

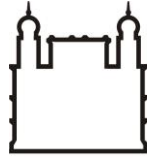
INSTITUTO OSWALDO CRUZ

Mestrado em Biologia Parasitária

CARACTERIZAÇÃO DE VARIANTES VIRAIS DE HIV-1 EM INDIVÍDUOS SOROPOSITIVOS COM PERFIL DE CONTROLE DA PROGRESSÃO PARA A AIDS E DA REPLICAÇÃO VIRAL: AVALIAÇÃO DA OCORRÊNCIA DE VARIANTES DE ESCAPE DA RESPOSTA IMUNE

DIOGO GAMA CAETANO

Rio de Janeiro
2016



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Pós-Graduação em Biologia Parasitária

Diogo Gama Caetano

Caracterização de variantes virais de HIV-1 em indivíduos soropositivos com perfil de controle da progressão para a AIDS e da replicação viral: avaliação da ocorrência de variantes de escape da resposta imune

Dissertação apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Mestre em Ciências

Orientador: Dr^a Mariza Gonçalves Morgado

Rio de Janeiro

2016

Ficha catalográfica elaborada pela
Biblioteca de Ciências Biomédicas/ ICICT / FIOCRUZ - RJ

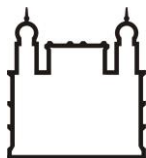
C128 Caetano, Diogo Gama

Caracterização de variantes virais de HIV-1 em indivíduos soropositivos com perfil de controle da progressão para a AIDS e da replicação viral: avaliação da ocorrência de variantes de escape da resposta imune / Diogo Gama Caetano. – Rio de Janeiro, 2016.
xx, 97 f. : il. ; 30 cm.

Dissertação (Mestrado) – Instituto Oswaldo Cruz, Pós-Graduação em Biologia Parasitária, 2016.
Bibliografia: f. 69-83

1. HIV. 2. LTNP. 3. Controladores. 4. Evolução viral. 5. Escape CTL.
I. Título.

CDD 616.9792



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Pós-Graduação em Biologia Parasitária

Diogo Gama Caetano

Caracterização de variantes virais de HIV-1 em indivíduos soropositivos com perfil de controle da progressão para a AIDS e da replicação viral: avaliação da ocorrência de variantes de escape da resposta imune

Orientador: Dr^a Mariza Gonçalves Morgado

Aprovada em: 03/03/2016

EXAMINADORES:

Dr^a Ana Carolina Paulo Vicente – **Presidente**

Dr. Carlos Roberto Alves

Dr. André Felipe Andrade dos Santos

Dr^a Nathalia Motta de Araújo - Suplente

Dr^a Luciana Jesus da Costa - Suplente

Rio de Janeiro

2016

III

Este trabalho foi realizado no
Laboratório de Aids e Imunologia
Molecular do Instituto Oswaldo Cruz,
FIOCRUZ, sob orientação da Dr^a
Mariza Gonçalves Morgado

*“Vão tentar derrubar, que é pra me ver crescer
E às vezes me matar, que é pra eu renascer
Como uma supernova que atravessa o ar
Eu sou a maré viva. Se entrar, vai se afogar”
Fresno*

Agradecimentos

Primeiramente a Deus, para o qual direcionei minha fé no desejo de ultrapassar todos os obstáculos que surgiram na minha caminhada obtendo o conforto que precisava nas horas de angústia e a força necessária para trilhar meu caminho;

À minha família, especialmente meus pais e meus avós, por me darem a base necessária para que eu seja o que sou hoje em dia e por compartilharem comigo de cada momento da vida. Obrigado pelo apoio, carinho e incentivo dados em momentos que eu nem percebia que precisava. Vê-los felizes e satisfeitos com as minhas vitórias é o melhor combustível que posso ter para buscar cada vez mais sucesso;

À minha orientadora, Dra. Mariza Morgado, pelo apoio e confiança no meu trabalho. Apesar de todos os problemas ocorridos durante esses dois anos e a distância forçada pelas dificuldades que fogem de nosso controle, agradeço pela participação nos momentos em que precisei e por confiar na condução desse estudo. Espero que possa ter correspondido com qualidade as expectativas.

A Dra. Fernanda Côrtes, pela co-orientação extra oficial e toda a ajuda em todas as etapas do trabalho. Obrigado pelas discussões e por ter aguentado todas as perturbações pelos dados prévios dos pacientes e artigos para discussão. Sua supervisão e vontade de ajudar, inclusive com coisas que eram novas pra você, me ajudaram muito e serão muito importantes para as próximas etapas deste trabalho.

A Dra. Monick Guimarães, por ter vindo a meu socorro nos momentos em que foi preciso, desde aos problemas com NGS até a revisão surpresa em cima da hora. Obrigado pela contribuição no trabalho e por, mesmo em cima da hora, ter ajudado com ótimas opiniões para que a escrita do trabalho pudesse ter qualidade.

Ao Dr. Edson Delatorre, por ter sido o meu guru do NGS. Obrigado por ter sido parte essencial na análise dos dados do sequenciamento e por todos os tutoriais e discussões que ajudaram no “parto” do resultado final.

Da mesma forma, agradeço ao Dr Gonzalo Bello pelo direcionamento nas análises filogenéticas, ajuda em discussões e nos aprendizados colhidos desde antes do meu mestrado e que possibilitaram minha confiança na condução desse trabalho.

A Dra Caroline Passaes, que continua sendo peça essencial pro meu sucesso profissional mesmo distante. Obrigado por toda a orientação pré Mestrado e pelo reconhecimento no artigo publicado que foi essencial para a minha indicação para a bolsa Nota 10 da Faperj.

À Dra Sylvia Teixeira e a Msc. Nathalia Beatriz, as quais realizaram as tipagens dos genes HLA dos indivíduos do presente estudo, os quais foram resultados essenciais para a realização de meu trabalho.

A todos os amigos e companheiros do LABAIDS, pela amizade, atenção e acolhimento durante os últimos dois anos. Seja pela supervisão e auxílio nas atividades, pelas discussões e ensinamentos teóricos, ou pelas brincadeiras nos momentos de descontração. Obrigado por fazerem parte do meu dia a dia e tornar os dias na sala 500, nos almoços na copa, no P2 ou BM mais divertidos e descontraídos. Especialmente obrigado às loucas Nathalia e Thaysse por todas as risadas e cumplicidade que tornam a sala 500 um dos ambientes mais legais da Fiocruz.

A todas as amizades inocentes que nasceram nas disciplinas do mestrado e que espero que possam ser continuadas daqui pra frente. Larissa, Oscar, Maria, Rodrigo, Pedro, Letícia e todos os outros que meu advogado me impede de citar os nomes: Vocês foram grande parte do meu escape diário e tornaram todo esse período no mestrado uma experiência de ensinamentos, medos e amadurecimento. Aprendi com vocês o que devo e o que não devo fazer e espero que nossa amizade possa continuar viva e eternizada (Seja em fotos ou seja em memes...).

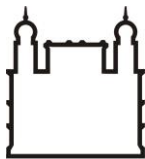
Da mesma forma, obrigado aos amigos, Vinicius e Larissa, por todo o apoio e por fazerem, dos momentos fora do trabalho, o escape da vida acadêmica que eu precisava;

Igualmente obrigado a amiga Tia Pipoca, por todos os momentos que tornaram o último ano inesquecível. Obrigado por me ensinar mais do que a vida acadêmica pode mostrar.

À Fundação Oswaldo Cruz e o Programa de Pós-Graduação em Biologia Parasitária, pela oportunidade acadêmica e experiências no mundo científico; Igualmente, Obrigado ao CNPQ pela concessão da bolsa regular durante o primeiro ano e a FAPERJ pela concessão da bolsa Nota 10 durante o último ano do Mestrado.

Aos pacientes que participaram deste estudo;

A todos esses e também aqueles que não foram citados, mas participaram desses 4 anos da graduação e alguma maneira, meus sinceros agradecimentos.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

CARACTERIZAÇÃO DE VARIANTES VIRAIS DE HIV-1 EM INDIVÍDUOS SOROPOSITIVOS COM PERFIL DE CONTROLE DA PROGRESSÃO PARA A AIDS E DA REPLICAÇÃO VIRAL: AVALIAÇÃO DA OCORRÊNCIA DE VARIANTES DE ESCAPE DA RESPOSTA IMUNE

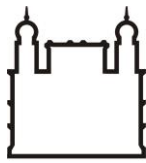
RESUMO

DISSERTAÇÃO DE MESTRADO

Diogo Gama Caetano

A infecção pelo HIV se caracteriza pela existência de um período de latência clínica (ou período assintomático) entre a infecção aguda e a fase de AIDS, cuja duração é extremamente variável, gerando distintos perfis de progressão entre os indivíduos infectados pelo HIV. Uma pequena fração de indivíduos infectados (5-10%), denominados não progressores de longo termo (LTNP), permanece clinicamente assintomática e com altos valores de TCD4⁺ (>500 cells/ μ l) durante mais de 8 anos de infecção na ausência de tratamento. Alguns desses indivíduos conseguem ainda bloquear quase totalmente a replicação e a evolução viral, gerando uma verdadeira latência evolutiva. O objetivo do presente estudo foi analisar o perfil mutacional do HIV-1 e a dinâmica de emergência e manutenção de variantes virais de escape da resposta imune em pacientes HIV-1 positivos classificados como LTNPs com diferentes níveis de controle da replicação viral. Para tal, utilizou-se uma casuística de 12 indivíduos LTNPs, classificados como controladores virêmicos (VC) e controladores de elite com (ECB) e sem *Blips* (EC) de acordo com os seus níveis de controle da infecção. Para cada indivíduo, foram extraídas amostras de DNA a partir de PBMCs obtidos durante a visita mais antiga (V1) e mais recente (VX) disponíveis. O DNA extraído foi utilizado para amplificação e sequenciamento, através de uma metodologia de NGS, dos genes *Gag* e *Nef*. As sequências obtidas permitiram o mapeamento do gene *nef* inteiro (cobertura média de 212.674 seqs/pb) e dos primeiros 1000pbs de *Gag* (cobertura média de 286.463 seqs/pb). Os consensos de cada mapeamento foram utilizados para as análises de divergência entre V1 e VX e variantes com frequências maior que 0,5% em cada pb mapeado foram consideradas verdadeiras. As análises de evolução evidenciaram o baixo nível de diversidade e divergência dos genes *Gag* e *Nef* entre os indivíduos controladores. Indivíduos ECs apresentaram menor perfil mutacional em ambos os genes analisados quando comparados com indivíduos ECB e VC. Enquanto *Gag* apresentou maior conservação que *Nef* em VCs, ambos os genes apresentaram semelhante variabilidade in ECs, indicando que o maior controle da viremia limita a evolução em nas duas regiões. Em alguns indivíduos ECBs, evidências de superinfecção e hipermutação mediada por APOBEC3G foi observada. As análises das mutações em regiões de epítomos restritos por CTL demonstraram a presença de mutações indicativas de escape da resposta imune surgidas entre V1 e VX na maioria dos indivíduos. Mutações não sinônimas foram encontradas em baixas frequências para todos os indivíduos, incluindo ECs, indicando que tais variantes surgem, mas a grande maioria não obtém sucesso evolutivo.

Palavras-chave: HIV, LTNP, Controladores, Evolução Viral, Escape CTL



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

CHARACTERIZATION OF HIV-1 VIRAL VARIANTS IN SOROPOSITIVE INDIVIDUALS WITH CONTROL OF PROGRESSION TO AIDS AND VIRAL REPLICATION: ASSESSMENT OF IMMUNE RESPONSE ESCAPE MUTANTS

ABSTRACT

DISSERTAÇÃO DE MESTRADO

Diogo Gama Caetano

The HIV infection is characterized by the existence of a clinical latency period (or asymptomatic period) between the acute infection and the AIDS phase, whose duration is extremely variable, generating distinct progression profiles between HIV infected individuals. A small fraction of infected individuals (5-10%), named long term non progressors (LTNP), remains clinically asymptomatic and with high CD4+ levels (>500cells/ul) during more than 8 years of infection in absence of treatment. Some of these individuals can even block almost all the viral replication and evolution, generating an evolutionary latency. The objective of this study was to analyze the mutational profile of HIV-1 and the emergence and maintenance dynamic of immune response escape mutants in HIV-1 positive patients classified as LTNPs and with different levels of viral replication control. For this end, a group of 12 LTNP patients, classified as viremic controllers (VC) and Elite Controllers with (ECB) and without Blips (EC) according to their level of viremia control, was used. For each individual, DNA samples were extracted from PBMCs obtained during the earliest (V1) and most recent visits (VX) available. The extracted DNA was used for amplification and sequencing, through a NGS methodology, of the *Gag* and *Nef* genes. The sequences obtained allowed to map the whole *Nef* gene (medium coverage of 286.463 reads/pb) and the first 1000bps of *Gag* (medium coverage of 212.674 reads/pb). The consensus sequences of each mapping were used for divergence analyses between V1 and VX and variants with frequencies higher than 0,5% in each mapped bp were considered real. The evolution analyses showed a lower degree of diversity and divergence of *Gag* and *Nef* genes between controllers. EC individuals presented lower mutational profile in both genes when compared with ECB and VC patients. While *Gag* presented higher conservation than *Nef* in VCs, both genes had showed similar variability in ECs, indicating that a higher viremia control limit the evolution in both regions. In some ECB individuals, evidence of superinfection and APOBEC3g mediated Hypermutation was observed. The analyses of mutation along regions of CTL-restricted epitopes had showed the presence of mutations arisen between V1 and VX, which were indicative of immune response escape, in most individuals. Non synonymous mutations had been found in lower frequencies in all individuals, including ECs, indicating that those escape variants arise, but most don't manage to obtain evolutionary success.

Keywords: HIV, LTNP, Controllers, Viral evolution, CTL escape

Lista de Figuras

Figura 1 Representação da Partícula Viral do HIV.....	3
Figura 2 Genoma do HIV-1 baseado na sequência de referência HXB2.....	3
Figura 3 Representação do Curso Clínico da Infecção pelo HIV, indicando as 3 fases e perfis de evolução clínica.....	9
Figura 4 Representação do Curso Clínico da Infecção pelo HIV em pacientes Controladores de Elite, sem e com Blips de carga viral.....	11
Figura 5 Posições dos fragmentos de gag e nef amplificados por PCR.....	20
Figura 6. Mapa de epítomos de Nef.....	23
Figura 7 Mapa de epítomos de Gag.....	24
Figura 8. Estatísticas de qualidade, tamanho e conteúdo de bases dos fragmentos referentes ao sequenciamento do Controle positivo HXB2.....	28
Figura 9. Estatísticas de distribuição de tamanho e qualidade das sequências das amostras ECB13V1 e VC14V1.....	29
Figura 10. Cobertura por par de base para o gene gag de cada visita de cada indivíduo estudado.....	32
Figura 11. Cobertura por par de base para o gene nef de cada visita de cada indivíduo estudado.....	33
Figura 12. Árvore filogenética Neighbor-Joining com as sequências consenso do gene gag referentes as visitas V1 e VX dos indivíduos estudados.....	35
Figura 13 Comparativo de Identidade entre as sequências V1 e Vx do gene gag para indivíduos Controladores de Elite.....	37
Figura 14 Comparativo de Identidade entre as sequências V1 e Vx do gene gag para indivíduos Controladores de Elite com Blips.....	37
Figura 15 Comparativo de Identidade entre as sequências V1 e Vx do gene gag para indivíduos Controladores Virêmicos.....	38

Figura 16 <i>Árvore filogenética Neighbor-Joining com as sequências das duas variantes isoladas para o indivíduo ECB09.</i>	39
Figura 17. <i>Árvore filogenética Neighbor-Joining com as sequências consenso do gene nef referentes as visitas V1 e VX dos indivíduos estudados.....</i>	40
Figura 18 <i>Comparativo de Identidade entre as sequências V1 e Vx do gene nef para indivíduos Controladores de Elite.</i>	42
Figura 19 <i>Comparativo de Identidade entre as sequências V1 e Vx do gene nef para indivíduos Controladores de Elite com Blips.....</i>	42
Figura 20 <i>Comparativo de Identidade entre as sequências V1 e Vx do gene nef para indivíduos Controladores Virêmicos.....</i>	43
Figura 21 <i>Comparativo de Identidade entre a sequência Vx do indivíduo EC17 e a sequência de referência HXB2 para o gene Nef.....</i>	43

Lista de tabelas

Tabela 1. Epítomos de Nef estudados e seus HLAs Associados	23
Tabela 2. Epítomos de Gag estudados e seus HLAs Associados	24
Tabela 3. Características Clínico Laboratoriais e Sócio-Demográficas dos Pacientes selecionados para o estudo.....	26
Tabela 4. Estatísticas das Sequências Pré e Pós-processamento	27
Tabela 5. Estatísticas de Mapeamento dos genes gag e nef	30
Tabela 6. Valores de Divergência entre visitas V1 e VX nos genes gag e nef ...	44
Tabela 7. Quantidade de SNPs e Características das Mutações associadas no gene gag dos indivíduos estudados	46
Tabela 8. Quantidade de SNPs e Características das Mutações associadas no gene NEF dos indivíduos estudados.....	48
Tabela 9. Mutações observadas em Gag e suas frequências em V1 e VX	84
Tabela 10. Mutações observadas em Nef e suas frequências em V1 e VX	94

Lista de abreviaturas, siglas e símbolos

A	Alanina
ACH2	Linhagem células contendo uma cópia de DNA Proviral
AIDS	Síndrome da Imunodeficiência Humana Adquirida
ANRS	Agência Nacional de Pesquisa em AIDS e Hepatite
APOBEC	Enzima editora de mRNA Apolipoproteína B
BAM	Arquivo binário contendo um mapa de alinhamento de sequências
C	Cisteína
CAAE	Certificado de Apresentação para Avaliação Ética
CCR5	Receptor de Alfa-quimiocina 5
CD4	Grupo de diferenciação 4
CDNA	Dna complementar
CEP	Comitê de Ética em Pesquisa
CONEP	Comissão Nacional de Ética em Pesquisa
CTL	Linfócito T Citotóxico
CTR	Complexo de Transcrição Reversa
CXCR4	Receptor de Beta-Quimiocina 4
D	Ácido Aspártico
DMSO	Dimetilsulfóxido
DNA	Ácido Desorribonucleico
E	Ácido Glutâmico
EC	Controlador de Elite
ECB	Controlador de Elite com Blip
EDTA	Ácido etilendiamino tetra-acético
ENV	Gene que codifica as proteínas do envelope viral
F	Fenilalanina
FIOCRUZ	Fundação Oswaldo Cruz
G	Glicina
<i>gag</i>	Gene do grupo antígeno do HIV
GP120	Glicoproteína de 120Kd
GP160	Glicoproteína de 160Kd

GP41	Glicoproteína de 41Kd
H	Histidina
HAART	Terapia Antirretroviral Altamente Ativa
HIV	Virus da Imunodeficiência humana
HLA	Antígeno Leucocitário Humano
HSH	Homem que faz sexo com homem
HXB2	Isolado Viral do subtipo B do HIV-1
I	Isoleucina
IN	Integrase
INI	Instituto Nacional de Infectologia
IPEC	Instituto de Pesquisa Clínica Evandro Chagas
IQR	Intervalo Interquartil
K	Lisina
L	Leucina
LEDGF	Fator de crescimento derivado do epitélio de lentes
LTNP	Não Progressores de Longo Termo
LTR	Terminações Longas Repetidas
M	Metionina
MHC	Complexo Principal de Histocompatibilidade
ml	Mililitro
Mm	milímetros
mRNA	RNA mensageiro
N	Asparagina
NaOH	Hidróxido de Sódio
<i>NEF</i>	Proteína Viral que atua na degradação de CD4
NGS	Sequenciamento de Nova Geração
Nm	Nanômetro
NPC	Complexo do Poro Nuclear
P	Prolina
P24	Proteína do capsídeo do HIV
P7	Proteína do Nucleocapsídeo
pb	Par de base

pb	Par de Base
PBMC	Células Mononucleares do sangue periférico
PCR	Reação em Cadeia da Polimerase
PIC	Complexo Pré-Integração
POL	Gene que codifica as enzimas virais
PR	Protease
Q	Glutamina
R	Arginina
R	Arginina
REV	Proteína reguladora da expressão viral
RJ	Rio de Janeiro
RNA	Acido ribonucleico
RRE	Elemento Resonsivo a VER
RT	Trasncriptase Reversa
S	Serina
SAM	Arquivo contendo um mapa de alinhamento de sequências
SGS	Sequenciamento de Genomas Únicos
T	Treonina
TAT	Transativador Transcricional viral
T-CD4+	Linfócitos T-CD4+
T-CD8+	Linfócito T-CD*+
TCR	Receptor de células T
tRNA	RNA transportador
UI	Microlitro
UNAIDS	Programa Conjunto das Nações Unidas sobre HIV/AIDS
V	Valina
V1	Primeira Visita
V3	Terceira região variavel da gp120 do envelope viral
VC	Controlador Virêmico
VIF	Fator viral de Infectividade
VPR	Proteína Viral R
VPU	Proteína Viral U

VX	Visita mais recente
W	Triptofano
WT	Selvagem
Y	Tirosina

Índice

1.	INTRODUÇÃO	1
1.1.	CARACTERÍSTICAS ESTRUTURAIS E GENOMA DO HIV	2
1.2.	O CICLO REPLICATIVO DO HIV	5
1.2.1.	FASE INICIAL: ENTRADA, RETROTRANSCRIÇÃO E INTEGRAÇÃO DO GENOMA VIRAL	5
1.2.2.	FASE TARDIA: TRANSCRIÇÃO, SÍNTESE PROTEICA E BROTAMENTO VIRAL	7
1.3.	O CURSO CLÍNICO E A HISTÓRIA NATURAL DA INFECÇÃO PELO HIV	8
1.3.1.	LTNPS E O CONTROLE DA REPLICAÇÃO VIRAL	10
1.4.	QUASIESPÉCIES E A COMPLEXIDADE DA POPULAÇÃO VIRAL	13
2.	OBJETIVOS	17
2.1.	OBJETIVO GERAL	17
2.2.	OBJETIVOS ESPECÍFICOS	17
3.	MATERIAIS E MÉTODOS	18
3.1.	CASUÍSTICA	18
3.2.	COLETA E PROCESSAMENTO DAS AMOSTRAS	19
3.3.	EXTRAÇÃO DE DNA	19
3.4.	ESTUDO DE DIVERSIDADE GENÉTICA VIRAL POR NGS	19
3.5.	ANÁLISES DE DADOS	21
4.	RESULTADOS	25
4.1.	CARACTERÍSTICAS CLÍNICO LABORATORIAIS E SÓCIO-DEMOGRÁFICAS DOS PACIENTES CONTROLADORES	25
4.2.	RENDIMENTO E ESTATÍSTICAS DO NGS	27
4.3.	MAPEAMENTO DOS GENES <i>GAG</i> E <i>NEF</i>	29
4.4.	ANÁLISES FILOGENÉTICAS	34
4.5.	IDENTIFICAÇÃO DE MUTAÇÕES NOS EPÍTOPOS RESTRITOS POR CTL	49
5.	DISCUSSÃO	55
5.1.	A EVOLUÇÃO VIRAL NOS INDIVÍDUOS CONTROLADORES	57
5.2.	EVOLUÇÃO EM EPÍTOPOS CTL E O ESCAPE VIRAL	62

6.	CONCLUSÃO	68
7.	REFERÊNCIAS BIBLIOGRÁFICAS	69
8.	ANEXOS	84
8.1.	TABELA DE VARIAÇÕES NO GENE GAG	84
8.2.	TABELA DE VARIAÇÕES NO GENE <i>NEF</i>	94

1. INTRODUÇÃO

Desde sua identificação no início dos anos 80 (BARRÉ-SINOUSSE et al., 1983; GALLO et al., 1984), o HIV-1 já infectou mais de 80 milhões de indivíduos em todo o mundo e foi associado à morte de cerca de 38 milhões desses, se tornando o patógeno de maior impacto do último século (UNAIDS, 2015). Segundo o último boletim epidemiológico, atualmente existem 36.9 milhões de indivíduos infectados e doenças relacionadas com a AIDS foram responsáveis pela morte de 1.2 milhões de indivíduos em 2014, sendo a África subsaariana o local onde a pandemia se mostra mais grave, acumulando a maior proporção dos casos (66% do total global de novas infecções em 2014) e a maior taxa de mortalidade (65% do total global de mortes em 2014).

Apesar de toda essa gravidade, os últimos dados epidemiológicos mostram que a pandemia vem sendo cada vez mais controlada, com uma redução de mais de 35% no número de novos casos por ano em comparação com o ano 2000 (3.1 milhões de para 2 milhões novas infecções). Uma semelhante queda também é observada para o número de mortes relacionadas a AIDS (2 milhões/ano em 2005 para 1.2 milhões/ano em 2014), configurando uma queda de 42% nos números totais. Paralelamente, o acesso ao tratamento com a terapia antirretroviral altamente ativa (HAART) tem aumentado progressivamente, sendo oferecido atualmente a cerca de 41% de todos os indivíduos infectados no mundo (UNAIDS, 2015).

Nos próximos anos, estima-se ainda o aumento da cobertura de oferecimento de medicamentos antirretrovirais devido à emergência de novas políticas de saúde pública, as quais incluem o oferecimento de terapia a grupos de risco como medida profilática, o que diminuiria a probabilidade de soroconversão segundo estudos (CHOOPANYA et al., 2013; COHEN et al., 2011; THIGPEN et al., 2012), e a indivíduos independente do estágio da doença e dos níveis de células T-CD4⁺, visto que o início precoce da terapia tem sido associado com melhores desfechos clínicos na fase crônica da doença (DING et al., 2015; OKULICZ et al.,

2015). Com base nessas políticas, estima-se a cobertura do tratamento com antirretrovirais em todo o mundo será total até 2030. (UNAIDS, 2015)

No Brasil, o número de infectados em 2014 foi estimado em mais de 700 mil pessoas, sendo a região sudeste um local de concentração de indivíduos infectados, possuindo 54,4% do total de casos identificados desde 1980. A região sudeste também abarca historicamente a maior proporção de mortalidade dentre todos os óbitos relacionados à AIDS no Brasil (61,8% do total entre 1980 e 2014). Dentre os estados do sudeste, o Rio de Janeiro é o que apresenta a maior taxa de detecção de HIV para a região (30,3 casos/100.000 habitantes), figurando como quarto lugar no cenário nacional. Apesar da histórica situação epidemiológica, a região sudeste é a única do Brasil a qual tem demonstrado tendência de queda na taxa de detecção de HIV nos últimos 10 anos. Paralelamente, observa-se uma diminuição de 26% na mortalidade nos últimos dez anos na região sudeste (MINISTÉRIO DA SAÚDE, 2014a).

Apesar de trinta anos já terem se passado desde a descrição da AIDS e da identificação de seu agente etiológico, até agora ainda não existem vacinas capazes de controlar a pandemia. Vários fatores têm limitado o avanço na área de vacinas sendo o polimorfismo viral e a falta de correlatos de imunidade associados à proteção contra a infecção ou evolução para a AIDS os obstáculos potencialmente mais relevantes.

1.1. CARACTERÍSTICAS ESTRUTURAIS E GENOMA DO HIV

O HIV é um retrovírus da família *Lentiviridae* que infecta linfócitos TCD4+, monócitos, macrófagos e células dendríticas. A sua partícula viral (**Figura 1**), ou vírion, apresenta formato esférico, medindo cerca de 100nm e apresenta em sua camada mais externa um envelope lipoproteico de mesma composição da membrana plasmática da célula hospedeira. (TURNER; SUMMERS, 1999). Distribuídas pela membrana existem as duas principais glicoproteínas de superfície do vírion, a gp120 e a gp41, além de várias proteínas celulares que são carregadas durante a formação do envelope e brotamento do vírion. (OTT, 2008). Envolto pelo

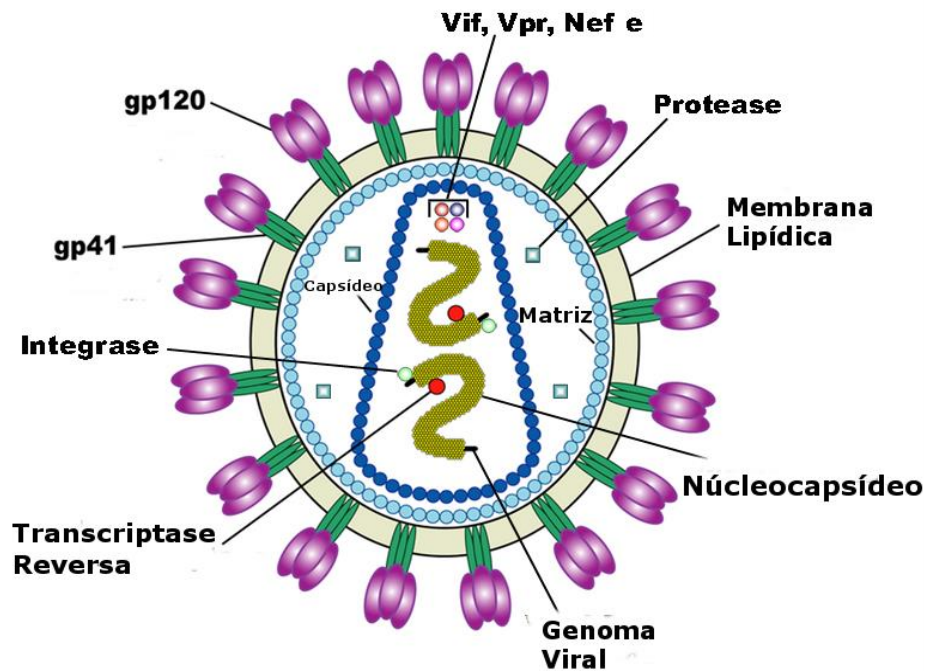


Figura 1 Representação da Partícula Viral do HIV. Fonte: Adaptado de National Institute of Allergy and Infectious Diseases (NIAID), 2005; Original disponível em: <http://web.archive.org/web/20050531012945/http://www.niaid.nih.gov/factsheets/howhiv.htm>

envelope está o capsídeo viral, o qual tem forma cônica e é formado pela proteína p24. No seu interior, o capsídeo abriga o genoma do HIV associado a proteínas do nucleocapsídeo (p7) e outras proteínas virais. (TURNER; SUMMERS, 1999). O genoma do HIV (**Figura 2**) consiste de duas fitas simples de RNA com polaridade positiva que medem cerca de 9.7kb cada e codificam 9 genes, dos quais 3 codificam as proteínas estruturais ou enzimas do vírus (*gag*, *pol* e *env*), 2 codificam proteínas regulatórias (*tat* e *rev*) e 4 codificam proteínas acessórias (*vif*, *Vpr*, *nef* e *vpu* ou *vpx*). Além desses genes, alguns elementos no genoma como as LTR (*Long*

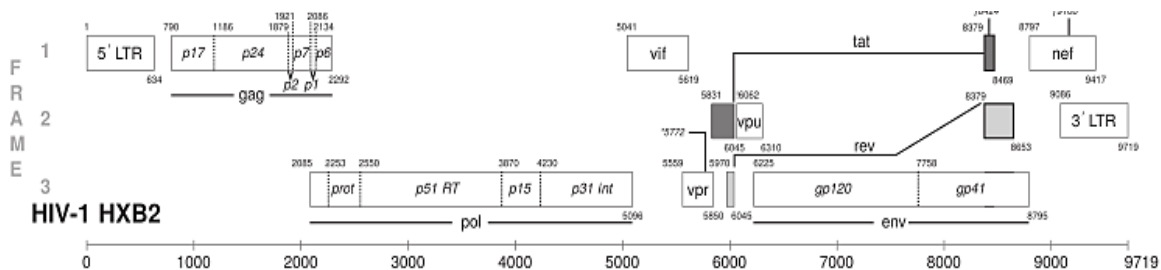


Figura 2 Genoma do HIV-1 baseado na sequência de referência HXB2. Fonte: LEITNER et al., 2013

Terminal Repeat) que flanqueiam o genoma, não são transcritos, mas agem como regiões regulatórias (LEITNER et al., 2013).

