.260 sequências parciais do gene *pol* foram subtipadas previamente como B, C e F1 e então incluídas no estudo. 2.107 sequências únicas controle foram selecionadas pelo blast contra as nossas sequências no banco de dados de Los Alamos e banco de dados de Portugal. Todas as sequências foram investigadas para a identificação de mutação de resistência (TDRM). Árvores filogenéticas foram construídas em RaxML para cada subtipo. Nós usamos Cluster Picker para análise da dinâmica de transmissão de acordo com os parâmetros de distância genética e *bootstrap*. Análises estatísticas foram utilizadas para identificar possíveis correlações dos clusters.

414 (18,3%) da nossa população foi incluída em cluster (2-12 indivíduos). A taxa de cluster não diferiu entre B e F1, no entanto as sequências do subtipo C se agruparam duas vezes mais ($p < 0.001$). Mais clusters foram identificados entre a população HSH quando comparada a de heterossexuais independente do subtipo ($p < 0.001$). A taxa de TDRM foi maior em cluster que fora de cluster nos subtipos B e F1 ($p < 0.001$ e $p = 0.009$), respectivamente. Apesar do alto número de clusters no subtipo C verificamos baixa prevalência de TDRM. Os indivíduos que estavam em clusters eram 5 anos mais jovens que os de fora de cluster nos subtipos B e C ($p < 0.001$ e 0.02, respectivamente).

Nossos resultados indicam que, independente do subtipo, a epidemia de HIV-1 em São Paulo é auto-sustentada pelos pacientes HSH virgens de tratamento. O subtipo C apresentou maior proporção de clusters. No entanto, mais análises são necessárias para clarificar se isto implica em alta taxa de transmissão do subtipo em São Paulo.

Palavras-chave: HIV-1. Redes de transmissão. Diversidade genética.

ABSTRACT

Pimentel, VF. Study of genetic diversity of HIV inter and intra-host in recent infected patients. [Ph.D thesis (Microbiology)]. São Paulo: Instituto de Ciências Biomédicas, Universidade de São Paulo; 2016.

The HIV-1 epidemic in Sao Paulo is changing, with a worrying increase in the number of new infections in young men and in the Men who have Sex with Men (MSM). The epidemic is dominated by subtypes B and F1, but there's been a recent increase in the prevalence of subtype C. The aim of this study was to identify and characterize HIV-1 transmission networks in Sao Paulo, comparing across subtype epidemics, risk factors, and HAART exposure therapy.

We gathered epidemiological, clinical and viral sequence data from HIV-1 infected individuals followed in Adolfo Lutz Institute/SP-BR from Jan 2004 through Feb 2015. After quality control, 2,260 sequences of the partial *pol* gene were subtyped as previously reported as B, C and F1 were included on this study. 2,107 unique background control sequences were selected by blasting our sequences against Los Alamos database and the Portuguese HIV-1 database. The dataset was screened for Drug Resistance Mutations to identify transmitted drug resistance mutations (TDRM). Maximum-Likelihood phylogenetic trees were built in RaxML for each subtype separately. We used Cluster Picker to analyze transmission dynamics according the thresholds: genetic distance (0.06) and bootstrap over 90%. Statistical analyzes were performed to identify possible correlates of clustering. Categorical variables were compared using Chi-square or Fisher's exact test and quantitative variables were analyzed using Mann-Whitney test (SPSS).

414 (18,3%) of our population were included in clusters (range: 2-12 individuals). The rate of clustering did not differ between subtypes B and F, however more subtype C sequences (40%) significantly clustered together ($p < 0.001$). Furthermore, more clusters were significantly found in the MSM group when compared to heterosexuals for all subtypes, ($p < 0.001$). Also, drug naïve patients were more likely to be in clusters when compared to treated patients in all subtypes ($p < 0.001$). TDRM were more prevalent in clustering than in non-clustering in subtypes B ($p < 0.001$) and F1 ($p = 0.009$), and more related to MSM group. Despite the higher number of clusters, subtype C presented a lower prevalence of TDRM, although without significant difference between cluster and non-cluster. Clustering individuals were also 5 years younger than non-clustering individuals for subtypes B and C ($p < 0.001$ and 0.02, respectively).

Our results indicate that, regardless of subtype, the HIV-1 epidemic in Sao Paulo is self-sustained by treatment-naïve MSM patients, who transmit the infection before starting treatment. Subtype C presents a higher proportion of patients that cluster together. However, further analyses are necessary to clarify whether this implies a higher transmission rate of this subtype in Sao Paulo.

Keywords: HIV-1. Transmission network. Genetic Diversity.

LISTA DE FIGURAS

FIGURA 1 - ESTIMATIVA DO NÚMERO DE ADULTOS E CRIANÇAS PORTADORAS DO HIV/AIDS, 2016	19
FIGURA 2 - REPRESENTAÇÃO ESQUEMÁTICA DA PARTÍCULA VIRAL DO HIV..	21
FIGURA 3 - REPRESENTAÇÃO ESQUEMÁTICA DA ESTRUTURA LINEAR PROVIRAL DE HIV-1	24
FIGURA 4 - REPRESENTAÇÃO ESQUEMÁTICA DO CICLO REPLICATIVO DO HIV-1.....	25
FIGURA 5 - HISTÓRIA NATURAL DA INFECÇÃO PELO HIV-1.....	28
FIGURA 6 - CLASSIFICAÇÃO ATUAL DO HIV.	30
FIGURA 7 - DISTRIBUIÇÃO GLOBAL DOS SUBTIPOS E RECOMBINANTES DE HIV-1 DO GRUPO M.....	32
FIGURA 8 - DISTRIBUIÇÃO DOS SUBTIPOS E FORMAS RECOMBINANTES CIRCULANTES DO HIV-1 NO BRASIL..	34
FIGURA 9 - ESQUEMA DE FORMAÇÃO DAS QUASI-ESPÉCIES.....	35
FIGURA 10 - MODELO DE TRANSMISSÃO DO HIV-1.....	36
FIGURA 11 - MAPA DE CALOR REPRESENTANDO A ORIGEM E QUANTIDADE DE AMOSTRAS POR LOCALIDADE.	39
FIGURA 12 - RESUMO METODOLÓGICO DAS ETAPAS DO SEQUENCIAMENTO DE NOVA GERAÇÃO.	46
FIGURA 13 - ANÁLISE QUE REPRESENTA A MEDIANA DA IDADE ENTRE OS DIFERENTES SUBTIPOS B, C E F1 DE HIV-1 LEVANDO EM CONSIDERAÇÃO A TERAPIA.	51
FIGURA 14 – ANÁLISE REPRESENTATIVA DA FREQUÊNCIA DAS CATEGORIA DE EXPOSIÇÃO AO RISCO NOS SUBTIPOS.....	51
FIGURA 15 - ANÁLISE QUE REPRESENTA A MEDIANA DA CONTAGEM DE LINFÓCITOS T CD4 ⁺ NOS GRUPOS <i>NAIVE</i> E TRATADOS	52
FIGURA 16 - ANÁLISE QUE REPRESENTA A MEDIANA DA CARGA VIRAL EM LOG/mL ENTRE OS INDIVÍDUOS QUE OS GRUPOS <i>NAIVE</i> E TRATADO NOS SUBTIPOS B, C E F1.	52
FIGURA 17 - ANÁLISE QUE REPRESENTA A MEDIANA DA IDADE ENTRE OS DIFERENTES SUBTIPOS B, C E F1 DE HIV-1 LEVANDO EM CONSIDERAÇÃO OS DOIS GRUPOS HSH E NÃO_HSH.	53
FIGURA 18 - ANÁLISE QUE REPRESENTA A FREQUÊNCIA DA RESISTÊNCIA A DROGAS TRANSMITIDA (TDR) ENTRE OS GRUPOS DE HSH E NÃO-HSH NOS SUBTIPOS.....	54
FIGURA 19 - ANÁLISE QUE REPRESENTA A CONTAGEM DE LINFÓCITOS T CD4 ⁺ NOS GRUPOS HSH E NÃO-HSH	54
FIGURA 20 - ANÁLISE FILOGENÉTICA CORRESPONDENTE AO GENE <i>POL</i> DE 3.708 SEQUENCIAS DO SUBTIPO B DE HIV-1.	57
FIGURA 21 - ANÁLISE FILOGENÉTICA CORRESPONDENTE AO GENE <i>POL</i> DE 300 SEQUENCIAS DO SUBTIPO C DE HIV-1.	58
FIGURA 22 - ANÁLISE FILOGENÉTICA CORRESPONDENTE AO GENE <i>POL</i> DE 359 SEQUENCIAS DO SUBTIPO F1 DE HIV-1.....	59
FIGURA 23 - ANÁLISE QUE REPRESENTA A MEDIANA DA IDADE ENTRE OS INDIVÍDUOS QUE ESTAVAM EM CLUSTERS DE TRANSMISSÃO E FORA DE CLUSTERS.....	60

FIGURA 24 - ANÁLISE QUE REPRESENTA A FREQUÊNCIA DOS INDIVÍDUOS QUE ESTÃO EM CLUSTER E FORA DE CLUSTER NOS GRUPOS DE PACIENTES <i>NAIVES</i> E TRATADOS NOS SUBTIPOS B, C E F1.....	60
FIGURA 25 - ANÁLISE QUE REPRESENTA A FREQUÊNCIA DOS INDIVÍDUOS QUE ESTÃO EM CLUSTER E FORA DE CLUSTERS NOS GRUPOS DE HSH E NÃO-HSH.....	61
FIGURA 26 - ANÁLISE QUE REPRESENTA A FREQUÊNCIA DA RESISTÊNCIA A DROGAS TRANSMITIDA (TDR) ENTRE OS INDIVÍDUOS QUE ESTÃO EM CLUSTER E FORA DE CLUSTER	61
FIGURA 27 - ANÁLISE QUE REPRESENTA A MEDIANA DA CONTAGEM DE LINFÓCITOS T CD4 ⁺ ENTRE OS INDIVÍDUOS QUE ESTÃO EM CLUSTER E FORA DELES.....	62
FIGURA 28 - ANÁLISE QUE REPRESENTA A MEDIANA DA CARGA VIRAL EM LOG/M _L ENTRE OS INDIVÍDUOS QUE ESTÃO EM CLUSTER E FORA DELES.	62
FIGURA 29 - TIPOS DE CLUSTERS GEOGRÁFICOS IDENTIFICADOS EM PACIENTES DO ESTADO DE SÃO PAULO.	65
FIGURA 30 - ÁRVORE FILOGENÉTICA DO GENE <i>POL</i> ENVOLVENDO OS 130 HAPLÓTIPOS RECONSTRUÍDOS.....	73
FIGURA 31 - ANÁLISE QUE REPRESETA A MEDIANA DA VARIABILIDADE GENÉTICA EM NUCLEOTÍDEOS E AMINOÁCIDOS	74
FIGURA 32 - ANÁLISE QUE REPRESENTA A FREQUÊNCIA DOS CÓDONS QUE TERMINAM COM A/U E C/G.	76

LISTA DE TABELAS

TABELA 1 - ESTRATÉGIA DE AMPLIFICAÇÃO DA REGIÃO <i>POL</i> DO HIV-1	40
TABELA 2 - INICIADORES A SEREM UTILIZADOS NA REAÇÃO DE SEQUENCIAMENTO SANGER DA REGIÃO <i>POL</i>	41
TABELA 3 – CARACTERÍSTICAS DOS PACIENTES ENVOLVIDOS NO ESTUDO.	50
TABELA 4 – CARACTERÍSTICAS DOS PACIENTES ESTUDADOS PELO SEQUENCIAMENTO DE NOVA GERAÇÃO	67
TABELA 5 – SUMÁRIO DA COMPOSIÇÃO DOS <i>READS</i> OBTIDOS APÓS SEQUENCIAMENTO DE NOVA GERAÇÃO.....	68
TABELA 6 – LISTA DE MUTAÇÕES DETECTADAS PELAS DUAS PLATAFORMAS DE SEQUENCIAMENTO GENÉTICO.....	70
TABELA 7 – MÉTRICA DE DIVERSIDADE GENÉTICA INTRA-HOSPEDEIRO.....	74

LISTA DE ABREVIATURAS

AIDS	Acquired Immunodeficiency Syndrome
cDNA	DNA complementar
CPR	Ferramenta de calibração da população com resistência
CRF	Forma recombinante circulante
dN	Sítio não sinônimo
dNTP	Desoxirribonucleotídeos trifosfatados
dS	Sítio sinônimo
FEL	Fixed effects likelihood
FUBAR	Fast unconstrained Bayesian approximation
GTR	Generalized time reversible model
HAART	Terapia antirretroviral altamente ativa
HIV-1	Vírus da imunodeficiência humana tipo 1
HET	Heterossexual
HSH	Homens que fazem sexo com homens
ML	Maximum Likelihood
NGS	Sequenciamento de nova geração
NRT	Inibidores da transcriptase reversa análogo de nucleosídeos
NNRT	Inibidores da transcriptase reversa não análogo de nucleosídeos
NI	Inibidor da integrase
PCR	Reação em cadeia da polimerase
RNA	Ácido ribonucléico
RSCU	Relativo uso de códons sinônimos
SLAC	Single likelihood ancestor counting
UDI	Usuário de drogas injetáveis
VT	Transmissão vertical

SUMÁRIO

1 INTRODUÇÃO	18
1.1 DADOS EPIDEMIOLÓGICOS.....	18
1.2 AGENTE ETIOLÓGICO DA AIDS	20
1.3 PARTÍCULA VIRAL	21
1.4 ESTRUTURA GENÔMICA	22
1.5 CICLO REPLICATIVO DO HIV	24
1.6 TERAPIA ANTIRRETROVIRAL	26
1.7 HISTÓRIA NATURAL DA INFECÇÃO PELO HIV-1	27
1.8 DIVERSIDADE GENÉTICA DO HIV E EPIDEMIOLOGIA MOLECULAR	29
1.9 EPIDEMIOLOGIA MOLECULAR DO HIV-1 NO BRASIL.....	32
1.10 TEORIA DAS QUASIESPÉCIES DO HIV.....	34
2.1 OBJETIVO GERAL.....	37
2.2 OBJETIVOS ESPECÍFICOS	37
3 MÉTODOS	38
3.1 CASUÍSTICA.....	38
3.2 EXTRAÇÃO DE RNA.....	39
3.3 AMPLIFICAÇÃO POR NESTED PCR DA REGIÃO POL.....	39
3.4 PURIFICAÇÃO E QUANTIFICAÇÃO DOS PRODUTOS DE PCR PARA SEQUENCIAMENTO	41
3.5 REAÇÃO DE SEQUENCIAMENTO AUTOMATIZADO	41
3.6 ANÁLISE E EDIÇÃO DE SEQUÊNCIAS OBTIDAS PELO MÉTODO SANGER.....	42
3.7 SUBTIPAGEM VIRAL E ANÁLISE FILOGENÉTICA	42
3.8 DINÂMICA GEOGRÁFICA DOS CLUSTERS INTER E INTRA ESTADO DE SÃO PAULO.....	43
3.9 ANÁLISE DE RESISTÊNCIA ANTIRRETROVIRAL	44
3.10 ANÁLISES ESTATÍSTICAS	44
3.11 SEQUENCIAMENTO LARGAMENTE PARALELIZADO (MiSEQ)	44
3.12 ANÁLISE DOS DADOS A PARTIR DO SEQUENCIAMENTO DE NOVA GERAÇÃO.....	46
3.13 RECONSTRUÇÃO HAPLOTÍPICA.....	47
3.14 USO DE CÓDON DIFERENCIAL	47
3.15 PRESSÃO DE SELEÇÃO	48
3.16 MÉTRICA DE DIVERSIDADE.....	48
4 RESULTADOS	49
4.1 DADOS CLÍNICOS E EPIDEMIOLÓGICOS.....	49
4.2 FATOR DE RISCO	53
4.3 INFERÊNCIA DOS CLUSTERS DE TRANSMISSÃO.....	55
4.4 DINÂMICA GEOGRÁFICA DOS CLUSTERS	63
4.5 ANÁLISE DO SEQUENCIAMENTO DE NOVA GERAÇÃO	66
4.6 ANÁLISE DE RESISTÊNCIA ASSOCIADA A DROGAS.....	69
4.7 RECONSTRUÇÃO HAPLOTÍPICA E MÉTRICA DE DIVERSIDADE.....	71
4.8 USO DE CÓDONS DIFERENCIAIS	75
4.9 SELEÇÃO NATURAL.....	76
5 DISCUSSÃO	78
5.1 DADOS CLÍNICOS E EPIDEMIOLÓGICOS E GRUPO DE RISCO	78
5.2 REDES DE TRANSMISSÃO	79
5.3 DINÂMICA DAS REDES DE TRANSMISSÃO.....	81
5.4 DETECÇÃO DE MUTAÇÃO DE RESISTÊNCIA E MÉTRICA DE DIVERSIDADE PELO NGS	83
5.5 USO DE CÓDONS DIFERENCIAL	87
5.6 PRESSÃO DE SELEÇÃO	89
6 CONCLUSÕES	91

REFERÊNCIAS92

APÊNDICE A - QUANTIFICAÇÃO DAS BIBLIOTECAS PELO PCR EM TEMPO REAL. 107
APÊNDICE B - ANÁLISE DO PRODUTO DAS BIBLIOTECAS REVELADO PELO BIOANALYZER. 108
APÊNDICE C- DESCRIÇÃO DAS VARIÁVEIS SEGUNDO O TRATAMENTO (NAIVE OU TRATADO)..... 109
APÊNDICE D - DESCRIÇÃO DAS VARIÁVEIS SEGUNDO OS PACIENTES QUE ESTÃO EM CLUSTERS 110
APÊNDICE E - DESCRIÇÃO DAS VARIÁVEIS SEGUNDO O GRUPO DE RISCO 111
APÊNDICE F - PADRÃO DE USO RELATIVO DE CÓDONS SINÔNIMOS (RSCU) 112
APÊNDICE G - A TABELA SUMARIZA OS VALORES OBTIDOS DE CAI (ÍNDICE DE ADAPTAÇÃO DO CÓDON).. 114

1 INTRODUÇÃO

1.1 DADOS EPIDEMIOLÓGICOS

O vírus da Imunodeficiência humana (HIV) é o agente etiológico da Síndrome da Imunodeficiência Adquirida (Aids) [1]. Passadas três décadas da identificação da Aids, a pandemia causada por este vírus tornou-se uma das mais devastadoras da história da humanidade e está mais estreitamente relacionada a populações economicamente ativas, o que leva a grandes problemas socioeconômicos em países emergentes [2].

Com base nos dados da UNAIDS estima-se que vivam no mundo cerca de 36,7 milhões de adultos e crianças infectados pelo HIV (Figura 1), deste conjunto, 2,1 milhões de novos casos foram descritos em 2015. O número de pessoas que morreram em consequência da Aids neste ano foi de 1,1 milhões de pessoas no mundo, o que aponta um decréscimo quando comparados aos dados de 2009, onde se estimou 1,8 milhões de mortes relacionadas à Aids. Este fato associa-se à maior disponibilidade de terapia antirretroviral e ao melhor suporte às pessoas infectadas com o HIV em países de média e baixa renda [3]. Com a intensificação dos programas de prevenção da Aids e o comportamento mais consciente por parte da população, a prevalência e incidência dos casos de HIV/AIDS nos últimos anos tenderam a se estabilizar, não obstante, um aumento do número de casos reflexo da alta densidade demográfica mundial e da melhora da expectativa de vida por em função da terapia antirretroviral também tem sido observado [3].

Ao longo de mais de três décadas de convívio com a infecção pelo HIV, constataram-se mudanças importantes no perfil epidemiológico da Aids. Nos primeiros anos da epidemia a maioria dos casos estava relacionada com determinados comportamentos/situações de risco tais como: homo/bissexuais masculinos, hemofílicos/receptores de sangue e usuários de drogas injetáveis.

Com o passar do tempo a disseminação da infecção pelo HIV alcançou outros segmentos da população através da transmissão heterossexual, a qual tem sido responsável pelo aumento no número de casos de Aids em mulheres e conseqüentemente em crianças [3]. Com base nos dados publicados pelo último boletim epidemiológico da UNAIDS, salienta-se que a África Subsaariana, epicentro da epidemia de Aids, concentra o maior número de casos de adultos e crianças infectados pelo HIV: cerca de 19,0 milhões, dos quais 960.000 são casos de pessoas recentemente infectadas, dentre eles, adultos e crianças [3].

Do total de pessoas infectadas, mais de 50% são mulheres, o que acarreta no número acentuado de crianças que contraíram o vírus por transmissão vertical nesta região.

Do total estimado de 36,7 milhões de pessoas vivendo com HIV no mundo, cerca de 2 milhões encontram-se nas Américas do Sul e Central (Figura 1), dos quais 100.000 adultos e crianças que foram recentemente infectados. Aproximadamente 50.000 pessoas morreram devido a AIDS em 2015 na América Latina [3].

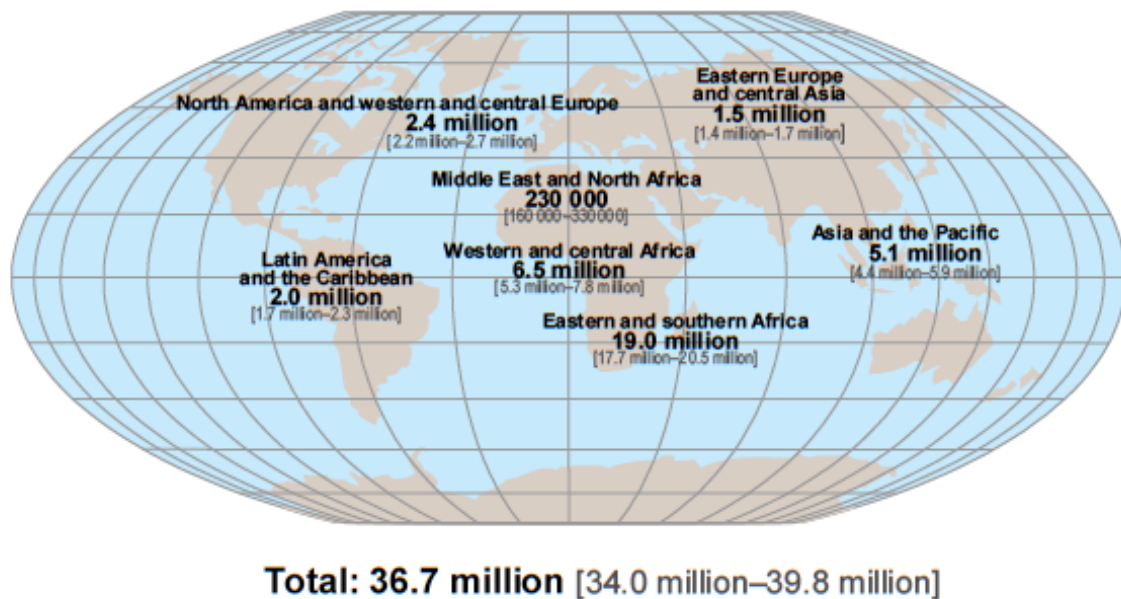


Figura 1 - Estimativa do número de adultos e crianças portadoras do HIV/AIDS, 2016. Fonte: [1]

Do total de casos notificados de pessoas portadoras do HIV na América do Sul, grande parte concentra-se no Brasil, devido a sua dimensão territorial e ao contingente populacional.

As políticas de prevenção à Aids, tais como, as campanhas de conscientização da população quanto aos riscos de contaminação, a distribuição gratuita de preservativos, seringas e agulhas, assim como as políticas de tratamento, que torna universal o acesso aos antirretrovirais, reduzindo assim diretamente as taxas de transmissão vertical e indiretamente as taxas de transmissão em consequência de um maior controle da carga viral, desde o início dos anos noventa contribuem para a contenção da epidemia no Brasil.

De acordo com o último Boletim epidemiológico, com dados da epidemia brasileira de 1980 a junho de 2015, foram notificados 798.366 casos acumulados de Aids em todo o país.

A taxa de incidência vem apresentando tendência à estabilização, em patamares elevados nos últimos anos e, em 2014, atingiu 20,5 por 100.000 habitantes. Contudo, os dados

indicam expressivas diferenças regionais, com declínio da taxa de incidência nas regiões Sudeste e Centro-Oeste, e aumento nas regiões Norte, Nordeste e Sul no período de 2005 a 2014. Na série histórica de 1980 até junho de 2015, foram identificados no Brasil 519.183 casos de Aids. A razão entre o número de casos segundo o sexo, masculino : feminino (M:F) sofreu significativa redução do início da epidemia até os dias atuais. Em 1986, a razão M:F era de 15,1:1 e, a partir de 2002, estabilizou-se em 1,5:1 [4]. No entanto, a partir de 2009, observou-se uma redução no número de casos em mulheres e um aumento entre os homens, refletindo na razão de sexos que passou a ser de 19:10 [5].

No Estado de São Paulo, onde se concentra a maior parte da epidemia nacional, (33% dos casos do país), foram registrados 228.698 casos de Aids até junho de 2013, e 103.267 óbitos por Aids. Estima-se, aproximadamente, 132.000 o número de portadores do HIV com idade entre 15-49 anos no estado [6]. Nestes 30 anos, a epidemia de HIV/Aids manteve-se concentrada nas grandes áreas metropolitanas, em adultos jovens, em especial nos grupos mais vulneráveis à infecção, tais como: Homens que fazem sexo com homens (HSH), profissionais do sexo e usuários de drogas ilícitas [6].

1.2 AGENTE ETIOLOGICO DA AIDS

O HIV é um retrovírus, que pertence à família *Retroviridae*, vírus com genoma RNA que possui seu genoma proviral reversamente transcrito pela transcriptase reversa, e ao gênero *Lentivirus*, que tem como importante característica o longo período de incubação entre a infecção e o início das manifestações clínicas, além de danos progressivos no sistema imune de seu hospedeiro.

Tanto o HIV-1 quanto o HIV-2 estão relacionados filogeneticamente a certos vírus de macacos africanos, isto é, aos Vírus da Imunodeficiência Símia (SIVs), com os quais apresentam uma similaridade na organização do genoma viral, coincidências geográficas e rotas de transmissão plausíveis reforçam a transmissão zoonótica destes lentivírus de primatas não humanos para os humanos [7].

Na tentativa de compreender melhor a origem do HIV-1, Gao e colaboradores em 1999 [7] analisaram a filogenia dos isolados de SIV conhecidos e identificaram duas linhagens principais e divergentes que infectam duas subespécies de chimpanzés, uma da África Central: *Pan troglodytes troglodytes*, e outra da África Oriental: *Pan troglodytes schweinfurthii*. Apenas a linhagem que infecta *Pan troglodytes troglodytes* mostrou-se estar

relacionada com o HIV-1 e parece ter originado as linhagens do grupo M e N. Outra forte evidência é que esta variante foi encontrada na mesma área geográfica da África onde podem ser encontrados todos os grupos do HIV-1 e seus subtipos. Assim, o chimpanzé *Pan troglodytes troglodytes* é considerado o reservatório primário do HIV-1[8].

1.3 PARTICULA VIRAL

A partícula viral madura e infecciosa do HIV (Figura 2) tem simetria icosaédrica e diâmetro de aproximadamente 125 nm. Externamente o HIV apresenta um envoltório de natureza bioquímica fosfolipídica derivada da célula hospedeira [9], onde são ancoradas as glicoproteínas de superfície e transmembranar (*gp120* e *gp41*) [9]. Internamente, uma matriz protéica (p17) envolve o capsídeo viral (p24). É nesta estrutura que encontramos o genoma diplóide do HIV que consiste de duas fitas simples de RNA similares, ambos de polaridade positiva, que interagem com as três enzimas virais (protease, transcriptase reversa e integrase) [10,11].

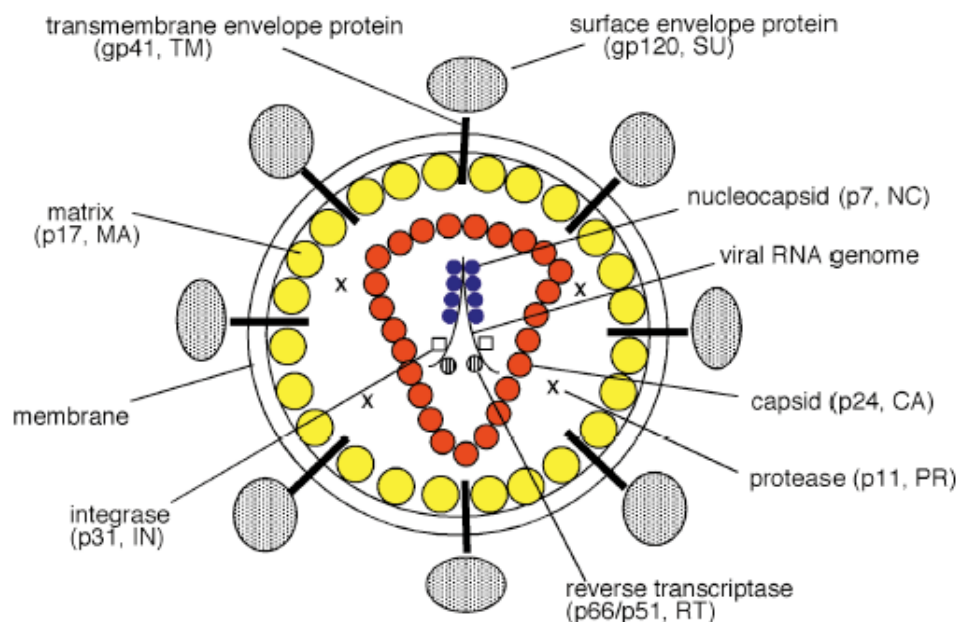


Figura 2 - Representação esquemática da partícula viral do HIV. Em destaque as proteínas estruturais, enzimas virais e genoma de RNA fita simples. Fonte [12].

1.4 ESTRUTURA GENÔMICA

Em adição aos genes estruturais (*gag*, *pol* e *env*), o HIV contém 4 genes acessórios (*vif*, *vpr*, *vpu* e *nef*) e dois genes regulatórios (*tat* e *rev*), cujos produtos são responsáveis por estabelecer sofisticadas interações entre o vírus e o hospedeiro. [13] (figura 3).

Como característica do DNA retroviral encontram-se nas porções terminais as LTRs (*Long Terminal Repeats*), que nada mais são que várias centenas de nucleotídeos (300-1800pb) composta de U3-R-U5 (5'- 3') Todos os sinais para expressão dos genes são encontrados nas LTR (promotor, iniciação de transcrição, terminador de transcrição e sinal de poliadenilação). Alguns sinais regulatórios da transcrição estão contidos na região U3 da 5' LTR e a região TATAA *Box* localiza-se a cerca de 25pb a partir do início da região R. A LTR 3' normalmente não é funcional como promotor, embora tenha a mesma sequência que a 5' LTR. Em vez disso, age na terminação da transcrição e poliadenilação, entretanto, quando a integridade da 5' LTR é comprometida, a 3' LTR funcione como promotor [14].

O primeiro gene localizado no terminal 5' do genoma do HIV é o gene *gag*. As proteínas formadas a partir de *gag*, são sintetizadas na forma de um precursor poliprotéico, que foi nomeado com base na massa molecular deste polipeptídeo, Pr55^{gag}. Logo após o brotamento da partícula viral, a protease do HIV-1 cliva Pr55^{Gag} em quatro proteínas de Gag maduras p17 (proteína de matriz, MA), p24 (proteína de capsídeo CA), p7 (proteína de nucleocapsídeo, NC) e p6, além de dois pepitídeos espaçadores p1 e p2 [12]. Durante a tradução de *gag* pode ocorrer um deslizamento do ribossomo na região p1, graças à presença de uma região rica em "U", levando à mudança na fase de leitura [15], gerando um precursor maior P160^{gag-pol}, este quando clivado proteoliticamente, gera as proteínas de *gag* e *pol*.

O *pol* é o segundo gene na organização genômica do HIV sendo o responsável por codificar as três enzimas do HIV: protease (PR), transcriptase reversa (RT) e integrase (IN), tendo os seus produtos derivados do precursor P160^{gag-pol} [14].

Das enzimas oriundas do gene da polimerase, a protease do HIV-1, é a responsável pelo processamento pós traducional das poliproteínas viral Gag (p24, p17 e p7) e Gag-pol (funcionalidade das enzimas RT e IN), transformando o vírus maduro [14–16]. A transcriptase reversa (RT) do HIV é capaz de converter o seu genoma de fita simples de RNA em uma dupla fita de DNA viral. A RT do HIV apresenta três atividades fundamentais: a DNA polimerase dependente de RNA, sintetizando a fita de DNA negativo, a atividade

RNAse H, utilizada para degradação do RNA viral, e a atividade de DNA polimerase dependente de DNA, gerando a fita positiva de DNA [17]. A integrase é enzima necessária para que o DNA viral seja inserido no genoma da célula hospedeira humana, processo indispensável para que o vírus possa utilizar o maquinário celular para produzir novos vírions. [14].

O gene *env* é o responsável por codificar as proteínas do envelope viral, o qual tem o seu precursor gp160, clivado por enzimas celulares e em *gp120* e *gp41*, proteínas de superfície e transmembranar, respectivamente. É na *gp120* que são encontrados os sítios de ligação para os receptores celulares e, portanto, importantes domínios a serem neutralizados com a finalidade de se interromper o ciclo viral [16]. Estas duas proteínas são fundamentais nas etapas de adsorção e fusão do vírus à célula hospedeira [18]. A partir do sequenciamento genético do gene *env* foi possível analisar e prever as estruturas secundárias, as quais definem importantes epítomos com papel de interação na resposta imune humoral e celular [19]

O gene *tat* (transativador da transcrição) é uma proteína multifuncional que contribui de maneira crítica na replicação do HIV-1. Tat é uma proteína transativadora robusta que induz uma variedade de efeitos, alterando os níveis de expressão de genes celulares e viral. Entre as funções de Tat destacam-se: super-expressão de citocinas, co-receptor CCR5 e receptor (CD25) de inter-leucina-2 [20–22], bem como a supressão do gene de MHC de classe I [23]

O gene que se sobrepõem a *tat* com fase +1 de leitura e codificando uma proteína de 116 aminoácidos é chamada de Rev e tem a função de regulação da expressão de proteínas do HIV pelo controle na exportação e de mRNA do núcleo para o citoplasma [14,24].

O gene *vif* (fator de infectividade do vírus) codifica uma proteína citoplasmática de 23 kD cuja função é aumentar significativamente a infecciosidade do vírion. A proteína é encontrada no interior das células infectadas pelo HIV, e a mesma atua interferindo no papel de uma proteína de restrição celular chamada APOBEC3G. Vif induz a célula a degradar a proteína APOBEC3G. Estudos in vitro mostraram que a construção de partículas virais sem o gene *vif* (delta *vif*) comprometiam o potencial desses vírus em infectar novas células [25–27].

O gene *vpr* codifica a proteína viral R com tamanho aproximado de 15 kD. Dentre as inúmeras funções conhecidas, esta acelera a expressão de proteínas do HIV, carrega "os sinais de localização nuclear" (seqüências de aminoácidos nas proteínas que são

reconhecidas pela maquinaria celular, indicando que ela deve ser transportado para o núcleo) e, em certo sentido, imita o comportamento de uma proteína chamada *beta* importina [28], Interrompe o ciclo celular na fase G2/M, que parece ser o cenário confortável para a replicação viral [29].

