



**Ministério da Saúde  
Instituto Nacional de Câncer  
Coordenação de Pós-graduação**

**INSTITUTO NACIONAL DE CÂNCER**  
**Pós-Graduação em Oncologia**

*PAULA DANIELA SOUZA ALVES*

Diversidade genética do vírus Epstein-Barr entre infecções naturais,  
linfoproliferações benignas e linfomas

**Orientador (es):** Prof<sup>a</sup>. Dr<sup>a</sup>. Rocio Hassan (*in memoriam*)  
Prof<sup>a</sup>. Dr<sup>a</sup>. Eliana Abdelhay  
Prof. Dr. Edson Delatorre

**RIO DE JANEIRO**  
**2021**



**Ministério da Saúde  
Instituto Nacional de Câncer  
Coordenação de Pós-graduação**

**INSTITUTO NACIONAL DE CÂNCER**  
**Pós-Graduação em Oncologia**

*PAULA DANIELA SOUZA ALVES*

Diversidade genética do vírus Epstein-Barr entre infecções naturais,  
linfoproliferações benignas e linfomas

Tese apresentada ao Instituto Nacional de Câncer  
como parte dos requisitos para obtenção do título  
de Doutor em Oncologia

**Orientador (es):** Prof<sup>a</sup>. Dr<sup>a</sup>. Rocio Hassan (*in memoriam*)  
Prof<sup>a</sup>. Dr<sup>a</sup>. Eliana Abdelhay  
Prof. Dr. Edson Delatorre

**RIO DE JANEIRO**  
**2021**

A474d Alves, Paula Daniela Souza

Diversidade genética do vírus Epstein-Barr entre infecções naturais, linfoproliferações benignas e linfomas / Paula Daniela Souza Alves. – Rio de Janeiro, 2021.  
115 f. : il. Color.

Tese (Doutorado em Oncologia) – Programa de Pós-Graduação em Oncologia, Instituto Nacional de Câncer José Alencar Gomes da Silva, 2021.

Orientadores: Profa. Dra. Rocio Hassan (in memoriam)  
Profa. Dra. Eliana Abdelhay  
Prof. Dr. Edson Delatorre

1. Vírus Epstein-Barr. 2. Diversidade Genética. 3. Oncovírus. 4. Linfoma. I. Rocio Hassan (Orient.). II. Abdelhay, Eliana (Orient.). III. Delatorre, Edson (Orient.). IV. Instituto Nacional de Câncer José Alencar Gomes da Silva. V. Título.

CDD 616.9112

Catálogo na fonte  
Kátia Simões CRB7/5952



Ministério da Saúde  
Instituto Nacional de Câncer  
Coordenação de Pós-graduação

**INSTITUTO NACIONAL DE CÂNCER**

**Pós-Graduação em Oncologia**

*PAULA DANIELA SOUZA ALVES*

Diversidade genética do vírus Epstein-Barr entre infecções naturais,  
linfoproliferações benignas e linfomas

**ORIENTADOR (ES):** Prof<sup>a</sup>. Dr<sup>a</sup>. Rocio Hassan (*in memoriam*)

Prof<sup>a</sup>. Dr<sup>a</sup>. Eliana Abdelhay

Prof. Dr. Edson Delatorre

Aprovada em: \_\_\_\_/\_\_\_\_/\_\_\_\_

**EXAMINADORES:**

**Prof. Dr. Martin Hernan Bonamino**

Instituto Nacional de Câncer (INCA)

**Prof<sup>a</sup>. Dr<sup>a</sup>. Tânia Regina Tozetto Mendoza**

Universidade de São Paulo (USP)

**Prof<sup>a</sup>. Dr<sup>a</sup>. Sheila Coelho Soares Lima**

Instituto Nacional de Câncer (INCA)

**Prof. Dr. Robson de Queiroz Monteiro**

Universidade Federal do Rio de Janeiro (UFRJ)

**Prof. Dr. Marcelo Alves Soares** – Suplente

Instituto Nacional de Câncer (INCA)

**Prof<sup>a</sup>. Dr<sup>a</sup>. Vanessa Salette de Paula** – Suplente

Fundação Oswaldo Cruz (FIOCRUZ)

**RIO DE JANEIRO**

**2021**



**Ministério da Saúde  
Instituto Nacional de Câncer  
Coordenação de Pós-graduação**

## **INSTITUTO NACIONAL DE CÂNCER**

### **DIVERSIDADE GENÉTICA DO VÍRUS EPSTEIN-BARR ENTRE INFECÇÕES NATURAIS, LINFOPROLIFERAÇÕES BENIGNAS E LINFOMAS**

#### **RESUMO**

#### **TESE DE DOUTORADO**

**Paula Daniela Souza Alves**

O vírus Epstein-Barr (EBV) é um  $\gamma$ -Herpesvírus transmitido pela saliva e que infecta assintomaticamente >90% da população adulta mundial. Entretanto, o EBV é associado etiologicamente ao desenvolvimento de linfoproliferações benignas e malignas. O EBV possui um genoma dsDNA com ~170kb e se divide em dois tipos principais com propriedades biológicas diferentes: EBV-1 e EBV-2. Os mecanismos carcinogênicos mediados pelo EBV não são compreendidos em sua totalidade, e postula-se que variantes oncogênicas virais podem atuar nesses mecanismos. A incidência dos cânceres associados ao EBV apresenta grande variação geográfica, podendo estar relacionada a fatores sociogeográficos, étnicos e genéticos, sendo importante diferenciar fatores virais restritos geograficamente daqueles associados a processos tumorigênicos específicos. Neste trabalho, caracterizamos a diversidade do EBV na população do Brasil (BR), com o intuito de entender seu papel no contexto da epidemia Sul-Americana. Foram analisadas 83 amostras EBV+, incluindo carreadores assintomáticos (CA, 54,2%) e patologias benignas e malignas (PB e PM respectivamente, 45,8%). A classificação através de assinaturas aminoacídicas da região codificante (cfs), de LMP1 indicou a presença de cinco diferentes cepas, com uma prevalência de 40% Med/AG876/Daudi (Med), 39% Raji/Argentine (Raji), 6% B95-8, 6% China1/CAO e 1% China2. Análises evolutivas demonstram que 79% das amostras se agruparam em dois clados com alto suporte: 40% relacionadas à variante Med (aLRT=0.95) e 39% relacionadas à variante Raji (aLRT=0.92). De forma surpreendente, as sequências oriundas dos CA e PB agruparam-se majoritariamente (61%) no clado Med, enquanto sequências relacionadas às patologias malignas agruparam-se em sua maioria (66,7%) no clado Raji,  $p=0,041$ . Polimorfismos da LMP1 de importância oncogênica (1124L/1152L, ins15pb e del30pb) foram mais frequentes no clado Raji do que no clado Med ( $p<0,0001$  para todos os polimorfismos). Além disso, foi observado uma frequência significativamente maior ( $p=0,02$ ) de cepas tipo1/V3 com alto poder replicativo no clado Raji ( $p=0,02$ ). Para colocar as novas sequências de LMP1 do EBV do BR em um contexto global, adicionamos 252 sequências do LMP1 disponíveis publicamente nas análises filogenéticas por máxima verossimilhança. As cepas brasileiras se agruparam com outras da América do Sul, caracterizando as linhagens Raji/Argentine (BR=35, Argentina (AR)=33; aLRT=1) e Med [BR=33, AR =11; aLRT=0.91] como as principais em circulação na América do Sul, reunindo 71,9% das sequências. Através de análises filogeográficas bayesianas, reconstruímos a dinâmica das migrações virais do clado Raji/Argentine, aqui redefinido como "South American". A filogenia mostrou que a linhagem derivada de Raji circulante na América do Sul se originou através de uma introdução viral proveniente do continente Africano para o Brasil entre os séculos XVII e XIX, e que posteriormente se disseminou para outros países da região (AR e Peru) e para a Europa. Nossos resultados demonstram uma grande diversidade genética do EBV circulante no Brasil, além de apontar a existência da associação entre mutações no *LMP1* presentes na variante Raji do BR com patologias malignas. O Brasil parece ter sido a origem e principal fonte de disseminação da variante Raji Sul-Americana derivada da cepa africana Raji. Em conjunto, esses dados destacam a grande diversidade genética do EBV, e a importância do monitoramento e avaliação das mutações no potencial patogênico das diferentes cepas.



**Ministério da Saúde  
Instituto Nacional de Câncer  
Coordenação de Pós-graduação**

**INSTITUTO NACIONAL DE CÂNCER**

**GENETIC DIVERSITY OF THE EPSTEIN-BARR VIRUS BETWEEN NATURAL  
INFECTIONS, BENIGNAL LYMPHOPROLIFERATIONS AND LYMPHOMAS**

**ABSTRACT**

**DOCTORAL THESIS**

**Paula Daniela Souza Alves**

Epstein-Barr virus (EBV) is a saliva-borne  $\gamma$ -Herpesvirus that asymptotically infects >90% of the world's adult population. However, EBV is etiologically associated with development of benign and malignant lymphoproliferations. EBV own a dsDNA genome nearly 170kb and is divided into two main types with different biological properties: EBV-1 and EBV-2. EBV-mediated tumorigenic mechanisms are not fully understood, and it is postulated that viral oncogenic variants may act on these mechanisms. The incidence of EBV-associated cancers varies widely and may be related to sociogeographic, ethnic and genetic factors, thus, it is important to differentiate geographically restricted viral factors from those specific tumorigenic-associated processes. In this work, we characterize the diversity of EBV circulating in Brazil (BR), in order to understand its role in the context of the South American epidemic. Eighty-three EBV+ samples were analyzed, including asymptomatic carriers (AC, 54.2%) and benign and malignant pathologies (BP and MP respectively, 45.8%). Classification by amino acid signatures of the coding region (cdr) of LMP1 oncoprotein indicated the presence of five different strains, with a prevalence of 40% Med/AG876/Daudi (Med), 39% Raji/Argentine (Raji), 6% B95-8, 6% China1/CAO and 1% China2. Evolutionary analyzes demonstrate that 79% of the samples clustered into two clades with high support: 40% related to the variant Med (aLRT=0.95) and 39% related to the variant Raji (aLRT=0.92). Surprisingly, sequences from AC and BP were mostly grouped (61%) in the Med clade, while sequences related to MP were mostly grouped (66.7%) in the clade Raji,  $p=0.041$ . LMP1 polymorphisms of oncogenic importance (I124L/I152L, ins15bp and del30bp) were more frequent among Raji clade than Med clade ( $p<0.0001$  for all polymorphisms). In addition, a significantly higher frequency of type1/V3 haplotypes, known to strongly increase lytic replication was observed in Raji clade ( $p=0.02$ ). In order to place our Brazilian LMP1 sequences in a global context, we added 252 publicly available LMP1 sequences in the maximum likelihood phylogenetic analyses. The Brazilian strains were grouped with others from South America, characterizing the Raji/Argentine (BR=35, Argentina (AR)=33; aLRT=1) and Med [BR=33, AR =11; aLRT=0.91] strains as the main ones in circulation in South America, gathering 71.9% of the sequences. Through Bayesian phylogeographic analyses, we reconstruct the dynamics of viral migrations of the Raji/Argentine clade, here redefined as South American. The phylogeny showed that the Raji-derived strain circulating in South America originated through a viral introduction from the African continent to Brazil between the 17th and 19th centuries, and which later spread to other countries in the region (AR and Peru) and to the Europe. Our results demonstrate a great genetic diversity of EBV circulating in Brazil, in addition to pointing out the existence of an association between mutations in LMP1 present in the Raji variant of BR with malignant pathologies. Brazil seems to have been the origin and main source of dissemination of the South American Raji variant derived from the African Raji strain. Together, these data highlight the great genetic EBV diversity, and the importance of monitoring and evaluating mutations in the pathogenic potential of the different strains.

“(...) aulas em que contava a história do Brasil,  
em que falava da mistura entre índios, negros e brancos, de como éramos felizes,  
de como nosso país era abençoado.”

- *Itamar Vieira Junior, Torto Arado*

"Nessa terra a dor é grande e a ambição pequena  
Carnaval e futebol  
Quem não finge, quem não mente  
Quem mais goza e pena  
É que serve de farol

Existe alguém em nós  
Em muito dentre nós esse alguém  
Que brilha mais do que milhões de sóis  
E que a escuridão conhece também  
Existe alguém aqui"

- *Caetano Veloso, A Luz de Tieta*

*À Claudia Esther Alicia Rocio Hassan.  
Que dentre tantos ensinamentos, me ensinou a amar o Brasil.*



## AGRADECIMENTOS

*Ao pessoal do Laboratório de Oncovirologia do CEMO:* Vanessa Emmel, Bianca Gama, Alessandra Teixeira, Regina Vieira e Paulo Rohan pela amizade e suporte a todo momento. Em especial à Marcella Larrate, cujo companheirismo foi fundamental durante essa tese.

*Àqueles que vieram antes de mim:* Gustavo Stefanoff, Mário Barros, Priscilla Segges, Maria Gabriela Lozada e Aruanã Garcia Costa. Que sem o árduo trabalho e dedicação, o atual trabalho não teria se desenvolvido até aqui.

*Àqueles que passaram por aqui:* Mariana Bizarro, Raquel Curtinhas, Elielson Veloso, Thiago Lima, Larissa Duarte, Flávia Krsticevic, Elizabeth Solorzano, Pavel Medina. Convivências que foram fundamentais para a minha formação.

*Aos professores das disciplinas de pós-graduação do INCa:* Luís Felipe Ribeiro, Sheila Coelho, Martin Bonamino, João Viola, Eliana Abdelhay, Stephany Correa, Renata Binato, Hector Nicolas, Leo Karam, Luiz Thuler, entre outros. Que me proporcionaram ricas discussões e me introduziram mais profundamente no conhecimento do câncer.

*Às colaborações científicas:* Luciana Britto (Hematologia – HCI/INCA), Luiz Thuler (CPq/INCA). Elizabeth Tapia, Joaquin Ezpeleta, Javier Murillo e Paolo Cacchiarelli da Universidade Nacional de Rosário. Em especial ao Edson Delatorre (UFES). As quais os ensinamentos foram essenciais para o desenvolvimento desse trabalho.

*Às pessoas que ajudaram na amostragem do estudo:*

Pacientes e profissionais do INCa, pela paciência e suporte. Organizadores do Curso de Verão do INCA 2019/20 e alunos(as) do Curso de Verão do INCA 2019/20. Colegas do CEMO, HCI e CPq - INCA, em especial: Everton Cruz. Aos voluntários que doaram saliva na UFRJ, em especial, integrantes do laboratório de Neuroplasticidade. Ao coordenador do curso de Biomedicina da UNIGRANRIO: Daniel Pereira Reynaldo da Unidade Nova Iguaçu e Caxias, que me proporcionou divulgar meu trabalho e a coleta de saliva de voluntários. Ao Pavilhão 26 e laboratório de Patologia, ambos do IOC-FIOCRUZ, em especial à Suwellen Azevedo e Jessica Silva. Enfim, às outras mais de 500 pessoas que doaram saliva e/ou apoiaram o presente trabalho: família, amigos, colegas de instituições científicas e desconhecidos. Meu grande agradecimento, pelo apoio e carinho à pesquisa científica em tempos tão difíceis para a ciência brasileira.

*Às líderes e aos trabalhadores do CEMO-INCA:* Pelo suporte e apoio em todo momento: Virgínia, Vânia, Karen, Moisés, Telminha, Gerson, entre outros. E Thais Hancio e Laura Robledo pelo suporte e alegria nas horas difíceis. Em especial à Eliana Abdelhay.

Agradecer não é o suficiente à Eliana Abdelhay e Edson Delatorre, que assumiram muito mais que suas responsabilidades durante os momentos mais difíceis do desenvolvimento dessa tese. Mas ainda assim, meu muito obrigada pela enorme paciência e o apoio sempre com muito carinho.

Por último e não menos importante, à pessoa que é responsável por esta tese existir, à minha querida **Rocio**, meu eterno agradecimento por ter tido a oportunidade e privilégio de estar em sua companhia.

## INDICE

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>1</b>
<b>1.1</b>	<b>Oncogênese mediada por vírus</b> .....	<b>1</b>
<b>1.2</b>	<b>O vírus Epstein-Barr</b> .....	<b>4</b>
1.2.1	<u>Estrutura</u> .....	6
1.2.2	<u>Ciclo replicativo</u> .....	7
1.2.3	<u>História natural da infecção</u> .....	9
1.2.4	<u>Diversidade genética do vírus Epstein-Barr</u> .....	10
1.2.4.1	A oncoproteína LMP1.....	15
1.2.5	<u>Neoplasias associadas ao EBV</u> .....	22
<b>2</b>	<b>OBJETIVOS</b> .....	<b>26</b>
<b>2.1</b>	<b>Objetivo geral</b> .....	<b>26</b>
<b>2.2</b>	<b>Objetivos secundários</b> .....	<b>26</b>
<b>3</b>	<b>METODOLOGIA</b> .....	<b>27</b>
<b>3.1</b>	<b>Desenho do estudo</b> .....	<b>27</b>
<b>3.2</b>	<b>Descrição da população estudada</b> .....	<b>28</b>
3.2.1	<u>Casos de doenças hematológicas EBV+</u> .....	28
3.2.2	<u>Doador assintomático</u> .....	28
3.2.3	<u>Dados clínicos e epidemiológicos</u> .....	29
3.2.3.1	Casos diagnosticados no INCA.....	29
3.2.3.2	Doadores assintomáticos.....	29
3.2.4	<u>Crítérios Éticos</u> .....	29
<b>3.3</b>	<b>Métodos para a extração de ácidos nucleicos</b> .....	<b>30</b>
3.3.1	<u>Extração de DNA a partir de tecido à fresco</u> .....	30
3.3.2	<u>Extração de DNA de saliva pelo método de "salting out"</u> .....	30
3.3.3	<u>Isolamento de células mononucleares a partir do sangue periférico e extração de DNA</u> .....	31
3.3.4	<u>Avaliação qualitativa e quantitativa do DNA extraído</u> .....	31
3.3.5	<u>Avaliação da amplificabilidade do DNA através de PCR para gene constitutivo</u> .....	31
<b>3.4</b>	<b>Métodos para a caracterização da associação com o EBV</b> .....	<b>32</b>
3.4.1	<u>Detecção e tipificação do EBV por PCR para o gene EBNA3C</u> .....	32
3.4.2	<u>Quantificação da carga viral do EBV por PCR quantitativo (Real-Time PCR)</u> .....	32
<b>3.5</b>	<b>Métodos para a análise molecular de polimorfismos do EBV</b> .....	<b>33</b>
3.5.1	<u>Amplificação do oncogene latente BNLF1 (LMP1)</u> .....	33

3.5.2 <u>Amplificação do promotor lítico do gene Z (Zp)</u> .....	34
3.5.3 <u>Purificação dos produtos de PCR</u> .....	34
3.5.4 <u>Sequenciamento pelo método de Sanger</u> .....	35
3.5.5 <u>Sequenciamento de nova geração para obtenção do genoma do EBV</u> .....	35
3.5.5.1 <u>Amplificação de gaps resultantes do NGS por sequenciamento de Sanger</u> ....	36
<b>3.6 Análise bioinformática</b> .....	<b>38</b>
3.6.1 <u>Análise de sequências geradas pelo método de Sanger</u> .....	38
3.6.1.1 <u>Reconstruções filogenéticas e filogeográficas da região codificante do gene LMP1</u> .....	38
3.6.1.2 <u>Caracterização de polimorfismos do promotor lítico Zp</u> .....	40
3.6.2 <u>Análise de sequências geradas por NGS</u> .....	40
<b>3.7 Análises estatísticas</b> .....	<b>41</b>
<b>4. RESULTADOS</b> .....	<b>41</b>
<b>4.1 Diversidade genética do EBV circulante no Brasil</b> .....	<b>41</b>
4.1.1 <u>Características da população estudada</u> .....	41
4.1.2 <u>Diversidade do EBV pelo oncogene LMP1</u> .....	43
4.1.2.1 <u>Características da população com sequência inteira do oncogene LMP1</u> .....	44
4.1.2.2 <u>Classificação de polimorfismos da LMP1</u> .....	45
4.1.2.3 <u>Diversidade e evolução da oncoproteína LMP1 na América do Sul</u> .....	53
4.1.2.3.1 <u>Relação dos polimorfismos genéticos com as variantes classificadas pela análise filogenética</u> .....	55
4.1.2.3.2 <u>Relação dos polimorfismos com as duas cepas principais circulantes no Brasil</u> .....	57
4.1.2.3.3 <u>Divergência entre as duas classificações aplicadas para a LMP1</u> .....	60
4.1.2.4 <u>Origem e diversidade do EBV circulante na América do Sul através do oncogene LMP1</u> .....	60
4.1.3 <u>Associação da variante Tipo1/V3 com as variantes da oncoproteína LMP1</u> .....	66
<b>4.2 Genoma parcial do EBV do subclado South American isolado de um linfoma de Burkitt</b> .....	<b>67</b>
<b>5 DISCUSSÃO</b> .....	<b>71</b>
5.1 <u>Diversidade do EBV circulante no Brasil pelo oncogene LMP1</u> .....	69
5.2 <u>Reavaliando a origem da variante Raji Sul-Americana</u> .....	75
<b>6 CONCLUSÕES</b> .....	<b>78</b>
<b>7 REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>79</b>
<b>ANEXO I. TABELAS SUPLEMENTARES</b> .....	<b>94</b>

<b>ANEXO II. PARECER DO COMITÊ DE ÉTICA EM PESQUISA.....</b>	<b>113</b>
<b>ANEXO III. QUESTIONÁRIO SOCIOECONÔMICO.....</b>	<b>114</b>

## LISTA DE TABELAS

<b>Tabela 1.1</b>	Oncovírus humanos. Ano de descoberta, família viral, tipo e tamanho do material genético, principais oncogenes, tropismo celular e neoplasias associadas.	3
<b>Tabela 1.2</b>	Perfis de latência e expressão gênica identificados em neoplasias associadas ao EBV.	5
<b>Tabela 1.3</b>	Variações aminoacídicas em sequências inteiras da LMP1 que definem protótipos (retirado de Edwards <i>et al</i> , 1999).	17
<b>Tabela 3.1</b>	Sequência dos primers utilizados para a amplificação dos gaps no genoma montado por NGS.	37
<b>Tabela 3.2</b>	Lista de sequências de referência utilizadas na reconstrução filogenética da região codificante da LMP1.	39
<b>Tabela 3.3</b>	Local e período da amostragem das sequências utilizadas nas análises filogenética mundial e filogeográfica.	40
<b>Tabela 4.1</b>	Características da população com sequência inteira da LMP1.	44
<b>Tabela 4.2</b>	Polimorfismos e classificação das variantes da oncoproteína LMP1.	47
<b>Tabela 4.3</b>	Relação dos polimorfismos da oncoproteína LMP1 com os casos investigados.	52
<b>Tabela 4.4</b>	Relação das variantes da oncoproteína LMP1 com a população estudada.	55
<b>Tabela 4.5</b>	Relação dos polimorfismos da oncoproteína LMP1 com o grupo DA+CB de variante Med/AG876/Daudi e o grupo CM de variante Raji/Argentine.	58

## LISTA DE TABELAS (continuação)

<b>Tabela 4.6</b>	Relação da classificação do Zp com as variantes da LMP1 estudadas.	66
<b>Tabela 4.7</b>	Relações haplotípicas de genomas brasileiros classificados como variante “South American” pela LMP1.	70
<b>Tabela S1</b>	Características sociodemográficas de carreadores assintomáticos.	94
<b>Tabela S2</b>	Carga viral dos casos de patologias hematológicas EBV+.	97
<b>Tabela S3</b>	Associação de fatores sociodemográficos com a detecção do EBV na 1ª amostra coletada da população sadia.	97
<b>Tabela S4</b>	Associação de fatores sociodemográficos com o tipo 1 ou 2 do EBV na população sadia.	102
<b>Tabela S5</b>	Clados do gene <i>LMP1</i> definidos pela filogenia de Máxima verossimilhança.	106
<b>Tabela S6</b>	Relação entre os polimorfismos e as variantes da oncoproteína LMP1.	107
<b>Tabela S7</b>	Relação dos polimorfismos da oncoproteína LMP1 com as variantes Med/AG876/Daudi e Raji/Argentine.	109
<b>Tabela S8</b>	Frequência das variantes do EBV pelo oncogene <i>LMP1</i> pelas classificações por Edwards <i>et al.</i> e filogenética.	110
<b>Tabela S9</b>	Comparação das classificações da LMP1 por Edwards <i>et al.</i> e filogenética.	111

## LISTA DE FIGURAS

<b>Figura 1.1</b>	Estrutura da partícula viral do EBV.	6
<b>Figura 1.2</b>	Esquema do ciclo replicativo do vírus Epstein-Barr.	8
<b>Figura 1.3</b>	Árvore de máxima parcimônia para o gene <i>UL5</i> de alguns membros da família herpesviridea.	11
<b>Figura 1.4</b>	Esquema da posição de alguns genes e miRNA virais na forma linear do genoma do EBV.	13
<b>Figura 1.5</b>	Esquema da posição de alguns genes, RNA não codificantes e promotores na forma circular (ou episomal) do genoma do EBV.	14
<b>Figura 1.6</b>	Árvore filogenética formada pelo método de <i>neighbor-joining</i> a partir de sequências nucleotídicas que definiram protótipos da LMP1 por Edwards <i>et al.</i>	18
<b>Figura 1.7</b>	Esquema aminoacídico da LMP1 do EBV.	19
<b>Figura 1.8</b>	Esquema da ativação de múltiplas vias de sinalização celular orquestradas pela LMP1.	21
<b>Figura 3.1</b>	Diagrama descritivo do estudo da diversidade molecular do EBV circulante no Brasil.	27
<b>Figura 3.2</b>	Esquema da região codificante do gene LMP1 mostrando a estratégia de amplificação e a localização dos primers.	33
<b>Figura 4.1</b>	Resultado ilustrativo da eletroforese em gel de agarose dos produtos da PCR para o gene viral <i>EBNA3C</i> e o gene da proteína constitutiva beta-globina.	42