O gene *pol* dá origem as 3 enzimas virais [Transcriptase reversa (RT), Integrase (IN) e Protease (PR)]. A enzima RT é constituída de duas subunidades (p51 e p66) e é responsável por, após a entrada do vírus na célula, transcrever o genoma viral em uma molécula de cDNA fita dupla, a qual posteriormente servirá de substrato da enzima IN que vai catalisar a integração do genoma viral ao genoma do hospedeiro. A PR, por sua vez, é responsável pela clivagem proteolítica pós-traducional das proteínas precursoras originadas na transcrição dos genes *gag* e *env* (GROTTO; PARDINI, 2006). O gene *gag* dá origem a uma poliproteína precursora denominada p55, a qual é posteriormente clivada pela PR nas proteínas p6, capsídeo (CA/p24), matriz (MA/p17) e nucleocapsídeo (NC/p7), sendo essenciais para a composição estrutural da partícula viral e sua montagem (GÖTTLINGER, 2000). O gene *env* dá origem a uma proteína precursora gp160 que é clivada posteriormente nas proteínas do envelope viral gp120 (superfície) e gp41 (transmembranar). Essas duas proteínas, quando traduzidas, inserem-se no envelope viral e medeiam a entrada do vírus na célula, através de interação com receptores e correceptores de entrada.

Os genes regulatórios (*tat* e *rev*) e acessórios (*vif*, *Vpr*, *nef* e *Vpu* ou *Vpx*) codificam proteínas que atuam na regulação da expressão gênica viral e participam como intermediários em várias fases do ciclo replicativo viral. Tat (do inglês *Transcriptional Activator*) se liga a elementos específicos do LTR do HIV, interage com diversos fatores de transcrição celulares e controla os processos de transcrição do DNA viral através de remodelamento da cromatina e recrutamento de complexos transcripcionais competentes. Rev (do inglês *Regulator of Virion gene expression*) se liga ao RRE (*Rev Responsive Element*) do RNA viral e medeia exportação de mRNA viral não processado por *splicing*, essencial para a montagem de novas partículas virais. *Nef* (do inglês *Negative effector*) atua na diminuição da expressão de CD4 na superfície da célula hospedeira e modula vias de sinalização de linfócitos e macrófagos, sendo crucial na determinação da infectividade viral. *Vpr* medeia a importação do complexo de pré-integração viral e o cDNA retrotranscrito

para o núcleo, atua na regulação da apoptose na célula hospedeira e age como transativador do LTR viral. Vif (do inglês *viral infectivity factor*) aumenta a infectividade viral e atua como antagonista do fator de restrição APOBEC3G. Vpu deflagra a degradação de moléculas de CD4 no retículo endoplasmático e aumenta a liberação de novas partículas virais pela célula, além de também atuar na regulação da apoptose (SEELAMGARI et al., 2004).

1.2. O CICLO REPLICATIVO DO HIV

O ciclo replicativo do HIV decorre através de duas fases. Na fase inicial, ocorrem a entrada da partícula viral na célula, a degradação do capsídeo, a retrotranscrição do genoma viral e a integração do provírus ao genoma da célula hospedeira. Na fase tardia, ocorre transcrição do provírus, síntese das proteínas virais e montagem e brotamento de novas partículas virais (TURNER; SUMMERS, 1999).

1.2.1. FASE INICIAL: ENTRADA, RETROTRANSCRIÇÃO E INTEGRAÇÃO DO GENOMA VIRAL

A fase inicial começa com a entrada do vírion na célula hospedeira. Essa entrada é mediada pelas proteínas virais gp120 e gp41 (Liu et al., 2008). Essas proteínas são responsáveis por interagir com moléculas de CD4, as quais atuam como receptores principais de entrada (DALGLEISH et al., 1984), e os receptores de quimiocina CCR5 ou CXCR4, com função de correceptores de entrada (MOORE; TRKOLA; DRAGIC, 1997), da célula hospedeira. A utilização de CCR5 ou CXCR4 é determinada pela estrutura primária da região loop V3 da gp120, influenciando no tropismo do vírus por macrófagos ou células TCD4⁺ respectivamente. Durante o processo de entrada, a gp120 se liga a molécula de CD4 e sofre mudanças conformacionais que expõe sítios ligação da proteína aos correceptores. A ligação do correceptor, por sua vez, induz a exposição da gp41, a qual medeia a fusão do envelope viral com a membrana plasmática (Blumenthal et al., 2012). Após a fusão, o capsídeo é liberado no interior da célula e é degradado, liberando o genoma e as enzimas virais no citoplasma.

Logo após a degradação do capsídeo, parte das proteínas virais liberadas se associa e forma um complexo enzimático denominado complexo de Transcrição reversa (CTR), o qual contém a enzima RT, o RNA viral, a IN, p24, p17, p7 e Vpr (FASSATI; GOFF, 2001). Através do CTR, principalmente por ação da RT, o RNA viral é transcrito em uma dupla fita de DNA viral denominada DNA proviral. Esse processo ocorre no citoplasma e utiliza o tRNA *Lys3* como *primer* e as duas fitas de RNA viral como *template* para a retrotranscrição, sendo ambas degradadas, durante o processo, pela atividade de RNase H da RT (HU; HUGHES, 2012).

Após a Transcrição reversa, o RTC se converte no Complexo de Pré-Integração (PIC), composto pela IN, p17, Vpr e o DNA proviral originado na etapa anterior (Miller et al., 1997). Todo esse complexo então é translocado para o núcleo através dos complexos de poro nuclear (NPC) em um processo mediado por Vpr (SEELAMGARI et al., 2004). Em seguida a translocação do PIC para o núcleo, o DNA proviral deve ser integrado ao genoma da célula hospedeira para efetivamente infectar a célula. Nesse processo, catalisado pela IN, o DNA genômico da célula hospedeira é clivado e o DNA proviral é covalentemente ligado, através dos LTRs dos flancos, no local de inserção (CRAIGIE; BUSHMAN, 2012). Apesar de não ser possível prever um local específico de integração, estudos indicaram que a integração parece ser direcionada, ocorrendo preferencialmente em genes ativamente transcritos após a infecção pelo HIV (SCHRÖDER et al., 2002) ou transativados pela LEDGF/p75 (LLANO et al., 2006), um fator de transcrição celular que interage com o PIC em células infectadas pelo HIV.

Uma vez integrado, o Provírus pode ser transcrito para produção de novos vírions, dando início a fase tardia do ciclo, ou permanecer latente, constituindo um reservatório viral. Essa latência se relaciona, principalmente, ao estado de ativação da célula hospedeira, ocorre em grande escala nas células TCD4+ de memória (DAHABIEH; BATTIVELLI; VERDIN, 2015; PACE et al., 2011; RUELAS; GREENE, 2013). A ativação dessas células de memória é um dos principais fatores que impede a erradicação do vírus em indivíduos em tratamento, visto que tais reservatórios atuam como fontes de novas partículas virais e podem permanecer

latentes por anos (IMAMICHI et al., 2014; PACE et al., 2011; SILICIANO et al., 2003).

1.2.2. FASE TARDIA: TRANSCRIÇÃO, SÍNTESE PROTEICA E BROTAMENTO VIRAL

Uma vez integrado, o provírus serve de molde para a transcrição de proteínas virais. Inicialmente, a maioria dos transcritos gerados é processado por *Splicing-múltiplo*, translocados para o citoplasma e traduzidos principalmente em *Nef*, *Tat* e *Rev* (Sierra et al., 2005; Turner & Summers, 1999). Pelas suas ações em aumentar o nível de ativação celular e transativar o LTR, respectivamente, a expressão de *Tat* e *Nef* acaba por gerar um feedback positivo que estimula a transcrição do provírus e a síntese de transcritos mais longos não processados ou parcialmente processados que codificam as proteínas estruturais *Env*, *Gag* e *Pol*. *Rev* será responsável por auxiliar a translocação para o citoplasma desses transcritos maiores, através da ligação com o RRE (Rev Responsive Element) dos mRNAs em questão, fornecendo substrato para o a montagem de novos vírions (SEELAMGARI et al., 2004).

A formação de novas partículas virais ocorre a partir do lado interno da membrana plasmática. Inicialmente, ocorre a transcrição e síntese da poliproteína gp160 (ENV), a qual é sintetizada no reticulo endoplasmático rugoso e liberada no interior do mesmo. A Gp160 segue então a via metabólica de secreção e é transportada para o complexo de Golgi, onde a região da poliproteína correspondente a gp120 é massivamente glicosilada, e depois clivada pela protease celular *furina*, resultando na gp120 e na gp41. Por fim, as duas proteínas do envelope são enviadas, via transporte vesicular exocítico, e inseridas na membrana plasmática (SUNDQUIST; KRÄUSSLICH, 2012). Para que não ocorra interações entre gp120 recém-inserida e moléculas de CD4 da própria célula hospedeira, a diminuição da expressão de CD4 pela célula hospedeira, a qual é mediada por *Nef* e *Vpu*, se torna essencial para o processo de brotamento (SEELAMGARI et al., 2004; SIERRA; KUPFER; KAISER, 2005).

A partir dos locais de concentração das proteínas de envelope, ocorre o processo de brotamento dos novos vírions. Para tal, conjuntos de poliproteínas precursoras de *Gag* se agrupam nos locais de brotamento, interagindo com os domínios citoplasmáticos da gp41 (TURNER; SUMMERS, 1999). O acúmulo de *Gag* inicia o brotamento e a formação da partícula viral imatura, ao mesmo tempo que sinaliza para o recrutamento dos outros componentes do vírus, como as duas fitas do genoma viral e a poliproteína precursora *Gag-Pol* que dá origem as enzimas virais. A presença desses elementos componentes da partícula viral estimula continuamente a evaginação e o brotamento da partícula viral imatura, ao mesmo tempo em que ativa a ação da PR viral responsável por clivar as poliproteínas precursoras *Gag* e *Gag-Pol*, culminando na formação do capsídeo viral e maturação do vírion. Com isso, uma nova partícula viral é liberada, estando livre para infectar outras células hospedeiras.

1.3.O CURSO CLÍNICO E A HISTÓRIA NATURAL DA INFECÇÃO PELO HIV

A história natural da infecção pelo HIV se caracteriza por 3 fases: fase aguda, fase de latência clínica e fase de AIDS (Figura 3). A fase aguda é caracterizada por um pico de viremia, paralelo a um decréscimo de células TCD4⁺, que ocorre cerca de 3-4 semanas após a aquisição do vírus. Seguido a esse pico, ocorre o progressivo aumento de células TCD4⁺ e diminuição da carga viral, causados pela montagem da resposta imune inicial contra o vírus. Cerca de 6 a 12 meses após a infecção, ocorre uma recuperação dos níveis de células TCD4⁺ e declínio da carga viral até um ponto basal denominado *setpoint* viral (KASSUTTO; ROSENBERG, 2004). A partir do estabelecimento do *setpoint* viral, se inicia um período de latência clínica (ou período assintomático), caracterizado na maioria dos indivíduos por um progressivo declínio de células TCD4⁺ e aumento da viremia durante anos, culminando na fase de AIDS, na qual a depleção massiva de células TCD4⁺ deixa o organismo vulnerável a infecções oportunistas e pode levar o indivíduo a morte na ausência de tratamento. A duração do período de latência clínica é

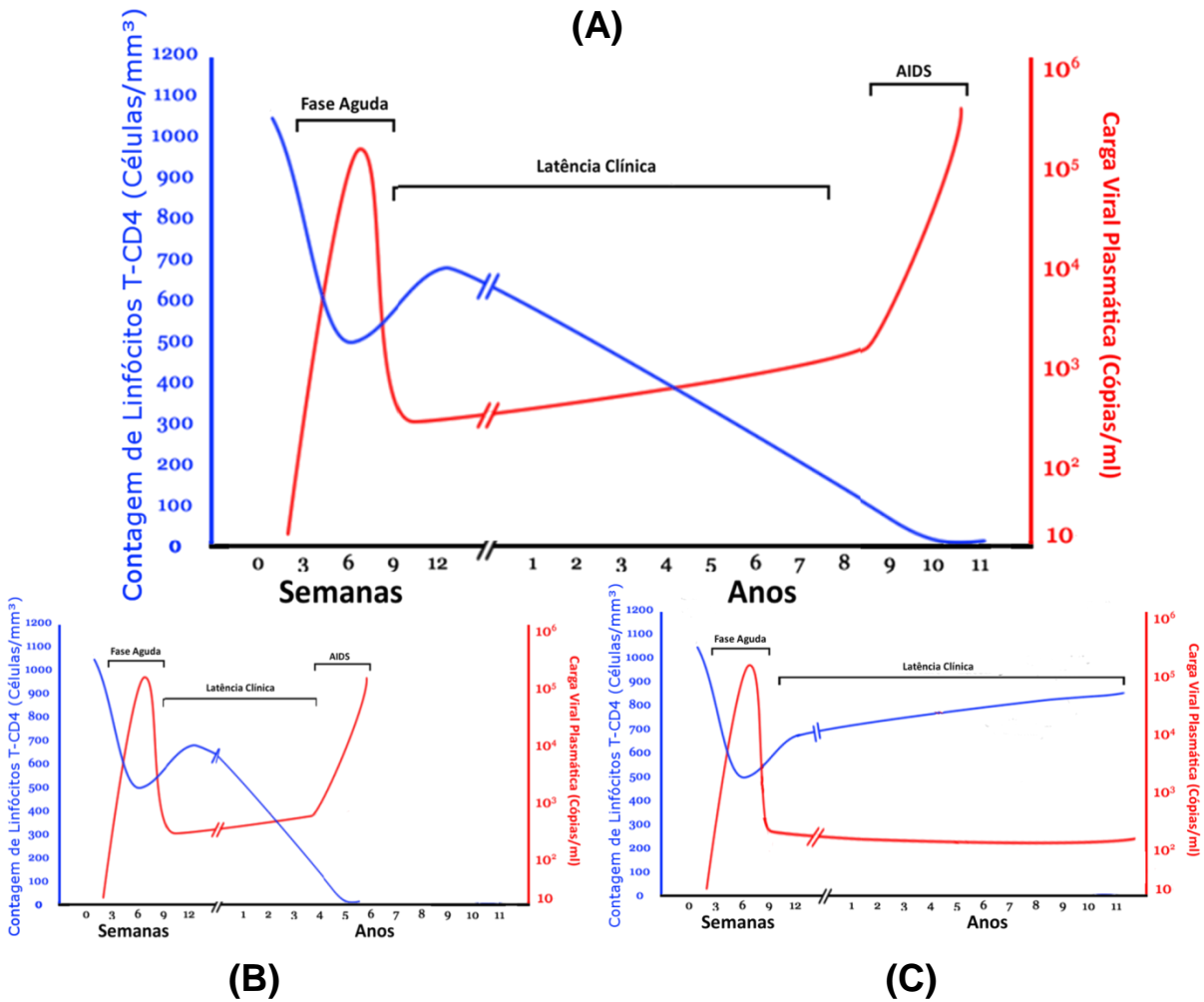


Figura 3 (A) Representação do Curso Clínico da Infecção pelo HIV, indicando as 3 fases. Esta representação é comum entre pacientes que apresentam a chamada progressão Típica; (B) Curso clínico em Progressores Rápidos; (C) Curso Clínico em LTNP. Este gráfico também é representativo da história natural de Controladores Virêmicos (VC). Fonte: Adaptado de Wikipedia. Figura baseada em *Pantaleo, G et al. (February 1993). "New concepts in the immunopathogenesis of human immunodeficiency virus infection". New England Journal of Medicine 328 (5): 327-335.*

extremamente variável e distintos perfis de progressão podem ser observados entre os indivíduos infectados pelo HIV (PANTALEO; FAUCI, 1996) (**Figura 3**). A maioria deles (70% a 80%), denominados progressores típicos, progride para a AIDS em cerca de 8 a 10 anos (**Figura 3-A**). Uma porcentagem significativa dos indivíduos (10% a 15%) evolui para a AIDS em um espaço de tempo mais curto, que varia de dois a cinco anos, e são chamados de progressores rápidos (**Figura 3-B**). Finalmente, uma pequena fração de indivíduos infectados (5-10%) permanece

cl clinicamente assintomática e com altos valores de TCD4⁺ (>500 cells/ μ l) durante mais de 10 anos de infecção na ausência de tratamento (**Figura 3-C**). Estes indivíduos são denominados progressores lentos ou não progressores de longo prazo (LTNPs, do inglês long-term nonprogressors). Com a introdução precoce da terapia antirretroviral, logo após a soroconversão, a tendência é a manutenção dos níveis iniciais das células TCD4⁺ e ausência de carga viral, retardando a evolução para a aids.

1.3.1. LTNPS E O CONTROLE DA REPLICAÇÃO VIRAL

Apesar do controle da progressão, estudos com diferentes coortes de indivíduos LTNPs tem demonstrado cada vez mais que esses indivíduos constituem uma população heterogênea, havendo variação de um indivíduo para o outro no nível de supressão da carga viral plasmática, no perfil genômico e transcriptômico e na resposta imunológica TCD4⁺ e TCD8⁺. (SÁEZ-CIRIÓ; PANCINO, 2013). Com base na capacidade de supressão da carga viral plasmática na ausência de terapia antirretroviral, tais indivíduos tem sido ainda subclassificados em populações distintas como LTNPs controladores e não-controladores (GAARDBO; HARTLING, 2012). Sabendo que a carga viral plasmática e seus níveis são um fator importante para a progressão da doença, tais indivíduos controladores ainda tem sido separados entre si de acordo com a eficiência de supressão viral, ao passo que estudos demonstram a heterogeneidade dentre tais indivíduos (PEREYRA et al., 2008; SAEZ-CIRION et al., 2009)

A maioria dos LTNPs controladores apresenta carga viral baixa (100 - 5000 cópias/ml), porém detectável, ao longo dos anos e são denominados Controladores Virêmicos (VC) (**Figura 3-C**). Uma pequena parte dos indivíduos LTNPs (<1%) apresenta carga viral plasmática indetectável pelos padrões de quantificação atuais (<50 cópias/ml) e são denominados de Controladores de Elite (EC) (**Figura 4-A**). Alguns desses EC apresentam carga viral indetectável em todas as medidas, enquanto outros podem apresentar episódios intermitentes e casuais de viremia

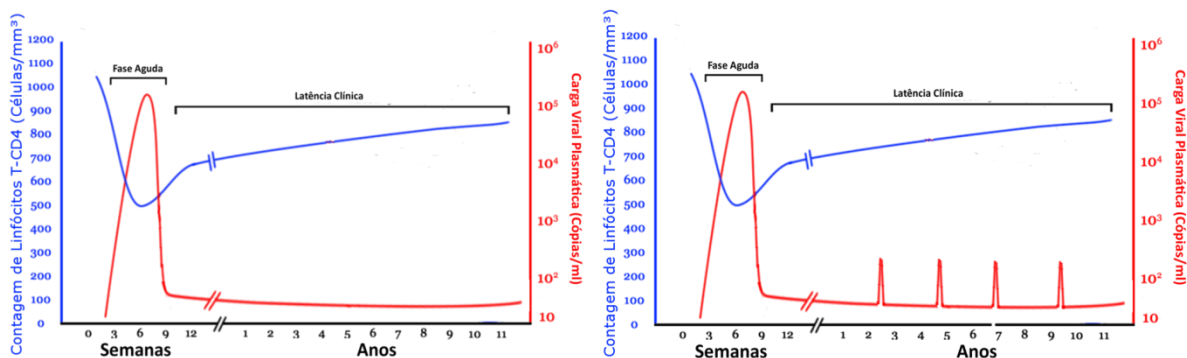


Figura 4 Representação do Curso Clínico da Infecção pelo HIV em pacientes Controladores de Elite, sem (A) e com Blips de carga viral (B). Fonte: Adaptado de Wikipedia. Figura baseada em *Pantaleo, G et al. (February 1993). "New concepts in the immunopathogenesis of human immunodeficiency virus infection". New England Journal of Medicine 328 (5): 327-335.*

chamados de “*Blips*” (ECB) (Figura 5-B). Vários fatores tanto relativos ao perfil genético e imunológico do hospedeiro humano, como características inerentes ao vírus têm sido associados aos diferentes perfis de progressão e controle da replicação viral.

Com base em estudos de perfil genético de indivíduos HIV-positivos, vários fatores já foram associados ao perfil de não progressão. Dentre eles, se destacam: a apresentação de alelos CCR5 Δ -32, o qual gera uma proteína truncada que impede a infecção por variantes virais M-trópicas em indivíduos homocigotos (DEAN et al., 1996; LIU et al., 1996); presença de alelos HLA protetivos tais como HLA B*57 e B*27 (ALTFELD et al., 2003; HENDEL et al., 1999; KASLOW et al., 1996) super-representados entre a população de indivíduos LTNPs; presença de alelos KIR3DS1 e 3DL1 (Tomescu et al., 2011), recentemente associados ao perfil de controle da progressão; e presença de polimorfismos em genes de proteínas celulares que atuam como fatores de restrição do vírus, como APOBEC, Tetherina, TRIM-5-alpha, SAMHD-1 e LEDGF/p75 (MESSIAEN et al., 2012; PASSAES et al., 2014; SANTA-MARTA et al., 2013)

O perfil da não progressão também pode ser decorrente de características inerentes do vírus. De fato, uma das primeiras coortes de LTNPs estudada e completamente caracterizada tinha o status de não progressão associada a infecção por uma variante viral atenuada. Nessa coorte, identificada na Austrália,

cinco indivíduos foram contaminados através de transfusões de sangue originadas de um único paciente soropositivo, que por sua vez era um LTNP (LEARMONT et al., 1992). O sequenciamento nucleotídico das variantes virais recuperadas destes indivíduos mostrou uma deleção comprometendo parte do gene *nef* e da região U3 do LTR (Deacon et al., 1995; Huang et al., 1995). Desde então, casos isolados de atenuação causada por mutações em *Nef* (KIRCHHOFF et al., 1999; RODÉS et al., 2004) ou outras variações genômicas virais também foram descritos, incluindo alterações atribuídas a genes acessórios (Michael et al., 1995; Mologni et al., 2006; Rajan, et al., 2006; Rodés et al., 2004) e estruturais (Huang et al., 1998).

Por fim, o perfil imunológico do hospedeiro também se mostra um fator essencial para a montagem de uma resposta eficiente contra o HIV. Indivíduos imunologicamente competentes tem sido caracterizados por apresentarem uma resposta TCD8⁺ mais eficiente, apesar de menores níveis de ativação celular e uma maior preservação de células T de memória (SÁEZ-CIRIÓN; PANCINO, 2013). Em adição, a produção de células TCD4 e TCD8 também se mostra maior em indivíduos controladores, enquanto a senescência da ativação imune e os níveis de apoptose se mostram menores, em comparação com indivíduos progressores (GAARDBO et al., 2013). Apesar desse perfil geral, os níveis de resposta são variáveis mesmo entre indivíduos com diferentes níveis de controle, de modo que a viremia persistente ou intermitente de ECBs e VCs parece influenciar níveis basais de ativação imune que se traduzem em maiores níveis de resposta TCD8⁺ nesses indivíduos (CÔRTEZ et al., 2015).

Além desses fatores, o conjunto de estudos imunológicos acerca dos diferentes perfis de controle parece demonstrar que a chave para o controle da infecção nesses indivíduos está numa resposta CTL vírus-específica eficiente (GAARDBO; HARTLING, 2012). Estas respostas tem sido amplamente associadas com menores cargas virais e perfis de controle da replicação, principalmente as direcionadas contra epítomos imunodominantes de *Gag* (KIEPIELA et al., 2007; MOTHE et al., 2012; MURAKOSHI et al., 2015; ZUÑIGA et al., 2006), o que tem motivado sua utilização como estratégia vacinal (KUNWAR et al., 2013).

1.4. QUASIESPÉCIES E A COMPLEXIDADE DA POPULAÇÃO VIRAL

O conceito de quasiespécies foi criado por Eigen e Schuster no fim da década de 70 e designava cada unidade de um conjunto de partículas auto-replicas e genotipicamente relacionadas, as quais formam uma população centrada no fenótipo de maior valor adaptativo geral (EIGEN; SCHUSTER, 1977). Inicialmente utilizado para designar entidades simples que poderiam ser importantes durante o evento de origem da vida na Terra, o conceito de quasiespécies passou a ser utilizado, com o passar dos anos, também nos estudos de genética de populações bacterianas e virais. A justificativa para o uso do modelo de quasiespécies nas pesquisas com modelos biológicos parte do princípio de que, em populações nas quais a taxa de mutação é alta entre os indivíduos, a seleção natural atua mais sobre um grupo de variantes intimamente relacionadas, do que unicamente sobre cada variante (WILKE, 2005). Dentro dessa premissa, a utilização da teoria das quasiespécies para descrever a dinâmica de populações de bactérias e vírus, principalmente de RNA, se torna viável.

O HIV é um dos modelos em que a teoria das quasiespécies se mostra mais importante ao descrever a dinâmica de evolução genética da população viral (VARTANIAN et al., 1992). Tal viabilidade do modelo se deve sobretudo ao alta potencial de variabilidade genética relativo ao vírus (NOWAK, 1992). De maneira geral, essa variabilidade genética do HIV pode ser relacionada diversos fatores, dentre os quais o principal é alta taxa de mutação causada por erros na retrotranscrição e na transcrição do provírus. A enzima transcriptase reversa não possui atividade de proofreading, bem como a RNA Polimerase II responsável pela transcrição do provírus, sendo estimada uma taxa de erro estimada em cerca de $1-5 \times 10^{-5}$ mutações/ciclo de replicação para o HIV (Abram et al., 2010).

Paralelamente, o processo de retrotranscrição não é contínuo, de modo que as duas fitas de RNA contidas em cada viri3n servem como molde em diferentes momentos do processo, gerando um cDNA híbrido dos dois moldes. Em células duplamente infectadas, tal processo pode gerar recombinantes entre diferentes variantes virais (HU; HUGHES, 2012). Associado ao fato de que células infectadas vivem 2 dias em média e podem liberar cerca de 10^{10} novos vírions por dia

(Perelson et al., 1996), essa propensão a mutação faz com que a emergência constante de novas variantes virais seja inevitável.

Embora a ocorrência de mutações naturais no genoma viral se dê ao acaso, a pressão seletiva de tais mutações ocorre de forma diferenciada de acordo com os genes estudados, de modo que o perfil de mutação dos diferentes genes do HIV não é tão aleatório como deveria. Análises de variação nos diferentes genes do HIV mostraram que o gene Env é o que apresenta maior variabilidade genética e dentro de cada gene ainda existem *hotspots* mutacionais (STARCICH et al., 1986). Depois de Env, as maiores taxas de evolução são observadas para *Nef* e os genes acessórios, seguida de *Gag* (YEBRA; BROWN, 2015). De todos, o gene Pol é o que apresenta menos variabilidade, sobretudo nas regiões de domínios enzimáticos e de ligação com outras proteínas, evidenciando uma conservação relacionada a manutenção da funcionalidade.

A susceptibilidade a alguns fatores de restrição também parece ter papel na geração de variabilidade genética no HIV. As proteínas da família APOBEC, que constituem citidina deaminases, são parte da maquinaria de proteção contra vírus das células humanas. A mais conhecida, APOBEC3G, induz mutações G->A no HIV, gerando partículas virais hipermutadas e ocasionalmente com *Stop códons* gerados devido a ação em códons TGG. Em ciclos replicativos normais, vif é responsável por antagonizar o efeito de APOBEC, inibindo a deaminação e impedindo a mutação viral (HARRIS; LIDDAMENT, 2004).

No conjunto, todo esse potencial de diversidade genética acaba atuando como combustível para uma intensa dinâmica de população viral, a qual é moldada não só por deriva genética e pressões estocásticas, mas principalmente pela ação do sistema imune. Essa pressão de seleção positiva se origina principalmente da ação de anticorpos neutralizantes, linfócitos TCD4⁺ e linfócitos TCD8⁺ (Rambaut et al., 2004), podendo ser experimentalmente observada na forma de um incremento no aparecimento de substituições nucleotídicas não sinônimas em regiões genômicas de "*hotspot*" de diversidade, como as regiões hipervariáveis de ENV (STARCICH et al., 1986; WILLIAMSON, 2003) e regiões de epítomos restritos pela resposta CTL (BRUMME; POON, 2011).

Uma forma de avaliar a intensidade e eficiência da pressão seletiva positiva é através das análises de padrões de evolução filogenética intra-hospedeiro, as quais levaram, em 1999, a proposição de um modelo de evolução das quasispécies do HIV-1 ao longo do tempo para progressores típicos (SHANKARAPPA et al., 1999). Segundo o estudo referenciado, a divergência (distância ao vírus que originou a infecção) e a diversidade (distância entre seqüências contemporâneas) viral se incrementam linearmente durante a maior parte do período assintomático, levando à ocorrência de quasispécies cada vez mais complexas ao longo do tempo. Esta dinâmica, associada a pressão seletiva causada pelo sistema imune explica a emergência de variantes virais mais citopáticas (TERSMETTE et al., 1989; TROYER et al., 2005) e/ou variantes de escape da resposta imune do hospedeiro (KOENIG et al., 1995; MAHALANABIS et al., 2009).

Estas últimas são definidas como variantes virais que apresentam mutações em regiões de epítomos restritos por CTL, o que pode vir a impedir o reconhecimento e ação citotóxica em células infectadas com tais vírus (BRUMME; POON, 2011). Tais mutações de escape tem sido amplamente estudadas e sua emergência tem sido identificada já durante a fase aguda da infecção (GOONETILLEKE et al., 2009). A participação das moléculas de MHC Classe I no processo de resposta citotóxica, por sua vez, ocasiona a restrição de tais mutações a sítios virais específicos, permitindo a identificação dessas variantes baseado no genótipo HLA de indivíduos infectados. Tais associações são tão importantes que estudos demonstram inclusive que tais pressões seletivas atuam como uma força de evolução a nível populacional, moldando as regiões de epítomos restritos por alelos HLA específicos (KAWASHIMA et al., 2009; LESLIE et al., 2004; MARTIN et al., 2014). O mesmo é verdadeiro para epítomos restritos por alelos HLA relacionados a não progressão clínica no HIV, os quais selecionam e demonstram padrões de emergência e reversão de mutações em tais sítios (KUNWAR et al., 2013; LESLIE et al., 2004; MIURA et al., 2009).

Por fim, a emergência de tais mutações de escape tem sido identificadas inclusive em indivíduos com fenótipo de controle da infecção, principalmente em

regiões de *Gag* e *Nef* (BAILEY et al., 2008; O'CONNELL et al., 2010; SALGADO et al., 2010), apesar da ocorrência de uma certa latência evolutiva, caracterizado por uma baixa diversidade entre sequências temporalmente distantes (BELLO et al., 2007). Apesar de tais populações de escape terem sido relacionadas a uma perda de fitness viral e conseqüentemente a sua atenuação replicativa, a ausência de reversão das mutações de escape em estudos in vitro mostram que a resposta CTL é uma força contínua de seleção viral (O'CONNELL et al., 2010), favorecendo o surgimento constante desses mutantes de escape, principalmente nos pacientes com alelos HLA relacionados a não progressão clínica. Em conjunto, os dados obtidos e a importância de tais regiões de maior pressão seletiva do vírus demonstra que a identificação de epítomos de escape em regiões imunodominantes, como as de *Gag* e *Nef*, em sequências virais de pacientes que naturalmente controlam a infecção pelo HIV pode contribuir para a seleção de alvos para o desenvolvimento de vacinas e terapias anti-HIV mais eficazes.