O gene *vpu* que codifica a proteína integral de membrana (proteína viral U) que tem como principais funções, auxiliar na montagem de novas partículas virais, no brotamento da partícula viral da célula hospedeira e na degradação das moléculas CD4 que ficam na superfície das células infectadas diminuindo o risco de superinfecção [30].

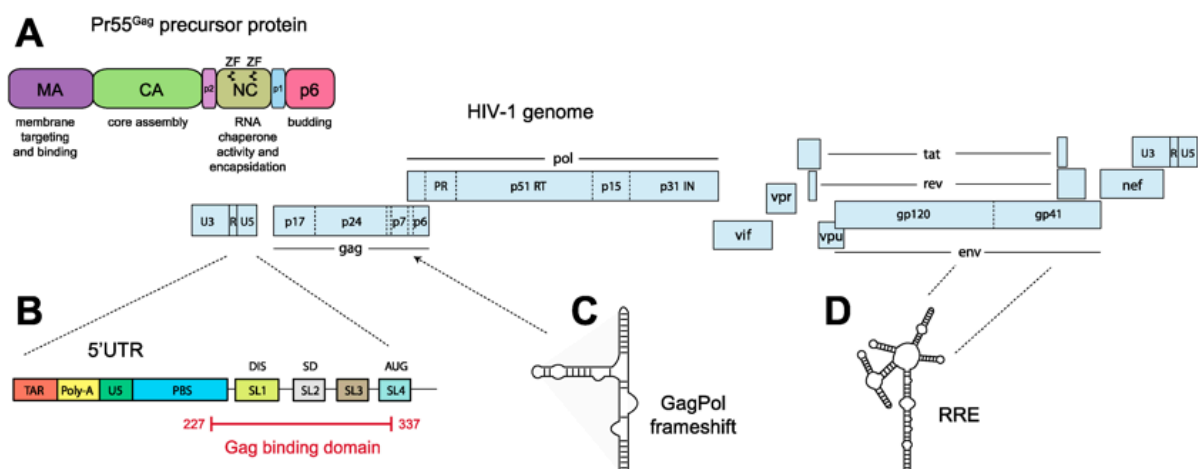


Figura 3 - Representação esquemática da estrutura linear proviral de HIV-1. Fonte: [31]

1.5 CICLO REPLICATIVO DO HIV

A infecção viral inicia-se quando a proteína gp120 do envelope viral, liga-se ao receptor CD4 da superfície celular na célula susceptível possibilitando a interação da gp120 ao correceptor celular CCR5/CXCR4. Após a interação com o co-receptor, e modificação conformacional de gp120, peptídeos hidrofóbicos da gp41 (peptídeo de fusão) são expostos e imediatamente inseridos na membrana da célula alvo. As regiões C- e N-terminal (HR1 e HR2) promovem a aproximação e contato entre a membrana celular e envoltório viral, formação de poros de fusão e liberação do capsídeo no citoplasma da célula hospedeira [16,32].

A transcrição reversa do RNA viral inicia-se a partir da ligação de um iniciador de tRNA complementar a um sítio próximo à extremidade 5' do RNA viral. A partir deste iniciador, a enzima transcriptase reversa sintetiza um segmento de DNA, complementar ao

RNA, na direção 5'. Este segmento de DNA sintetizado é translocado para o término 3' da molécula, para completar a síntese da fita de DNA. Concomitantemente a enzima RNase H é necessária para degradar o RNA do híbrido DNA-RNA. A síntese da fita de DNA complementar é realizada pela atividade DNA-polimerase RNA dependente da transcriptase reversa (RT).

O DNA viral dupla fita recém sintetizado é translocado para o núcleo celular a partir do complexo pré-integração. Uma vez no núcleo o DNA viral é integrado ao genoma celular através da ação da integrase (IN) [33]. Após a integração, o DNA viral agora denominado provírus, funciona como molde para a síntese das proteínas virais pela RNA polimerase II do hospedeiro quando a célula infectada for ativada.

Os mRNAs retrovirais são sintetizados, processados e então transportados para o citoplasma, onde são traduzidos para produzir as proteínas virais. As proteínas que formam o core viral, isto é produtos dos genes *gag* e *pol*, inicialmente montadas dentro de um nucleocapsídeo imaturo juntos com duas cópias do genoma completo de RNA viral. Como essas estruturas brotam da membrana plasmática da célula, eles se tornam envolvidos pela membrana plasmática da célula hospedeira albergando as glicoproteínas do envelope. Coincidente com o brotamento, a terceira enzima viral protease (PT) cliva as proteínas do core em suas formas finais. Tornando o vírus maduro e infeccioso [14,34–36].

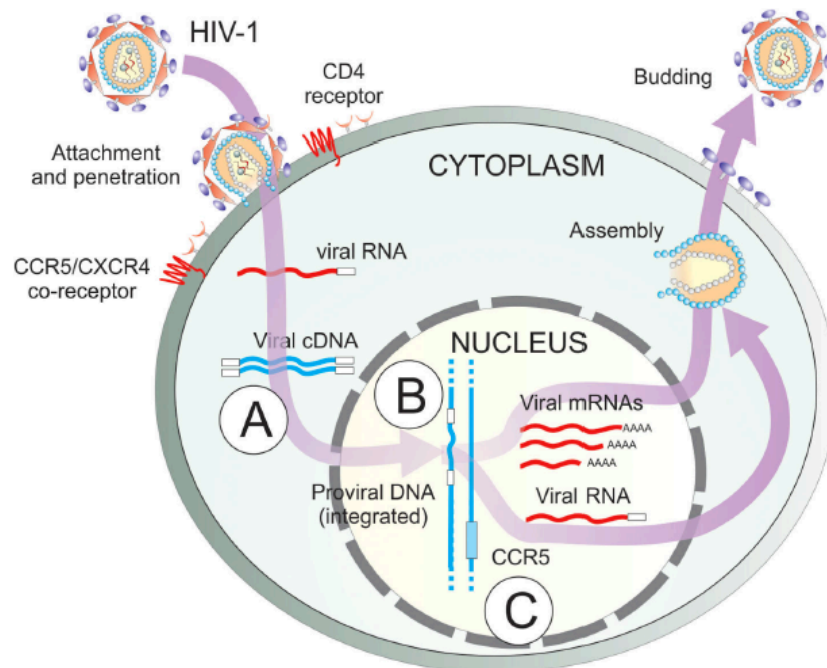


Figura 4 - Representação esquemática do ciclo replicativo do HIV-1. A= retrotranscrição do RNA viral, B = integração do genoma viral e C = síntese dos transcritos virais. Fonte: [36].

1.6 TERAPIA ANTIRRETROVIRAL

É sabido que a terapia antirretroviral de alta potência, denominada TARV, é um dos grandes ganhos no combate à Aids. Um dos principais objetivos é reduzir os níveis de carga viral ao indetectável, tornando o sistema imune menos vulnerável às doenças oportunistas, reduzir a mortalidade associada à Aids, além de reduzir drasticamente a disseminação viral para outros indivíduos suscetíveis [37].

Até os dias de hoje um arsenal de 27 drogas foram aprovadas e disponíveis pelo *Food and Drugs Administration* (FDA) para tratamento da infecção pelo HIV. Essas drogas são distribuídas em 6 classes, segundo os seus mecanismos moleculares e perfil de resistência [38,39]:

- (a) inibidores da transcriptase reversa análogos de nucleosídeos (INRTs);
- (b) inibidores da transcriptase reversa não análogos de nucleosídeos (INNRTs);
- (c) inibidores da integrase (INs);
- (d) inibidores de protease (IPs);
- (e) Inibidores de fusão;
- (f) Antagonistas de co-receptor/inibidores de entrada.

Os análogos de nucleosídeos, INRTs, formam a primeira classe de drogas a ser aprovada pelo FDA [40]. São administrados como pró-drogas, os quais requerem fosforilação ao entrarem na célula por quinases celulares [41]. A ausência do grupo hidroxil na porção 3' do açúcar (2'-desoxiribosil) meio pelo qual os INTRs impede a formação de uma ligação fosfodiéster 3'-5' entre os INTRs e o próximo 5' trifosfato nucleotídeo, resultando na terminação da cadeia crescente de DNA [42]. Atualmente, existem 8 INTRs disponíveis e aprovadas pelo FDA.

Os inibidores da transcriptase, não nucleosídeos, INNRTs, inibem a RT do HIV-1 pela ligação e indução de uma bolsa hidrofóbica proximal, mas não sobreposta ao sítio catalítico da enzima. A ligação do INNRTs muda a conformação espacial do sítio de ligação ao substrato e reduz a atividade da polimerase [43,44].

A integrase foi a mais recente enzima do HIV-1 com sucesso a ser alvo terapêutico [45,46]. Raltegravir (RAL) foi a primeira droga inibidora da integrase a ser licenciada em 2007 [47]. Sua ação inibitória dar-se de duas formas: ligando-se apenas ao específico complexo entre a integrase e o DNA viral e interagindo com o cofator íon magnésio no sítio ativo da integrase. Atualmente, existem 3 drogas disponíveis e aprovadas [48].

Em 1995, o primeiro IP foi aprovado. A protease é responsável pela clivagem das proteínas precursoras de gag e gag-pol durante a maturação viral. Os IPS atuam ligando-se ao sítio ativo da protease, impedindo a clivagem dos precursores e como consequência a maturação da partícula viral [49].

Os inibidores de fusão foram desenhados baseados na descoberta de dois domínios homólogos da proteína viral gp41 que devem interagir um com o outro para promover a fusão, e essa imitação de um desses domínios pela proteína heteróloga pode se ligar e interromper as interações da proteína viral [49].

O antagonista de CCR5 opera pela ligação à proteína de superfície de células da imunidade. Uma vez estabelecida essa ligação, certas variantes do HIV-1, chamadas R5 trópicas não podem se ligar ao correceptor, e portanto, não conseguem entrar na célula susceptível [41,49].

Tal como acontece com todas as terapias antirretrovirais, o tratamento com qualquer um destes agentes muitas vezes resulta no aparecimento de variantes de HIV-1 com reduzida sensibilidade às drogas. Os mecanismos de resistência associados à classe INRTs são mediados pelos (1) mecanismo ATP-dependente da pirofosforilase, que é a remoção do INRT da extremidade 3' da cadeia nascente e o aumento da discriminação entre o substrato nativo de desoxiribonucleotídeo e o inibidor [50].

A resistência aos INNRTs geralmente resulta da substituição de aminoácidos tais como L100, K101, K103, E138, V179 e Y188 na bolsa de ligação dos inibidores a RT [51,52].

A resistência aos inibidores de integrase quase sempre estão associados aos resíduos de aminoácidos que coordenam os co-fatores de magnésio essenciais. Em estudos clínicos, a resistência ao raltegravir está associada a um conjunto de quatro mutações independentes no gene da integrase (Y143, N155, 121 ou Q148) [53–55].

Todos os IPs compartilham relativamente uma estrutura química semelhante, assim, resistência cruzada é comumente observada. Para a maioria dos IPs, mutação de resistência primária agrupadas próximas ao sítio ativo da enzima, na posição localizada no sítio de ligação do substrato/inibidor (D30N, G48V, I50V, V82A ou I84V, entre outros) [56,57].

1.7 HISTÓRIA NATURAL DA INFECÇÃO PELO HIV-1

A infecção primária ou fase aguda é geralmente assintomática e é caracterizada pelo sequencial aparecimento de marcadores viral e anticorpos no sangue. Atualmente, testes de

quarta-geração mais sensíveis são capazes de detectar precocemente tanto antígenos virais quanto os anticorpos e reduzem o período de janela imunológica por volta de 5 dias [58,59].

Clinicamente as manifestações são vagas e inespecíficas tais como: febre, linfadenopatia, cefaleia, mialgia, náuseas, diarreia e exantema. A fase aguda é limitada, rápida e a sua severidade está associada à progressão rápida da doença. Reagentes de fase aguda incluem elevados níveis de proteína amiloide A sérica (3 a 5 dias após infecção). Os níveis de carga viral aumentam coincidem com a explosão de citocinas inflamatórias. A resposta de células T CD8⁺ leva ao aparecimento de mutantes de escape. Complexos imunes de anticorpos, precedem os primeiros anticorpos livres da gp41. Anticorpos espécie-específicos que neutralizam o vírus não aparecem até o octogésimo dia de infecção [60–62].

Com o fim da fase aguda, o paciente entra na fase crônica da infecção assintomática, conhecida como fase de latência clínica. Esta fase pode durar de meses a anos, com progressiva destruição de células da imunidade (linfócitos T CD4⁺, macrófagos e células dendríticas). Uma vez que o estado imunológico do indivíduo infectado se torna crítico, com contagem de T CD4⁺ inferior a 200 cel/mm³ emergem as doenças definidoras de Aids [63]. A Figura 5 resume a infecção natural pelo HIV-1.

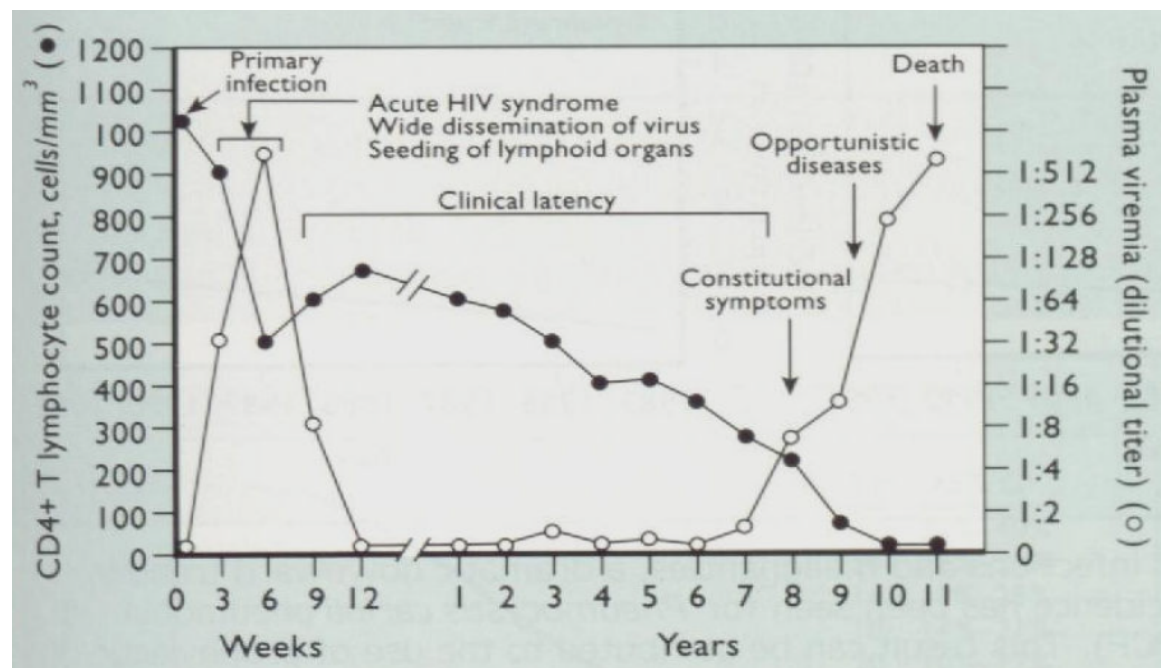


Figura 5 - História natural da infecção pelo HIV-1. Fonte: [64]

1.8 DIVERSIDADE GENÉTICA DO HIV E EPIDEMIOLOGIA MOLECULAR

O HIV se caracteriza por um enorme polimorfismo genético. Inicialmente as variantes do vírus eram identificadas como americanas ou africanas de acordo com o seu local de origem. Mais recentemente fez-se necessária a classificação baseada nas relações filogenéticas de genomas completo, em grupos, subtipos, sub-subtipos, formas recombinantes circulantes (CRFs) e de formas recombinantes únicas (URFs) [65]. O HIV-1 é classificado filogeneticamente em quatro grupos M, N, O e P, refletindo separadas introduções do vírus da imunodeficiência símia (SIVs) de primatas não humanos para os seres humanos [8]. Destes, apenas o grupo M sofreu uma disseminação pandêmica. Fatores virais e dos hospedeiros, poderiam ter influenciados na relativa transmissibilidade dos diferentes grupos de HIV-1 [8]. Modelagem matemática de vírus sugerem que o grupo M tenham sofrido diversificação entre 1910-1930 [66,67], seguido pela disseminação endêmica do vírus na população humana na África centro-ocidental.

O grupo M é o mais representativo sendo dividido em nove subtipos A, B, C, D, F, G, H, J, K [65]. As variantes dos subtipos A e F podem ainda ser classificadas em sub-subtipos [A1-A5] e [F1-F2] [7,68,69]. A Figura 6 apresenta de forma esquemática a classificação do HIV.

Eventos de recombinação são comuns em áreas onde existe uma co-circulação de subtipos gerando linhagens mosaicas as quais são responsáveis por 10-20% de novas sequências geradas. Dois tipos de genomas mosaicos podem ser gerados. As formas recombinantes circulantes (CRFs), quando identificadas em pelo menos três indivíduos HIV infectados sem vínculo epidemiológico direto e que apresentam os mesmos pontos de recombinação ao longo do genoma, estas geralmente apresentam importância epidemiológica e as formas recombinantes únicas (URFs).

Atualmente 79 formas recombinantes circulantes foram descritas [70]. Além deste conjunto, existem as formas recombinantes únicas (URFs) que são formas recombinantes entre os subtipos, mas que são encontradas em apenas um indivíduo infectado. Centenas de URFs foram encontradas com base na análise parcial das sequências do genoma e algumas possuem análise completa do genoma, porém o seu potencial de propagação epidêmico é pouco conhecido [71,72]

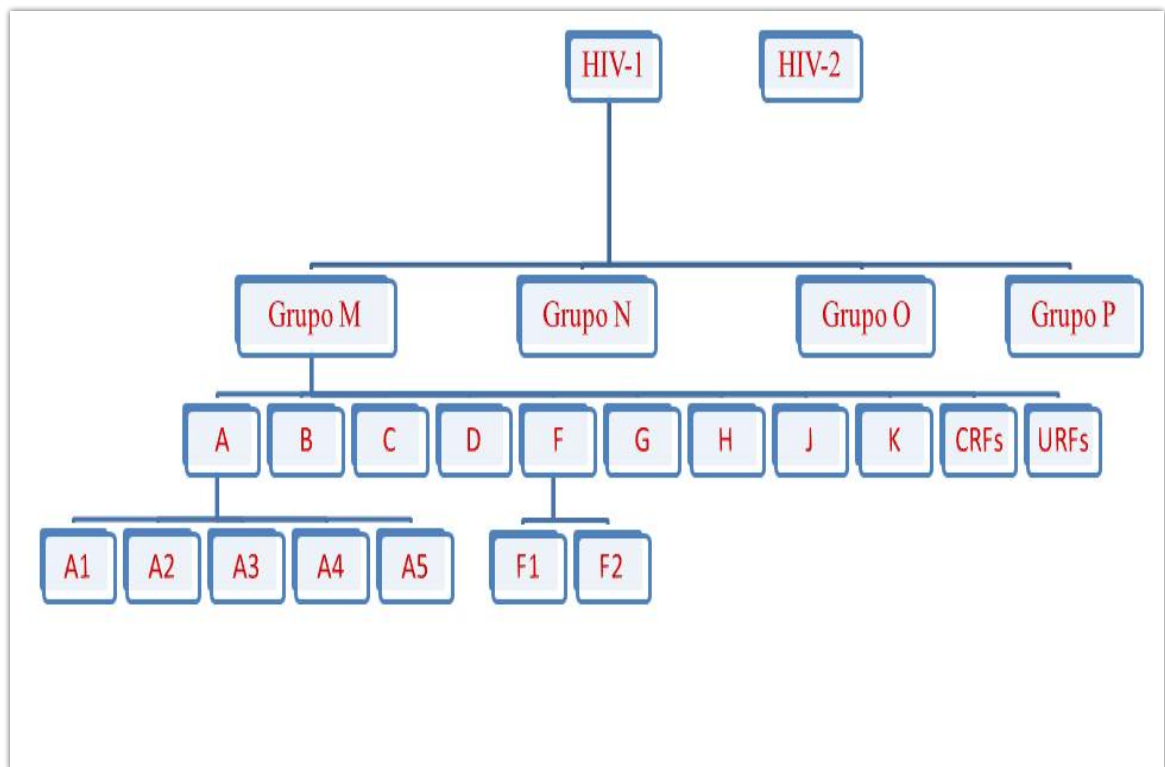


Figura 6 - Classificação atual do HIV. Fonte: adaptado de [65].

A notável variabilidade genética dentro e entre os subtipos do grupo M de HIV-1, pode alcançar 25%-35% no gene *env*, sendo possível um controle preciso da transmissão dos HIV-1 entre populações numa escala global [73], e entre indivíduos numa micro escala [74].

A vigilância global da distribuição dos subtipos e formas recombinantes circulantes (CRFs), revelou em 2011 que o subtipo C representa por volta de 50% do total dos infectados no mundo inteiro e os subtipos A, B, D e G correspondem a 12%, 10%, 3% e 6%, respectivamente, e os subtipos F, H, J e K juntos correspondem a 1% dos casos [75].

Na América do Sul, responsável por 4% das infecções neste continente, o subtipo prevalente é o B, que representa mais de 70% dos casos. Os subtipos F1, C e recombinantes BF1 são também encontrados nessa região [76]. O subtipo F1 está presente em grande parte dos países do cone sul, tendo como maior destaque o Brasil [77,78], Argentina [79], Bolívia [80] e Uruguai [81]. No entanto, com exceção do Brasil, amostras F1 puras são raramente encontradas, na maioria dos casos, o subtipo F1 é encontrado na sua forma recombinante BF1 e destas, dentre os CRFs destacamos o CRF12_BF e em menor extensão o CRF17_BF encontrados na Argentina, Uruguai e Paraguai e mais recentemente o CRF38_BF foi identificado no Uruguai [82,83].

Uma plêiade de subtipos e formas recombinantes de HIV-1 descritos, em diferentes perímetros geográficos (Figura 7) pode ter como consequência uma poderosa barreira antigênica que afeta o sucesso da terapia antirretroviral bem como o desenvolvimentos de protótipos vacinais eficazes.

Apesar deste ambiente fértil para mutação, existe pouca evidência de diferenças entre os subtipos de HIV-1 nos seus padrões ou eficiência de transmissão. Por outro lado, existe uma observação epidemiológica que o vírus do subtipo D de Uganda e regiões vizinhas possam mostrar mais comumente vírus R5/X4 dual trópico mais que outros subtipos de HIV-1, mas isso deve ter um maior efeito na patogênese viral que na transmissão em si [84].

A transmissão do HIV pode, também, ser modulado por outros fatores, incluindo doenças sexualmente transmissíveis, particularmente, estas resultam em inflamações e úlceras que podem elevar a suscetibilidade de contrair o HIV em 2 a 11 vezes [85].

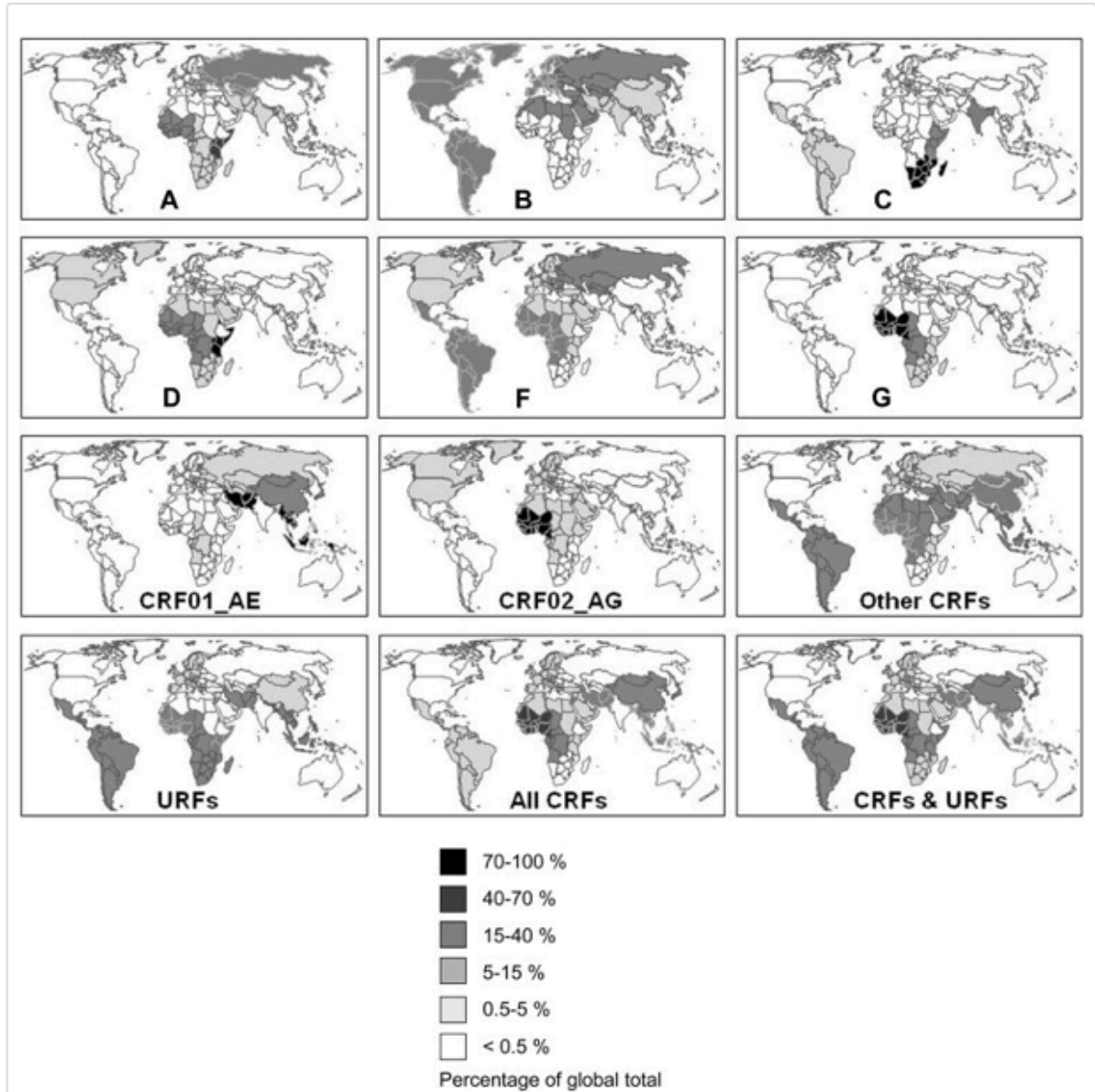


Figura 7 - Distribuição global dos subtipos e recombinantes de HIV-1 do grupo M. Para cada subtipo/recombinante um painel único. Fonte: [72].

1.9 EPIDEMIOLOGIA MOLECULAR DO HIV-1 NO BRASIL

O monitoramento da diversidade genética do HIV-1 se faz útil no Brasil, a fim de compreender a disseminação da epidemia, bem como perceber as características que compõem os vírus circulantes.

Inúmeros trabalhos envolvendo coorte de pacientes de diversas regiões geográficas do país acusam uma predominância do clado B (~70%), porém, diferentes dos países que compõem a América do Norte e alguns países da Europa ocidental, o Brasil apresenta um

expressivo número de infecções causadas pelos clados não B (F1, C, formas recombinantes BF1, BC, além do recombinantes complexos) [77,86,87].

Um cenário diferente do observado na maior parte do território nacional, a região do Sul do país apresenta uma elevada prevalência de mais de 50% das infecções estão relacionadas ao subtipo C [88,89]. Cabe ressaltar que a depender da cidade do sul do país esse número pode ser entorno de 80% das infecções [90].

A co-circulação dos diferentes subtipos em uma localidade aliada ao crescimento da epidemia em muitas regiões do país, além do crescente número de estudos de epidemiologia molecular do HIV-1 em importantes centros urbanos no Brasil, têm gerado um volume importante de informações no que tange as formas recombinantes virais. A maioria dos recombinantes descritos são URFs, não obstante alguns recombinantes foram capazes de se espalhar na epidemia e deram origem aos diferentes CRFs descritos no Brasil tais como: CRF28_BF1, CRF29_BF1, CRF31_BC, CRF39_BF1, CRF40_BF1, CRF46_BF1, CRF_72BF [91–94].

Estudos sobre a evolução do HIV no Brasil sugerem que o primeiro subtipo a ser introduzido na população brasileira foi o subtipo B (por volta de 1965-1970), seguido pelo subtipo F1 (~1975-1980) [95]. As primeiras descrições do subtipo F1 em amostras brasileiras são de dois trabalhos de 1994 com amostras do início dos anos 90 [77,96] sendo também da mesma época a identificação do primeiro genoma recombinante B/F [97].

O subtipo C de HIV-1 tem se disseminado eficientemente em estados do sul do Brasil. Estudos de filogeografia indicam que a epidemia do subtipo C no sul No Brasil foi iniciada pela introdução de um único vírus fundador em algum momento entre 1960 e 1980, mas pouco se sabe sobre a dinâmica espacial da disseminação viral [90,98].

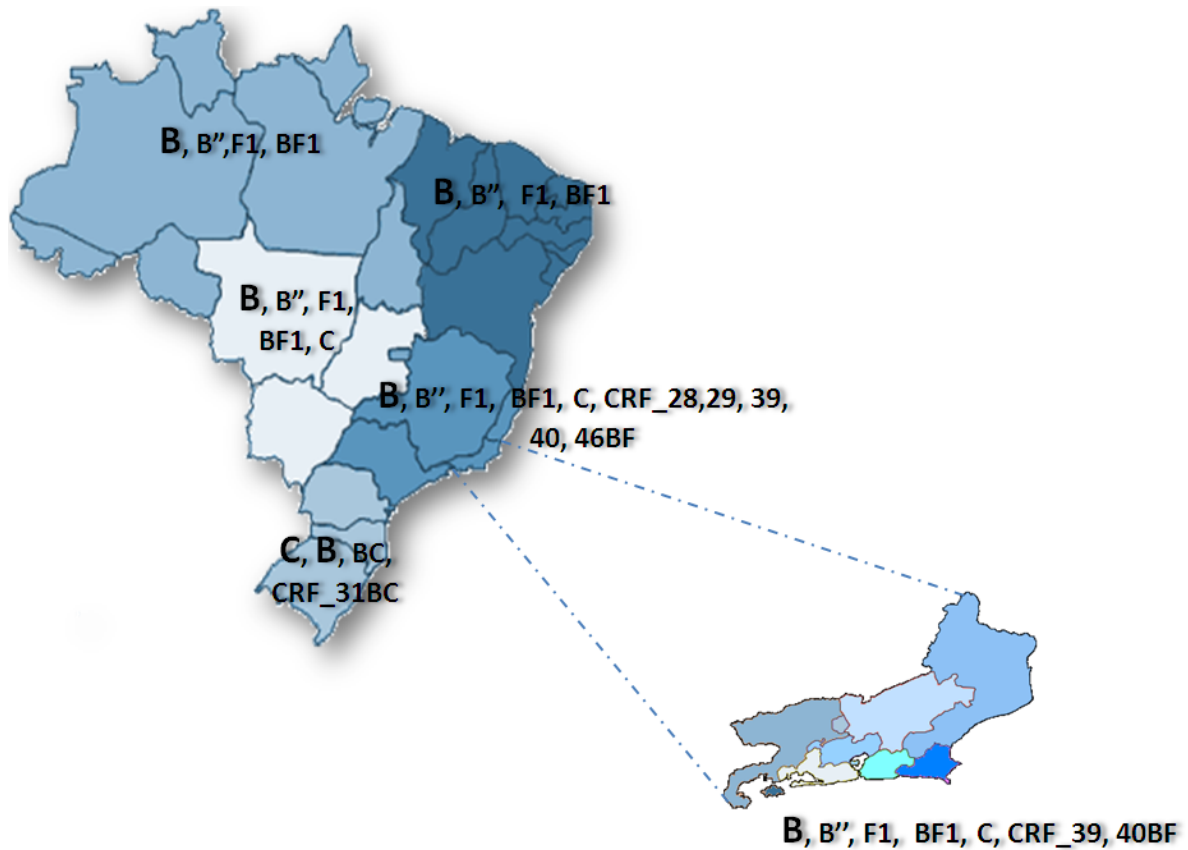


Figura 8 - Distribuição dos subtipos e formas recombinantes circulantes do HIV-1 no Brasil. Fonte: Adaptado de [78].

1.10 TEORIA DAS QUASIESPÉCIES DO HIV

A dinâmica evolutiva dos vírus de RNA são complexas o tamanho populacional quase infinito, as altas taxas de mutação, e a rápida cinética de replicação dos vírus de RNA impossibilitam sua descrição pela genética de populações clássica, no qual o *fitness* (adaptabilidade ou cinética de replicação) de um fenótipo característico é principalmente o produto da habilidade do sucesso reprodutivo [99,100]. Em vírus de RNA tal qual o HIV, a taxa de erro extremamente elevada durante o curso da replicação viral cria um grande *swarms* de variantes geneticamente diversas.

Coletivamente, este *swarm* de variantes ainda similar, do qual emergem de um tronco único fruto da infecção de uma variante (vírus fundador), forma as quasiespécies (Figura 9). Dentro das quasiespécies, o *fitness* de um fenótipo depende da habilidade reprodutiva de uma variante individual e da sua capacidade de mutação.

Para ser considerado quasispécies, o vírus deve conter os seguintes critérios: uma variante individual dentro de uma população deve ter divergido da sequência master de um vírus fundador por mutação, uma variante individual dentro de uma população deve responder à pressão de seleção dentro do hospedeiro e por fim, cada variante deve contribuir para melhorar os atributos da população total [101].

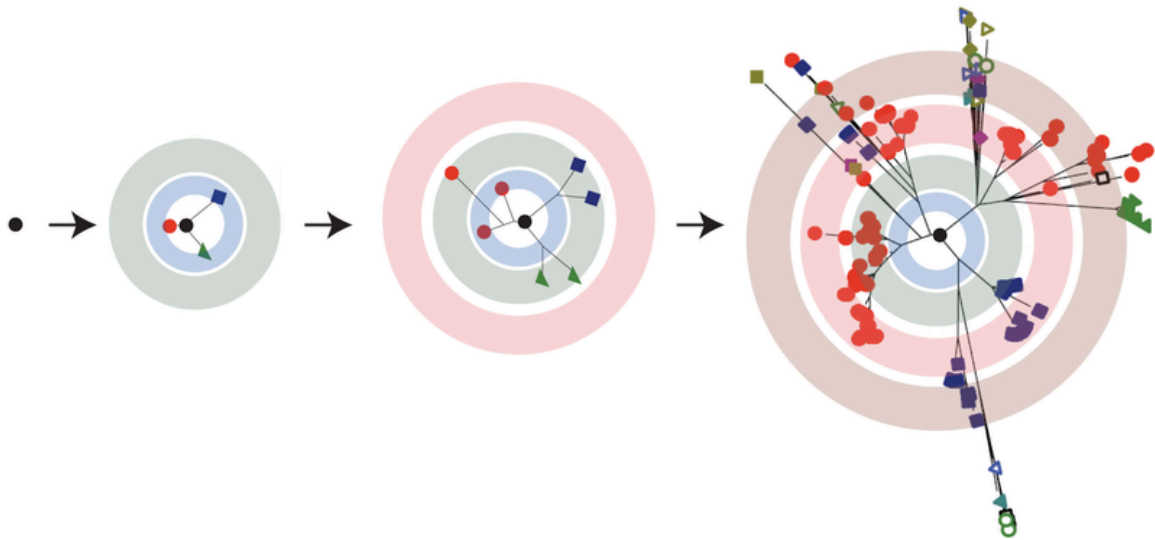


Figura 9 - Esquema de geração das quasi-espécies. As variantes individuais que compõem as quasispécies são geneticamente ligadas a sequência master (assinaladas como círculo preto). Fonte: [102].