## LISTA DE FIGURAS (continuação)

- Figura 4.2** Frequência das variantes do LMP1 classificadas pelo método de Edwards *et al.* (1999) nas amostras estudadas. 45
- Figura 4.3** Árvore filogenética de máxima verossimilhança das sequências completas da oncoproteína LMP1 de casos e carreadores assintomáticos (em preto). 53
- Figura 4.4** Frequência das variantes do LMP1 classificadas por filogenia de máxima verossimilhança nas amostras estudadas. 55
- Figura 4.5** Frequência dos polimorfismos genéticos da LMP1 de importância oncológica presente nos clados. 57
- Figura 4.6** Frequência dos polimorfismos genéticos da LMP1 de importância oncológica nos dois clados principais. 58
- Figura 4.7** Relação dos polimorfismos nos dois principais clados com o desfecho clínico. 59
- Figura 4.8** Árvore filogenética de máxima verossimilhança do gene LMP1 das amostras brasileiras deste estudo (n = 83) e as disponíveis em bancos de dados (n = 252). 62
- Figura 4.9** Frequência do país de origem das sequências do *LMP1* pertencentes aos dois principais clados da América do Sul. 63
- Figura 4.10** Filogeografia da variante Raji/Argentine circulante na América do Sul. 64
- Figura 4.11** Esquema demonstrando as migrações virais e origem da linhagem Raji/Argentine. 65
- Figura 4.12** Identificação do haplótipo Tipo1/V3 nos clados da LMP1 do presente estudo. 67



<b>Figura 4.13</b>	Células do linfoma de Burkitt do isolado CEMO3.	68
<b>Figura 4.14</b>	Esquema do genoma parcial CEMO3.	69
<b>Figura S1</b>	Representação da classificação dos polimorfismos do Zp.	112

## LISTA DE ABREVIATURAS

°C	Graus Celsius
AIDS	Síndrome da imunodeficiência adquirida (do inglês: <i>acquired immunodeficiency syndrome</i> )
ANVISA	Agência Nacional de Vigilância Sanitária
BamHI-N	do inglês: <i>BamHI-N Left Fragment 1</i>
BARF	do inglês: <i>BamHI A rightward open reading frame</i>
BART	do inglês: <i>BamHI A rightward transcripts</i>
BCL-2	do inglês: <i>B-cell lymphoma 2</i>
BCR	Receptor de células B (do inglês: <i>B cell receptor</i> )
BHRF1	do inglês: <i>BamHI H fragment rightward open reading frame 1</i>
BRLF1	do inglês: <i>BamHI R fragment leftward open reading frame 1</i>
BZLF1	do inglês: <i>BamHI Z fragment leftward open reading frame 1</i>
CalHV3	<i>Callitrichine herpesvirus 3</i>
CCL17	do inglês: <i>chemokine C-C motif ligand 17</i>
CD	Conjunto de diferenciação (do inglês: <i>cluster of differentiation</i> )
CeHV15	<i>Cercophitecine herpesvirus 15</i>
CEMO	Centro de Transplante de Medula Óssea
CG	Centro germinativo
CM	Casos malignos
CSF1	Fator estimulante de colônias 1 (do inglês: <i>colony stimulating factor 1</i> )
CTAR	Regiões de ativação no carboxi-terminal (do inglês: <i>C-terminal activation regions</i> )
C-ter	Carboxi-terminal
DNA	Ácido desoxirribonucleico (do inglês: <i>deoxyribonucleic acid</i> )
DNTPs	Desorribonucleotídeos trifosfatados
DP	Depleção linfocitária
dsDNA	DNA linear de dupla fita
EBER	Pequeno RNA não codificantes pelo vírus Epstein-Barr (do inglês: <i>Epstein-Barr virus encoded small RNA</i> )
EBNA	Antígeno nuclear do Epstein-Barr (do inglês: <i>Epstein-Barr nuclear antigen</i> )
EBV	Vírus Epstein-Barr

EN	Esclerose Nodular
gB	Glicoproteína B
gH	Glicoproteína H
gp	Glicoproteína
GWA	Estudo de associação ampla do genoma (do inglês: <i>genome-wide association study</i> )
HBV	Vírus da hepatite B (do inglês: <i>hepatitis B virus</i> )
HCV	Vírus da hepatite C (do inglês: <i>hepatites C virus</i> )
HHV-4	Herpesvírus humano-4 (do inglês: <i>human herpesvírus 4</i> )
HHV-8	Herpesvírus humano-8 (do inglês: <i>human herpesvírus 8</i> )
HIV	Vírus da imunodeficiência humana (do inglês: <i>human immunodeficiency vírus</i> )
HLA	Antígeno leucocitário humano (do inglês: <i>human leukocyte antigen</i> )
HPV	Papilomavírus humano
HR	Hiperplasia reativa
H-RS	Hodgkin e Reed-Sternberg
HTLV-1	Vírus linfotrópico de células T humanas-1 (do inglês: <i>human T-cell lymphotropic virus type 1</i> )
ID2	Proteína inibidora de ligação ao DNA (do inglês: <i>inhibidor of DNA-binding</i> )
IDH	Índice de desenvolvimento humano
IL	Interleucina
INCA	Instituto Nacional de Câncer
IR	Repetições interna (do inglês: <i>internal repetition</i> )
ISH	Hibridização <i>in situ</i> (do inglês: <i>in situ hybridization</i> )
JAK3	Janus kinase 3
LB	Linfoma de Burkitt
LCL	Linfoblastoides
LCV	Linfocriptovirus
LH	Linfoma de Hodgkin
LHc	Linfoma de Hodgkin clássico
LHPLN	Linfoma de Hodgkin predomínio linfocítico nodular
LMP	Proteína latente de membrana (do inglês: <i>latent membrane protein</i> )
LNH	Linfoma não Hodgkin
LP	Proteína líder (do inglês: <i>leader protein</i> )

MAT	Microambiente tumoral
MCPvV	Poliomavírus das células de Merkel
MHC	Complexo principal de histocompatibilidade (do inglês: <i>major histocompatibility complex</i> )
miRNA	do inglês: <i>microRNA</i>
mL	Mililitro
mM	Milimolar
MNI	Mononucleose infecciosa
NF-κB	do inglês: <i>nuclear factor kappa-light-chain-enhancer of activated B cells</i>
NGS	Sequenciamento de nova geração (do inglês: <i>next generation sequencing</i> )
NK	Células exterminadoras naturais (do inglês: <i>natural killer cells</i> )
NTC	Controle sem amostra (do inglês: <i>no template control</i> )
N-Ter	Amino-terminal
OMS	Organização Mundial da Saúde
OR	Razão de chance (do inglês: <i>odds ratio</i> )
ORF	Fase de leitura aberta (do inglês: <i>open reading frame</i> )
<i>P</i>	Valor de <i>P</i>
pb	Pares de base
PBS	Tampão fosfato salino (do inglês: <i>phosphate-buffered saline</i> )
PCR	Reação em cadeia da polimerase (do inglês: <i>polymerase chain reaction</i> )
PE	Patologias hematológicas EBV+
PTLD	Doença linfoproliferativa pós-transplante (do inglês: <i>post-transplant lymphoproliferative disease</i> )
qPCR	PCR quantitativa
RL	Rico em linfócitos
RNA	Ácido ribonucleico (do inglês: <i>ribonucleic acid</i> )
rpm	Rotações por minuto
SNP	Polimorfismos de base única (do inglês: <i>single nucleotide polymorphism</i> )
SPSS	do inglês: <i>Statistical Package for Social Sciences</i>
Th	Células T auxiliares (do inglês: <i>T helper cells</i> )
TM	Transmembrana

TMA	do inglês: tissue microarray
TNFRSF	Superfamília do receptor de fator de necrose tumoral (do inglês: <i>tumor necrosis factor receptor superfamily</i> )
TNF- $\alpha$	Fator de necrose tumoral alfa (do inglês: <i>tumor necrosis factor alpha</i> )
TR	Repetições terminais (do inglês: <i>terminal repetition</i> )
TRADD	do inglês: <i>TNF-receptor type 1-associated death domain protein</i>
TRAF	do inglês: <i>TNF-receptor associated factors protein</i>

# 1 INTRODUÇÃO

## 1.1 Oncogênese mediada por vírus

Câncer é a segunda causa de mortalidade no mundo e no Brasil é estimado cerca de 600 mil novos casos por ano (INCA, 2021). Neoplasias surgem através do acúmulo de diversas alterações genéticas e epigenéticas no genoma do hospedeiro, de modo que uma complexa cadeia de eventos responsável por essas alterações seja capaz de alterar funções celulares de genes responsáveis por controlar a homeostase celular e tecidual. Caso a desregulação molecular da célula ocorra, e esta não reparar através do seu sistema de reparo ao dano de DNA ou através do sistema imune, o acúmulo dessas alterações pode levar o estágio de uma célula pré-maligna ao desenvolvimento do câncer. Assim, o câncer pode ter diferentes origens celulares, levando a patologias de fenótipos diversos, sendo definido atualmente como um conjunto de mais de 277 doenças, de base genética e natureza multifatorial (HASSANPOUR *et al*, 2017).

A oncogênese multifatorial deve-se a 5-10% dos casos de cânceres possuírem alterações genéticas herdadas, enquanto 90-95% tem fatores ambientais envolvidos (DE MARTEL *et al*, 2020). Dentre esses fatores ambientais, infecções parasitárias podem contribuir de 5 a 40% dos casos dependendo da localidade geográfica, fatores socioeconômicos, étnicos, culturais e genéticos (PLUMMER *et al*, 2016; DE MARTEL *et al*, 2020). No Brasil, 20% dos casos de câncer estão associados a infecções parasitárias, sendo 15% infecções virais.

Na infecção viral em indivíduos imunocompetentes, a replicação viral após a primo-infecção é geralmente diminuída ou ausente, com o vírus permanecendo no núcleo da célula como um ácido nucleico na forma episomal ou integrado ao genoma. A carcinogênese mediada por vírus não surge agudamente após uma infecção e pode levar anos ou décadas após a infecção primária para se desenvolver (MOORE; CHANG, 2010). Os vírus associados ao câncer, ou oncovírus, podem ser classificados usualmente como agentes carcinogênicos diretos ou indiretos. Os vírus que são definidos como oncogênicos diretos possuem na sua estrutura oncogenes virais que manipulam diretamente vias da transformação celular, enquanto os oncovírus indiretos causam primariamente inflamação crônica, que pode favorecer a transformação celular (GUVEN-MAIOROV *et al*, 2017).

Os oncovírus humanos têm características em comum, como a coevolução de longa data entre o vírus e o seu hospedeiro; são capazes de estabelecer uma infecção crônica e comumente assintomática ao longo da vida do hospedeiro (WALLACE &

GALLOWAY, 2016). Apesar dessa natureza em comum, são vírus que possuem genomas e ciclos replicativos completamente diferentes, assim como compreendem diversas famílias virais (Tabela 1.1). Até o momento, não se conhece oncovírus humanos que sempre induzam transformação neoplásica em seu ciclo de replicativo de maneira obrigatória. Desta forma, a carcinogênese mediada por oncovírus é resultado da seleção natural viral na tentativa de maximizar a sua persistência tecido específica, ao utilizarem estratégias envolvidas em vias das células hospedeiras durante seu ciclo replicativo, como favorecer a sobrevivência celular, controle da proliferação, indução do ciclo celular, reprogramação metabólica, evasão imune etc., ou seja, a consequência dos mecanismos virais orquestrados na célula hospedeira em conjunto com outros fatores, como pré-disposição genética, imunossupressão, exposição à carcinógenos, entre outros fatores, podem levar a alterações genéticas e epigenéticas celulares contribuindo para a instabilidade genética, crescimento celular, inflamação, entre outras características que favorecem o desenvolvimento do câncer (GRIVENNIKOV *et al*, 2010; HANAHAN & WEINBERG, 2011). Sendo assim, os oncovírus são apenas uma etapa na complexa cadeia de eventos que favorecem o desenvolvimento do câncer. Os oncovírus são necessários, mas não suficientes para causar o câncer.

O potencial oncogênico de alguns vírus está cada vez mais estabelecido na literatura, desta forma, muitos tipos de vírus têm sido alvos para prevenção ou tratamento do câncer. Uma vez que a imperícia das terapias tradicionais de câncer, como a radiação e a quimioterapia falham em distinguir células normais de células cancerígenas, podendo levar a altas toxicidades e desvantagens no tratamento. Assim, grandes esforços têm sido realizados para o desenvolvimento de novas estratégias que utilizam oncoproteínas virais como um alvo que possa diferir as células normais das neoplásicas. Outros estudos, podem ajudar a prevenir a infecção primária ou desenvolver terapias que possam melhorar a resposta imune contra as infecções virais, como o sucesso das vacinas do papilomavírus humano (HPV) e do vírus da hepatite B (HBV), dito isso, o estudo de oncovírus podem gerar estratégias efetivas para evitar o desenvolvimento ou a progressão tumoral.

Atualmente, através de evidências epidemiológicas e moleculares, foram identificados sete vírus oncogênicos humanos: cinco são vírus que têm como material genético o ácido desoxirribonucleico (DNA) e dois, o ácido ribonucleico (RNA). Os vírus de DNA incluem o herpesvírus humano-4 (HHV-4) ou vírus Epstein-Barr (EBV), o herpesvírus humano-8 (HHV-8), o HPV, o HBV e o poliomavírus de células de Merkel

(MCPyV). Os vírus de RNA incluem o vírus da hepatite C (HCV) e o vírus linfotrópico de células T humanas-1 (HTLV-1) (ZU HAUSEN, 2009; MOORE, CHANG, 2010) (Tabela 1.1).

Tabela 1.1 Oncovírus humanos. Ano de descoberta, família viral, tipo e tamanho do material genético, principais oncogenes, tropismo celular e neoplasias associadas.

Vírus	Ano	Taxonomia Viral	Tipo do genoma	Tamanho do genoma	Principais Oncogenes	Tropismo Celular	Neoplasias humanas
EBV	1964	Herpesviridae	dsDNA	172 kb (~90 ORFs)	LMP1	Células epiteliais da orofaringe, linfócitos B	Linfoproliferações, linfomas (LNH e LH) e Carcinomas (NPC e CG).
HBV	1965	Hepadnoviridae	ssDNA e dsDNA	3,2 kb (4 ORFs)	HBx	Hepatócitos, leucócitos	Carcinoma hepatocelular
HTLV-1	1980	Retroviridae	ssRNA	9 kb (6 ORFs)	Tax	Linfócitos T	Leucemia-linfoma de células T do adulto
HPV	1983-84	Papillomaviridae	dsDNA	8 kb (8-10 ORFs)	E5, E6, E7	Células epiteliais escamosas	Oral, cervical, peniano e anal
HCV	1989	Flaviviridae	(+)ssRNA	9,4 kb (9 ORFs)	NS5A	Hepatócitos e linfócitos	Carcinoma hepatocelular e linfomas
KSHV	1994	Herpesviridae	dsDNA	165 kb (~90 ORFs)	LANA, vflp, vBcl2 e outros	Linfócitos B	Sarcoma de Kaposi, linfoma primário de cavidades
MCPyV	2008	Polyomaviridae	dsDNA	5,3 kb (3 ORFs)	antígenos T	Células de Merkel	Carcinoma das células de Merkel

EBV: vírus Epstein-Barr (do inglês: *Epstein-Barr virus*); ds: fita dupla (do inglês: *double-stranded*); kb: quilobases (do inglês: *kilobases*); ORF: fase de leitura aberta (do inglês: *open read frame*); LMP1: proteína latente de membrana 1 (do inglês: *latent membrane protein 1*); LNH: Linfomas não-Hodgkin; LH: linfoma de Hodgkin; NPC: carcinoma de nasofaringe (do inglês: *nasopharyngeal carcinoma*); CG: carcinoma gástrico; HBV: vírus da Hepatite B (do inglês: *Hepatitis B virus*); HBx: proteína X do HBV (do inglês: *HBV X protein*); ss: fita simples (do inglês: *single-stranded*); HTLV-1: vírus T-linfotrópico humano tipo 1 (do inglês: *human T-lymphotropic virus type 1*); Tax: transativador da região do gene X (do inglês: *transactivator from the X-gene region*); HPV: papilomavírus humano (do inglês: *human papillomavirus*); HCV: vírus da Hepatite C (do inglês: *Hepatitis C virus*); NS5A: proteína não estrutural 5A (do inglês: *nonstructural protein 5A*); KSHV: herpesvírus associado ao sarcoma de Kaposi (do



inglês: *Kaposi's sarcoma-associated herpesvirus*); LANA: antígeno nuclear associado à latência (do inglês: *latency-associated nuclear antigen*); vflip: Proteína inibidora de FLICE viral (do inglês: *viral FLICE-inhibitory protein*); vBcl2: proteína viral do linfoma de célula B 2 (do inglês: *viral protein B-cell lymphoma 2*); MCPyV: poliomavírus das células de Merkel (do inglês: *Merkelcell polyomavirus*).

## 1.2 O vírus Epstein-Barr

Em 1958, Denis Burkitt descreveu um linfoma infantil endêmico na África Equatorial, o qual se suspeitava de uma etiologia associada a um agente biológico, como um vírus, em conjunto com a malária, visto que a prevalência geográfica do linfoma parecia depender da temperatura e da precipitação (BURKITT, 1958; 1962a;1962b). Em 1964, Michael Anthony Epstein, Ivonne Barr e Bert Achong por intermédio de Burkitt, conseguiram cultivar as células provenientes do linfoma, e através da microscopia eletrônica, descobriram partículas virais semelhantes aos herpesvírus. Identificou-se então um novo membro da família *Herpesviridae*, nomeado inicialmente como vírus Epstein-Barr (EBV) e atualmente denominado *Human gammaherpesvirus 4* (EPSTEIN; ACHONG; BARR, 1964; ICTV). Este vírus pertence a subfamília *Gammaherpesvirinae*, sendo o protótipo do gênero *Lymphocryptovirus* (LCV) (ICTV: <https://ictv.global/taxonomy/>)

Apesar da relação entre vírus e oncogênese já existir a partir de estudos com outros animais (ROUS, 1911), o sucesso da identificação do EBV no tumor (EPSTEIN; ACHONG; BARR, 1964) não era suficiente para entender seu papel, seja como um passageiro inócuo ou etiologicamente associado. No entanto, pouco tempo depois, o EBV isolado das células cultivadas de linfoma de Burkitt se mostrou capaz de transformar e imortalizar *in vitro* linfócitos B *naïves* (HENLE *et al*, 1967). A capacidade de transformação de células B *in vitro* e posteriormente observada *in vivo*, é um dos principais marcos que levaram a identificar o EBV como o primeiro candidato a vírus oncogênico humano (HENLE, 1967; POPE *et al*, 1968).

Em 1997, o EBV foi classificado pela Agência Internacional de Pesquisa em Câncer da Organização Mundial da Saúde (IARC; WHO, 1997), como um carcinógeno humano do grupo I, devido a sua associação causal com várias malignidades humanas, as quais possuem uma diversa distribuição geográfica (HUMANS; CANCER, 1997).

O papel do EBV na patogênese da maioria das neoplasias às quais é associado ainda não é claro (YOUNG; RICKINSON, 2004; FARRELL, 2019), mas entende-se que os genes virais expressos durante a transformação celular alteram o programa de transcrição gênica normal, ativando vias importantes de sinalização celular. Uma vez estabelecido o processo neoplásico, o vírus estabelece um padrão de expressão gênica particular conhecido como latências (explicado em detalhe abaixo) em cada um dos diferentes tumores que infecta (Tabela 1.2) (KUPPERS, 2003).

Tabela 1.2 Perfis de latência e expressão gênica identificados em neoplasias associadas ao EBV.

Tipos de Latência	RNAs não codificantes	mRNAs	Neoplasias associadas
Latência tipo I	EBERs BART miRNAs	EBNA1	Linfoma de Burkitt
Latência tipo II	EBERs BART miRNAs	EBNA1 LMP1 LMP2A LMP2B EBNA-LP EBNA2	Linfoma de Hodgkin Linfoma T/NK periférico Carcinoma de Nasofaringe Carcinoma Gástrico
Latência tipo III	EBERs BART miRNAs BHRF1 miRNAs	EBNA3s EBNA1 LMP1 LMP2A LMP2B Wp-BHRF1	Linfomas no hospedeiro imunocomprometido Linfomas associados a AIDS Linfoproliferações pós-transplante

EBER: pequenos RNAs codificados pelo vírus Epstein-Barr (do inglês: *Epstein-Barr virus-encoded small RNAs*); BART: transcritos BamHI-A para a direita (do inglês: *BamHI-A Rightward Transcript*); miRNA: pequeno RNA não-codificante (do inglês: *microRNA*); EBNA: antígeno nuclear do Epstein-Barr (do inglês: *Epstein-Barr nuclear antigen*); LMP: proteína latente de membrana (do inglês: *latent membrane protein*); LP: proteína líder (do inglês: *leader protein*); BHRF1: fase de leitura aberta à direita do BamHI-H (do inglês: *BamHI-H right reading frame*); AIDS: síndrome da imunodeficiência adquirida (do inglês: *acquired immunodeficiency syndrome*).

A incidência dos cânceres associados ao EBV denota grande variação geográfica, como a alta incidência de carcinoma de nasofaringe EBV+ no Sudeste da China e a alta incidência de linfoma de Burkitt EBV+ na África Subsaariana, onde em

ambas as regiões o EBV está quase sempre associado aos casos tumorais (ABDEL-HAMID *et al*, 1992; NICHOLLS *et al*, 1997; SANDVEJ *et al*, 2000; KLUMB *et al*, 2004; YOUNG; RICKINSON, 2004; CARBONE *et al*, 2008; NEVES *et al*, 2017; CORREIA *et al*, 2018). Acredita-se que as diferenças geográficas ou demográficas na incidência das doenças associadas ao EBV podem ser explicadas por cofatores ambientais, como a variação genética do EBV e a variação imunogenética do hospedeiro (DIEPSTRA *et al.*, 2005; NIENS *et al.*, 2007; HUANG, *et al* 2012), uma vez que as interações entre vírus, célula e sistema imune são cruciais no desenvolvimento do câncer.

Os indivíduos com imunossupressão severa encontram-se em risco de desenvolver linfoproliferações associadas ao EBV, devido à queda da imunovigilância do hospedeiro permitindo a expressão de produtos virais de latência III, os quais possuem um alto potencial transformador (BIBAS & ANTINORI, 2009). Os outros tumores EBV-positivos com padrões mais restritos de latência apresentam uma patogênese mais complexa que pode envolver a ação transformadora de oncoproteínas virais específicas. As proteínas latentes codificadas pelo EBV que medeiam a transformação celular incluem principalmente os seis antígenos nucleares (EBNAs) e três proteínas de membrana latentes (LMPs). Nos tumores linfoides e epiteliais, as expressões virais dos genes latentes têm padrão distinto. Ao contrário da célula benigna cronicamente infectada pelo EBV, essas proteínas latentes podem estar super expressadas nas células neoplásicas associadas ao vírus (FARRELL, 2019).

### 1.2.1 Estrutura

O EBV é composto por um genoma de DNA linear de dupla fita (dsDNA) e um capsídeo proteico icosaédrico de 162 capsômeros, circundado por uma matriz proteica, chamada tegumento (Figura 1.1). Sendo um herpesvírus, o EBV possui um envelope adquirido da membrana nuclear interna das células infectadas, que possui glicoproteínas virais (como a gp350/220, gH, gB e gp42) importantes para seu tropismo celular por linfócitos B e células epiteliais (JOHANNSEN *et al*, 2004).

O ciclo biológico do EBV é bifásico, contendo uma fase replicativa ou produtiva, e uma latente, gerando uma infecção crônica nos linfócitos B de memória que persistirá durante a vida do hospedeiro. Recentemente, foi demonstrado que o EBV é capaz de infectar linfócitos T durante a primo-infecção, entretanto ainda não está

claro, se é uma infecção passageira ou se o EBV é capaz manter uma infecção persistente (BARROS *et al*, 2019).

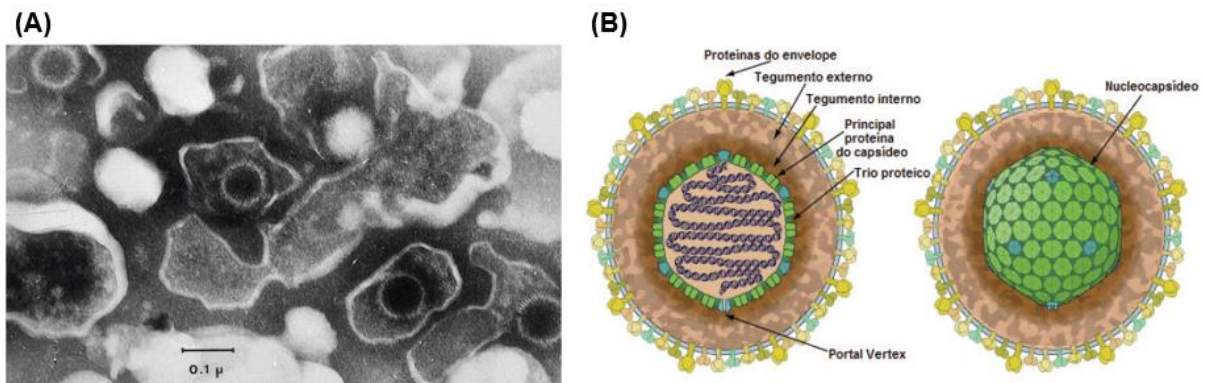


Figura 1.1 - **Estrutura da partícula viral do EBV.** (A) Microscopia eletrônica do EBV. Fonte: BORNKAMM, 2016. (B) Esquema da estrutura do EBV. Fonte: Adaptado de ViralZone, c2017.