2. OBJETIVOS

2.1. OBJETIVO GERAL

Estudar o perfil mutacional do HIV-1 e a dinâmica de emergência e manutenção de mutantes virais de escape da resposta imune em pacientes HIV-1 positivos classificados como LTNPs com diferentes níveis de controle da replicação viral.

2.2. OBJETIVOS ESPECÍFICOS

- Analisar as sequências, obtidas por sequenciamento de nova geração (NGS), referentes às regiões *Gag* e *Nef* das diversas quasiespécies componentes da população proviral de pacientes LNTP, descrevendo-a em termos de diversidade e frequência de variantes.
- Caracterizar a variação de aminoácidos em regiões internas ou flangeadoras de epítomos relacionados à resposta T-CD8⁺ específica com base nos HLAs dos pacientes em estudo.
- Identificar a ocorrência de alterações genéticas (mutações, deleções ou inserções) que poderiam caracterizar defectividade das sequências virais de pacientes controladores de HIV.
- Verificar a frequência de hipermutação nas sequências provirais dos pacientes, bem como as taxas de diversidade sinônimas e não sinônimas nas regiões verificadas.

3. MATERIAIS E MÉTODOS

3.1. CASUÍSTICA

Foram incluídos no estudo 12 indivíduos que estão em acompanhamento regular no Instituto Nacional de Infectologia Evandro Chagas (**INI**) da FIOCRUZ, RJ. Todos os indivíduos são classificados como LTNPs e subdivididos em Controladores de Elite (EC, $n = 3$), Controladores de Elite com “Blips” (ECB, $n = 4$) e Controladores Virêmicos (VC, $n = 5$), de acordo com os critérios de inclusão apresentados a seguir.

Pacientes LTNPs foram definidos como aqueles com diagnóstico de infecção pelo HIV documentada há ≥ 8 anos que, na ausência de terapia antirretroviral, permaneceram sem evidência de progressão clínica e com contagens de células T CD4 superiores a $500/\text{mm}^3$ em mais de 90% das mensurações. Controladores de Elite (EC) são definidos como indivíduos com carga viral plasmática <80 cópias/ml em 100% das mensurações. Controladores de Elite com “Blips” (ECB) são definidos como indivíduos com carga viral plasmática <80 cópias/ml em até 80% das mensurações, apresentando episódios intermitentes de viremia. Controladores Virêmicos (VC) são definidos como LTNPs com carga viral baixa, porém detectável (81-5000 cópias/ml) em até 80% das mensurações.

Todos os indivíduos forneceram consentimento livre e esclarecido e tem se submetido a punções venosas regulares. O presente estudo é um subprojeto do estudo intitulado “Caracterização de pacientes HIV-positivos não progressores de longo termo (LTNP) do Rio de Janeiro – Contribuição de fatores virológicos, genéticos e imunológicos”, o qual foi conduzido no Instituto Nacional de Infectologia Evandro Chagas, denominado anteriormente Instituto de Pesquisa Clínica Evandro Chagas (IPEC), quando da apresentação e aprovação do CEP/CONEP. O mesmo foi aprovado no CONEP em 29/10/2008 e referenciado como CAAE 1717.0.000.009-07.

3.2. COLETA E PROCESSAMENTO DAS AMOSTRAS

A cada visita foram coletados 60 ml de sangue por punção venosa divididos em tubos contendo como anticoagulantes EDTA ou heparina. O sangue foi centrifugado 30 min a 400xg em gradiente de Ficoll-Hypaque para separação das células mononucleares e plasma. As células periféricas mononucleares (PBMCs) foram criopreservadas em nitrogênio líquido na concentração de $5-10 \times 10^6/\text{ml}$ em solução contendo 90% de soro bovino fetal e 10% de DMSO. Para cada paciente foram utilizados os PBMCs correspondentes as visitas mais antiga (V1) e mais recente (VX) disponíveis.

3.3. EXTRAÇÃO DE DNA

Para cada ponto analisado, O DNA Total foi purificado a partir dos PBMCs por extração com DNAzol (Invitrogen). As células descongeladas foram ressuspensas em 1ml de DNAZOL e incubadas a 4°C por 72 horas. Após incubação e lise, o DNA foi precipitado pela adição de 500ul de Etanol 100% e centrifugado a 12000 rpm por 8 minutos. O sobrenadante foi descartado e o pellet formado foi lavado utilizando 1.4ml de Etanol 75%, sendo centrifugado, em seguida, a 12000rpm por 5 minutos. O sobrenadante foi novamente descartado e o pellet seco foi ressuspensão em 50ul de NaOH 8mm e incubado a temperatura ambiente por 30 minutos para solubilização. Posteriormente, foram adicionados mais 50ul de Buffer EB (Qiagen) e a amostra incubada por mais 30 minutos a 56°C. O DNA obtido foi armazenado a -20°C.

3.4. ESTUDO DE DIVERSIDADE GENÉTICA VIRAL POR NGS

A partir do DNA total extraído dos PBMCs, os fragmentos referentes as regiões 408-1844 (compreendendo *Gag*) e 8697-9639 (compreendendo *Nef*) do genoma do HIV (em relação ao HXB2) foram amplificados por Nested PCR (**Figura 5**).



Figura 5 Posições dos fragmentos de *gag* e *nef* amplificados no presente estudo. Trecho em azul corresponde ao amplicon resultante do 2o Round do Nested PCR, enquanto trechos em vermelho são resultantes do 1o Round. A figura foi obtida a partir da ferramenta HIV Sequence Locator (<http://www.hiv.lanl.gov/content/sequence/LOCATE/locate.html>)

Para o gene *gag*, foram utilizados os primers SCAOSD (GGGACTTTCCGCTGGGGACTTTC) (SIERRA et al., 2005) e G17 (TCCACATTTCCAACAGCCYTTTTT) (HEYNDRICKX et al., 2000) para o primeiro round da PCR e os primers SCANSO (CGAGCCCTCAGATGCTGCATATAAGC) (SIERRA et al., 2005) e P24-1 (CCCTGRCATGCTGTCATCA) (LEE et al., 1998) para o segundo round.

Para o gene *nef*, o conjunto de primers foi NEF-1 (CCTGTGCCTCTTCAGCTACCACCG) (ARTENSTEIN et al., 1996) e SCDOAD (AGTCACACAACAGACGGGCACACAC) (SIERRA et al., 2005) para o primeiro round e NEF-3 (TGRACAGATAGRRTYATAGAA) (SALVI et al., 1998) e SCDNAD (ACGTGCCCTCAAGGCAAGCTTTATTGAGGCT) (SIERRA et al., 2005) para o segundo round. A amplificação dos fragmentos foi realizado utilizando a enzima Platinum® Taq DNA Polimerase High Fidelity (Invitrogen), de acordo com as instruções do fabricante. Como controle positivo foi utilizado DNA total proveniente da linhagem celular ACH2 (CLOUSE et al., 1989; FOLKS et al., 1989), a qual contém uma cópia do HXB2 integrada por célula.

Os produtos de PCR obtidos foram purificados utilizando o kit Illustra GFX PCR DNA and Gel Band Purification Kit (GE Life Sciences), de acordo com as instruções do fabricante, e quantificados utilizando um fluorômetro Qubit® 2.0 (Invitrogen) e os reagentes do kit Qubit dsDNA BR Assay (Invitrogen).

Com base nas quantificações, os produtos purificados de *Gag* e *Nef* para cada visita de cada paciente foram unidos em pools equimolares, diluídos para a

concentração de 0,2ng/ul e submetidos ao sequenciamento na Plataforma de Sequenciamento de Nova Geração Fiocruz/RJ.

Os amplicons multiplexados foram utilizados para construção da biblioteca genômica utilizando os kits Nextera XT DNA Library (Illumina) e Nextera XT Index Kit (Illumina), de acordo com as instruções do fabricante. As bibliotecas geradas foram normalizadas e submetidas a clusterização utilizando o kit HiSeq SR Rapid Cluster V2. O sequenciamento foi realizado utilizando o kit HiSeq Rapid SBS v2 de 200 ciclos (Illumina) em um sequenciador Illumina HISEQ2500.

3.5. ANÁLISES DE DADOS

Os dados brutos do sequenciamento foram pré-processados e compilados em arquivos FASTQ em colaboração com a Plataforma de Bioinformática Fiocruz/RJ. A qualidade das sequências geradas foi avaliada utilizando o software FastQC (BABRAHAM BIOINFORMATICS, 2015). As sequências foram escaneadas para limpeza de sequências remanescentes de adaptadores e retirada dos 10 primeiros pares de base utilizando o software Trimmomatic (Bolger, et al, 2014). O software Sickle (JOSHI; FASS, 2011) foi utilizado para seleção de sequências com tamanho maior que 150pb e qualidade maior que 30.

As sequências de cada visita foram mapeadas separadamente frente a sequências de referência dos genes *Gag* e *Nef* através do software Geneious 9.0.5 (KEARSE et al., 2012), utilizando o algoritmo de mapeamento Geneious e 5 iterações para ajuste fino. A referência utilizada para o mapeamento de *Gag* consistiu na sequência referente as posições 690-2292 do genoma viral HXB2 (K03455), enquanto a referência utilizada para o mapeamento de *Nef* consistiu na sequência referente as posições 8697-9517 do mesmo genoma.

Os mapeamentos foram posteriormente exportados no formato SAM e em seguida convertidos, ordenados e indexados no formato BAM através do software Samtools (LI et al., 2009). O comando Samtools Depth foi utilizado para obter as estatísticas de cobertura de mapeamento.

As sequências consenso de cada alinhamento foram geradas através do software Geneious e um limite de cobertura de 20X foi utilizado para limitar o

tamanho dos consensos de *gag*. As sequências consenso para cada visita e cada paciente foram manualmente alinhadas entre si e com sequências de referência para o subtipo B (K03455) e subtipo F1 (AF005494) através do software Mega 6 (TAMURA et al., 2013). Através do mesmo software foram realizadas as estimativas de distância genética entre as sequências de visitas diferentes de um mesmo paciente, caracterizando a divergência genética no período, e foi construída a árvore filogenética Neighbor Joining utilizando 1000 replicatas de bootstrap. O modelo de substituição utilizado para ambas as análises foi o Tamura Nei, selecionado mediante verificação do melhor modelo através do software Jmodeltest (DARRIBA et al., 2012). A identificação de epítomos relacionados a resposta CTL relevantes para análise foi realizada com base no banco de dados de epítomos disponibilizados no Los Alamos HIV Immunology Database (LOS ALAMOS NATIONAL LABORATORY, 2015). Para cada indivíduo, foram gerados alinhamentos das sequências deduzidas de aminoácidos contendo as sequências de epítomos ótimos ou regiões de polimorfismos associados com a resposta citotóxica, ambos em acordo com o genótipo HLA-B do paciente. As tabelas com a descrição dos epítomos de *Gag* e *Nef* selecionados para análise e os mapas contendo suas respectivas localizações estão representados nas Tabelas 1 e 2 e nas figuras 6 e 7.

Paralelamente, Variações e SNPs entre as sequências para cada mapeamento foram identificados através do algoritmo de “*Variant Call*” do software Geneious, sendo considerados para as análises apenas variações com qualidade de “*base call*” maior que 30, cobertura maior que 100X e frequência maior que 0,5%. Os alinhamentos contendo os epítomos foram unidos às sequências consenso obtidas através do sequenciamento de nova geração e cruzados com as informações obtidas por análise de variantes nas sequências obtidos, de modo a identificar possíveis mutações de escape.

Tabela 1. Epítomos de Nef estudados e seus HLAs Associados

Epítopo	Posição	Sequência	HLA associado
RA9	19 - 27	RMRRRAEPAA	B*15
EP9	64 - 74	EEVGFPVKP	B*45
RF9	69 - 79	RPMTFKGAF	B*42
RM9	69 - 79	RPQVPLRPM	B*81
KF9	82 - 90	KAAFDLSFF	B*57/B*58
KL9	92 - 100	KEKGGLEGL	B*44
QL9	104 - 112	QRRQDILD	B*14/B*44
HW9	116 - 124	HTQGYFPDW	B*57/B*58/B*15
YY8	120 - 127	YFPDWQNY	B*58
FT9	120 - 128	FFPDWKNYT	B*51/B*57
YY9	127 - 135	YTPGPGIRY	B*57/B*42
YF9	135 - 143	YPLTFGWCF	B*49/B*57/B*15
WF9	183 - 191	WRFDSRLAF	B*15/B*51
RR9	188 - 196	RLAFHHMAR	B*52/B*51

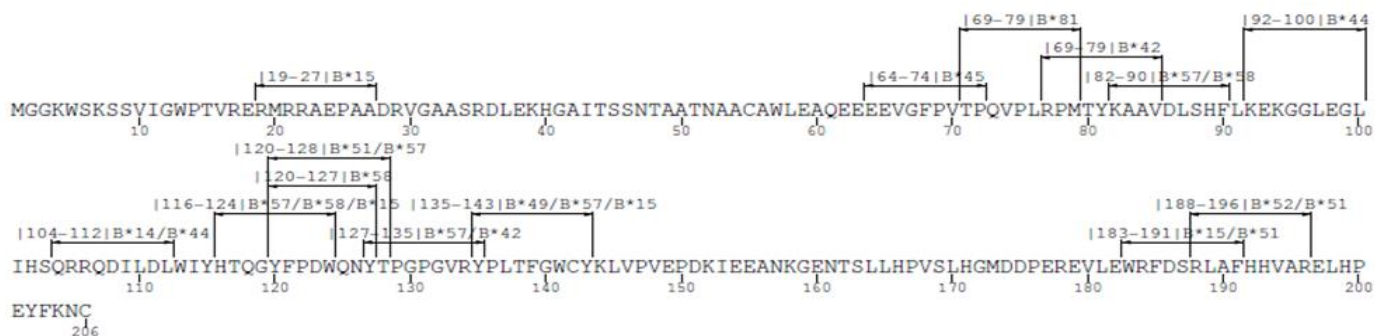


Figura 6. Mapa de epítomos de Nef. As regiões marcadas correspondem a locais com epítomos restritos pelos diferentes alelos HLA-B presentes nos pacientes selecionados para o estudo. No topo de cada marcação estão representadas as posições dos aminoácidos referentes ao epítopo e o HLA relacionado.

Tabela 2. Epítópos de Gag estudados e seus HLAs Associados

Epítopo	Posição	Sequência	HLA associado
RY10	20 - 29	RLRPGGKKHY	B*42
RM9	22 - 30	RPGGKKHYM	B*42/B*51
LF11	34 - 44	LWASRELERF	B*57
RY11	76 - 86	RSLYNTVATLY	B*57/B*58/B*44
KK9	114 - 122	KTQQAADK	B*57
G18	140 - 147	GQMVHQAI	B*48
VL10	143 - 152	VHQAISPRTL	B*15
AW10	146 - 155	AISPRTLNAW	B*57/B*58/B*15/B*81
VF9	156 - 164	VKVIIEKAF	B*15
II11	159 - 169	IIEKAFSPEVI	B*44/B*45/B*57/B*58
LL10	175 - 184	LSEGATPQDL	B*44
TM8	180 - 187	TPQDLNTM	B*81/B*42/B*58
GL9	193 - 201	GHQAAMQML	B*15
DW10	203 - 212	DTINEEAAEW	B*58
TW10	240 - 249	TSTLQEQIAW	B*58/B*57
GY9	269 - 277	GLNKIVRMV	B*15
R18	275 - 282	RMYSPTSI	B*52
RL11	294 - 304	RDYVDRFYKTL	B*44/B*58/B*15/B*57/B*14
RV9	305 - 313	RAEQASQEV	B*51
AW11	306 - 316	AEQASQDVKNW	B*44/B*57
VL9	313 - 321	VKNWMTETL	B*48
NI9	325 - 333	NANPDCKTI	B*51
DL9	329 - 337	DCKTILKAL	B*57
GL9	355 - 363	GPSHKARVL	B*42
AS10	364 - 373	AEAMSQVTNS	B*45
GF9	385 - 393	GPKRMKCF	B*42
CC9	405 - 413	CRAPRKKGC	B*14/B*57
R19	429 - 437	RQANFLGKI	B*48



Figura 7 Mapa de epítópos de Gag. As regiões marcadas correspondem a locais com epítópos restritos pelos diferentes alelos HLA-B presentes nos pacientes selecionados para o estudo. No topo de cada marcação estão representadas as posições dos aminoácidos referentes ao epítopo e o HLA relacionado.

4. RESULTADOS

4.1. CARACTERÍSTICAS CLÍNICO LABORATORIAIS E SÓCIO-DEMOGRÁFICAS DOS PACIENTES CONTROLADORES

As características clínico-laboratoriais e sócio demográficas para os 12 pacientes incluídos na casuística estão apresentadas na Tabela 3. Os indivíduos apresentaram a mediana de idade de 48 anos (IQR 45-51). Indivíduos do sexo feminino constituíram 2/3 da casuística e a categoria de exposição Heterossexual (58%) foi mais prevalente. As medianas de ano de diagnóstico e de tempo de supressão do HIV dos indivíduos foi, respectivamente, 1998 (IQR 1996-2000) e 15 anos (IQR 14-18).

Em relação aos padrões imunológicos, todos os pacientes apresentaram contagens medianas de Células T-CD4⁺ e Células T-CD8⁺ compatíveis com os valores observados para indivíduos saudáveis, evidenciando a estabilidade imunológica dos mesmos.

Dentre os 12 pacientes, a genotipagem viral revelou que apenas um apresentou infecção por um vírus de subtipo não-B, sendo genotipado como subtipo F (VC14). A genotipagem do HLA, por sua vez, identificou três pacientes com o alelo B*57 (EC08, VC15 e VC16), o qual é fortemente associado com a não progressão. Outros pacientes foram identificados com alelos B*51 (ECB42) e B*58 (EC18), os quais também têm sido, em menor nível, associados com a progressão lenta. Os pacientes VC05, EC02 e ECB09 apresentaram o alelo B*52, o qual foi recentemente associado a progressão lenta para a AIDS por estudos de nosso grupo (TEIXEIRA et al., 2014). Ademais, foram identificados dois indivíduos heterozigotos para a deleção $\Delta 32$ do CCR5 (EC11 e VC06), também relacionado com o controle do HIV. No conjunto da casuística, apenas os pacientes EC17 e VC14 não apresentaram um perfil genético associado com o controle da infecção e progressão para a AIDS.

Tabela 3. Características Clínico Laboratoriais e Sócio-Demográficas dos Pacientes selecionados para o estudo

Indivíduo	Sexo	Idade	Ano de Diagnóstico HIV	Anos conhecidos de Supressão HIV	Categoria de exposição	Mediana de CD4 (IQR)	Mediana de CD8 (IQR)	Frequência de Carga Viral			Alelos HLA-B	Genótipo CCR5	Visita 1 (Mês/Ano)	Visita X (Mês/Ano)	Diferença V1 x VX (Meses)
								<80cps/ml	81-400cps/ml	401-5000cps/ml					
EC02	F	52	1997	15	HETERO	1229,0 (1099-1406)	1510,0 (1412 - 1706)	100%			B*48, B*52	WT/WT	11/2008	08/2012	45
EC17	F	65	2000	15	NI	1813,0 (1588 - 2131)	815,0 (706 - 1011)	100%			B*14, B*15	WT/WT	09/2009	10/2013	49
EC08	F	43	1997	18	HETERO	1286,0 (1100 - 1421)	453,0 (422 - 527)	100%			B*45, B*57	WT/WT	02/2009	08/2013	54
ECB09	M	46	2001	14	HSH	835,0 (744 - 1031)	630,0 (558 - 773)	78%	22%		B*49, B*52	WT/WT	02/2009	11/2014	69
ECB11	F	48	1995	20	HETERO	1074,0 (1015 - 1216)	939,0 (823 - 1096)	94%		6%	B*49, B*81	WT/Δ32	12/2009	04/2012	28
ECB18	F	82	2001	9	HETERO	809,0 (684 - 930)	664,0 (595 - 749)	87%	13%		B*44, B*58	WT/WT	10/2009	09/2010	11
ECB13	F	61	1993	22	HETERO	966,0 (870 - 1117)	774,0 (533 - 932)	60%	40%		B*15, B*51	WT/WT	12/2009	11/2014	59
VC10	M	51	1991	24	NI	1231,0 (1113 - 1456)	837,0 (642 - 955)	20%	44%	36%	B*15, B*52	WT/WT	02/2009	06/2013	52
VC06	M	37	2000	11	HSH	1068,0 (864 - 1212)	970,0 (801 - 1127)	25%	35%	40%	B*15, B*48	WT/Δ32	01/2009	05/2011	28
VC14	F	45	1999	16	HETERO	694,0 (632 - 731)	671,0 (596 - 735)	60%	40%		B*42, B*44	WT/WT	04/2009	11/2014	67
VC15	F	41	2001	14	HETERO	701,0 (674 - 793)	842,0 (770 - 990)			100%	B*56, B*57	WT/WT	08/2009	02/2013	42
VC16	M	48	1998	17	HSH	553,0 (526 - 624)	794,0 (723 - 940)	26%	45%	29%	B*14, B*57	WT/WT	09/2009	09/2014	60

IQR=Amplitude Interquartil; F=Feminino; M=Masculino; HETERO=Heterossexual; HSH=Homem que faz sexo com homens; NI=Não Informado; WT= Wild-Type (Selvaagem)

4.2 RENDIMENTO E ESTATÍSTICAS DO NGS

As Estatísticas gerais para as sequências obtidos antes e após as etapas de triagem de qualidade estão descritas na Tabela 4. Após o sequenciamento, foram gerados uma média de 4.667.122 sequencias (IQR 5.491.122 – 3.688.230) medindo 35-200pbs por visita de cada indivíduo.

Tabela 4. Estatísticas das Sequências Pré e Pós-processamento

Amostras		Dado Bruto			Dados Pós-Processamento			
Indivíduo	Visita	Número de Reads	Tamanho (pb)	%GC	Número de Reads	Tamanho (pb)	%GC	Porcentagem do dado Bruto
EC02	v1	8821210	35-200	47	5003182	150-190	47	56,72%
	vx	4580712	35-200	47	3505033	150-190	47	76,52%
EC08	v1	5503207	35-200	48	3711980	150-190	48	67,45%
	vx	4503507	35-200	47	3228117	150-190	47	71,68%
EC17	v1	4978965	35-200	47	3598666	150-190	47	72,28%
	vx	4601957	35-200	46	3521212	150-190	46	76,52%
EC18	v1	6562425	35-200	47	4325987	150-190	47	65,92%
	vx	4040144	35-200	47	3238646	150-190	47	80,16%
ECB09	v1	6242473	35-200	47	4294573	150-190	47	68,80%
	vx	3688230	35-200	45	2782675	150-190	45	75,45%
EC11	v1	5491122	35-200	47	3968360	150-190	47	72,27%
	vx	2995589	35-200	46	2385946	150-190	46	79,65%
ECB13	v1	5871097	35-200	48	1684093	150-190	48	28,68%
	vx	3242877	35-200	45	2482542	150-190	45	76,55%
VC06	v1	5360962	35-200	46	3756561	150-190	46	70,07%
	vx	3921045	35-200	46	3010452	150-190	46	76,78%
VC10	v1	6285675	35-200	47	4226988	150-190	47	67,25%
	vx	3687296	35-200	46	2783816	150-190	46	75,50%
VC14	v1	2524293	35-200	46	861491	150-190	46	34,13%
	vx	3123780	35-200	45	2391996	150-190	45	76,57%
VC15	v1	4531209	35-200	47	3389165	150-190	47	74,80%
	vx	3073206	35-200	47	2321730	150-190	47	75,55%
VC16	v1	5282854	35-200	48	3758835	150-190	48	71,15%
	vx	4053824	35-200	46	3014330	150-190	46	74,36%
C+ HXB2		3710409	35-200	46	2825252	150-190	46	76,14%

Ainda na primeira etapa, a análise de qualidade através do programa FastQC (Figura 8) demonstrou que a maior parte das sequências apresentavam alta qualidade (entre 38 e 40 pontos) e tamanho próximo de 200pbs. No entanto, um certo nível de divergência no conteúdo nucleotídico foi identificado no início das sequências, sendo identificado como artefatos gerados durante o preparo da biblioteca com kits Nextera XT.

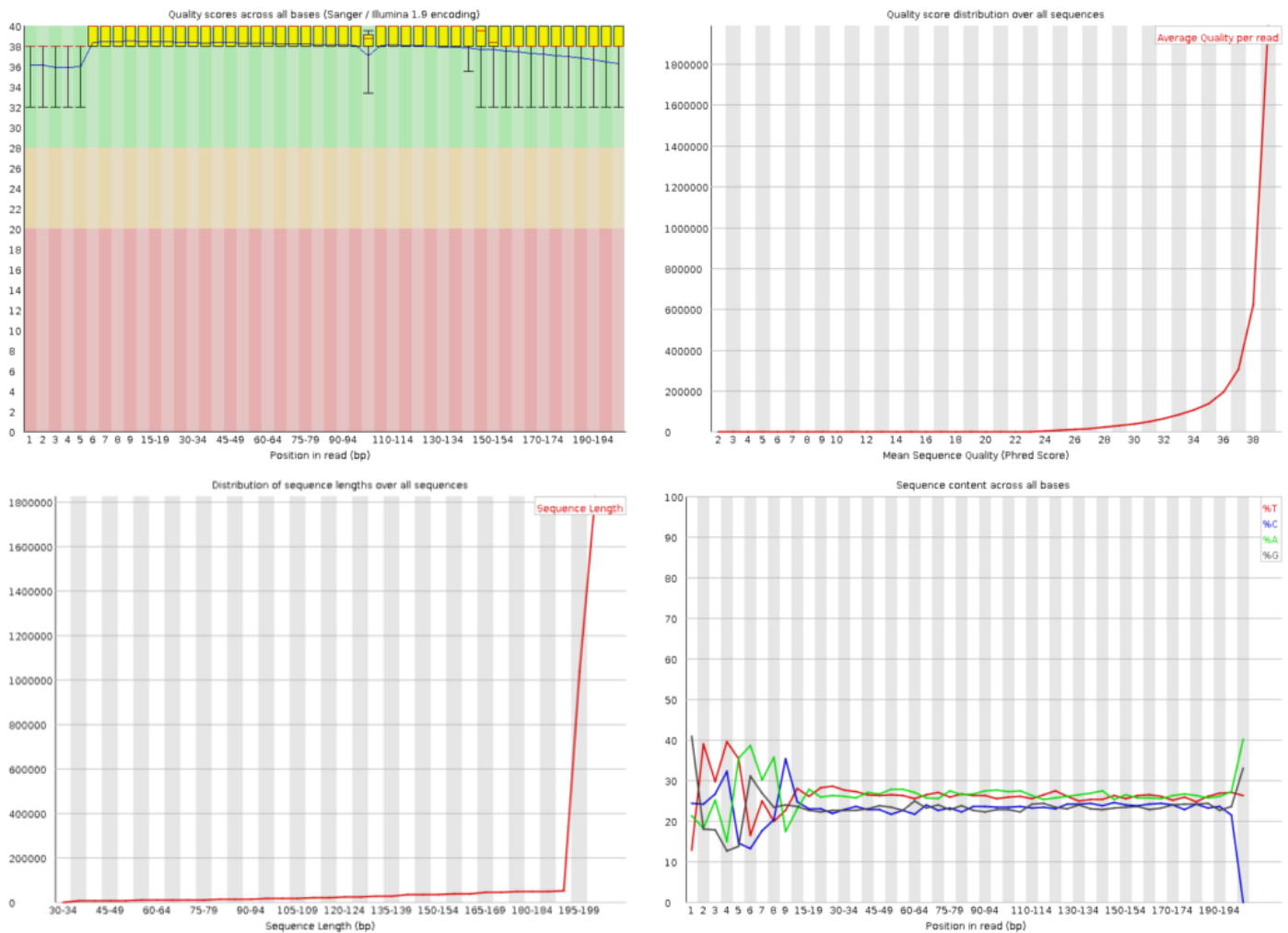


Figura 8. Estatísticas de qualidade, tamanho e conteúdo de bases dos fragmentos referentes ao sequenciamento do Controle positivo HXB2. Os gráficos são semelhantes e a figura apresentada é representativa para a maior parte das amostras sequenciadas.

De modo a melhorar a qualidade do mapeamento posterior, foi realizada a limpeza de fragmentos remanescentes de adaptadores e triagem de fragmentos com qualidade maior que 30 e tamanho maior que 150pbs. Devido aos artefatos no início das sequências, os primeiros 10 nucleotídeos de cada read também foram retirados. Ao fim do processamento, restaram para cada visita de paciente, uma média de 70% (IQR 68% - 76%) da quantidade inicial das sequências com tamanho entre 150-190pbs. As amostras ECB13V1 e VC14V1 apresentaram menor rendimento (cerca de 30%) após o processamento devido a maior distribuição de sequências menores que 150pbs, apesar de alta qualidade (Figura 9).

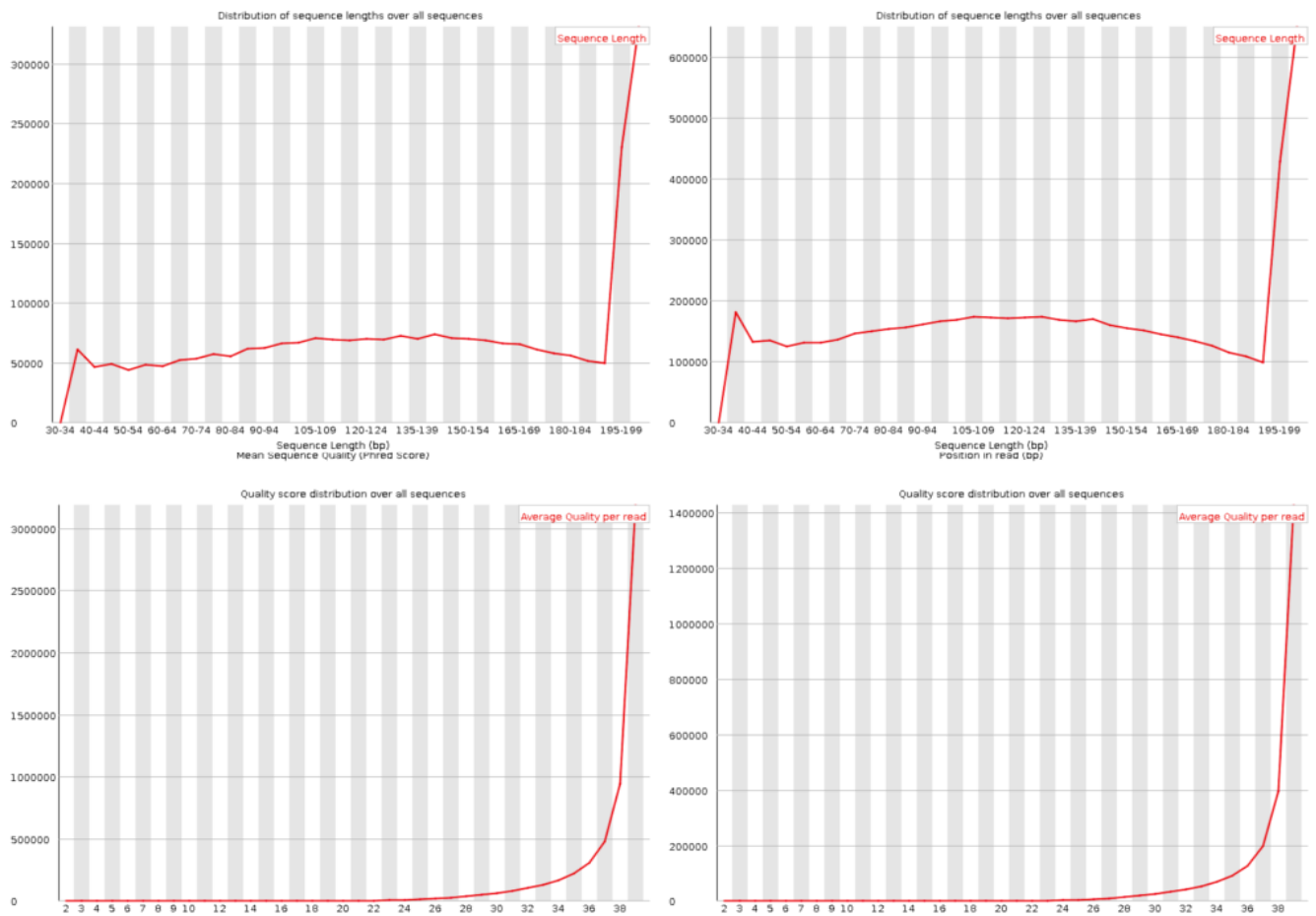


Figura 9. Estatísticas de distribuição de tamanho (Acima) e qualidade (Abaixo) das sequências das amostras ECB13V1 (À Esquerda) e VC14V1 (À Direita).