O entendimento da ligação entre a teoria da quasispécie e o efeito do comportamento do HIV dentro do hospedeiro pode ajudar a elucidar estratégias terapêuticas de atenuação ou eliminação do patógeno dentro do indivíduo infectado [103].

Apesar dos inúmeros estudos sobre epidemiologia molecular do HIV-1, ainda há uma lacuna do conhecimento no tocante à diversidade genética intra-hospedeiro na população de indivíduos recentemente infectados em especial no Brasil, onde há co-circulação de diferentes subtipos, além de um número importante de formas recombinantes.

Muitos estudos de diversidade genética do HIV-1 no Brasil estão atrelados apenas à vigilância das mutações de resistência aos antivirais. Todavia, o presente estudo se propõem a estudar questões de variabilidade intra-hospedeiro, pressão de seleção agindo nas populações virais, e descrição de mutações associadas à resistência aos antirretrovirais que mantêm-se escondidas entre as sub-populações minoritárias dentro do paciente infectado.

Um outro olhar não menos importante, que reverbera da diversidade genética intra-hospedeiro é o efeito do *bottleneck* (gargalo) da diversidade viral durante a transmissão inter-hospedeiro, e as consequência de vírus de baixo *fitness* $R_0 < 1$, sendo portanto a infecção não produtiva e quando o vírus apresenta maior *fitness* $R_0 > 1$, a infecção clínica é estabelecida com altos títulos viral na ordem de 10^6 partículas/mL (Figura 10). No contexto do Brasil, não há qualquer trabalho que descreva as malhas de transmissão inter-hospedeiro associadas aos subtipos virais, terapia e grupos de risco.

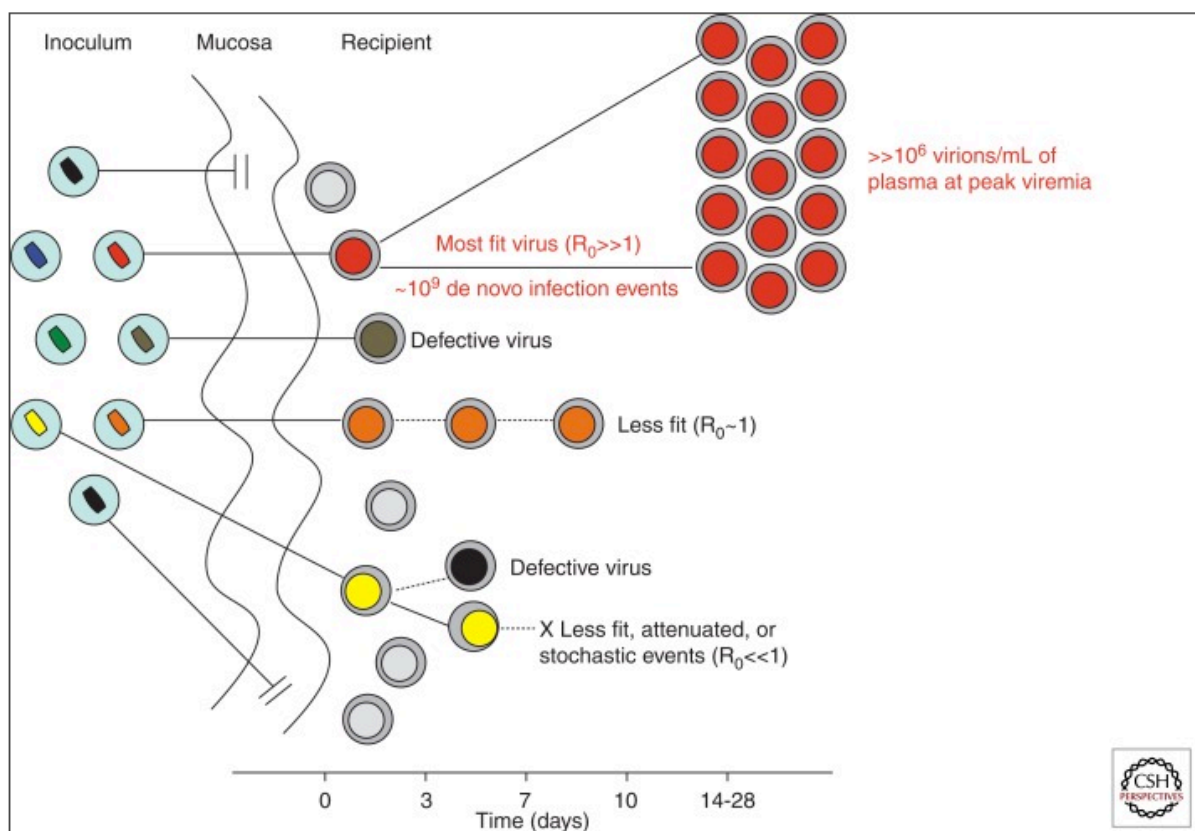


Figura 10 - Modelo de Transmissão do HIV-1 Fonte: [102].

2 OBJETIVOS

2.1 OBJETIVO GERAL

O presente estudo tem como objetivo descrever a dinâmica da diversidade genética do HIV intra-hospedeiro, levando em consideração os marcadores resistência primária ou secundária aos antirretrovirais em voga em soroconvertedores recentes bem como descrever as redes de transmissão do HIV-1 no estado de São Paulo.

2.2 OBJETIVOS ESPECÍFICOS

- Descrever as variantes do HIV-1 intra-hospedeiro através da tecnologia do sequenciamento largamente paralelizado usando a plataforma MiSeq-Illumina.
- Detectar as possíveis mutações de resistências primária ou secundária que se apresentavam escondidas pelo método de sequenciamento populacional (Sanger).
- Descrever posições no gene *pol* que estão sob efeito de evolução purificadora ou adaptativas.
- Descrever a frequência de códons diferenciais entre os haplótipos reconstruídos.
- Descrever a dinâmica da epidemia de Aids no estado de São Paulo a partir de inferências filogenéticas e dados clínicos e demográficos.
- Descrever a frequência de clusters endógenos segundo os subtipos B, C e F1 de HIV-1 ao longo do território do estado de São Paulo.

3 MATERIAL E MÉTODOS

3.1 CASUISTICA

A população alvo deste estudo consiste de indivíduos infectados pelo HIV-1 (n=2.518) oriundos de 53 cidades do estado de São Paulo, sendo a maioria da região metropolitana. Nesta coorte incluímos indivíduos infectados pelo HIV-1 em diferentes fases da doença (agudos e crônicos). As amostras de sangue total foram encaminhadas ao Instituto Adolfo Lutz (IAL-SP) para os testes de resistência genotípica conduzidos entre janeiro de 2004 a fevereiro de 2015. A figura 11 sumariza a origem das amostras ao longo do território do estado de São Paulo.

Nesta etapa do estudo, todas sequencias parciais do gene *pol* (PR/RT) foram coletadas pelo sequenciamento populacional (Sanger). Os indivíduos que tiveram as suas sequencias genéticas disponíveis foram incluídas no estudo. Neste contexto nós compilamos os dados demográficos, geográficos e clínicos quando disponíveis a fim de compreender a dinâmica da epidemia de Aids no estado de São Paulo. A partir desses esforços foi possível compreender a relação em cadeia de transmissibilidade entre os pacientes em diferentes fases da doença na presença ou ausência da TARV e segundo a sua exposição ao risco.

A fim de estudar a diversidade genética intra hospedeiro bem como mutações associadas a resistência que não foram detectadas pelo método clássico de sequenciamento genético. Um sub conjunto de 24 amostras foi selecionado baseada no critério de diagnóstico recente entre 2011-2015 e com perfil de soroconversão recente seja por:

- (a) intervalo inferior a 2 anos entre o último teste anti-HIV negativo e o primeiro positivo;
- (b) Teste de avidéz para detecção de infecção recente do HIV-1;
- (c) Sinais clínicos especificamente síndrome retroviral aguda.

Este sub-conjunto de amostras foi submetido ao método de sequenciamento genético de nova geração (NGS-Illumina) pela abordagem de amplicom do gene *pol*.

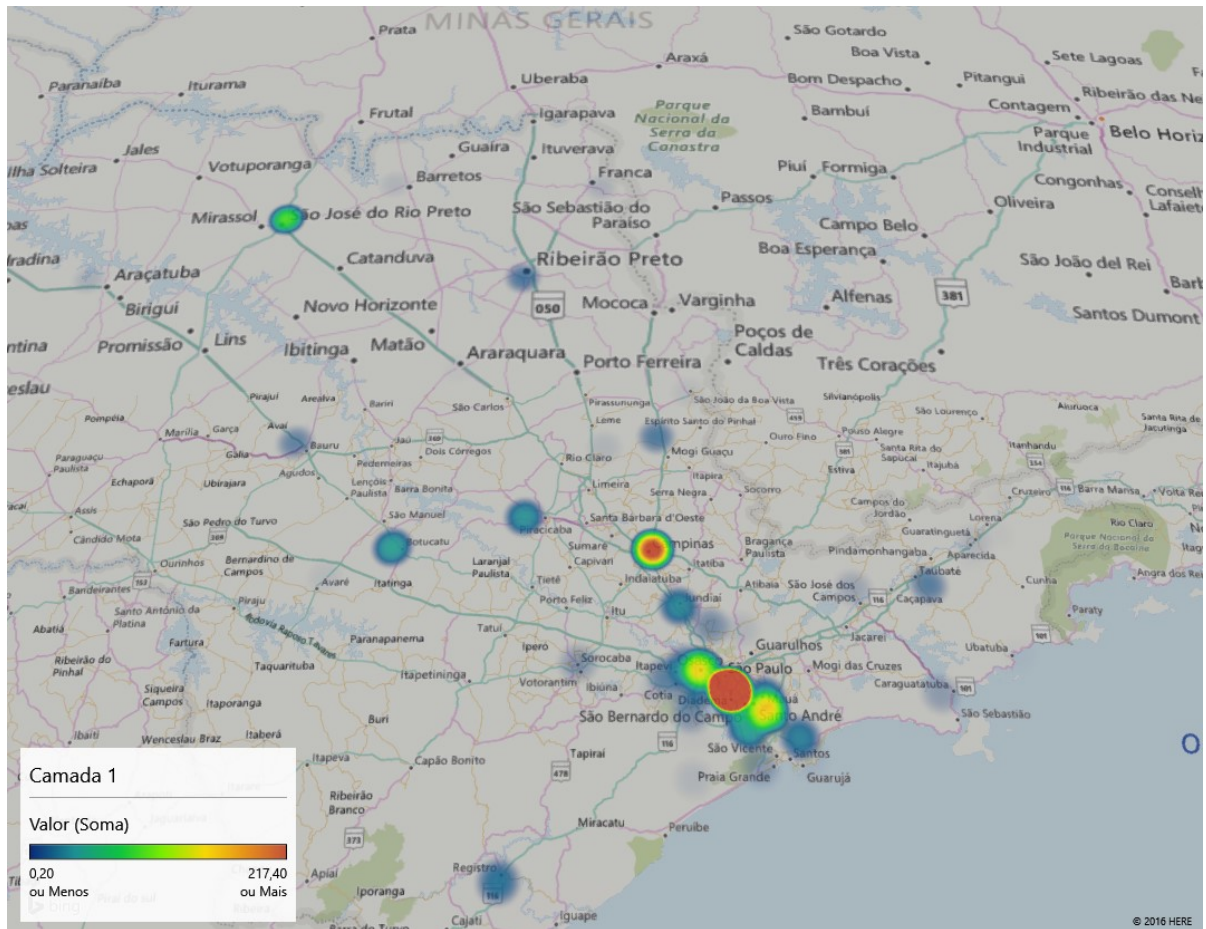


Figura 11 - Mapa de calor representando a origem e quantidade de amostras por localidade dentro do estado de São Paulo.

3.2 EXTRAÇÃO DE RNA

A extração do RNA viral a partir do plasma e de células mononucleares do sangue periférico procedeu-se pelo método cromatográfico (colunas de sílica), utilizando-se o QIAamp DNA Blood Mini Kit (QIAGEN, Alemanha), conforme indicação do fabricante. Neste kit, a lise celular ocorre através do uso de tampões contendo Isotiocianato de Guanidina, e a digestão dos componentes protéicos da célula ocorre sob ação da enzima Protease (QIAGEN). O RNA extraído foi identificado e estocado a $-80\text{ }^{\circ}\text{C}$.

3.3 AMPLIFICAÇÃO POR nested PCR DA REGIÃO POL

Com o intuito de caracterizar os subtipos/variantes virais em indivíduos HIV-1 positivos em São Paulo recém diagnosticados com o perfil de soroconversão recente,

utilizamos a técnica de RT-PCR seguido da nested PCR para amplificar a região *pol* do HIV-1.

Para amplificação de um fragmento de 1.112 pb (correspondente a posição 2252-3574 em relação a cepa HXB2) da região da polimerase viral, região gênica que compreende a protease e parcialmente a RT, utilizamos como iniciadores externos GAG2-R137 e como iniciadores internos PR3/RT12 (Tabela 1).

Os reagentes utilizados em todas as reações de RT-PCR para cada amostra foram: Tampão 10X da Enzima high fidelity; 10 mM Tris-HCl (pH 8.3); 50 mM MgSO₄; 200 mM de cada dNTP; 10 pmol de cada iniciador (senso e anti-senso); 2,5 U Taq polimerase high fidelity; Inibidor de RNase 40 U/ μ L; 0,1 mM de DTT, Superscript III e água DEPC. A essa solução adicionamos 2.5 μ L do RNA de cada amostra. Para a amplificação das regiões alvo do estudo a ciclagem da RT-PCR foi a seguinte: 1 ciclo de 50 °C por 30 min, 1 ciclo de 94 °C por 2 min, seguindo-se de 40 ciclos de 94 °C por 15 seg, 55 °C por 30 seg, 68 °C por 2,5 min e uma extensão final de 68 °C por 10 min. Para a amplificação do produto de nested PCR os reagentes utilizados foram: Go Taq colorless master mix 2X (Promega), iniciadores 10 pMol, água DEPC.

Após a primeira reação de PCR foram transferidos 5 μ l do produto amplificado para a segunda reação de PCR, totalizando um volume de 50 μ l.

Para verificar se houve amplificação dos fragmentos foram aplicados 5 μ l do produto da segunda reação de PCR adicionado de 1 μ l do corante Azul de Bromofenol (Azul de Bromofenol 28%; EDTA 10mM pH 8,0 e Sacarose 40%) em gel de agarose 1% corado com brometo de etídeo 0,001% (v/v). As amostras foram submetidas à eletroforese (80V por uma hora e trinta minutos) utilizando o TBE 1X (Tris-HCl 1M, Ácido Bórico 1 M e EDTA 20 mM) como tampão de corrida. O fago Φ X174 digerido com a enzima Hae III (1353, 1078, 872, 603, 310 pb) foi utilizado como padrão de peso molecular. Todas as reações foram monitoradas através da utilização de controles positivos (DNAs de amostras HIV positivas amplificáveis) e negativos (todos os reagentes do PCR excluindo DNA).

Tabela 1 - Estratégia de amplificação da região *pol* do HIV-1

Região genômica	Iniciadores	Posição (HXB2)	Tamanho
Pol	GAG2-R137	2020-3937	1.917pb
	PR3-RT12	2148-3260	1.112pb

3.4 PURIFICAÇÃO E QUANTIFICAÇÃO DOS PRODUTOS DE PCR PARA SEQUENCIAMENTO

As amostras amplificadas para a região *pol* foram purificadas para o uso nas reações de sequenciamento por método cromatográfico, através de colunas de purificação, utilizando-se o kit em colunas (GE Healthcare) seguindo as recomendações do fabricante.

A quantificação dos produtos purificados foi realizada em gel de agarose (2%, corado com brometo de etídeo 0,001% (v/v)). Foram aplicados 2 μ l do produto purificado adicionado de 1 μ l do corante Azul de Bromofenol (Azul de Bromofenol 28%; EDTA 10 mM pH 8,0 e Sacarose 40%) no gel. As amostras foram submetidas à eletroforese (80V por duas horas) utilizando o tampão TBE 1X como tampão de corrida. Como padrão de peso molecular, utilizamos 2 μ l do Low DNA Mass Ladder. Este padrão apresenta bandas de 100, 200, 400, 800, 1200 e 2000pb que correspondem a 5, 10, 20, 40, 60 e 100 ng de DNA, respectivamente.

A intensidade das bandas correspondentes às amostras foi comparada com as bandas do padrão, a fim de estimar a - concentração, em nanogramas, do produto purificado e o volume do mesmo a ser utilizado na reação de sequenciamento.

3.5 REAÇÃO DE SEQUENCIAMENTO AUTOMATIZADO

Na reação de sequenciamento, utilizamos 1 μ l do Big Dye v.3.1 (Ready Reaction Mix), 1,5 μ l do tampão de diluição 5X (BigDye Terminator v1.1/3.1 Sequencing Buffer), 1 μ l do oligonucleotídeo iniciador a 5pmol/ μ l e a essa solução adicionamos 60 a 100 ng do produto da segunda reação de PCR purificada. Água Milli-Q foi adicionada à reação para completar um volume final de 10 μ l.

O emprego dos iniciadores na reação de sequenciamento das região gênica descrita (*pol*) está sumarizado na tabela 2.

Tabela 2 - Iniciadores a serem utilizados na reação de sequenciamento Sanger da região *pol*

Fragmento	Região do Genoma	Iniciadores utilizados no sequenciamento
A	<i>pol</i>	DP10/DP11/DP16/DP17/F1/RT2882/RT2949/RT12

A ciclagem utilizada para a reação de sequenciamento foi a seguinte 25 ciclos: 96 °C por 30 segundos, 50 °C por 20 segundos e 60 °C por 4 min.

Após a reação de sequenciamento, foi realizada a etapa de precipitação do produto sequenciado. A cada poço da placa de sequenciamento foram adicionados 80 μ l de Isopropanol 75%. A placa foi incubada ao abrigo da luz por 15 min e então centrifugada em centrífuga com rotor adaptado para placas (Eppendorf 5804-R) a 4000 rpm por 45 min. Após a centrifugação o Isopropanol foi removido, vertendo-se a placa e o excesso removido através da centrifugação da placa invertida a 900 rpm por um min, protegida com papel absorvente. Com a finalidade de garantir a ausência de Isopropanol nos poços da placa, a mesma foi levada ao termociclador a 95 °C por 5 min. A partir de então a placa foi mantida ao abrigo da luz a -20 °C até o dia em que seria submetida ao sequenciador.

Antes de serem colocadas no sequenciador, as amostras foram ressuspensas em 10 μ l de Formamida HI-DI (Applied Biosystems, EUA), aquecidas em termociclador por 5 min a 95 °C, para garantir que o “template” fosse desnaturado, e então submetidas imediatamente ao gelo para evitar que as fitas de DNA voltassem a parear.

O sequenciador DNA Sequencer modelo 3130 (Applied Biosystems, EUA) foi utilizado para o sequenciamento automatizado das amostras. O protocolo utilizado para o sequenciamento seguiu a recomendação do fabricante para a versão 3.1 do kit Big Dye Terminator.

3.6 ANÁLISE E EDIÇÃO DE SEQUENCIAS OBTIDAS PELO MÉTODO SANGER

Os cromatogramas gerados a partir do sequenciamento foram visualizados e editados, utilizando o pacote de programas DNASTAR [104]. Utilizando o software SeqMan 4.0, as sequências geradas foram alinhadas e editadas manualmente, gerando uma sequência consenso correspondente à região amplificada.

3.7 SUBTIPAGEM VIRAL E ANÁLISE FILOGENÉTICA

A fim de estudar as redes de transmissão do HIV-1 envolvendo os diferentes subtipos B, C e F1 no estado de São Paulo, submetemos o conjunto total de sequências sem duplicatas a diferentes algoritmos de subtipagem viral (Rega v.2.0, Rega V.3.0, Comet, Scuel e jpHMM). Esta etapa foi fundamental para remoção das sequências genéticas com

algum sinal de recombinação homóloga comum entre os retrovírus que poderiam causar ruídos nas análises.

Algumas das amostras que apresentaram alguma divergência entre os diferentes algoritmos foram submetidas a análise de recombinação segundo o programa Simplot 3.5.1[105]. Esta metodologia utiliza o método de NJ com fator de substituição de nucleotídeos K-2 [106] para fazer a análise filogenética do fragmento alvo, esta se dá através de fragmentos menores (janelas), que são analisados por sobreposição (passos). O suporte dos grupamentos filogenéticos foi realizado pelo *bootstrap* de 100 pseudoréplicas. Os parâmetros utilizados na análise de *bootscan* foram: janela = 300 pb e passos de 20 pb. Nossas amostras foram analisadas frente às sequências disponíveis no banco de dados de Los Alamos consideradas puras para cada subtipo viral (B_BZ167, C_AY145829, F1_BZ163).

Após a subtipagem e controle de qualidade das sequências, nós dividimos as mesmas segundo os subtipos puros B, C e F1 e então selecionamos 2.107 sequências controle globais únicas obtidas a partir do *blast* das nossas sequências frente ao banco de dados de Los Alamos (<http://hiv-web.lanl.gov>) e o banco de dados de Portugal. Assim, nós alinhamos as sequências pertinente a cada subtipo separadamente com as 10 sequências globais mais próximas usando o *muscle* [107]. Uma análise filogenética preliminar usando o FastTree usando o modelo de *Maximum-likelihood*, modelo de substituição nucleotídica GTR [108]. Os códons associados a resistência foram removidos das sequências [109].

Para a análise da dinâmica da transmissão foi realizado a reconstrução filogenética pelo critério de *máximum likelihood* (ML) usando o RAxML versão 7.4.8 [110] sobre o modelo de substituição GTR + Γ sem partições. Nós usamos o programa Cluster Picker para definir as redes de transmissão dos *clusters* a partir dos parâmetros do suporte de bootstrap ≥ 90 e distância genética ≤ 6.0 .

3.8 DINÂMICA GEOGRÁFICA DOS CLUSTERS INTER E INTRA ESTADO DE SÃO PAULO

Com base na origem do serviço do paciente, clusters que contem $\geq 66\%$ das sequências com a mesma área do estado de São Paulo foi classificado como cluster endógeno. No caso de clusters com apenas duas sequências, ambas devem pertencer a mesma localidade para serem definidas como cluster endógeno. São identificados clusters externos quando pelo menos 34% das sequências que os compõem são de áreas diferentes.

3.9 ANÁLISE DE RESISTÊNCIA ANTIRRETROVIRAL

As sequencias parciais do gene *pol* (PR/RT) foram investigadas quanto a presença de mutações associadas a resistência a drogas de acordo com os preceitos estabelecidos pela Organização Mundial de Saúde (OMS) utilizando a última atualização de 2009 do CPR (Calibrated Population Resistance) [109].

3.10 ANÁLISES ESTATÍSTICAS

As análises estatísticas foram realizadas em SPSS [111]. As tabelas de frequências foram feitas para as variáveis qualitativas. O teste de normalidade Kolmogorv-Smirnov foi aplicado a fim de averiguar o tipo de distribuição das variáveis. Mediana e intervalo interquartis (IQR) foram usados para variáveis quantitativas e foram comparadas usando o teste de Mann-Whitney U ou o teste de Kruskal Wallis. As diferenças entre as proporções relacionadas as variáveis qualitativas foram realizadas a partir do teste do Qui quadrado ou teste de Fisher.

3.11 SEQUENCIAMENTO LARGAMENTE PARALELIZADO (MiSeq)

O sequenciamento de nova geração (NGS) foi realizado no equipamento de disponível pelo *core facilitie* do Instituto de ciências biomédicas 2 , a fim de identificar a as quasi-espécies do HIV-1 intra hospedeiro que é de grande valia para uma melhor compreender a dinâmica da evolução viral bem como detectar os polimorfismos que conferem algum tipo de resistência aos antirretrovirais em uso no país.

O amplicon do gene *pol* foi quantificado no Qubit de acordo com o manual do fabricante e aproximadamente 1ng do amplicon foi usado na reação de fragmentação usando o Kit de preparação da biblioteca Nextera XT de acordo com a recomendação do fabricante. A fragmentação de cada amostra foi simultaneamente realizado pela incubação por 5 min a 55°C, seguido pela incubação por 5 min a 10°C em tampão de neutralização. Na sequencia foi realizada uma PCR de 12 ciclos contendo a combinação ideal entre os adaptadores para subsequente geração de clusters e sequenciamento.

O DNA amplificado foi purificado usando o *Agencourt AMPure XP beads* (Beckman Coulter), que elimina fragmentos muito curtos da biblioteca. Após esta etapa, a quantificação de cada biblioteca foi normalizada pela medida equimolar. Antes da geração de clusters as bibliotecas foram normalizadas e quantificadas pelo qPCR usando o *SYBR Fast Illumina Library Quantification Kit* (KAPA Biosystems) a partir das instruções do fabricante. O qPCR foi corrido na *7.500 Fast-Time PCR System* (Applied Biosystems). As condições de ciclagem foram: Um ciclo de desnaturação a 95 °C por 5 min seguido de 35 ciclos de [30 s a 95 °C e 45s a 60 °C].

Após a quantificação das Bibliotecas, submetemos cada amostra ao Bioanalyzer de acordo com instruções do fabricante a fim de obter o tamanho médio da fragmentação dos amplicons para obtenção dos dados mais fieis obtidos pelo Real Time, que nos possibilitou fazer a normalização equimolar entre as amostras gerando um *pool* final de aproximadamente 750pM. Como etapa final adicionamos 10 µL do nosso *pool* em 590 µL do tampão HT1 e carregamos o volume total de 600 µL no Illumina MiSeq clamshell style cartridge kit V2 (500 ciclos). O Esquema metodológico está sumarizado na figura 12.

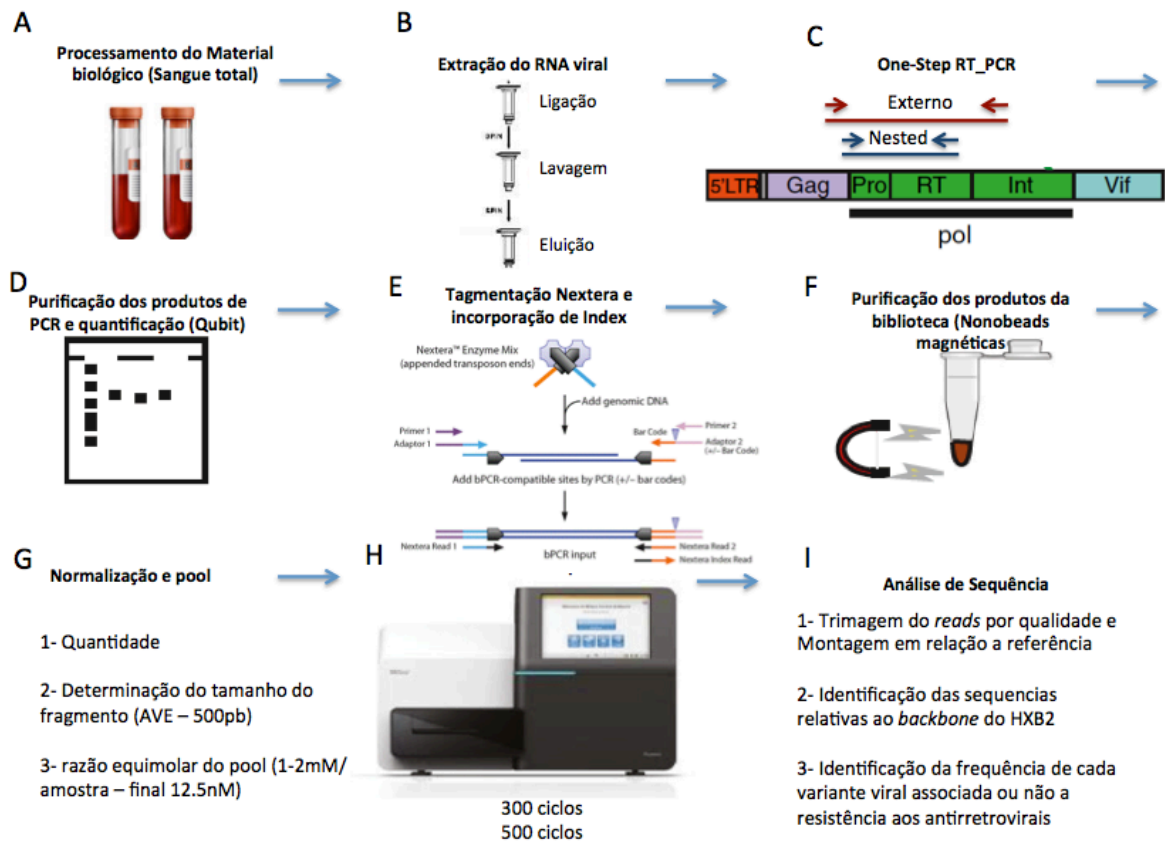


Figura 12 - Resumo metodológico das etapas do sequenciamento de nova geração.

3.12 ANÁLISE DOS DADOS A PARTIR DO SEQUENCIAMENTO DE NOVA GERAÇÃO

Os dados oriundos do sequenciamento de nova geração da plataforma Illumina foi primeiramente analisados usando o software Geneious [112]. A sequencia de referência usada foi o HXB2 (número de acesso #NC_001802) e todas as mutações de resistência foram anotadas manualmente a partir dos *reads* que compuseram a montagem da sequencia consenso.

As sequencias foram desmultiplexadas automaticamente no MiSeq e então emparelhadas num arquivo (.fastq) que compõem cada amostra representadas por *reads forward* e *reverse*. Estes arquivos foram importados no Geneious e então pareadas. Por seguinte as sequencias foram trimadas nas extremidades como parte do processo de montagem. A taxa de erro adotada para montagem foi neste caso de 0.01. As sequências foram então mapeadas frente a sequencia de referencia segundo os parâmetros: Foram permitidos *Gaps* de no máximo 15% por *read*, o máximo de ausência de combinação por

read foi de 25% e a identidade mínima de sobreposição foi de 80%. O mapeador usado foi o Geneious, cuja sensibilidade selecionada foi de média a rápida. Toda variante nucleotídica representada obteve pelo menos 50 vezes de cobertura com frequência mínima de 1%.

Variantes e suas frequências foram exportadas em um documento de excel e então filtramos as posições de aminoácidos conhecidos para então correlacionar com a resistência a drogas baseada na lista do banco de dados disponível em *Stanford Drug Resistance*. A frequência de cada variante, o número de sequências contendo o nucleotídeo mutado bem como a qualidade média da posição onde houve variação nucleotídica foram assinados e obtidos a partir do pacote de programas disponíveis em [112].

3.13 RECONSTRUÇÃO HAPLOTÍPICA

A reconstrução haplotípica foi inferida a partir do software QuasiRecomb. Em resumo o modelo probabilístico implementado em QuasiRecomb é baseado em HMM o qual assume que a diversidade subjacente é gerada através de mutação e recombinação [113]. Para estimar a distribuição haplotípica nós levamos em consideração os valores de PHRED por base e apenas haplótipo com frequência $\geq 1\%$ foram selecionados assim, reduzindo as chances de variantes falso-positivas induzidas por erros experimentais.

3.14 USO DE CÓDON DIFERECIAL

Com o objetivo de investigar se um códon é preferencialmente expresso em detrimento de outro dentro da população de HIV-1, nós calculamos o Uso Relativo de Códon Sinônimos (RSCU), este índice é definido como a razão entre a frequência observada do códon pela frequência esperada dado que todos os códon sinônimos tem a mesma probabilidade de serem usados para o mesmo aminoácido [114]. Os valores de RSCU não tem relação com o uso do aminoácido e a abundância da taxa de códon sinônimos, o qual pode refletir diretamente o viés de uso de códon sinônimos.

O índice de códon adaptativo (CAI) foi usado para estimar a extensão do viés de códon que foi conhecido por ser preferido em genes altamente expressos. Então CAI tem sido largamente usado para medir os níveis de expressão genica. Os valores de CAI variam entre 0 e 1 e quanto maior o valor mais poderoso é o uso do códon e maior os níveis de expressão. Todas essas análises foram realizadas no Mega [115].

3.15 PRESSÃO DE SELEÇÃO

A construção do banco de sequencia para a análise de pressão de seleção foi baseada na no gene *pol* entre os haplótipos reconstruídos, cujo *cutoff* da frequência dos haplótipos foi de 1%. As sequencias foram alinhadas no muscle usando os parâmetros default. Nós avaliamos a pressão de seleção agindo posição a posição ao longo do gene *pol*, e pelo uso de 4 abordagens: SLAC (counting substitutions), FEL (counting substitutions), RFL (random effects) e FUBAR (hierarchical Bayesian method). Nós incluímos o método de MEME para detectar a seleção dos locais individuais seguindo a recomendação do desenvolvedor. Em adição, mudanças em linhagens específicas foram testadas usando as abordagens de BUSTED e aBSREL.

Todas as análises foram realizadas como o software HyPhy no servidor Datamonkey [116].

3.16 MÉTRICA DE DIVERSIDADE

As sequencias obtidas a partir da reconstrução haplotípica foram comparadas pelo uso da distância. Para esta proposta nós estimamos a diferença da proporção de nucleotídeos e aminoácidos. A taxa global dN/dS (seleção positiva e a estatística de Tajima's D, foram realizadas no Mega [117].

4 RESULTADOS

4.1 DADOS CLÍNICOS E EPIDEMIOLÓGICOS

Um conjunto de 2,518 sequências elegíveis oriundas de 53 municípios do estado de São Paulo datadas entre jan 2004 e fev de 2015 foram estudadas. Após o controle de qualidade das sequências, 25 delas foram descartadas (1,1%). Seguida a subtipagem viral, excluímos 233 (10,3%) com algum sinal de recombinação inter-subtipo. Assim, 2.260 sequências de pacientes infectados pelo HIV-1 foram incluídas no estudo. Um total de 1,993 (88%) foram caracterizadas como subtipo B, 185 (8.1%) F1 e 82 (4%) C. Quando comparamos as variáveis epidemiológicas e clínicas entre os subtipos mais prevalentes (Tabela 3), verificamos que a mediana da idade dos indivíduos infectados pelo subtipo C são pelo menos 6 anos mais novos que os indivíduos infectados pelos subtipos B e F1 ($P = 0.02$). Diferenças na taxa de resistência primária envolvendo os diferentes subtipos com uma prevalência de 18.2% para o subtipo F1, 12.2% para o subtipo B e baixa prevalência em torno de 3.8% no subtipo C, porém sem diferenças estatísticas. Quanto ao status clínico, os pacientes infectados pelo subtipo B, C e F1 apresentaram contagem de T CD4+ célula de 268, 395 e 274, respectivamente. ($P = 0.004$).