### 1.2.2 Ciclo replicativo

A infecção do EBV ocorre principalmente através da transmissão de vírions pela saliva, mas também pode ser transmitido por contato sexual, transfusão de sangue ou transplante de órgãos (CRAWFORD, 2001; CRAWFORD *et al*, 2002). O modelo atual para explicar a infecção e persistência do EBV é através da infecção de linfócitos B naíves residentes nos tecidos linfoides do epitélio tonsilar (Figura 1.2).

Na infecção primária após a transmissão através da saliva, o vírus invade as células epiteliais da orofaringe, estabelecendo um ciclo replicativo que resulta na produção de novos vírions, os quais irão invadir linfócitos B através dos seus receptores celulares específicos (CRAWFORD, 2001). Nestas células, o EBV estabelece um programa de latência, expressando um conjunto de genes: seis antígenos nucleares (EBNA1, EBNA2, EBNA3A, 3B e 3C e EBNA LP), três proteínas integrais de membrana (LMP1, LMP2A e LMP2B), dois RNAs não poliadenilados e não traduzíveis (EBER1 e EBER2) e um conjunto de microRNAs com funções pouco esclarecidas (BARTs). Este programa de expressão é denominado de latência III ou padrão de crescimento e tem como resultado a indução da transformação (blástica) e a proliferação dos linfócitos B (BORNKAMM; HAMMERSCHMIDT, 2001; RICKINSON, A., 2002; KUPPERS, 2003; FARRELL, 2019).

A resposta imune em hospedeiros imunocompetentes é principalmente por células T citotóxicas, as quais são ativadas e elicitam um vigoroso estímulo contra as células infectadas que expressam o programa de latência III (HISLOP *et al.*, 2007).

Isto é possível porque as proteínas expressas nesse programa são altamente antigênicas, especialmente as EBNA2, EBNA3A, EBNA3B e EBNA3C (THORLEY-LAWSON, 2001; HISLOP *et al.*, 2007). Durante a primo-infecção, os vírus que obtiverem sucesso em reduzir a expressão dessas proteínas altamente imunogênicas seguem a via normal de desenvolvimento das células B, utilizando um padrão de expressão gênica mais restrito de latência (II) nos linfócitos do centro germinativo (CG) (SOUZA *et al.*, 2005).

Na latência II, após a estimulação antigênica das células B infectadas pelo EBV, a expressão da proteína EBNA1 assegura a manutenção episomal durante a divisão celular e as proteínas latentes de membrana (LMP) 1 e 2 fornecem um sinal “auxiliar” artificial T (CD40) e B (BCR, receptor de células B), respectivamente, garantindo a sobrevivência do linfócito infectado no CG (UCHIDA *et al.*, 1999; THORLEY-LAWSON, 2001). Finalmente, após inibir a via de diferenciação para células plasmáticas, as células B EBV+ chegam ao compartimento de memória (HOCHBERG *et al.*, 2004) onde toda a expressão gênica viral é silenciada (latência 0), com o DNA viral se mantendo latente como epissoma no núcleo celular das células B de memória (RICKINSON, 2002). Assim, o EBV torna-se invisível para o sistema imune, estado que persistirá ao longo da vida do hospedeiro. A proteína EBNA1 será expressa novamente somente em células em divisão para assegurar a manutenção genômica do EBV (latência 1) (TIERNEY *et al.*, 1994; JOSEPH; BABCOCK; THORLEY-LAWSON, 2000; THORLEY-LAWSON; GROSS, 2004). Esporadicamente, o vírus pode recomeçar o ciclo replicativo ou lítico, garantindo a renovação da infecção em novas células B e adicionalmente ser excretado pela saliva, o que permite a transmissão para novos hospedeiros (BABCOCK *et al.*, 1998).

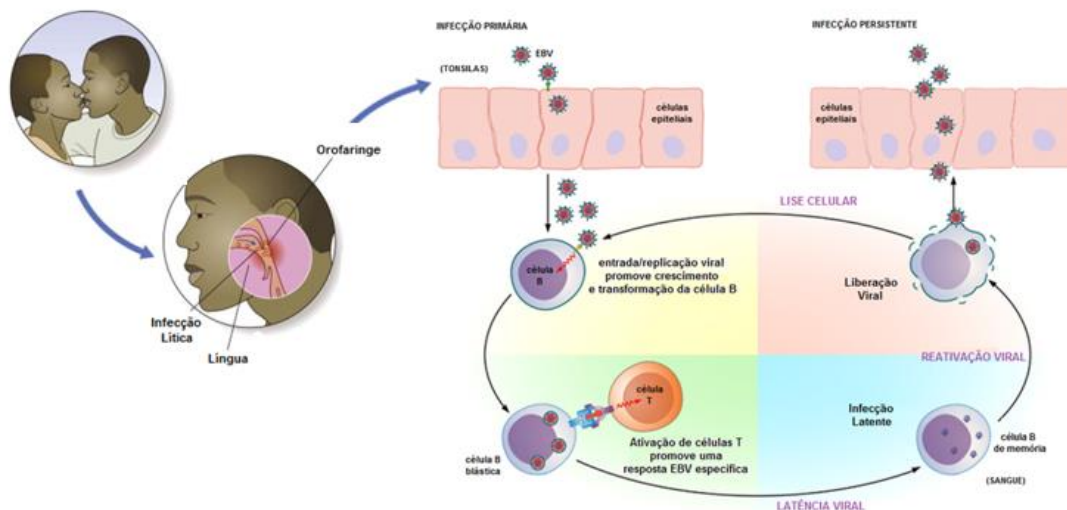


Figura 1.2 - **Esquema do ciclo replicativo do vírus Epstein-Barr** (adaptado de KORSMAN *et al.*, 2012; ODUMADE *et al.*, 2011).

### 1.2.3 História natural da infecção

O EBV estabelece no hospedeiro uma infecção latente assintomática, estando presente em mais de 90% da população humana adulta (CRAWFORD, 2001). Na história natural da infecção pelo EBV, a infecção primária ocorre pelo contato com secreções orais usualmente durante a primeira década de vida e transcorre de forma assintomática (HADINOTO *et al.*, 2008). De 20% a 50% dos casos em que o primeiro contato ocorre na adolescência, a infecção viral pode gerar uma resposta imune vigorosa de células T CD8+, denominada mononucleose infecciosa (MNI) (CRAWFORD, 2001; VETSIKA; CALLAN, 2004). Ademais, a identificação da exposição inicial é dificultada pelo período de incubação entre a exposição e a apresentação dos sintomas, o qual na MNI pode levar de 30 a 60 dias (PANNUTI *et al.*, 1980).

É observado que a alta incidência da MNI em adolescentes e jovens adultos nos países com elevado índice de desenvolvimento humano (IDH) pode ser devida a soroconversão tardia nesses países. Anticorpos contra o EBV são encontrados em 50% das crianças na idade de 5 anos e em 90% na idade 25 anos, mostrando que a maioria dos adultos já desenvolveu infecção assintomática ou subclínica (FLEISHER *et al.*, 1979; HADINOTO *et al.*, 2008). Em países com baixo IDH, a infecção primária ocorre entre algumas semanas e os primeiros anos de vida, onde na idade de seis anos a soroconversão é quase universal (BIGGAR *et al.*, 1978; CHAN *et al.*, 2001). Já em países com médio IDH a curva de soroconversão pelo EBV é intermediária à observada nos países com baixo e elevado IDH (PANNUTI *et al.*, 1980; FIGUEIRA-SILVA; PEREIRA, 2004).

Apesar da sua natureza latente durante grande parte da vida do indivíduo, o EBV pode esporadicamente recomeçar seu ciclo replicativo. A ativação dos programas de expressão dos produtos do ciclo lítico parece ocorrer em células B de memória, que estão recirculando através do tecido linfático associado à mucosa orofaríngea, no entanto os mecanismos de reativação *in vivo* ainda não estão claros. Além disso, indivíduos saudáveis portadores do vírus podem eliminar o vírus continuamente, onde se presume que o vírus esteja se replicando em células epiteliais orais (HADINOTO *et al.*, 2009).

Embora a maioria das infecções pelo EBV seja benigna e o desenvolvimento de neoplasias seja raro, o EBV está associado causalmente a aproximadamente 1,5% das neoplasias em todo o mundo (CRAWFORD, 2001; PARKIN *et al.*, 2006.; FARRELL, 2019). Existem fortes evidências de que o desenvolvimento da MNI é um fator chave para a predisposição a algumas doenças, como LHC e a esclerose (HJALGRIM *et al.*, 2003; THACKER; MIRZAEI; ASCHERIO, 2006; HJALGRIM *et al.*, 2007). Estudos têm relatado um aumento do risco de desenvolvimento de LHC de até 5 vezes nos cinco anos seguintes ao diagnóstico de MNI (HJALGRIM *et al.*, 2007). O mecanismo para essa associação ainda não é bem entendido, mas dados sugerem que a persistência de muitas células B infectadas pelo EBV e a replicação viral depois da MNI, pode ser um fator crítico para o desenvolvimento do LHC EBV+ (HJALGRIM *et al.*, 2010). De fato, em relação a outro câncer associado ao EBV, como o linfoma de Burkitt, já foi relatada a presença crônica de níveis aumentados de anticorpos anti-VCA do EBV anos antes do desenvolvimento da doença (GESER *et al.*, 1982).

#### 1.2.4 Diversidade genética do vírus Epstein-Barr

O EBV é biologicamente similar aos Linfocriptovírus (LCVs) que infectam naturalmente primatas do Velho Mundo, como o McHV-4 (*Macacine gammaherpesvirus 4*, anteriormente conhecido como *Cercopithecine gammaherpesvirus 15*) e do Novo Mundo, como o CalHV3 (*Callitrichine gammaherpesvirus 3*), entre outros (WANG *et al.*, 2001; LACOSTE *et al.*, 2010). Reconstruções filogenéticas demonstram que os herpesvírus de mamíferos têm origem monofilética, a qual radiou para a formação das três subfamílias (*Alpha*, *Beta* e *Gammaherpesvirinae*) conhecidas atualmente (Figura 1.3) (MCGEOCH; RIXON; DAVISON, 2006). A divergência das três subfamílias foi estimada aproximadamente há 400 milhões de anos, tendo os gêneros dentro das subfamílias aparecido antes da radiação dos mamíferos, por volta de 60–80 milhões de anos, onde as espécies virais teriam surgido durante os últimos 60 milhões de anos (MCGEOCH *et al.*, 1995). Desta forma, o estabelecimento dessas três subfamílias teria ocorrido durante a dispersão dos mamíferos e postula-se que o processo de diversificação dos vírus atuais tenha ocorrido simultaneamente com a evolução dos hospedeiros hominídeos (MCGEOCH *et al.*, 1995; ALBA *et al.*, 2001).

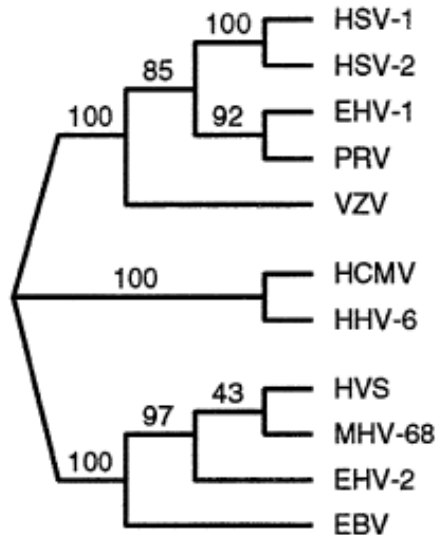


Figura 1.3 - **Árvore de máxima parcimônia para o gene *UL5* de alguns membros da família herpesviridea.** A árvore é mostrada com os valores de *bootstrap*, de forma não enraizada com a separação monofilética das subfamílias de alfa herpesvírus (HSV-1, HSV-2, EHV-1, PRV, VZV), a subfamília de beta herpesvírus (HCMV e HHV-6) e a subfamília de gamma herpesvírus (HVS, MHV-68, EHV-2 e EBV) indicados por espaçamento (retirado de MCGEOCH *et al.*, 1995). HSV-1: herpes simplex vírus tipo 1; HSV-2: herpes simplex vírus tipo 2; EHV-1: equine herpesvírus tipo 1; PRV: pseudorabies vírus; VZV: varicella-zoster vírus; HCMV: human cytomegalovírus; HHV-6: human herpesvírus 6; HVS: herpesvírus saimiri; MHV-68: murine herpesvírus; EHV-2: equine herpesvírus; EBV: Epstein-Barr vírus.

Como todo herpesvírus, o EBV é considerado um agente biológico com material genético estável e de evolução lenta devido a latência estabelecida no hospedeiro, no qual replica seu genoma com a DNA polimerase eucariota. No entanto, o longo período de coevolução com o hospedeiro pode ter levado ao desenvolvimento de mecanismos de adaptação viral frente às condições impostas pela defesa antiviral do hospedeiro (MCGEOCH *et al.*, 1995; MCGEOCH *et al.*, 2006; EHLERS *et al.*, 2010). Assim, genes que codificam proteínas do ciclo lítico envolvidas na replicação do DNA viral, estrutura da partícula viral e empacotamento do DNA viral demonstram uma alta conservação com outros herpesvírus. Contudo, outros genes são compartilhados apenas entre a subfamília *gamma herpesvirinae*, por exemplo os que codificam para *BZLF1* e *BRLF1*, reguladores iniciais imediatos do ciclo lítico, e as proteínas latentes LMP1 e LMP2 (KIEFF e RICKINSON, 2007). E outros genes têm semelhanças com os do hospedeiro, como: *BZLF1*, *BHRF1* e *BCRF1* que são semelhantes aos genes *c-FOS*, *BCL-2* e *IL-10* do hospedeiro, respectivamente (PEARSON *et al.*, 1987; FARRELL *et al.*, 1989; MOORE *et al.*, 1990).



O EBV é classificado principalmente em tipo 1 e tipo 2 (ou tipo A e tipo B), devido a variações nas sequências dos genes *EBNA-2*, *EBNA-3A*, *-3B* e *3C* (DAMBAUGH *et al.*, 1984; ROWE *et al.*, 1989; SAMPLE *et al.*, 1990). O gene *EBNA-2* é o mais divergente no genoma do EBV, resultando em 70% de identidade a nível nucleotídico e 54% a nível de proteína entre o tipo 1 e o tipo 2 (TZELLOS *et al.*, 2012). O protótipo do EBV tipo 1 é a variante B95-8, uma linhagem que foi produzida a partir da infecção de linfócitos de *Saguinus oedipus* com o vírus proveniente de uma linhagem linfoblastoide gerada a partir de linfócitos B EBV+ de um paciente norte-americano, diagnosticado com mononucleose infecciosa (MILLER *et al.*, 1972; 1973). Em 1990, uma região de 11.8kb deletada do genoma da linhagem B95-8 foi preenchida com a sequência do genoma da linhagem Raji, tornando-se o protótipo tipo 1 do genoma do EBV até hoje (PARKER *et al.*, 1990), com o número de acesso NC\_007605 pelo Genbank. O protótipo do tipo 2 foi estabelecido da linhagem celular AG786 oriunda de um linfoma de Burkitt africano (PIZZO *et al.*, 1978) (nº de acesso do Genbank DQ279927). A principal diferença funcional entre os tipos é a maior capacidade de imortalização e estabelecimento das linhagens celulares linfoblastoides (LCL) *in vitro* das variantes do EBV do tipo 1 em comparação com os EBVs tipo 2 (RICKINSON *et al.*, 1987). Foi demonstrado que essa diferença está no gene *EBNA-2*, o qual possui um papel principal na habilidade que as LCLs do tipo 1 têm de serem capazes de exibir uma alta taxa de crescimento e densidade celular, quando comparado a LCLs do tipo 2 (TOMKINSON *et al.*, 1992).

O EBV possui um genoma de dsDNA com aproximadamente 172 kb, podendo exibir até 85 fases de leituras abertas (ORF, do inglês: *open reading frame*) e vários RNAs funcionais não-codificantes (EBERs, BART miRNAs e BHRF1 miRNAs). A grande maioria das ORFs é expressa apenas durante o ciclo lítico e apenas 11 delas são transcritas durante a latência, dos quais apenas 9 são traduzidos. O genoma tem uma organização contendo várias repetições diretas em série idênticas de 0,5 kb em ambos os terminais do genoma (TR), possuindo inclusive várias repetições diretas internas (IR) de 3 kb, que constam nos promotores de latência e em sequências longas e curtas únicas ao longo do genoma. Essas sequências contêm quase todo o conteúdo de codificação (RICKINSON *et al.*, 2007).

O genoma apresenta uma forma linear na partícula viral (Figura 1.3), no entanto, circulariza-se nas células B infectadas via TRs estabelecendo-se na forma epissomal (Figura 1.4) (BANKIER *et al.*, 1983). Em 1984, foi sequenciado o primeiro genoma completo da linhagem B95-8 (nº de acesso do Genbank V01555) (BAER *et*

al., 1984). O genoma do EBV foi sequenciado a partir de fragmentos das enzimas de restrição *EcoRI* e *BamHI*, desta forma, a sequência se inicia em 1 nucleotídeo à esquerda do sítio *EcoRI*, que separa *EcoRI* Dhet de *EcoRI* I (o primeiro A de AGAATTC). Conseqüentemente, os fragmentos foram nomeados alfabeticamente por seus tamanhos e os genes intitulados de acordo com a orientação a fase de leitura do fragmento *BamHI* correspondente, por exemplo, o gene *BNLF1* é a primeira fase de leitura para a esquerda começando no fragmento *BamHI*-N no genoma (*BNLF1*, do inglês: **BamHI-N Left Fragment 1**) (Figura 1.3).

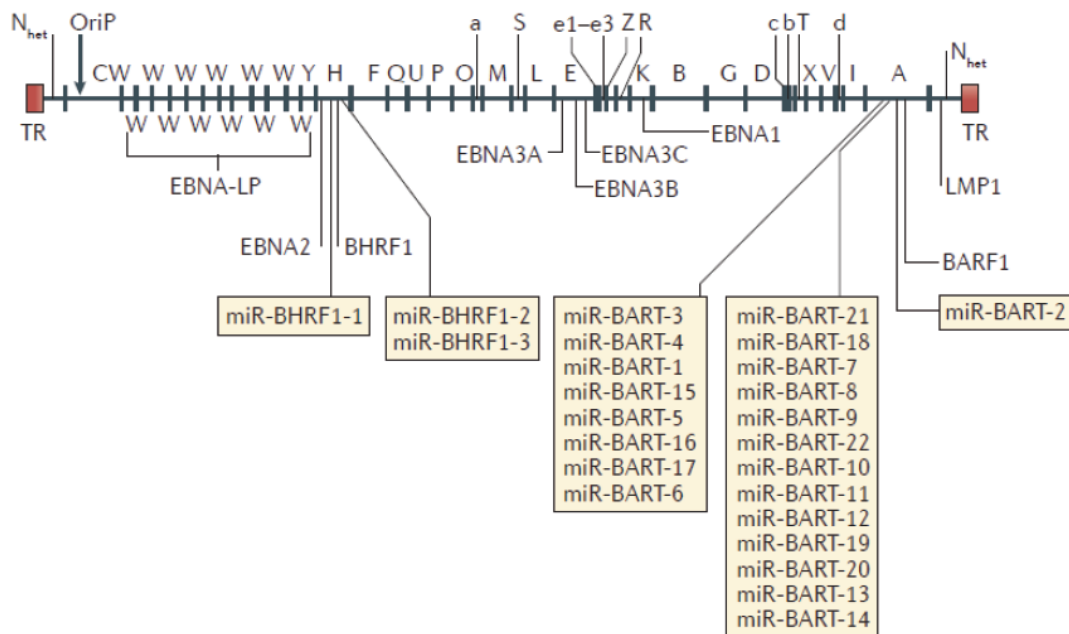


Figura 1.4 – **Esquema da posição de alguns genes e miRNA virais na forma linear do genoma do EBV.** As letras denotam fragmentos de digestão da enzima de restrição *BamHI* que são nomeados de acordo com o tamanho, sendo o “A” o maior, e as letras minúsculas os menores. Os quadrados vermelhos são as repetições terminais em cada extremidade do genoma. miR: microRNAs (retirado de YOUNG, 2016).

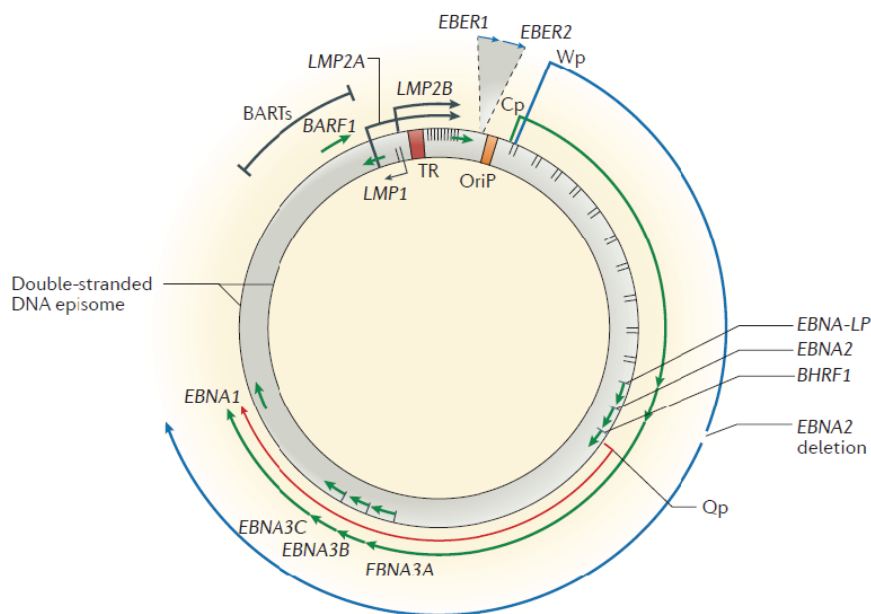


Figura 1.5 – **Esquema da posição de alguns genes, RNA não codificantes e promotores na forma circular (ou episomal) do genoma do EBV.** Genes: oncoproteínas latentes de membrana (LMP) e EBNA; RNA não codificantes: Bart, Barf e Eber; promotores: Wp, Cp e Qp, regiões que promovem o início da transcrição; quadrado vermelho: TR; quadrado laranja: OriP, região que regula a transcrição de genes vizinhos (retirado de YOUNG, 2016).

Até recentemente, devido ao seu genoma extenso e heterogêneo, apenas 8 genomas inteiros do EBV eram conhecidos (ZENG *et al.*, 2005; DOLAN *et al.*, 2006; LIU *et al.*, 2011; KWOK *et al.*, 2012; LIN *et al.*, 2013). A partir de 2014 com o advento dos métodos de captura híbrida, o número de sequências do genoma completo do EBV cresceu e muitos esforços têm sido realizados para caracterizar por epidemiologia molecular a distribuição das variantes do EBV no mundo, processo que acarretou mais de 200 genomas completos caracterizados do EBV em todo o mundo (KWOK *et al.*, 2014; PALSER *et al.*, 2015; CORREIA *et al.*, 2018; KANDA *et al.*, 2019). Entretanto, muitas áreas geográficas ainda estão sub-representadas, como a América do Sul, possuindo apenas nove genomas completos caracterizados até o momento, sendo três da Argentina e seis do Brasil, todos relacionados ao tipo 1 do vírus e originários de casos com Linfoma de Burkitt (LEI *et al.*, 2015). A falta de representação geográfica pode levar a uma dificuldade de monitoramento de variantes, uma vez que o sistema de classificação genômico do EBV é recente, não existindo um consenso ou metodologia preconizada para a classificação. Por exemplo, análises de diferentes alvos virais por filogenia podem gerar topologias filogenéticas diferentes, então, apesar da diversidade do EBV ser classificada através dessa topologia, geralmente

são gerados muitos clados parafiléticos que dificultam a definição de clados para uma classificação adequada (ZANELLA *et al.*, 2019). Desta forma, torna-se importante a investigação do genoma do EBV de diferentes regiões geográficas, na busca de classificações mais robustas a nível genômico, a fim de entendermos a complexa interação EBV-hospedeiro, no intuito de destacar potenciais oncogênicos virais.

Os métodos mais utilizados no intuito de identificar variantes com potencial oncogênicos são polimorfismos específicos em diferentes oncogenes, como os dos genes *EBNA2/3C*, classificando os tipos 1 e 2, onde o tipo 1 possui uma maior capacidade do EBV em promover a proliferação e sobrevivência de células B (CALCIAN *et al.*, 2011; TZELLOS *et al.*, 2014). Tem sido demonstrado que os tipos 1 e 2 têm restrição geográfica de acordo principalmente com o grau de imunocompetência da população ou grupo estudado, com o tipo 1 mais prevalente no mundo em assintomáticos e doenças associadas, e o tipo 2 mais presente em indivíduos HIV+, no entanto, os tipos *per se*, não são associados à doença. No entanto, mais recentemente foi relatada uma associação de haplótipos virais, como o Tipo1 e o polimorfismo V3 em variantes do EBV associadas ao tumor (BRISTOL *et al.*, 2018). O polimorfismo V3, está localizado no promotor do gene lítico *Z* (*Zp*), responsável pela troca do perfil latente para lítico, que quando presente, essa variante V3 confere um aumento funcional na reativação lítica, podendo contribuir para a prevalência de variantes associadas a malignidades (BRISTOL *et al.*, 2018). Além disso dos *EBNAs*, outros oncogenes latentes são utilizados a classificar a diversidade de variantes do EBV no mundo, na tentativa de identificar variantes com maior potencial oncogênico (HU *et al.*, 1993; ZUECHER *et al.*, 2012; BRISTOL *et al.*, 2018).