4.3 MAPEAMENTO DOS GENES GAG E NEF

As estatísticas de mapeamento das sequências frente às referências de *gag* e *nef* estão demonstradas na Tabela 5.

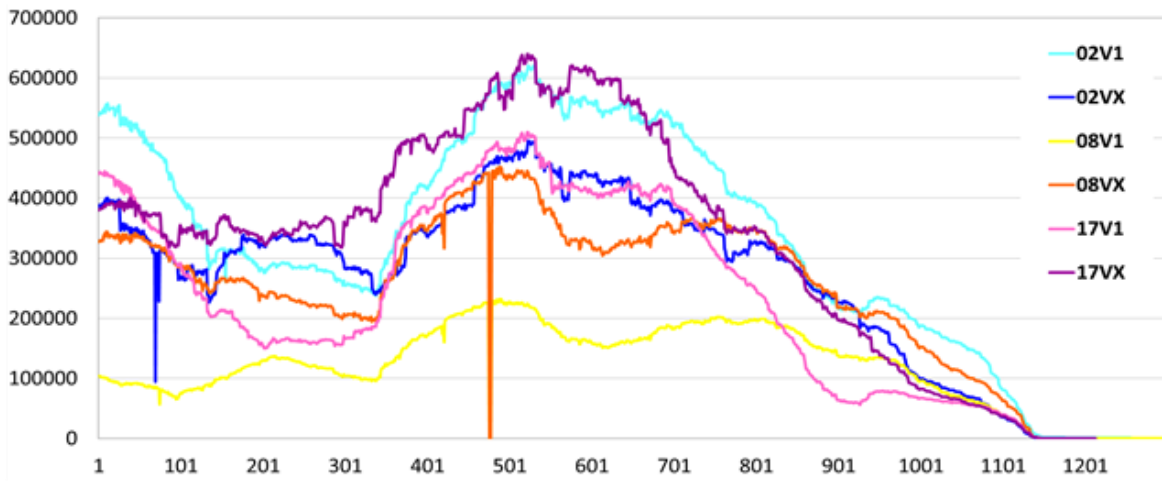
Tabela 5. Estatísticas de Mapeamento dos genes *gag* e *nef*

Amostras		Número de Reads	Gag		Nef	
Indivíduo	Visita		Cobertura por pb (IQR)	Tamanho do consenso (nb)	Cobertura por pb (IQR)	Tamanho do consenso (pb)
EC02	V1	5.003.182	313.475 (185.803 – 496.466)	1 - 1238	481.172 (426.211 – 563.297)	1-821
	VX	3.505.033	278.739 (224.901 – 378.457)	1 - 1115	261.540 (187.512 – 347.389)	1-821
EC08	V1	3.711.980	117.678 (77.763 – 174.696)	1 - 1233	411.984 (335.414 – 503.150)	1-821
	VX	3.228.117	265.931 (207.848 – 344.114)	1 - 1084	269.424 (220.768 – 333.011)	1-821
EC17	V1	3.598.666	250.903 (94.798 - 407.124)	1 - 1065	-----	-----
	VX	3.521.212	334.115 (193.692 – 491.751)	1 - 1113	100.458 (63.477 – 135.286)	1-821
ECB09	V1	4.294.573	-----	-----	-----	-----
	VX	2.782.675	225.504 (58.510 – 383.411)	1 - 1223	179.882 (163.469 – 203.604)	1-821
EC11	V1	3.968.360	259.633 (142.248 – 378.219)	1 - 1095	478.446 (434.887 – 534.803)	1-821
	VX	2.385.946	195.481 (135.711 – 273.048)	1 - 1062	283.701 (264.096 – 312.581)	1-821
ECB13	V1	1.684.093	-----	-----	64.325 (56.038 – 71.145)	1-821
	VX	2.482.542	204.823 (150.037 – 276.734)	1 - 1054	229.115 (207.804 – 258.725)	1-821
EC18	V1	4.325.987	308.385 (105.213 – 535.980)	1 - 1188	299.952 (276.520 – 329.667)	1-821
	VX	3.238.646	183.912 (63.292 – 272.838)	1 - 1213	426.834 (400.251 – 478.804)	1-821
VC06	V1	3.756.561	365.832 (139.187 – 571.490)	1 - 1098	-----	-----
	VX	3.010.452	249.588 (138.606 – 399.083)	1 - 1162	123.688 (109.534 – 140.590)	1-821
VC10	V1	4.226.988	140.795 (57.552 – 225.829)	1 - 1244	586.848 (543.124 – 646.658)	1-821
	VX	2.783.816	228.698 (68.823 – 364.659)	1 - 1241	207.286 (192.997 – 233.045)	1-821
VC14	V1	861.491	70.910 (51.944 – 90.448)	1 - 1069	65.595 (63.448 – 78.992)	1-821
	VX	2.391.996	246.708 (205.839 – 332.977)	1 - 1117	18.498 (17.311 – 21.300)	1-821
VC15	V1	3.389.165	184.682 (86.656 – 297.503)	1 - 1235	456.541 (421.875 – 517.799)	1-821
	VX	2.321.730	122.896 (82.646 – 173.119)	1 - 1246	323.997 (281.137 – 386.994)	1-821
VC16	V1	3.758.835	111.700 (62.243 – 172.716)	1 - 1232	611.240 (473.587 – 753.073)	1-821
	VX	3.014.330	214.445 (69.671 – 329.870)	1 - 1252	236.389 (133.982 – 332.199)	1-821
C+ HXB2		2.825.252	190.203 (5336 – 324.901)	1 - 1503	224.970 (207.361 – 247.079)	1-821

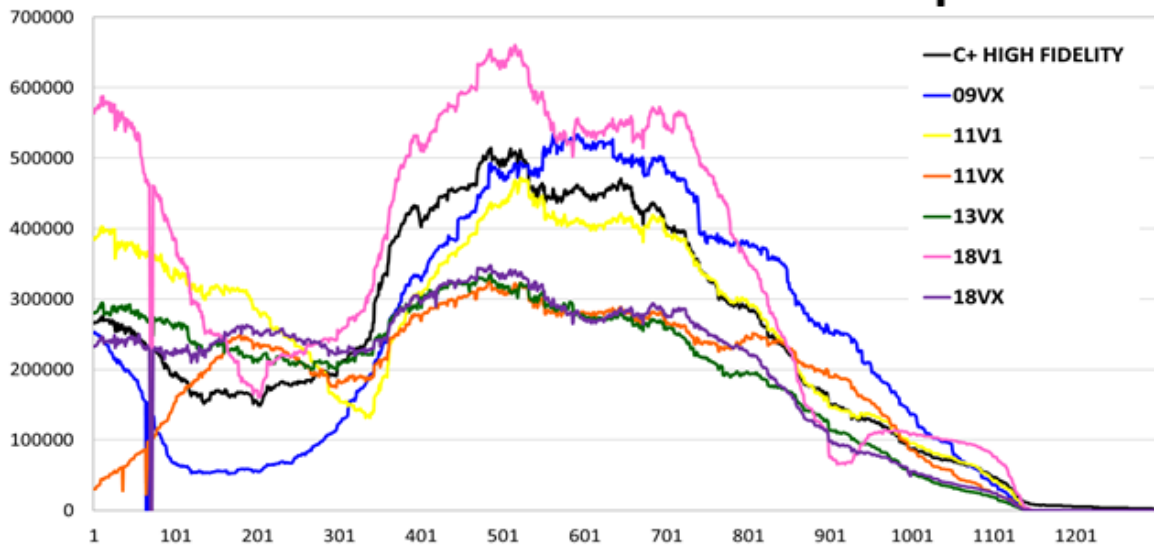
Para o gene *gag*, a sequência de referência utilizada no mapeamento HXB2 compreendia 100 bps antes do início de *Gag* e 1500pbs do gene *gag* (correspondendo as posições 690-2190 do HXB2). Para este gene, a cobertura média foi de 212.674 sequências por pb (IQR 173.132 – 261.207). A **Figura 10** apresenta os gráficos correspondentes a cobertura por par de base para *gag* de cada amostra mapeada. Estabelecendo um limite de cobertura de 20X, foi possível reconstruir consensos compreendendo, no mínimo, os primeiros 1054pbs do gene (ECB13 VX). Para as amostras ECB09 V1, EC17 V1 e ECB13 V1 não foi possível mapear corretamente o gene *gag*.

Para o gene *nef*, a sequência de referência utilizada no mapeamento compreendia o gene *nef* somado a 100 bps antes e depois do gene (correspondendo as posições 8697-9517 do HXB2). A **Figura 11** apresenta os gráficos correspondentes a cobertura por par de base para *Nef* de cada amostra mapeada. Para este gene a cobertura média foi de 286.463 leituras por pb (IQR 179.882 – 426.834) e os consensos gerados possuem toda a sequência de *Nef*. Para este gene, não foi possível mapear corretamente as amostras ECB09 V1 e VC06 V1.

Controladores de Elite



Controladores de Elite com Blips



Controladores Virêmicos

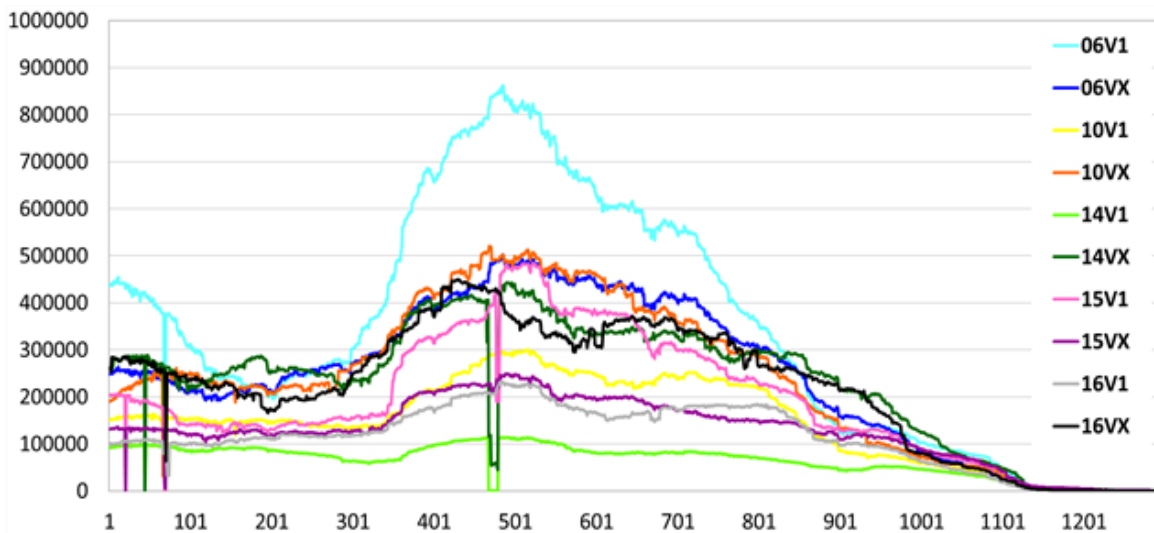
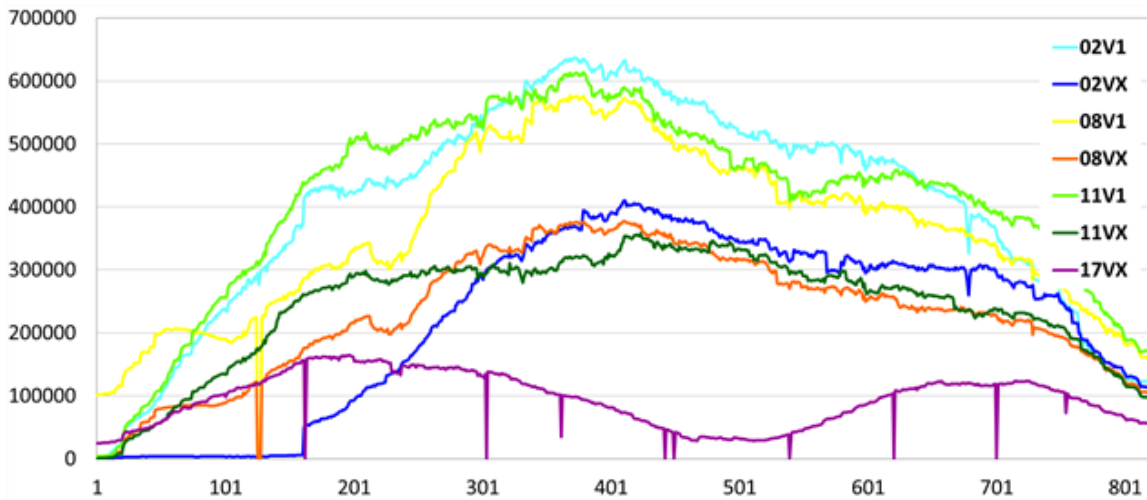
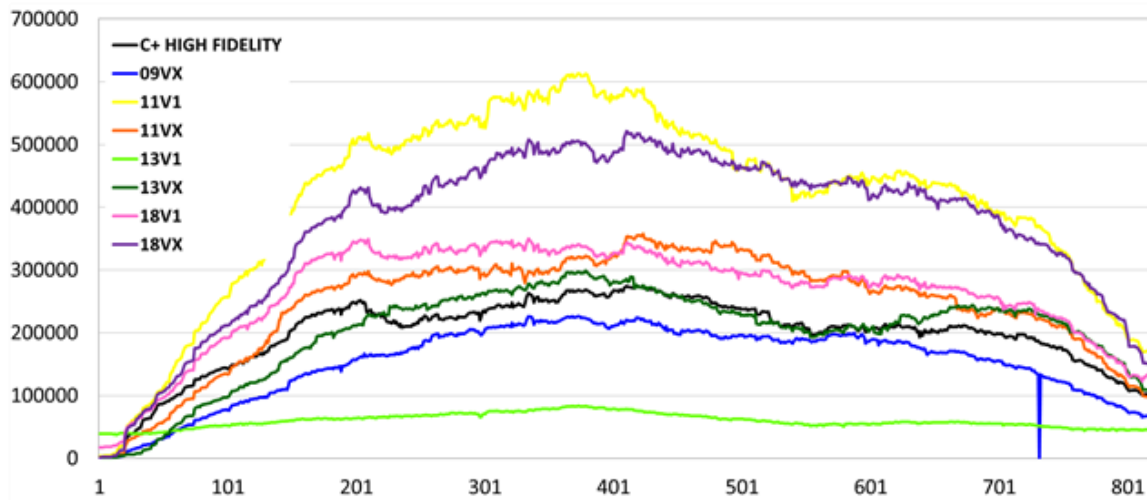


Figura 10. Cobertura por par de base para o *gene gag* de cada visita de cada indivíduo estudado. Os dados referentes a Visita 1 (V1) e a visita mais recente (VX) estão coloridos de acordo com a legenda.

Controladores de Elite



Controladores de Elite com Blips



Controladores Virêmicos

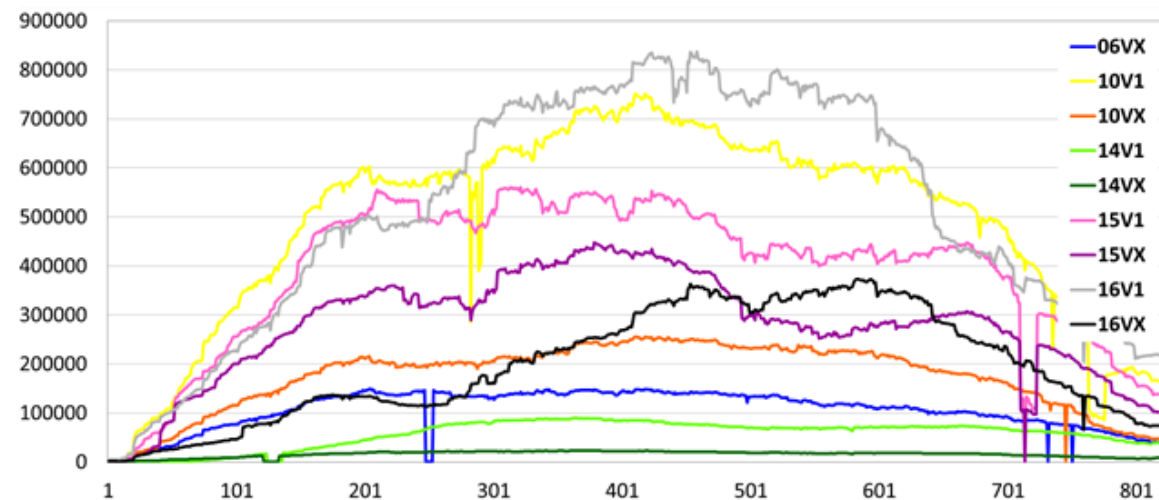


Figura 11. Cobertura por par de base para o *gene nef* de cada visita de cada indivíduo estudado. Os dados referentes a Visita 1 (V1) e a visita mais recente (VX) estão coloridos de acordo com a legenda.

4.4 ANÁLISES FILOGENÉTICAS

A partir dos mapeamentos obtidos, sequências consenso para cada gene foram obtidas para as visitas V1 e VX respectivas de cada indivíduo. Embora o sequenciamento da região *Nef* referente a primeira visita dos indivíduos ECB09 e VC06 não tenha tido sucesso, sequências do mesmo ponto, obtidas por meio de sequenciamento pelo método de Sanger durante estudos anteriores de nosso grupo, estavam disponíveis e foram utilizadas para as análises de diversidade. O mesmo foi verdadeiro para o gene *gag* na primeira visita dos indivíduos ECB09 e ECB13. A Árvore filogenética Neighbour Joining obtida para o gene *gag* está representada na **Figura 12**.

Como esperado, sequências referentes a V1 e VX de cada paciente agruparam-se em clusters próprios, confirmando a identidade dos pares. Comparativamente, as sequências das duas visitas para cada paciente demonstraram divergência entre si, com diferenças nos comprimentos dos ramos, caracterizando algum nível de evolução viral ao longo do período. Para os indivíduos VC06 e ECB09, os clusters referentes a cada indivíduo também agruparam entre si com um alto suporte, indicando origem filogenética próxima. No entanto, a comparação das sequências atuais com sequências obtidas anteriormente demonstrou que tal relação não é fruto de contaminação das amostras. Por fim, as sequências do indivíduo VC14 agruparam junto com a referência para o subtipo F1, confirmando o genótipo viral no gene *gag*.

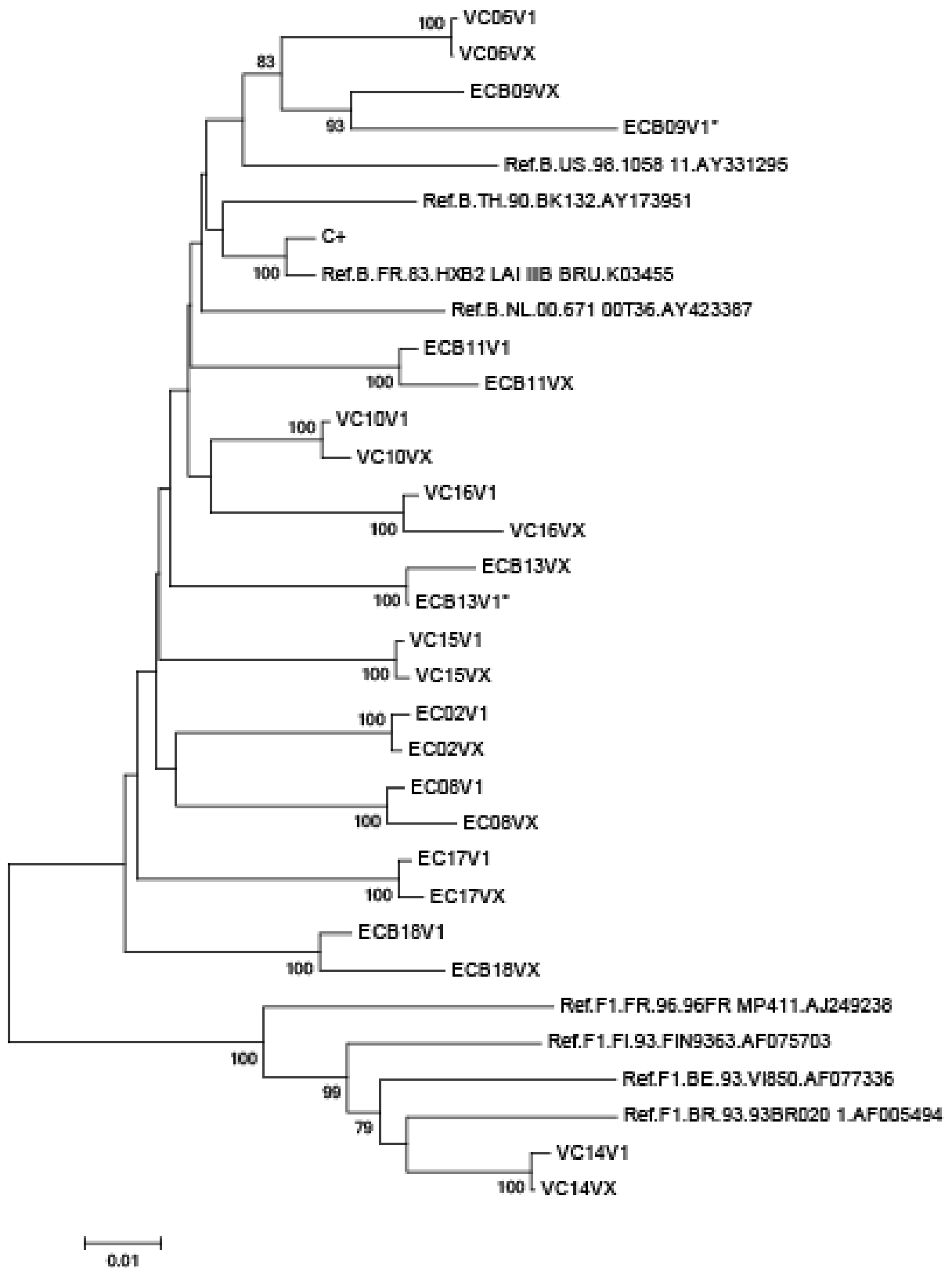


Figura 12. Árvore filogenética Neighbor-Joining com as sequências consenso do *gene gag* referentes as visitas V1 e VX dos indivíduos estudados. Sequências marcadas por um asterisco foram obtidas a partir de estudos anteriores. A árvore foi enraizada no ponto médio. Sequências de referência para o subtipo B e subtipo F1 do banco de dados de Los Alamos foram utilizadas. Valores de Bootstrap menores que 75 foram omitidos.

Para melhor demonstrar diferenças entre as sequências referentes a V1 e VX em *Gag*, as figuras 13, 14 e 15 apresentam os gráficos de identidade entre sequências V1 e VX para cada indivíduo. No geral, ocorreram poucas posições diferentes entre os consensos de cada indivíduo, evidenciando um baixo nível de mutação no gene. Tal padrão de baixa variação não foi observado apenas para os indivíduos ECB09 e ECB18, os quais apresentaram um grande número de diferenças entre os consensos V1 e VX.

Análises mais minuciosas do mapeamento do gene *gag* para a visita VX do indivíduo ECB09 demonstraram que, nas regiões com grande número de diferenças em comparação com o consenso V1, as sequências obtidas podiam ser divididas em dois grupos, indicando a presença de variantes distintas. Sequências características de cada uma das variantes foram então manualmente separadas do mapeamento e utilizadas para geração de um consenso para cada uma das variantes. Dessa forma foi obtida uma variante encontrada no mapeamento com frequências próximas de 60% (ECB09VX Variante 1) e uma variante com frequências próximas a 40% (ECB09VX Variante 2).

A figura 16 exhibe as relações filogenéticas entre as sequências em questão e a sequência consenso para a V1 do paciente ECB09. As três sequências apresentaram íntima relação filogenética, agrupando no mesmo cluster, apesar de grande distância genética. A Variante 1 obtida se mostrou mais próxima da sequência V1, com 2,7% de distância genética, enquanto a variante 2 apresentou 5% de distância genética em comparação com a sequência V1. Estudos paralelos do nosso grupo realizados a pouco tempo e visando a análise dos padrões de evolução viral intra-hospedeiro no gene *Env* revelaram uma possível superinfecção o gene do envelope Viral desse mesmo indivíduo e, apesar dos dados obtidos para *gag* não permitirem afirmar o mesmo, os altos valores de distância genética podem estar relacionados.

Já para o indivíduo ECB18, a análise de SNPs evidenciou uma mutação na visita VX no códon 215, resultando numa alteração de Triptofano para um Stop códon. Tal mutação aparentemente decorreu de uma mutação G->A, podendo ser associado com a ação de APOBEC3G, de forma que a ocorrência de hipermutação pode explicar o alto número de diferenças entre as sequências V1 e VX.

Controladores de Elite (EC)

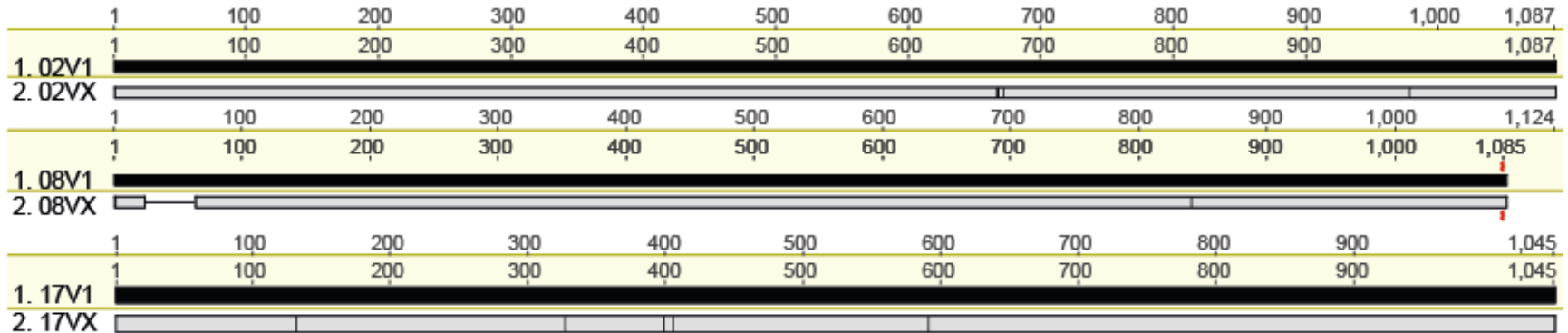


Figura 13 Comparativo de Identidade entre as seqüências V1 e Vx do *gene gag* para indivíduos Controladores de Elite. As seqüências referentes a Visita 1 foram utilizadas como referência e traços ao longo da seqüência Vx indicam posições com diferenças entre as duas seqüências. Inserções ou deleções aparecem como retas ao longo do trecho diferente. Os gráficos foram gerados através do programa Geneious 9.0.5.

Controladores de Elite com Blips (ECB)

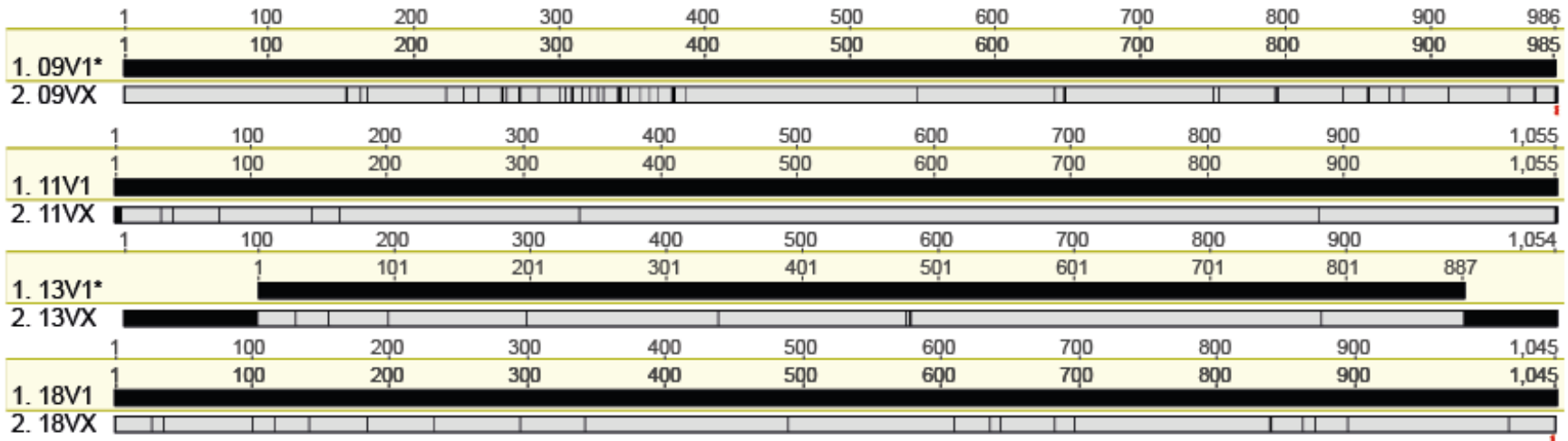


Figura 14 Comparativo de Identidade entre as seqüências V1 e Vx do *gene gag* para indivíduos Controladores de Elite com Blips. As seqüências referentes a Visita 1 foram utilizadas como referência e traços ao longo da seqüência Vx indicam posições com diferenças entre as duas seqüências. Os gráficos foram gerados através do programa Geneious 9.0.5.

Controladores Virêmicos (VC)

	1	100	200	300	400	500	600	700	800	900	1,052
1. 06V1	1	100	200	300	400	500	600	700	800	900	1,052
2. 06VX	[Sequência Vx com traços de diferenças]										
1. 10V1	1	100	200	300	400	500	600	700	800	900	1,087
2. 10VX	1	100	200	300	400	500	600	700	800	900	1,087
1. 14V1	1	100	200	300	400	500	600	700	800	900	1,072
2. 14VX	[Sequência Vx com traços de diferenças]										
1. 15V1	1	100	200	300	400	500	600	700	800	900	1,093
2. 15VX	1	100	200	300	391	491	591	691	791	891	1,084
1. 16V1	1	100	200	300	400	500	600	700	800	900	1,087
2. 16VX	[Sequência Vx com traços de diferenças]										

Figura 15 Comparativo de Identidade entre as sequências V1 e Vx do *gene gag* para indivíduos Controladores Virêmicos. As sequências referentes a Visita 1 foram utilizadas como referência e traços ao longo da sequência Vx indicam posições com diferenças entre as duas sequências. Inserções ou deleções aparecem como retas ao longo do trecho diferente. Os gráficos foram gerados através do programa Geneious 9.0.5.