Quando estratificamos os subtipos segundo o tratamento verificamos que os indivíduos *naives* eram 10 anos mais jovens que os indivíduos sob regime terapêutico para os subtipos B ($P < 0.001$), C ($P = 0.012$) e 6 e anos mais jovens que os tratados dentro do subtipo F1 (0.334) (Figura 13, Tabela suplementar 1). Independente do subtipo viral, a população de pacientes *naive* eram majoritariamente representada pelo grupo MSM, (74%) no subtipo B, (71.7%) C e (57.1%) F1 (Figura 14). A contagem de células T CD4+ entre a População de pacientes não tratados foi bem superior à de pacientes em terapia nos subtipos B, C ($P < 0.001$) e F1 ($P = 0.003$) (Figura 15). Não observamos diferença estatística nos níveis de carga viral (log/mL) entre os grupos tratados e não tratados (Figura 16). A tabela suplementar 1 reúne com mais detalhes os dados analisados).

Tabela 3 – Características dos pacientes envolvidos no estudo.

Variáveis	Subtipo_B n (3708)	Subtipo_C n (300)	Subtipo_F1 n (359)	Total
Mediana da idade (IQR)	38 (29-46)	33 (28-40)	40 (31-45)	38 (29-46)
Gênero, n (%)				
Masculino	1688 (68,1%)	85 (62%)	115 (53%)	1888 (66,6%)
Feminino	791 (31,9%)	52 (38%)	102 (47%)	945 (33,4%)
Fator de risco, n (%)				
Het	825 (52,6%)	69 (61,1%)	58 (54,2%)	952 (53,2%)
HSH	552 (35,2%)	41 (36,3%)	31 (8,6%)	624 (34,9%)
UDI	69 (4,4%)	0 (0%)	3 (2,8%)	72 (4%)
tv	122 (7,8%)	3 (2,7%)	15 (14%)	140 (7,8%)
Population, n (%)				
Caso	1993 (53,7%)	82 (27,3%)	185 (51,5%)	2260 (51,8%)
Controle	1715 (46,3%)	218 (72,7%)	174 (48,5%)	2107 (48,2%)
Terapia, n (%)				
Naive	775 (28,8%)	140 (78,7%)	62 (26,8%)	977 (31,5%)
Tratado	1919 (71,2%)	38 (21,3%)	169 (73,2%)	2126 (68,5%)
Cadeias, n (%)				
Clusters	797 (21,5%)	72 (24%)	74 (20,6%)	943 (21,6%)
Non clusters	2911 (78,5%)	228 (76%)	285 (79,4%)	3424 (78,4%)
Tipo de cluster				
pares	464 (58,2%)	34 (47,2%)	40 (54,1%)	538 (57,1%)
>2 seq clusters	333 (41,8%)			405 (42,9%)
Mutações CPR, n (%)	2150 (58%)	50 (16,7%)	209 (58,2%)	2409 (55,2%)
Perfil das mutações CPR, n (%)				
Ips	1057 (28,5%)	13 (4,3%)	91 (25,3%)	1161 (26,6%)
NRTs	1923 (51,9%)	38 (12,7%)	177 (49,3%)	2138 (49%)
NNRTs	1257 (33,9%)	32 (10,7%)	125 (34,8%)	1414 (32,4%)
Mutações não CPR, n (%)				
T74S	112 (3%)	16 (5,3%)	21 (5,8%)	159 (3,6%)
E138A	88 (2,4%)	26 (8,7%)	19 (5,3%)	133 (3%)
E138K	8 (0,2%)	6 (2%)	2 (0,6%)	16 (0,4%)
Mediana do CD4 (IQR)	279 (139-466)	419 (223-595)	311 (177-504)	286 (144-474)
Mediana CV (log) (IQR)	4,38 (3,79-4,95)	4,52 (3,77-5,17)	4,42 (3,70-4,98)	4,39 (3,79-4,96)

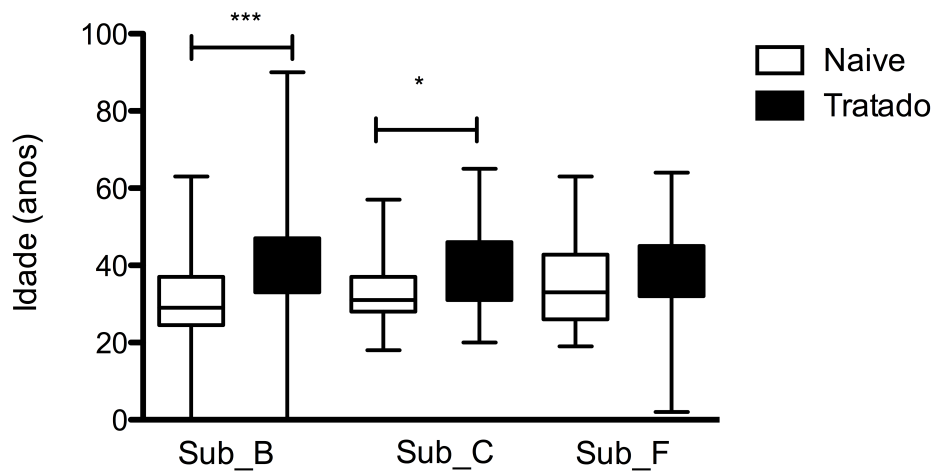


Figura 13 - Análise que representa a mediana da idade entre os diferentes subtipos B, C e F1 de HIV-1 levando em consideração a terapia: Naive – pacientes não expostos à terapia antirretroviral e tratados- pacientes previamente expostos à terapia antirretroviral. O valor significativo está representado por * ($p < 0.001$).

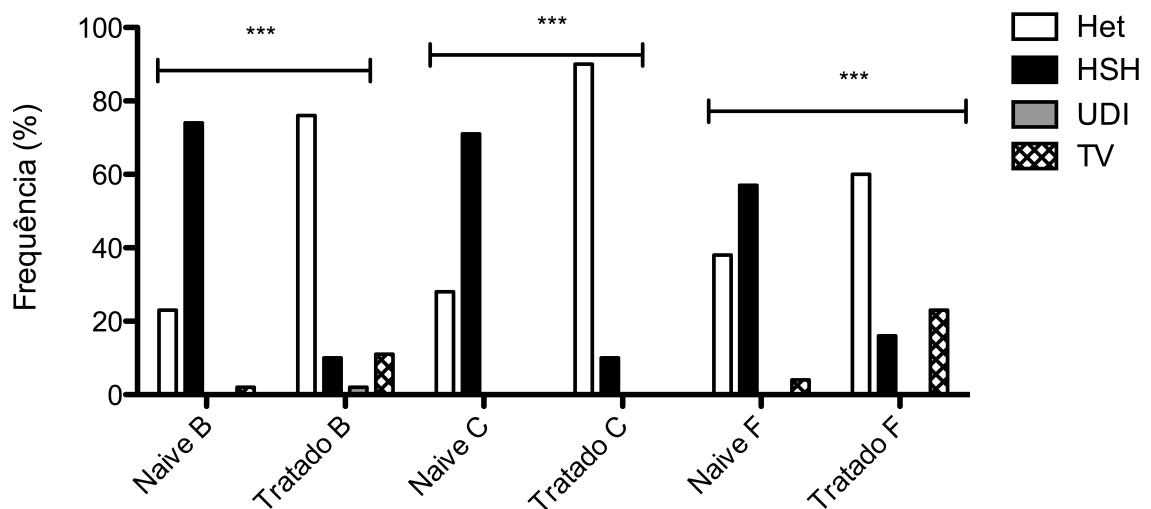


Figura 14 – Análise representativa da frequência das categoria de exposição ao risco nos subtipos B, C e F1. Het heterossexuais, HSH- homens que fazem sexo com homens, UDI- usuários de drogas injetáveis, TV- transmissão vertical. Valores significativos estão representados por * (< 0.001).

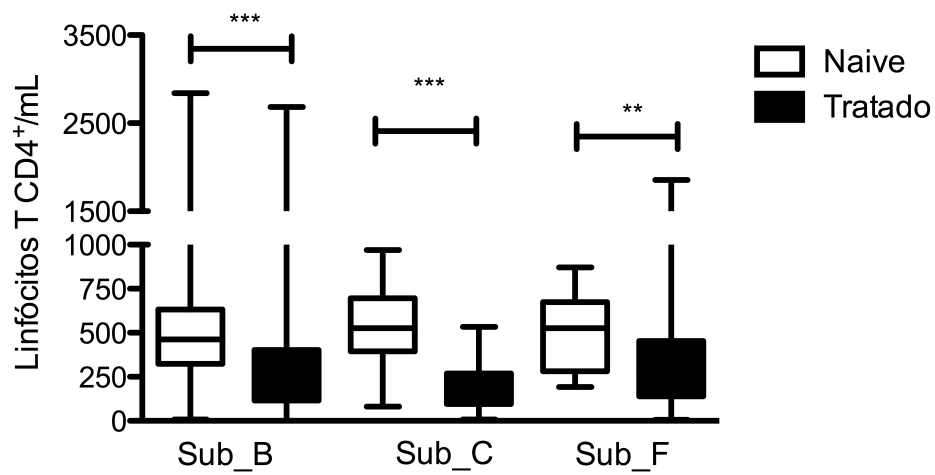


Figura 15 - Análise que representa a mediana da contagem de linfócitos T CD4⁺ nos grupos *naive* e tratados nos subtipos B, C e F1. Os valores significativos está representado por * ($p < 0.001$).

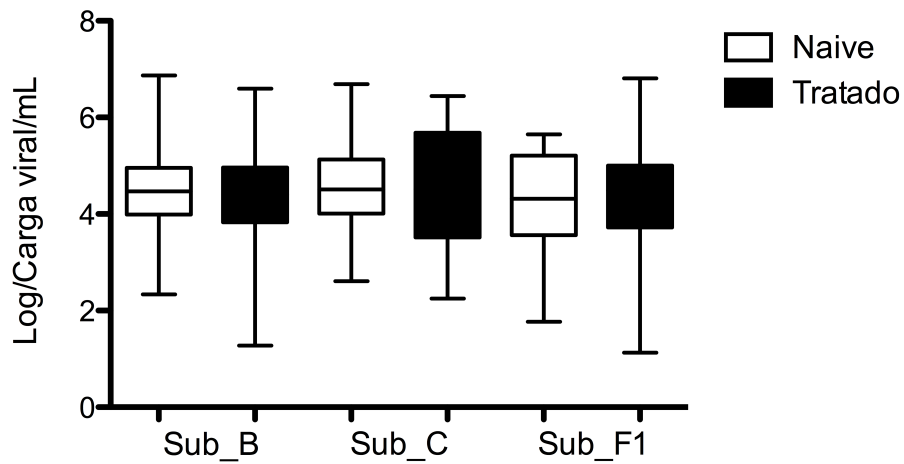


Figura 16 - Análise que representa a mediana da carga viral em log/mL entre os indivíduos que os grupos *naive* e tratado nos subtipos B, C e F1.

4.2 FATOR DE RISCO

Quando estratificamos a população do estudo segundo a categoria de risco (HSH- Homens que fazem sexo com homens e Não-HSH- heterossexuais, usuário de drogas injetáveis e transmissão vertical), notamos que em todos os subtipos a população de HSH era mais jovem, porém apenas no subtipo B essa diferença era significativa, onde a mediana da idade era 8 anos mais jovem que da população Não-HSH ($P < 0.001$) (Figura 17). Sendo a população de HSH majoritariamente composto por pacientes virgens de tratamento.

Um dado que merece destaque é que a população composta por HSH do subtipo B apresentou uma taxa de resistência primária alta em torno de 8.5% ao passo que a população de indivíduos Non-MSM apenas 1.5% ($P < 0.001$). O mesmo cenário foi observado entre os indivíduos do subtipo F, onde a taxa de resistência primária foi de 21% entre os MSM e 0% entre a população Non-MSM ($P = 0.006$). Para o subtipo C a taxa de resistência primária foi baixa e sem diferença entre os grupos (Figura 18). Independente do subtipo viral os níveis de T CD4+ foram superiores entre os indivíduos do grupo MSM apontando significância estatística apenas no subtipo B ($P < 0.001$) (Figura 19). Os dados pertinente aos grupos HSH e Não-HSH podem ser observados em maior detalhe na tabela suplementar 2.

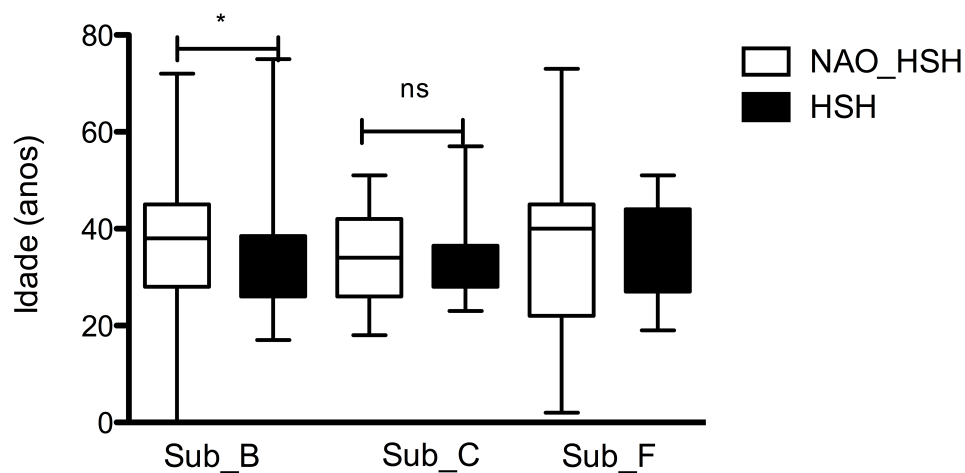


Figura 17 - Análise que representa a mediana da idade entre os diferentes subtipos B, C e F1 de HIV-1 levando em consideração os dois grupos HSH (homens que fazem sexo com homens) e Não_HSH (heterossexuais, usuários de drogas e transmissão vertical). O valor de p significativo está representado por * ($p < 0.001$).

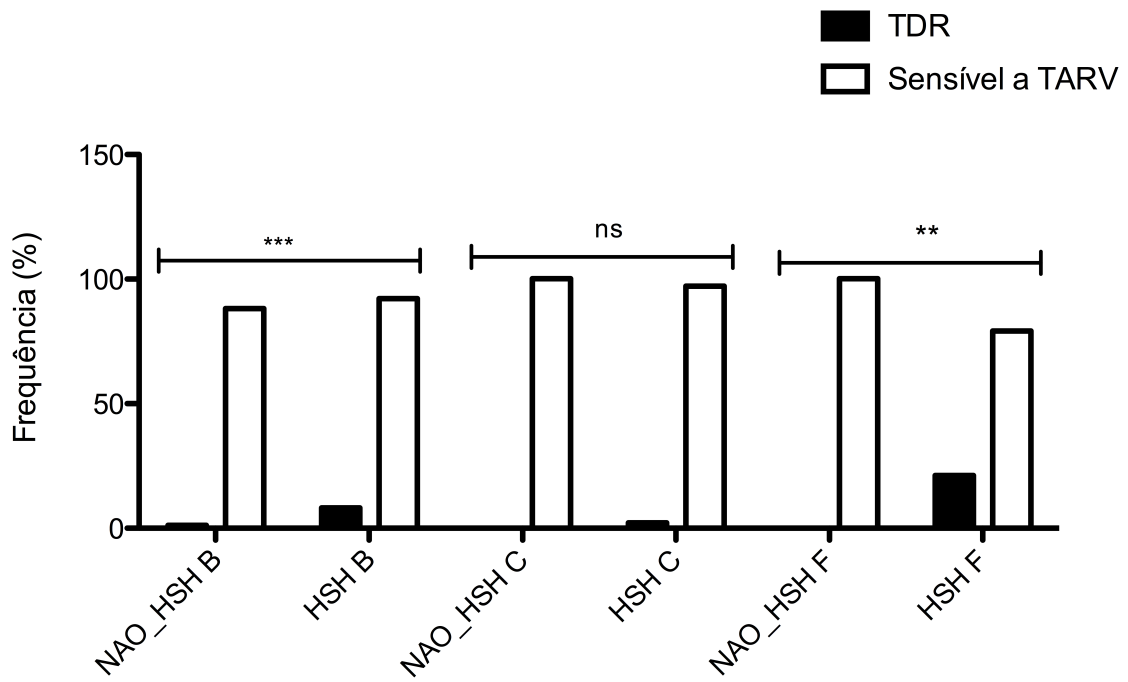


Figura 18 - Análise que representa a frequência da resistência a drogas transmitida (TDR) entre os grupos de HSH e Não-HSH nos subtipos B, C e F1. Valores significativos estão representados por * (<0.001 e 0.006), respectivamente.

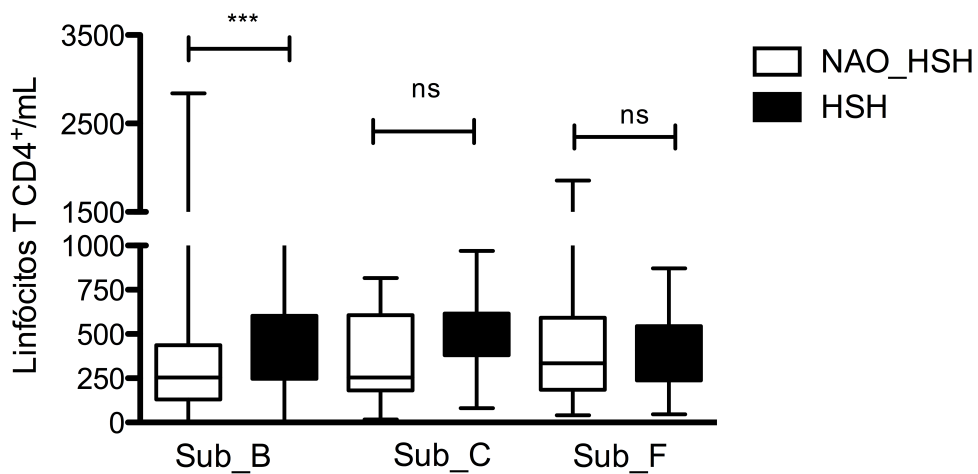


Figura 19 - Análise que representa a mediana da contagem de linfócitos T CD4⁺ nos grupos HSH e Não-HSH nos subtipos B, C e F1. O valor significativo está representado por * (p <0.001).

4.3 INFERÊNCIA DOS CLUSTERS DE TRANSMISSÃO

A etapa deste estudo que concerne análise e inferência de *clusters* de transmissão foi baseada em reconstruções filogenéticas da região da polimerase viral, especificamente (PR/RT), contemplando um fragmento de aproximadamente 860 nucleotídeos. Ressaltamos que para evitar o viés de agrupamento removemos os códons associados a resistência. O critério de reconstrução foi o Maximum likelihood (ML), com método de substituição nucleotídica de GTR.

As análises filogenéticas pertinentes a cada subtipo B, C e F1 encontram-se nas figuras 20, 21 e 22, representadas por árvores circulares e enraizadas. Nestas, podemos observar as amostras objeto deste estudo se agruparem de forma consistente com as respectivas sequências puras de cada subtipo viral. Foram adotados parâmetros que identificam putativos *clusters* de transmissão baseados no suporte dos ramos (*bootstrap* ≥ 90 e distância genética ≤ 0.06). Os respectivos *clusters* podem ser identificados como clados assinalados em vermelho nas árvores filogenéticas.

Usando a nossa definição de *cluster*, 18,3% (414/2.260) das sequências se clusterizaram com uma ou mais sequências. Estes *clusters* foram menos pronunciados nos subtipos B (17,7%) e F1 (15,1%) comparados com o subtipo C (40,2%) apontando diferenças estatísticas de conectividade entre os subtipos ($P < 0.001$) (Tabela 3). A idade também foi significativamente associada a *clusters* para os subtipos B e C. A idade foi também estatisticamente significante entre os subtipos B e C. A mediana da idade dos indivíduos clusterizados no subtipo B (33 yo) 6 anos mais jovens que os indivíduos não clusterizados (39 yo) ($P < 0.001$). A mediana da idade entre os indivíduos em cluster do subtipo C foi de (30 yo), sendo 10 anos mais jovens que os indivíduos não clusterizados ($P < 0.001$) (Figura 23).

Vale destacar que dentro do subtipo B, 38,5% dos pacientes *naives* estavam dentro da rede transmissão contra apenas 12,1% dos indivíduos tratados ($P < 0.001$). Na População do subtipo C (59,6%) dos *naives* estavam em redes de transmissão contra apenas 4,2% dos tratados em *clusters* ($P < 0.001$) e por fim, na população do subtipo F1 36,4% dos *naives* formavam redes de transmissão contra apenas 10% dos tratados ($P = 0.004$) (Figura 24).

Independente do subtipo viral, indivíduos identificados como HSH foram mais propensos a se agruparem que Het, UDI ou TV. Dos putativos *clusters* detectados pelo subtipo B 51% das sequências foram identificadas como MSM, 41,4% Het, 6% VT e 1,3%

UDI ($P < 0.001$). No subtipo C 85.7% dos *clusters* eram compostos por sequências identificadas como HSH e 14.3% Het ($P < 0.001$) Para o subtipo F1 63.6% eram HSH e 36.4% Het ($P = 0.011$) (Tabela suplementar 3). Ressaltando que dentro do grupo de não *clusters* apontava um perfil oposto, sendo representada majoritariamente por indivíduos heterossexuais (Figura 25).

A taxa de TDR dentre os indivíduos que se clusterizaram foi muito superior aos que não se clusterizaram nos subtipos B ($P > 0.001$) e F1 ($P = 0.009$). Não observamos diferenças da taxa de TDR entre os grupos de *clusters* e não *clusters* para o subtipo C (Figura 26). Somente no subtipo C houve diferença na contagem de células T CD4+ entre os indivíduos em cluster e não clustering. Sendo duas vezes maior entre os indivíduos em clustering ($P = 0.003$) (Figura 26).

Com relação aos níveis de carga viral não observamos diferenças estatísticas entre o grupo indivíduos que estavam em *clusters* e o que estavam fora de *clusters*. (Figura 27).

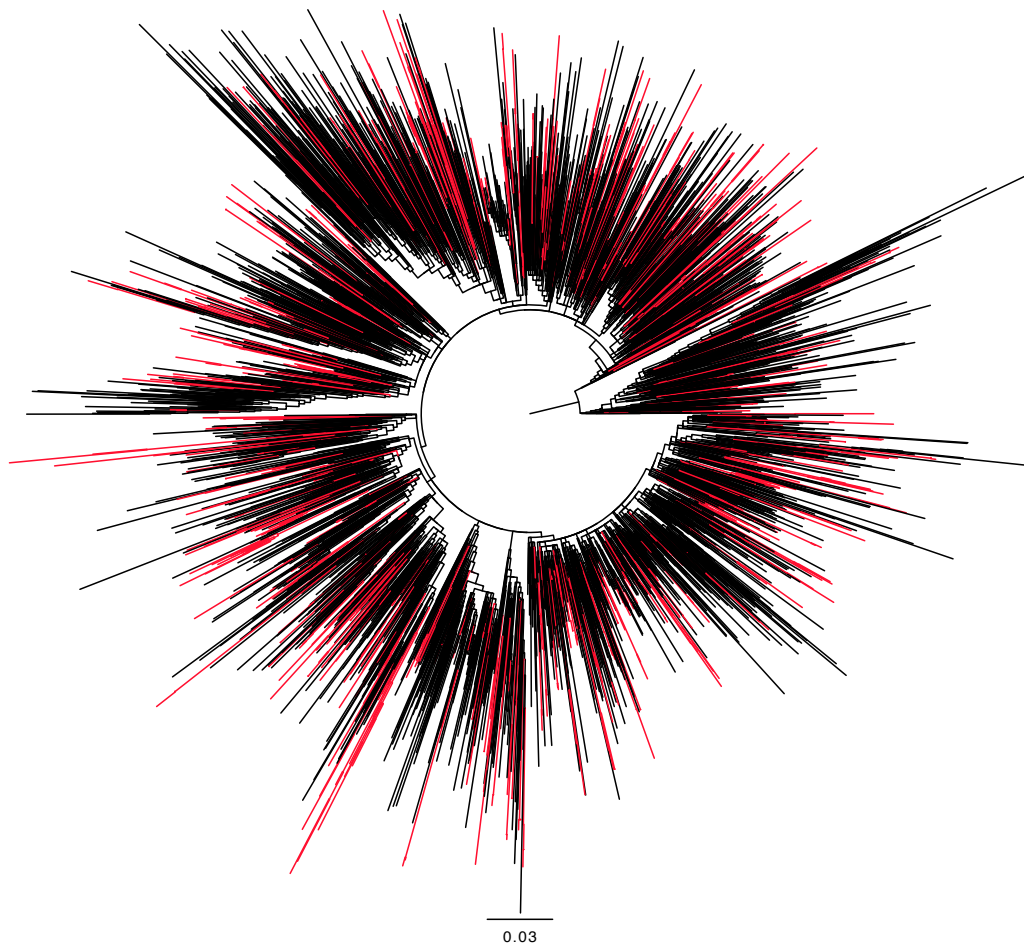


Figura 20 - Análise filogenética correspondente ao gene *pol* de 3.708 sequências do subtipo B de HIV-1. As sequências foram alinhadas frente sequências controles globais oriundas do *blast* coletadas a partir do banco de dados de Los Alamos. Três sequências de referência do subtipo C puro foram usadas como grupo externo. Valores de *Bootstrap* igual ou superior a 90% e distância genética inferior a 0.06 estão representadas na árvore como *clusters* assinalados em vermelho. A barra de escala indica 3% de divergência na sequência nucleotídica.

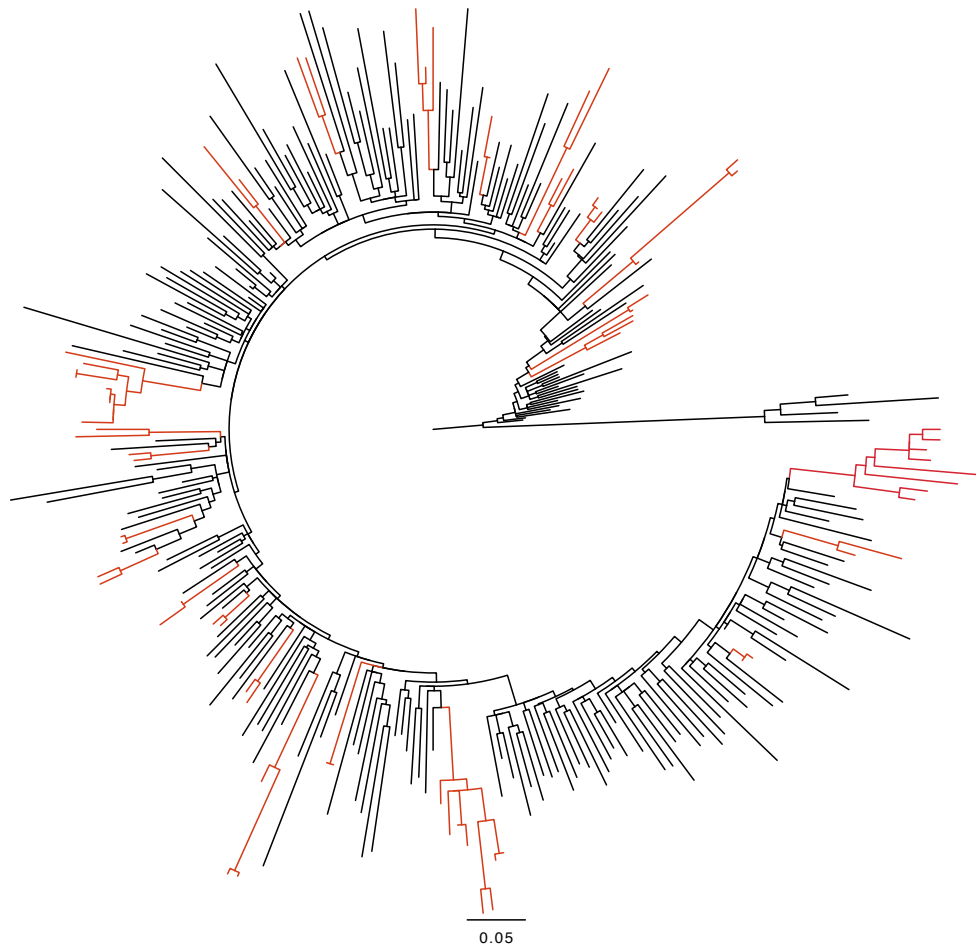


Figura 21 - Análise filogenética correspondente ao gene *pol* de 300 sequências do subtipo C de HIV-1. As sequências foram alinhadas frente sequências controles globais oriundas do *blast* de coletadas a partir do banco de dados de Los Alamos. Três sequências de referência do subtipo B puro foram usadas como grupo externo. Valores de *Bootstrap* igual ou superior a 90% e distância genética inferior a 0.06 estão representadas na árvore como *clusters* assinalados em vermelho. A barra de escala indica 5% de divergência na sequência nucleotídica.

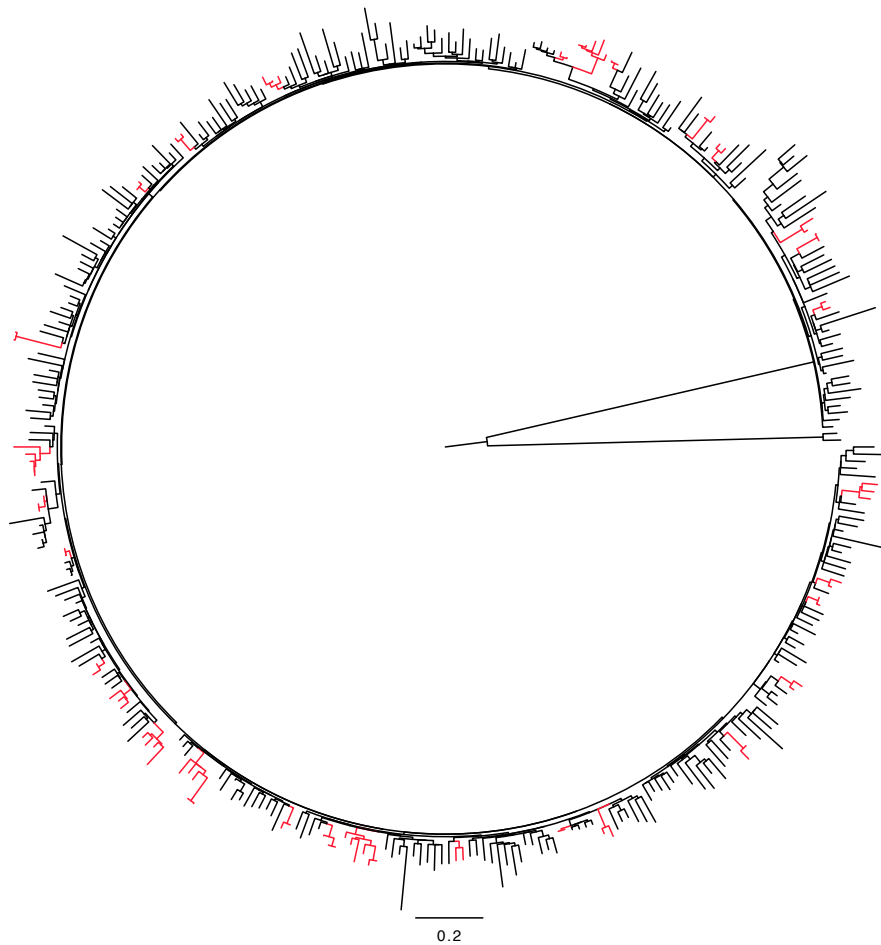


Figura 22 - Análise filogenética correspondente ao gene *pol* de 359 sequências do subtipo F1 de HIV-1. As sequências foram alinhadas frente sequências controles globais oriundas do *blast* de coletadas a partir do banco de dados de Los Alamos. Três sequências de referência do subtipo B puro foram usadas como grupo externo. Valores de *Bootstrap* igual ou superior a 90% e distancia genética inferior a 0.06 estão representadas na árvore como clusters assinalados em vermelho. A barra de escala indica 2% de divergência na sequência nucleotídica.

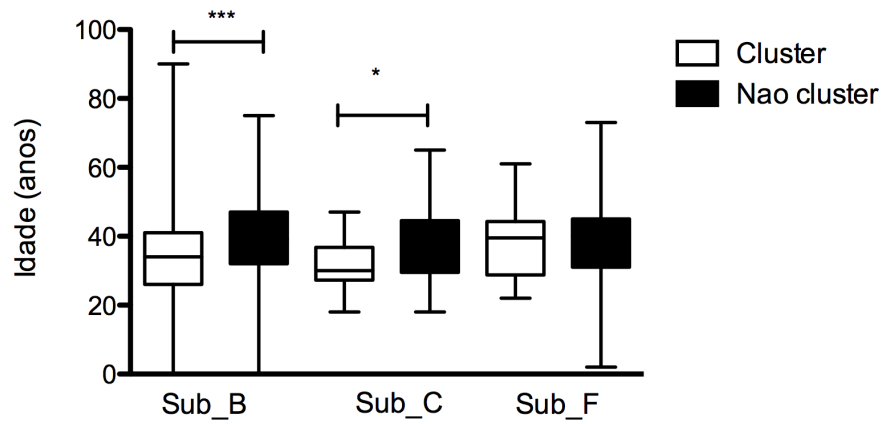


Figura 23 - Análise que representa a mediana da idade entre os indivíduos que estavam em clusters de transmissão e fora de clusters nos subtipos B, C e F1 de HIV-1. Os valores significativos de p estão representados por * ($p < 0.001$ e 0.037), respectivamente.

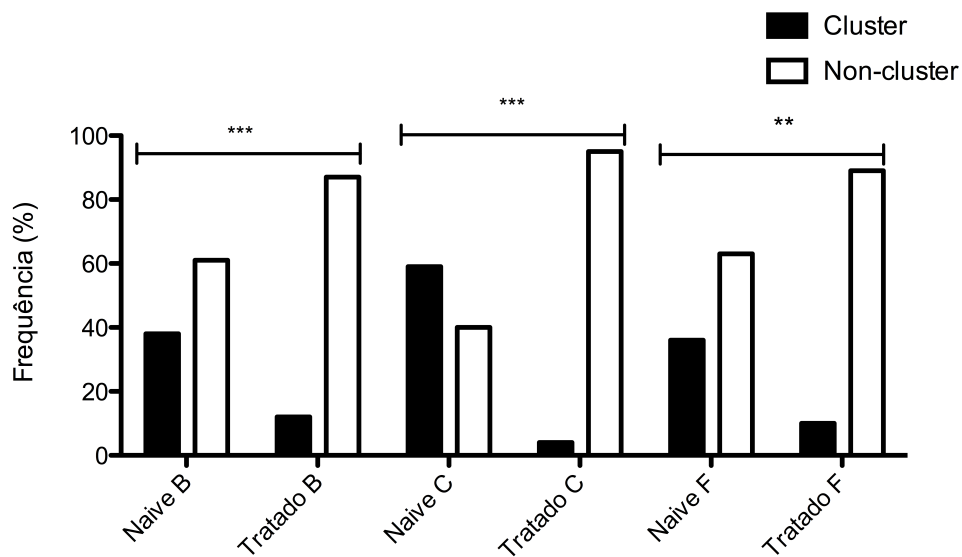


Figura 24 - Análise que representa a frequência dos indivíduos que estão em cluster e fora de cluster nos grupos de pacientes *naives* e tratados nos subtipos B, C e F1. Valores significativos de p estão representados por * (< 0.001 , < 0.001 e 0.004), respectivamente.