#### 1.2.4.1 A oncoproteína LMP1

A proteína latente de membrana 1 ou LMP1, codificada pela gene *BNLF1* ou *LMP1*, é uma oncoproteína que tem sido estudada extensivamente devido à sua alta variabilidade intrahospedeiro e seu impacto na oncogênese mediada pelo EBV, por ter a capacidade de transformar fibroblastos de roedores *in vitro* e por ser essencial para a transformação de linfócitos B infectados pelo EBV (WANG; LIEBOWITZ; KIEFF, 1985; KAYE; IZUMI; KIEFF, 1993). Diferente do que ocorre nas células benignas infectadas pelo EBV, a LMP1 é expressa em altos níveis na grande maioria das neoplasias associadas ao EBV (MIDDELDORP; PEGTEL, 2008).

À nível genético, o *LMP1* é um dos genes mais variáveis do EBV, que além de possuir regiões repetitivas internas, é localizado na porção 3' do genoma à montante

das regiões repetitivas terminais, assim, o *LMP1* está localizado em uma região do genoma do EBV com alta instabilidade genômica (YOUNG, 2016). Ainda que o EBV tenha um genoma estável de dsDNA, processos mutacionais, como substituição nucleotídica e recombinação podem ocorrer, principalmente levando em consideração o tipo celular em que o EBV estabelece sua infecção latente: linfócitos B, os quais possuem uma alta taxa de recombinação homóloga em comparação a outros tipos celulares, à vista disso, os principais mecanismos genéticos que podem contribuir para a variabilidade da *LMP1* são mutação pontual, deleção e duplicação e recombinação homóloga intra/intercepta (WALLING *et al.*, 1999; GUIRETTI *et al.*, 2008).

Devido a sua importância oncogênica, foram propostas diversas classificações no intuito de verificar a associação de variantes da *LMP1* com patologias relacionadas ao EBV, sob a hipótese que a variação a nível genético poderia afetar a capacidade de transformação celular (SANDVEJ *et al.*, 1997; EDWARDS *et al.*; SEILLIER-MOISEWITSCH *et al.*; RAAB-TRAUB *et al.*, 1999; WALLING *et al.*, 1999; Lei *et al.*, 2015; GANTUZ *et al.*, 2017). No entanto, não existe um consenso para a classificação das diferentes variantes da *LMP1*, e diferentes sistemas de classificações continuam sendo propostos até os dias atuais (Lei *et al.*, 2015). Alguns dos sistemas de classificações sugeridos até o momento podem utilizar a região promotora, assim como parte ou o gene inteiro da proteína, e comumente uma amostragem de uma região geográfica específica. Como o sistema de classificação proposto por Walling e colaboradores (WALLING *et al.*, 1999) que utiliza parte da região C-terminal, e identificou 22 padrões distintos. Sandvej e colaboradores (SANDVEJ *et al.*, 1997), por sua vez, identificaram quatro variantes, A, B, C e D em indivíduos europeus. E mais recentemente, Lei *et al.* (LEI *et al.*, 2015), através da análise de 112 genomas propôs um sistema de classificação incluindo a região promotora e codificante da *LMP1*, através de 4 padrões (Padrão A, B, C, e D) incluindo 23 polimorfismos, na tentativa de identificar padrões associados a oncogenicidade. No entanto, o sistema de classificação mais utilizado é o de Edwards *et al.* (EDWARDS *et al.*, 1999) baseado em mudanças aminoacídicas da região C-terminal em conjunto com agrupamento filogenético, os autores utilizaram amostras de diferentes regiões geográficas, propondo sequências prototípicas que são comumente utilizadas como referências para a classificação da *LMP1*, como os protótipos B95.8, Med+/-, China 1 e 2, Alaskan e NC (Tabela 1.3 e Figura 1.6) (PALSER *et al.*, 2015; CORREIA *et al.*, 2017; GANTUZ *et al.*, 2017).



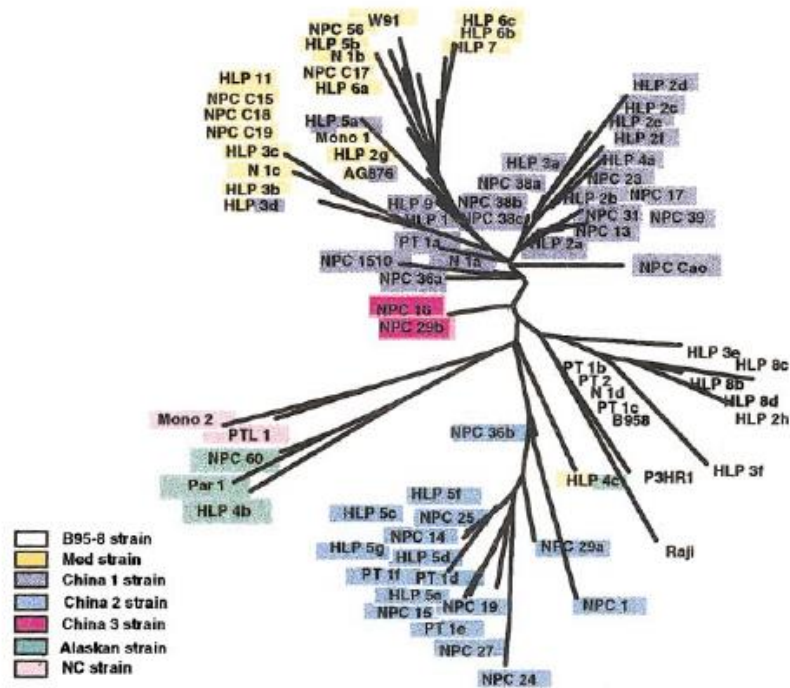


Figura 1.6 – **Árvore filogenética formada pelo método de *neighbor-joining* a partir de seqüências nucleotídicas que definiram protótipos da LMP1 por Edwards *et al.*** Os isolados agrupam-se caracteristicamente de acordo com as cepas prototípicas (retirado de EDWARDS *et al.*, 1999).

Funcionalmente, a LMP1 mimetiza a proteína transmembranar CD40, agindo como um receptor independente de ligante constitutivamente ativo (YOUNG; RICKINSON, 2004). A LMP1 possui 386 aminoácidos (aa) e 63 kDa (FENNEWALD *et al.*, 1984). O gene que codifica a LMP1 (*BNLF1*) encontra-se numa região de grande instabilidade genômica, na parte final do genoma linear do EBV, próximo das repetições terminais do genoma (YOUNG; RICKINSON, 2004). A LMP1 contém um curto domínio N-terminal (N-Ter) intracelular (1-23 aa), seguido por um domínio transmembrana (TM) com seis regiões intramembranares hidrofóbicas (24-186 aa) e um domínio intracelular denominado C-terminal (C-ter) (200 aa) (Figura 1.5).

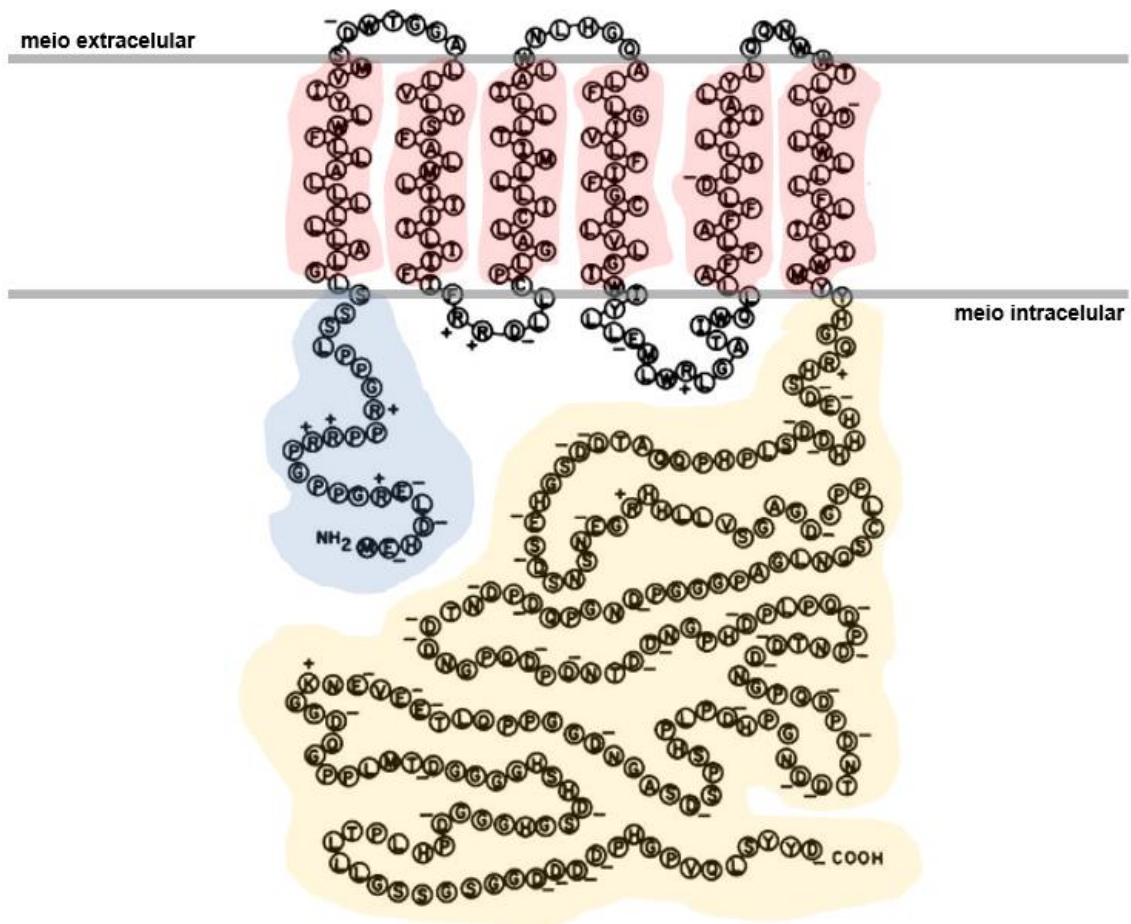


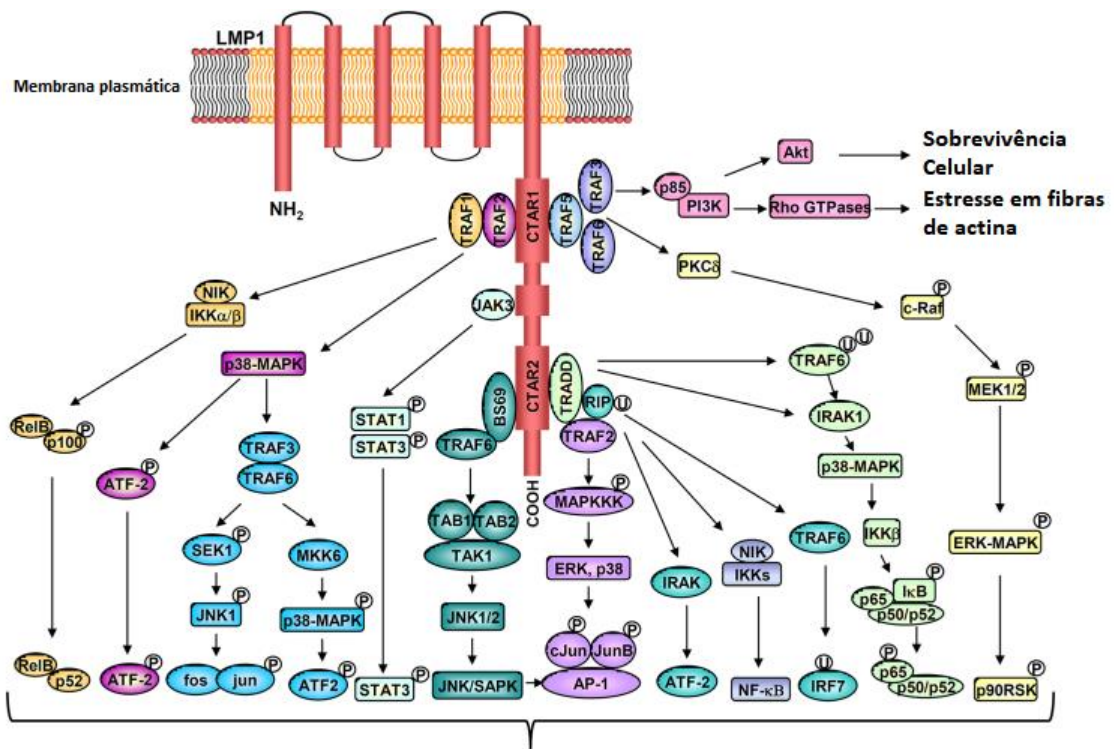
Figura 1.7 - **Esquema aminoacídico da LMP1 do EBV.** Cada círculo identifica um dos aminoácidos contidos na estrutura da proteína do protótipo B95-8. Os sinais positivos e negativos caracterizam a carga do aminoácido. Os aminoácidos coloridos de azul, vermelho e amarelo pertencem aos domínios N-terminal, intramembranares e C-terminal da proteína, respectivamente. Adaptado de FENNEWALD *et al.*, 1984.

Os domínios da LMP1 apresentam características e funções distintas na biologia viral e na transformação celular. O domínio N-ter tem um papel importante na ancoragem da LMP1 na membrana, além de regular a degradação proteossomal da proteína por ubiquitinação (IZUMI *et al.*, 1994; AVIEL *et al.*, 2000; CHUNG *et al.*, 2002). A região TM, por sua vez, é crucial para o tráfico intracelular e a oligomerização da molécula nos *rafts* lipídicos da membrana, sendo essencial para a capacidade de transformação da LMP1. Além disso, essa região é crítica na estimulação da via NF- $\kappa$ B (KAYE *et al.*, 1993; KAYKAS; WORRINGER; SUGDEN, 2002). Na região C-Ter encontram-se os principais motivos de sinalização intracelular, que conferem à LMP1 sua capacidade de transformação. Os motivos estão localizados em regiões denominadas de CTAR1 e 2 (do inglês: *C-terminal activation regions*) (ELIOPOULOS; YOUNG, 2001), permitindo a ligação de proteínas de fatores associados ao receptor de TNF (TRAFFs) e de receptores de fator de necrose tumoral



tipo1 (TRADDs), da família TNFRSF que regulam apoptose e sobrevivência celular, sendo capaz de ativar a via de NF- $\kappa$ B (Figura 1.6) (DEVERGNE *et al.*, 1996; DEVERGNE *et al.*, 1998).

O motivo CTAR1 contém um motivo de interação com a proteína TRAF, cuja interação com TRAF1, 2 e 3 é crítica para sua função na sinalização de NF- $\kappa$ B e JNK/AP (DEVERGNE *et al.*, 1996; BRODEUR *et al.*, 1997; MILLER *et al.*, 1998). O motivo CTAR2 contém um motivo menor de ligação à TRAF (379 aa), que é crítico para JNK e sinalização do NF- $\kappa$ B pela via canônica, embora a ligação direta de TRAFs ainda não tenha sido demonstrada (FLOETTMANN *et al.*, 1997). CTAR2 também atua recrutando a proteína do domínio DEATH associada ao TRADD por meio de um motivo (384 aa) (SCHNEIDER *et al.*, 2008). Em suma, o CTAR1 ativa principalmente as vias de NF- $\kappa$ B não canônicas (via p100 e p52) (LUFTIG *et al.*, 2004; THORNBURG *et al.*, 2007), enquanto a CTAR2 é responsável pela via NF- $\kappa$ B canônica (via I $\kappa$ B e p65) (SAITO *et al.*, 2003). A região entre CTAR1 e CTAR2 contém dois motivos para Janus Kinase 3 (JAK3) e uma região de repetição de 33 pb (rep33 pb). As cepas da LMP1 geralmente contêm 2-6 cópias da rep33 pb (11 aa), sendo 3-4 e 5 o número mais comum de repetições (WALLING *et al.*, 1999; GUIRETTI *et al.*, 2008; GANTUZ *et al.*, 2017). As regiões da JAK3, são definidas como CTAR3, e são responsáveis pela transformação celular (HIGUCHI *et al.*, 2002), ainda que na literatura a ativação por JAK/STAT ainda seja controversa (GIRES *et al.*, 1999; HIGUCHI *et al.*, 2011). Estudos recentes têm mostrado que a ativação pelo motivo de CTAR3 está envolvido na migração celular através da interação e regulação da proteína IRF7 por induzir sua sumoilação (BENTZ *et al.*, 2011; 2012). Uma parte das vias de sinalizações citadas acima, está esquematizada na Figura 1.6.



**Sobrevivência celular, Proliferação, Angiogênese, Invasão, Metástase e indução de citocinas**

Figura 1.8 - **Esquema da ativação de múltiplas vias de sinalização celular orquestradas pela LMP1.** O domínio C-terminal da LMP1 é capaz de interagir com diferentes vias de sinalização que podem promover sobrevivência celular, proliferação, angiogênese, invasão metástase e indução de produção de citocinas (Adaptado de LO *et al*, 2021).

O *LMP1* é um dos genes mais variáveis do genoma do EBV (EHLERS *et al.*, 2010; BERENSTEIN *et al*, 2018), e possui marcadores polimórficos, como a deleção de 30 pb (del30), associado a linfomas e linfomas associados a AIDS em populações específicas (HU *et al*, 1993; GUIRETTI *et al*, 2007; DOLCETTI *et al*, 2011; SPINA *et al*, 2011; LORENZETTI *et al*, 2012), uma inserção de 15 pb (ins15) codificando um motivo de sinalização para Janus Kinase 3 (JAK3), número variável de repetições de 33 pb em tandem e *hotspots* de mutações no seu domínio C-ter (KNECHT *et al.*, 1995; SANDVEJ *et al.*, 1997; WALLING *et al.*, 1999; BERGER *et al.*, 1999; CHANG *et al.*, 2009; EDWARDS *et al.*, 2004; ROTHENBERGER; ZUERCHER, 2012).

Pesquisas em busca de variantes mais oncogênicas associadas aos polimorfismos da *LMP1* levaram a um achado positivo com relação a variante CAO, derivada de um carcinoma de nasofaringe, cuja características polimórficas são altamente oncogênicas e menos imunogênicas *in vitro* que o protótipo B95.8, sendo capaz de formar tumores em modelos *in vivo* (CHEN *et al.*, 1992; HU *et al.*, 1993; TRIVEDI *et al.*; 1994). Ademais, nosso grupo identificou uma associação do

polimorfismo de 30 pb, localizado no domínio C-ter da LMP1 em casos de Linfoma de Hodgkin clássico, levantando a necessidade de classificar a diversidade do EBV pela LMP1 na nossa região geográfica, a fim de tentar identificar variantes com potenciais oncogênicos. Outros estudos de amostragem brasileira ainda são poucos, e focados principalmente em polimorfismos específicos, onde a maioria são de EBV oriundos de amostras neoplásicas *versus* casos benignos (GUIRETTI *et al.*, 2007; CHABAY *et al.*, 2020). E apenas mais recentemente, amostras brasileiras foram classificadas por toda sua extensão codificante da LMP1 em conjunto com a região promotora, sendo classificadas pelos padrões definidos por Lei e colaboradores (2015). O estudo identificou um padrão de polimorfismos em amostras brasileiras, argentinas e ganenses de linfoma de Burkitt. Posteriormente, Liao e colaboradores (2018) confirmaram esse padrão (Padrão A) na maioria da sua amostragem composta de linfomas de Burkitt oriundos do Brasil, Argentina e Gana, em conjunto com a coorte de Lei *et al.*, assim, compararam com amostras de indivíduos “saudáveis” oriundos de Uganda, onde o padrão esteve presente em 10% das sequências. Liao *et al.*, associaram o Padrão A de classificação de Lei *et al.*, como um padrão de polimorfismos do gene e promotor da LMP1 como potencialmente oncogênico, por estarem presentes na maioria dos casos de linfoma de Burkitt dessas três regiões e em menor parte no grupo comparativo (LIAO *et al.*, 2018). Apesar da presença em mais de 90% da população mundial, a variação natural do EBV em carreadores assintomáticos no mundo ainda é pouco investigada (CORREIA *et al.*, 2017), onde muitos estudos utilizam controles de doenças não neoplásicas como grupo comparativo (GUIRETTI *et al.*, 2007; LORENZETTI *et al.*, 2014; GANTUZ *et al.*, 2017). No Brasil, ainda não se conhece a diversidade do EBV circulante em carreadores assintomáticos, o que pode dificultar a identificação de variantes oncogênicas.

#### 1.2.5 Neoplasias associadas ao EBV

A infecção primária por EBV ocorre cedo infância, em torno dos 5 anos de idade, e geralmente é subclínica, levando a infecção latente, onde o EBV irá persistir nas células B de memória ao longo da vida do hospedeiro. Caso a infecção inicial seja retardada até a adolescência ou idade adulta, pode ocorrer o desenvolvimento de uma doença infecciosa benigna, chamada mononucleose infecciosa (MNI), também conhecida como febre glandular. MNI é uma doença imunopatológica autolimitada causada por uma resposta robusta de células T ao EBV e é caracterizada por febre, letargia, linfadenopatia, esplenomegalia e amigdalite. Embora a maioria das infecções

por EBV sejam benignas e o desenvolvimento de tumor seja raro, o EBV é associado a aproximadamente 1,5% das neoplasias malignas em todo o mundo. O EBV foi associado a várias doenças malignas linfoproliferativas humanas, como linfoma de Hodgkin clássico, linfoma de Burkitt, doença linfoproliferativa pós-transplante e linfoma de células NK e T (PATTLE *et al*, 2006). EBV também está associado a doenças malignas epiteliais, como carcinomas de nasofaringe (RAAB-TRAUB, 2002; YOUNG *et al*, 2014) e aproximadamente 10% dos cânceres gástricos (GC) (TAKADA, 2000). No entanto, os mecanismos carcinogênicos mediados pelo EBV não são compreendidos em sua totalidade.

A epidemiologia das doenças associadas ao EBV denota uma distribuição geográfica incomum, que podem depender de fatores socioeconômicos, etnográficos, genéticos do hospedeiro e virais (FARRELL, 2019; CHABAY *et al*, 2020), sendo importante diferenciar fatores virais restritos geograficamente daqueles associados a processos tumorigênicos específicos. O Brasil é um país de território extenso, que possui cinco macrorregiões geográficas que se diferem em sua heterogeneidade cultural, ancestralidade, clima e desenvolvimento socioeconômico, onde a idade de soroconversão do EBV depende de fatores socioeconômicos (FIGUEIRA-SILVA *et al*, 2004). Adicionalmente, nosso grupo demonstrou que existem padrões epidemiológicos específicos de patologias malignas como o linfoma de Hodgkin clássico e o linfoma de Burkitt no sudeste brasileiro (HASSAN *et al*, 2008; BARROS *et al*, 2011; 2012), em relação aos padrões comumente relatados na literatura, que geralmente levam em conta dados de países com baixo ou elevado IDH (JARRET *et al*, 2014).

A relação entre o EBV e o linfoma de Hodgkin clássico (LHc) foi estabelecida inicialmente através de evidências epidemiológicas e sorológicas, sendo confirmadas posteriormente pela detecção do material genético clonal do vírus nas células neoplásicas Hodgkin e Reed-Sternberg (H-RS). Em conjunto com o conhecimento da capacidade transformadora do EBV, o material genético clonal do vírus nas células H-RS indicou que a infecção pode ter ocorrido antes da transformação maligna das células tumorais (MASSINI *et al.*, 2009). O EBV apresenta diferentes padrões de expressão gênica (latências I, II ou III) dependendo da patologia associada, no caso do LHc, o EBV apresenta-se com o padrão de latência II, onde as proteínas LMP1 e LMP2A estabelecem funções importantes na transformação celular (KUPPERS, 2009a). A frequência de associação do EBV com o LHc é bastante variável, dependendo de diversos fatores, como a idade e o nível socioeconômico. Em países com elevado IDH, LHc se desenvolve

principalmente em adultos e a associação varia de 30 a 40% com o EBV. Em contraste, a prevalência do EBV em células H-RS em países com baixo ou médio IDH pode ser extremamente elevada, como em Honduras (100%), China (100%), México (65%) e Malásia (93%), onde demonstram uma forte correlação entre o pico de incidência etária em crianças com o LHc EBV+ (HERBST *et al.*, 1991; AMBINDER *et al.*, 1993; CHANG *et al.*, 1993; ENGEL *et al.*, 2000; CHABAY *et al.*, 2020). A epidemiologia do LHc é complexa, e definida atualmente em um modelo que representa quatro doenças distintas, levando em consideração idade de incidência da doença, a presença do EBV, fatores imunes e socioeconômicos. Esse modelo de quatro doenças, é definido: pela incidência do LHc EBV+ muito cedo na infância características de países com baixo IDH; incidência do LHc EBV+ em jovens adultos, característica de países com elevado IDH, onde se hipotetiza que a MNI pode ser um risco; a incidência de LHc sem associação com EBV, em adultos mais velhos, cenário comum em países com elevado IDH; e uma incidência do LHc EBV+ no indivíduo mais idoso, onde a senescência do sistema imune pode levar reativação constante viral e aumento do risco (JARRET *et al.*, 2014). Dito isso, nosso grupo de pesquisa observou que a epidemiologia do LHc associado ao EBV na nossa região é intermediária da observada em países com baixo e elevado IDH, visto que a associação é em 50% dos casos de LHc em uma mediana de idade de 14 anos (BARROS *et al.*, 2011; 2012). Em conjunto com outros estudos que apontam que o local de origem tem impacto no desenvolvimento da doença (GLASER *et al.*, 2014). Dessa forma, podemos concluir que o modelo epidemiológico mais atual de quatro doenças do LHc, não representa a epidemiologia observada na nossa região geográfica, onde dados relativos a outras áreas geográficas podem não representar a complexa interação entre o EBV-hospedeiro no Brasil.