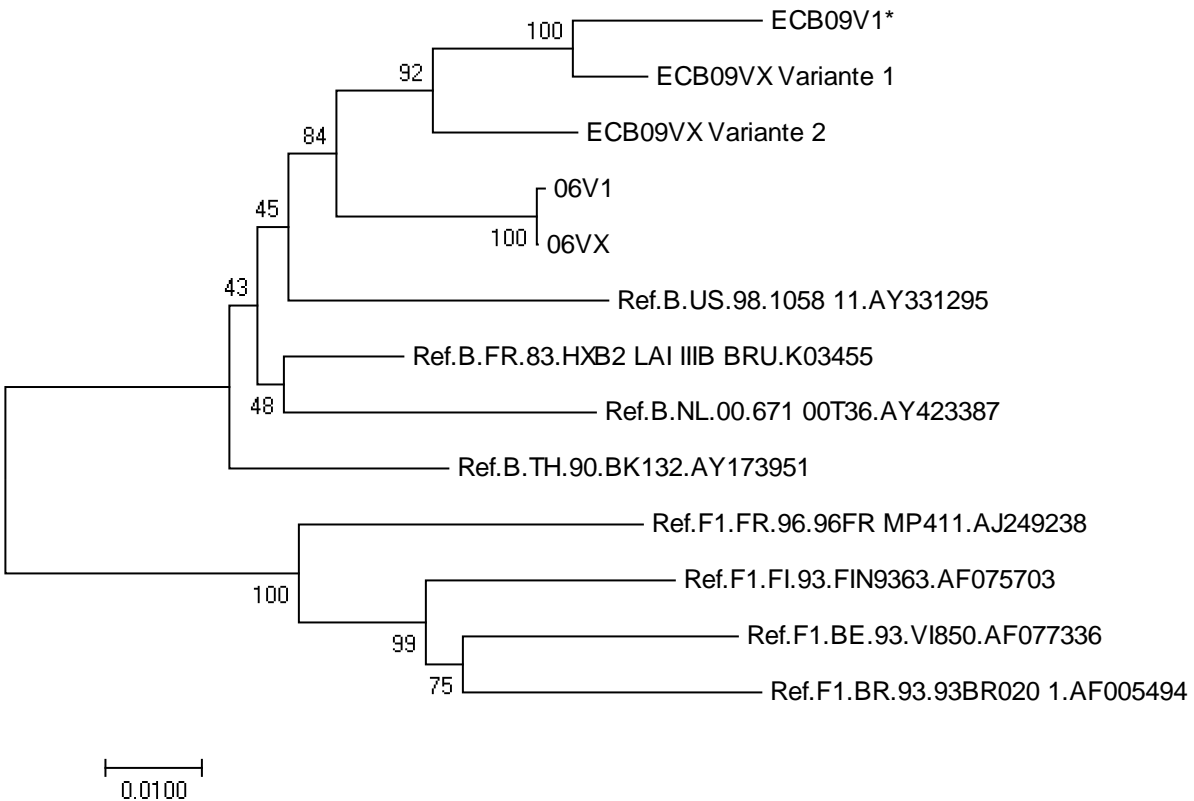


Figura 16 Árvore filogenética Neighbor-Joining com as sequências das duas variantes isoladas para o indivíduo ECB09. A árvore foi enraizada no ponto médio. Sequências de referência para o subtipo B e subtipo F1 do banco de dados de Los Alamos foram utilizadas. Valores de Bootstrap menores que 75 foram omitidos.

Para o gene *nef*, por sua vez, a árvore filogenética Neighbor Joining está representada na Figura 17. Assim como para Gag, sequências referentes a V1 e VX para cada indivíduo também agruparam em cluster únicos e as diferenças de comprimento entre os ramos também foram indicativas de evolução viral na maioria dos casos. A estreita relação filogenética entre os vírus dos indivíduos ECB09 e VC06 também pode ser observada para o gene *nef*. Diferente de Gag, no entanto, padrões filogenéticos diferenciados puderam ser observados para as sequências 13VX e 17VX.

Para a sequência ECB13VX, análises da sequência proteica traduzida a partir do consenso evidenciaram a presença de Códons de parada substituindo os aminoácidos nas posições 13, 57, 113, 141 e 183. Tais defeitos foram associados com mutações G->A em códons TGG, sendo indicativos de hipermutação mediada por APOBEC3G. A análise

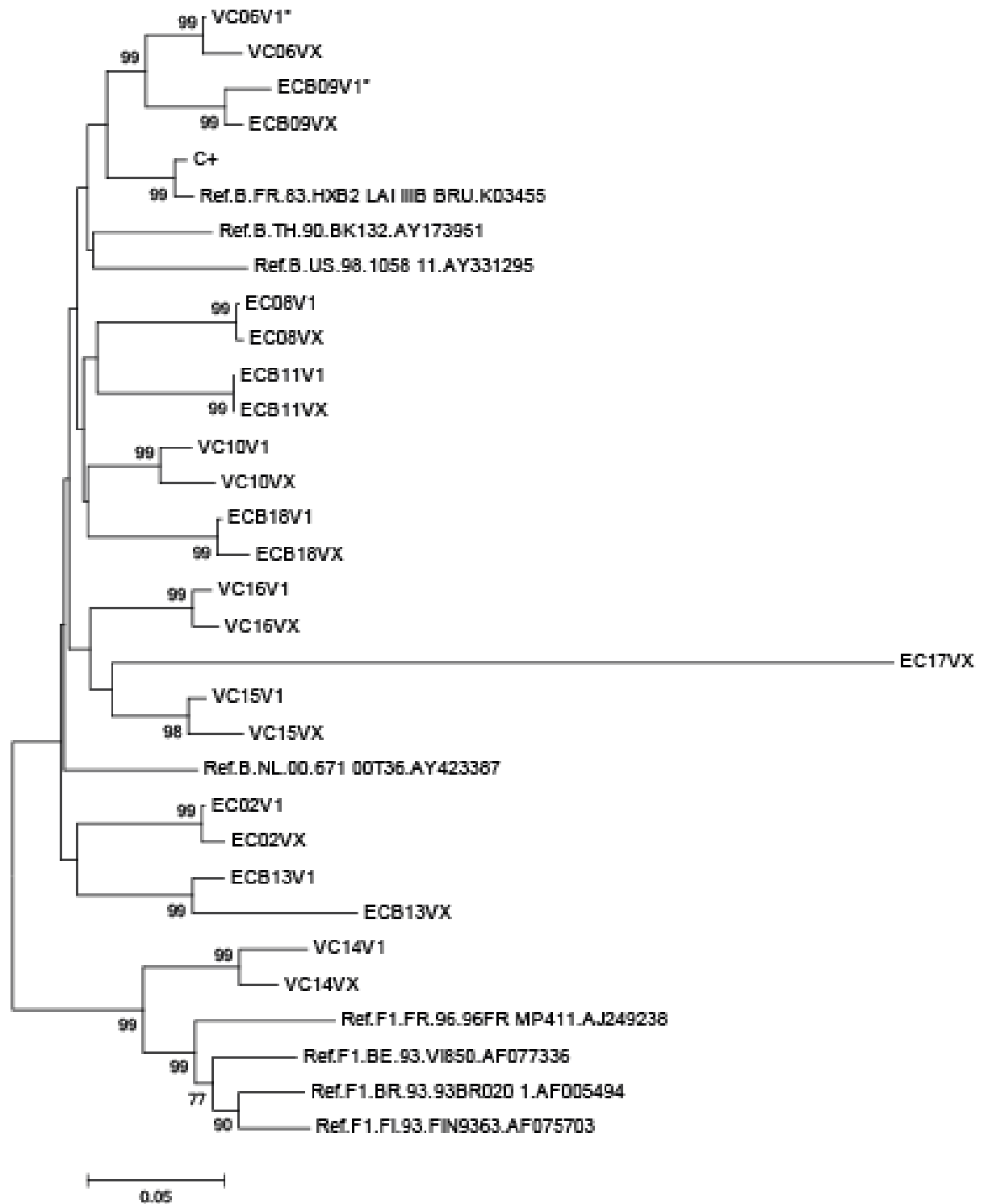


Figura 17. Árvore filogenética Neighbor-Joining com as sequências consenso do *gene nef* referentes as visitas V1 e VX dos indivíduos estudados. Sequências marcadas por um asterisco foram obtidas a partir de estudos anteriores. A árvore foi enraizada clado de sequências F. Sequências de referência para o subtipo B e subtipo F1 do banco de dados de Los Alamos foram utilizadas. Valores de Bootstrap menores que 75 foram omitidos.

de SNPs para essas posições demonstrou uma frequência entre 88,6% e 92,6% da mutação defectiva entre o total de sequências analisados em cada posição, indicando predominância de variantes virais defectivas na amostra VX do indivíduo ECB13.

Já para a sequência EC17VX, a análise da sequência proteica evidenciou grande número de inserções e deleções ao longo do gene *nef*, causando mutações “Frame Shift”. Em adição a essas mutações pontuais, um grande trecho a partir da metade do gene se mostrou mutado, gerando problemas observados no mapeamento. A partir do mapeamento, sequências cobrindo a região problemática foram extraídas e realinhadas separadamente, permitindo identificar a presença de uma deleção de 135pbs. No conjunto, estes dados evidenciam um possível defeito genético em *Nef* na visita VX do indivíduo EC17.

Para evidenciar as diferenças entre V1 e Vx, como feito para *Gag*, os gráficos de identidade entre as visitas V1 e VX de *Nef* para cada um dos pacientes está representada nas figuras 17, 18 e 19. Comparativamente, as sequências do gene *nef* apresentaram visualmente maior número de diferenças do que observado para o gene *gag*. Além disso, diferenças entre as sequências referentes a V1 e VX se mostraram mais frequentes nos indivíduos Controladores Virêmicos, em comparação com os Controladores de Elite com e sem Blips. Tal fato evidencia um maior nível de evolução no gene *nef* em comparação com o gene *gag*, o qual se mostrou mais conservado. Em adição, a figura 21 apresenta a comparação da sequência obtida para a visita VX do indivíduo EC17 com a sequência de referência HXB2, indicando posições com mutações FrameShift e a região da deleção encontrada.

Controladores de Elite (EC)

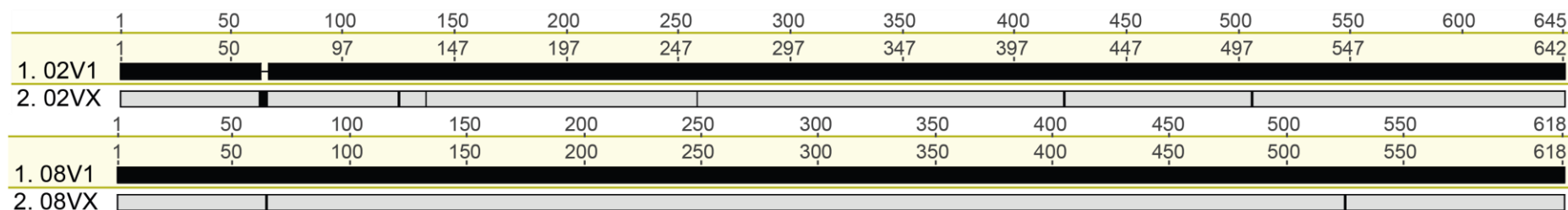


Figura 19 Comparativo de Identidade entre as sequências V1 e Vx do *gene nef* para indivíduos Controladores de Elite. As sequências referentes a Visita 1 foram utilizadas como referência e traços ao longo da sequência Vx indicam posições com diferenças entre as duas sequências. Inserções ou deleções aparecem como retas ao longo do trecho diferente. Os gráficos foram gerados através do programa Geneious 9.0.5.

Controladores de Elite com Blips (ECB)

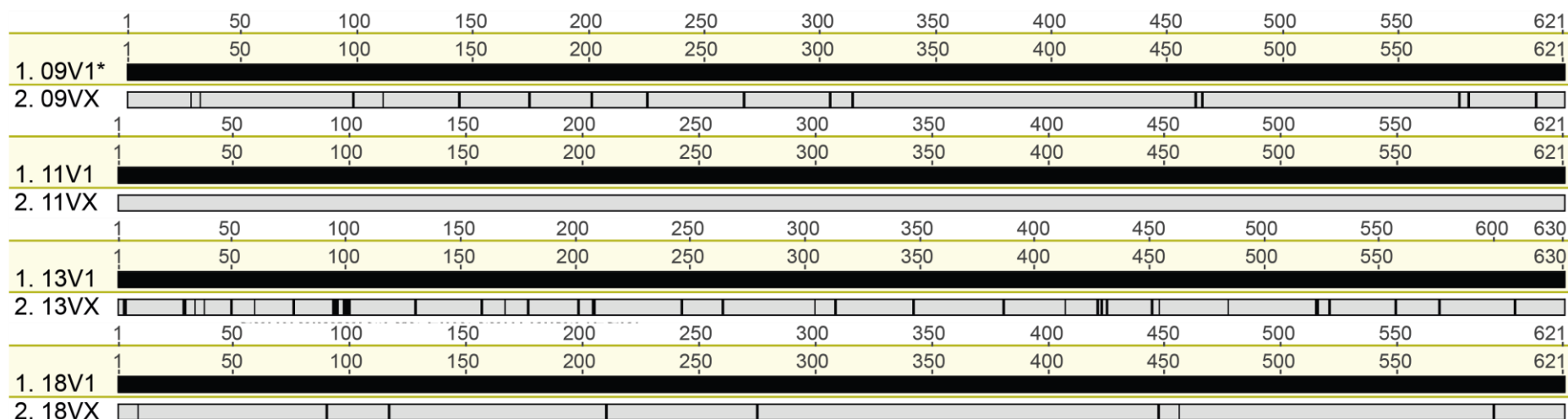


Figura 18 Comparativo de Identidade entre as sequências V1 e Vx do *gene nef* para indivíduos Controladores de Elite com Blips. As sequências referentes a Visita 1 foram utilizadas como referência e traços ao longo da sequência Vx indicam posições com diferenças entre as duas sequências. Os gráficos foram gerados através do programa Geneious 9.0.5.

Controladores Virêmicos (VC)



Figura 20 Comparativo de Identidade entre as sequências V1 e Vx do *gene nef* para indivíduos Controladores Virêmicos. As sequências referentes a Visita 1 foram utilizadas como referência e traços ao longo da sequência Vx indicam posições com diferenças entre as duas sequências. Inserções ou deleções aparecem como retas ao longo do trecho diferente. Os gráficos foram gerados através do programa Geneious 9.0.5.



Figura 21 Comparativo de Identidade entre a sequência Vx do indivíduo EC17 e a sequência de referência HXB2 para o gene Nef. Regiões com um traço horizontal representam gaps no alinhamento. Quadros vermelhos indicam regiões com inserções ou deleções associadas a mutações Frameshift. Os gráficos foram gerados através do programa Geneious 9.0.5.

De modo a melhor caracterizar a dinâmica de evolução viral, a divergência entre as sequências V1 e VX para cada indivíduo foram obtidas através do cálculo da distância genética entre as mesmas, sendo representadas na Tabela 6. Para o gene *gag*, os valores obtidos para a grande maioria dos pacientes oscilaram próximo a casa dos 0,5% de divergência entre as duas visitas. Comparando entre os grupos, a média da distância genética foi semelhantes entre ECs e ECBs (0,7% x 0,5%, respectivamente), mas maiores em ECBs (2,1%). Tal diferença ocorreu devido aos altos valores de divergência para os indivíduos ECB09 e ECB18, os quais apresentaram evidências de superinfecção e hipermutação, como mencionado anteriormente. As relações se mantiveram mesmo quando os valores de distância genética foram normalizados com base no tempo entre V1 e VX para cada indivíduo.

Tabela 6. Valores de Divergência entre visitas V1 e VX nos genes *gag* e *nef*

Grupo	Indivíduo	Gag			Nef		
		Distância Genética	Desvio Padrão	Distância Normalizada*	Distância Genética	Desvio Padrão	Distância Normalizada*
EC	EC02	0,004	0,002	0,082	0,009	0,004	0,210
	EC08	0,011	0,003	0,207	0,003	0,002	0,060
	EC17	0,005	0,002	0,099	-	-	-
	Média	0,007	0,002	0,129	0,006	0,003	0,135
ECB	ECB09	0,049	0,010	0,712	0,023	0,006	0,335
	ECB11	0,008	0,003	0,273	0,000	0,000	0,000
	ECB13	0,009	0,003	0,154	0,072	0,011	1,221
	ECB18	0,020	0,004	1,860	0,013	0,005	1,188
	Média	0,022	0,005	0,750	0,027	0,005	0,686
VC	VC06	0,001	0,001	0,034	0,015	0,006	0,530
	VC10	0,005	0,002	0,089	0,031	0,007	0,597
	VC14	0,003	0,002	0,042	0,039	0,008	0,581
	VC15	0,003	0,002	0,066	0,026	0,007	0,629
	VC16	0,015	0,004	0,249	0,016	0,027	0,272
	Média	0,005	0,002	0,096	0,026	0,011	0,522

*Os valores normalizados equivalem a distância genética dividida pelo número de meses entre V1 e VX multiplicado por 1000.

Para o gene *nef*, os valores de divergência encontrados foram, em geral, maiores do que os observados para o gene *gag* (acima de 1%), indicando uma menor conservação genética do gene. No entanto, para indivíduos ECs, a média de divergência

foi menor do que para ECBs e VCs (0,6% x 2,7% x 2,5%, respectivamente). Da mesma forma que para o gene *gag*, a média de distância genética entre os indivíduos ECBs foi alta, comparável a dos indivíduos VCs, mas tal dado se mostra desviado pelo alto valor de divergência do indivíduo ECB13 (7,2%), o qual também apresentou evidências de hipermutação no gene. Para esse gene, a relação observada entre os grupos também se manteve após normalização dos valores com base no tempo entre V1 e VX.

Ainda, como uma medida da diversidade intrapopulação de cada paciente, as estatísticas de quantidade de SNPs e as características de suas mutações associadas foram calculadas. Para este resultado e com base nos resultados de variação para o controle positivo sequenciado, uma frequência de 0,5% foi considerada mínima para considerar uma variante verdadeira.

Para o gene *gag* (Tabela 7), indivíduos VCs apresentaram, em média, maior número de mutações (193 variantes) e posições variáveis (96 posições) do que ECs e ECBs. Entre ECs e ECBs, os valores médios foram maiores para o segundo grupo (128 variantes em 64 posições) do que para o primeiro (109 variantes em 54 posições), porém próximos. A razão entre a quantidade de posições mutadas e o total de posições analisadas também foi maior nos VCs (0,08), em comparação com os ECBs (0,06) e ECs (0,05), e similares entre os dois últimos grupos.

Analisando a natureza das mutações geradas, observou-se um balanço entre as médias de mutações sinônimas x mutações não sinônimas para o grupo dos VCs (46 sinônimas x 44 não sinônimas). Tal balanço, no entanto não foi observado nas médias para ECs (18 sinônimas x 29 não sinônimas) e ECBs (22 sinônimas x 36 não sinônimas), indicando predominância de substituições de aminoácidos nas posições mutadas. As razões entre mutações sinônimas e mutações não sinônimas foi maior no grupo dos VCs (1,20), porém similar entre ECs (0,70) e ECBs (0,68).

Com base nas frequências relativas de cada SNP, foi observado que variações presentes em frequências acima de 1% correspondiam, respectivamente, a uma média de 35%, 49% e 63% do total de mutações em ECs, ECBs e VCs. No entanto, nesse estrato dos dados, a proporção de mutações sinônimas foi maior do que a de mutações não sinônimas em ECs (Razão sinônimas/não sinônimas de 1,54) e VCs (Razão

Tabela 7. Quantidade de SNPs e Características das Mutações associadas no *gene gag* dos indivíduos estudados

Grupo	Indivíduo	Visita	Total							Maiores que 1%			
			Número de Variantes	Número de Posições	Posições mutadas/ Total	Mutações Sinônimas	Mutações não Sinônimas	Defectivas	Sinônimas/ Não Sinônimas	Mutações Sinônimas	Mutações não Sinônimas	Defectivas	Sinônimas / Não Sinônimas
EC	EC 08	V1	140	69	0,06	19	39	5	0,49	4	25	5	0,16
		VX	110	55	0,05	10	29	14	0,34	2	1	0	2,00
	EC 02	V1	106	53	0,05	20	29	0	0,69	7	4	1	1,75
		VX	98	49	0,04	16	27	2	0,59	8	10	2	0,80
	EC 17	V1	-	-	-	-	-	-	-	-	-	-	-
		VX	92	46	0,04	26	19	0	1,37	15	5	0	3,00
Média			109	54	0,05	18	29	4	0,70	7	9	2	1,54
ECB	ECB 09	V1	-	-	-	-	-	-	-	-	-	-	-
		VX	336	167	0,15	51	97	20	0,53	33	40	1	0,83
	ECB 11	V1	66	33	0,03	12	20	1	0,60	5	5	0	1,00
		VX	50	25	0,02	10	15	0	0,67	1	9	0	0,11
	ECB 13	V1	-	-	-	-	-	-	-	-	-	-	-
		VX	146	73	0,06	31	32	7	0,97	23	26	6	0,88
	ECB 18	V1	100	49	0,04	14	33	2	0,42	8	26	2	0,31
		VX	69	34	0,03	16	18	1	0,89	5	4	0	1,25
Média			128	64	0,06	22	36	5	0,68	13	18	2	0,73
VC	VC 06	V1	52	26	0,02	10	16	0	0,63	3	6	0	0,50
		VX	62	31	0,03	9	20	2	0,45	5	6	6	0,83
	VC 10	V1	187	92	0,08	45	46	1	0,98	26	15	1	1,73
		VX	216	107	0,10	50	53	4	0,94	45	29	4	1,55
	VC 14	V1	346	173	0,15	34	104	32	0,33	13	34	10	0,38
		VX	104	51	0,05	35	17	0	2,06	21	11	0	1,91
	VC 15	V1	181	90	0,08	57	30	5	1,90	49	22	2	2,23
		VX	300	149	0,13	104	39	7	2,67	97	29	6	3,34
	VC 16	V1	250	122	0,11	62	64	0	0,97	45	43	0	1,05
		VX	229	114	0,10	55	49	6	1,12	48	43	6	1,12
Média			193	96	0,08	46	44	6	1,20	35	24	4	1,46

sinônimas/não sinônimas de 1,46). Para ECBs, mutações não sinônimas continuaram mais frequentes que sinônimas ((Razão sinônimas/não sinônimas de 0,73).

No geral, mutações defectivas em *Gag*, gerando Frame-shifts, Stop-Códons ou Extensões, ocorreram na maioria dos indivíduos, inclusive em frequências maiores que 1%. Em comparação com a proporção de mutações não sinônimas e sinônimas, no entanto, os níveis de mutações defectivas foram baixos, exceto para as amostras ECB09VX, EC08VX e VC14V1.

Para o gene *nef* (Tabela 8), a quantidade média de variantes e posições variáveis apresentou diferenças entre os três grupos, sendo maior entre os VCs (179 variantes em 88 posições), seguida pelos ECBs (108 variantes em 54 posições) e menor entre os ECs (66 variantes em 32 posições). A razão entre a quantidade de posições mutadas e o total de posições analisadas também seguiu a mesma tendência, sendo progressivamente maior entre ECs, ECBs e VCs (0,05 x 0,09 x 0,14). Estas razões foram ainda maiores para *Nef*, em comparação com *Gag*, para VCs (0,14 x 0,08) e ECBs (0,09 x 0,06). Para ECs, a razão entre número de variantes e posições mutadas foi igual para *Gag* e *Nef*. Analisando a natureza das mutações geradas, os três grupos apresentaram maior proporção média de mutações não sinônimas em comparação com mutações sinônimas. A razão entre mutações sinônimas não sinônimas foi maior no grupo dos ECBs (0,83), mas próxima a observada para o grupo dos VCs (0,76). Tal razão, no entanto, se mostrou menor no grupo dos ECs (0,52), indicando maior prevalência de mutações não sinônimas.

Com base nas frequências relativas de cada SNP, foi observado que variações presentes em frequências acima de 1% correspondiam, respectivamente, a uma média de 36%, 43% e 62% do total de mutações em ECs, ECBs e VCs. Tais médias foram semelhantes as observadas para o gene *gag*.

Também como em *Gag*, considerando apenas mutações com frequências maiores que 1%, a proporção de mutações sinônimas foi expressivamente maior do que a de mutações não sinônimas em ECs (Razão sinônimas/não sinônimas de 1,05). Em VCs, foi observado um aumento da razão em níveis muito baixos (0,76 para 0,83), enquanto para ECBs foi observada uma diminuição (0,83 para 0,72). As alterações em ECBs e VCs, no entanto foram pequenas e comparáveis entre os dois grupos.

Tabela 8. Quantidade de SNPs e Características das Mutações associadas no gene NEF dos indivíduos estudados

Grupo	Indivíduo	Visita	Total							Frequências maiores que 1%			
			Numero de Variantes	Numero de Posições	Posições mutadas/ Total	Mutações Sinônimas	Mutações não sinônimas	Defectivas	Sinônimas / Não sinônimas	Mutações Sinônimas	Mutações não sinônimas	Defectivas	Sinônimas / Não sinônimas
EC	EC02	V1	66	33	0,05	7	26	0	0,27	3	7	0	0,43
		VX	54	27	0,04	11	16	0	0,69	5	2	0	2,50
	EC08	V1	45	21	0,03	7	14	1	0,50	4	7	0	0,57
		VX	98	48	0,08	18	28	3	0,64	7	10	1	0,70
	Média		66	32	0,05	11	21	1	0,52	5	7	0	1,05
ECB	ECB09	V1	-	-	-	-	-	-	-	-	-	-	-
		VX	233	116	0,19	18	82	15	0,22	8	24	1	0,33
	ECB11	V1	94	47	0,08	16	27	3	0,59	11	19	3	0,58
		VX	54	26	0,04	15	11	0	1,36	7	5	0	1,40
	ECB13	V1	107	52	0,08	19	20	3	0,95	3	6	0	0,50
		VX	174	87	0,14	28	52	6	0,54	14	28	5	0,50
	ECB18	V1	24	12	0,02	6	6	0	1,00	1	2	0	0,50
		VX	73	36	0,06	19	17	1	1,12	10	8	1	1,25
Média		108	54	0,09	17	31	4	0,83	8	13	1	0,72	
VC	VC06	V1	-	-	-	-	-	-	-	-	-	-	-
		VX	128	63	0,10	15	48	2	0,31	3	22	0	0,14
	VC10	V1	60	29	0,05	13	15	3	0,87	10	10	2	1,00
		VX	303	151	0,24	50	85	5	0,59	39	43	1	0,91
	VC14	V1	413	202	0,33	53	120	27	0,44	19	55	5	0,35
		VX	200	99	0,16	41	47	0	0,87	19	22	0	0,86
	VC15	V1	106	50	0,08	22	28	0	0,79	14	20	0	0,70
		VX	189	91	0,15	51	35	6	1,46	41	23	5	1,78
	VC16	V1	107	53	0,09	20	30	3	0,67	17	23	1	0,74
		VX	107	53	0,09	24	27	2	0,89	20	21	1	0,95
Média		179	88	0,14	32	48	5	0,76	20	27	2	0,83	

No geral, mutações defectivas também foram observadas em *Nef*, porém seguindo a mesma tendência de baixa proporção observada em *Gag*. Valores maiores de mutações defectivas também foram detectadas nas amostras ECB09VX e VC14V1, assim como para *Gag*. A amostra EC08VX, no entanto não apresentou a alta quantidade de defeitos como observado para *Gag*.

4.5 IDENTIFICAÇÃO DE MUTAÇÕES NOS EPÍTOPOS RESTRITOS POR CTL

Com base na posição dos epítomos ótimos descritos no *Los Alamos HIV Immunology Database* (LOS ALAMOS NATIONAL LABORATORY, 2015) e restritos pelos alelos HLA de cada indivíduo, mutações possivelmente correlacionadas com pressão seletiva do sistema imune foram identificadas nos genes *Gag* e *Nef*.

Para o gene *NEF*, os indivíduos EC02 e EC08 foram os que apresentaram menor quantidade de mutação. Enquanto o primeiro não apresentou mutações com frequência acima de 1%, o segundo apresentou apenas 3 mutações com mais de 1% de frequência, sendo duas sinônimas (a mais frequente com 5%) e uma não sinônima (V85L – 1%) na Visita VX.

Dentre os indivíduos ECBs, três perfis diferentes foram observados para *Nef*. O indivíduo ECB11 foi o que apresentou menor variabilidade, com apenas duas mutações sinônimas mais expressivas nos códons 72 (CCT->CCC em 11% da VX) e 73 (CAG->CAA em 3,2% da V1).

Os indivíduos ECB09 e ECB18 também apresentaram poucas mutações com frequência maior que 1%, mas algumas mutações foram fixadas na visita VX. Para o indivíduo ECB09, foi observada uma fixação da mutação Q192H (H com 99% de frequência em VX) no epítopo *Nef* RR9 (RLAFHHMAR), restrito por B52, indicando uma possível mutação de escape. As mutações G132W (1,9%), G132V (0,9%) e L137P (3,1%) também foram identificadas na Visita V1 do ECB09, mas não foram detectadas em VX.

Para o indivíduo ECB18, a mutação K92R foi encontrada com frequência de 68,5% na visita VX, porém não foi encontrada na V1. O códon mutado fica na adjacência do epítopo *Nef* KF9 (KAAFDSLFF), restrito por B58, e caracteriza uma troca por aminoácidos

de carga semelhante. Outras mutações encontradas foram L87P (1,2% na V1) e G96R (6,9% na VX)

O indivíduo ECB13 destoou do observado para os outros indivíduos e apresentou um padrão de mutações com indícios da ação de APOBEC3G. As mutações R17K, M20I, D28A, G29K, S126N, F135Y, G140R, assim como 3 mutações sinônimas nas posições 140, 189 e 200, foram associadas com trocas G->A e apresentaram todas frequências próximas a 90% na VX, apesar de não terem sido detectadas na amostra V1. As frequências são semelhantes às observadas para as mudanças defectivas G->A que geraram códons de parada na amostra VX.

Entre os indivíduos VC, a quantidade de mutações foi, no geral, maior do que o observado para ECs e ECBs, assim como a proporção de mutações com frequências maiores que 1%. O indivíduo VC10 foi o que apresentou menor quantidade de mutações e com as frequências mais baixas. Na amostra V1, as maiores frequências foram observadas para as mutações E197L (1,5%) e H199T (2,1%), em regiões adjacentes do epítipo restrito por B52 *Nef* RR9 (RLAFHHMAR), sendo que ambas não foram detectadas na VX. Na VX, no entanto, foi detectada a mutação H192Q (1,5%), observada também para o ECB09, e a mutação G188R (8,4%), representando uma possível emergência de reversão para o aminoácido consenso para o subtipo B nessa posição.

Para o indivíduo VC06, foram detectadas as mutações R21K (14,2%) R21T (74,2%), e M20T (1%) e V30A (1,1%) no interior e adjacências do epítipo *Nef* RA9 (RMRRAEPAA) restrito por B15. Sequências da V1 nessa posição não estavam disponíveis e, por isso, a troca das posições com maior frequência não pôde ser avaliada. Também foi detectada a mutação V194M (13,7%) na adjacência do epítipo *Nef* WF9 (WRFDSRLAF), restrito por B15, caracterizando uma troca de aminoácidos de mesma carga.