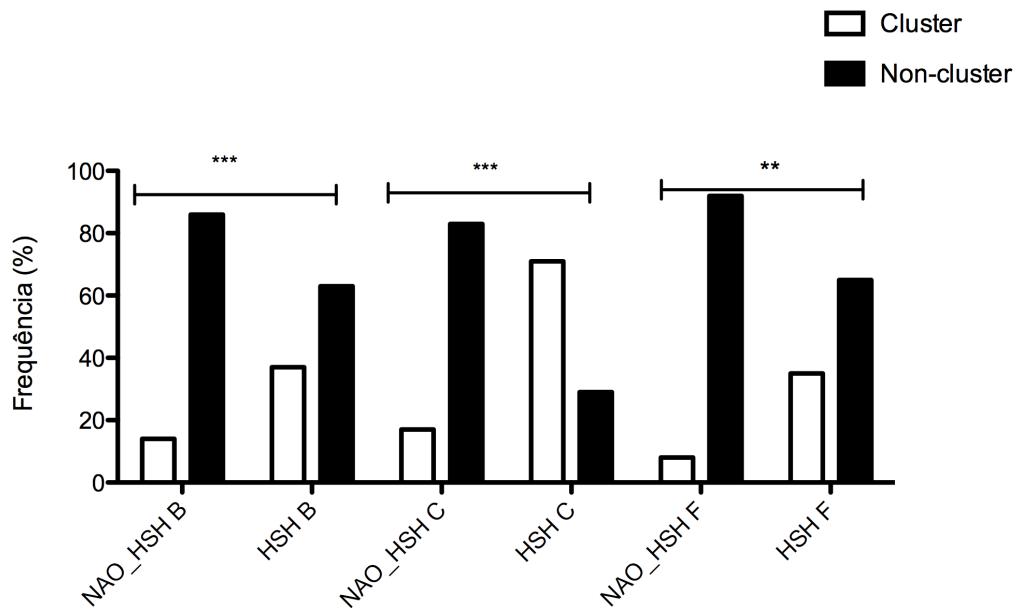


Figura 25 - Análise que representa a frequência dos indivíduos que estão em cluster e fora de clusters nos grupos de HSH e Não-HSH entre os subtipos B, C e F1. Valores significativos de p estão representados por * (<0.001 , <0.001 e 0.008), respectivamente.

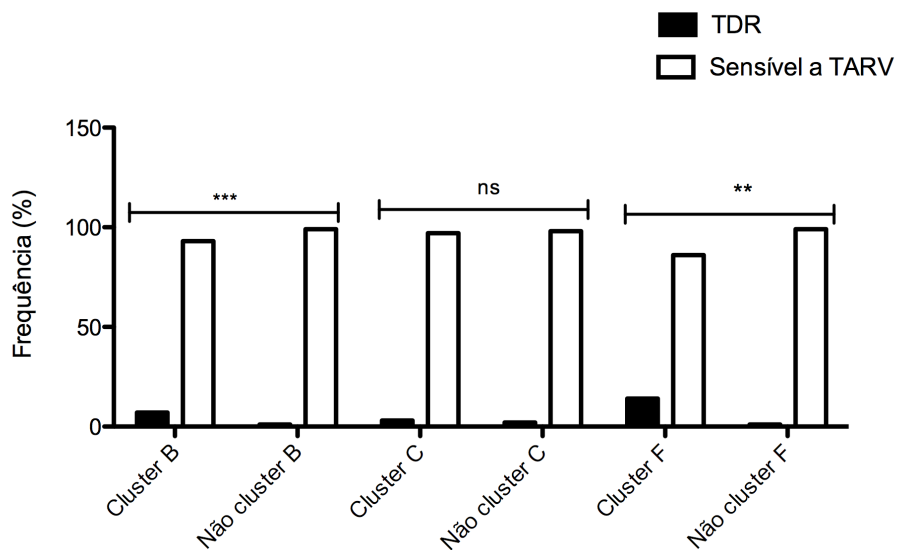


Figura 26 - Análise que representa a frequência da resistência a drogas transmitida (TDR) entre os indivíduos que estão em cluster e fora de cluster nos subtipos B, C e F1. Valores significativos estão representados por * (<0.001 e 0.009).

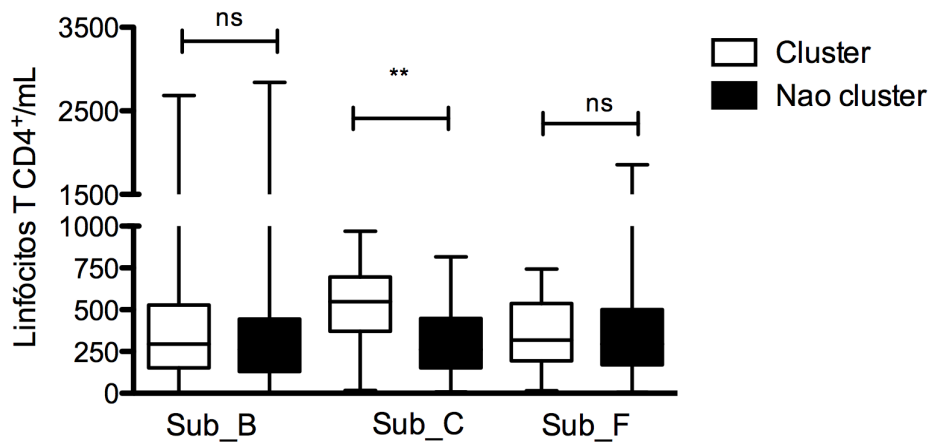


Figura 27 - Análise que representa a mediana da contagem de linfócitos T CD4⁺ entre os indivíduos que estão em cluster e fora deles nos subtipos B, C e F1. O valor significativo está representado por * ($p < 0.001$).

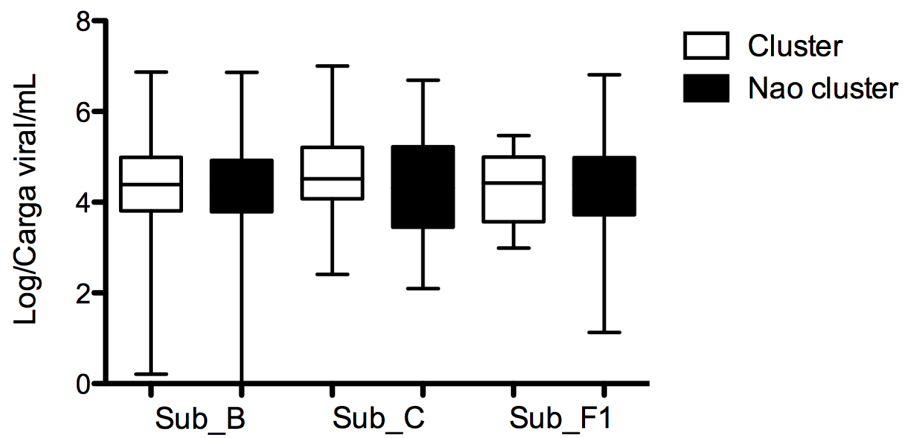


Figura 28 - Análise que representa a mediana da carga viral em log/mL entre os indivíduos que estão em cluster e fora deles nos subtipos B, C e F1.

4.4 DINÂMICA GEOGRÁFICA DOS CLUSTERS

Nós definimos clusters endógenos quando pelo menos 66% dos membros que os compõem são da mesma localidade geográfica, significando que clusters compostos por dois indivíduos, ambos devem ser da mesma localidade, e clusters externos são compostos de menos de 66% dos casos locais. Sob esta luz, nós tentamos inferir sobre a origem dos clusters de transmissão do HIV-1 entre os subtipos mais prevalentes no estado de São Paulo.

É neste cenário que verificamos que mais de 81% das infecções do subtipo B aconteceram dentro do estado de São Paulo. Quando nós consideramos clusters externos, nós podemos observar que 11% são associados com clusters de outros estados brasileiro e 8% das infecções com sequencias de fora do país. No contexto do subtipo C verificamos que 64% das transmissões aconteceram dentro do estado ao passo que 18% estão relacionadas a clusters de outros estados brasileiros e 18% a clusters de outros países. Quanto ao subtipo F1 menos de 60% das transmissões aconteceram dentro do estado, 30% aconteceram em outros estados brasileiros e apenas 11% das transmissões associados a outros países (Figura 29A).

Para melhor entender micro-epidemias no estado de São Paulo, dividimos o mapa do estado segundo o conceito geopolítico, mantendo unido os municípios mais relacionados. Verificamos que a área 1 (Nordeste, NE) e 4 (Metropolitana, Metro), que estão mais bem representada no estudo apresentaram um perfil muito semelhante, onde aproximadamente 60% das sequencias do subtipo B estão relacionados à clusters endógenos, sugerindo transmissões dentro dessas regiões. Quanto ao subtipo C, na área NE, mais de 30% das transmissões do aconteceram dentro desta localidade, ao passo, que na área metropolitana mais de 50% das transmissões ocorreram dentro desta localidade. Quanto ao subtipo F1 houve uma menor presença de clusters endógenos nas referidas áreas 1 e 4. No que diz respeito as áreas 2 (Noroeste, NW) e 3 (Sul, SU) apresentaram um perfil muito semelhante, onde cerca de 50% das sequencias do subtipo B formaram clusters endógenos, mas sem identificação de clusters endógenos para os subtipos C e (Figura 29B).

A área metropolitana de São Paulo contemplou o maior número de amostras e por representar a região mais importante dentro do Estado foi dividida em 3 áreas (região da periferia, Região do ABCD e cidade de São Paulo).

A região periférica da grande São Paulo (27C- A), mostrou uma diferença de perfil de origem de clusters de transmissão, comparada as regiões (27C-B e 27C-C). Nós

encontramos nesta região populosa um elevado nível de clusters endógenos para os subtipos B e C. Merece destacar que a região da cidade de São Paulo teve 46% de clusters endógenos do subtipo C. Interessantemente, a região periférica é responsável por 33% dos clusters endógenos do subtipo F1, similar ao observado na área NE.

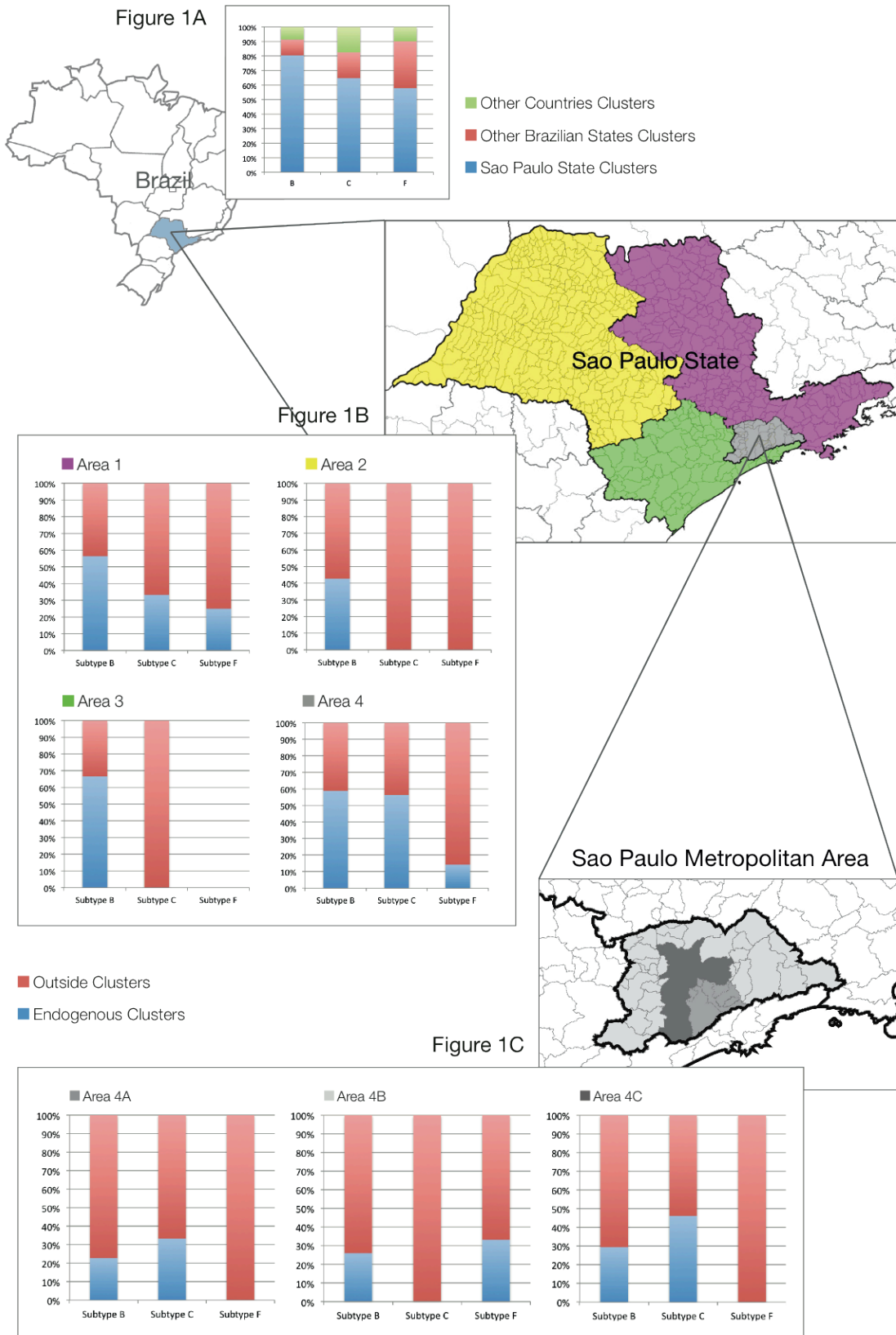


Figura 29 – Tipos de clusters geográficos identificados em pacientes do estado de São Paulo. A figura retrata a composição dos clusters pelos subtipos circulantes de acordo com a origem das sequências que compõem os clusters como (1) clusters endógenos (barras em azul)- clusters que incluem pelo menos 66% das sequências compostas por membros de uma mesma localidade geográfica, (2) clusters de fora (barras em

vermelho), compostos por menos de 66% dos casos locais. Figura 29 b (B) mostra a divisão geográfica para a análise das redes de transmissão do HI-1 em 4 áreas menores. A zona destacada em púrpura representa a área 1, amarelo área 2, verde área 3 e cinza área 4. (C) Um zoom na área metropolitana de São Paulo (zona 4). Cinza claro representa a região periférica da área metropolitana, cinza intermediário representa região do ABCD e cinza escuro representa a cidade de São Paulo.

4.5 ANÁLISE DO SEQUENCIAMENTO DE NOVA GERAÇÃO

Um sub-conjunto de 24 amostras de indivíduos de diagnóstico recente com um perfil de infecção recente foi estudado na região da polimerase (PR/RT) (Tabela 4) que é a região gênica responsável pela síntese das proteínas virais do HIV-1, além de ser alvo da terapia antirretroviral de alta potencia. É neste contexto que se faz útil a análise minudente do fragmento parcial do gene, a fim de detectar variantes minoritárias com mutação de resistência à TARV.

Segundo os nossos dados 79% dos pacientes foram infectados com o subtipo B e em menor escala 21% dos pacientes pelo subtipo C. A maioria das infecções ocorreram através do contato homossexual (87%) no subtipo B e (75%) no subtipo C. Baseado no teste de avidéz previamente conduzido pelo nosso grupo 9 amostras foram identificadas como incidente (0-6 meses de infecção), 9 amostras como prevalentes não tratados (infecção superior a 6 meses) e 5 amostras que resultaram do *follow-up* dos pacientes, porém já sob regime terapêutico (Tabela 4).

Nesta etapa do estudo sequenciamos 24 amostras derivadas do plasma pelo método *amplicon*, um fragmento de aproximadamente 1322pb (posição relativa ao HXB2 2252-3574). Com base no sequenciamento genético de nova geração (NGS) todas as amostras obtiveram uma significativa proporção de *reads* mapeados em razão do número total de *reads*. Nós obtivemos uma cobertura média por base > 6.932 dentro dos genomas montados frente a sequência de referência de HIV-1 (protótipo HXB2) (Tabela 5).

Tabela 4 – Características dos pacientes estudados pelo sequenciamento de nova geração

Amostra	Ano	Clado	Origem do paciente	Fator de risco	Último negativo	Primeiro positivo	Estágio da infecção	CV/log cópias/mL	CD4	Tratamento
317_11	2011	B	Santo André	HSH	15/09/11	14/10/11	Incident	6,33	152	Naïve
70_13	2013	B	Santo André	HSH	01/11/11	10/12/12	Prevalent	5,43	967	Naïve
89_13	2013	C	Santo André	HET	09/03/12	21/01/13	Prevalent	3,95	644	Naïve
104_13	2013	B	Santo André	HET	15/09/12	26/02/13	Incident	>LIM	607	Naïve
FO1_70_13_219_13	2013	B	Santo André	HSH	01/11/11	10/12/12	Prevalent	2,84	MD	Tratado
258_13	2013	C	Santo André	HSH	15/06/212	25/04/13	Prevalent	4,11	MD	Naïve
FO1_104_13_272_13	2013	B	Santo André	HET	15/09/12	26/02/13	Incident	3,68	MD	Tratado
FO1_89_13_406_13	2013	C	Santo André	HET	09/03/12	21/01/13	Prevalent	3,84	770	Tratado
451_13	2013	B	Santo André	HSH	29/11/12	02/08/13	Incidente	5,51	391	Naïve
599_13	2013	B	Sao Paulo	HSH	MD	04/03/13	Prevalent	4,59	965	Naïve
600_13	2013	B	Sao Paulo	HSH	07/09/12	10/09/13	Incident	4,16	655	Naïve
FO1-258_13_639_13	2013	C	Santo André	HSH	15/06/212	25/04/13	Prevalent	MD	MD	Tratado
864_13	2013	B	Santo André	HSH	15/10/11	02/12/13	Prevalent/Chronic	6,48	44	Naïve
FO1_864_13_113_14	2014	B	Santo André	HSH	15/10/11	02/12/13	Prevalent/Chronic	2,38	110	Tratado
137_14	2014	C	Santo André	HSH	15/07/13	20/01/14	Incident	4,08	479	Naïve
139_14	2014	B	Santo André	HSH	13/03/13	29/01/14	Prevalent	5,09	649	Naïve
169_14	2014	B	Santo André	HSH	15/06/13	30/01/14	Incident	MD	1405	Naïve
582_14	2014	B	Santo André	HSH	MD	MD	Incident	5,1	432	Naïve
650_14	2014	B	Santo André	HET	MD	MD	Prevalent	4,35	1234	Naïve
717_14	2014	B	Sao Paulo	HSH	MD	15/01/14	Incident	3,96	599	Naïve
750_14	2014	B	Sao Paulo	HSH	MD	15/08/13	Prevalent	4,1	479	Naïve
760_14	2014	C	Sao Paulo	HSH	MD	15/04/2014	Prevalent	5,94	429	Naïve
1270_14	2014	B	Santo André	HSH	MD	MD	Incident	4,57	MD	Naïve
1534_14	2014	B	Santo André	HSH	MD	MD	Incident	5,11	MD	Naïve

HSH- Homens que fazem sexo com homens, Het- heterossexual, Incidente- infecção 0-6 meses, prevalente- infecção superior a 6 meses.

Tabela 5 – Sumário da composição dos *reads* obtidos após sequenciamento de nova geração.

Nome	N of reads	cobertura média	Trimmed (%)	reads mapeados	reads não mapeados	PHRED score
317_11	138.251	16.628	99.3	136.748	1.503	≥30
70_13	121.776	4.312	97.2	41.156	80.620	≥30
89_13	129.924	5.111	99.6	127.353	2.571	≥30
104_13	125.962	4.286	99.7	35.747	90.215	≥30
FO1_70_13_219_13	92.190	5.609	98.8	55.211	36.979	≥30
258_13	105.474	4.527	99.3	49.794	55.680	≥30
FO1_104_13_272_13	85.694	7.780	97.9	83.055	2.639	≥30
FO1_89_13_406_13	159.856	6.176	98.9	155.840	4.016	≥30
451_13	108.608	3.223	99.7	39.868	68.740	≥30
599_13	127.680	5.139	99.8	53.128	74.552	≥30
600_13	109.122	3.318	98.7	74.615	34.507	≥30
FO1-258_13_639_13	102.492	1.263	99.8	71.965	30.527	≥30
80_14	42.154	5.111	99.7	41.813	341	≥30
864_13	134.526	4.322	98.4	52.077	82.449	≥30
FO1_864_13_113_14	134.526	5.518	99.7	52.058	82.468	≥31
137_14	54.644	6.008	99.8	53.902	742	≥30
139_14	116.110	4.682	99.8	60.754	55.356	≥30
169_14	13.386	1.536	98.9	13.260	126	≥30
582_14	61.180	6.587	99.2	60.769	411	≥30
650_14	57.648	1.740	99.7	17.307	40.341	≥30
717_14	185.520	25.205	99.5	178.582	6.938	≥30
723_14	497.137	15.054	97.2	489.084	8.053	≥30
750_14	53.389	7.282	99.6	52.135	1.254	≥30
760_14	240.988	15.498	98.9	145.709	95.279	≥30
1270_14	226.052	6.959	99.9	223.895	2.157	≥30
1534_14	78.546	7.358	99.6	76.830	1.716	≥30

4.6 ANÁLISE DE RESISTÊNCIA ASSOCIADA A DROGAS

A partir da análise comparativa das abordagens de sequenciamento genético, observamos que a plataforma Sanger identificou 4 códons de resistência associados à drogas não análogos de nucleosídeos (NNRT) em 4 amostras dentre eles: K103N em 3 amostras (70_13, FO1_70_13, 1534_14), P225H em 1 amostra (70_13), G190S em 1 amostra (FO1_864_13_113_14). Identificamos 2 códons associados à resistência análogo de nucleosídeos (NRT) em dois pacientes: M184V na amostra (FO1_70_13_219_13) e K65R na amostra (FO1_864_13_113_14).

As mutações avaliadas a partir do sequenciamento largamente paralelizado identificamos todas as mutações observadas no sanger nas respectivas amostras, além da mutação K103N na amostra FO1_104_13_219_13 que estava presente em 1.6% dos *reads* com valor de Phred de 38.

O sequenciamento de nova geração conseguiu detectar, também, um maior número de amostras com resistência à drogas para os NRTs. A mutação M184V esteve presente na amostra 70_13 em 18.8% dos *reads* com valor de Phred de 36. A mutação K65R foi detectado na amostra 137_14 em 1.1% dos *reads* com valor de Phred de 36. A mutação D67N esteve presente na amostra 760_14 em 4% dos *reads* com valor de Phred de 37. Por fim, detectamos a mutação relacionada aos inibidores de protease (IP) M46L na amostra 139_14, porém em uma frequência inferior a 1%. Os dados estão sumarizados na tabela 6.

Tabela 6 – Lista de mutações detectadas pelas duas plataformas de sequenciamento genético.

Nome	Perfil de mutações			Mutações minoritárias detectadas pelo Illumina
	DRM (PI)	DRM (NRT)	DRM (NNRT)	
317_11				
70_13		M184V	K103N,P225H	M184V - ATG ->GTG ; Cob 4615 (18,8%); phred 36 (transicao)
89_13				
104_13				
FO1_70_13_219_13		M184V	K103N	
258_13				
FO1_104_13_272_13			K103N	K103N - AAA->AAC; Cob 10078 (1,6%) phred 38 (transicao)
FO1_89_13_406_13				
451_13				
599_13				
600_13				
FO1-258_13_639_13				
864_13				
FO1_864_13_113_14		K65R	G190S	
137_14		K65R		K65R - AAA->AGA; Cob 6479 (1,1%) phred 36
139_14	M46L			M46L PRESENTE MENNOS DE 1%
169_14				
582_14				
650_14				
717_14				
750_14				
760_14		D67N		D67N - GAT->AAT; Cob 12650 (4%), phred 37
1270_14				
1534_14			K103N	

DRM- mutação de resistência à drogas. PI- inibidor de protéase, NRT- análogo de nucleosídeo da transcriptase reversa, NNRTA- não análogo de nucleosídeo da transcriptase reversa. As mutações assinaladas em negrito foram detectadas em ambas plataformas.

4.7 RECONSTRUÇÃO HAPLOTÍPICA E MÉTRICA DE DIVERSIDADE

Nesta etapa do estudo consistiu da identificação da composição das variantes virais intra-hospedeiro, é neste contexto que foi possível detectar o potencial da diversidade genética presente dentro de cada indivíduo infectado. Do conjunto de pacientes estudados (n=24) foi possível detectar o conjunto de 130 variantes virais. No estudo, foi possível identificar uma flutuação da frequência haplotípica entre os pacientes (1-19 variantes). A mediana da frequência haplotípica detectada foi de 6. Uma vez reconstruído as variantes virais, foi possível traçar a relação filogenética da região parcial do gene *pol*.

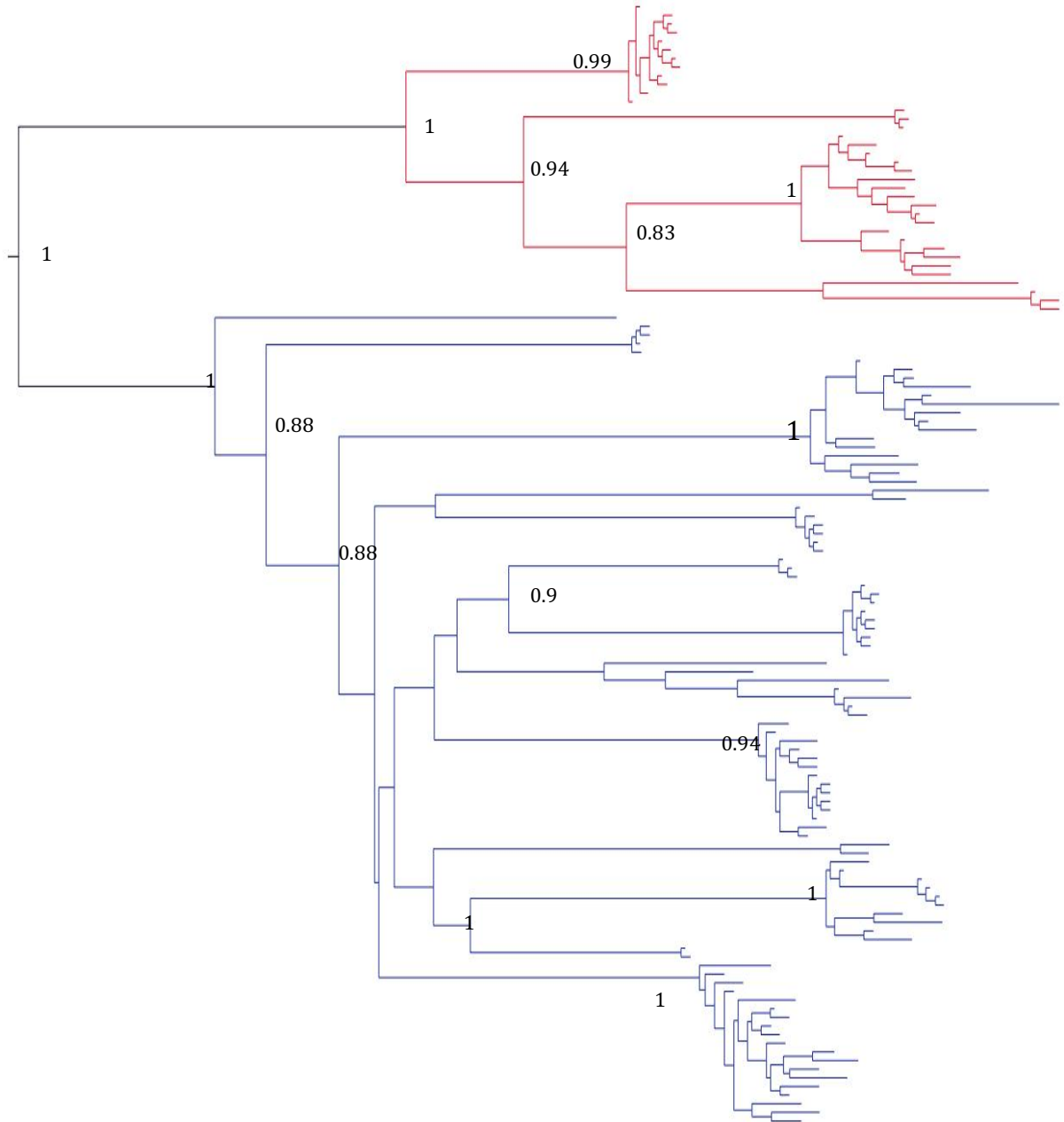
A análise filogenética das 130 variantes envolvidas neste estudo está representada na Figura 30a, uma árvore radial, não enraizada, onde identificamos dois clados referentes aos subtipos B e C com suporte de *bootstrap* de 100%. Esta análise permitiu observar que as variantes dentro de cada hospedeiro se agruparam de maneira consistente com valores de *bootstrap* superior a 95%. O *followup* dos pacientes se agruparam em ramos adjacentes as respectivas amostras mais antigas com valores de *bootstrap* igualmente altos (superior a 90%). As variantes das amostras 599_13 e 600_13 se agruparam entre si formando um clado monofilético com suporte de (100%) (Figura 30b). As variantes 864_13 (sensíveis aos antirretrovirais) e o *followup* FO1_864_13_113_14 (com duas mutações de resistência) se agruparam entre si com suporte de 85% apontando a possível variante que tenha gerado os vírus resistente (Figura 30c).

A fim de estabelecer a extensão da variação genética em pacientes de diagnóstico recente com perfil de soroconversão inferior a dois anos. Nós avaliamos diferentes métricas de diversidade como pode ser visualizada na tabela 7. A mediana da distância genética em nucleotídeos e aminoácido foram 0.005 e 0.006, respectivamente.

De maneira análoga, verificamos que a porcentagem de sítios polimórficos em nucleotídeo foi de 1.10% e de aminoácidos foi de 1.07%. Ao analisar a mediana da razão de dN/dS foi de 0.15, sugerindo uma importante seleção negativa agindo sobre o gene *pol* de HIV-1.

Ao segregar as nossas amostras segundo os critério de incidente (<6 meses de infecção) e prevalente (>6 meses de infecção), verificamos diferença estatística entres os grupos no que diz respeito a distância genética em nucleotídeo (Figura 31a), mas não observamos diferenças estatísticas no que dizem respeito a distancia genética em aminoácidos (Figura 31b) e razão dN/dS.

A



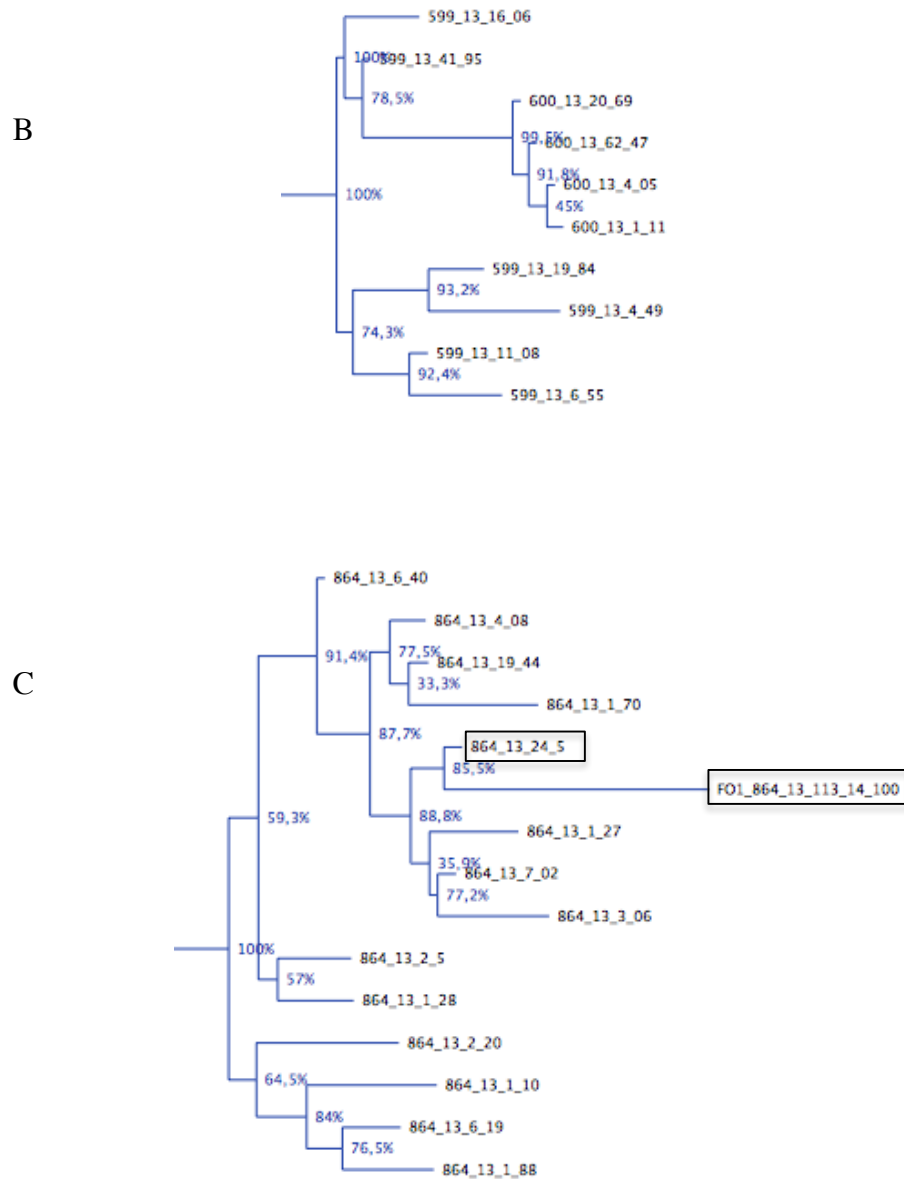


Figura 30 – (A) árvore filogenética por ML com modelo de substituição GTR do gene *pol* envolvendo as 130 haplótipos. Os ramos assinalados em azul reportam-se ao subtipo B e os ramos em vermelho o subtipo C. (B) Zoom do ramo da árvore total contemplando as variantes das amostras 599_13 e 600_13. (C) Zoom do ramos da árvore total contemplado as variantes da amostra 864_13 e seu segmento (*follow-up*). Os retângulos identificam qual variante da amostra 864_13 está mais relacionada ao seguimento (FO1_864_13_113_14).

Tabela 7 – Métrica de diversidade genética intra-hospedeiro.