Pela oncogênese mediada por vírus ter como característica o surgimento e seleção de variantes com potenciais oncogênicos diferentes, como exemplificado pelos tipos de HPV de alto e baixo risco (ZUR HAUSEN, 2009), e os mecanismos carcinogênicos mediados pelo EBV não serem compreendidos em sua totalidade, postula-se que variantes oncogênicas virais podem atuar nesses mecanismos, podendo favorecer o desenvolvimento do câncer, como já demonstrado no caso de cepas com maior potencial oncogênico em regiões geográficas específicas (HU *et al.*, 1993; ZUERCHER *et al.*, 2012; LORENZETTI *et al.*, 2012; SUEUR *et al.*, 2014; KUPPERS, 2003). Contudo, não existem dados na literatura sobre a secreção do EBV, a respeito detecção, carga viral na saliva (principal forma de transmissão), e a

diversidade viral da população assintomática brasileira, o que pode dificultar a associação de variantes com potencial oncogênico na nossa região. Ademais, dados sobre a diversidade do EBV circulante no Brasil em casos patológicos são insipientes na literatura, sendo nossa região geográfica pouco representada em estudos globais sobre a variabilidade do EBV (PALSER *et al*, 2015; Lei *et al* 2015; CHIARA *et al*, 2016; GANTUZ *et al*, 2017; LIAO *et al* 2018; BLAZQUEZ *et al*, 2021).

Dito isso, são levantadas as seguintes questões: pelo Brasil ser um país heterogêneo em termos socioeconômicos, étnicos, culturais, que acabam por refletir em padrões epidemiológicos específicos em diferentes doenças associadas ao EBV (ABREU *et al*, 1997; HASSAN *et al*, 2008; BARROS *et al*, 2011; 2012; FERREIRA *et al*, 2012), e uma vez identificados alguns polimorfismos de importância oncológica da LMP1 no Brasil em associação com casos de linfomas (GUIRETTI *et al*, 2007; LEI *et al*, 2015; LIAO *et al*, 2018), poderia ter alguma variante da LMP1 associada à linfomas no Brasil? Outros oncogenes, ou conjunto de genes virais poderiam estar associados a doenças malignas na nossa população?

A partir dessas considerações, o presente trabalho tem como objetivo estudar a diversidade do EBV circulante no Brasil tanto em casos de patologias benignas e malignas hematológicas associadas ao EBV, quanto em portadores assintomáticos EBV+, na tentativa de contribuir com o conhecimento da diversidade do EBV na América do Sul e destacar potenciais determinantes genéticos de oncogenicidade, mediados pelo EBV na nossa região geográfica.

## **2. OBJETIVOS**

### **2.1 Objetivo geral**

Avaliar a diversidade genética do vírus Epstein-Barr circulante no Brasil através da amostragem de indivíduos com infecções naturais, linfoproliferações benignas e malignas

### **2.2 Objetivos secundários:**

- I. Classificar o EBV circulante na nossa região geográfica através da análise de sequências aminoacídicas da oncoproteína LMP1 nos casos de patologias benignas e malignas, e carreadores assintomáticos EBV+.
- II. Investigar a relação de polimorfismos da LMP1 de importância oncológica com as variantes de LMP1 classificadas no presente trabalho.
- III. Caracterizar a diversidade do EBV circulante na América do Sul, através de análises filogeográficas bayesianas do oncogene *LMP1*.
- IV. Investigar a presença do haplótipo oncogênico viral Tipo/V3 nas variantes de LMP1 classificadas no presente trabalho.
- V. Realizar a montagem do genoma do EBV da nossa região geográfica.

### 3 METODOLOGIA

#### 3.1 Desenho do estudo

Com o objetivo de caracterizar a diversidade genética do EBV circulante no Brasil, foram investigados genótipos virais em dois grupos amostrais: (i) indivíduos sem câncer (doadores assintomáticos), residentes no Brasil, através da detecção do EBV na saliva; (ii) amostras de sangue periférico ou biópsia de pacientes atendidos no CEMO/INCA com doenças linfoproliferativas EBV+ (Figura 1.7).

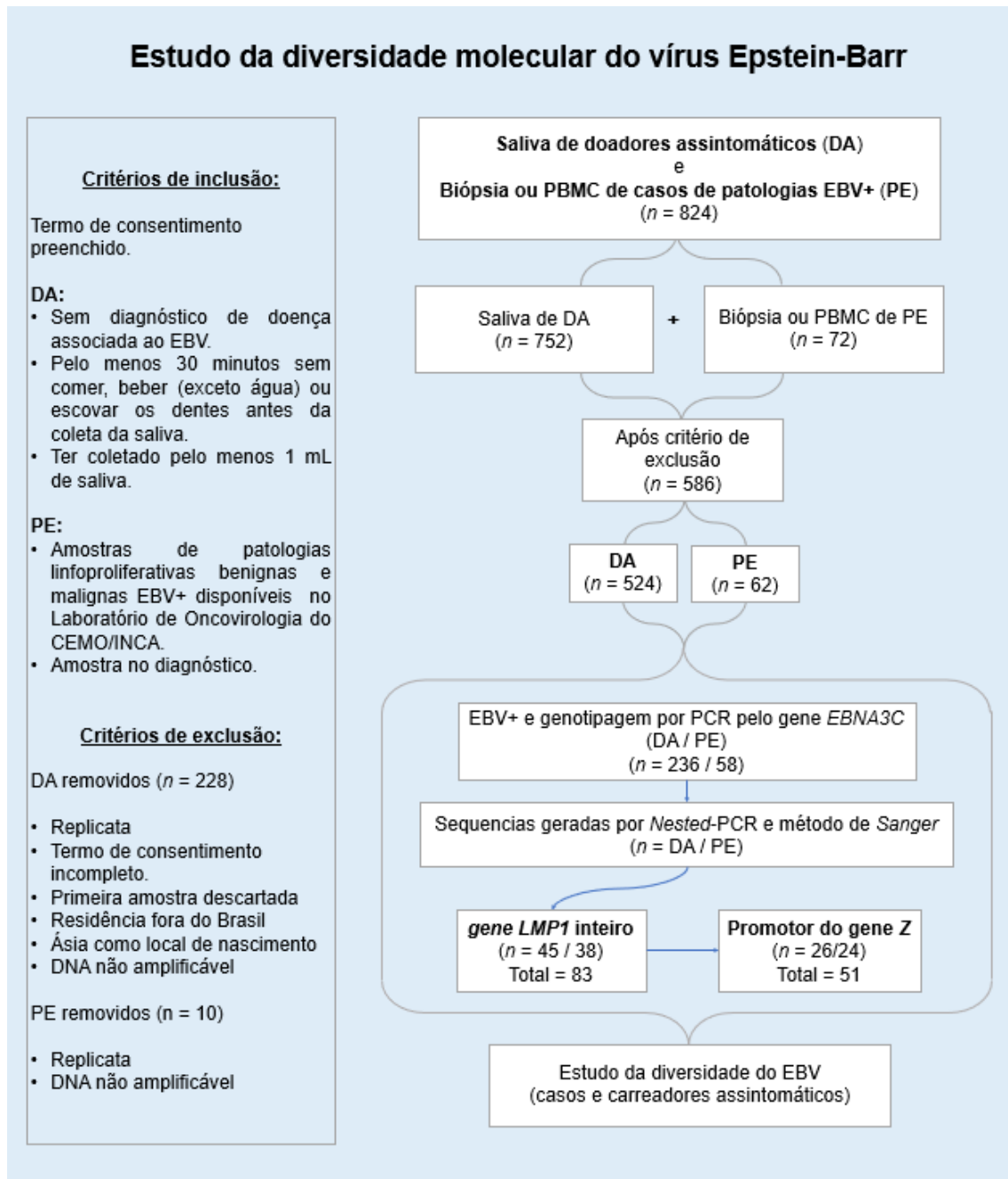


Figura 3.1 - Diagrama descritivo do estudo da diversidade molecular do EBV circulante no Brasil. DA: doadores assintomáticos; PBMC: células mononucleares do sangue periférico (do inglês: *peripheral blood mononuclear cells*); EBV: vírus



Epstein-Barr; PE: patologias EBV+; DNA: ácido desoxirribonucleico (do inglês: *deoxyribonucleic acid*); EBNA3C: antígeno nuclear do Epstein-Barr 3C (do inglês: *Epstein-Barr nuclear antigen*); PCR: reação em cadeia da polimerase (do inglês: *polymerase chain reaction*); qPCR: PCR quantitativo; LMP1: proteína latente de membrana 1 (do inglês: *latent membrane protein*);

## **3.2 Descrição da população estudada**

### **3.2.1 Casos de doenças hematológicas EBV+**

Para a análise da diversidade do EBV no Brasil, foram selecionadas 38 amostras de pacientes diagnosticados no INCA enviadas ao Laboratório de Oncovirologia do CEMO/INCA, onde foi confirmada a positividade para o EBV por PCR em tempo real. Estes casos são classificados em a) linfoproliferações benignas (n=4), diagnosticadas com hiperplasia linfoide reativa (HLR); b) doença linfoproliferativa pós-transplante (PTLD) (n=1); c) casos de linfomas (n=38); sendo 26 diagnosticados com LHc e 8 com LB. A mediana de idade foi 26 anos (3 – 79).

Os pacientes com linfomas e processos reativos linfoides foram diagnosticados na Divisão Integrada de Patologia (DIPAT) do INCA, no Rio de Janeiro - RJ, segundo critérios morfológicos e imunohistoquímicos, estabelecidos pela classificação da OMS (SWERDLOW SH, 2008). Nos casos de LHc, foi assumida uma associação com o EBV devido à alta carga tumoral na biópsia do tumor fresco (ANEXO I – Tabela S2).

Os tipos de amostras estudadas variaram de acordo com o tipo de doença. Foram estudados biópsia de tecidos frescos (linfonodos) nos casos de LHc, LB, e HLR. Já nos casos PTLD, foram estudadas amostras de sangue periférico.

### **3.2.2 Doador assintomático**

Para o desenvolvimento do estudo preliminar de detecção do EBV na saliva de indivíduos assintomáticos, foram incluídos a 1º coleta de saliva de 524 doadores de saliva voluntários de diferentes localidades do estado do Rio de Janeiro. Desses participantes, foram coletadas até o momento, 752 amostras de saliva. O número de amostras é superior ao número de voluntários, uma vez que a liberação do vírus na saliva é um evento aleatório, e, deste modo, indivíduos que não evidenciaram a presença do vírus na primeira coleta foram convidados a fazer uma nova doação na tentativa de genotipar a variante viral para o estudo da diversidade do EBV. As coletas foram realizadas por amostragem de conveniência, e sua inclusão ou exclusão no estudo seguiram os critérios apresentados na Figura 1.7.

Para a coleta de amostras de saliva dos indivíduos residentes no estado do Rio de Janeiro, foi utilizado tubo plástico estéril e solicitados pelo menos 5 mL de saliva líquida. A doação foi realizada sempre após intervalo superior a 30 minutos após alimentação, bebida (exceto água) ou após escovar os dentes. A duração da coleta foi de no máximo 30 minutos, e, em seguida, a amostra foi identificada e acondicionada em um saco de coleta único para cada doador. As amostras foram armazenadas imediatamente após a coleta em gelo ou geladeira, onde permaneceram por até três dias para a aliquotagem e posterior armazenamento em freezer a -20°C por tempo indeterminado.

### 3.2.3 Dados clínicos e epidemiológicos

#### 3.2.3.1 Casos diagnosticados no INCA

Os dados clínicos e demográficos dos pacientes diagnosticados no INCA foram obtidos através dos prontuários médicos. As informações coletadas incluem idade, sexo, local de procedência do paciente.

#### 3.2.3.2 Doadores assintomáticos

Os dados sociodemográficos da população sadia foram obtidos através do questionário socioeconômico preenchido no ato da coleta de amostra (saliva). As informações solicitadas foram: sexo, idade, autodeclaração da cor da pele de acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE), estado civil, filhos, ancestrais familiares, se costuma beijar pais e/ou filhos na boca, local de nascimento, local de residência até 18 anos, com quantos e qual(is) parente(s)/criador(es) residiu até 18 anos, grau de escolaridade do próprio e parente(s)/criador(es), local de residência atual, com quantas pessoas reside atualmente, faixa da renda da casa, assim como dados referentes a histórico de saúde de mononucleose e/ou doença crônica e/ou câncer (ANEXO III).

Embora não tenhamos avaliado o estado sérico da infecção pelo EBV, visto que a soroconversão pelo EBV no Brasil e na América Latina ocorre no início da infância (MENDOZA *et al.*, 1989; MARTINEZ *et al.*, 2012; FERRÉS *et al.*, 1995; FIGUEIRA-SILVA *et al.*, 2004), consideramos local de origem até 18 anos como representativo do EBV circulante nessas regiões. Desta forma, os 524 residentes no Brasil foram classificados de acordo à região em que viveram até os 18 anos.

#### 3.2.4 Critérios Éticos

Foram adotadas medidas que proporcionam confidencialidade e privacidade dos pacientes e doadores, assim como sigilo e a segurança dos dados obtidos. Esse estudo foi aprovado pelo Comitê de Ética e pesquisa do INCA (CAAE nº 3.56999916.5.0000.5274).

### **3.3 Métodos para a extração de ácidos nucleicos**

Condições livres de DNase ou RNase foram mantidas no local de extração, assim como nos instrumentos utilizados.

#### **3.3.1 Extração de DNA a partir de tecido à fresco**

A extração do DNA a partir de amostras de tecido à fresco foi feita com o kit comercial Wizard® Genomic DNA Purification Kit (Promega). Cerca de 10 a 20 mg de tecido foi cortado em pequenos pedaços e macerado. O tecido foi adicionado em um microtubo, junto com 600 µL da solução de lise e 17,5 µL de Proteinase K (20mg/mL) e incubado à 65°C por 12 horas. Após a incubação, foi adicionado 200 µL da solução de precipitação de proteína. A solução foi homogeneizada com o auxílio de um vortex, incubada no gelo por 5 minutos e centrifugada por 4 minutos à 12.000 rpm. O sobrenadante foi transferido para um novo tubo, seguido pela adição de 600 µL de isopropanol e centrifugação por 1 minuto à 12.000 rpm. O sobrenadante foi descartado e 600 µL de etanol foi adicionado e centrifugado por mais 1 minuto à 12.000 rpm. O etanol foi retirado e as amostras foram deixadas em temperatura ambiente para a evaporação residual do etanol. O DNA foi finalmente reidratado com 30-50 µL de solução de reidratação de DNA, de acordo ao tamanho do pellet e armazenado à 4°C até a sua utilização.

#### **3.3.2 Extração de DNA de saliva pelo método de "salting out"**

A extração foi realizada através de um método padronizado especialmente para o presente trabalho, visando contar com um método rápido e de baixo custo para processamento de um grande número de amostras. O volume inicial de saliva foi de 500 µL. Após o descongelamento, a amostra foi centrifugada em de 17000 x g por 5 minutos. Posteriormente o sobrenadante foi descartado e adicionou-se 500 µL de NaOH 50 mM. Em seguida, a mistura foi homogeneizada com auxílio de um vortex e incubada em termobloco por 5 minutos à 95°C. A amostra foi então homogeneizada em vortex e neutralizada com Tris-HCl (1 M, pH 6,8 – 10% v/v de NaOH), seguida por homogeneização por inversão (10 vezes) e posterior centrifugação à 8609 x g por 5

minutos. Após a centrifugação, o sobrenadante foi transferido para um novo tubo e o sedimento descartado. O sobrenadante foi armazenado de 2 a 8°C, caso fosse utilizado em até duas semanas, e à -20°C para o armazenamento final.

### 3.3.3 Isolamento de células mononucleares a partir do sangue periférico e extração de DNA

A partir da coleta do sangue periférico (SP) com tubo estéril contendo EDTA, foi utilizado o reagente Ficoll Histopaque (Sigma-Aldrich®) para a separação das células mononucleares do SP por centrifugação em gradiente de densidade. Para tal, o reagente Ficoll foi utilizado na proporção 1:1 do SP. Em um novo tubo, o SP foi colocado vagarosamente acima do reagente Ficoll e então o tubo foi centrifugado à 2000 rpm em temperatura de 25 °C por 25 minutos, sem freio ou aceleração. Logo após, o anel de células mononucleares foi retirado e colocado em um novo tubo, para realização de duas lavagens com PBS 1X à 1200 rotações por minuto (rpm) por 10 minutos. Posteriormente, procedeu-se à contagem de células em câmara de Neubauer e a proporção foi ajustada para 10-20 x 10<sup>6</sup> células para cada 1 mL do reagente DNAzol, em seguida o DNA foi extraído de acordo com o protocolo do fabricante e armazenado de 2 a 8 °C.

### 3.3.4 Avaliação qualitativa e quantitativa do DNA extraído

A integridade do DNA extraído foi avaliada por eletroforese em gel de agarose à 0,8% em tampão TAE (Tris-Acetato-EDTA) (SAMBROOK; RUSSELL, 2001). A quantificação do DNA foi realizada em um espectrofotômetro Nanodrop® 9000 (Thermo Scientific®), por meio de leituras feitas nos comprimentos de onda ( $\lambda$ ) 260 nm e 280 nm. A pureza das amostras foi estimada pela relação 260/280. Amostras com valores entre 1,8 e 2,0 foram consideradas com grau de pureza satisfatório e desta forma utilizadas nas análises posteriores.

### 3.3.5 Avaliação da amplificabilidade do DNA através de PCR para gene constitutivo

Para determinar a amplificabilidade do DNA e testar a presença de inibidores da PCR que pudessem levar ao aparecimento de falsos negativos, foram realizadas reações com iniciadores específicos para o gene constitutivo  $\beta$ -globina (STEFANOFF *et al*, 2003). As reações de PCR foram desenvolvidas em uma etapa, em volumes finais de 30  $\mu$ L, com 35 ciclos, temperatura de "annealing" de 60°C, uso de 1,5 mM de MgCl<sub>2</sub>, concentração de 0,33  $\mu$ M de cada primer e 1U de Taq DNA Polimerase

(Invitrogen®). O tamanho esperado do produto de PCR foi de 167 pb. Os produtos foram avaliados em géis de agarose à 2% contendo brometo de etídio numa concentração de 0,5 µg/mL.

### **3.4 Métodos para a caracterização da associação com o EBV**

#### **3.4.1 Detecção e tipificação do EBV por PCR para o gene *EBNA3C***

A reação tem como objetivo a detecção, mesmo em carga viral baixa, e a tipificação viral (Tipo 1 e/ou Tipo 2). Os oligonucleotídeos utilizados foram EBNA3CS 5'-AGAAGGGGAGCGTGTGTTGT-3' e EBNA3CAS 5'-GGCTCGTTTTTGACGTCGGC-3' (SAMPLE et al., 1990), que permitem a detecção do EBV1 (produto de 153 pb) e EBV2 (produto de 246 pb) na mesma reação. As reações continham 50 µL de volume final, 2 mM de MgCl<sub>2</sub>, 0,33 µM de cada primer, 1U de Taq DNA polimerase Platinum® (Invitrogen®) e 200-300 ng de DNA genômico. O perfil térmico consistiu em 40 ciclos com 94°C por 45 segundos, 67°C por 1 minuto e 72° por 45 segundos. Para avaliar o número mínimo de cópias do vírus que podem ser detectadas por esta técnica, em uma massa determinada de DNA genômico, foi realizado um teste de sensibilidade com construção de uma curva, utilizando diluições seriadas de DNA da linhagem celular Namalwa (CRL-1432; American Type Culture Collection), que contém 2 cópias do genoma do EBV por célula. A reação mostrou ter uma sensibilidade para detectar até 30 cópias do DNA viral.

#### **3.4.2 Quantificação da carga viral do EBV por PCR quantitativo (*Real-Time* PCR)**

Para quantificação do EBV presente em saliva ou em células mononucleares infectadas pelo vírus, foi utilizada a metodologia de PCR em tempo real com sondas TaqMan® como descrito por NIESTERS *et al* 2000. Para a quantificação absoluta de carga viral foi empregada uma curva padrão externa (DNA plasmidial) com pelo menos 5 logs de amplitude. Para a amplificação e quantificação do EBV foram utilizados primers e sondas específicos e complementares à região gênica BNT p143 (baseado no protótipo B95.8 V01555), utilizando os seguintes primers e sondas EBV BN-F 5'-GGAACCTGGTCATCCTTTGC-3', EBV BN-R 5'-ACGTGCATGGACCGGTTAAT-3' e Sonda (5' 3' FAM) 5'-CGCAGGCACTCGTACTGCTCGCT-3' (3' MGB/BHQ).

A reação constituía-se dos seguintes reagentes: 1X do TaqMan Universal PCR Mastermix, 900nM de primers, 200nM de sonda e 5 uL do DNA da amostra. Cada reação foi realizada com duplicata técnica e a leitura foi feita através do aparelho Viia7

(Applied Biosystems). Para o cálculo de cópias de EBV por número de células, é preciso calcular também a quantidade presente de células, através da qPCR de um gene constitutivo, onde n° de cópias do gene constitutivo é dividido por dois por causa dos dois alelos por célula. Assim:  $(106 \times \text{número de cópias do vírus EBV}) / (\text{número de cópias do gene humano B2MG/2})$ , o resultado é obtido para quantidade de EBV por cada 106 células.

Nas amostras oriundas de saliva, a extração do DNA foi feita através de 500 uL de saliva e o volume final da extração foi de 500 uL de DNA, desse DNA utilizou-se 5 uL de DNA (100x menos), então, para se obter o resultado em 500 mL de saliva, foi multiplicado o resultado vezes o fator de diluição 100, com o objetivo de se obter o resultado final em mL de saliva, o resultado em 500mL foi multiplicado x2.

### 3.5 Métodos para a análise molecular de polimorfismos do EBV

#### 3.5.1 Amplificação do oncogene latente *BNLF1* (*LMP1*)

A estratégia de amplificação da região codificante do gene *BNLF1* consistiu em amostragens com >10.000 EBV/mL de saliva, amplificando um fragmento de 1.590 pb incluindo a sequência completa do gene, seguido por três nested-PCRs, para as regiões N-terminal (N-ter), transmembrana (TM) e C-terminal (C-ter), utilizando 2 µL do produto da primeira etapa (Figura 1.8).

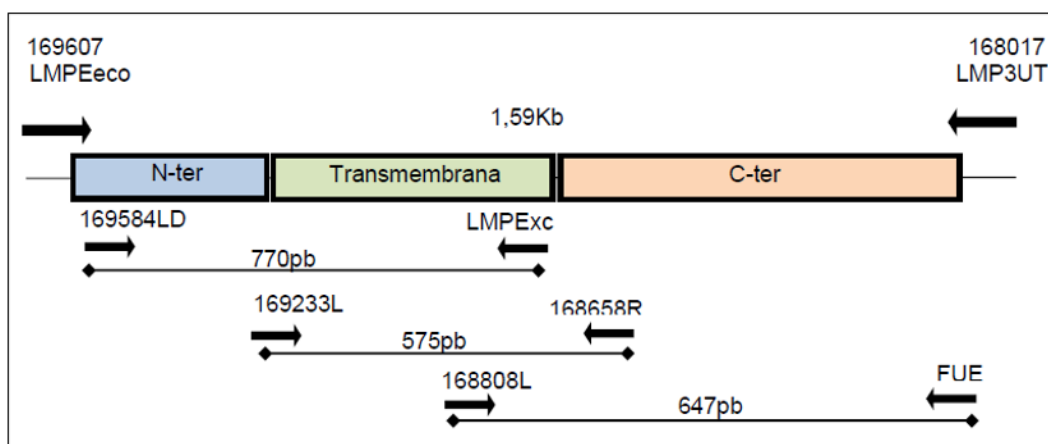


Figura 3.2 - Esquema da região codificante do gene *LMP1* mostrando a estratégia de amplificação e a localização dos primers. As caixas indicam as três regiões distintas do gene *LMP1* (N-ter, Transmembrana e C-Ter), enquanto as barras indicam o tamanho dos fragmentos e as setas o sentido de cada primer (Retirada de Garcia-Costa, INCA, 2012).

Para a primeira amplificação, foram utilizados os *primers* LMP3UT e LMPEco, que amplifica um produto de 1.590 pb, em uma reação de 50 µL de volume final.

Utilizou-se 2,5 mM de MgSO<sub>4</sub>, solução *enhancer* 1X [betaína 2,7 M (Sigma B-0300); DMSO 6,7% (Sigma D-8418); DTT 6,7 mM (Sigma)], 0,6 µM de *primers* e 2U da enzima *Taq Polymerase High Fidelity Platinum*® (Invitrogen). A região N-ter foi amplificada com os primer LMPExc e 169584Ld, que amplifica um produto de 779 pb, com 2 µL do produto de PCR da primeira etapa em uma reação de 50 µL de volume final. Adicionou-se 2 mM de MgSO<sub>4</sub>, 1 µL de DMSO (Sigma), 0,6 µM de *primers* e 2U da enzima *Taq Polymerase High Fidelity Platinum*® (Invitrogen). A região transmembrana foi amplificada com os *primers* 168658R e 169251L, que amplifica um produto de 575 pb, a partir de 2 µL do produto da reação para o gene inteiro, em uma reação de 50 µL de volume final. Incluiu-se 2 mM de MgSO<sub>4</sub>, 1 µL do coadjuvante DMSO, 0,4 µM de *primers* e 2U da enzima *Taq Polymerase High Fidelity Platinum*® (Invitrogen). A região C-ter foi amplificada com os *primers* FUE e 168808L, gerando um produto de 647 pb a partir de 2 µL do produto da reação para o gene inteiro, em uma reação de 50 µL de volume final. Utilizou-se 2 mM de MgSO<sub>4</sub>, 1 µL de DMSO, 0,4 µM de *primers*, e 2U de enzima *Taq Polymerase High Fidelity Platinum*® (Invitrogen). Ao final, os produtos de PCR foram avaliados em eletroforese de géis de agarose 2%, corados com *Diamond*® para a visualização em luz ultravioleta e armazenados a -20°C para posterior purificação e sequenciamento. As sequências dos primers citados acima foram retirados de Edwards *et al*, 2004.