Os indivíduos VC14, VC15 e VC16 foram os que demonstraram evidências de maior variabilidade nas regiões de epítipos. O indivíduo VC14 apresentou trocas não sinônimas entre as visitas V1 e VX geradas pelas mutações R71T (4,6% para 97%), P78Q (0,7% para 97,6%), P129Q (5,7% para 97%), V133I (5,2% para 96,5%) nas regiões dos epítipos *Nef* RF9 (RPMTFKGAF) e *Nef* YY9 (YTPGPGVRY), ambos restritos por B42. Similarmente, trocas sinônimas com o mesmo perfil foram localizadas nos códons 74 (GTG para GTA de 5,2% para 97,6%) e 77 (AAG->AGA de 5,2% para 97,6%). As

mutações R71I (0,9%), M79I (2,7%), Q125H (1,25%), foram detectadas apenas na V1. Uma mutação sinônima no códon 80 também foi identificado em frequência de 40% da VX (ACT->ACG), indicando uma possível troca em andamento. Outras mutações sinônimas foram detectadas nas duas visitas, mas com pouca variação nas frequências relativas nos códons 138 (TGT->TGC de 4,1% para 2,6%) e 131 (CCA->CCC de 0,9% para 1,2%).

Para o indivíduo VC15 houve maior prevalência de mutações sinônimas nos epítomos *Nef* KF9 (KAAFDLSFF), *Nef* HW9 (HTQGYFPDW), *Nef* YY9 (YTPGPGIRY), *Nef* YF9 (YPLTFGWCF), todos restritos por B57. A diminuição de frequência entre uma visita ou outra foi observada nos códons 80 (ACU->ACC de 37,1% para 6,3%), 119 (GGC->GGT de 67,5% para 54,9%), 133 (CCT->CCC de 76% para 47%) e 137 (CUG->TUG de 40,6% para 12,1%), indicando uma troca de prevalência de população. No codon 126, uma mutação AAC->AAT foi estimada com uma frequência de 40% em ambas as visitas. No códon 85, foi detectada a mutação V85L passando de frequência de 66% na V1 para 99% na VX. Exceto a última, apenas mais uma mutação não sinônima com frequência maior que 1% foi identificada (Q125R com 1,7% de frequência em VX).

Para o indivíduo VC16, uma alta quantidade de mutações não sinônimas foi observada para o mesmo conjunto de epítomos restritos por B57 citados acima. As mutações K92R (19%), D111G (1,2%), Y135H (14%) foram observadas apenas na V1, indicando os momentos finais de uma troca de predominância nessas posições. Para o códon 81, foi observada a mutação Y81F variando de 54,3% para 47,6% entre V1 e VX, o que pode indicar uma possível reversão em andamento para a Tirosina selvagem na posição. As mutações H89Y (29%), Y102H (48%) e T128I (13,3%), bem como mutações sinônimas nos códons 135 (CCA->CCG em 15,3%) e 136 (CCA->CCT em 11%), foram detectadas apenas em VX, mas as altas frequências evidenciam uma possível troca de predominância nos próximos anos. Por fim, as posições 105 (AAA->AAG em 32%) e 116 (N116H em 37%) tiveram mutações em frequências similares em V1 e VX, indicando ausência de pressão seletiva nessas posições ao longo do intervalo V1-VX.

Para o gene *gag* os padrões de diversidade observado entre grupos foi semelhante ao observado para *Nef*, com o menor número de mutações e menores frequências de variação nos indivíduos ECs. No entanto, as mutações em regiões de epítomos foram

observadas em maior quantidade do que visto para *Nef*, indicando uma possível maior pressão seletiva no gene *gag*.

Entre os indivíduos ECs, o EC02 foi, novamente, o que apresentou aparentemente a menor diversidade. Apenas a mutação N137S (3% em V1), na adjacência do epítipo *Gag* GI8 (GQMVHQAI) restrito por B48, apresentou maior representatividade entre as mutações não sinônimas identificadas. Dentre as mutações sinônimas, no entanto, três foram detectadas com maiores frequências nos códons 139(CAG->CAA com 30% em V1 e VX), 275 (AGA->AGG em 23,8% de VX) e 278 (AGC->TGC em 22,8% de VX), sendo as duas últimas referentes ao epítipo *Gag* RI8 (RMYSPTSI), restrito por B52.

Para o indivíduo EC08, as mutações G248E (1%), D295N(1,3%), D298G (2%), D298N (1,7%), bem como mutações sinônimas nos códons 308 (1,1%) e 33 (2,2%), foram as únicas mais expressivas, sendo detectadas apenas na V1 e em epítipos restritos por B57. Para o indivíduo EC17, por fim, apenas mutações sinônimas nos códons 195 (CAA->CAG em 3,3% de VX) e 197 (GCC para GCT em 56,5% na VX), no epítipo *Gag* GL9 (GHQAAMQML) restrito por B15, alcançaram frequências mais expressivas.

Os ECBS, assim como observado para *Nef*, apresentou um perfil mutacional variado. O indivíduo ECB11 foi novamente o que apresentou perfil mais conservado com apenas uma mutação expressiva, caracterizada por uma troca de AGG para AGA no códon 294, identificada com uma proporção de 48,2% na V1 e 99% da VX.

Nas posições referentes dos epítipos *Gag* RY11 (RSLYNTVATLY), *Gag* IW9 (ISPRTLNAW), *Gag* II11 (IEEKAFSPEVI) e *Gag* RL11 (RDYVDRFYKTL), o indivíduo ECB18 apresentou as mutações R76K (20%), T84A (3%), M147I (17,2%), E161K (18,8%), D295N (25,3%), D295G (4,5%), D213N (25%) e L215M (13,8%), além de uma mutação sinônima no códon 181 (CCA->CCG em 15,5%), as quais foram observadas apenas na V1, não sendo detectadas na VX. Outras 3 mutações, sinônimas nos códons 163, 203 e 298, apresentaram proporções entre 13 e 20% durante a V1 e, ao contrário das anteriores, foram fixadas na VX com mais de 99% de frequência. Além dessas, uma mutação L215M foi detectada com 13,8% na V1, se fixando posteriormente na VX com 99% de frequência.

O indivíduo ECB13, por sua vez, foi o ECB que apresentou perfil de maior diversidade, embora as mutações observadas não apresentassem aparente ação de

APOBEC3G, como visto para *Nef*. As mutações K26N (6,3% em VX), K28R (25,3% em VX), G24E (6,1% em VX), Q30R (25% em VX) G192R (10% em VX) e Q311R (1% em VX), bem como uma mutação sinônima no códon 301 (TAT->TAC em 20,3% de VX), foram observadas com maiores frequências na VX. Os dados da V1 não foram utilizados para comparação devido aos problemas no mapeamento já citados.

Entre os indivíduos VC, por sua vez, o perfil de maior diversidade, em comparação com os outros grupos, também foi observado para *Gag*. Para este gene, o indivíduo VC06 foi o que apresentou perfil de menor diversidade, sendo observadas com maiores frequências apenas as mutações Q136R (1,3% em VX), F300L (3,2% na V1), E307G (1,4% em VX) e E319G (2,5% em VX), nas adjacências dos epítomos *Gag* GI8 (GQMVHQAI restrito por B48) e *Gag* RL11 (RDYVDRFFKTL, restrito por B15), e no interior do epítopo *Gag* VL9 (VKNWMTETL, restrito por B48). Mutações sinônimas nos códons 141 (CAA->CAG em 6,7% de VX) e 194 (CAT->CAC em 5,4% de V1) também foram observadas, mas só foram registradas em uma das visitas.

A segunda menor diversidade foi observada para o indivíduo VC10, o qual também apresentou perfil mais conservado em *Nef*. As mutações M142V (3,1% em VX), G192E (1% em VX), M276I (1,3% em VX), bem como mutações sinônimas nos códons 158 (GTA->GTG em 1,5% de VX), 203 (GAG->GAA em 7,8% de VX e GAG->GAT em 1,5% de VX) foram detectadas em baixas frequências apenas na visita VX).

Mudanças na representatividade entre as duas visitas foram observadas para as mutações não sinônimas S146P (40% para 37%), I280S (44,5%- para 66,9%), I280T (1,2% para 13,5%) e R268K (1,2% para 9,2%). O mesmo padrão foi observado para mutações sinônimas nos códons 144 (38% para 22%), 154 (1% para 6%), 157 (1,5% para 6,3%), 159 (0,6% para 1,4%) e 278 (2,3% para 9%).

Para o indivíduo VC14, *Gag* apresentou menor diversidade do que foi observado para *Nef* do mesmo indivíduo. A visita X apresentou mutações sinônimas nos códons 22, 78, 153 e 184 com frequências próximas a 20%. Com proporções semelhantes, foram encontradas as mutações não sinônimas I82V e T306S. Além dessas, as mutações M31I (1,3% em V1), Y87C (2,5% em VX), P177T (1,1% em V1), D291Y (1% em V1), D291G (1% em VX) e M313I (1,8% em V1), foram encontradas em baixas frequências em apenas uma das visitas.

Para o indivíduo VC15, 90% das mutações com grande representatividade encontradas caracterizaram trocas sinônimas de códon. Apenas as mutações T242N (22% em VX) e P328L (4% em VX) geraram substituições aminoacídicas.

Para o indivíduo VC16, novamente foi encontrada a maior diversidade, com grande quantidade de mutações sinônimas e não sinônimas. As mutações I34L (35% em V1), V35I (35% em V1), V46I (30% em V1), V82L (25% em V1), A115E (25% em V1), A118P (25% em V1), A119T (25% em V1), S125R (30% em V1), P146A (22% em V1), N242T (21% em V1), I247B (31,6% em V1), N252S (32,3% em V1), S310T (17% em V1) foram observadas apenas na V1, mas não caracterizaram fixação do aminoácido menos prevalente na visita VX. As mutações R39K (5,6% em VX), S111G (1,4% em VX), E160K (19,3% em VX), E307K (10,8% em VX) e E312K (12% em VX) foram observadas apenas na visita VX e podem caracterizar novas variantes de escape. As mutações A122T (4,8% para 72%), N124S (0,7% para 35,5%), V159I (22 para 9,3%), G163A (21 para 25%), N165S (21 para 25%), S173T (37,7 para 74,3%) e K335R (49% para 90%) caracterizaram trocas de representatividade entre as duas visitas.

5 DISCUSSÃO

Pacientes HIV-positivos com perfil de controle da progressão e da viremia são uma parcela rara da população e os mecanismos responsáveis por esse fenótipo ainda não foram completamente elucidados. Assim como pacientes em tratamento com HAART, muitos destes indivíduos conseguem manter a carga viral em níveis abaixo do limite de detecção (<50 cópias /ml) na grande maioria dos testes clínicos disponíveis. No entanto, a detecção de DNA e RNA viral, utilizando testes ultrasensíveis, mesmo após décadas de infecção demonstra que estes pacientes apresentam ao menos níveis basais de replicação (SÁEZ-CIRIÓN; PANCINO, 2013). Estes níveis basais, por sua vez, devem permitir algum nível de evolução viral, devido ao alto potencial de variabilidade genética do HIV.

Paralelamente, os diversos estudos de caracterização imunológica de controladores têm demonstrado que a resposta CTL nestes indivíduos é diferenciada (GAARDBO et al., 2013; GAARDBO; HARTLING, 2012). Tal resposta citotóxica mais efetiva, por sua vez, deveria refletir numa maior pressão seletiva na população viral desses indivíduos controladores, principalmente em regiões imunodominantes. Dessa forma, a pressão seletiva da resposta CTL atuaria como uma força contínua ao longo de toda a infecção, influenciaria o surgimento de mutações de escape da resposta imune, principalmente nos pacientes com alelos HLA relacionados a não progressão clínica.

Com base nisso, o presente estudo teve como foco o estudo da evolução viral nos genes *Gag* e *Nef*, os quais possuem uma alta concentração de regiões imunodominantes, em uma coorte de 12 pacientes com perfil de não-progressão de longo termo e controle variado da replicação viral. Para tal, foram utilizadas amostras celulares de duas visitas num intervalo médio de 47 meses para obtenção de DNA e sequenciamento dos genes *Gag* e *Nef* através da plataforma de NGS Illumina HiSeq.

Nos últimos anos, o advento do sequenciamento de nova geração possibilitou que a caracterização das variantes virais componentes do conjunto de quasiespécies em pacientes HIV positivos fosse cada vez mais precisa, sendo possível detectar variantes minoritárias que passam despercebidas através de sequenciamento de Sanger (CHABRIA; GUPTA; KOZAL, 2014; GALL et al., 2012; HENN et al., 2012). Devido a sua maior robustez, esta técnica tem sido amplamente utilizada e sua maior eficácia

constatada em diferentes estudos de detecção e dinâmica de emergência de mutações de resistência do HIV-1 a antirretrovirais em pacientes sob HAART (RAM et al., 2015; ROZERA et al., 2014) e de variantes de escape da resposta imune de HIV-1 e SIV, durante a fase aguda e crônica da infecção (BIMBER et al., 2009; DENG et al., 2015; FISCHER et al., 2010; HENN et al., 2012; LOVE et al., 2010) .

Apesar da maior eficiência metodológica das técnicas de NGS na detecção de variantes minoritárias, não existem, até a data, estudos publicados que utilizem esta modalidade de sequenciamento para análise da diversidade viral e da emergência de mutantes de escape em pacientes controladores de HIV. Na pesquisa bibliográfica, os estudos mais próximos encontrados realizaram as análises em primatas infectados por SIV e com perfil de controle da progressão (HUGHES et al., 2012; MUDD et al., 2012).

Apesar do número relativamente baixo de pacientes envolvidos no presente estudo, todos os pacientes selecionados apresentam um perfil de controle do HIV bem definido e caracterizado. Cabe ressaltar que estes LTNPs controladores representam <de 1% dos indivíduos infectados e que é necessário um longo tempo de acompanhamento da infecção, em cada indivíduo, para a correta classificação dos mesmos. Outro fator complicador é que os critérios utilizados para classificação desses divergem muito entre os diversos grupos de pesquisa, dificultando a comparação de resultados em alguns casos (GAARDBO; HARTLING, 2012). Por exemplo, a coorte francesa ANRS CO21 classifica controladores como pacientes sem terapia antirretroviral, com mais de 5 anos de infecção e carga viral abaixo de 400 cópias/ml ao menos nas últimas 5 quantificações (SÁEZ-CIRIÓN; PANCINO, 2013), agrupando, de acordo com os parâmetros do presente estudo, pacientes EC e VC. A coorte do Estudo de História Natural do HIV do Departamento de Defesa dos EUA, por sua vez, classifica VCs e ECs como pacientes com ao menos um ano de acompanhamento e com as últimas 3 mensurações de carga viral abaixo de 2000 ou 50 cópias, respectivamente (OKULICZ et al., 2009).

No presente estudo, os indivíduos incluídos possuem uma mediana de 15 anos de supressão conhecida e fazem parte de uma coorte que tem sido acompanhado pelo nosso grupo pelos últimos 15 anos. Com base nessa coorte e incluindo dados de grande parte dos indivíduos deste estudo, nosso grupo já foi capaz de verificar a limitada diversidade viral desses controladores (BELLO et al., 2007; CAETANO et al., 2014), assim como uma reduzida

resposta de anticorpos e de ativação de linfócitos T quando comparados a indivíduos não controladores (BELLO et al., 2009). Além destes marcadores, já verificamos a presença de SNPs no gene da LEDGF no grupo de LTNP, (PASSAES et al., 2014) além de reduzida carga proviral, ausência de DNA não integrado e redução da ativação de células T e da resposta celular a alvos antigênicos (CÔRTEZ et al., 2015), além da presença aumentada de alelos protetores da progressão para a aids (TEIXEIRA et al., 2014). Em relação ao vírus, a análise do envelope viral demonstrou uma baixa diversidade tanto em relação às quasiespécies identificadas num determinado momento, como na comparação destas ao longo do tempo, além da evidência da presença de hipermutação G->A nas amostras analisadas (CAETANO, 2013).

Com a identificação de que o início precoce da terapia gera melhores resposta imunológicas (DING et al., 2015; OKULICZ et al., 2015), a HAART tem sido oferecida cada vez mais cedo a indivíduos infectados. No Brasil, por exemplo, a HAART já é oferecida a todos os indivíduos infectados independente dos níveis de CD4 desde o final de 2013 (MINISTÉRIO DA SAÚDE, 2014b). Estas medidas permitirão um melhor manejo da infecção pelo HIV no país, culminando em menores taxas de transmissão, mas impossibilitarão a identificação de novos indivíduos com perfil de controle espontâneo. Além disso, tais medidas ainda tendem a diminuir as coortes já existentes, devido ao oferecimento da terapia principalmente a indivíduos como os VCs, que apresentam cargas virais detectáveis, mesmo que baixas, permitindo um controle máximo da viremia. Para indivíduos ECs e ECBs, o início do tratamento se mostra talvez menos necessário, devido a carga viral indetectável representativa deste grupo, porém pode estar associado a diminuição dos níveis de reservatório viral (CHUN et al., 2013).

5.1. A EVOLUÇÃO VIRAL NOS INDIVÍDUOS CONTROLADORES

Dentre os estudos realizados anteriormente pelo nosso grupo, um estudo de evolução intra-hospedeiro do gene do envelope viral do HIV-1 conduzido em colaboração com o Laboratório do Dr Cecílio Lopez Galindez demonstrou que alguns indivíduos EC conseguiriam bloquear em grande parte a evolução do vírus, confirmando a hipótese da existência de uma verdadeira latência evolutiva do HIV nesses indivíduos (BELLO et al., 2007). Nos últimos anos, este estudo tem ampliado com adição de novos indivíduos em

diferentes perfis de controle da viremia e atualização dos dados de pacientes antigos, demonstrando valores de diversidade e divergência até dez vezes menores que o esperado em ECs (CAETANO et al., 2014; CAETANO, 2013). O conjunto de nossos resultados então tem demonstrado que a evolução viral é observável pelo menos no gene do envelope viral de indivíduos controladores (Dados não publicados).

No presente estudo a utilização da plataforma Illumina HiSeq permitiu a obtenção de dados que possibilitam uma profundidade única na análise de quasiespécies, a qual não pode ser alcançada com o uso de metodologias de diluição (SGS) ou clonagem. O sequenciamento realizado, após pré-processamento e mapeamento dos genes, rendeu uma cobertura média superior a 219.000 sequências e 291.000 sequências por par de base para *Gag* e *Nef*, respectivamente. Cale e colaboradores, em 2011, utilizaram dados obtidos com a plataforma 454 e, através de modelo binomial, argumentaram que uma cobertura total de 50.000 sequências era o suficiente para detectar variantes com frequência de 0,006% com pvalor significativo. Isto demonstra a grande eficácia da metodologia utilizada no presente estudo.

Tentativas de reconstrução das sequências referentes as quasiespécies de cada paciente em cada ponto analisado foram realizadas com diferentes ferramentas de bioinformática, mas o grande número de sequências e a necessidade de grande poder de processamento computacional necessário para sua análise impediram a obtenção de resultados até o presente momento. Devido a essa impossibilidade, a análise de evolução viral foi realizada com base na distância genética entre sequências consenso das visitas V1 e VX de cada indivíduo e no número de variantes encontradas em cada posição foi mapeada. Para esta última análise, apenas variantes encontradas com frequências maiores que 0,5% foram consideradas, levando em conta os dados obtidos pela avaliação da variação encontrada no sequenciamento do controle positivo contendo DNA proveniente da linhagem viral ACH2.

As análises de diversidade demonstraram um baixo nível de variação ao longo do tempo tanto no gene *gag* quanto no gene *nef*. As análises de divergência entre as sequências das duas visitas de cada paciente demonstraram, valores médios próximos a 0,5% para o gene *gag* de ECs e VCs, indicando a maior conservação deste gene. Gijssbers e colaboradores, em 2015, realizaram estimativas de divergência em pacientes

na fase crônica da infecção e obtiveram valores entre 0,9 e 1,9% entre sequências obtidas em intervalos de cerca de 10 meses, o que corrobora menores níveis de evolução viral de nossa coorte. Para o gene *nef*, no entanto, a média de divergência se manteve próxima a 0,5% para ECs, mas foi maior em VCs (2,5%). Para o grupo dos ECBs, os valores médios observados foram altos para os dois genes, mas tal aumento foi decorrente de eventos de hipermutação ou possíveis associações com superinfecção.

O indivíduo ECB09, por exemplo, apresentou variantes virais com grande distância genética entre si as quais podem estar associados, de acordo com resultados obtidos paralelamente pro gene do envelope viral, com indícios de uma superinfecção na visita X, definida como a presença de variantes virais filogeneticamente distintas surgidas em períodos distintos (Redd et al., 2013). As análises do mapeamento do gene *gag* demonstraram a presença de sequências muito divergentes entre si e cuja identidade não foi compatível com nenhuma das sequências dos outros indivíduos do mesmo estudo, descartando contaminação, embora não observada em *Nef*. A identificação de duplas variantes também já foi relatada em outros estudos de evolução viral com LTNPs (Bello et al., 2005; Casado et al., 2007; Clerc et al., 2010; Mens et al., 2010; Rachinger et al., 2008). Os efeitos na patogênese do HIV em pacientes infectados por duas linhagens ainda não estão claros, mas alguns estudos revelaram que, em alguns casos, a super ou dupla infecção pode ser responsável por maiores quedas de linfócitos TCD4⁺ e maiores cargas virais (Redd et al., 2013; Smith et al., 2005). Em pacientes LTNPs, os efeitos patogênicos também não são conclusivos, visto que estudos já associaram esse estado à perda (CLERC et al., 2010), manutenção (CASADO et al., 2007) ou retomada (RACHINGER et al., 2008) do controle da progressão. No indivíduo ECB09, a superinfecção pode estar associada com aumentos da viremia observados desde 2012 (de níveis indetectáveis para um pico de 581 cópias/ml no final de 2015), o que motivou inclusive o início do tratamento deste indivíduo com antirretrovirais ao final deste ano, de acordo com a nova política de oferecimento de tratamento independente dos níveis de CD4.

Os indivíduos ECB13 e ECB18, por sua vez, apresentaram indícios de hipermutação mediado por proteínas APOBEC. Essas proteínas são responsáveis por gerar mutações G->A no HIV antes de sua integração, resultando em provírus que se

integram, mas podem apresentar defeitos que impedem a produção de partículas virais competentes (HARRIS; LIDDAMENT, 2004). Para o indivíduo ECB13 várias mutações TGG->TGA foram observadas em *Gag* com frequências acima de 80% na visita VX, culminando na geração de códons de parada ao longo de vários pontos do gene. Para o indivíduo ECB18, uma mutação majoritária do mesmo tipo foi observada também no códon 215 do mesmo gene. Tal achado não é tão inesperado visto que os últimos estudos tem demonstrado a influência de APOBEC3G na composição de reservatórios virais não funcionais (HO et al., 2013) e detectado que pacientes controladores tendem a apresentar uma grande proporção de vírus hipermutados (EYZAGUIRRE et al., 2013; FOURATI et al., 2014). Nos dois indivíduos citados anteriormente, o aparecimento de tais mutações na visita VX podem ter sido resultado de novos ciclos de replicação, ocorridos em alguns dos Blips, ou da diminuição de representatividade de reservatórios funcionais. Uma vez que os reservatórios induzíveis e competentes para a replicação vão decaindo, isso pode facilitar o ressurgimento e aumento da predominância de reservatórios defectivos arquivados (HO et al., 2013). Análises posteriores mais aprofundadas dos dados desses indivíduos, assim como da frequência de hipermutação nos demais indivíduos desse estudo, possibilitarão melhores resoluções acerca da influência de APOBEC nesse estudo.

Os padrões de variabilidade de cada gene foram melhor identificados através da contagem de SNPs e mutações. Uma maior conservação do gene *gag*, em comparação com *Nef*, pode ser constatada por menores razões entre o número de mutações e o número de posições variáveis em *Gag*. Este resultado é compatível com as menores taxas de evolução em *Gag* estimada por alguns estudos (STARCICH et al., 1986; YEBRA; BROWN, 2015) e as observações feitas em estudos de emergência de mutações de escape (BRUMME et al., 2008b; RADEBE et al., 2015b).

Tal diferença na variabilidade de cada gene só não foi bem observada para indivíduos ECs, os quais apresentaram razões de mutações por posições variáveis iguais para *Gag* e *Nef*, demonstrando a maior latência evolutiva nesse grupo. Indivíduos ECBs apresentaram valores da mesma razão muito similares aos de ECs em *Gag*, porém maiores em *Nef*, indicando uma maior taxa de evolução da população viral nesse grupo.

Indivíduos VCs por sua vez apresentaram, para os dois genes analisados, as maiores quantidades de posições variáveis e de mutações dentre os três grupos.

Analisando a natureza das mutações provocadas pelos SNPs identificados (Sinônimas X Não sinônimas), foi notável uma maior quantidade de mutações não sinônimas, principalmente no gene *nef*. Para este gene as razões entre mutações sinônimas e não sinônimas foi menor que 1 em todos os três grupos analisados. Estes valores mais baixos se mantiveram para ECBs e VCs mesmo quando mutações com frequências menores que 1% foram desconsideradas. Para ECs, essa mesma tendência se manteve, mas a razão média para o grupo foi desviada por altos valores da visita VX do indivíduo EC02. Para o gene *gag*, o número de mutações não sinônimas também foi alto, mas essas não foram as mutações mais prevalentes. Indivíduos VCs apresentaram a prevalência de mutações sinônimas tanto no total, quanto somente entre as mutações com frequências maiores que 1%. Indivíduos ECs e ECBs apresentaram razão média de mutações sinônimas/não sinônimas menores que 1 considerando todas as mutações, mas esta razão aumentou para cerca de 1,5 para os ECs.

No conjunto estes dados demonstraram a maior conservação do gene *gag* em comparação com o gene *nef* nestes indivíduos controladores, assim como observado nos indivíduos em geral (STARCICH et al., 1986; YEBRA; BROWN, 2015). Apesar do surgimento de mutações observada em *Gag*, a maior parte das mutações fixadas na população proviral são mutações sinônimas, indicando que esse gene aceita menor variação genética em comparação com *Nef*.

Estudos realizados com indivíduos controladores indicaram que, apesar da evolução viral ocorrer, ela estaria mais confinada aos vírus plasmáticos, visto que sequências provirais apresentavam pouca ou nenhuma variação (BAILEY et al., 2009; O'CONNELL et al., 2010; SALGADO et al., 2010). No presente estudo, a utilização do NGS e maior profundidade de análise resultante nos permitiu identificar que embora menos visível pelas metodologias mais usuais, a evolução no compartimento proviral também ocorre efetivamente.

5.2. EVOLUÇÃO EM EPÍTOPOS CTL E O ESCAPE VIRAL

De modo a caracterizar possíveis mutantes de escape e a pressão seletiva da resposta imune, realizamos a análise da variação dentro dos epítomos restritos pelas moléculas HLA-B de cada paciente. Mutações de escape podem ser caracterizadas por alterações de aminoácidos nas regiões centrais dos epítomos CTL (gerando menor reconhecimento por moléculas TCR), nos aminoácidos da ponta dos epítomos (gerando alterações no ponto de ancoragem do epítomo) ou em regiões adjacentes ao epítomo (prejudicando o processamento antigênico) (Revisto por Jonathan M. Carlson, Le, Shahid, & Brumme, 2015). Apesar dessas mutações serem relacionadas a custos no fitness viral e na capacidade replicativa do vírus, elas ocorrem em grande quantidade, demonstrando o quão efetiva é a resposta CTL em exercer pressão seletiva nessas regiões (MIGUELES et al., 2003). Prova disso é a existência de mutações comuns a indivíduos com o mesmo perfil de HLA, o que gera padrões de emergência de mutantes em nível populacional (CARLSON; BRUMME, 2008; DILERNIA et al., 2008; JUAREZ-MOLINA et al., 2014). Considerando ainda a existência de alelos protetivos contra o HIV, estudos tem demonstrado que essa pressão seletiva é ainda maior em indivíduos manifestando estes genótipos, favorecendo o surgimento mais rápido e de maior quantidade de mutantes (ROBERTS et al., 2015) e de mutações em regiões de âncora ou multisítios (CARLSON et al., 2012). Estas últimas são importantes visto que o escape da resposta CTL ocorre, na maioria dos casos, não pelo surgimento de mutações pontuais, mas do acúmulo de mutações em diversas posições (MUDD et al., 2012). Nos últimos anos este tipo de análise tem sido extensivamente utilizada para caracterizar as dinâmicas de emergência de mutações de escape em indivíduos ainda na fase aguda. Com isso, estudos conseguiram identificar que a maior parte das mutações indentificadas nas primeiras semanas da fase aguda são resultado da pressão seletiva da resposta CTL (BRUMME et al., 2008a; FISCHER et al., 2010; GOUNDER et al., 2015; HENN et al., 2012; RADEBE et al., 2015a). Analisando a resposta CTL nos diferentes genes, foi possível ainda estimar que a resposta imune direcionada a *Nef* nos primeiros momentos da infecção corresponde a 40% do total, passando a ser majoritariamente direcionada a *Gag* após um ano (RADEBE et al., 2015a). Funcionalmente isso se correlaciona com observações de que 60% do total de mutações surgidas em *NEF* no primeiro ano de

infecção podem ser relacionadas com a resposta CTL. Para *Gag* esse percentual corresponde a 30%.(BRUMME et al., 2008a).

Em indivíduos com perfil de controladores, são poucos os estudos publicados que verificaram a emergência de mutantes de escape e a pressão seletiva em *Gag* e *Nef*. Migueles e colaboradores analisaram *Gag* de indivíduos CEs e VCs que possuíam o alelo HLA B*57, e verificaram poucos mutantes de escape no epítipo KF11 de indivíduos controladores, apesar de grande reconhecimento da resposta CTL nesse ponto, indicando que mutações no epítipo poderiam ser deletérias (MIGUELES et al., 2003). Mens e colaboradores avaliaram sequencias de *Env* e *Pol* de controladores e não-controladores, observando mutações em cerca de 10 a 20% dos epítipos já descritos para cada um dos genes (MENS et al., 2010).

Um grupo da Universidade de medicina Jonh Hopkins, por sua vez, publicou uma série de estudos em que descreve a evolução dos genes *Gag* e *Nef* em indivíduos controladores de elite com genótipo HLA B*57. Nestes, os dois genes foram amplificados e sequenciados por SGS a partir de PBMCs e Plasma, possibilitando observar que as mutações de escape CTL eram majoritárias nas sequências de *Gag* dos vírus plasmaticas, mas não encontradas no compartimento proviral. Em *Nef*, a pressão seletiva observada foi menor, com a mesma discordância entre sequências de plasma e provirus observada para este gene, embora em menor nível. Além disso, o grupo também observou que muitas das respostas para epítipos selvagem suprimiam os vírus, observando inclusive casos em que as respostas geradas por mutantes de escape era maior que a gerada pela sequência selvagem, favorecendo uma maior supressão viral (BAILEY et al., 2006, 2009; O'CONNELL et al., 2010; SALGADO et al., 2010). O mesmo grupo ainda demonstrou mais tarde um caso de um indivíduo EC que apresentou diversos escapes em epítipos restritos por HLA B*57 e conseguiu montar rapidamente novas respostas CTL específicas contra os epítipos mutantes, indicando um alto nível de adaptabilidade da resposta imune nesse paciente (O'CONNELL et al., 2011). Os resultados vão de acordo com outro estudo que utilizou pseudovírus com mutações nos epítipos IW9 (*Gag* 147-155), KF11 (*Gag* 162-172) e TW10 (*Gag* 240-249) para demonstrar que células com genótipo B*57 eram capazes de suprimir todos os vírus mutantes (POHLMAYER et al., 2013).