Sample Id	N de Haplótipos >1	N de nt variáveis	Π nt	N de aa variáveis	Π aa	dN/dS	Infecção
317_11	1	0 (0,00%)	0	0 (0,00%)	0	0	incidente
70_13	8	17 (1,78%)	0.006	7 (2,19%)	0.01	0,30	prevalente
89_13	9	16 (1,6%)	0.008	2 (0,63%)	0.008	0,15	prevalente
104_13	3	9 (0,94%)	0.006	4 (1,25%)	0.008	0,26	incidente
FO1_70_13_219_13	6	5 (0,52%)	0.002	2 (0,63%)	0.002	0,26	prevalente
258_13	5	15 (1,57%)	0.007	3 (0,94%)	0.004	0,06	prevalente
FO1_104_13_272_13	4	20 (2,09%)	0.034	12 (3,76%)	0.028	0,10	prevalente
FO1_89_13_406_13	2	8 (0,84%)	0.008	1 (0,31%)	0.003	0,03	prevalente
451_13	2	15 (1,57%)	0.016	4 (1,25%)	0.013	0,12	incidente
599_13	6	24 (2,51%)	0.011	16 (5,02%)	0.023	0,60	prevalente
600_13	4	2 (0,21%)	0.001	1 (0,31%)	0.002	0,26	incidente
FO1-258_13_639_13	1	0 (0,00%)	0	0 (0,00%)	0	0,00	prevalente
864_13	14	39 (4,08%)	0.012	9 (2,82%)	0.009	0,09	prevalente
FO1_864_13_113_14	1	0 (0,00%)	0	0 (0,00%)	0	0,00	prevalente
137_14	4	5 (0,52%)	0.002	1 (0,31%)	0.003	0,10	incidente
139_14	4	3 (0,31%)	0.002	1 (0,31%)	0.003	0,21	prevalente
169_14	3	3 (0,31%)	0.001	2 (0,63%)	0.006	0,26	incidente
582_14	2	8 (0,84%)	0.008	2 (0,63%)	0.006	0,08	incidente
650_14	9	7 (0,73%)	0.002	2 (0,63%)	0.001	0,22	prevalente
717_14	6	5 (0,52%)	0.002	2 (0,63%)	0.002	0,26	incidente
750_14	19	45 (4,70%)	0.01	8 (2,51%)	0.004	0,04	prevalente
760_14	3	2 (0,21%)	0.001	2 (0,63%)	0.004	0	prevalente
1270_14	2	1 (0,10%)	0.001	1 (0,31%)	0.003	0	incidente
1534_14	12	5 (0,52%)	0.002	2 (0,63%)	0.003	0,22	incidente

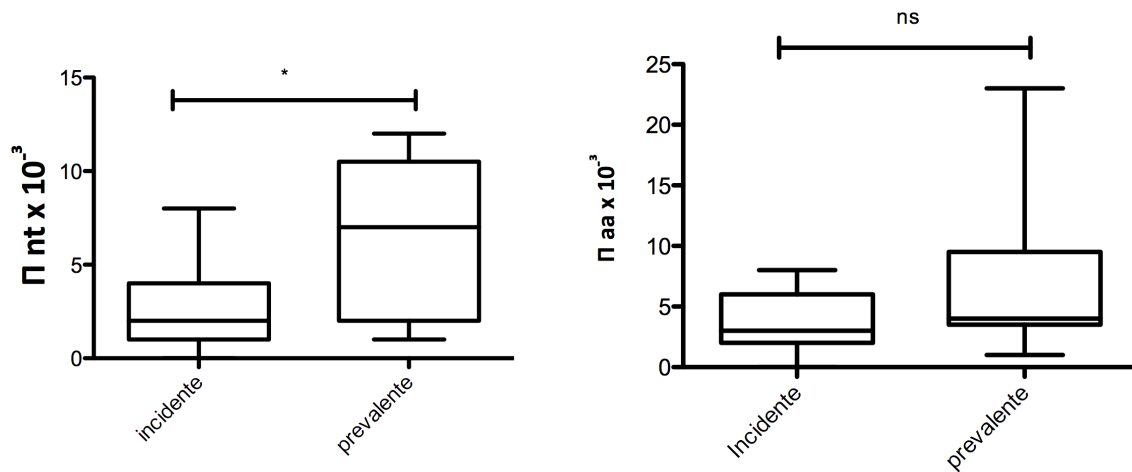


Figura 31 – (A) Análise que representa a mediana da variabilidade genética em nucleotídeos (B) e aminoácidos entre os grupos incidente versus prevalente.

4.8 USO DE CÓDONS DIFERENCIAIS

Para decifrar o grau pelo qual o códon terminando com A/U pode ser preferencialmente usado e para determinar os padrões de uso de códon sinônimo, foi calculado os valores de RSCU de cada códon das 130 variantes virais intra-hospedeiro pertencentes aos subtipos B e C e comparado com a espécie hospedeira (*Homo sapiens*).

Entre os 21 códons mais extensivamente usados no subtipo B de HIV-1, todos (UUU, UUA, AUA, GUA, UCA, CCA, CCU, ACU, ACA, GCA, UAU, CAU, CAA, AAU, AAA, GAU, GAA, UGU, AGA, AGU e GGA), terminaram com A/U (12 terminando com A e 9 terminando com U). Interessantemente, um perfil muito semelhante foi observado no subtipo C, onde 24 códons foram preferencialmente usados, dos quais 21 terminaram com A/U (UUU, UUA, AUA, GUA, UCA, CCU, CCA, ACU, ACA, GCA, UAU, CAU, CAA, AAU, AAA, GAU, GAA, UGU, AGA, , AGU, GGA) (12 terminando com A e 9 terminando com U), e 3 códons terminando com G (UUG, AGG e GGG) e nenhum com C. Não houve diferenças estatísticas entre os subtipos de HIV-1 quanto a preferência de códons terminando com A ou U. Não obstante, fica bem aparente que baseado na análise de RSCU no gene *pol* de HIV-1 há diferenças marcantes quando ao uso diferencial de códons terminando em A/U em detrimento de G/C (Tabela suplementar).

Para uma análise mais detalhada, dividimos os valores de RSCU em 3 categorias: (A) códons com valores de RSCU ≤ 0.6 (sub-representados), (B) códons com RSCU > 1 (preferencialmente usados) e (C) códons com RSCU ≥ 1.6 (super representados). É muito interessante notar que entre os 12 códons super-representados no subtipo B (10 terminam em A ao passo que apenas 2 terminam em U) quanto ao subtipo C dos 10 códons super-representados (8 códons terminam com A e 2 com U). Ao analisar os códons sub-representados 22 a maioria dos códons terminam com G/C 18 terminam com G/C e apenas 4 códons terminam com A/U (Tabela suplementar e)

Nós não detectamos nenhum códon que fosse super representado nos dois subtipos de HIV-1 e no hospedeiro. Como por exemplo no aminoácido valina códon GUA cujos valores de RSCU são 2.43 e 2.29 nos subtipos B e C, respectivamente e no hospedeiro o RSCU é 0.47, portanto sub-representado.

Neste estudo nós comparamos o uso de códon preferenciais nos subtipos B e C com o espécie hospedeira usando o índice de adaptação do códon (CAI). O teste de Kruskal-Wallis revelou que independente do subtipo o uso de códon estimado pelo hospedeiro foi

estatisticamente significativo ($p < 0.05$). O valor de CAI esperado foi de 0.714 e os valores de CAI obtidos entre as nossas variantes variou de 0.75 a 0.78 (tabela suplementar).

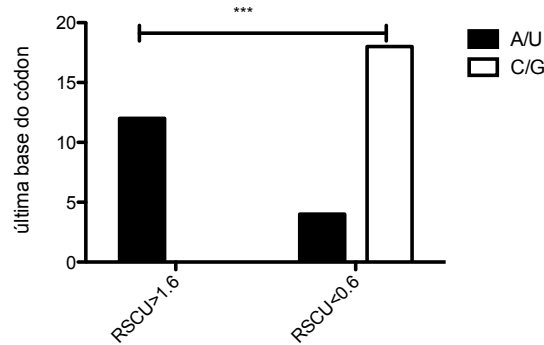


Figura 32 – Análise que representa a frequência dos códons que terminam com A/U e C/G nos grupos de aminoácidos que foram super-representados ($RSCU > 1.6$) e sub-representados ($RSCU < 0.6$).

4.9 SELEÇÃO NATURAL

As variantes virais detectadas a partir da análise da secção 4.7 foram submetidas a análise de seleção natural. As estimativas de pressão de seleção foram determinadas na sequência global codificante que revelou seleção negativa, dados mostrados na tabela 7 ($\omega < 1$). Por outro lado, quando analisamos a adaptação do códon do gene *pol* parcial nas posições do códon 1 da protease ao códon 220 da transcriptase reversa verificamos diferentes forças de seleção agindo sobre os mesmos. A seleção positiva foi observada em dois códons: 229 nos três métodos SLAC, FEL e IFEL e 37 apenas no método IFEL. O regime de seleção negativa atuou de maneira intensa e sob uma gama de códons como podemos observar na tabela 8.

Tabela 8 – Sítios individuais de aminoácidos sobre seleção no gene parcial *pol*.

Método	Haplótipos	
	Sítio selecionado positivamente	N de sítios selecionado negativamente
SLAC	229	64 (20%)
FEL	229	108 (34%)
IFEL	37, 229	99 (31%)

5 DISCUSSÃO

5.1 DADOS CLÍNICOS E EPIDEMIOLÓGICOS E GRUPO DE RISCO

Para melhorar o conhecimento sobre a dinâmica da epidemia de Aids no estado de São Paulo este estudo considerou um grande conjunto de dados, que examina as características demográficas e clínicas dos indivíduos infectados pelos subtipos de HIV-1 mais prevalentes no Brasil. Este trabalho descreveu o grau de relacionamento das sequencias de pacientes crônicos e de infecção recente pelo HIV-1 nos últimos 10 anos da epidemia.

A prevalência dos subtipos foi similar ao observado em estudos prévios conduzidos na região com predomínio dos subtipo B , seguido pelo subtipo F1 e em menor escala o subtipo C Por outro lado, quando avaliamos apenas os pacientes de diagnostico recentes entre Jan 2013 a fev de 2015 verificamos uma frequência do subtipo C em torno de 10%, dado bem contrastante com o cenário do início dos anos 2000, cuja frequência deste subtipo era em torno de 3.8% [118]. Em congruência com estudos mais recentes de diversidade viral do HIV que apontam um aumento deste subtipo no contexto da epidemia de São Paulo [119,120]. Ainda no âmbito do subtipo C, algo que nos chamou a atenção foi a elevada proporção de pacientes *naïves* em relação aos pacientes tratados, como previamente descrito [121]. Estes achados sustentam a hipótese de introdução recente deste subtipo no estado de São Paulo [122].

A comparação dos dados clínicos entre os subtipos B com os outros subtipos mostrou que o subtipo B exibiu menor mediana de contagem de células T CD4⁺ em relação aos subtipos F1 e C, como previamente descrito [123]. Embora, sem nenhuma diferença entre os níveis de carga viral entre os subtipos. Este estudo não realizou nenhuma análise de sobrevida envolvendo os três subtipos, não obstante alguns autores sugerem associação entre subtipos virais e progressão à Aids [124–127]

Este estudo foi o pioneiro na inclusão de amostras oriundas de 53 cidades de pequena, média e grande área a fim de compreender de forma global a prevalência de resistência transmitida no estado. Estudos de vigilância da resistência primaria conduzidos no Brasil estimam uma prevalência entre 5% a 15% dependendo da região geográfica [128–130]. Mais recentemente, Guimarães e cols. [131] em 2015 detectou uma taxa de resistência primária de 9.2% em amostras de pacientes com diagnostico recente da região metropolitana de São Paulo.

O presente estudo demonstrou uma distribuição heterogênea da prevalência de TDR entre os subtipos virais, sendo o subtipo F1 com maior taxa de resistência 18.2%, subtipo B 12.2% e o subtipo C com apenas 3.8%. Comparativamente, um estudo conduzido no Sul do país envolvendo uma coorte de pacientes virgens de tratamento demonstrou uma menor taxa de mutações associadas à resistência no subtipo C em comparação ao subtipo B [132].

Embora, vários trabalhos reportem a taxa de resistência primária conduzidos em diversos estados brasileiros, ainda, há uma lacuna do conhecimento sobre essa taxa em diferentes grupos de risco. É sabido, que no Brasil a população de HSH é 29 vezes mais susceptível de serem infectados que a população em geral [133]. E na América Latina este grupo representa o principal grupo de exposição ao risco à infecção do HIV-1 sobretudo, representando a maior parcela dos novos diagnósticos [134].

O estudo de Bermúdez-Aza EH [135], avaliaram 3515 homens que fazem sexo com homens de 9 cidades brasileiras dos quais 299 foram positivos para o HIV-1 e a taxa de resistência primária nesta população foi elevada (19.3%). Nossos resultados mostraram-se congruentes com os dados de Bermúdez, onde verificamos diferenças estatísticas entre o grupo de HSH e a população em geral nos subtipos B e F1, o qual apresentaram uma elevada taxa de TDR na população de HSH.

5.2 REDES DE TRANSMISSÃO

O conceito de clusters de transmissão é controverso e ainda não há um consenso estabelecido. A definição de clusters adotada neste estudo foi determinada pelo suporte de bootstrap $\geq 90\%$ e distância genética $\leq 6.0\%$. Apesar de termos ciência que muitos estudos adotam a distância genética de até 4.5%, acreditamos que o parâmetro adotado é necessariamente mais flexível, visto que, a nossa estudo considera 10 anos de epidemia, onde os pacientes estudados estavam em diferentes fases da doença, e com diferentes tempos de evolução viral intra-hospedeiro. A fim de validar os clusters detectados, nós separamos todas as sequências que haviam se agrupado entre si e corremos uma análise bayesiana que confirmou o agrupamento das amostras (dado suplementar 2).

O nosso estudo detectou 18.3% putativos *clusters*, embora um baixo percentual de conectividade da nossa população, isto deve-se sobretudo as características da nossa amostragem. Nós queríamos representar mais propriamente o cenário atual do HIV-1

circulante em São Paulo, e portanto, incluímos o maior número de sequências disponíveis (HIVDB-SP).

Nós encontramos diferenças estatísticas entre o percentual de cluster de transmissão de acordo com os subtipos virais. No subtipo C detectamos 40% de clusters de transmissão, um percentual muito superior ao encontrado nos subtipos B e F1. Uma hipótese para esta diferença deve-se ao fato deste subtipo ter sido introduzido recentemente e está se espalhando pelo efeito do fundador.

Estudos mais recentes envolvendo filogenia para descrever redes de transmissão são cada vez mais empregados em países desenvolvidos. O trabalho de Lewis F [136] em 2008 descreve redes de transmissão do HIV-1 em Londres. O estudo destacou que 25% da população de HSH estavam associadas com vários *clusters*. Por outro lado, um estudo conduzido no Reino Unido numa coorte de heterossexuais reporta que apenas 5% da população estuda estava relacionada a *clusters* de transmissão [137]. Embora nós tenhamos usado um parâmetro mais relaxado para definir *cluster*, é importante notar que o espalhamento da infecção entre homens que fazem sexo com homens é muito superior a população heterossexual.

A partir da análise minudente dos *clusters* de transmissão verificamos que independente do subtipo viral os clusters são majoritariamente representados pelo grupo HSH, virgens de tratamento e mais jovens. Este grupo é provavelmente o responsável por dirigir a epidemia no estado de São Paulo. É sabido que fatores de risco individual na população HSH já está bem documentado. Eles incluem alta frequência de diferentes parceiros e sexo desprotegido [138,139]. Em adição aos fatores acima mencionados a biologia do sexo anal, bem como o comportamento aliado a ele, aumentam a eficiência da transmissão. A probabilidade de transmissão do HIV via sexo anal é 18 vezes maior do que a estimado pelo sexo vaginal [140].

Um outro achado interessante foi que os pacientes carregando vírus com resistência primária para os subtipos B e F1 teve significativamente maior associação com redes de transmissão. Em outras palavras, muitos indivíduos carregando vírus com TDR, mesmo removendo os códons associados à resistência aos antirretrovirais estavam relacionados a outros pacientes da nossa coorte indicando a domesticação da transmissão ou eventos de transmissão endógenos. Paradoxalmente, as mutações associadas a resistência que prevalecem nos clusters parecem não reduzir o *fitness* viral. Estes resultados alertam para a importância dos testes de genotipagem pré tratamento entre os indivíduos de diagnóstico

recente a fim de reduzir a transmissão de vírus resistente para os indivíduos não tratados. Nesta linha, o estudo conduzido numa coorte suíça mostrou que a os indivíduos virgens de tratamento são as principais responsáveis pela transmissão de vírus resistente na população de HSH [141].

5.3 DINÂMICA DAS REDES DE TRANSMISSÃO

A fim de melhor compreender a dinâmica da transmissão do HIV-1 no estado de São Paulo, baseado no critério de clusters endógenos, onde pelo menos 66% das sequencias de uma mesma região se agrupam juntas, a percentagem da origem dos mesmos foi estimada. Neste estudo contendo sequencias coletadas nos últimos 10 anos de epidemia de HIV/Aids no estado de São Paulo, verificamos que 81% dos clusters no subtipo B são endógenos, 11% dos cluster tem origem em outros estados brasileiros e 8% dos clusters tenham origem em outro país. Estes achados possam ser explicados pela introdução precoce deste subtipo no inicio dos anos 60 no Brasil [142] e em função disto a epidemia dentro do estado parece auto-manter-se.

Dentre os subtipos estudados, o subtipo C é o que apresentou maior frequência de cluster cuja origem é fora do Brasil, tendo também uma importante influência de clusters cuja origem são dos estados que compõem a região Sul do País. Isso também, talvez possa ser explicado pela sua recente introdução no estado de São Paulo e por este subtipo ser o mais prevalente no mundo [75]. Outras análises também suportam a origem do subtipo C em São Paulo das cidades do Sul do Brasil [122]. O subtipo F1, teve uma parcela importante de clusters cuja origem são de outros estados brasileiros e um baixo percentual de clusters fora do Brasil, talvez por causa da sua baixa prevalência no mundo [75].

Para elucidar a questão da origem dos clusters de transmissão dentro do estado, dividimos o mesmo em 4 áreas geopolíticas. Esta divisão revelou que as regiões Nordeste (NE) e área metropolitana (Metro) têm perfis muito similares. Elas refletem epidemias independentes o qual mais de 50% dos clusters do subtipo B tem origem local. Essas diferenças parecem ser causadas pelo fato de cada uma dessas regiões possuírem seus próprios centros econômicos independentes e uma relativa maturidade, embora epidemia concentrada. O subtipo C apresentou uma parcela importante de cluster endógenos, também encontramos uma associação com a população de HSH. Este resultado é parcial similar ao

encontrado no Sul do país, onde a epidemia relativa a este subtipo é associada a ambos ps grupos de risco: Heterossexual e HSH [143].

Apesar, das regiões 2- Noroeste (NO) e 3- Sul (SU) diferirem das outras regiões descritas acima, apenas no subtipo B encontramos uma importante proporção de *clusters* endógenos. Essas diferenças talvez deva-se a sub-representação das amostras nestas regiões e ou a baixa prevalência dos subtipos não-B nestas regiões geográficas. Embora, nós avaliamos um grande banco de dados da maioria das regiões analisadas. Nós levamos em consideração o número de casos de Aids por localização, de acordo com o último boletim epidemiológico. Então, a percentagem de pacientes infectados pelo localização segue a seguinte distribuição: área 1 (Nordeste, NE) (0.78%), área 2 (Noroeste, NO) (0.72%), área 3 (Sul, SU) (0.4%) e área 4 (Metropolitana, Metro) (1.4%). Destacando que a área com menor representatividade (área30, no qual o intervalo de amostragem varia de 0.4 a 0.7%, caso excluirmos a cidade de Santos, que por questões logísticas não tiveram amostras representadas na nossa coorte.

Tendo em vista que a região metropolitana de São Paulo que compreende uma população de aproximadamente 14 milhões de habitantes e com maior importância econômica dentro do estado, sub-dividimos a mesma em 3 diferentes áreas, a fim de compreender o papel de cada uma delas na epidemia local.

Um achado que nos chama bastante atenção deve-se ao fato de haver apenas clusters endógenos do subtipo C nas áreas 4a (ABCD região) e na região 4c (cidade de São Paulo), o que significa que as transmissões desse subtipo na região 4 deve-se a essas duas sub-áreas. Muito provável que esse subtipo tenha sido introduzido nessas duas regiões de maneira independente e tenha se espalhado. Por outro lado, a epidemia do subtipo F1 é mantida na região metropolitana apenas pelos indivíduos da região periférica (4b) e deve ter se espalhado para as outras áreas dentro da região metropolitana. Nós reconhecemos algumas limitações no nosso estudo e potenciais vies nos resultados. Sobretudo pelo vies da amostragem, método de coleta dos dados e desproporção de amostras entre os municípios do estado de São Paulo. Percebemos, também, que análises de filogeografia/migração, poderia nos dar dicas para melhor compreender a questão da transmissão entre as cidades.

Em conclusão, os nossos dados sugerem que a epidemia do estado de São Paulo é dirigida pelo grupo de HSH com elevado nível de resistência primária, independente do subtipo e a epidemia é sustentada pelos pacientes virgens de tratamento. A epidemia do subtipo C está crescendo dentro do estado em especial na região metropolitana de São Paulo,

onde os indivíduos infectados pelo subtipo C estão 2 vezes mais ligados a redes de transmissão do que os subtipos B e F1.

5.4 DETECÇÃO DE MUTAÇÃO DE RESISTÊNCIA E METRICA DE DIVERSIDADE PELO NGS

Vírus de RNA bem como o HIV são geneticamente diversos, devido a sua alta taxa de replicação, genoma relativamente pequeno e erros da polimerase viral [144,145]. Mesmo dentro do hospedeiro, o HIV existe como uma população complexa, composta de vários sub-grupos relacionados, então chamado de quasi-espécies [146,147]. Além disso, fatores do hospedeiro e resposta imune aplicam uma pressão de seleção que amplifica a diversificação das quasi-espécies. Por exemplo, a transcriptase reversa do HIV-1 é um importante alvo da terapia antirretroviral que pode carregar várias mutações de resistência a drogas, ocorrendo predominantemente entre os 350 primeiros aminoácidos [148].

É sabido, no entanto, que o sequenciamento sanger do gene de interesse a partir do produto de PCR tem sido aplicado por muitos anos como padrão ouro para detecção de variantes de resistência à drogas em amostras clínicas. Não obstante, este tipo de sequenciamento seja hábil apenas em detectar populações majoritárias com frequências entre 20-30%.

Várias revisões destacam a crescente importância do sequenciamento massivamente paralelizado para diferentes aplicações em virologia [149–151]. A principal aplicação do então sequenciamento de nova geração é a detecção de variantes minoritária dentro de uma população viral e o seu impacto clínico na terapia antirretroviral [152–154]. Atualmente, tem-se utilizado esta plataforma para detecção dessas variantes com um frequência de até 1% que parece ser suficiente para gerar um impacto clínico [155,156]. Kim Thys e cols. [157] em 2015 ao comparar diferentes plataformas adota que o Illumina é capaz de detectar populações minoritárias com frequência de até 0.5% e ter uma boa acuraria.

Há trabalhos na literatura que demonstraram claramente que os pacientes com taxa de mutação entre 1% e 20%, particularmente em inibidores não análogos de nucleosídeos da transcriptase reversa, são mais propensos a falharem a terapia [158].

Diferentes abordagens metodológicas podem ser empregadas na obtenção do sequenciamento de nova geração tais como: método amplicon, que permite construir bibliotecas de HIV-1 baseado em produtos específico de PCR e por isso costuma ser bem específico e enriquecer a amostra com uma alta cobertura média dos reads. E o método

sequenciamento a partir de cDNA ou DNA dupla fita está abordagem possibilita uma análise mais acurada, pois envolve menos passos para o processamento do material a ser sequenciado o que reflete em menos viés de erros pré e pós analítico. Além de possibilitar ter uma percepção global do genoma em questão permitindo compreender inclusive associações pleiotrópicas e epistáticas.

Nesta parte do estudo, nós analisamos 23 amostras dos quais 18 amostras reportam-se a pacientes de diagnóstico recente com perfil de soroconvertor recente sem exposição prévia à TARV e 5 amostras sendo o *follow-up* de alguns desses indivíduos. O sequenciamento foi realizado pelo método *amplicon*, visto que o nosso grupo já havia validado o PCR *one step* do gene *pol* e este gene é valioso para o *screening* de variantes com resistência aos antirretrovirais.

A princípio tínhamos como objetivo sequenciar mais de um momento dos 18 pacientes elegíveis para o NGS, porém já na segunda visita a maioria dos pacientes já se encontrava sob regime terapêutico e como se esperava estavam indetectáveis. Desta maneira, a amplificação do gene de interesse a partir da amostra de plasma seria infrutífera. Por outro lado, teríamos maiores chances de amplificar os produtos do gene a partir de DNA total, porém entendemos que este produto não necessariamente reporta-se ao vírus circulante, visto que, o DNA proviral em indivíduo avirêmico simbolize o vírus em estado de latência e portanto com a sua evolução estacionada. Uma outra observação que advogue negativamente ao uso de DNA proviral a estes fins é o efeito da proteína célula APOBEC3G que incorpora G-> A ao longo do genoma durante a etapa de transcrição reversa. Este tipo de viés pode ser crítico nas análises de detecção de população minoritária com perfil de resistência [159].

Baseado nos nossos achados, todas as mutações de resistência observadas pela plataforma sanger foram detectadas pelo sequenciamento de nova geração, no entanto as mutações com frequências inferior a 20% do total de *reads* foram negligenciadas pelo método Sanger. Do conjunto de 12 mutações de resistência observada pelos métodos sanger mais Illumina, 4 delas foram apenas detectadas pelo NGS. Merece destaque a amostra de paciente 70_13 que apresentou 3 mutações de resistência pelo Illumina, onde detectamos a mutação M184V que estava frequente em 18% dos *reads* e que no seu *follow-up* esta mutação estava presente em 50% dos *reads*. Esta observação nos remete a um questionamento cuja resposta parece não ter sido respondida. Quanto tempo uma mutação com frequência inferior a 20% é capaz de se tornar prevalente na presença da droga? O paciente em questão com o laudo de genotipagem sanger pre-tratamento foi indicado a tomar

tenofovir + lamivudina + kaletra e com um intervalo inferior a 3 meses a população viral com a mutação M184V já superava 50%. Cabe ressaltar contudo que esta mutação confere reduzido *fitness* ao vírus. Além disso, mais recentemente Oliveira e cols. [160] em [2016] relataram que esta mutação pode prevenir a mutação de resistência contra dolutegravir a droga mais recentemente licenciada no tratamento de HIV/Aids no Brasil.

Ainda sobre o mesmo vértice do efeito de populações minoritárias, o indivíduo 864_13 realizou o teste de genotipagem sanger pre-tratamento apresentando um laudo de vírus sensível. A mesma análise foi conduzida com o Illumina e tomando o *cut-off* 1% as mutações não foram detectadas, não obstante ao relaxar a frequência das mutações abaixo de 1% detectamos as mutações K65R bem como a mutação G190S estas com frequência de 0.3% e 0.15, respectivamente, ambas com o valor de Phred superior a 30. O que mais nos chamou a atenção é que após 49 dias de uso da terapia antirretroviral as variantes com as respectivas mutações de resistência correspondiam a 100% dos *reads*. Estes achados abrem espaço para repensar nos cálculos estatísticos desenvolvidos pioneiramente por Beerenwinkel e Zagordi [161] em [2011] e posteriormente por Mytsuya e cols. [50] em [2011] que definem valores abaixo de 1% de representatividade são resultados de taxa de erro do sequenciamento.

Não fugindo a regra a amostra 104_13 que é um caso de infecção aguda, de estágio de Fiebig II, mostra claramente que tanto a primeira amostra como o follow-up mostraram-se sensíveis aos TARVs, não obstante quando avaliamos pelo NGS a primeira amostra tinha uma mutação K103N em 0.3% dos *reads*, porém passado um intervalo inferior a 2 meses a amostra de *follow-up* detectou essa mutação na ordem de 1.6% do total de *reads* Isto representa um crescimento na frequência desta mutação na ordem de 5 vezes. Vale salientar que o parceiro desta paciente, também HIV-1 positivo está sob regime de efavirenz, responsável pela seleção desta mutação.

Estas questões trazidas à baila enobrecem o conceito de quasi-espécies, onde populações muito minoritárias tem um efeito *sine qua non* na adaptabilidade da população viral. Os nossos dados corroboram com estudos prévios que reforçam a ideia que populações minoritárias carregam mutações que sobre o efeito de pressão de seleção se tornam hábeis em escapar do sistema imune ou dos agentes antivirais [99,146].

A análise da detecção das mutações associadas a resistência antirretroviral, deu-se a partir da análise dos *reads* que compõem a sequência consenso alinhada e mapeadas frente ao protótipo HXB2. Porém toda a métrica de diversidade, bem como as análises sucessória

de uso de códons diferenciais e seleção descritas neste trabalho foram realizadas após a reconstrução haplotípica das variantes virais. Embora, fosse idealmente aplicada a genomas completos, neste estudo focamos na diversidade gene específico (*pol*). Este gene é responsável pela codificação de proteínas vitais à biologia viral tais como: protéase, transcriptase reversa e integrase. E por isso as principais classes de drogas agem especificamente sobre esta região.

Uma questão que nos alertou após a reconstrução haplotípica seria a possibilidade de contaminação cruzada durante a exequibilidade experimental. Esta hipótese foi totalmente refutada a medida que a análise filogenética mostrou *clusters* monofiléticos apenas entre haplótipos de um mesmo paciente. Não obstante as amostras 599_13 e 600_13 formaram um grupo único, onde alguns haplótipos do paciente 599_13 se agruparam de maneira consistente com os haplótipos do paciente 600_13 e isto deve-se ao fato desses indivíduos fazerem parte da mesma cadeia epidemiológica por serem parceiros. Uma outra explicação pode ser razoavelmente plausível é de que o indivíduo 600_13 era incidente, ou seja, com infecção inferior a 6 meses e os vírus que ele albergava não obteve tempo suficiente para se diferenciarem dos vírus do paciente índice.

Os haplótipos da amostras *follow-up* dos pacientes estudadas se agruparam de maneira consistente com os haplótipos da primeira visita descartando a possibilidade de super-infecção.

A evolução do HIV-1 bem como a dinâmica populacional tem sido extensivamente investigada especialmente pelo seu potencial papel na patogenese (e porque eles constituem o principal obstáculo para o desenvolvimento de uma vacina [162].

É sabido que a variabilidade genética do HIV-1 pode alcançar 5% dentro de um indivíduo infectado e isto leva a população viral rapidamente se adaptar as mudanças ambientais [163]. As infecções são tipicamente caracterizadas em três fases: fase aguda, latência clínica e progressão para a Aids, durante o qual a divergência genética viral e diversidade jogam com um consistente padrão associado a progressão da doença [164].

Nós estratificamos os pacientes segundo o critério de incidentes (menos de 6 meses de infecção) e prevalente (superior a 6 meses de infecção). Em linhas gerais, dois métodos nos possibilitaram a esta divisão: o intervalo entre o último teste negativo para o HIV e o primeiro teste positivo inferior a 6 meses e/ou teste de avidéz desenvolvido por [165] que apesar de viés de uma sorologia há vários trabalhos na literatura que usam como um parâmetro importante de segregação entre infecção de curto e longo prazo.

Nossos dados mostram que os indivíduos incidentes apresentavam um grau de diversidade genética nucleotídica muito inferior ao observado entre os indivíduos prevalentes. E estes achados podem ser explicados pelo fato de um indivíduo ao se tornar infectado pelo HIV, a população viral é inicialmente homogênea, visto que a infecção é usualmente associada a um significativo *bottleneck* populacional [166–168]. A infecção pode ser estabelecida por um único virion [169,170]. Existem estudos que suportam a hipótese que o número de variantes transmitida parece ser dependente da rota de transmissão e que o *bottleneck* pode ser mais severo na transmissão sexual. Estudos apontam que a transmissão do HIV entre a população HSH ocorre por múltiplos haplotipos ao passo que entre os heterossexuais esse número de haplotipos é bem menor [162,171,172].

Quando direcionamos o nosso olhar aos pacientes incidentes a grande maioria era composta por homens que fazem sexo com homens, onde a maioria carregava múltiplas variantes. Porém, dois indivíduos apresentaram apenas uma única variante viral que possivelmente representa a variante transmitida. Ao passo no grupo de incidentes detectamos uma única mulher heterossexual que apresentou 4 variantes virais, que foram as prováveis variantes transmitida, visto que esta paciente encontrava-se em estágio de Fiebig II, ou seja, tinha sorologia negativa para o HIV-1, porém carga viral positiva e neste caso superior a 1 milhão de cópias.

Este tipo de análise comparativa entre HSH e Het quanto ao número de possíveis variantes transmitida aqui neste estudo passa a ser um limitante, haja vista, que o n de amostras de Het é bem reduzido. Porém, ressaltamos que embora a maioria da população de HSH apresentasse múltiplas variantes (mediana de 5 variantes) na fase inicial da doença, não podemos excluir a possibilidade de transmissão de variantes únicas nesta população.

Ao analisar a variabilidade genética em aminoácidos não verificamos diferenças entre as populações de incidentes e prevalentes o que reforça a ideia que a variabilidade é preferencialmente observada na última base do códon, refletindo em mutações sinônimas, sendo a seleção e deriva genética a principal força agindo sobre a evolução do vírus intra-hospedeiro [173].

5.5 USO DE CÓDONS DIFERENCIAL

O código genético é degenerado com múltiplos códons codificando um mesmo aminoácido, muitos organismos exibem diferentes composições de bases e significativo uso de códon diferencial, que outras palavras significa uso desigual de códons sinônimos.

Geralmente mutações que levam mudanças de aminoácidos são estudados como medidas de seleção. Mutações sinônimas podem mudar a composição do gene sem alterar a proteína correspondente. Assim, mutações sinônimas usualmente são neutras ou quase neutras e seus efeitos estão para melhor serem compreendidos. Estudos indicam que uso de códons preferenciais estão sob seleção fraca e desta forma estão dirigindo a evolução em inúmeros organismos. Genes que são enriquecidos por um códon preferido são conhecidos por aumentar a eficiência da tradução [174].

No sistema bacteriófago parece que muitos dos seus genes são enriquecidos por códons que, também, são preferenciais pelo seu hospedeiro. Não obstante, já foi relatado que o HIV-1 difere do seu hospedeiro na composição de base bem como no uso de códons diferenciais [175–177].

O genoma do HIV-1 é rico no nucleotídeo A, que conta com o apoio de uma proteína celular da família APOBEC que hipermuta o genoma viral de G → A durante a etapa de retrotranscrição reversa [178]. Ao passo que o genoma do hospedeiro é rico em GC em regiões codificantes.

Nossos resultados corroboram com os achados da literatura onde detectamos inúmeros códons sendo preferencialmente usados e a maioria terminando com A ou U ao passo que não detectamos nenhum códons que fosse, também preferencialmente usado entre o seu hospedeiro [175]. Tal observação, talvez possa ser explicada pelo fato de nossos pacientes de diagnóstico recente foram caracterizados como sorovonvertedores recentes com infecção documentada inferior a 2 anos. Segundo Peter e cols. [175] hipotetizaram a respeito do perfil de RSCU variando segundo o estadiamento da doença, isto quer dizer que os indivíduos recém infectados apresentaram um perfil de RSCU bem diferentes do observado no hospedeiro, porém a media que evoluem com a doença o RSCU tende a se aproximar do RSCU do hospedeiro. Segundo Peter, esta mudança pode ser atribuída a seleção dirigida que reflete no aumento da eficiência da tradução. Uma segunda explicação seria o mero resultado da pressão de mutação na sequencia viral.