### 3.5.2 Amplificação do promotor do gene Z (Zp)

Para a amplificação da região do promotor Zp (-221 à +12), foram utilizados os *primers* Zp1: 5'- CGAGTTTGGGTCCATCATCT -3' (103138–103157) e Zp2: 5'- CTGACCCCCGAACCTTAATGA -3' (103457–103476) (LORENZETTI *et al.*, 2009), que amplificou um produto final de 299 pb, em uma reação de 50 uL de volume final. Para a reação de PCR foi utilizado 2 mM de MgCl<sub>2</sub>, solução *enhancer* 0,5X [betaína 2,7 M (Sigma B-0300); DMSO 6,7% (Sigma D-8418); DTT 6,7 mM (Sigma)], 0,6 µM de *primers* e 1,5U da enzima *Taq Polymerase Platinum*® (Invitrogen). O perfil térmico consistiu em 1 ciclo de 94 °C de 2 minutos, 35 ciclos a 94°C de 30 segundos, 58°C por 30 segundos, 72°C por 45 segundos e 1 ciclo de 72 °C de 7 minutos. Os produtos de PCR foram avaliados em géis de agarose 2% contendo brometo de etídeo (0,5 µg/mL) e armazenados a -20°C para posterior purificação e sequenciamento.

### 3.5.3 Purificação dos produtos de PCR

De acordo com a intensidade do produto de PCR foram utilizadas duas estratégias. Na presença de bandas de intensidade forte, a purificação dos produtos foi realizada com o Kit comercial *QIAquick PCR Purification*® (Qiagen), seguindo o protocolo do fabricante. Na presença de bandas de intensidade fraca, procedeu-se com a técnica de *ExoSAP*, que utiliza duas enzimas, *Exonuclease I + shrimp Alkaline Phosphatase (SAP)* (ThermoFisher®). Para cada volume de produto de PCR que foi utilizado na reação de sequenciamento, acrescentou-se 1U da *Exonuclease I* e 0,25 U da SAP. O perfil térmico da reação foi de 37°C por 30 minutos e 80°C por 15 minutos. Ao final da reação, segue-se para a reação de sequenciamento.

#### 3.5.4 Sequenciamento pelo método de *Sanger*

A quantificação do produto purificado é feita em um gel de agarose 2% com a utilização de 2 µL do marcador *Low DNA Mass Ladder* (Invitrogen®) misturado com 2 µL de amostra. As intensidades das bandas foram comparadas para se obter uma quantificação aproximada das amostras. As reações de sequenciamento foram realizadas em volume final de 10 µL, utilizando para cada reação 1 µL do reagente BigDye® v.3.1 (*Applied Biosystems*), 1,5 µL do tampão de sequenciamento, 3,2 pmol de um dos *primers* (F ou R) e um volume de amostra que variou conforme sua concentração, para garantir uma concentração ótima de 5-20 ng de DNA. A reação de sequenciamento consistiu em 40 ciclos de 96°C por 15 segundos, 50°C por 5 segundos e extensão por 60°C por 4 minutos. Ao término da reação de sequenciamento, foi dado um rápido *spin* na placa, e sua vedação foi removida. A seguir, a reação foi purificada com EDTA 125 mM, pH 8,0 e etanol 100% (Merck), em seguida, a placa foi vortexada cuidadosamente por no máximo 10 segundos, seguido por incubação a temperatura ambiente por 15 minutos. Posteriormente, a placa foi centrifugada a 3.700 rpm por 45 minutos a 4°C. Após centrifugação, a vedação da placa foi retirada e o sobrenadante foi eliminado. Depois disso, acrescentou-se etanol 70% fresco, a placa foi vortexada cuidadosamente por no máximo 10 segundos e centrifugada à 4 °C por 15 minutos na rotação de 3.700 rpm. O sobrenadante foi novamente eliminado por inversão. Para a secagem, a placa ficou por 10 minutos à 60 °C em um termobloco com a tampa do aparelho aberta. Após a secagem, a placa foi vedada, acondicionada em papel alumínio para proteger da luz e armazenada no freezer a -20°C. O sequenciamento foi realizado em um instrumento ABI 3130xl (*Applied Biosystems*).



### 3.5.5 Sequenciamento de nova geração para obtenção do genoma do EBV

A amostra CEMO3, oriunda de um paciente com linfoma de Burkitt foi progressivamente diluída em água ultrapura e aferida por método fluorimétrico Qubit® (Invitrogen, EUA) até alcançarem a concentração de 0,4 ng/uL. Em seguida, foram preparadas as bibliotecas utilizando o *kit Nextera XT*®, que faz uso do método de fragmentação enzimática por transposases (Illumina, EUA). As etapas posteriores de adição dos adaptadores, colocação do par de índices e a normalização das bibliotecas foram realizados seguindo os procedimentos estabelecidos pelo fabricante, a partir de uma concentração total de 2 ng de DNA total da amostra. A normalização das bibliotecas foi realizada através da utilização de microesferas (*beads*) magnéticas com capacidade limitada de carregamento, desta forma “nivelando” a concentração das bibliotecas. Em seguida, as bibliotecas foram quantificadas separadamente por Qubit®, diluídas em Tris-HCL pH 8,0 para a concentração final de 0,4 nM e distribuídas igualmente em um *pool* final para o sequenciamento. O sequenciamento ocorreu sob a concentração de 13 pM com fragmentos entre 400 e 500 pb, seguindo as recomendações do fabricante. O sequenciamento foi realizado na plataforma Illumina *HiSeq* (Illumina, EUA), com a configuração de *Rapid V2 -Paired end* - 200 ciclos.

#### 3.5.5.1 Amplificação de *gaps* resultantes do NGS por sequenciamento de *Sanger*

Foram identificadas 17 regiões (*gaps*) após a montagem do genoma do EBV da amostra CEMO3 que não tiveram coberturas pelas sequências geradas pelo método de NGS. Para sanar este problema técnico, procedeu-se com a amplificação dessas regiões por PCR e posterior sequenciamento de *Sanger*. Para tal, 4 *gaps* foram cobertos por primers encontrados na literatura (primers Gap01, 02, 03 e 04). Para os outros 13 *gaps*, primers foram desenhados como auxílio do programa Primer3 web version 4.0 e ApE – A plasmid editor version 2.0.55. Os 17 conjuntos de primers cobriram um total de 8672 pb. Os primers, a referência, assim como o tamanho dos fragmentos estão disponíveis na Tabela 3.1 abaixo.

Para a amplificação das regiões dos primers Gap01, 02, 04, 16 e 17, foi utilizada uma reação de 50 µL de volume final, com 2,5 mM de MgSO<sub>4</sub>, 0,2mM dNTPs, 0,6 µM de primers e 2U da enzima Taq Polymerase High Fidelity Platinum® (Invitrogen). O perfil térmico caracterizou-se por 5 minutos à 94°C, 35 ciclos de 94°C por 45 segundos, 58°C por 30 segundos, 68°C por 90 segundos e 7 minutos de extensão à 68°C. Para as regiões restantes, foi utilizada uma reação de 50 µL de volume final,

com 2 mM de MgCl<sub>2</sub>, 0,2mM dNTPs, 0,6 µM de primers e 2U de enzima Taq DNA Polimerase Platinum® (Invitrogen). O perfil térmico consistiu em 5 minutos à 95°C, 35 ciclos de 95°C por 45 segundos, 58°C por 30 segundos, 72°C por 45 segundos e 7 minutos de extensão à 72°C. Os produtos de PCR foram avaliados em géis de agarose 3% corados com Diamond® para a visualização em luz ultravioleta e armazenados a -20°C para posterior purificação e sequenciamento.

O preenchimento dos gaps do sequenciamento da amostra 7788 encontra-se em curso, de acordo com as etapas descritas acima para a amostra CEMO3.

Tabela 3.1 - Sequência dos primers utilizados para a amplificação dos gaps no genoma montado por NGS.

	<b>ID primer</b>	<b>Sequência do primer 5'- 3'</b>	<b>Região coordenada</b>	<b>Produto</b>
<b>Gap01*</b>	60125	AACAGGCGGGCGAATGTGTAAT	60125	1272
	61397	ACCTTTCATCCGAACCTCCTCAGGT	61397	
<b>Gap02*</b>	96177	TTGCATTGGCTGCAAAGGGG	96177	614
	96791	CACCACGTCCACGACCTCTC	96791	
<b>Gap03*</b>	157382	TCAGCACCATGTTCTGCCTCTT	157382	507
	157889	TTCGTGACCAACACAACCGT	157889	
<b>Gap04*</b>	51199	CCGAAATAGGGCCTTGCCATCAAT	51199	1046
	52245	ATTCAGGACTACCTGCGCGACTT	52245	
<b>Gap05**</b>	39911	CTGGTTCTGGGGCTCCTCT	39911	221
	40112	CAGGTCGGACTAGCGGATG	145435	
<b>Gap06**</b>	40944	CACCCTTCCTCCTTCCGTTT	40944	482
	41405	AAATCACCCCAAGTCCCCTTC	41405	
<b>Gap07**</b>	48386	CATTCAGTGGATGTGGTACCG	48386	242
	48608	TGAAACACTTGTAGGCCCGG	48608	
<b>Gap08**</b>	63017	CGGCCCGGCGTCTAAATATT	63017	290
	63287	GAACGGGTCACCTCTGCTTGA	62287	
<b>Gap09**</b>	64555	CGGAGATTGAGAGTGACATGG	64555	290
	64971	GTGGCTGTTGATCATCCTCAT	64971	
<b>Gap10**</b>	87439	CAGTGAAGCGCACAAATGTTA	87439	306
	87721	TATTTGGCTGTGACACTGACG	87721	

<b>Gap11**</b>	88592	GTCTCAAGATCCTGTGCATGG	88803	452
	89023	CCTGTGCTTCAGCAGTAGCTT	88808	
<b>Gap12**</b>	100708	AATTCTTTTGGAGCCAGGTGT	100708	355
	101042	TTTGGACACCTCATGAAGACC	101042	
<b>Gap13**</b>	105498	AATTCGTAGAGCCACCAAGTG	105498	364
	105841	CAGATTTCCAGCCCTATGTCC	105841	
<b>Gap14**</b>	108015	CTCCTCCGTTTTACCGATGAC	108015	426
	108421	GGTAAAGGGGGAGGCATAAC	108421	
<b>Gap15**</b>	136417	ATGCTATCAGGTAACGCAGGA	136417	470
	136867	TCTCTTTTCCGCTTGTTGGAT	136867	
<b>Gap16**</b>	143795	TGGGTGGAGCTAGGTAGGATT	143795	656
	144430	TGTGTACCCAGGGTGAGAGAC	144430	
<b>Gap17**</b>	168997	CTCTCAAGGTCGTGTTCCATC	168997	679
	169655/170	CCTCCACTTTTTCCAGGAATG	169655	

\*Referência: SIMBIRI *et al.*, 2015. \*\*Referência: este trabalho.

### 3.6 Análise bioinformática

#### 3.6.1 Análise de sequências geradas pelo método de *Sanger*

As sequências foram alinhadas com uma sequência de referência da região estudada, utilizando como base o protótipo B95.8, número de acesso do Genbank NC\_007605, utilizando os programas BioEdit v.7.1.11 (HALL, 1999) e SeqMan Pro (DNASTAR ©1988-2013, versão 11.1.0(59),419), onde a qualidade dos eletroferogramas das sequências senso e antisenso do mesmo fragmento foi avaliada. As sequências senso e antisenso de cada amostra e cada fragmento da LMP1 foram agrupadas e alinhadas em um único arquivo, formando assim um sequência consenso (*contig*) contendo as três regiões da proteína, a qual foi salva em formato fasta. A partir desses alinhamentos os *inputs* foram montados para as análises subsequentes, onde as variações dos nucleotídeos foram verificadas.

##### 3.6.1.1 Reconstruções filogenética e filogeográfica da região codificante do gene *LMP1*

Para a classificação filogenética da oncoproteína LMP1, foi construído um banco de dados com as sequências obtidas dos casos patológicos, grupo comparativo e as sequências referências disponíveis no GenBank (Tabela 3.2). Este banco de

dados foi alinhado com o algoritmo MAFFT (KATO et al, 2012) e inspecionado no programa MEGA 7 (KUMAR et al, 2016) para identificação dos íntrons e regiões repetitivas.

As análises evolutivas de máxima verossimilhança foram realizadas no programa PhyML (GUIDON et al, 2003) e suporte dos ramos calculado a partir do teste aproximado da razão de verossimilhança (aLRT). As árvores geradas foram visualizadas e anotadas no programa FigTree (<http://tree.bio.ed.uk/software/figtree/>).

No intuito de explorar a diversidade das nossas amostras frente a diversidade mundial, foram adicionadas sequências do gene inteiro da oncoproteína LMP1 disponíveis no Genbank em agosto de 2021, utilizando o seguinte código na barra de busca: LMP1[Gene Name] AND Human herpesvirus 4[Organism] AND 1100 : 99999999[SLen], ou seja, sequências com anotação de LMP1 no nome do gene, de EBV e que tinham no mínimo 110 nucleotídeos de extensão. O local de procedência das sequências em relação à continente, país, e o período da amostragem se encontram descritos na Tabela 3.3.

Para a análise filogenética mundial foi escolhido o modelo de substituição nucleotídica GTR + I + G, e para análise filogeográfica do clado Raji/Argentine foi utilizado o modelo GTR + G, ambos modelos selecionados através do programa JModelTest.

Para a análise filogeográfica foi utilizado o programa *Beast*, com os seguintes parâmetros: população constante, relógio restrito e o modelo filogeográfico discreto simétrico. A corrida de MCMC teve 50.000.00 de passos. Ademais, foram utilizados a localidade de nascimento e moradia até os 18 anos dos indivíduos como local de origem do EBV, assumindo que a primo-infecção do EBV ocorreu na infância. Desta forma, duas amostras coletadas no Brasil foram assumidas uma como africana e outra peruana.

Tabela 3.2 - Lista de sequências de referência utilizadas na reconstrução filogenética da região codificante da LMP1.

ID	Fonte	Número de acesso	País de origem
B95.8	NCBI	V01555.2	EUA
GD1	NCBI	AY961628.3	China
GD2	NCBI	HQ020558.1	China
Akata	NCBI	KC207813.1	Japão

CAO	NCBI	X58140.1	China
Raji	NCBI	KF717093.1	Nigéria
AG876	NCBI	DQ279927.1	Ghana
Mutu	NCBI	KC207814.1	Kenya
Daudi	NCBI	LN827545	Kenya
China1	NCBI	AY337723.1	China
China2	NCBI	AY337724.1	China
Med+	NCBI	AY337722.2	EUA
Med-	NCBI	AY337721.2	EUA
NC	NCBI	AY337726.2	EUA
Alaskan	NCBI	AY337725.1	EUA
HKNPC1	NCBI	JQ009376.2	Hongkong
M81	NCBI	KF373730.1	Hongkong
LMP1_RH14	NCBI	KY118032	Argentina

ID: identificação da variante; NCBI: do inglês: *National Center for Biotechnology Information*; EUA: Estados Unidos da América;

Tabela 3.3 – Local e período da amostragem das sequências utilizadas nas análises filogenética mundial e filogeográfica.

Continente	N	País	N	Intervalo temporal
<b>América do Sul</b>	72	Brasil	4	1980
		Argentina	68	1980 - 2017
<b>América do Norte</b>	27	EUA	27	2008 - 2010
<b>Europa</b>	35	Suíça	35	2010
<b>África</b>	4	Ghana	4	1976 - 1990
<b>Ásia</b>	103	China	70	2015 - 2018
		Taiwan	26	2020
		Coreia do Sul	3	2008 - 2021
		Japão	3	2019
		Hong Kong	1	ND
–	–	Sem país	11	ND

N: número; EUA: Estados Unidos da América; ND: não disponível.

### 3.6.1.2 Caracterização de polimorfismos do promotor lítico Zp

Para a análise do promotor Zp, foram utilizados os protótipos B95.8 do tipo 1 (número de acesso: NC\_007605) e AG876 do tipo 2 (número de acesso: NC\_009334).

Uma vez que as duas variantes de maior importância oncológica descritas são: V1 que é considerado a protótipo, sendo a sequência presente no protótipo do tipo 1 do EBV; e a variante V3, que difere em três posições a respeito à sequência protótipo, [-141 (A>G), -106 (A>G), -100 (T>G)], presente na protótipo do tipo 2 do EBV (AG876) (Figura S1; Bristol *et al.*, 2018).

### 3.6.2 Análise de sequências geradas por NGS

Manipulações de arquivos de texto contendo sequências geradas por NGS foram feitas principalmente com auxílio de *scripts* com linguagem de programação Perl e Python. Após a retirada dos barcodes, a qualidade das sequências geradas foi avaliada por FastQC (ANDREWS, 2010). Os algoritmos BWA (LI e DURBIN, 2009) e Bowtie2 (LANGMEAD *et al.*, 2009b; LANGMEAD e SALZBERG, 2012) foram utilizados para a manipulação das sequências e montagem dos genomas baseada em uma sequência referência. Empregou-se a estratégia de uma filtragem direta de sequências semelhantes ao EBV, a partir de todas as sequências geradas pelo NGS, utilizando um banco de 113 genomas inteiros do EBV publicados e depositados no Genbank até 2017 (BAER *et al.*, 1984; PARKER *et al.*, 1990; DE JESUS *et al.*, 2003; ZENG *et al.*, 2005; DOLAN *et al.*, 2006; LIU *et al.*, 2011; KWOK *et al.*, 2012; LIN *et al.*, 2013; TSAI *et al.*, 2013; KWOK *et al.*, 2014; PALSER *et al.*, 2015; LIU *et al.*, 2016; TU *et al.*, 2017). Para análise das regiões sequenciadas por método de Sanger, conforme anteriormente descrito, e anotação do genoma montado utilizou-se o programa Geneious®. A sequência referência utilizada para a montagem do genoma e a anotação foi o protótipo B95.8 (Genbank acesso: NC\_007605).

### 3.7 Análises estatísticas

A existência de associações entre variáveis categóricas foi testada utilizando o teste de chi-quadrado de *Pearson* ou o teste exato de *Fisher*, dependendo do número de amostras avaliado. Para variáveis contínuas com distribuição não paramétrica foi utilizado o valor da mediana e os valores mínimo e máximo. Para medir a força de associação das variáveis categóricas, foi utilizada a medida de associação do tipo razão de chance (OR, do inglês: *odds ratio*), onde o valor =1, não tem associação, OR> 1 pode conferir um maior risco, e OR<1 pode conferir proteção. Nos testes, o valor de *p* menor do que 0,05 foi indicativo de significado estatístico. Todas as análises estatísticas foram realizadas com o programa SPSS, versão 20.0 (IBM®).

## 4. RESULTADOS

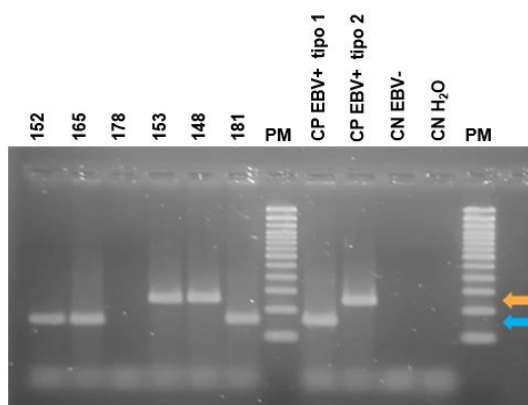
### 4.1 Caracterização da diversidade genética do oncovírus Epstein-Barr circulante no Brasil

#### 4.1.1 Características da população estudada

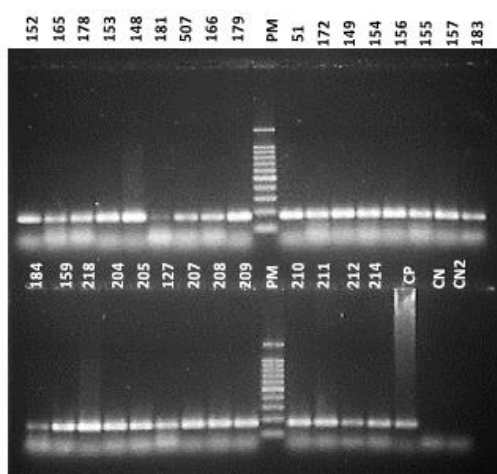
A diversidade genética do EBV foi estudada utilizando três abordagens: a tipagem pelo gene *EBNA3C*, a região codificante completa da oncoproteína LMP1, e o promotor do gene lítico *Z* (*Zp*). A população EBV+ estudada consistiu em um total de 83 indivíduos: 45 indivíduos sem diagnóstico de câncer (doadores assintomáticos – DA; com mediana de 27, mínimo de 5 e máxima de 94 anos, razão entre sexos F:M=2:1) e 38 indivíduos com doenças linfoproliferativas benignas [4 com hiperplasia linfoide reativa (HLR), com mediana de idade de 37 anos e razão entre sexos M:F=2] e malignas EBV+ [1 linfoproliferações pós-transplante (PTLD), com idade de 55, e sexo feminino; 7 linfomas de Burkitt (LB) e 26 linfomas de Hodgkin clássico (LHc), os casos de linfomas tiveram A mediana de idade foi 26 anos (3 – 79) e a razão entre sexos M:F=2,5:1].

O grupo de DA EBV+ foi proveniente de uma pesquisa preliminar em uma população amostral de 524 indivíduos de diferentes regiões do Brasil, cujas amostras de saliva foram utilizadas para detectar a liberação do EBV (Figura 4.1). O EBV foi encontrado em 240 amostras (46%), das quais 236 foram caracterizadas geneticamente em tipo 1 (79,2%) e tipo 2 (20,8%). As informações sociodemográficas dos indivíduos amostrados estão descritas na tabela suplementar (S) 1. A associação da presença e frequência do EBV na saliva, assim como a tipagem do EBV encontrado estão listadas nas tabelas S3 e S4, respectivamente.

(A)



(B)



**Figura 4.1 - Resultado ilustrativo da eletroforese em gel de agarose dos produtos da PCR para o gene viral *EBNA3C* e o gene da proteína constitutiva beta-globina.** (A) Resultado da PCR com alvo em um fragmento do gene *EBNA3C*. A seta azul aponta o fragmento de 153 pb relativo ao tipo 1 do EBV e a seta laranja representa o fragmento de 246 pb referente ao tipo 2 do EBV. (B) Resultado da PCR com alvo em um fragmento do gene constitutivo *beta-globina* de tamanho 167 pb. Em ambas as reações o peso molecular (PM) foi de 100 pb. CP: controle positivo; CN: controle negativo.

Indivíduos que autodeclararam ancestralidade materna de origem africana e ameríndia tiveram possibilidade significativamente maior de detecção do EBV na saliva (OR=2,067 e OR=1,920 –  $p=0,023$  e  $p=0,025$ , respectivamente) em comparação com indivíduos de ancestralidade materna autodeclarada de origem europeia. Neste trabalho, alguns indivíduos estrangeiros residentes no Brasil foram amostrados, sendo 15 da África (1 de Guiné-Bissau e 14 de Angola) e 13 da América do Sul (3 da Argentina, 5 da Venezuela, 1 do Uruguai, 1 da Colômbia, 1 de Cuba e 1 do Peru). Estes estrangeiros residentes no Brasil, principalmente os de origem africana até os 18 anos, tiveram maior detecção do EBV na saliva em comparação com os brasileiros, 86,7% *versus* 44,6%, respectivamente,  $p=0,001$ . A frequência do EBV em indivíduos com origem até 18 anos foi significativamente menor na cidade do Rio de Janeiro (principal local de coleta das amostras; OR=0,719 –  $p=0,037$ ). No mais, foi observada uma maior possibilidade de secreção do EBV na saliva em indivíduos cuja escolaridade paterna foi baixa ( $\leq 1^{\circ}$  grau completo), com OR=1,430 e  $p=0,046$ . A idade foi um fator associado ao tipo do EBV encontrado, com indivíduos  $\leq 18$  anos com a menor detecção do tipo 1 do EBV, (53,8% *versus* 80,7% em indivíduos  $\geq 19$  anos: OR=0,278, IC95%=0,089-0,872 –  $p=0,032$ ). Cabe ressaltar que na avaliação da carga viral do EBV foi encontrada  $4,55 \times 10^3$  de EBV/ $1 \times 10^6$  de células MN para as amostras de HLR. Nas patologias hematológicas malignas as medianas de carga viral foram de  $3,10 \times 10^4$ ,  $5,57 \times 10^8$  e  $3,26 \times 10^5$  EBV/ $1 \times 10^6$  de células MN para PTLN, LB e LHC, respectivamente. Na saliva de 240 doadores sadios EBV+, a mediana de carga viral foi  $3,35 \times 10^3$  EBV/mL de saliva, com o mínimo de 0,20 e máximo de  $1,82 \times 10^7$  cópias de EBV/mL de saliva (Tabela S2).

#### 4.1.2 Diversidade do EBV pelo oncogene *LMP1*

Para entender o padrão de diversidade do EBV circulante no Brasil, um subgrupo das amostras avaliadas no tópico anterior foi selecionado para



sequenciamento do gene *LMP1*. Foram obtidas 83 novas sequências nucleotídicas provenientes da saliva de carreadores assintomáticos EBV+ (54,2%) e de linfonodos e PBMC de indivíduos com patologias hematológicas benignas e malignas EBV+ (45,8%) (Tabela 4.2). As sequências completas do oncogene *LMP1* foram alinhadas juntamente com sequências completas do *LMP1* provenientes de linhagens celulares EBV+ e de isolados com especificidade geográfica, reconhecidos por caracterizar diferentes variantes, disponíveis no Genbank (Tabela 3.2).