No presente estudo, mutações não sinonimas e sinônimas em regiões de epitopos CTL foram encontradas nos genes *Gag* e *Nef* em todos os indivíduos, apesar de a análise ter sido realizada no compartimento proviral. A ocorrência de mutações em baixas frequências únicas para cada uma das visitas de um mesmo paciente demonstram que, ao contrario do argumentado por Bailey e colaboradores (BAILEY et al., 2006), possíveis mutantes gerados pela resposta CTL replicando no plasma conseguem obter sucesso em se integrar a células hospedeiras. Embora tais variantes de baixa frequência possam não parecer significantes, é preciso ressaltar que novas variantes virais surgem como resultado dos processos de variabilidade genética inerentes ao vírus e necessitam obter sucesso em todas as etapas do ciclo celular para se firmarem no pool de quasiespécies. Tal processo é regido pelas diversas pressões seletivas atuando na população viral e indicam que, na grande maioria dos casos, novas variantes não surgem com frequências massivas, mesmo que o seu fitness seja muito maior que a de outras populações majoritárias. Cale e Colaboradores exemplificam tal conceito ao demonstrarem a detecção, através de sequenciamento na plataforma 454, de mutações não sinônimas com frequências de 0,1%, associadas com o escape CTL, que se tornam majoritárias ao longo do intervalo de alguns meses (CALE et al., 2011). Isto foi observado no nosso estudo por exemplo, para as mutações P78Q em *Nef* do indivíduo VC14 (0,7% em V1 para 97% em VX) e K198L (1,3% em V1 para 99% em VX) em *Nef* do indivíduo ECB13, além de outras.

Paralelamente, embora grande parte das mutações de baixa frequência pareça ser eliminada com o tempo, outras demonstram não sofrer aparente seleção, ficando aparentemente arquivadas nos reservatórios. Tal teoria pode ser exemplificada por várias mutações encontradas em ambas as visitas de nosso estudo e que não sofreram grandes alterações de frequência como, por exemplo, a mutação M194V (0,7% em V1 e 0,5% em VX) em *Nef* do indivíduo VC10.

Com base nas mutações encontradas, os indivíduos ECs foram os que demonstraram menor grau de variação nas regiões de epítomos dos dois genes, apesar de mutações não sinônimas terem sido observadas em baixas frequências nos indivíduos EC08 e EC17. Entre as duas visitas observadas, todos os indivíduos ECs apresentaram

conservação da sequência proteica nos epítomos, sendo observado em maiores frequências apenas mutações sinônimas.

Indivíduos ECB apresentaram o perfil mais diversificado também na análise isolada dos epítomos. O Indivíduo ECB11 apresentou um perfil muito semelhante ao dos ECs, com muito poucas mutações sinônimas em maior frequência em *Gag* e *Nef*. Este perfil não é tão inesperado levando-se em conta que, tal paciente, apresenta apenas um Blip de carga viral desde o início do seu acompanhamento em 2002.

Para o indivíduo ECB13, uma alta frequência de mutações G->A indicativa da ação de APOBEC3G também foi observada nas regiões de epítomos de *Nef*, gerando mutações não sinônimas que não puderam ser associadas diretamente a pressão seletiva do sistema imune. Para *Gag* uma alta variabilidade foi observada no epítopo RL10 restrito por B*51, indicada pela presença das mutações G24E (6,1%), K26S (93,6%), K26N (6,3%), K28R (25,5%), K30Q (24,4%) e K30R (74,7%). É curioso observar que as posições 26 e 30 foram associadas a uma mutação de maior frequência e outra de menor frequência, gerando a substituição do aminoácido selvagem para o subtipo B nessas posições (K). A ausência da sequência referente a V1 nessa posição não nos permite indicar qual era o aminoácido predominante em tal ponto no tempo, mas o padrão observado na VX é indicativo de que as mutações de menor frequência possam ser um escape secundário emergindo no mesmo aminoácido. Nenhum dos outros epítomos analisados nesse indivíduo mostraram dados semelhantes, indicando que a pressão seletiva nessa região é aparentemente maior e poderia estar associada a restrição pelo alelo B*51, facilitando o surgimento de mutações de escape.

Os indivíduos ECB09 e ECB18, por sua vez, mostraram poucos indicativos de variabilidade nas regiões de epítomos. Embora a análise de *Gag* para o indivíduo ECB09 não tenha sido feita devido aos problemas de possível superinfecção já discutidos, o gene *nef* apresentou uma troca Q->K na posição 192 correspondente ao epítopo AL9, restrito por B52. Chama a atenção também a presença das mutações V133T, Y135F e Y143F associadas ao epítopo YF9, restrito por B*52, as quais representam diferenças marcantes entre o vírus desse indivíduo e o consenso para o subtipo B. Curiosamente, as mutações Y->F descritas ocorrem ambas nas posições referentes as âncoras dos epítomos e podem estar associadas a um escape mediado por B52. Associado as outras mutações

observadas nos indivíduos com este mesmo genótipo, tais dados podem ser um indicativo do efeito protetivo do HLA B*52 descrito por outros estudos (MURAKOSHI et al., 2015; TEIXEIRA et al., 2014)

O indivíduo ECB18, por sua vez, não apresentou mudanças nas regiões de *Gag* e *Nef* entre as duas visitas analisadas. Para *Gag*, no entanto, uma série de mutações em frequências entre 10 e 25% na V1 foram encontradas em quase todos os epítomos analisados. Nenhuma das mutações se fixou na visita VX e podem estar associadas com a pressão imune por B*44 e B*58 ou a algum Blip anterior a V1 que não foi detectada nos exames regulares.

Os indivíduos VC, por sua vez, demonstraram o maior grau de variabilidade nas regiões de epítomos, com mutações sugestivas de escape imune em todos os pacientes. O indivíduo VC06 apresentou as mutações R21T (74% em VX) no epítomo *Nef*-RA9 e a mutação I114V na adjacência do epítomo *Nef*-TQ9, ambos restritos por B*15. O indivíduo VC10 apresentou a mutação S146P no epítomo *Nef*-VL10 em cerca de 15% das sequências nas duas visitas, indicando a co-circulação de variantes com os dois aminoácidos. Para esse mesmo indivíduo foram observadas as mutações I280S (44,5% em V1 para 66,9% em VX) e I280T (1,2% em V1 para 9,2% em VX) no epítomo *Nef*-RI9 o que pode ser um indicativo de uma maior pressão seletiva exercida por B*52 nesse códon. Em *Gag*, foi observado ainda uma mutação G188R em 8,4% das sequências da visita VX que caracteriza uma reversão em uma das âncoras do epítomo *Gag*-RR9 restrito por B*52. O indivíduo VC14 não apresentou mudanças sugestivas de escape entre as visitas em *Gag*, porém a grande mudança de proporção nas mutações R71T (4,6% em V1 para 97% em VX), P129Q (5,7% em V1 para 96% em VX) e V133I (5,2% para 97%) são sugestivas de escape em epítomos restritos por B*42. O indivíduo VC15 apresentou as mutações V82I (43,4% em V1 para 84,9% em VX) em *Gag* e V85L (66,8% em V1 para 99% em VX) em *Nef*, ambas em epítomos restritos Por B*57 e sugestivas de escape. Uma alta quantidade de mutações sinônimas neste indivíduo também chamou atenção e pode ser indicativo de pressão seletiva purificadora.

Embora a quantidade de mutações tenha sido mais alta em todos os VCs, a grande quantidade de mutações observadas no indivíduo VC16 chamou a atenção devido ao aumento de mutações sugestivas de escape e cocirculação de variantes. As mutações

Y81F (54,3% em V1 e 47,6% em VX) e H116N (63,4% em V1 e 63,1% em VX) em *Nef*, bem como as mutações G163A (21,4% em V1 e 24,4% em VX) e N165S (21,7% em V1 e 24% em VX) em *Gag*, mantiveram frequências semelhantes nas duas visitas indicando a sua co-circulação e manutenção no reservatório.

No códon 133 de *Nef* deste paciente, a visita VX apresentou sequências com a presença de T (10%), I (74,3%) e V (13,3%) enquanto a visita V1 possuía o aminoácido T predominantes na posição. Considerando que o aminoácido V na referida posição é o consenso para o subtipo B, os dados observados são indicativos de um escape T->I entre as duas visitas que poderá ser seguido por uma reversão para o V selvagem. A mutação H102Y (0% em V1 a 48,6% em VX) também foi sugestiva de escape em andamento.

Em *Gag*, a baixa conservação gênica nas regiões de epítomos do VC16 se fez mais evidente. As mutações I34L, A115E, A118P, P146A, N242T foram observadas com frequências entre 20 e 30% apenas na V1 e podem ter sido resultado de um escape anterior a visita V1. Já as mutações T122A (4,8% em V1 para 72% em VX) e K335R (50% em V1 e 99% em VX) são sugestivas de um escape ocorrida entre as duas visitas. Considerando que todos os supostos escapes descritos ocorreram em epítomos restritos por B*57, esta maior variação pode ser um sinal da maior pressão seletiva causada por esse alelo já observada em outros estudos.

Para os indivíduos EC08, ECB18, VC15 e VC16, os quais possuem o genótipo B*57 ou B*58, a análise dos epítomos IW9 (*Gag* 147-155), KF11 (*Gag* 162-172) e TW10 (*Gag* 240-249) demonstrou o menor grau de evolução dos indivíduos ECs e ECBs, mesmo nesses epítomos descritos por apresentarem grande pressão seletiva. Enquanto EC08 e ECB18 apresentarem as sequências de epítomos selvagens para o subtipo B, o indivíduo VC15 apresentou as mutações I147L, A146P e T242N e o indivíduo VC16 apresentou as mutações A146P, A163G e T242N. As mutações I147L, T242N e G248A são observadas em diversos estudos e são bem descritas por surgirem na grande maioria de indivíduos soropositivos com os genótipos B*57 e B*58 (BAILEY et al., 2006; CARLSON et al., 2015), apesar de causarem uma diminuição do fitness viral (POHLMAYER et al., 2013).

6. CONCLUSÃO

No presente estudo, a utilização de uma metodologia de NGS foi essencial e eficiente para descrever padrões de evolução discretos ocorridos nos indivíduos com o perfil de controle da progressão e da viremia, permitindo a observação de variantes virais com frequências mínimas de 0,5%.

Pelas análises filogenéticas, ficou evidente o baixo nível de diversidade e divergência dos genes *Gag* e *Nef* na população viral de cada indivíduo analisado. Embora em baixa quantidade nos ECs, mutações únicas para cada visita foram encontradas em todos os indivíduos, indicando que realmente existem baixos níveis de replicação viral e que as variantes geradas nesse processo conseguem efetivamente se integrar.

De modo geral, a quantidade de mutações e posições variáveis para os dois genes demonstraram uma maior conservação do gene *gag* em comparação com gene *nef*. No entanto, este padrão foi muito menos expressivo em indivíduos EC, indicando que o maior controle da viremia limita a variabilidade em ambos genes.

Como observado em estudos anteriores para o gene *Env*, indivíduos com perfil VC são os que apresentam maiores níveis de evolução, enquanto ECs apresentam os menores. ECBs de modo geral se encontram no meio termo, porém as análises para esses indivíduos foram desviadas deste padrão intermediário devido hipermutação mediada por APOBEC3G e um possível caso de co-infecção.

As análises de regiões de epítomos restritos por CTL demonstraram a presença de mutações indicativas de escape na maioria dos indivíduos. Embora apenas mutações com expressiva mudança de frequência tenham sido consideradas como possível escape, mutações não sinônimas foram encontradas em baixas frequências para todos os indivíduos, indicando que tais variantes surgem, mas a grande maioria não obtém sucesso evolutivo.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ABRAM, M. E. et al. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. **Journal of virology**, v. 84, n. 19, p. 9864–78, out. 2010.

ALTFELD, M. et al. Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. **AIDS (London, England)**, v. 17, n. 18, p. 2581–91, 5 dez. 2003.

ARTENSTEIN, A. W. et al. Sequences and phylogenetic analysis of the *nef* gene from Thai subjects harboring subtype E HIV-1. **AIDS research and human retroviruses**, v. 12, n. 6, p. 557–60, 10 abr. 1996.

BABRAHAM BIOINFORMATICS. **FastQC - A quality control tool for high throughput sequence data.**, 2015.

BAILEY, J. R. et al. Maintenance of viral suppression in HIV-1-infected HLA-B*57+ elite suppressors despite CTL escape mutations. **The Journal of Experimental Medicine**, v. 203, n. 5, p. 1357–1369, 2006.

BAILEY, J. R. et al. Transmission of human immunodeficiency virus type 1 from a patient who developed AIDS to an elite suppressor. **Journal of virology**, v. 82, n. 15, p. 7395–7410, 2008.

BAILEY, J. R. et al. Evidence of CD8+ T-cell-mediated selective pressure on human immunodeficiency virus type 1 *nef* in HLA-B*57+ elite suppressors. **Journal of virology**, v. 83, n. 1, p. 88–97, 2009.

BARRE-SINOUSSE, F. et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). **Science**, v. 220, n. 4599, p. 868–871, 20 maio 1983.

BELLO, G. et al. A subset of human immunodeficiency virus type 1 long-term non-progressors is characterized by the unique presence of ancestral sequences in the viral population. **The Journal of general virology**, v. 86, n. Pt 2, p. 355–64, fev. 2005.

BELLO, G. et al. Plasma viral load threshold for sustaining intrahost HIV type 1 evolution. **AIDS research and human retroviruses**, v. 23, n. 10, p. 1242–50, out. 2007.

BELLO, G. et al. Immune activation and antibody responses in non-progressing elite

controller individuals infected with HIV-1. **Journal of medical virology**, v. 81, n. 10, p. 1681–90, out. 2009.

BIMBER, B. N. et al. Ultradeep pyrosequencing detects complex patterns of CD8+ T-lymphocyte escape in simian immunodeficiency virus-infected macaques. **Journal of virology**, v. 83, n. 16, p. 8247–8253, ago. 2009.

BLUMENTHAL, R. et al. HIV entry and envelope glycoprotein-mediated fusion. **The Journal of biological chemistry**, v. 287, n. 49, p. 40841–9, 30 nov. 2012.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics (Oxford, England)**, p. btu170–, 28 abr. 2014.

BRUMME, Z. L. et al. Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) *Gag*, *Pol*, and *Nef* cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. **Journal of virology**, v. 82, n. 18, p. 9216–9227, 2008a.

BRUMME, Z. L. et al. Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) *Gag*, *Pol*, and *Nef* cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. **Journal of virology**, v. 82, n. 18, p. 9216–27, set. 2008b.

BRUMME, Z. L.; POON, A. F. Y. Identifying HLA-Associated Polymorphisms in HIV-1. n. October, p. 3–16, 2011.

CAETANO, D. G. **CARACTERIZAÇÃO DE PACIENTES HIV-POSITIVOS NÃO PROGRESSORES DE LONGO TERMO (LTNP) DO RIO DE JANEIRO - CONTRIBUIÇÃO DE FATORES VIROLÓGICOS**. [s.l.] Universidade Federal do Estado do Rio de Janeiro, 2013.

CAETANO, D. G. . et al. LACK OF INTRA-HOST VIRAL EVOLUTION IN HIV CONTROLLERS: A 10-13 YEARS FOLLOW-UP STUDY. **Virus Reviews & Research (Annals of XXV BCV)**, v. 19, 2014.

CALE, E. M. et al. Epitope-specific CD8+ T lymphocytes cross-recognize mutant simian immunodeficiency virus (SIV) sequences but fail to contain very early evolution and eventual fixation of epitope escape mutations during SIV infection. **Journal of virology**,

v. 85, n. 8, p. 3746–57, 2011.

CARLSON, J. M. et al. Correlates of Protective Cellular Immunity Revealed by Analysis of Population-Level Immune Escape Pathways in HIV-1. **Journal of Virology**, v. 86, n. 24, p. 13202–13216, 2012.

CARLSON, J. M. et al. HIV-1 adaptation to HLA: a window into virus–host immune interactions. **Trends in Microbiology**, v. 23, n. 4, p. 212–224, 2015.

CARLSON, J. M.; BRUMME, Z. L. HIV evolution in response to HLA-restricted CTL selection pressures: a population-based perspective. **Microbes and infection / Institut Pasteur**, v. 10, n. 5, p. 455–61, abr. 2008.

CASADO, C. et al. Coinfection and superinfection in patients with long-term, nonprogressive HIV-1 disease. **The Journal of infectious diseases**, v. 196, n. 6, p. 895–9, 15 set. 2007.

CHABRIA, S. B.; GUPTA, S.; KOZAL, M. J. Deep Sequencing of HIV: Clinical and Research Applications. **Annual review of genomics and human genetics**, n. April, p. 1–31, 2014.

CHOOPANYA, K. et al. Antiretroviral prophylaxis for HIV infection in injecting drug users in Bangkok, Thailand (the Bangkok Tenofovir Study): A randomised, double-blind, placebo-controlled phase 3 trial. **The Lancet**, v. 381, n. 9883, p. 2083–2090, 2013.

CHUN, T.-W. et al. Effect of antiretroviral therapy on HIV reservoirs in elite controllers. **The Journal of infectious diseases**, v. 208, n. 9, p. 1443–7, 1 nov. 2013.

CLERC, O. et al. HIV-1 elite controllers: beware of super-infections. **Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology**, v. 47, n. 4, p. 376–8, 1 abr. 2010.

CLOUSE, K. A. et al. Monokine regulation of human immunodeficiency virus-1 expression in a chronically infected human T cell clone. **Journal of immunology (Baltimore, Md. : 1950)**, v. 142, n. 2, p. 431–8, 15 jan. 1989.

COHEN, M. S. et al. Prevention of HIV-1 Infection with Early Antiretroviral Therapy. **New England Journal of Medicine**, v. 365, n. 6, p. 493–505, 11 ago. 2011.

CÔRTEZ, F. H. et al. HIV Controllers With Different Viral Load Cutoff Levels Have Distinct Virologic and Immunologic Profiles. **JAIDS Journal of Acquired Immune Deficiency Syndromes**, v. 68, n. 4, p. 377–385, 2015.

CRAIGIE, R.; BUSHMAN, F. D. HIV DNA integration. **Cold Spring Harbor perspectives in medicine**, v. 2, n. 7, p. a006890, jul. 2012.

DAHABIEH, M. S.; BATTIVELLI, E.; VERDIN, E. Understanding HIV Latency: The Road to an HIV Cure. **Annual review of medicine**, v. 66, p. 407–421, 14 jan. 2015.

DALGLEISH, A. G. et al. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. **Nature**, v. 312, n. 5996, p. 763–7, 1984.

DARRIBA, D. et al. jModelTest 2: more models, new heuristics and parallel computing. **Nature methods**, v. 9, n. 8, p. 772, ago. 2012.

DEACON, N. J. et al. Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients. **Science (New York, N.Y.)**, v. 270, n. 5238, p. 988–91, 10 nov. 1995.

DEAN, M. et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE . **Science (New York, N.Y.)**, v. 273, n. 5283, p. 1856–62, 27 set. 1996.

DENG, K. et al. Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. **Nature**, v. 517, n. 7534, p. 381–385, 7 jan. 2015.

DILERNIA, D. A. et al. HLA-driven convergence of HIV-1 viral subtypes B and F toward the adaptation to immune responses in human populations. **PloS one**, v. 3, n. 10, p. e3429, jan. 2008.

DING, Y. et al. Timing of antiretroviral therapy initiation after diagnosis of recent human immunodeficiency virus infection and CD4(+) T-cell recovery. **Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases**, 25 nov. 2015.

EIGEN, M.; SCHUSTER, P. The hypercycle. A principle of natural self-organization. Part

A: Emergence of the hypercycle. **Die Naturwissenschaften**, v. 64, n. 11, p. 541–65, nov. 1977.

EYZAGUIRRE, L. M. et al. Elevated hypermutation levels in HIV-1 natural viral suppressors. **Virology**, v. 443, n. 2, p. 306–12, 1 set. 2013.

FASSATI, A.; GOFF, S. P. Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. **Journal of virology**, v. 75, n. 8, p. 3626–35, abr. 2001.

FISCHER, W. et al. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. **PloS one**, v. 5, n. 8, p. e12303, jan. 2010.

FOLKS, T. M. et al. Tumor necrosis factor alpha induces expression of human immunodeficiency virus in a chronically infected T-cell clone. **Proceedings of the National Academy of Sciences of the United States of America**, v. 86, n. 7, p. 2365–8, abr. 1989.

FOURATI, S. et al. Differential impact of APOBEC3-driven mutagenesis on HIV evolution in diverse anatomical compartments. **AIDS (London, England)**, v. 28, n. 4, p. 487–91, 20 fev. 2014.

GAARDBO, J. C. et al. Different immunological phenotypes associated with preserved CD4+ T cell counts in HIV-infected controllers and viremic long term non-progressors. **PloS one**, v. 8, n. 5, p. e63744, jan. 2013.

GAARDBO, J.; HARTLING, H. Thirty Years with HIV Infection—Nonprogression Is Still Puzzling: Lessons to Be Learned from Controllers and Long-Term Nonprogressors. **AIDS research and ...**, v. 2012, p. 161584, jan. 2012.

GALL, A. et al. Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. **Journal of clinical microbiology**, v. 50, n. 12, p. 3838–44, dez. 2012.

GALLO, R. et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. **Science**, v. 224, n. 4648, p. 500–503, 4 maio 1984.

GIJSBERS, E. F. et al. Viral evolution in HLA-B27-restricted CTL epitopes in human

immunodeficiency virus type 1-infected individuals. **Journal of General Virology**, v. 96, n. 8, p. 2372–2380, 2015.

GOONETILLEKE, N. et al. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. **The Journal of experimental medicine**, v. 206, n. 6, p. 1253–72, 8 jun. 2009.

GÖTTLINGER, H. G. HIV-1 *Gag*: a Molecular Machine Driving Viral Particle Assembly and Release. p. 2–28, 2000.

GOUNDER, K. et al. High frequency of transmitted HIV-1 *Gag* HLA class I-driven immune escape variants but minimal immune selection over the first year of clade C infection. **PLoS one**, v. 10, n. 3, p. e0119886, jan. 2015.

GROTTO, R.; PARDINI, M. Biologia molecular do HIV-1 e genética da resistência humana à AIDS. **Arquivos de Ciências da Saúde**, v. 13, n. 3, p. 61–64, 2006.

HARRIS, R. S.; LIDDAMENT, M. T. Retroviral restriction by APOBEC proteins. **Nature reviews. Immunology**, v. 4, n. 11, p. 868–77, nov. 2004.

HENDEL, H. et al. New class I and II HLA alleles strongly associated with opposite patterns of progression to AIDS. **Journal of immunology (Baltimore, Md. : 1950)**, v. 162, n. 11, p. 6942–6, 1 jun. 1999.

HENN, M. R. et al. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. **PLoS pathogens**, v. 8, n. 3, p. e1002529, 3 jan. 2012.

HEYNDRICKX, L. et al. Simplified strategy for detection of recombinant human immunodeficiency virus type 1 group M isolates by *gag/env* heteroduplex mobility assay. Study Group on Heterogeneity of HIV Epidemics in African Cities. **Journal of virology**, v. 74, n. 1, p. 363–70, jan. 2000.

HO, Y.-C. et al. Replication-Competent Noninduced Proviruses in the Latent Reservoir Increase Barrier to HIV-1 Cure. **Cell**, v. 155, n. 3, p. 540–551, 2013.

HU, W.-S.; HUGHES, S. H. HIV-1 reverse transcription. **Cold Spring Harbor perspectives in medicine**, v. 2, n. 10, p. 55–70, out. 2012.

HUANG, Y.; ZHANG, L.; HO, D. D. Biological characterization of *nef* in long-term survivors of human immunodeficiency virus type 1 infection. **Journal of virology**, v. 69, n. 12, p. 8142–6, dez. 1995.

HUANG, Y.; ZHANG, L.; HO, D. D. **Characterization of gag and pol sequences from long-term survivors of human immunodeficiency virus type 1 infection.** **Virology**, 5 jan. 1998.

HUGHES, A. L. et al. SIV Genome-Wide Pyrosequencing Provides a Comprehensive and Unbiased View of Variation within and outside CD8 T Lymphocyte Epitopes. **PLoS ONE**, v. 7, n. 10, 2012.

IMAMICHI, H. et al. Lifespan of effector memory CD4+ T cells determined by replication-incompetent integrated HIV-1 provirus. **AIDS (London, England)**, v. 28, n. 8, p. 1091–9, 2014.

JOSHI, N.; FASS, J. **Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files**, 2011.

JUAREZ-MOLINA, C. I. et al. Impact of HLA Selection Pressure on HIV Fitness at a Population Level in Mexico and Barbados. **Journal of virology**, v. 88, n. 18, p. 10392–8, 15 set. 2014.

KASLOW, R. A. et al. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. **Nature medicine**, v. 2, n. 4, p. 405–11, abr. 1996.

KASSUTTO, S.; ROSENBERG, E. Primary HIV type 1 infection. **Clinical Infectious Diseases**, v. 38, p. 1447–1453, 2004.

KAWASHIMA, Y. et al. Adaptation of HIV-1 to human leukocyte antigen class I. **Nature**, v. 458, n. 7238, p. 641–5, 2 abr. 2009.

KEARSE, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics (Oxford, England)**, v. 28, n. 12, p. 1647–9, 15 jun. 2012.

KIEPIELA, P. et al. CD8+ T-cell responses to different HIV proteins have discordant

associations with viral load. **Nature medicine**, v. 13, n. 1, p. 46–53, jan. 2007.

KIRCHHOFF, F. et al. Sequence variations in human immunodeficiency virus type 1 *Nef* are associated with different stages of disease. **Journal of virology**, v. 73, n. 7, p. 5497–508, jul. 1999.

KOENIG, S. et al. Transfer of HIV-1-specific cytotoxic T lymphocytes to an AIDS patient leads to selection for mutant HIV variants and subsequent disease progression. **Nature medicine**, v. 1, n. 4, p. 330–6, abr. 1995.

KUNWAR, P. et al. Superior control of HIV-1 replication by CD8+ T cells targeting conserved epitopes: implications for HIV vaccine design. **PloS one**, v. 8, n. 5, p. e64405, jan. 2013.

LEARMONT, J. et al. Long-term symptomless HIV-1 infection in recipients of blood products from a single donor. **The Lancet**, v. 340, n. 8824, p. 863–867, 1992.

LEE, C. N. et al. HIV type 1 env subtype A variants in Taiwan. **AIDS research and human retroviruses**, v. 14, n. 9, p. 807–9, 10 jun. 1998.

LEITNER, T. et al. **HIV Sequence Compendium 2013**. [s.l.] Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, 2013.

LESLIE, A. J. et al. HIV evolution: CTL escape mutation and reversion after transmission. **Nature medicine**, v. 10, n. 3, p. 282–9, mar. 2004.

LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics (Oxford, England)**, v. 25, n. 16, p. 2078–9, 15 ago. 2009.

LIU, J. et al. Molecular architecture of native HIV-1 gp120 trimers. **Nature**, v. 455, n. 7209, p. 109–13, 4 set. 2008.

LIU, R. et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. **Cell**, v. 86, n. 3, p. 367–77, 9 ago. 1996.

LLANO, M. et al. An essential role for LEDGF/p75 in HIV integration. **Science (New York, N.Y.)**, v. 314, n. 5798, p. 461–4, 20 out. 2006.

LOS ALAMOS NATIONAL LABORATORY. **Los Alamos Immunology Database**. Disponível em: <<http://www.hiv.lanl.gov/content/immunology>>.

LOVE, T. M. T. et al. Mathematical modeling of ultradeep sequencing data reveals that acute CD8⁺ T-lymphocyte responses exert strong selective pressure in simian immunodeficiency virus-infected macaques but still fail to clear founder epitope sequences. **Journal of virology**, v. 84, n. 11, p. 5802–14, jun. 2010.

MAHALANABIS, M. et al. Continuous viral escape and selection by autologous neutralizing antibodies in drug-naive human immunodeficiency virus controllers. **Journal of virology**, v. 83, n. 2, p. 662–72, jan. 2009.

MARTIN, E. et al. Early immune adaptation in HIV-1 revealed by population-level approaches. **Retrovirology**, v. 11, n. 1, p. 64, 29 ago. 2014.

MENS, H. et al. HIV-1 continues to replicate and evolve in patients with natural control of HIV infection. **Journal of virology**, v. 84, n. 24, p. 12971–81, dez. 2010.

MESSIAEN, P. et al. Characterization of LEDGF/p75 genetic variants and association with HIV-1 disease progression. **PloS one**, v. 7, n. 11, p. e50204, jan. 2012.

MICHAEL, N. L. et al. Defective accessory genes in a human immunodeficiency virus type 1-infected long-term survivor lacking recoverable virus. **Journal of virology**, v. 69, n. 7, p. 4228–36, jul. 1995.

MIGUELES, S. A et al. The differential ability of HLA B*5701+ long-term nonprogressors and progressors to restrict human immunodeficiency virus replication is not caused by loss of recognition of autologous viral *gag* sequences. **Journal of virology**, v. 77, n. 12, p. 6889–6898, 2003.

MILLER, M. D.; FARNET, C. M.; BUSHMAN, F. D. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. **Journal of virology**, v. 71, n. 7, p. 5382–90, jul. 1997.

MINISTÉRIO DA SAÚDE. Boletim epidemiológico Brasil. 2014a.

MINISTÉRIO DA SAÚDE. **Protocolo clínico e diretrizes terapêuticas para manejo da infecção pelo hiv em adultos**, 2014b.

MIURA, T. et al. HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare *gag* variants associated with reduced viral replication capacity and strong

cytotoxic T-lymphocyte [corrected] recognition. **Journal of virology**, v. 83, n. 6, p. 2743–55, mar. 2009.

MOLOGNI, D. et al. Vpr and HIV-1 disease progression: R77Q mutation is associated with long-term control of HIV-1 infection in different groups of patients. **AIDS (London, England)**, v. 20, n. 4, p. 567–74, 28 fev. 2006.

MOORE, J. P.; TRKOLA, A.; DRAGIC, T. Co-receptors for HIV-1 entry. **Current opinion in immunology**, v. 9, n. 4, p. 551–62, ago. 1997.

MOTHE, B. et al. CTL responses of high functional avidity and broad variant cross-reactivity are associated with HIV control. **PLoS one**, v. 7, n. 1, p. e29717, jan. 2012.

MUDD, P. A et al. Escape from CD8(+) T cell responses in Mamu-B*00801(+) macaques differentiates progressors from elite controllers. **J Immunol**, v. 188, n. 7, p. 3364–3370, 2012.

MURAKOSHI, H. et al. Clinical Control of HIV-1 by Cytotoxic T Cells Specific for Multiple Conserved Epitopes. **Journal of virology**, v. 89, n. 10, p. 5330–9, maio 2015.

NOWAK, M. A. What is a quasispecies? **Trends in ecology & evolution**, v. 7, n. 4, p. 118–21, abr. 1992.

O'CONNELL, K. A. et al. Control of HIV-1 in elite suppressors despite ongoing replication and evolution in plasma virus. **Journal of virology**, v. 84, n. 14, p. 7018–28, jul. 2010.

O'CONNELL, K. A. et al. Viral suppression of multiple escape mutants by de novo CD8(+) T cell responses in a human immunodeficiency virus-1 infected elite suppressor. **Retrovirology**, v. 8, p. 63, jan. 2011.

OKULICZ, J. F. et al. Clinical outcomes of elite controllers, viremic controllers, and long-term nonprogressors in the US Department of Defense HIV natural history study. **The Journal of infectious diseases**, v. 200, n. 11, p. 1714–23, 1 dez. 2009.

OKULICZ, J. F. et al. Influence of the timing of antiretroviral therapy on the potential for normalization of immune status in human immunodeficiency virus 1-infected individuals. **JAMA internal medicine**, v. 175, n. 1, p. 88–99, jan. 2015.

OTT, D. E. Cellular proteins detected in HIV-1. **Reviews in medical virology**, v. 18, n. 3,

p. 159–75, 2008.

PACE, M. J. et al. HIV reservoirs and latency models. **Virology**, v. 411, n. 2, p. 344–54, 15 mar. 2011.