Um quadro que ilustra muito bem as diferenças entre o RSCU do HIV-1 e do hospedeiro é o aminoácido arginina, cujo códons AGA teve o maior RSCU de 4.7 sendo este mesmo códon em humano com valor de 1.29. No entanto, não fomos hábeis em detectar o padrão de RSCU em momentos diferentes da evolução clínica o que nos possibilita apenas reiterar as diferenças observadas no uso códons entre patógeno e hospedeiro.

Por outro lado, Aridaman e cols. [174] fizeram uma análise em larga escala sobre o padrão do uso de códons nos 9 genes dos HIV-1 em 1.357 genomas completos contendo sequências ao longo de 23 anos. O grupo mostrou que apenas 3 genes (*env*, *rev* e *tat*) sofreram uma maior mudança na composição do uso de códons ao longo do tempo, tornando-se mais próximos ao padrão observado pelos hospedeiros.

5.6 PRESSÃO DE SELEÇÃO

A variação genética dentro do hospedeiro varia ao longo do tempo, como resultado de pelo menos 3 processos. O primeiro deles, envolve a introdução de uma nova mutação que entra e interage com o ciclo de vida do vírus e uma vez, integrado ao genoma da célula pode acarretar em uma nuvem de genomas semelhantes que compõem um grande espaço de sequências com competência replicativa [101]. O segundo grande modulador da variabilidade intra-hospedeiro são os efeitos de pressão seletiva resultantes da resposta imune, intervenções terapêuticas, a adesão aos medicamentos e outros fatores ambientais que afetam todas as frequências das variantes virais pré-existentes [179,180]. E o terceiro processo seria baseado no escape viral ou liberação do HIV-1 a partir dos reservatórios celulares ou anatômicos que continuam a replicar o vírus em baixa concentração [181,182].

Em acordo com o estudo prévio [183], nossos resultados demonstraram que a seleção natural é predominantemente negativa no gene *pol* do HIV-1, isto quer dizer que a seleção natural elimina a boa parte dos vírus que albergam mutantes que implicam em troca de aminoácido. Em outras palavras, isto significa que existem ao longo do gene *pol* um grande número de mutações que são deletérias ao vírus. E portanto, o gene está sobre o efeito preferencial de seleção purificadora. A explicação mais plausível é que este gene sintetiza as principais enzimas do HIV-1 e isto mudanças de aminoácidos poderiam implicar em mudanças conformacionais afetando o fitness viral ou ainda mesmo a sua capacidade de infectar novas células. Apesar do gene está sob o regime de seleção purificadora, nós detectamos dois sites sob regime de seleção positiva o códon 37 da protease viral e o códon 130 da transcriptase reversa. Quanto ao códon 37, ele não tem qualquer relação de associação direta com resistência à TARV.

O códon 130 que corresponde a posição 229 do nosso alinhamento de aminoácidos é um domínio da transcriptase reversa que também não tem associação com mutação de resistência principal ou acessória.

Apesar dos códons que estão sob efeito de seleção positiva/diversificadora não estarem associados a alguma resistência documentada aos antirretrovirais, não quer dizer que eles não aumentem a capacidade adaptativa do vírus. Em resumo a vantagem de manter diversidade deste códon, leva ao aumento da capacidade da população viral de conferir um fenótipo que possa adquirir vantagens em condições específicas [183,184].

6 CONCLUSÕES

- A epidemia do estado de São Paulo é dirigida pelo grupo de homens que fazem sexo com homens com elevada taxa de resistência primária.
- Independente do subtipo a epidemia é auto-sustentada pelos pacientes virgens de tratamento.
- A epidemia do subtipo C está crescendo dentro do estado de São Paulo sobretudo na região metropolitana.
- Os indivíduos infectados pelo subtipo C estão duas vezes mais ligados a redes de transmissão que os indivíduos infectados pelos subtipos B e F1
- O método do sequenciamento de nova geração foi capaz de detectar 4/12 (30%) mutações de baixa frequência entre 1-20% que foram negligenciadas pelo método de sanger, chegando a 50% quando relaxamos o parâmetro para abaixo de 1%.
- Verificamos diferenças significativas de diversidade genética em nucleotídeo dentro do grupo de soroconvertores recentes (incidentes versus prevalentes).
- O gene *pol* do HIV-1 exibe uma grande gama de códons preferencialmente expressos e sobretudo terminando em A/T, no grupo de soroconvertores recentes, diferentemente do observado no hospedeiro.
- Os códons 36 da protease e o códon 130 do domínio da transcriptase reversa estão em regime de seleção positiva, muito embora a maioria dos códons estão sob efeito de seleção purificadora.

REFERENCIAS*

1. Coffin J, Haase A, Levy J, Montagnier L, Oroszlan S, Teich N, et al. Human immunodeficiency viruses. *Science*. 1986;232(4751):697–697.
2. AIDS EPIDEMIC UPDATE. UNAIDS. Published jointly by UNAIDS and the World health Organization. [internet]. 2008. [capturado em 31 de maio e 2015]. disponível em: <http://www.unaids.org>.
3. AIDS EPIDEMIC UPDATE. UNAIDS. Published jointly by UNAIDS and the World health Organization. [internet]. 2016. [capturado em 15 de agosto de 2015]. disponível em: <http://www.unaids.org>.
4. Boletim Epidemiológico. Programa Nacional de DST e AIDS. Ministério da saúde. [internet]. Brasil; 2010. [capturado em 27 de março de 2010]. Disponível em <http://www.aids.gov.br>.
5. Boletim Epidemiológico. Programa Nacional de DST e AIDS. Ministério da saúde. [internet]. Brasil; 2015. [capturado em 21 de agosto de 2015]. Disponível em <http://www.aids.gov.br>.
6. Boletim Epidemiológico. Programa Nacional de DST e AIDS. Ministério da saúde do estado de São Paulo [internet]. Brasil; 2014. [capturado em 06 de dezembro de 2014]. Disponível em <http://www.saude.sp.gov.br>.
7. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, et al. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodyte*. *Nature*. 1999;397(6718):436–41.
8. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med*. 2011;1(1):a006841.
9. Goto T, Nakai M, Ikuta K. The life-cycle of human immunodeficiency virus type 1. *Micron*. 1998; 29(2–3):123–38.
10. Marx PA, Munn RJ, Joy KI. Computer emulation of thin section electron microscopy predicts an envelope-associated icosadeltahedral capsid for human immunodeficiency virus. *Lab Invest*. 1988;58(1):112–8.
11. Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem*. 1998;67:1–25.
12. Freed EO. HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology*. 1998;251(1):1–15.
13. Fang X, Wang J, O’Carroll IP, Mitchell M, Zuo X, Wang Y, et al. An unusual topological structure of the HIV-1 Rev response element. *Cell*. 2013;155(3):594–605.

14. Freed EO. HIV-1 replication. *Somat Cell Mol Genet*. 2001;26(1–6):13–33.
15. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*. 1988;331(6153):280–3.
16. Turner BG, Summers MF. Structural biology of HIV. *J Mol Biol*. 1999;285(1):1–32.
17. Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, et al. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science*. 2011;334(6053):242–5.
18. Polzer S, Dittmar MT, Schmitz H, Meyer B, Müller H, Kräusslich HG, et al. Loss of N-linked glycans in the V3-loop region of gp120 is correlated to an enhanced infectivity of HIV-1. *Glycobiology*. 2001;11(1):11–9.
19. Donahoe SM, Moretto WJ, Samuel R V., Metzner KJ, Marx PA, Hanke T, et al. Direct Measurement of CD8+ T Cell Responses in Macaques Infected with Simian Immunodeficiency Virus. *Virology*. 2000;272(2):347–56.
20. Zhang G, Liu R, Zhong Y, Plotnikov AN, Zhang W, Zeng L, et al. Down-regulation of NF- κ B transcriptional activity in HIV-associated kidney disease by BRD4 inhibition. *J Biol Chem*. 2012;287(46):38956–38956.
21. Kumar B, Arora S, Ahmed S, Banerjea AC. Hyperactivation of mammalian target of rapamycin complex 1 by HIV-1 is necessary for virion production and latent viral reactivation. *FASEB J*. 2016; pii: fj.201600813R.
22. Stettner MR, Nance JA, Wright CA, Kinoshita Y, Kim W-K, Morgello S, et al. SMAD proteins of oligodendroglial cells regulate transcription of JC virus early and late genes coordinately with the Tat protein of human immunodeficiency virus type 1. *J Gen Virol*. 2009;90(8):2005–14.
23. Matsui M, Warburton RJ, Cogswell PC, Baldwin AS, Frelinger JA. Effects of HIV-1 Tat on expression of HLA class I molecules. *J Acquir Immune Defic Syndr Hum Retrovirol*. 1996;11(3):233–40.
24. Malim MH, Tiley LS, McCarn DF, Rusche JR, Hauber J, Cullen BR. HIV-1 structural gene expression requires binding of the rev trans-activator to its RNA target sequence. *Cell*. 1990;60(4):675–83.
25. Rose KM, Marin M, Kozak SL, Kabat D. The viral infectivity factor (Vif) of HIV-1 unveiled. *Trends Mol Med*. 2004;10(6):291–7.
26. Navarro F, Landau NR. Recent insights into HIV-1 Vif. *Curr Opin Immunol*. 2004;16(4):477–82.
27. Argyris EG, Pomerantz RJ. HIV-1 Vif versus APOBEC3G: newly appreciated warriors in the ancient battle between virus and host. *Trends Microbiol*.

- 2004;12(4):145–8.
28. Heinzinger NK, Bukrinsky MI, Haggerty SA, Ragland AM, Kewalramani V, Lee MA, et al. The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proc Natl Acad Sci U S A*. 1994;91(15):7311–5.
 29. He J, Choe S, Walker R, Di Marzio P, Morgan DO, Landau NR. Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. *J Virol*. 1995;69(11):6705–11.
 30. Bour S, Strebel K. The HIV-1 Vpu protein: a multifunctional enhancer of viral particle release. *Microbes Infect*. 2003;5(11):1029–39.
 31. Mailler E, Bernacchi S, Marquet R, Paillart J-C, Vivet-Boudou V, Smyth R. The Life-Cycle of the HIV-1 Gag–RNA Complex. *Viruses*. 2016;8(9):248.
 32. Levitt JT, Brown TA, Orsillo SM, Barlow DH. The effects of acceptance versus suppression of emotion on subjective and psychophysiological response to carbon dioxide challenge in patients with panic disorder. *Behav Ther*. 2004;35(4):747–66.
 33. Bowerman B, Brown PO, Bishop JM, Varmus HE. A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev*. 1989;3(4):469–78.
 34. BARBER AM, HIZI A, MAIZEL JV, HUGHES SH. HIV-1 Reverse Transcriptase: Structure Predictions for the Polymerase Domain. *AIDS Res Hum Retroviruses*. 1990;6(9):1061–72.
 35. Hottiger M, Hübscher U. Human Immunodeficiency Virus type 1 reverse transcriptase. *Biol Chem Hoppe Seyler*. 1996;377(2):97–120.
 36. Saayman S, Ali SA, Morris K V, Weinberg MS. The therapeutic application of CRISPR/Cas9 technologies for HIV. *Expert Opin Biol Ther*. 2015;15(6):819–30.
 37. Shehu-Xhilaga M, Tachedjian G, Crowe S, Kedzierska K. Antiretroviral Compounds: Mechanisms Underlying Failure of HAART to Eradicate HIV-1. *Curr Med Chem*. 2005;12(15):1705–19.
 38. Jülg B, Barouch DH. Novel immunological strategies for HIV-1 eradication. *J virus Erad*. 2015;1(4):232–6.
 39. Stein J, Storcksdieck Genannt Bonsmann M, Streeck H. Barriers to HIV Cure. *HLA*. 2016;88(4):155–63.
 40. Bangsberg DR, Charlebois ED, Grant RM, Holodniy M, Deeks SG, Perry S, et al. High levels of adherence do not prevent accumulation of HIV drug resistance mutations. *AIDS*. 2003;17(13):1925–32.
 41. Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. *Cold Spring Harb Perspect Med*. 2012;2(4):a007161.

42. Richman DD. HIV chemotherapy. *Nature*. 2001 Apr 19;410(6831):995–1001.
43. Esnouf R, Ren J, Ross C, Jones Y, Stammers D, Stuart D. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nat Struct Biol*. 1995;2(4):303–8.
44. Xia Q, Radzio J, Anderson KS, Sluis-Cremer N. Probing nonnucleoside inhibitor-induced active-site distortion in HIV-1 reverse transcriptase by transient kinetic analyses. *Protein Sci*. 2007;16(8):1728–37.
45. Hazuda DJ, Young SD, Guare JP, Anthony NJ, Gomez RP, Wai JS, et al. Integrase Inhibitors and Cellular Immunity Suppress Retroviral Replication in Rhesus Macaques. *Science*. 2004; 305(5683):528-32.
46. Hazuda DJ, Anthony NJ, Gomez RP, Jolly SM, Wai JS, Zhuang L, et al. A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proc Natl Acad Sci U S A*. 2004;101(31):11233–8.
47. Shimura K, Kodama E, Sakagami Y, Matsuzaki Y, Watanabe W, Yamataka K, et al. Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137). *J Virol*. 2008;82(2):764–74.
48. Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature*. 2010 11;464(7286):232–6.
49. Pomerantz RJ, Horn DL. Twenty years of therapy for HIV-1 infection. *Nat Med*. 2003;9(7):867–73.
50. White KL, Chen JM, Feng JY, Margot NA, Ly JK, Ray AS, et al. The K65R reverse transcriptase mutation in HIV-1 reverses the excision phenotype of zidovudine resistance mutations. *Antivir Ther*. 2006;11(2):155-63.
51. Demeter LM, Shafer RW, Meehan PM, Holden-Wiltse J, Fischl MA, Freimuth WW, et al. Delavirdine susceptibilities and associated reverse transcriptase mutations in human immunodeficiency virus type 1 isolates from patients in a phase I/II trial of delavirdine monotherapy (ACTG 260). *Antimicrob Agents Chemother*. 2000;44(3):794–7.
52. Azijn H, Tirry I, Vingerhoets J, de Béthune M-P, Kraus G, Boven K, et al. TMC278, a next-generation nonnucleoside reverse transcriptase inhibitor (NNRTI), active against wild-type and NNRTI-resistant HIV-1. *Antimicrob Agents Chemother*. 2010 Feb;54(2):718–27.
53. Métifiot M, Vandegraaff N, Maddali K, Naumova A, Zhang X, Rhodes D, et al. Elvitegravir overcomes resistance to raltegravir induced by integrase mutation Y143. *AIDS*. 2011;25(9):1175–8.

54. Hatano H, Lampiris H, Fransen S, Gupta S, Huang W, Hoh R, et al. Evolution of integrase resistance during failure of integrase inhibitor-based antiretroviral therapy. *J Acquir Immune Defic Syndr*. 2010;54(4):389–93.
55. Quashie PK, Mesplède T, Wainberg MA. Evolution of HIV integrase resistance mutations. *Curr Opin Infect Dis*. 2013;26(1):43-9.
56. Clavel F, Hance AJ. HIV Drug Resistance. *N Engl J Med*. 2004;350(10):1023–35.
57. Liu F, Kovalevsky AY, Louis JM, Boross PI, Wang Y-F, Harrison RW, et al. Mechanism of Drug Resistance Revealed by the Crystal Structure of the Unliganded HIV-1 Protease with F53L Mutation. *Journal of Molecular Biology*. 2006; 358(5):1191-9.
58. Jurriaans S, Van Gemen B, Weverling GJ, Van Strup D, Nara P, Coutinho R, et al. The Natural History of HIV-1 Infection: Virus Load and Virus Phenotype Independent Determinants of Clinical Course? *Virology*. 1994;204(1):223–33.
59. Jaffar S, Grant AD, Whitworth J, Smith PG, Whittle H. The natural history of HIV-1 and HIV-2 infections in adults in Africa: a literature review. *Bull World Health Organ*. 2004;82(6):462–9.
60. Cohen MS, Shaw GM, McMichael AJ, Haynes BF. Acute HIV-1 Infection. *N Engl J Med*. 2011;364(20):1943–54.
61. Garber DA, Silvestri G, Barry AP, Fedanov A, Kozyr N, McClure H, et al. Blockade of T cell costimulation reveals interrelated actions of CD4+ and CD8+ T cells in control of SIV replication. *J Clin Invest*. 2004;113(6):836–45.
62. Staprans S, Feinberg MB. The roles of nonhuman primates in the preclinical evaluation of candidate AIDS vaccines. *Expert Rev Vaccines*. 2004;3(sup1):S5–32.
63. Karon JM, Buehler JW, Byers RH, Farizo KM, Green TA, Hanson DL, et al. Projections of the number of persons diagnosed with AIDS and the number of immunosuppressed HIV-infected persons--United States, 1992-1994. *MMWR Recomm Rep*. 1992;41(RR-18):1–29.
64. Mindel A, Tenant-Flowers M. ABC of AIDS: Natural history and management of early HIV infection. *BMJ*. 2001;322(7297):1290–3.
65. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, et al. HIV-1 nomenclature proposal. *Science*. 2000;288(5463):55–6.
66. Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, et al. Timing the Ancestor of the HIV-1 Pandemic Strains. *Science*. 2000;288(5472).
67. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature*.

- 2008;455(7213):661–4.
68. Triques K, Bourgeois A, Vidal N, Mpoudi-Ngole E, Mulanga-Kabeya C, Nzilambi N, et al. Near-Full-Length Genome Sequencing of Divergent African HIV Type 1 Subtype F Viruses Leads to the Identification of a New HIV Type 1 Subtype Designated K. *AIDS Res Hum Retroviruses*. 2000;16(2):139–51.
 69. Niama FR, Toure-Kane C, Vidal N, Obengui P, Bikandou B, Ndoundou Nkodia MY, et al. HIV-1 subtypes and recombinants in the Republic of Congo. *Infect Genet Evol*. 2006;6(5):337–43.
 70. Foley B, Leitner T, Apetrei C, Hahn B, Mizrachi I, Mullins J, Rambaut A, Wolinsky S and KB. HIV Sequence Compendium 2013 [Internet]. 2013. Available from: <https://www.hiv.lanl.gov>.
 71. Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. The Challenge of HIV-1 Subtype Diversity. *N Engl J Med*. 2008 Apr 10;358(15):1590–602.
 72. Hemelaar J. Implications of HIV diversity for the HIV-1 pandemic. *J Infect*. 2013;66(5):391–400.
 73. Gilbert MTP, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A*. 2007;104(47):18566–70.
 74. Haaland RE, Hawkins PA, Salazar-Gonzalez J, Johnson A, Tichacek A, Karita E, et al. Inflammatory Genital Infections Mitigate a Severe Genetic Bottleneck in Heterosexual Transmission of Subtype A and C HIV-1. Trkola A, editor. *PLoS Pathog*. 2009;5(1):e1000274.
 75. Hemelaar J, Gouws E, Ghys PD, Osmanov S, WHO-UNAIDS Network for HIV Isolation and Characterisation. Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS*. 2011;25(5):679–89.
 76. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS*. 2006 Oct;20(16):W13–23.
 77. Morgado MG, Sabino EC, Shpaer EG, Bongertz V, Brigido L, Guimaraes MDC, et al. V3 Region Polymorphisms in HIV-1 from Brazil: Prevalence of Subtype B Strains Divergent from North American/European Prototype and Detection of Subtype F. *AIDS Res Hum Retroviruses*. 1994;10(5):569–76.
 78. Morgado M, Guimarães M, Neves Jr I, Veloso dos Santos V, Linhares-de-Carvalho M, Castello-Branco L, et al. Molecular Epidemiology of HIV in Brazil: Polymorphism of the Antigenically Distinct HIV-1 B Subtype Strains. *Mem Inst Oswaldo Cruz*. 1998;93(3):383–6.
 79. Thomson MM, Villahermosa ML, Vázquez-de-Parga E, Cuevas MT, Delgado E, Manjón N, et al. Widespread circulation of a B/F intersubtype recombinant form

- among HIV-1-infected individuals in Buenos Aires, Argentina. *AIDS*. 2000;14(7):897–9.
80. Velarde-Dunois KG, Guimarães ML, La Fuente C, Andrade R, Arévalo R, Pantoja S, et al. Molecular Characterization of Human Immunodeficiency Virus Type 1-Infected Individuals from Bolivia Reveals the Presence of Two Distinct Genetic Subtypes B and F. *AIDS Res Hum Retroviruses*. 2000;16(17):1921–6.
 81. Hierholzer J, Montano S, Hoelscher M, Negrete M, Hierholzer M, Avila MM, et al. Molecular Epidemiology of HIV Type 1 in Ecuador, Peru, Bolivia, Uruguay, and Argentina. *AIDS Res Hum Retroviruses*. 2002;18(18):1339–50.
 82. Aulicino PC, Gómez-Carrillo M, Bello G, Rocco C, Mangano A, Carr J, et al. Characterization of full-length HIV-1 CRF17_BF genomes and comparison to the prototype CRF12_BF strains. *Infect Genet Evol*. 2012;12(2):443–7.
 83. Ruchansky D, Casado C, Russi JC, Arbiza JR, Lopez-Galindez C. Identification of a new HIV Type 1 circulating recombinant form (CRF38_BF1) in Uruguay. *AIDS Res Hum Retroviruses*. 2009;25(3):351–6.
 84. Ssemwanga D, Nsubuga RN, Mayanja BN, Lyagoba F, Magambo B, Yirrell D, et al. Effect of HIV-1 subtypes on disease progression in rural Uganda: a prospective clinical cohort study. *PLoS One*. 2013;8(8):e71768.
 85. Galvin SR, Cohen MS. The role of sexually transmitted diseases in HIV transmission. *Nat Rev Microbiol*. 2004 Jan;2(1):33–42.
 86. Guimarães ML, Marques BCL, Bertoni N, Teixeira SLM, Morgado MG, Bastos FI, et al. Assessing the HIV-1 Epidemic in Brazilian Drug Users: A Molecular Epidemiology Approach. *PLoS One*. 2015;10(11):e0141372.
 87. Machado LFA, Ishak MOG, Vallinoto ACR, Lemos JAR, Azevedo VN, Moreira MRC, et al. Molecular Epidemiology of HIV Type 1 in Northern Brazil: Identification of Subtypes C and D and the Introduction of CRF02_AG in the Amazon Region of Brazil. *AIDS Res Hum Retroviruses*. 2009;25(10):961–6.
 88. Brindeiro RM, Diaz RS, Sabino EC, Morgado MG, Pires IL, Brigido L, et al. Brazilian Network for HIV Drug Resistance Surveillance (HIV-BResNet): a survey of chronically infected individuals. *AIDS*. 2003;17(7):1063–9.
 89. Soares EAJM, Martínez AMB, Souza TM, Santos AFA, Da Hora V, Silveira J, et al. HIV-1 subtype C dissemination in southern Brazil. *AIDS*. 2005;19 Suppl 4:S81–6.
 90. Gräf T, Pinto AR. The increasing prevalence of HIV-1 subtype C in Southern Brazil and its dispersion through the continent. *Virology*. 2013;435(1):170–8.
 91. De Sá Filho DJ, Sucupira MCA, Caseiro MM, Casiero MM, Sabino EC, Diaz RS, et al. Identification of two HIV type 1 circulating recombinant forms in Brazil. *AIDS Res Hum Retroviruses*. 2006;22(1):1–13.

92. Guimarães ML, Couto-Fernandez JC, Eyer-Silva W de A, Teixeira SLM, Chequer-Fernandez SL, Morgado MG. Analysis of HIV-1 BF pr/rt recombinant strains from Rio de Janeiro/Brazil reveals multiple unrelated mosaic structures. *Infect Genet Evol.* 2010;10(7):1094–100.
93. Sanabani SS, Pastena ER de S, Neto WK, Martinez VP, Sabino EC. Characterization and frequency of a newly identified HIV-1 BF1 intersubtype circulating recombinant form in São Paulo, Brazil. *Virology*. 2010;7:74.
94. Pessôa R, Carneiro Proietti AB de F, Busch MP, Sanabani SS. Identification of a Novel HIV-1 Circulating Recombinant Form (CRF72_BF1) in Deep Sequencing Data from Blood Donors in Southeastern Brazil. *Genome Announc.* 2014;2(3).
95. Bello G, Guimarães ML, Morgado MG. Evolutionary history of HIV-1 subtype B and F infections in Brazil. *AIDS.* 2006 Mar 21;20(5):763–8.
96. Louwagie J, Delwart EL, Mullins JI, Mccutchan FE, Eddy G, Burke DS. Genetic Analysis of HIV-1 Isolates from Brazil Reveals Presence of Two Distinct Genetic Subtypes. *AIDS Res Hum Retroviruses.* 1994;10(5):561–7.
97. Sabino EC, Shpaer EG, Morgado MG, Korber BT, Diaz RS, Bongertz V, et al. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J Virol.* 1994;68(10):6340–6.
98. Bello G, Zanotto PM de A, Iamarino A, Gräf T, Pinto AR, Couto-Fernandez JC, et al. Phylogeographic Analysis of HIV-1 Subtype C Dissemination in Southern Brazil. Sandstrom P, editor. *PLoS One.* 2012;7(4):e35649.
99. Domingo E, Escarmís C, Sevilla N, Moya A, Elena SF, Quer J, et al. Basic concepts in RNA virus evolution. *FASEB J.* 1996;10(8):859–64.
100. Smith DB, Mcallister J, Casino C, Simmonds P. Printed in Great Britain Virus ' quasispecies ' : making a mountain out of a molehill? *J Gen Virol.* 1997;78:1511–9.
101. Dampier W, Nonnemacher MR, Mell J, Earl J, Ehrlich GD, Pirrone V, et al. HIV-1 Genetic Variation Resulting in the Development of New Quasispecies Continues to Be Encountered in the Peripheral Blood of Well-Suppressed Patients. *PLoS One.* 2016;11(5):e0155382.
102. Luring AS, Andino R. Quasispecies Theory and the Behavior of RNA Viruses. Manchester M, editor. *PLoS Pathog.* 2010;6(7):e1001005.
103. Zhang J, Xing H. [Application of Next-generation Sequencing Techniques in the Dynamics of HIV-1 Quasispecies]. *Bing Du Xue Bao.* 2015;31(5):573–8.
104. Bezuidt OKI, Makhalanyane TP, Gomri MA, Kharroub K, Cowan DA. Draft Genome Sequence of Thermophilic *Geobacillus* sp. Strain Sah69, Isolated from Saharan Soil,

- Southeast Algeria. *Genome Announc.* 2015;3(6).
105. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol.* 1999;73(1):152–60.
 106. Kimura M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature.* 1977;267(5608):275–6.
 107. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7.
 108. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One.* 2010;5(3):e9490.
 109. Bennett DE, Camacho RJ, Otelea D, Kuritzkes DR, Fleury H, Kiuchi M, et al. Drug Resistance Mutations for Surveillance of Transmitted HIV-1 Drug-Resistance: 2009 Update. Nixon DF, editor. *PLoS One.* 2009;4(3):e4724.
 110. Stamatakis A, Ludwig T, Meier H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics.* 2005;21(4):456–63.
 111. O’connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer’s MAP test. *Behav Res Methods, Instruments, Comput.* 2000;32(3):396–402.
 112. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28(12):1647–9.
 113. Töpfer A, Zagordi O, Prabhakaran S, Roth V, Halperin E, Beerenwinkel N. Probabilistic inference of viral quasispecies subject to recombination. *J Comput Biol.* 2013;20(2):113–23.
 114. Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol.* 1986;24(1–2):28–38.
 115. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* 2007;24(8):1596–9.
 116. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW. GARD: a genetic algorithm for recombination detection. *Bioinformatics.* 2006;22(24):3096–8.
 117. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 1989;123(3):585–95.
 118. Barreto CC, Nishyia A, Araújo L V, Ferreira JE, Busch MP, Sabino EC. Trends in

- antiretroviral drug resistance and clade distributions among HIV-1--infected blood donors in Sao Paulo, Brazil. *J Acquir Immune Defic Syndr*. 2006;41(3):338–41.
119. Brígido LFM, Ferreira JLP, Almeida VC, Rocha SQ, Ragazzo TG, Estevam DL, et al. Southern Brazil HIV type 1 C expansion into the state of São Paulo, Brazil. *AIDS Res Hum Retroviruses*. 2011;27(3):339–44.
 120. Alencar CS, Sabino EC, Carvalho SMF, Leao SC, Carneiro-Proietti AB, Capuani L, et al. HIV genotypes and primary drug resistance among HIV-seropositive blood donors in Brazil: role of infected blood donors as sentinel populations for molecular surveillance of HIV. *J Acquir Immune Defic Syndr*. 2013;63(3):387–92.
 121. Gonzalez CR, Alcalde R, Nishiya A, Barreto CC, Silva FES, de Almeida A, et al. Drug resistance among chronic HIV-1-infected patients naïve for use of anti-retroviral therapy in Sao Paulo city. *Virus Res*. 2007;129(2):87–90.
 122. Véras NMC, Gray RR, Brígido LF de M, Rodrigues R, Salemi M. High-resolution phylogenetics and phylogeography of human immunodeficiency virus type 1 subtype C epidemic in South America. *J Gen Virol*. 2011;92(Pt 7):1698–709.
 123. Lima K, de Souza Leal É, Cavalcanti AMS, Salustiano DM, de Medeiros LB, da Silva SP, et al. Epidemiological, Clinical and Antiretroviral Susceptibility Characterization of Human Immunodeficiency Virus Subtypes B and Non-B in Pernambuco, Northeast Brazil. *PLoS One*. 2016;11(5):e0155854.
 124. Easterbrook PJ, Smith M, Mullen J, O’Shea S, Chrystie I, de Ruiter A, et al. Impact of HIV-1 viral subtype on disease progression and response to antiretroviral therapy. *J Int AIDS Soc*. 2010;13:4.
 125. Cilliers T, Nhlapo J, Coetzer M, Orlovic D, Ketas T, Olson WC, et al. The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C. *J Virol*. 2003;77(7):4449–56.
 126. Coetzer M, Nedellec R, Cilliers T, Meyers T, Morris L, Mosier DE. Extreme genetic divergence is required for coreceptor switching in HIV-1 subtype C. *J Acquir Immune Defic Syndr*. 2011;56(1):9–15.
 127. Leite TCNF, Campos DP, Coelho AB, Teixeira SLM, Veloso V, Morgado MG, et al. Impact of HIV-1 Subtypes on AIDS Progression in a Brazilian Cohort. *AIDS Res Hum Retroviruses*. 2016; 28(3):339–44.
 128. Ferreira JLP, Rodrigues R, Lança AM, de Almeida VC, Rocha SQ, Ragazzo TG, et al. Transmitted Drug Resistance among People Living with HIV/Aids at Major Cities of Sao Paulo State, Brazil. *Adv Virol*. 2013;2013:878237.
 129. Inocencio LA, Pereira AA, Sucupira MCA, Fernandez JCC, Jorge CP, Souza DF, et al. Brazilian Network for HIV Drug Resistance Surveillance: a survey of individuals recently diagnosed with HIV. *J Int AIDS Soc*. 2009;12:20.

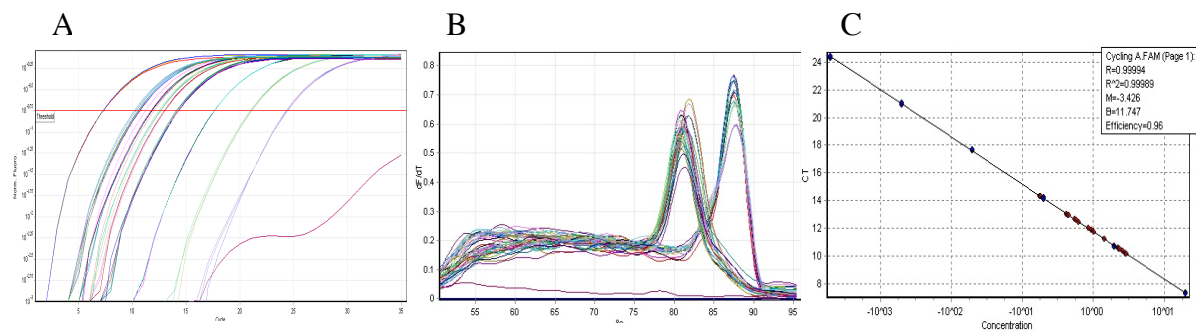
130. Sanabani SS, Pastena ÉR de S, da Costa AC, Martinez VP, Kleine-Neto W, de Oliveira ACS, et al. Characterization of partial and near full-length genomes of HIV-1 strains sampled from recently infected individuals in São Paulo, Brazil. *PLoS One*. 2011;6(10):e25869.
131. Guimarães PM de S, Ferreira JL de P, Coelho LPO, Cavalcanti J de S, Lopes GISL, Matsuda EM, et al. Transmitted Drug Resistance Among Recently Diagnosed Adults and Children in São Paulo, Brazil. *AIDS Res Hum Retroviruses*. 2015;31(12):1219–24.
132. de Medeiros RM, Junqueira DM, Matte MCC, Barcellos NT, Chies JAB, Matos Almeida SE. Co-circulation HIV-1 subtypes B, C, and CRF31_BC in a drug-naïve population from Southernmost Brazil: analysis of primary resistance mutations. *J Med Virol*. 2011;83(10):1682–8.
133. Baral S, Sifakis F, Cleghorn F, Beyrer C. Elevated risk for HIV infection among men who have sex with men in low- and middle-income countries 2000–2006: a systematic review. *PLoS Med*. 2007;4(12):e339.
134. Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. *AIDS*. 2005;19 Suppl 2:S67-72.
135. Bermúdez-Aza EH, Kerr LRFS, Kendall C, Pinho AA, de Mello MB, Mota RS, et al. Antiretroviral drug resistance in a respondent-driven sample of HIV-infected men who have sex with men in Brazil. *J Acquir Immune Defic Syndr*. 2011;57 Suppl 3:S186-92.
136. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med*. 2008;5(3):e50.
137. Hughes GJ, Fearnhill E, Dunn D, Lycett SJ, Rambaut A, Leigh Brown AJ, et al. Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. *PLoS Pathog*. 2009;5(9):e1000590.
138. German D, Sifakis F, Maulsby C, Towe VL, Flynn CP, Latkin CA, et al. Persistently high prevalence and unrecognized HIV infection among men who have sex with men in Baltimore: the BESURE study. *J Acquir Immune Defic Syndr*. 2011;57(1):77–87.
139. Rosenberg ES, Sullivan PS, DiNenno EA, Salazar LF, Sanchez TH. Number of casual male sexual partners and associated factors among men who have sex with men: Results from the National HIV Behavioral Surveillance system. *BMC Public Health*. 2011;11(1):189.
140. Baggaley RF, White RG, Boily M-C. HIV transmission risk through anal intercourse: systematic review, meta-analysis and implications for HIV prevention. *Int J Epidemiol*. 2010;39(4):1048–63.
141. Drescher SM, von Wyl V, Yang W-L, Böni J, Yerly S, Shah C, et al. Treatment-naive individuals are the major source of transmitted HIV-1 drug resistance in men who

- have sex with men in the Swiss HIV Cohort Study. *Clin Infect Dis*. 2014;58(2):285–94.
142. Bello G, Eyer-Silva WA, Couto-Fernandez JC, Guimarães ML, Chequer-Fernandez SL, Teixeira SLM, et al. Demographic history of HIV-1 subtypes B and F in Brazil. *Infect Genet Evol*. 2007;7(2):263–70.
 143. Gräf T, Passaes CPB, Ferreira LGE, Grisard EC, Morgado MG, Bello G, et al. HIV-1 genetic diversity and drug resistance among treatment naïve patients from Southern Brazil: an association of HIV-1 subtypes with exposure categories. *J Clin Virol*. 2011;51(3):186–91.
 144. Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH. Nature, Position, and Frequency of Mutations Made in a Single Cycle of HIV-1 Replication. *J Virol*. 2010;84(19):9864–78.
 145. O’Neil PK. Mutational Analysis of HIV-1 Long Terminal Repeats to Explore the Relative Contribution of Reverse Transcriptase and RNA Polymerase II to Viral Mutagenesis. *J Biol Chem*. 2002;277(41):38053–61.
 146. Domingo E, Sheldon J, Perales C. Viral Quasispecies Evolution. *Microbiol Mol Biol Rev*. 2012;76(2):159–216.
 147. Pawlotsky JM. Hepatitis C virus population dynamics during infection. *Curr Top Microbiol Immunol*. 2006;299:261–84.
 148. Johnson VA, Calvez V, Günthard HF, Paredes R, Pillay D, Shafer R, et al. 2011 update of the drug resistance mutations in HIV-1. *Top Antivir Med*. 2011;19(4):156–64.
 149. Barzon L, Lavezzo E, Militello V, Toppo S, Palù G. Applications of Next-Generation Sequencing Technologies to Diagnostic Virology. *Int J Mol Sci*. 2011;12(12):7861–84.
 150. Capobianchi MR, Giombini E, Rozera G. Next-generation sequencing technology in clinical virology. *Clin Microbiol Infect*. 2013;19(1):15–22.
 151. Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. *J Gen Virol*. 2012;93(Pt_9):1853–68.
 152. Fisher R, van Zyl GU, Travers SAA, Kosakovsky Pond SL, Engelbrech S, Murrell B, et al. Deep Sequencing Reveals Minor Protease Resistance Mutations in Patients Failing a Protease Inhibitor Regimen. *J Virol*. 2012;86(11):6231–7.
 153. Lauck M, Alvarado-Mora M V., Becker EA, Bhattacharya D, Striker R, Hughes AL, et al. Analysis of Hepatitis C Virus Intrahost Diversity across the Coding Region by Ultradeep Pyrosequencing. *J Virol*. 2012;86(7):3952–60.