#### 4.1.2.1 Características da população com sequência inteira do oncogene *LMP1*

Das 83 amostras selecionadas para caracterização das sequências nucleotídicas do oncogene *LMP1*, 45 (54,2%) foram selecionadas de carreadores assintomáticos e 38 (45,8%) de indivíduos com patologias hematológicas EBV+ [1 (1,2%) de PTL, 4 (4,8%) de HLR, 7 (8,4%) de LB e 26 (31,3%) de LHc]. A faixa etária e o sexo foi significativamente diferente entre os dois grupos: com um predomínio de <18 anos e sexo masculino no grupo patológico e >18 anos e sexo feminino no grupo assintomático ( $p < 0,05$ ). A tipagem do EBV das amostras que foram obtidas sequências inteiras da oncoproteína consistiu em 84,3% tipo 1 e 15,7% tipo 2 do EBV, sem diferença significativa entre os grupos (Tabela 4.1).

Tabela 4.1 - Características da população com sequência inteira da *LMP1*.

Características	Total	N (%)	
		DA	PE
<b>N° de sequências inteiras do <i>LMP1</i></b>	83 (100)	45 (54,2)	38 (45,8)
<b>Sexo</b>			
Feminino	39 (47,0)	28 (63,6)	10 (25,6)
Masculino	44 (53,0)	16 (36,4)	29 (74,4)
<b>Idade</b>			
<= 18 anos	19 (22,9)	6 (31,6)	13 (68,4)
>= 19 anos	64 (77,1)	39 (60,9)	25 (39,1)
<b>Casos Benignos</b>			
HLR	—	—	4 (4,8)
<b>Casos Malignos</b>			
LHc	—	—	26 (31,3)
LB	—	—	7 (8,4)
PTLD	—	—	1 (1,2)
<b>Doadores assintomáticos</b>	45 (54,2)	—	—

### Tipagem do EBV pelo gene *EBNA3C*

Tipo 1	70 (84,3)	35 (77,8)	35 (92,1)
Tipo 2	13 (15,7)	10 (22,2)	3 (7,9)

DA: doador assintomático; N: número; PE: patologia EBV+; LMP1: proteína latente de membrana 1; HLR: hiperplasia linfoide reativa; LHc: linfoma de Hodgkin clássico; LB: linfoma de Burkitt; PTLT: doença linfoproliferativa pós transplante (do inglês: *post transplant lymphoproliferative disease*); EBV: vírus Epstein-Barr; EBNA3C: antígeno nuclear do Epstein-Barr 3C (do inglês: *Epstein-Barr nuclear antigen 3C*).

#### 4.1.2.2 Classificação de polimorfismos da LMP1

Inicialmente, as variantes da LMP1 foram caracterizadas utilizando como base variantes descritas por Edwards *et al.* (1999), considerado o sistema de classificação mais utilizado até o momento a considerar a oncoproteína LMP1 (Figura 4.2 e Tabela 4.2) (EDWARDS *et al.*, 1999; PALSER *et al.*, 2015; CORREIA *et al.*; GANTUZ *et al.*, 2017). Foi avaliada a presença de polimorfismos presentes nos três diferentes domínios da oncoproteína LMP1, N-terminal (N-ter), transmembranar (TM) e C-terminal (C-ter), e a sua relação de associação com casos benignos e carreadores assintomáticos *versus* casos de doenças malignas.

O polimorfismo mais estudado da região N-ter é a presença do sítio de restrição da enzima *XhoI* (c↓tcgag) (LIN *et al.*, 2003; NGUYEN-VAN *et al.*, 2008). O sítio *XhoI* foi detectado em 98% das amostras de doadores assintomáticos (DA) e casos benignos (CB), e em 100% dos casos malignos (CM), não havendo diferença significativa entre estes grupos ( $p > 0,05$ ; Tabela 4.3).

Dentre os polimorfismos da região transmembrana, dois grupos de polimorfismos principais, I124V/I152L e F144I/D150A/L151I, foram descritos por caracterizar um aumento da atividade do NF- $\kappa$ B *in vitro* (ZUERCHER *et al.*, 2012). O primeiro, I124V/I152L, foi identificado em 48,5% dos CM, e em 28,6% dos DA+CB ( $p = 0,1$ ). No caso do grupo de polimorfismos F144I/D150A/L151I, a mutação L151I foi mais prevalente que a variante selvagem (wt/wt/wt), tanto no grupo CM quanto nos DA+CB, totalizando uma fração de 70% e 55,1%, respectivamente. Enquanto a tríade F144I/D150A/L151I foi observada apenas em 5,3% dos CM *versus* 6,7% em DA+CB. Nenhum dos polimorfismos citados acima foram associados a casos de malignidade ( $p > 0,05$ ; Tabela 4.3).

Devido à sua capacidade de alterar a atividade de sinalização, polimorfismos no domínio C-ter são conhecidos por definir variantes da LMP1, (EDWARDS *et al.*, 1999; 2004). Dentre esses polimorfismos está o número de repetições de 33 pb em tandem (rep33), e a presença da inserção de 15 pb (ins15) que codifica um motivo de

sinalização para Janus Kinase 3 (JAK3), na qual o locus pode variar em umas das repetições de 33 pb, e a deleção de 30 pb (del30). Em relação ao número de repetições em tandem de 33pb, a mediana do grupo CM foi de 4, com mínimo de 2 e máxima de 6, onde a mais frequente foram 4 repetições em 50%, seguido de 5 repetições (35,3%), 6 repetições (8,8%), e 2 e 3 repetições com 2,6% cada. No grupo DA+CB, a mediana foi de 5, com mínimo de 1 e máxima de 8. Nesse grupo as 5 repetições de 33 pb em tandem foram as mais frequentes em 42,9% dos indivíduos, seguida por 4 repetições em 32,7%, 6 repetições em 12,2%, 3 repetições em 6,1%, e 1, 7 e 8 repetições em 2% ( $p>0,05$ ; Tabela 4.3). A ins15 que codifica um motivo de sinalização para JAK3, foi localizada principalmente na 4ª repetição de 33pb tanto em CM quanto DA+CB ( $p>0,05$ ), entretanto, a ins15 foi identificada majoritariamente nas amostras de CM (70,6%) e em menor proporção (57,1%) no grupo DA+CB,  $p=0,015$  (Tabela 4.3). A ocorrência da del30 na nossa amostragem foi de 53,1% nos DA+CB e 55,9% dos CM. De forma surpreendente, uma deleção de 69 pb foi encontrada em 6,7% das amostras de DA não relacionados ( $p>0,05$ ; Tabela 4.3).

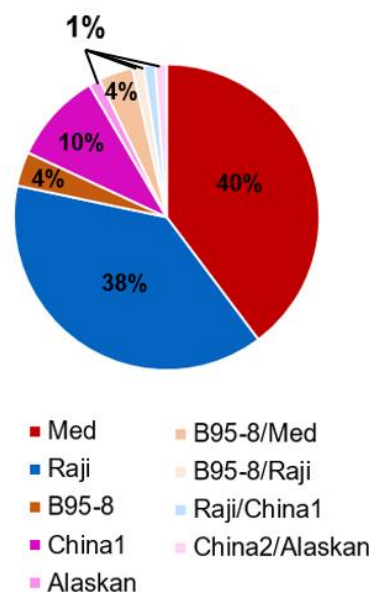


Figura 4.2 – Frequência das variantes do LMP1 classificadas pelo método de Edwards *et al.* (1999) nas amostras estudadas.

Tabela 4.2 - Polimorfismos e classificação das variantes da oncoproteína LMP1.

N	Isolado	Caso	Amostra	N-ter	Transmembrana			C-ter				Variante pela filogenia
				Xhol ctcgag	posição 169055 (C>T)	I124V/I152L	F144I/D150A/L1 51I	rep 33	ins15 (rep33)	del30	variante C-ter por Edwards <i>et al</i> , 1999	
1	2	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	6	sim (4°)	del30	Raji	Raji/Argentine
2	6	DA	Saliva	Sim	wt	wt/wt	wt/wt/L151I	4	não	wt	Raji	Raji/Argentine
3	8	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
4	38	DA	Saliva	Sim	wt	wt/wt	wt/wt/L151I	3	não	wt	Med	Med/AG876/Daudi
5	91	DA	Saliva	Sim	C>T	wt/wt	F144I/D150A/L1 51I	5	no	del30	China2/Alaskan	Asiática
6	132	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	3	no	wt	Med	Med/AG876/Daudi
7	136	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	4	sim (3°)	wt	B95-8/Med	B95-8
8	143	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	4	não	wt	Med	Med/AG876/Daudi
9	151	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
10	174	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	5	não	<b>del69</b>	Med	Med/AG876/Daudi
11	195	DA	Saliva	Sim	wt	wt/wt	wt/wt/L151I	4	sim (3°)	wt	Raji	Raji/Argentine
12	206	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	7	não	del30	China1	Med/AG876/Daudi
13	224	DA	Saliva	Sim	C>T	wt/wt	F144I/D150A/L1 51I	5	não	del30	China1	Asiática
14	227	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine

15	247	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
16	271	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	4	sim (3°)	wt	Raji	Raji/Argentine
17	280	DA	Saliva	Sim	wt	wt/wt	wt/wt/L151I	4	não	wt	Med	Med/AG876/Daudi
18	284	DA	Saliva	Sim	wt	wt/wt	wt/wt/L151I	4	não	wt	Med	Não classificada
19	302	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
20	317	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	8	não	del30	Med	Med/AG876/Daudi
21	326	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	4	não	del30	China1	Med/AG876/Daudi
22	343	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
23	351	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	6	não	wt	Med	Med/AG876/Daudi
24	365	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
25	383	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	5	não	wt	Med	Med/AG876/Daudi
26	386	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	6	não	wt	Med	Med/AG876/Daudi
27	389	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	1	não	del69	Med	Med/AG876/Daudi
28	407	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
29	447	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	5	não	del69	Med	Med/AG876/Daudi
30	456	DA	Saliva	CTCTAG	wt	wt/wt	F144I/wt/L151I	5	não	wt	Alaskan	Asiática
31	467	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	6	sim (3°)	del30	Raji	Raji/Argentine
32	478	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	4	não	del30	China1	Med/AG876/Daudi
33	479	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	5	não	wt	Med	Med/AG876/Daudi
34	489	DA	Saliva	Sim	wt	wt/wt	wt/wt/L151I	4	não	wt	Med	Med/AG876/Daudi

35	503	DA	Saliva	Sim	C>T	wt/wt	F144I/D150A/L151I	4	não	del30	China1/China2	Asiática
36	505	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	3	não	wt	Med	Med/AG876/Daudi
37	510	DA	Saliva	Sim	wt	wt/wt	wt/wt/L151I	4	não	wt	Med	Não classificada
38	519	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	6	sim (4°)	del30	Raji	Raji/Argentine
39	527	DA	Saliva	Sim	wt	wt/wt	wt/wt/L151I	4	não	wt	Med	Med/AG876/Daudi
40	560	DA	Saliva	Sim	wt	wt/wt	wt/wt/L151I	6	não	wt	B95-8/Med	Med/AG876/Daudi
41	589	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (3°)	del30	Raji	Raji/Argentine
42	626	DA	Saliva	Sim	wt	wt/wt	wt/wt/L151I	5	não	wt	Med	Med/AG876/Daudi
43	639	DA	Saliva	Sim	wt	wt/wt	wt/wt/L151I	5	não	wt	Med	Med/AG876/Daudi
44	657	DA	Saliva	Sim	wt	wt/wt	wt/wt/wt	5	sim (2°)	wt	B95-8/Med	B95-8
45	665	DA	Saliva	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (2°)	wt	Raji	Raji/Argentine
46	3633	HLR	LN	Sim	wt	wt/wt	wt/wt/wt	5	não	del30	China1	Med/AG876/Daudi
47	7607	HLR	LN	Sim	wt	wt/wt	wt/wt/L151I	4	não	wt	Med	Med/AG876/Daudi
48	9290	HLR	LN	Sim	wt	wt/wt	wt/wt/L151I	4	não	wt	Med	Med/AG876/Daudi
49	9893	HLR	LN	Sim	wt	wt/wt	wt/wt/wt	4	não	wt	Med	Med/AG876/Daudi
50	49	LB	LN	Sim	wt	wt/wt	wt/wt/wt	3	não	wt	Med	Med/AG876/Daudi
51	2438	LB	LN	Sim	wt	wt/wt	wt/wt/L151I	5	não	wt	Med	Não classificada
52	4756	LB	LN	Sim	wt	wt/wt	wt/wt/wt	4	não	wt	Med	Med/AG876/Daudi
53	7788	LB	LN	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	V-Raji	Raji/Argentine

54	10179	LB	LN	Sim	wt	I124V/I152L	wt/wt/L151I	4	não	wt	V-Raji	Raji/Argentine
55	12583	LB	LN	Sim	wt	wt/wt	wt/wt/L151I	6	sim (5°)	del30	China1	Não classificada
56	CEMO3	LB	LN	Sim	wt	I124V/I152L	wt/wt/L151I	4	sim (3°)	del30	Raji	Raji/Argentine
57	3566	LHc	LN	Sim	wt	wt/wt	wt/wt/wt	4	não	del30	Med	Med/AG876/Daudi
58	4203	LHc	LN	Sim	wt	wt/wt	wt/wt/wt	5	sim (4°)	wt	B95-8	B95-8
59	4468	LHc	LN	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	V-Raji	Raji/Argentine
60	4482	LHc	LN	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
61	4803	LHc	LN	Sim	wt	wt/wt	wt/wt/L151I	4	não	wt	Med	Med/AG876/Daudi
62	5010	LHc	LN	Sim	wt	wt/wt	wt/wt/wt	4	sim (3°)	wt	B95-8	B95-8
63	5134	LHc	LN	Sim	wt	wt/wt	wt/wt/L151I	6	sim (5°)	wt	Med	Não classificada
64	6408	LHc	LN	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
65	7087	LHc	LN	Sim	wt	wt/wt	wt/wt/wt	4	não	wt	Med	Med/AG876/Daudi
66	7218	LHc	LN	Sim	wt	I124V/I152L	wt/wt/L151I	4	sim (3°)	wt	Raji	Raji/Argentine
67	7581	LHc	LN	Sim	wt	wt/I152L	wt/wt/L151I	4	sim (3°)	wt	B95-8/V-Raji	Não classificada
68	7854	LHc	LN	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
69	7887	LHc	LN	Sim	wt	wt/wt	wt/wt/wt	4	sim (3°)	wt	B95-8	B95-8
70	8202	LHc	LN	NA	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
71	8228	LHc	LN	Sim	wt	I124V/I152L	wt/wt/L151I	4	não	wt	Raji	Raji/Argentine
72	8458	LHc	LN	Sim	C>T	wt/wt	F144I/D150A/L151I	5	não	del30	Raji/China1	Asiática

73	8909	LHc	LN	Sim	C>T	wt/wt	F144I/D150A/L151I	2	não	del30	China1	Asiática
74	8943	LHc	LN	Sim	wt	I124V/I152L	wt/wt/L151I	6	sim (4°)	del30	Raji	Raji/Argentine
75	8962	LHc	LN	Sim	wt	I124V/I152L	wt/wt/L151I	4	sim (3°)	wt	Raji	Raji/Argentine
76	9134	LHc	LN	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
77	9146	LHc	LN	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
78	9417	LHc	LN	Sim	wt	I124V/I152L	wt/wt/L151I	4	não	wt	Raji	Raji/Argentine
79	9577	LHc	LN	Sim	wt	wt/wt	wt/wt/L151I	4	sim (3°)	wt	Med	Med/AG876/Daudi
80	9646	LHc	LN	Sim	wt	wt/wt	wt/wt/L151I	4	sim (3°)	wt	Med	Med/AG876/Daudi
81	9863	LHc	LN	Sim	wt	wt/wt	wt/wt/wt	4	sim (3°)	wt	Med	Med/AG876/Daudi
82	10564	LHc	LN	Sim	wt	I124V/I152L	wt/wt/L151I	5	sim (4°)	del30	Raji	Raji/Argentine
83	26366	PTLD	PB	Sim	wt	wt/wt	wt/wt/L151I	4	não	wt	Med	Não classificada

ND: Não disponível; PBMC: células mononucleares do sangue periférico (do inglês: *peripheral blood mononuclear cells*); DA: doador assintomático; LN: linfonodo; HLR: hiperplasia linfóide reativa; LB: linfoma de Burkitt; LHc: linfoma de Hodgkin clássico; LMP1: proteína latente de membrana 1 (do inglês: *latente membrane protein 1*); N-ter: domínio amino terminal; *Xho*I: endonuclease do tipo 2, de *Xanthomonas holcicola*; wt: tipo selvagem (do inglês: *wildtype*); C: citosina; T: timina; I: isoleucina; V: valina; L: leucina; F: fenilalanina; D: ácido aspártico; A: alanina; rep 33: repetição de 33 pb; pb: pares de base; ins15: inserção de 15 pb; del30: deleção de 30 pb; del69: deleção de 69 pb; N-ter: amino-terminal; TM: transmembrana; C-ter: carboxi-terminal.



Tabela 4.3 - Relação dos polimorfismos da oncoproteína LMP1 com os casos investigados

Polimorfismos da LMP1	N (%)		p - valor
	DA+CB	CM	
<b>Polimorfismo do N-terminal</b>			
Sítio de restrição da <i>Xho</i> I (c↓tcgag)	48 (98)	33 (100)	1
Ausência do sítio de restrição da <i>Xho</i> I	1 (2)	0 (0)	
<b>Polimorfismos da Transmembrana</b>			
Posição 169055 wt	3 (6,1)	2 (5,9)	1
Posição 169055 C>T	46 (93,9)	32 (94,1)	
wt/wt	35 (71,4)	17 (51,5)	0,1
I124V/I152L	14 (28,6)	16 (48,5)	
wt/wt/wt	18 (36,7)	8 (23,5)	0,45
wt/wt/L151I	27 (55,1)	24 (70,6)	
F144I/wt/L151I	1 (2,0)	0 (0,0)	
F144I/D150A/L151I	3 (6,1)	2 (5,9)	
<b>Polimorfismos do C-terminal</b>			
N° de repetições 33 pb			
x1	1 (2,0)	0 (0,0)	0,51
x2	0 (0,0)	1 (2,9)	
x3	3 (6,1)	1 (2,9)	
x4	16 (32,7)	17 (50,0)	
x5	21 (42,9)	12 (35,3)	
x6	6 (12,2)	3 (8,8)	
x7	1 (2,0)	0 (0,0)	
x8	1 (2,0)	0 (0,0)	
wt	28 (57,1)	10 (29,4)	0,015
Ins15 pb	21 (42,9)	24 (70,6)	
wt	20 (40,8)	15 (44,1)	0,33
del30	26 (53,1)	19 (55,9)	
del69	3 (6,1)	0 (0,0)	

LMP1: proteína latente de membrana 1 (do inglês: *latent membrane protein 1*); DA: doador assintomático; CB: casos benignos; CM: casos malignos; *Xho*I: endonuclease do tipo 2, de *Xanthomonas holcicola*; wt: tipo selvagem (do inglês: *wildtype*); C: citosina; T: timina; I: isoleucina; V: valina; L: leucina; F: fenilalanina; D: ácido aspártico; A: alanina; rep: repetição; pb: pares de base; ins15: inserção de 15 pb; del30: deleção de 30 pb; del69: deleção de 69 pb;

#### 4.1.2.3 Diversidade e evolução da oncoproteína LMP1 na América do Sul

Adicionalmente, reconstruções filogenéticas utilizando o método de máxima verossimilhança foram realizadas com a finalidade de classificar as linhagens presentes na nossa amostragem pela região codificante inteira. A classificação filogenética da oncoproteína LMP1 das 83 amostras investigadas, indicou a presença de diferentes linhagens, distribuídas em cinco clados com alto suporte ( $aLRT > 0.81$ ; (Tabela S5 e Figura 4.3). Aproximadamente 79% das amostras de EBV brasileiras se agruparam em dois clados principais: 40% das amostras se agruparam em um clado que reuniu as linhagens Med/AG876/Daudi ( $aLRT=0.95$ ), enquanto 39% das amostras ramificaram juntamente com as variantes Raji/Argentine ( $aLRT=0.92$ ). As amostras restantes formaram linhagens minoritárias, reunindo 6% juntamente com a variante B95-8 ( $aLRT=0.96$ ), 6% com as variantes China1/CAO ( $aLRT=0.96$ ), 1% com a variante China2 ( $aLRT=0.81$ ) e 8% não puderam ser classificadas através da filogenia (Figura 4.4 e Tabela S5). Em relação à diversidade das cepas encontradas, não houve diferença significativa entre o grupo de doadores assintomáticos e casos patológicos benignos (DA+CB) e os casos patológicos malignos (CM),  $p=0,07$ . No entanto, quando os dois principais clados (reunindo 79% da amostragem estudada) são analisados, de maneira surpreendente, é possível identificar que o clado monofilético Med/Ag876/Daudi reunia majoritariamente as amostras do grupo DA+CB (61%), enquanto a maior parte das amostras derivadas de CM se agrupam no clado monofilético Raji/Argentine (66,7%),  $p = 0,041$  (Tabela 4.4).

Um total de sete amostras não se agruparam em clados bem definidos, portanto, não foram classificadas com o presente método utilizado. Dentre essas amostras, cinco são CM e duas são DA, totalizando 4,1% de DA+CB e 14,7% de CM ( $p=0,1$ ), sugerindo a existência de linhagens não contempladas dentre as variantes referência representativas ou possivelmente recombinantes intervariantes.

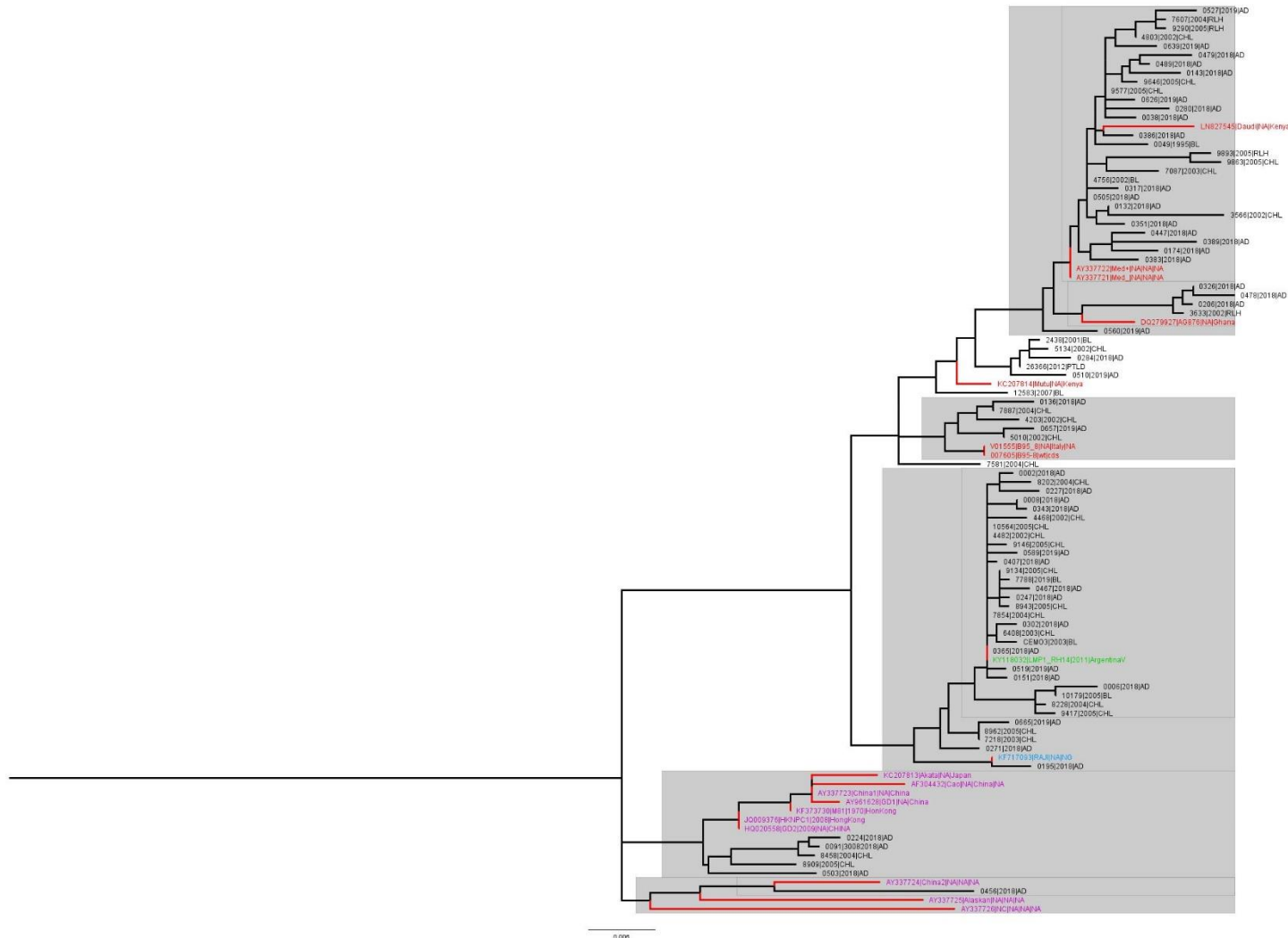


Figura 4.3 - **Árvore filogenética de máxima verossimilhança das seqüências completas da oncoproteína LMP1 de casos e carreadores assintomáticos (em preto).** Sequências referências incluídas na análise pertencem às variantes: B95-8, Med +/- e AG876 em vermelho; China1, China2, GD1, GD2, Akata, Cao, NC, Alaskan, HKNPC e M81 em rosa; Raji em azul; e Argentine em verde. Foram formados 5 clados principais (sombas cinzas). Os valores de suporte aLRT são indicados nas ramificações principais e os comprimentos dos ramos estão em escala com a barra inferior indicando substituições nucleotídicas por sítio.