PANTALEO, G.; FAUCI, A. S. Immunopathogenesis of HIV infection. **Annual review of microbiology**, v. 50, p. 825–54, jan. 1996.

PASSAES, C. P. B. et al. Association of single nucleotide polymorphisms in the lens epithelium-derived growth factor (LEDGF/p75) with HIV-1 infection outcomes in Brazilian HIV-1+ individuals. **PloS one**, v. 9, n. 7, p. e101780, jan. 2014.

PERELSON, A. S. et al. HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time. **Science**, v. 271, n. 5255, p. 1582–1586, 15 mar. 1996.

PEREYRA, F. et al. Genetic and Immunologic Heterogeneity among Persons Who Control HIV Infection in the Absence of Therapy. **The Journal of Infectious Diseases**, v. 197, n. 4, p. 563–571, 2008.

POHLMAYER, C. W. et al. CD8+ T cells from HLA-B*57 elite suppressors effectively suppress replication of HIV-1 escape mutants. **Retrovirology**, v. 10, p. 152, jan. 2013.

RACHINGER, A. et al. Recovery of viremic control after superinfection with pathogenic HIV type 1 in a long-term elite controller of HIV type 1 infection. **Clinical infectious diseases : an official publication of the Infectious Diseases Society of America**, v. 47, n. 11, p. e86–9, 1 dez. 2008.

RADEBE, M. et al. Broad and persistent Gag-specific CD8+ T-cell responses are associated with viral control but rarely drive viral escape during primary HIV-1 infection. **AIDS**, v. 29, n. 1, p. 23–33, jan. 2015a.

RADEBE, M. et al. Broad and persistent Gag-specific CD8+ T-cell responses are associated with viral control but rarely drive viral escape during primary HIV-1 infection. **AIDS (London, England)**, v. 29, n. 1, p. 23–33, 2 jan. 2015b.

RAJAN, D. et al. Effect of R77Q, R77A and R80A changes in Vpr on HIV-1 replication and CD4 T cell depletion in human lymphoid tissue ex vivo. **AIDS (London, England)**, v. 20,

n. 6, p. 831–6, 4 abr. 2006.

RAM, D. et al. Evaluation of GS Junior and MiSeq next-generation sequencing technologies as an alternative to Trugene population sequencing in the clinical HIV laboratory. **Journal of virological methods**, v. 212, p. 12–16, 2015.

RAMBAUT, A. et al. The causes and consequences of HIV evolution. **Nature reviews. Genetics**, v. 5, n. 1, p. 52–61, jan. 2004.

REDD, A. D.; QUINN, T. C.; TOBIAN, A. A. R. Frequency and implications of HIV superinfection. **The Lancet infectious diseases**, v. 13, n. 7, p. 622–8, jul. 2013.

ROBERTS, H. E. et al. Structured observations reveal slow HIV-1 CTL escape. **PLoS genetics**, v. 11, n. 2, p. e1004914, 2 fev. 2015.

RODÉS, B. et al. Differences in disease progression in a cohort of long-term non-progressors after more than 16 years of HIV-1 infection. **AIDS (London, England)**, v. 18, n. 8, p. 1109–16, 21 maio 2004.

ROZERA, G. et al. Evolution of HIV-1 tropism at quasispecies level after 5 years of combination antiretroviral therapy in patients always suppressed or experiencing episodes of virological failure. **The Journal of antimicrobial chemotherapy**, p. 1–10, 10 jul. 2014.

RUELAS, D. S.; GREENE, W. C. **An integrated overview of HIV-1 latency**Cell, 2013.

SAEZ-CIRION, A. et al. Heterogeneity in HIV Suppression by CD8 T Cells from HIV Controllers: Association with *Gag*-Specific CD8 T Cell Responses. **The Journal of Immunology**, v. 182, n. 12, p. 7828–7837, 2009.

SÁEZ-CIRIÓN, A.; PANCINO, G. HIV controllers: a genetically determined or inducible phenotype? **Immunological reviews**, v. 254, n. 1, p. 281–94, jul. 2013.

SALGADO, M. et al. Evolution of the HIV-1 *nef* gene in HLA-B*57 positive elite suppressors. **Retrovirology**, v. 7, n. 1, p. 94, jan. 2010.

SALVI, R. et al. Grossly defective *nef* gene sequences in a human immunodeficiency virus type 1-seropositive long-term nonprogressor. **Journal of virology**, v. 72, n. 5, p. 3646–57, maio 1998.

SANTA-MARTA, M. et al. Host Factors and HIV-1 Replication: Clinical Evidence and

Potential Therapeutic Approaches. **Frontiers in Immunology**, v. 4, p. 343, jan. 2013.

SCHRÖDER, A. R. W. et al. HIV-1 integration in the human genome favors active genes and local hotspots. **Cell**, v. 110, n. 4, p. 521–9, 23 ago. 2002.

SEELAMGARI, A. et al. Role of viral regulatory and accessory proteins in HIV-1 replication. **Frontiers in bioscience : a journal and virtual library**, v. 9, p. 2388–413, 1 set. 2004.

SHANKARAPPA, R. et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. **Journal of virology**, v. 73, n. 12, p. 10489–502, dez. 1999.

SIERRA, M. et al. The analysis of near full-length genome sequences of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Chile, Venezuela and Spain reveals their relationship to diverse lineages of recombinant viruses related to CRF12_BF. **Infection, Genetics and Evolution**, v. 5, n. 3, p. 209–217, abr. 2005.

SIERRA, S.; KUPFER, B.; KAISER, R. Basics of the virology of HIV-1 and its replication. **Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology**, v. 34, n. 4, p. 233–44, dez. 2005.

SILICIANO, J. D. et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. **Nature medicine**, v. 9, n. 6, p. 727–728, 2003.

SMITH, D. M.; RICHMAN, D. D.; LITTLE, S. J. HIV superinfection. **The Journal of infectious diseases**, v. 192, n. 3, p. 438–44, 1 ago. 2005.

STARCICH, B. R. et al. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. **Cell**, v. 45, n. 5, p. 637–48, 6 jun. 1986.

SUNDQUIST, W. I.; KRÄUSSLICH, H.-G. HIV-1 assembly, budding, and maturation. **Cold Spring Harbor perspectives in medicine**, v. 2, n. 7, p. a006924, jul. 2012.

TAMURA, K. et al. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. **Molecular biology and evolution**, v. 30, n. 12, p. 2725–9, dez. 2013.

TEIXEIRA, S. L. M. et al. Association of the HLA-B*52 allele with non-progression to AIDS in Brazilian HIV-1-infected individuals. **Genes and immunity**, v. 15, n. 4, p. 256–62, abr. 2014.

TERSMETTE, M. et al. Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the pathogenesis of acquired immunodeficiency syndrome: studies on sequential HIV isolates. **Journal of virology**, v. 63, n. 5, p. 2118–25, maio 1989.

THIGPEN, M. C. et al. Antiretroviral Preexposure Prophylaxis for Heterosexual HIV Transmission in Botswana. **New England Journal of Medicine**, v. 367, n. 5, p. 423–434, 2012.

TOMESCU, C.; ABDULHAQQ, S.; MONTANER, L. J. Evidence for the innate immune response as a correlate of protection in human immunodeficiency virus (HIV)-1 highly exposed seronegative subjects (HESN). **Clinical and experimental immunology**, v. 164, n. 2, p. 158–69, maio 2011.

TROYER, R. M. et al. Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression. **Journal of virology**, v. 79, n. 14, p. 9006–18, jul. 2005.

TURNER, B. G.; SUMMERS, M. F. Structural biology of HIV. **Journal of molecular biology**, v. 285, n. 1, p. 1–32, 8 jan. 1999.

UNAIDS. **How AIDS changed everything**. [s.l.: s.n.].

VARTANIAN, J. P. et al. High-resolution structure of an HIV-1 quasispecies: identification of novel coding sequences. **AIDS (London, England)**, v. 6, n. 10, p. 1095–8, out. 1992.

WILKE, C. O. Quasispecies theory in the context of population genetics. **BMC evolutionary biology**, v. 5, p. 44, jan. 2005.

WILLIAMSON, S. Adaptation in the env gene of HIV-1 and evolutionary theories of disease progression. **Molecular biology and evolution**, v. 20, n. 8, p. 1318–25, ago. 2003.

YEBRA, G.; BROWN, A. J. L. Evolutionary Rates of HIV 1 Accessory Genes from Full length Datasets Across Subtypes. **22nd HIV Dynamics & Evolution Workshop**, 2015.

ZUÑIGA, R. et al. Relative dominance of *Gag* p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control. **Journal of virology**, v. 80, n. 6, p. 3122–5, mar. 2006.

8. ANEXOS

8.1. TABELA DE VARIAÇÕES NO GENE GAG

Tabela 9. Mutações observadas em *Gag* e suas frequências em V1 e VX

Indivíduo	Códon	Mutação	Frequência	
			V1	VX
EC02	135	V -> A		0,5%
	137	N -> S		3%
	139	CAG -> CAA	28,2%	29,7%
	147	I -> V		0,6%
	160	GAA -> GAG		0,5%
	165	AGC -> AGT		1,4%
	170	CCC -> CCA	0,7%	
	275	R -> G		0,8%
	275	AGA -> AGG		23,8%
	278	AGC -> TGC		22,8%
	280	T -> A	0,6%	
	EC08	43	R -> L	
44		F -> L		0,7%
80		AAT -> AAC	0,5%	
83		GCG -> GCT		0,8%
83		GCG -> GCA	0,5%	
88		GTA -> GTG	0,6%	
120		T -> A	0,6%	
147		TCA -> TCG	0,7%	
149		R -> I		0,6%
153		GCA -> GCG		0,8%
156		AAG -> AAA	0,5%	
164		S -> G		0,6%
168		I -> V		0,7%
248		G -> E	1%	
250		M -> I		0,9%
295		D -> N	1,3%	
298		D -> Y		0,6%
298		D -> N	1,7%	
298		D -> G	2%	
300		F -> L	0,5%	0,6%
301	Y -> C	0,6%		
304	L -> I		0,5%	
305	AGA -> AGG	0,5%		

	308	CAA	->	CAG	1,1%	
	312	E	->	G	0,7%	
	317	M	->	I		0,7%
	326	GCA	->	GCG	0,5%	0,5%
	331	AAG	->	AAA	2,2%	
	339	M	->	V		0,5%
EC17	142	M	->	T	2,6%	
	147	L	->	I	1,9%	
	151	T	->	A		0,5%
	151	ACC	->	ACT	2,0%	
	158	GTG	->	GTA	1,8%	
	160	GAA	->	GAG		0,7%
	161	E	->	K	1,4%	
	162	K	->	R		0,6%
	163	GCT	->	GCC		0,5%
	164	TTT	->	TTC	1,9%	
	166	CCG	->	CCA	1,8%	
	191	GTG	->	GTA	1,6%	
	195	CAA	->	CAG		3,3%
	197	GCC	->	GCT	2,2%	56,5%
	199	CAA	->	CAG	2,0%	
	204	T	->	A		0,5%
	267	ATT	->	ATC	1,6%	
	268	L	->	I	1,3%	
	269	G	->	E	1,1%	
	271	N	->	T	1,4%	
	275	AGA	->	AGG		0,5%
	277	Y	->	C		0,6%
	280	V	->	A	0,6%	
	297	GTA	->	GTG	2,2%	
	298	D	->	G		0,9%
	300	F	->	L	2,1%	
	303	ACT	->	ACC	2,1%	
	304	TTA	->	CTA	2,1%	
	305	R	->	K	1,6%	
	307	E	->	G		0,6%
ECB09	279	P	->	I		1,8%
	280	I	->	T		39,1%
	283	CTG	->	CTT		0,5%
ECB11	141	Q	->	R		0,9%
	157	AAA	->	AAG	0,5%	
	190	ACA	->	ACG	0,6%	
	203	E	->	G	0,6%	

	264	R	->	G	0,6%	
	267	I	->	V	0,8%	
	277	Y	->	H	0,5%	
	294	AGG	->	AGA	48,2%	
	300	F	->	L	0,6%	
	303	V	->	A	0,6%	
	24	G	->	E	6,1%	
	26	K	->	S	93,6%	
	26	K	->	N	6,3%	
	28	K	->	R	25,3%	
	30	K	->	Q	24,4%	
	30	K	->	R	74,7%	
	146	GCT	->	GCC	9,3%	
	152	TTA	->	CTA	0,6%	
	161	GAG	->	GAA	90,3%	
	192	G(GGA)	->	R(AGA)	9,9%	
ECB13	193	GGG	->	GGA	14,1%	
	268	CTA	->	CTG	95,2%	
	278	AGC	->	AGT	2%	
	298	GAC	->	GAT	92%	
	301	TAT	->	TAC	20,3%	
	310	S	->	T	99,4%	
	311	Q	->	R	1%	
	322	TTG	->	TTA	10,3%	
	325	N	->	D	1%	
	335	AAA	->	AAG	0,5%	
	338	GGA	->	GGG	92,7%	
	338	G(GGG)	->	R(AGG)	7,3%	
	76	R	->	K	20,2%	
	84	T	->	A	3,1%	
	85	L	->	P	1%	0,5%
	147	M	->	I	17,2%	
	148	S	->	P	0,6%	
	153	N	->	D	2,0%	
	161	E	->	K	18,8%	
ECB18	163	GCT	->	GCC	17,5%	0,99%
	165	S	->	G	0,7%	0,7%
	175	TTA	->	CTA	1,1%	
	181	CCA	->	CCG	15,2%	
	189	N	->	H	13,6%	
	190	ACA	->	ACG	0,7%	
	203	GAG	->	GAA	14,3%	99%
	207	E	->	G	0,5%	

	208	GAA	->	GAG		0,5%
	210	GCA	->	GCG		0,6%
	213	D	->	N	25,3%	
	215	L	->	M	13,8%	
	251	ACA	->	ACG		0,5%
	295	D	->	N	22,3%	
	295	D	->	G	0%	
	298	GAC	->	GAT	16,1%	
	308	CAA	->	CAG	0,9%	
	314	AAG	->	AGG		0,5%
	136	Q	->	R		1,3%
	138	L	->	P	0,6%	
	139	Q	->	R	0,8%	
	139	Q	->	R		0,6%
	160	E	->	G	0,6%	
	169	I	->	V		0,5%
	169	I	->	M	1,4%	
	194	CAT	->	CAC	5,4%	
VC06	204	ACT	->	ACC	0,5%	
	295	D	->	G		1,6%
	300	F	->	L	3,2%	
	302	R	->	I		0,8%
	303	ACT	->	ACC		0,7%
	307	E	->	G		1,4%
	311	Q	->	R		0,5%
	312	E	->	G		0,5%
	318	T	->	A		0,6%
	319	E	->	G		2,5%
	141	CAA	->	CAG		6,7%
	142	M	->	V	3,1%	
	143	V	->	A	0,6%	
	144	CAT	->	CAC	38%	22,5%
	146	S	->	P	39,9%	37,6%
	146	S	->	A	0,7%	
	147	L	->	I	0,8%	
VC10	154	GCA	->	GCG	1%	6,3%
	154	GCA	->	GCC	1,4%	
	155	W	->	R		0,7%
	157	AAA	->	AAG	1,5%	6,3%
	158	GTA	->	GTG		1,5%
	159	GTA	->	ATA	0,6%	1,4%
	161	GAG	->	GAA		1%
	166	CCA	->	CCG	0,5%	

	191	GTG	->	GTA	1,2%	
	192	G	->	E	1,1%	
	193	GGA	->	GGG	0,6%	
	203	GAG	->	GAA	7,8%	
	203	GAG	->	GAT	1,5%	
	268	CTA	->	CTG	1,2%	
	271	N	->	S	0,7%	
	276	M	->	I	1,3%	
	278	S	->	G	0,7%	
	278	AGT	->	GGT	2,3%	9%
	280	I	->	S	44,5%	66,9%
	280	I	->	T	1,2%	13,5%
	286	R	->	K	1,2%	9,2%
	20	CGG	->	AGG	1,5%	
	20	R	->	L	1%	
	22	AGG	->	AGA		23,1%
	24	GGG	->	GGT	0,5%	
	25	G	->	V	0,6%	
	26	K	->	N	0,7%	
	30	R	->	I	0,9%	
	31	M	->	I	1,3%	
	32	AAA	->	AAG		0,5%
	76	R	->	I	0,5%	
	78	TTA	->	TTG		22,8%
	82	I	->	V		22%
	84	GTC	->	GTA	0,7%	
	87	Y	->	C	2,5%	
	89	H	->	T		0,6%
VC14	153	AAG	->	AAA		19,6%
	154	GTG	->	GTT	1,2%	
	159	GCT	->	GCC		0,8%
	162	P	->	Q	0,6%	
	166	GAA	->	GAG		0,5%
	168	CCC	->	CCA	0,5%	
	169	TTT	->	TTC		1%
	175	GCC	->	GCA	0,5%	
	177	P	->	T	1,1%	
	177	P	->	Q	0,6%	
	179	D	->	Y	0,8%	
	182	ACC	->	ACA	0,8%	
	184	TTA	->	CTA		20,3%
	290	R	->	I	0,5%	
	291	D	->	Y	1,4%	

291	D	->	G	1,7%
294	D	->	Y	1,1%
294	D	->	G	1,2%
296	F	->	L	0,8%
299	ACC	->	ACA	0,5%
301	R	->	I	0,8%
306	T	->	S	18,6%
308	E	->	G	0,6%
308	GAA	->	GAG	1,4%
312	W	->	L	0,8%
313	M	->	I	1,8%
314	ACA	->	ACG	0,7%
315	D	->	Y	0,8%

VC15	31	UUA	->	CUA	40,5%	40,1%
	31	UUA	->	CUT	0,6%	
	34	I	->	M	0,5%	
	38	S	->	G	3,1%	
	39	AGG	->	AGA	1,0%	2,5%
	40	GAG	->	GAA		8,0%
	43	CGA	->	CGG	1,8%	7,3%
	46	GUU	->	GUC	11,2%	4,8%
	47	AAU	->	AAC	1,0%	5,7%
	73	GAA	->	GAG		5,4%
	76	AGA	->	AGG		6,5%
	77	UCA	->	UCC	12,5%	37,8%
	82	V	->	I	43,4%	84,9%
	82	GUA	->	GUG	25,1%	10,0%
	85	CUC	->	CUT		1,8%
	88	GUG	->	GUA		7,3%
	111	S	->	G		1,0%
	111	S	->	N		3,1%
	111	AGU	->	AGC		3,7%
	113	K	->	E	1,4%	
	115	A	->	T		4,4%
	115	A	->	V	11,1%	2,5%
	118	GCA	->	GCG		3,4%
	119	A	->	T	10,3%	
	119	GCA	->	GCT	1,1%	
	120	A	->	Q	1,0%	
	120	GCU	->	GCA	0,9%	
	120	GCU	->	GCG		6,4%
	121	D	->	A	30,6%	81,8%
	121	GAC	->	GAT		5,3%

122	T	->	A	49,2%	82,1%
124	H	->	N	59,7%	8,2%
124	H	->	S	0,8%	
124	H	->	Y		0,6%
124	H	->	R	2,4%	3,7%
145	Q	->	R	0,6%	
147	I	->	Y		3,6%
148	S	->	K		3,5%
148	UCA	->	UCG	2,9%	
149	P	->	T		0,6%
149	P	->	R		3,1%
150	AGA	->	AGG		17,3%
151	T	->	A	0,8%	
151	T	->	K		3,0%
151	ACU	->	ACA		6,9%
153	N	->	S		0,6%
154	A	->	G		2,5%
156	V	->	Q		2,4%
158	V	->	P		2,1%
159	V	->	C		2,0%
160	E	->	W		1,9%
161	E	->	S		1,9%
161	E	->	D		8,3%
163	A	->	S		7,4%
163	GCU	->	GCC	0,5%	6,6%
163	GCU	->	GCC		
164	UUC	->	UUT	4,4%	3,2%
164	UUC	->	UUT		
165	AGC	->	AGT	0,7%	41,9%
165	AGC	->	AGT		
168	GUG	->	GUT		2,8%
174	GCA	->	GCG		3,5%
242	N	->	T		22,7%
243	CUU	->	CUC		8,5%
244	CAG	->	CAA	1,5%	
251	ACA	->	ACC		14,3%
251	ACA	->	ACT		1,6%
302	AAA	->	AAG		25,0%
303	ACU	->	ACC	2,4%	
303	ACU	->	ACG	2,6%	2,6%
304	CUA	->	TUA	7,2%	
306	GCC	->	GCT		12,7%
308	CAA	->	CAG	42,1%	

	309	GCU	->	GCC		0,6%
	310	S	->	P		0,5%
	312	GAG	->	GAA	29,5%	2,9%
	313	GUA	->	GUC	3,6%	
	323	GUC	->	GUT	7,4%	
	326	GCG	->	GCT	6,2%	19,8%
	326	GCG	->	GCA		6,9%
	328	P	->	L		4,0%
	329	GAU	->	GAC		1,3%
	331	AAG	->	AAA	63,3%	
	334	UUG	->	UAA		2,9%
	336	GCA	->	GCC		7,8%
	337	UUG	->	UUA	2,6%	16,7%
	338	GGA	->	GGG	2,7%	12,1%
	339	CCA	->	CCT	1,1%	1,9%
	31	TTA	->	CTA		28,9%
	34	I	->	L	32,2%	
	35	V	->	I	32,2%	
	35	GTC	->	GTA	0,5%	
	35	GTC	->	GTT	30,0%	76%
	36	W	->	R		0,5%
	39	R	->	K		5,6%
	40	GAG	->	GAA		5,6%
	41	CTA	->	CTG		0,8%
	42	GAA	->	GAG		11,7%
	45	GCA	->	GCG	0,6%	
	46	V(GTT)	->	I(ATT)	29,3%	
	46	GTT	->	GTC		5,4%
	47	AAT	->	AAC		7,4%
	75	CTC	->	CTT	27,3%	
	76	AGA	->	AGG	4,7%	
	80	AAT	->	AAC	26%	
	82	V	->	L	25,9%	
	111	S	->	G		1,4%
	113	AAA	->	AAG	1,2%	0,6%
	115	A	->	E	25,0%	
	118	A	->	P	24,9%	
	118	GCA	->	GCG	0,5%	41,1%
	119	A	->	T	24,9%	
	122	T	->	A	4,8%	72%
	123	G	->	K	24,0%	25,1%

VC16

123	G	->	A	2,2%	
124	N	->	S	0,7%	35,5%
125	S	->	R	30,9%	
146	P	->	A	22,1%	
146	CCT	->	CCC	0,6%	
151	T	->	A		0,6%
152	TTA	->	CTA	0,7%	
157	AAA	->	AAG	0,6%	
159	V	->	I	22,0%	9,3%
160	E	->	K		19,3%
161	GAG	->	GAA	21,7%	
162	AAG	->	AAA		5,2%
163	G	->	A	21,4%	24,4%
164	TTC	->	TTT	15,5%	
165	N	->	D	0,6%	
165	N	->	S	21,7%	24,9%
166	CCA	->	CCG	21,6%	
167	E	->	G	1,5%	
168	V	->	T		22,9%
170	CCC	->	CCT	8,7%	
173	S	->	T	37,7%	74,3%
242	N	->	T	21,3%	
246	CAA	->	CAG	14,7%	
247	I	->	V	31,6%	
251	ACA	->	ACC		14,9%
252	N	->	S	32,3%	
295	D	->	G	1,5%	
295	GAC	->	GAT	4,0%	69,8%
298	GAC	->	GAT	20,7%	
299	CGG	->	AGG	20,5%	
301	TAT	->	TAC	5,8%	
303	T	->	A		0,9%
306	GCC	->	GCT		40,3%
307	E	->	K		10,8%
309	GCT	->	GCC		29,1%
310	S	->	T	17,1%	
312	E	->	K		12,1%
313	V	->	I	0,8%	
314	AAA	->	AAG	15,8%	
315	N	->	S	0,7%	

318	T	->	A	0,9%	
319	E	->	G	0,9%	
319	GAA	->	GAG		25,5%
326	TCG	->	TCA		7,1%
327	AAC	->	AAT	17,5%	
329	GAT	->	GAC	47,0%	70%
331	AAG	->	AAA		11,5%
334	TTA	->	CTA		1,6%
334	TTA	->	TTG	21,8%	
335	K	->	R	49,2%	
337	TTA	->	TTG	0,9%	60%
338	GGA	->	GGG	24,5%	
340	GGA	->	GGG		0,5%

8.2. TABELA DE VARIAÇÕES NO GENE NEF

Tabela 10. Mutações observadas em *Nef* e suas frequências em V1 e VX

Indivíduo	Códon	Mutação	Frequência	
			V1	VX
EC02	186	D -> G	0,8%	
	187	S -> G		0,8%
	196	CGA -> CGG		0,5%
	197	E -> G		0,5%
EC08	70	GUC -> GUA		1,1%
	71	T -> R	0,7%	
	71	T -> I		0,6%
	74	GUA -> GUG		5,5%
	80	T -> A		0,7%
	80	ACU -> ACC		0,6%
	85	V -> L		1,6%
	89	H -> R	0,7%	
	92	AAA -> AAG	0,6%	
	117	T -> A		0,9%
	121	F -> L	0,6%	
	125	Q -> H		0,6%
	127	Y -> C	0,9%	
	137	CUG -> CUT		0,7%
	143	Y -> L		0,6%
	147	P -> Q		0,5%
148	V -> L		0,7%	
ECB09	137	CUG -> CUT		0,8%
	138	ACC -> ACA		0,8%
	141	W -> L		0,6%
	147	P -> Q		0,6%
	186	D -> Y		0,8%
	186	D -> G		0,5%
	192	H -> N		0,5%
	196	CGA -> AGA		1,2%
	197	E -> G		0,6%
200	P -> Q		0,7%	
ECB11	72	CCU -> CCC		11,7%
	73	CAG -> CAA	3,2%	
	77	AGA -> AGG		0,6%
	82	K -> R		0,5%
	92	AAA -> AAG		0,5%
	98	GAA -> GAG	0,5%	

	103	S -> P	0,5%	
	106	R -> G	0,5%	
	106	AGA -> AGG	0,6%	
	114	I -> A	0,6%	
ECB13	17	R -> K	99%	9,4%
	19	R -> G		0,8%
	20	M -> I	99%	9%
	28	D -> A	99%	8,9%
	29	G -> K	99,5%	8,7%
	115	Y -> H	1,7%	
	115	UAC -> UAT	0,5%	
	116	H -> N	0,6%	
	116	H -> R		0,6%
	117	T -> A		0,5%
	126	N -> S	10,0%	99,2%
	128	ACA -> ACC	0,5%	
	128	ACA -> ACT	99,1%	
	129	CCA -> CCT		2,2%
	135	Y -> F	99,2%	10,4%
	138	T -> A		0,6%
	140	G -> R		88,5%
	140	GGA -> GGG	99,1%	10,7%
	143	Y -> L		0,6%
	148	V -> I		0,7%
	181	UUA -> CUA		3,8%
	185	F -> L	0,7%	
	186	D -> G		0,5%
	188	R -> S	25,1%	
	189	CUA -> CUG	99,4%	9,4%
	191	F -> S		0,7%
	196	CGA -> CGG	0,7%	
	197	E -> G		0,5%
	198	L -> K	98,7%	0,8%
	200	CCG -> CCA		91,0%
ECB18	79	M -> I		0,6%
	87	L -> P	1,2%	
	92	K -> R		68,5%
	92	AAA -> AAG	0,9%	
	96	G -> R		6,9%
	98	GAA -> GAG	0,6%	
	100	CUA -> CUG		1,2%
	103	S -> P	0,5%	
	108	D -> G	0,5%	

	117	ACA -> ACG	0,5%	
VC06	19	R -> G		0,6%
	20	M -> T		1,0%
	21	R -> K		14,2%
	21	R -> T		74,2%
	28	D -> G		0,6%
	30	V -> A		1,1%
	114	I -> V		60,3%
	117	T -> A		0,6%
	128	T -> A		0,6%
	133	V -> A		0,5%
	134	AGA -> AGG		0,9%
	137	L -> P		0,6%
	145	L -> P		0,6%
	145	CUA -> CUG		0,5%
	186	D -> G		0,6%
	190	GCA -> GCG		0,7%
191	F -> S		0,6%	
194	V -> M		13,7%	
VC10	187	S -> R		0,6%
	188	G -> R	0,2%	8,4%
	189	L -> P		0,5%
	192	H -> Q		1,5%
	194	M -> V	0,7%	0,5%
	195	GCG -> GCA	0,5%	
	196	CGA -> AGA	0,5%	
	197	E -> L	1,5%	
	197	E -> K		0,5%
	197	E -> V	0,5%	
	197	E -> G		0,5%
	199	H -> T	2,1%	
VC14	68	F -> L	0,5%	0,5%
	69	P -> T	0,6%	
	70	GTC -> GTA	2,8%	
	70	V -> A		0,6%
	71	R -> T	4,6%	97%
	71	R -> I	0,9%	
	74	GTG -> GTA	5,2%	97,6%
	77	AGG -> AGA	5,2%	
	78	P -> Q	0,7%	97,6%
	79	M -> I	2,7%	
	80	ACT -> ACC		0,8%
	80	ACT -> ACG		39,5%

	81	TAT	->	TAC		1,0%
	82	K	->	N	0,5%	
	85	V	->	L		0,5%
	86	D	->	Y	0,7%	
	125	Q	->	H	1,4%	
	129	P	->	Q	5,7%	96,7%
	131	CCA	->	CCC	0,9%	1,2%
	133	V	->	I	5,2%	96,5%
	133	GTC	->	GTA	0,8%	
	133	GTC	->	GTT	3,4%	
	133	I	->	T		0,5%
	135	TAC	->	TAT	5,6%	97,4%
	136	P	->	Q	0,6%	
	137	CTG	->	TTG		0,7%
	138	C	->	R	0,8%	
	138	TGT	->	TGC	4,1%	2,6%
	138	C	->	T		0,8%
	140	G	->	V	0,5%	
	79	M	->	I		0,5%
	80	ACU	->	ACC	37,3%	6,3%
	81	UAC	->	UAT		6,7%
	85	V	->	L	66,8%	99%
	85	GUA	->	GUG	32,9%	
	93	E	->	G		0,5%
	116	H	->	D		0,8%
	119	GGC	->	GGT	67,5%	54,9%
	123	GAU	->	GAC		0,7%
	125	Q	->	R		1,7%
	125	CAG	->	CAA		1,6%
VC15	126	AAC	->	AAT	39,9%	39,7%
	127	Y	->	C		0,5%
	129	CCA	->	CCG		1,2%
	133	CCC	->	CCT	23%	51%
	133	GUC	->	GUT		0,9%
	134	AGA	->	AGG		0,6%
	137	CUG	->	TUG	40,6%	12,1%
	139	UUU	->	UUC		1,4%
	140	GGA	->	GGC	38,5%	34,7%
	143	Y(UAC)	->	F(UTT)	0,8%	2,7%
	145	CUA	->	CUG	1,8%	
	148	GUU	->	GUA		2,9%
VC16	81	Y	->	F	54,3%	47,6%
	84	GCU	->	GCC		0,5%

89	H -> Y	29,3%	
92	K -> R	19,0%	
102	H -> Y	99,9%	48,6%
105	AAA -> AAG	31,6%	32,6%
106	AGA -> AGG	1,6%	
111	D -> G	1,9%	
116	H -> N	63,4%	63,1%
117	T -> A	0,6%	
121	F -> L	0,5%	
133	T -> V		13,3%
133	T -> I		74,3%
134	AGA -> AGG		1,0%
135	Y -> H	14,0%	
135	UAU -> UAC		75,7%
136	CCA -> CCG		15,3%
136	CCA -> CCT		11,1%
143	Y -> L		0,6%