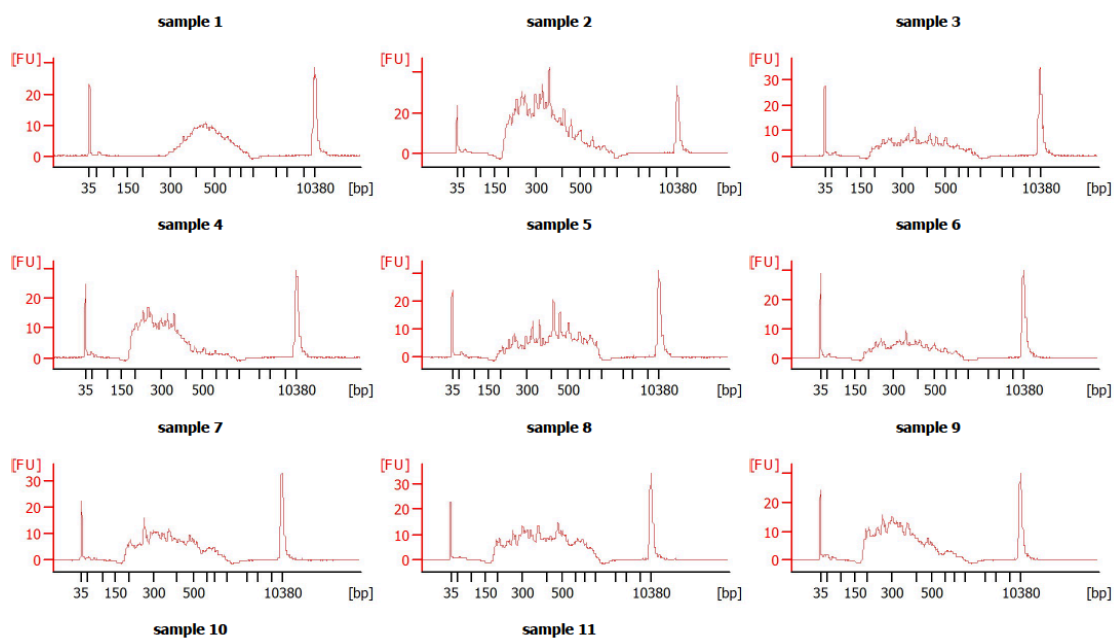
154. Simen BB, Braverman MS, Abbate I, Aerssens J, Bidet Y, Bouchez O, et al. An international multicenter study on HIV-1 drug resistance testing by 454 ultra-deep pyrosequencing. *J Virol Methods*. 2014;204:31–7.
155. Di Giallonardo F, Zagordi O, Duport Y, Leemann C, Joos B, Künzli-Gontarczyk M, et al. Next-Generation Sequencing of HIV-1 RNA Genomes: Determination of Error Rates and Minimizing Artificial Recombination. Wainberg M, editor. *PLoS One*. 2013 18;8(9):e74249.
156. Poveda E. Multi-step inhibition against HIV lifecycle-underlying the “magic” of protease inhibitors. *AIDS Rev*. 2014;16(1):52.
157. Thys K, Verhasselt P, Reumers J, Verbist BMP, Maes B, Aerssens J. Performance assessment of the Illumina massively parallel sequencing platform for deep sequencing analysis of viral minority variants. *J Virol Methods*. 2015;221:29–38.
158. Simen BB, Simons JF, Hullsiek KH, Novak RM, MacArthur RD, Baxter JD, et al. Low-Abundance Drug-Resistant Viral Variants in Chronically HIV-Infected, Antiretroviral Treatment–Naive Patients Significantly Impact Treatment Outcomes. *J Infect Dis*. 2009;199(5):693–701.
159. Noguera-Julian M, Cozzi-Lepri A, Di Giallonardo F, Schuurman R, Däumer M, Aitken S, et al. Contribution of APOBEC3G/F activity to the development of low-abundance drug-resistant human immunodeficiency virus type 1 variants. *Clin Microbiol Infect*. 2016;22(2):191–200.
160. Oliveira M, Ibanescu RI, Pham HT, Brenner B, Mesplède T, Wainberg MA. The M184I/V and K65R nucleoside resistance mutations in HIV-1 prevent the emergence of resistance mutations against dolutegravir. *AIDS*. 2016;30(15):2267–73.
161. Beerenwinkel N, Zagordi O. Ultra-deep sequencing for the analysis of viral populations. *Curr Opin Virol*. 2011;1(5):413–8.
162. Salemi M. The intra-host evolutionary and population dynamics of human immunodeficiency virus type 1: a phylogenetic perspective. *Infect Dis Rep*. 2013;5(Suppl 1):e3.
163. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. *Nat Rev Genet*. 2004;5(1):52–61.
164. Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol*. 1999;73(12):10489–502.
165. Granade TC, Nguyen S, Kuehl DS, Parekh BS. Development of a novel rapid HIV test for simultaneous detection of recent or long-term HIV type 1 infection using a single testing device. *AIDS Res Hum Retroviruses*. 2013;29(1):61–7.
166. Edwards CTT, Holmes EC, Wilson DJ, Viscidi RP, Abrams EJ, Phillips RE, et al.

- Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1. *BMC Evol Biol.* 2006;6:28.
167. Zhang LQ, MacKenzie P, Cleland A, Holmes EC, Brown AJ, Simmonds P. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *J Virol.* 1993;67(6):3345–56.
 168. Zhu T, Mo H, Wang N, Nam DS, Cao Y, Koup RA, et al. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science.* 1993;261(5125):1179–81.
 169. Tully DC, Ogilvie CB, Batorsky RE, Bean DJ, Power KA, Ghebremichael M, et al. Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus. *PLoS Pathog.* 2016;12(5):e1005619.
 170. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, et al. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *J Virol.* 2008;82(8):3952–70.
 171. Li H, Bar KJ, Wang S, Decker JM, Chen Y, Sun C, et al. High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men. *PLoS Pathog.* 2010;6(5):e1000890.
 172. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, et al. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J Exp Med.* 2009;206(6):1273–89.
 173. Plikat U, Nieselt-Struwe K, Meyerhans A. Genetic drift can dominate short-term human immunodeficiency virus type 1 nef quasispecies evolution in vivo. *J Virol.* 1997;71(6):4233–40.
 174. Pandit A, Sinha S. Differential Trends in the Codon Usage Patterns in HIV-1 Genes. Kim P, editor. *PLoS One.* 2011;6(12):e28889.
 175. Meintjes PL, Rodrigo AG. Evolution of Relative Synonymous Codon Usage in Human Immunodeficiency Virus Type-1. *-Pacific Bioinforma Conf Conf Res Pract Inf Technol.* 2004;29.
 176. Lin Y-Z, Sun L-K, Zhu D-T, Hu Z, Wang X-F, Du C, et al. Equine schlafen 11 restricts the production of equine infectious anemia virus via a codon usage-dependent mechanism. *Virology.* 2016;495:112–21.
 177. Ahn I, Son HS. Epidemiological comparisons of codon usage patterns among HIV-1 isolates from Asia, Europe, Africa and the Americas. *Exp Mol Med.* 2006;38(6):643–51.
 178. Neogi U, Shet A, Sahoo PN, Bontell I, Ekstrand ML, Banerjea AC, et al. Human APOBEC3G-mediated hypermutation is associated with antiretroviral therapy failure in HIV-1 subtype C-infected individuals. *J Int AIDS Soc.* 2013;16:18472.

179. Pennings PS, Holmes SP, Shafer RW. HIV-1 transmission networks in a small world. *J Infect Dis.* 2014;209(2):180–2.
180. Briones C, Domingo E. Minority report: hidden memory genomes in HIV-1 quasispecies and possible clinical implications. *AIDS Rev.* 10(2):93–109.
181. Churchill M, Nath A. Where does HIV hide? A focus on the central nervous system. *Curr Opin HIV AIDS.* 2013;8(3):165–9.
182. Kanmogne GD, Kennedy RC, Grammas P. Analysis of human lung endothelial cells for susceptibility to HIV type 1 infection, coreceptor expression, and cytotoxicity of gp120 protein. *AIDS Res Hum Retroviruses.* 2001;17(1):45–53.
183. Iamarino A, de Melo FL, Braconi CT, Zanotto PM de A. BF integrase genes of HIV-1 circulating in São Paulo, Brazil, with a recurrent recombination region. *PLoS One.* 2012;7(4):e34324.
184. Ross HA, Rodrigo AG. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. *J Virol.* 2002;76(22):11715–20.

APÊNDICE A – QUANTIFICAÇÃO DAS BIBLIOECAS PELO PCR EM TEMPO REAL.

APÊNDICE B – ANÁLISE DO PRODUTO DAS BIBLIOTECAS REVELADO PELO BIOANALYZER.



APÊNDICE C- DESCRIÇÃO DAS VARIÁVEIS SEGUNDO O TRATAMENTO (NAIVE OU TRATADO).

Variables	Subtype_B				Subtype_C n(82)				Subtype_F1 n(185)			
	Naive n (353)	Treated n (1435)	Total	p	Naive n (52)	Treated n (24)	Total	p	Naive n (22)	Treated n (124)	Total	p
Age median (IQR)	30 (25-38)	40 (32-47)	38 (30-46)	< 0,001	31,5 (28-38,5)	40 (31-46)	34 (28-42)	0,012	34 (26-47,5)	40,5 (33-45)	34 (28-42)	0,334
Gender, n (%)				< 0,001				0,020				0,151
Male	292 (83,7%)	864 (62,1%)	1156(66,4%)		41 (80,4%)	10 (52,6%)	51 (72,9%)		15 (68,2%)	64 (51,6%)	79 (54,1%)	
Female	57 (16,3%)	528 (37,9%)	585 (33,6%)		10 (19,6%)	9 (47,4%)	19 (27,1%)		7 (31,8%)	60 (48,4)	67 (45,9%)	
Risk Factor, n (%)				< 0,001				< 0,001				0,002
Het	74 (23,2%)	550 (76,6%)	624 (60,2%)		13 (28,3%)	9 (90%)	22 (39,3%)		8 (38,1%)	26 (60,5%)	34 (53,1%)	
MSM	236 (74%)	73 (10,2%)	309 (29,8%)		33 (71,7%)	1 (10%)	34 (60,7%)		12 (57,1%)	7 (16,3%)	19 (29,7%)	
PWID	1 (0,3%)	16 (2,2%)	17 (1,6%)		0 (0%)	0 (0%)	0 (0%)		0 (0%)	0 (0%)	0 (0%)	
VT	8 (2,5%)	79 (11%)	87 (8,4%)		0 (0%)	0 (0%)	0 (0%)		1 (4,8%)	10 (23,3%)	11 (17,2%)	
Chains, n (%)				< 0,001				< 0,001				*0,004
Clusters	136 (38,5%)	173 (12,1%)	309 (17,3%)		31 (59,6%)	1 (4,2%)	32 (42,1%)		8 (36,4%)	14 (10,6%)	22 (14,3%)	
Non clusters	217 (61,5%)	1262 (87,9%)	1479 (82,7%)		21 (40,4%)	23 (95,8%)	44 (57,9%)		14 (63,6%)	118 (89,4%)	132 (85,7%)	
Type of cluster, n (%)				< 0,001				1				*0,006
Pair clusters	76 (55,9%)	133 (76,9%)	209 (67,6%)		15 (48,4%)	1 (100%)	16 (50%)		1 (12,5%)	11 (78,6%)	12 (54,5%)	
>2 seq clusters	60 (44,1%)	50 (23,1%)	100 (32,4%)		16 (51,6%)	0 (0%)	16 (50%)		7 (87,5%)	3 (21,4%)	10 (45,5%)	
CPR mutations, n (%)	43 (12,2%)	1260 (87,8%)	1303 (72,9%)	< 0,001	2 (3,8%)	20 (83,3%)	22 (28,9%)	< 0,001	4 (18,2%)	113 (85,6%)	117 (76%)	< 0,001
CPR profile mutations, n (%)												
PI mutations	8 (2,3%)	642 (44,7%)	650 (36,4%)	< 0,001	1 (1,9%)	4 (16,7%)	5 (6,6%)	0,032	0 (0%)	56 (42,4%)	56 (36,4)	< 0,001
NRTI mutations	14 (4%)	1172 (81,7%)	1186 (66,3%)	< 0,001	1 (1,9%)	17 (70,8%)	18 (23,7%)	< 0,001	2 (9,1%)	100 (75,8%)	102 (66,2%)	< 0,001
NNRTI mutatinos	28 (7,9%)	834 (58,1%)	862 (48,2%)	< 0,001	1 (1,9%)	14 (58,3%)	15 (19,7%)	< 0,001	2 (9,1%)	79 (59,8%)	81 (52,6%)	< 0,001
Non CPR mutations, n (%)												
T74S	1 (0,3%)	83 (5,8%)	84 (4,7%)	< 0,001	8 (15,4%)	1 (4,2%)	9 (11,8%)	*0,258	0 (0%)	14 (10,6%)	14 (9,1%)	*0,222
E138A	10 (2,8%)	49 (3,4%)	59 (3,3%)	*0,585	2 (3,8%)	2 (8,3%)	4 (5,3%)	*0,587	1 (4,5%)	7 (5,3%)	8 (5,2%)	*1
E138K	1 (0,3%)	3 (0,2%)	4 (0,2%)	*0,585	2 (3,8%)	1 (4,2%)	3 (3,9%)	*1	1 (4,5%)	0 (0%)	1 (0,6%)	*0,143
CD4 median (IQR)	464 (324-642)	241 (117-404)	268 (136-452)	< 0,001	545 (400-703)	204 (106-274)	395 (204-577)	< 0,001	526 (282-673,7)	285 (157-454)	395 (204-577)	0,003
VL(log) median (IQR)	4,39 (3,18-4,89)	4,40 (3,82-4,95)	4,37 (3,79-4,94)	0,874			4,45 (3,69-5,20)	0,820	4,42 (3,57-5,3)	4,3 (3,72-4,9)	4,43 (3,69-5,2)	0,899

APÊNDICE D – DESCRIÇÃO DAS VARIÁVEIS SEGUNDO OS PACIENTES QUE ESTÃO EM CLUSTER E FORA DE CLUSTER.

Variables	Subtype_B n (1993)				Subtype_C n(82)				Subtype_F1 n (185)			
	Non cluster n (1640)	Cluster n (353)	Total	p	Non cluster n (49)	Cluster n (33)	Total	p	Non cluster n (157)	Cluster n (28)	Total	p
Age median (IQ)	39 (31-47)	33 (26-41)	38 (30-46)	<0,001	35 (29,5-44,5)	30 (27,2-36,7%)	34 (28-42)	0,037	40 (31-45)	39,5 (28,7-44,2)	40 (31-45)	0,987
Gender, n (%)				0,001				0,012				0,124
Male	1002 (64,1%)	246 (73,7%)	1248 (65,8%)		26 (60,5%)	27 (87,1%)	53 (71,6%)		74 (51,4%)	17 (68%)	91 (53,8%)	
Female	560 (35,9%)	88 (26,3%)	648 (34,2%)		17 (39,5%)	4 (12,9%)	21 (28,4%)		70 (48,6%)	8 (32%)	78 (46,2%)	
Risk Factor, n (%)				<0,001				<0,001				*0,011
Het	592 (67,3%)	96 (41,4%)	688 (61,9%)		20 (66,7%)	4 (14,3%)	24 (41,4%)		37 (60,7%)	4 (36,4%)	41 (56,9%)	
MSM	200 (22,8%)	119 (51,3%)	319 (28,7%)		10 (33,3%)	24 (85,7%)	34 (58,6%)		13 (21,3%)	7 (63,6%)	20 (27,8%)	
PWID	14 (1,6%)	3 (1,3%)	17 (1,5%)		0 (0%)	0 (0%)	0 (0%)		0 (0%)	0 (0%)	0 (0%)	
VT	73 (8,3%)	14 (6%)	87 (7,8%)		0 (0%)	0 (0%)	0 (0%)		11 (18%)	0 (0%)	11 (15,3%)	
Type of cluster, n (%)												
Pair clusters		240 (68%)	240 (68%)			16 (48,5%)	16 (48,5%)			15 (53,6%)	15 (53,6%)	
>2 seq clusters		113 (32%)	113 (32%)			17 (51,5%)	17 (51,5%)			13 (46,4%)	13 (46,4%)	
CPR mutations, n (%)	1257 (76,6%)	197 (55,8%)	1454 (73%)	<0,001	21 (42,9%)	2 (6,1%)	23 (28%)	<0,001	118 (75,2%)	21 (75%)	139 (75,1%)	1
TDR, n (%)	22 (1,5%)	21 (6,8%)	43 (2,1%)	<0,001	1 (2,3%)	1 (3,1%)	2 (2,4%)	1	1 (0,8%)	3 (13,6%)	4 (2,1%)	*0,009
CPR profile mutations, n (%)												
PI mutations	652 (39,8%)	78 (22,1%)	730 (36,6%)	<0,001	4 (8,2%)	1 (3%)	5 (6,1%)	*0,643	58 (36,9%)	7 (25%)	65 (35,1%)	0,223
NRTI mutations	1150 (70,1%)	169 (47,9%)	1319 (66,2%)	<0,001	19 (38,8%)	0 (0%)	19 (23,2%)	<0,001	104 (66,2%)	18 (64,3%)	122 (65,9%)	0,841
NNRTI mutatinos	827 (50,4%)	145 (41,1%)	972 (48,8%)	<0,001	14 (28,6%)	2 (6,1%)	16 (19,5%)	0,012	83 (52,9%)	14 (50%)	97 (52,4%)	0,78
Non CPR mutations, n (%)												
T74S	82 (5%)	9 (2,5%)	91 (4,6%)	0,045	2 (4,1%)	7 (21,2%)	9 (10,98%)	*0,027	15 (9,6%)	0 (0%)	15 (8,1%)	*0,132
E138A	56 (3,4%)	11 (3,1%)	67 (3,4%)	0,778	2 (4,1%)	3 (9,1%)	5 (6,1%)	*0,387	7 (4,5%)	1 (3,6%)	8 (4,3%)	*1
E138K	2 (0,1%)	3 (0,8%)	5 (0,3%)	*0,042	2 (4,1%)	1 (3%)	3 (3,7%)	*1	1 (0,6%)	1 (3,6%)	2 (1,1%)	*0,28
CD4 median (IQ)	263 (131-446)	299 (153-516)	268 (136-452)	0,039	261,5 (153-448)	549 (371-697)	395 (204-577)	0,003	294 (170-500)	318 (194-537)	303 (175-504)	0,748
VL(log) median (IQ)	4,37 (3,79-4,92)	4,39 (3,81-4,99)	4,30 (3,79-4,94)	0,372	4,31 (3,45-5,22)	4,52 (4,08-5,21)	4,45 (3,69-5,20)	0,254	4,39 (3,72-4,98)	4,42 (3,57-4,99)	4,39 (3,70-4,98)	0,94

APÊNDICE E – DESCRIÇÃO DAS VARIÁVEIS SEGUNDO GRUPO DE RISCO.

Variables	Subtype_B n (1111)				Subtype_C n(58)				Subtype_F1 n (72)			
	Non MSM n (792)	MSM n (319)	Total	p	Non MSM n (24)	MSM n (34)	Total	p	Non MSM n (52)	MSM n (20)	Total	p
Age median (IQR)	38 (28-45)	30 (26-38,5)	35 (27-44)	< 0,001	34 (26-42)	30 (28-36,5)	31 (27,2-37,7)	0,499	40 (22-45)	37 (27-44)	39 (26-45)	0,751
Gender, n (%)				< 0,001				< 0,001				< 0,001
Male	144 (18,2%)	319 (100%)	462 (41,6%)		3 (12,5%)	34 (100%)	37 (63,8%)		12 (23,1%)	20 (100%)	32 (44,4%)	
Female	648 (81,8%)	0 (0%)	648 (58,4%)		21 (87,5%)	0 (0%)	21 (36,2%)		40 (76,9%)	0 (0%)	40 (55,6%)	
Therapy, n (%)				< 0,001				< 0,001				0,001
Naive	83 (11,4%)	236 (76,4%)	319 (30,8%)		13 (59,1%)	33 (97,1%)	46 (82,1%)		9 (20%)	12 (63,2%)	21 (32,8%)	
Treated	645 (88,6%)	73 (23,6%)	718 (69,2%)		9 (40,9%)	1 (2,9%)	10 (17,9%)		36 (80%)	7 (36,8%)	43 (67,2%)	
Chains, n (%)				< 0,001				< 0,001				*0,008
Clusters	113 (14,3%)	119 (37,3%)	232 (20,8%)		4 (16,7%)	24 (70,6%)	28 (48,3%)		4 (7,7%)	7 (35%)	11 (15,3%)	
Non clusters	679 (85,7%)	200 (62,7%)	879 (79,2%)		20 (83,3%)	10 (29,4%)	30 (51,7%)		48 (92,3%)	13 (65%)	61 (84,7%)	
Type of cluster, n (%)				< 0,001				*0,6				*0,576
Pair clusters	94 (83,2%)	57 (47,9%)	151 (65,1%)		3 (75%)	12 (50%)	15 (53,6%)		2 (50%)	2 (28,6%)	4 (36,4%)	
>2 seq clusters	19 (16,8%)	62 (52,1%)	81 (34,9%)		1 (25%)	12 (50%)	13 (46,4%)		2 (50%)	5 (71,4%)	7 (63,6%)	
CPR mutations, n (%)	615 (77,7%)	95 (29,8%)	710 (63,9%)	< 0,001	7 (29,2%)	1 (2,9%)	8 (13,8%)	*0,007	39 (75%)	9 (45%)	48 (66,7%)	0,016
TDR, n (%)	11 (1,5%)	27 (8,7%)	38 (3,7%)	< 0,001	0 (0%)	1 (2,9%)	1 (1,8%)	*1	0 (0%)	4 (21%)	4 (6,3%)	*0,006
CPR profile mutations, n (%)												
PI mutations	304 (38,4%)	44 (13,8%)	348 (31,3%)	< 0,001	0 (0%)	1 (2,9%)	1 (1,7%)		17 (32,7%)	2 (10%)	19 (26,4%)	0,05
NRTI mutations	551 (69,6%)	69 (21,6%)	620 (55,8%)	< 0,001	5 (20,8%)	0 (0%)	5 (8,6%)		35 (67,3%)	7 (35%)	42 (58,3%)	0,013
NNRTI mutatinos	406 (51,3%)	58 (18,2%)	464 (41,8%)	< 0,001	6 (25%)	1 (2,9%)	7 (12,1%)		26 (50%)	6 (30%)	32 (44,4%)	0,126
Non CPR mutations, n (%)												
T74S	41 (5,2%)	3 (0,9%)	44 (4%)	0,001	2 (8,3%)	6 (17,6%)	8 (13,8%)	*0,449	3 (5,8%)	0 (0%)	3 (4,2%)	*0,555
E138A	19 (2,4%)	13 (4,1%)	32 (2,9%)	0,131	2 (8,3%)	1 (2,9%)	3 (5,2%)	*0,564	0 (0%)	2 (10%)	2 (2,8%)	*0,074
E138K	2 (0,3%)	1 (0,3%)	3 (0,3%)	*1	0 (0%)	2 (5,9%)	2 (3,4%)	*0,506	0 (0%)	1 (5%)	1 (1,4%)	*0,278
CD4 median (IQR)	255 (130-436,2)	407,5 (247,2-603,7)	296 (147,7-482,5)	< 0,001	254 (182,2-606,5)	526 (380-615,5)	450 (260-614)	0,066	334 (185-592)	410 (238-544)	379 (201,7-548)	0,636
VL(log) median (IQR)	4,37(3,81-4,94)	4,42 (3,85-4,91)	4,37 (3,82-4,93)	0,793	3,77 (3,20-5,14)	4,63 (4,11-5,17)	4,41 (3,74-5,14)	0,024	4,59 (3,73-5,1)	4,31 (3,30-5,09)	4,51 (3,69-5,1)	0,784

APÊNDICE F – PADRÃO DE USO RELATIVO DE CÓDONS (RSCU) PARA OS DOIS SUBTIPOS (B E C) DE HIV-1 E SEU HOSPEDEIRO.

Aminoácido	Códons	HIV-1 B	HIV-1 C	Humano
	UUU	1.32	1.3	0.93
Fen	UUC	0.68	0.67	1.07
	UUA	2.76	2.23	0.46
	UUG	1	1.2	0.77
	CUU	0.7	0.89	0.79
	CUC	0.15	0.46	1.17
	CUA	0.63	0.78	0.43
Leu	CUG	0.74	0.44	2.37
	AUU	0.9	0.92	1.08
	AUC	0.49	0.49	1.41
Ile	AUA	1.61	1.58	0.51
Met	AUG	1	1	1
	GUU	0.66	0.5	0.73
	GUC	0.4	0.5	0.95
	GUA	2.43	2.29	0.47
Val	GUG	0.5	0.7	1.85
	UCU	0.51	0.55	1.13
	UCC	0.1	0	1.31
	UCA	2.47	2.55	0.9
Ser	UCG	0.11	0	0.33
	CCU	1.1	1.13	1.15
	CCC	0.58	0.39	1.29
	CCA	2.3	2.43	1.11
Pro	CCG	0.02	0.04	0.45
	ACU	1.49	1.29	0.99
	ACC	0.33	0.37	1.42
	ACA	2.16	2.28	1.14
Tre	ACG	0.02	0.07	0.46
	GCU	0.89	0.86	1.06
Ala	GCC	0.35	0.53	1.60

	GCA	2.72	2.51	0.91
	GCG	<i>0.04</i>	<i>0.1</i>	<i>0.42</i>
	UAU	1.38	1.97	0.89
Tir	UAC	0.62	<i>0.03</i>	1.11
	CAU	1.8	1.47	0.84
His	CAC	<i>0.2</i>	<i>0.53</i>	1.16
	CAA	1.26	1.52	0.84
Glu	CAG	0.74	<i>0.48</i>	1.16
	AAU	1.52	1.26	0.94
Asn	AAC	<i>0.48</i>	0.74	1.06
	AAA	1.47	1.34	0.87
Lis	AAG	<i>0.53</i>	0.66	1.13
	GAU	1.35	1.03	0.93
Asp	GAC	0.65	0.97	1.07
	GAA	1.62	1.58	0.84
Glu	GAG	<i>0.38</i>	<i>0.42</i>	1.16
	UGU	1.17	1.46	0.91
Cis	UGC	0.83	<i>0.54</i>	1.09
Trp	UGG	1	1	1
	CGU	<i>0.02</i>	<i>0</i>	<i>0.48</i>
	CGC	<i>0</i>	<i>0</i>	1.10
	CGA	<i>0.48</i>	<i>0.48</i>	0.65
	CGG	<i>0</i>	<i>0</i>	1.21
	AGA	4.6	4.27	1.29
Arg	AGG	0.91	1.26	1.27
	AGU	2.16	2.23	0.99
Ser	AGC	0.67	0.67	1.50
	GGU	<i>0.53</i>	<i>0.28</i>	0.65
	GGC	<i>0.26</i>	<i>0.23</i>	1.35
	GGA	2.16	2.25	1
Gli	GGG	1.0	1.23	1

APÊNDICE G – A TABELA SUMARIZA OS VALORES OBTIDOS DE CAI (ÍNDICE DE ADAPTAÇÃO DO CÓDON).

Name	Length	CAI	Nc
70_13_47_27	957	0,781	43,1
70_13_27_52	957	0,781	43,5
70_13_9_48	957	0,777	43,8
70_13_4_70	957	0,779	42,6
70_13_4_14	957	0,779	43
70_13_1_95	957	0,782	42,9
70_13_1_5	957	0,784	42,4
70_13_1_31	957	0,781	43,1
89_13_30_37	957	0,769	40,4
89_13_21_32	957	0,774	40,2
89_13_14_92	957	0,771	41,5
89_13_11_84	957	0,771	40,9
89_13_7_42	957	0,77	41,1
89_13_4_66	957	0,775	40,9
89_13_2_62	957	0,771	40,4
89_13_2_37	957	0,772	40,3
89_13_1_92	957	0,771	41,1
104_13_95_79	957	0,772	40,9
104_13_1_12	957	0,774	41
104_13_0_47	957	0,769	41,2
137_14_80_62	957	0,765	43,1
137_14_15_76	957	0,765	43,7
137_14_1_59	957	0,769	44,6
137_14_0_80	957	0,765	44,1
139_14_98_26	957	0,767	41,5
139_14_0_62	957	0,765	41,5
139_14_0_32	957	0,768	41,6
139_14_0_29	957	0,768	41,6

169_14_20	957	0,782	38,1
169_14_2_36	957	0,782	37,9
169_14_1_84	957	0,782	38,1
FO1_70_13_219_13_8_5	957	0,783	43,2
FO1_70_13_219_13_3_41	957	0,783	43,3
FO1_70_13_219_13_1_4	957	0,78	43,5
FO1_70_13_219_13_1_16	957	0,78	43,5
FO1_70_13_219_13_1_13	957	0,781	43,5
FO1_70_13_219_13_1_07	957	0,782	43,2
258_13_88_43	957	0,772	39,7
258_13_5_16	957	0,771	40
258_13_2_34	957	0,773	39,7
258_13_0_61	957	0,773	39,5
258_13_0_42	957	0,772	39,9
FO1_104_13_272_13_96_39	957	0,776	41
FO1_104_13_272_13_0_50	957	0,777	40,2
FO1_104_13_272_13_0_48	957	0,772	42,6
FO1_104_13_272_13_0_25	957	0,77	40,9
317_11_100_0	957	0,774	42,2
FO1_89_13_406_13_69_27	957	0,769	40,3
FO1_89_13_406_13_29_13	957	0,769	41,2
451_13_99_44	957	0,774	41,2
451_13_0_55	957	0,771	41,6
582_14_99_55	957	0,77	41,2
582_14_0_44	957	0,776	41,3
599_13_41_95	957	0,773	40,7
599_13_19_84	957	0,776	41,4
599_13_16_06	957	0,773	41
599_13_11_08	957	0,775	40,7
599_13_6_55	957	0,774	40,8
599_13_4_49	957	0,777	41,3

600_13_62_47	957	0,78	41,2
600_13_20_69	957	0,777	41,1
600_13_4_05	957	0,777	41,4
600_13_1_11	957	0,774	41,3
FO1_258_13_639_13_100_0	957	0,767	41,2
650_14_33_47	957	0,773	40,1
650_14_8_6	957	0,772	40,8
650_14_7_2	957	0,773	40,2
650_14_5_08	957	0,773	40
650_14_3_75	957	0,772	40,1
650_14_3_47	957	0,773	39,6
650_14_2_9	957	0,773	40
650_14_2_2	957	0,771	40
650_14_1_5	957	0,773	40,8
717_14_56_4	957	0,775	40,3
717_14_16_4	957	0,774	40,1
717_14_1_2	957	0,774	40,1
717_14_1_02	957	0,774	40,2
717_14_1_1	957	0,774	39,9
717_14_1_0	957	0,772	40,1
750_14_14_85	957	0,774	37,5
750_14_12_79	957	0,771	37,6
750_14_12_07	957	0,774	37,6
750_14_8_234	957	0,775	37,6
750_14_5_95	957	0,774	37,6
750_14_5_84	957	0,772	37,8
750_14_5_42	957	0,772	37,9
750_14_4_70	957	0,774	38
750_14_3_25	957	0,772	37,1
750_149_3_13	957	0,772	37,9
750_14_2_89	957	0,769	38,2

750_14_2_40	957	0,774	37,8
750_14_2_19	957	0,778	38,2
750_14_1_98	957	0,777	37,4
750_14_1_84	957	0,766	38
750_14_1_68	957	0,775	38,6
750_14_1_44	957	0,774	37,6
750_14_1_41	957	0,77	38,1
750_14_1_10	957	0,768	38,4
760_14_51_82	957	0,774	42
760_14_12_05	957	0,774	42,1
760_14_4_32	957	0,774	42
864_13_24_5	957	0,761	37,6
864_13_19_44	957	0,758	38,2
864_13_7_02	957	0,759	38,5
864_13_6_40	957	0,765	36,6
864_13_6_19	957	0,763	37,2
864_13_4_08	957	0,757	37,5
864_13_3_06	957	0,762	38,2
864_13_2_5	957	0,766	37,1
864_13_2_20	957	0,754	35,7
864_13_1_88	957	0,758	36,5
864_13_1_70	957	0,755	38
864_13_1_28	957	0,766	37,2
864_13_1_27	957	0,762	38,8
864_13_1_10	957	0,76	37,1
1270_14_33	957	0,781	41,7
1270_14_19_2	957	0,781	41,6
1534_14_40_2	957	0,773	43,5
1534_14_12_5	957	0,771	43,5
1534_14_7_0	957	0,772	43,4
1534_14_5_8	957	0,772	43,5

1534_14_5_6	957	0,772	43,7
1534_14_5_2	957	0,774	43,6
1534_14_4_5	957	0,771	43,6
1534_14_3_96	957	0,771	43,3
1534_14_3_85	957	0,771	43,4
1534_14_3_6	957	0,773	43,5
1534_14_1_91	957	0,772	43,5
1534_14_1_87	957	0,772	43,3
FO1_864_13_113_14_100	957	0,759	38,5