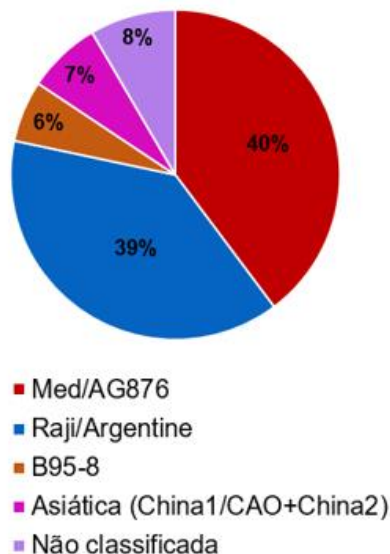


Figura 4.4 – Frequência das variantes do LMP1 classificadas por filogenia de máxima verossimilhança nas amostras estudadas.

Tabela 4.4 - Relação das variantes da oncoproteína LMP1 com a população estudada.

Classificação pela filogenia	N (%)			<i>p</i> - valor
	total	DA+CB	CM	
<b>Clados identificados</b>				
Med/AG876/Daudi	33 (40)	25 (51,0)	8 (23,5)	0,07
Raji/Argentine	32 (39)	16 (32,7)	16 (47,1)	
B95-8	5 (6)	2 (4,1)	3 (8,8)	
Asiática (China1/CAO+China2)	6 (7)	4 (8,2)	2 (5,9)	
Não classificada	7 (8)	2 (4,1)	5 (14,7)	
<b>Clados majoritários no Brasil</b>				
Med/AG876/Daudi	33 (40)	25 (61,0)	8 (33,3)	0,041
Raji/Argentine	32 (39)	16 (39,0)	16 (66,6)	

N: número de amostras; DA: doador assintomático; CB: casos benignos; CM: casos malignos.

#### 4.1.2.3.1 Relação dos polimorfismos genéticos com as variantes classificadas pela análise filogenética

Quando avaliada a relação dos polimorfismos de importância oncológica já descritos na literatura com as linhagens identificadas pela análise filogenética, foi observado que o polimorfismo principal do domínio N-ter (a perda do sítio de restrição da *Xho*I) e o polimorfismo do domínio TM (a posição 169055 C>T), estavam presentes apenas nas variantes Asiáticas (China1/CAO e China2), o primeiro em 16,7% e o segundo em 83,3% das amostras, em contrapartida da ausência dos mesmos nas

linhagens Med/AG876/Daudi, Raji/Argentine e B95-8 ( $p=0,009$  para o primeiro, e  $p<0,0001$  para o segundo; Tabela S6).

Em relação aos outros dois polimorfismos do domínio TM: I124V/I152L e F144I/D150A/L151I, o primeiro estava presente quase que exclusivamente nas variantes Raji/Argentine (93,8%) e ausente nas demais variantes ( $p<0,0001$ ). O segundo polimorfismo, F144I/D150A/L151I, estava majoritariamente presente nas variantes Asiáticas (83,3%) e ausente nas demais variantes ( $p<0,0001$ ). Adicionalmente, as variantes Raji/Argentine possuem em o polimorfismo L151I em 100% da amostragem ( $p<0,0001$ ) (Figura 4.5 e Tabela S6).

Na avaliação dos polimorfismos do C-ter, não houve diferença estatística no número de rep33 identificadas nas diferentes variantes, e a mediana variou de quatro rep33 para as variantes Med/AG876/Daudi (mín. 1 – máx. 8) e B95-8 (mín. 4 – máx. 5), e mediana de 5 rep 33 pb para as variantes Raji/Argentine (mín. 4 – máx. 6) e Asiática (mín. 2 – máx. 5);  $p = 0,061$  (Tabela S7).

A presença da ins15, que codifica um motivo de sinalização para JAK3, foi identificada na maioria das amostras classificadas nos clados Raji/Argentine (87,5%), B95-8 (100%) e Asiáticas (83,3%), enquanto no clado Med/AG876/Daudi, 87,9% das amostras não possuem essa inserção ( $p <0,0001$ ) (Figura 4.5 e Tabela S6).

A del30 foi identificada na maioria das amostras dos clados Raji/Argentine (71,9%) e B95-8 (100%), e ausente na maioria das amostras dos clados Med/AG876/Daudi (72,7%) e Asiáticas (83,3%),  $p <0,0001$ . Apesar da identificação da del69 em 9,1% das amostras do clado Med/AG876/Daudi, sua presença não foi estatisticamente significativa como um marcador desse clado ( $p=0,64$ ) (Figura 4.5 e Tabela S6).

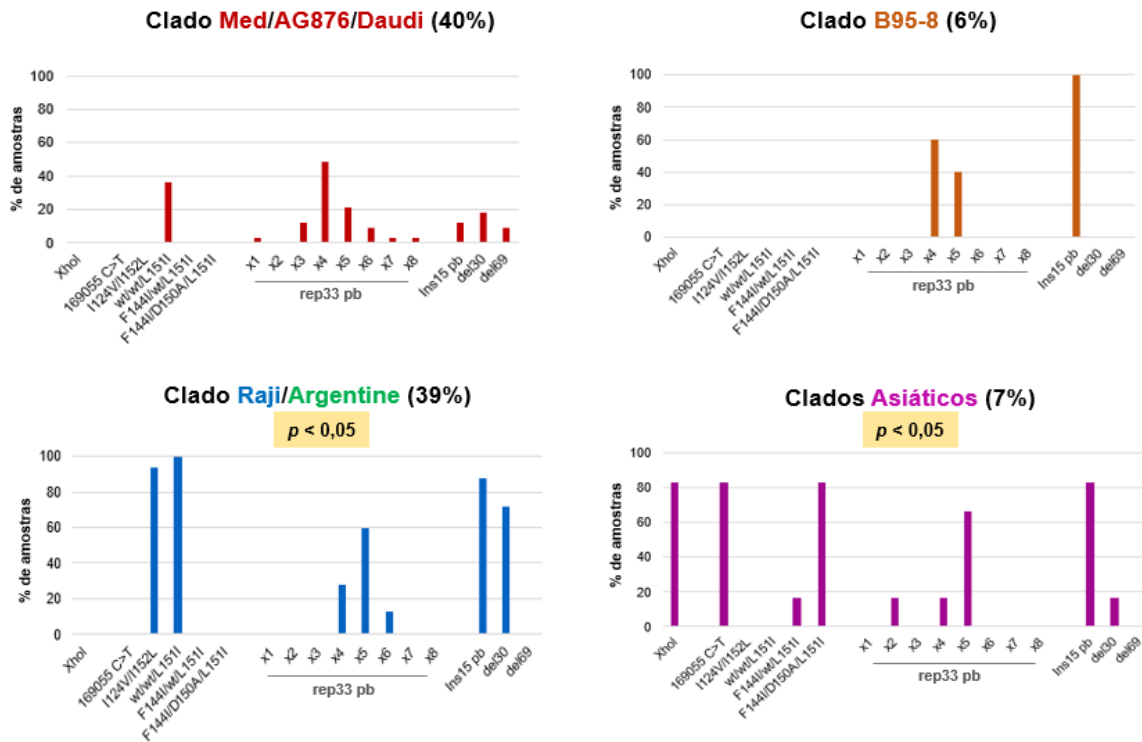


Figura 4.5 – Frequência dos polimorfismos genéticos da LMP1 de importância oncológica presente nos clados. No eixo Y está identificada a porcentagem da amostragem de cada clado e no eixo X os polimorfismos estudados. A porcentagem ao lado do nome dos clados é relacionada a quantidade total da amostragem em relação a todos os clados. XhoI: ausência do sítio de restrição da XhoI; pb: pares de base; rep33 pb: repetição de 33 pb; ins15 pb: inserção de 15 pb; del30: deleção de 30 pb; del69: deleção de 69 pares de base.

#### 4.1.2.3.2 Relação dos polimorfismos com as duas cepas principais circulantes no Brasil

Quando avaliado a presença dos polimorfismos de importância oncológica considerando somente os dois principais clados identificados no Brasil (Med/AG876/Daudi e Raji/Argentine, que abrigam 79% das sequências BR nossa amostragem, a significância estatística se manteve, demonstrando que estes polimorfismos são marcadores moleculares importantes do clado Raji/Argentine. As sequências que agruparam neste clado possuem majoritariamente os polimorfismos: I124V/I152L (93,8%), wt/wt/L151I (100%), ins15 (87,5%) e del30 (71,9%), em contrapartida com as variantes Med/AG876/Daudi, que em sua maioria apresenta as versões *wildtype* dos polimorfismos ( $p < 0,0001$  para todos, (Figura 4.6 e Tabela S7).

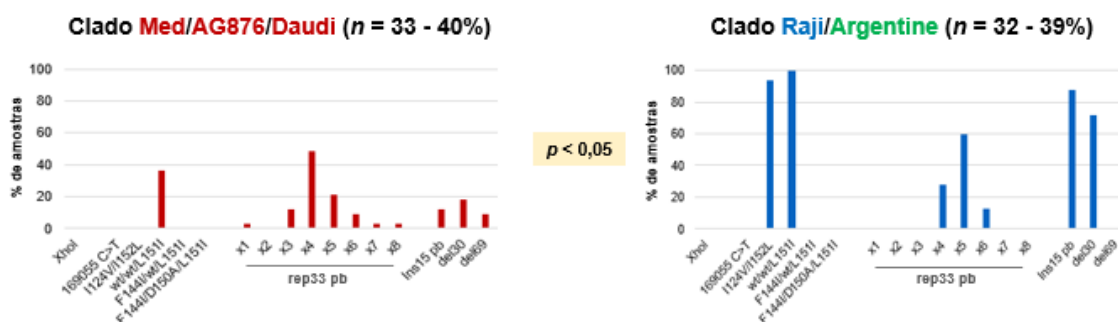


Figura 4.6 – Frequência dos polimorfismos de importância oncológica da LMP1 nos dois principais clados identificados. *XhoI*: ausência do sítio de restrição da *XhoI*; pb: pares de base; rep33 pb: repetição de 33 pb; Ins15 pb: inserção de 15 pb; del30: deleção de 30 pb; del69: deleção de 69 pares de base.

Em seguida, foi investigado se os grupos amostrais (DA+CB e CM) majoritários nesses dois principais clados, mantêm essa disparidade estatística relativo à presença dos polimorfismos investigados. Desta forma, foi observado, que o grupo de CM definidos como variante Raji/Argentine (66,7% das amostras no clado) se manteve possuindo majoritariamente, e de maneira significativa, os polimorfismos I124V/I152L (100%), wt/wt/L151I (100%), ins15 (81,3%) e del30 (68,8%), em contrapartida com o grupo de DA+CB definidos como variante Med/AG876/Daudi (61% das amostras no clado), possuindo em menor parte os polimorfismos I124V/I152L (ausente), wt/wt/L151I (33,3%), Ins15 pb (4,8%) e del30 (19%),  $p < 0,0001$  para todos (Tabela 4.5 e Figura 4.7).

Tabela 4.5 - Relação dos polimorfismos da oncoproteína LMP1 com o grupo DA+CB de variante Med/AG876/Daudi e o grupo CM de variante Raji/Argentine

Polimorfismos da LMP1	N (%)		<i>p</i> - valor
	Med DA+CB	Raji CM	
<b>Sequências inteiras</b>	21 (56,8)	16 (43,2)	
<b>Polimorfismo do N-terminal</b>			
Sítio de restrição da <i>XhoI</i>	33 (100)	32 (100)	<i>p</i> *
Ausência do sítio de restrição da <i>XhoI</i>	0 (0,0)	0 (0,0)	
<b>Polimorfismos da Transmembrana</b>			
Posição 169055 wt	33 (100)	32 (100)	<i>p</i> *
Posição 169055 C>T	0 (0,0)	0 (0,0)	
wt/wt	21 (100)	0 (0,0)	$p < 0,0001$
I124V/I152L	0 (0,0)	16 (100)	

wt/wt/wt	14 (66,7)	0 (0,0)	$p < 0,0001$
wt/wt/L151I	7 (33,3)	16 (100)	
F144I/wt/L151I	0 (0,0)	0 (0,0)	
F144I/D150A/L151I	0 (0,0)	0 (0,0)	

### Polimorfismos do C-terminal

Nº de repetição 33 pb

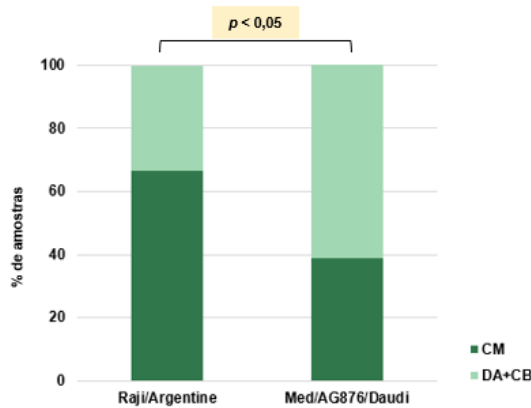
x1	1 (4,8)	0 (0,0)	$p = 0,31$
x2	3 (14,3)	0 (0,0)	
x3	6 (28,6)	6 (37,5)	
x4	6 (28,6)	9 (56,2)	
x5	3 (14,3)	1 (6,2)	
x6	1 (4,8)	0 (0,0)	
x7	1 (4,8)	0 (0,0)	
x8			

wt	20 (95,2)	3 (18,7)	$p < 0,0001$
Ins15 pb	1 (4,8)	13 (81,3)	

wt	14 (66,7)	5 (31,2)	$p = 0,007$
del30	4 (19,0)	11 (68,8)	
del69	3 (14,3)	0 (0,0)	

p\*: não há cálculo estatístico, visto que a variável é constante; LMP1: ; Med : Med/AG876/Daudi; Raji : Raji/Argentine; DA+CB ; CM

(A)



(B)

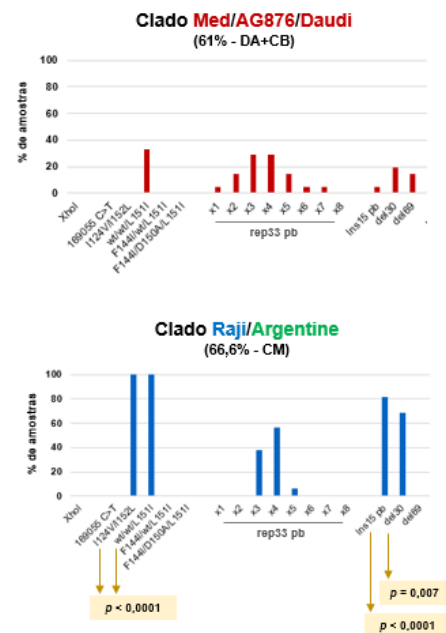


Figura 4.7 – **Relação dos polimorfismos nos dois principais clados com o desfecho clínico.** (A) Relação dos dois principais clados com o desfecho clínico. Eixo Y é a porcentagem da amostragem de cada clado (eixo X). (B) Relação dos polimorfismos genéticos na amostragem assintomática e benigna no clado



Med/AB876/Daudi e nos casos malignos presentes no clado Raji/Argentine. CM: casos malignos; DA: doadores assintomáticos; CB: casos benignos; *Xhol*: ausência do sítio de restrição da *Xhol*; pb: pares de base; rep33 pb: repetição de 33 pb; Ins15 pb: inserção de 15 pb; del30: deleção de 30 pb; del69: deleção de 69 pares de base.

#### 4.1.2.3.3 Divergência entre as duas classificações aplicadas para a LMP1

Atualmente, a classificação da LMP1 por Edwards *et al.* (1999) é a mais utilizada na literatura para definir as variantes de LMP1 (PALSER *et al.*, 2015, CORREIA *et al.*, 2017; GANTUZ *et al.*, 2017). A comparação desta classificação com a classificação filogenética presente neste trabalho revelou que para o clado Raji/Argentine, B95-8 e algumas amostras Asiática não houve diferença na classificação, entretanto, 5 amostras classificadas como Med pela classificação por Edwards *et al.*, não se agruparam a um clado específico, ficando de fora do clado Med/AG876/Daudi, desta forma, estas amostras através da análise do gene inteiro são sugeridas de serem recombinantes. Da mesma forma, uma amostra classificada como China1 pela classificação por Edwards *et al.*, foi sugerida de ser recombinante por não se agrupar a um clado específico pela nossa análise filogenética. De maneira inversa, duas amostras consideradas recombinantes entre as variantes B95-8/Med pela análise do domínio C-ter por Edwards *et al.*, se agruparam dentro do clado B95-8 pela análise filogenética presente neste trabalho (Tabela S8 e S9).

#### 4.1.2.4 Origem e diversidade do EBV circulante na América do Sul através do oncogene *LMP1*

Com o intuito de caracterizar a diversidade do EBV circulante no Brasil em seu contexto global, foi realizada uma análise filogenética de máxima verossimilhança das novas sequências do gene LMP1 geradas neste estudo com 252 sequências disponíveis no Genbank. As sequências se distribuíram em seis clados importantes com alto suporte (>70%) representando a diversidade do LMP1 do EBV que agruparam em alguns casos mais de uma variante utilizada como referência e que foram denominados da seguinte forma: Med, suporte aLRT = 0,91 - agrupando as referências Med+, Med-, Daudi e Mutu (JONES *et al.*, 1984; EDWARDS *et al.*, 1999; LIN *et al.*, 2013); B95-8, suporte aLRT = 0,97 – exibindo somente esta variante (BAER *et al.*, 1984); Raji/Argentine, suporte aLRT = 1 – reunindo as variantes Raji e Argentine (HATFULL *et al.*, 1988; GANTUZ *et al.*, 2017); Asiático 1, suporte aLRT = 0,81 – agrupando as referências HKNPC1, M81, Cao, Akata, GD2 e China1; Asiático 2,

suporte aLRT = 0,93 - agrupando as referências NC, Alaskan e China2 (EDWARDS *et al*, 1999); e América/Europa, suporte aLRT = 0,77 – nenhuma referência estava presente neste clado (Figura 4.8).

As cepas brasileiras juntamente com outras da América do Sul se agruparam majoritariamente nos clados Raji/Argentine (Brasil = 35, Argentina = 33) e Med (Brasil = 33, Argentina = 11), caracterizando estas duas linhagens como as principais em circulação na América do Sul, reunindo 72,3% das sequências totais oriundas do continente (Figura 4.9). Outro clado importante na América do Sul, porém em menor prevalência, foi o B95-8, que reuniu cinco amostras do Brasil (6%) e seis amostras da Argentina (8,8%). Uma amostra do Brasil se agrupou no clado América/Europa, juntamente com outras duas amostras da Argentina. Embora seis sequências da Argentina tenham se agrupado no clado Asiático 2, nenhuma amostra brasileira estava presente. Outras sete amostras brasileiras (8,4%) e cinco argentinas (7,3%) não se agruparam em nenhum dos clados principais, podendo representar linhagens minoritárias ou recombinantes intervariantes. É digno de nota que uma amostra brasileira e uma argentina se agruparam formando uma díade com alto suporte (aLRT = 0.99), representando uma linhagem exclusiva da América do Sul, embora aparentemente com baixa prevalência.

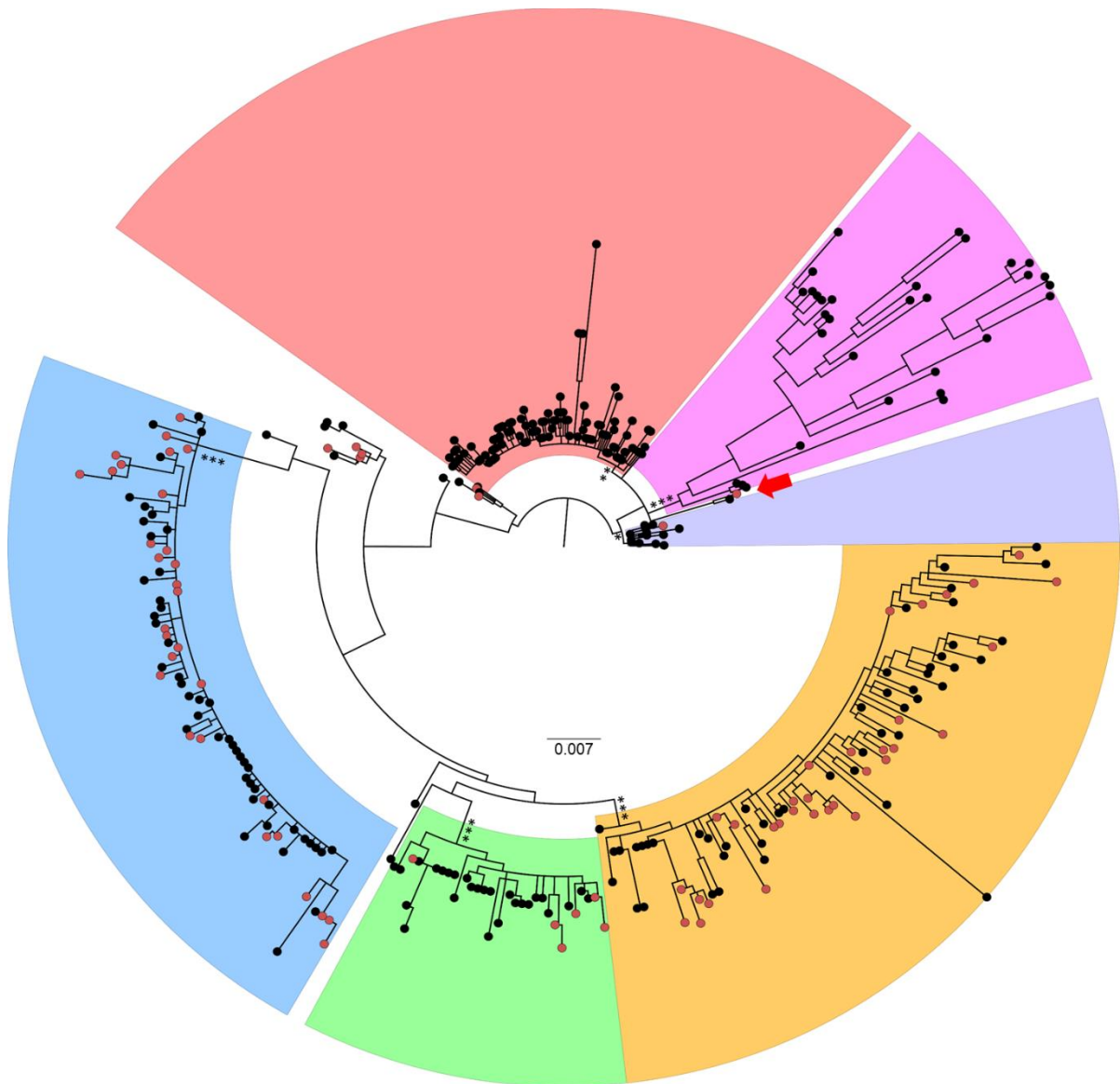


Figura 4.8 - **Árvore filogenética de máxima verossimilhança do gene LMP1 das amostras brasileiras deste estudo (n = 83) e as disponíveis em bancos de dados (n = 252).** Foram observados 6 clados importantes assinalados pelos diferentes sombreados: laranja – Med (Agrupando as referências Med+, Med-, Daudi e Mutu; verde – B95-8; azul – Raji/Argentine; vermelho – Asiático 1 (agrupando as referências HKNPC1, M81, Cao, Akata, NPC e China1); rosa – Asiático 2 (agrupando as referências NC, Alaskan e China2); e lilás (clado agrupando seqüências de diferentes países sem uma referência). A seta vermelha marca a posição da díade reunindo as seqüências brasileiras e argentinas. Asteriscos representam os valores de suporte aLRT das ramificações principais (\* >70%, \*\* >80% e \*\*\* >90%) e os comprimentos dos ramos estão em escala com a barra inferior indicando substituições nucleotídicas por sítio.

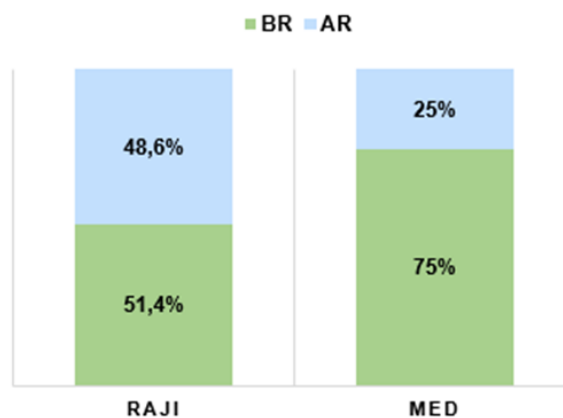


Figura 4.9 – Frequência do país de origem das sequências do *LMP1* pertencentes aos dois principais clados da América do Sul. BR: Brasil; AR: Argentina.

Com a finalidade de entender a origem geográfica do clado da variante Raji/Argentine que predomina no Brasil e na Argentina foi realizada a análise filogeográfica bayesiana para reconstruir a dinâmica das migrações virais do clado Raji (Figura 4.10). A reconstrução espaço-temporal bayesiana estimou a taxa do relógio molecular do gene LPM1 em  $4,8 \times 10^{-5}$  (Intervalo de credibilidade:  $2,2 \times 10^{-5}$  –  $7,6 \times 10^{-5}$ ) substituições/sítio/ano. De acordo com a filogenia reconstruída, as cepas da variante Raji circulante na América do Sul (Brasil, Argentina e Peru) originaram-se através de uma introdução viral proveniente do continente Africano para o Brasil (probabilidade posterior de 0,46) entre os séculos XVII e XIX. Posteriormente, o Brasil (e em menor intensidade a Argentina) atuou como uma fonte de exportação desta linhagem para outros países do continente americano (Argentina e Peru) e para a Europa (Suíça) (Figura 4.11). Embora em nossa reconstrução filogeográfica indique o retorno desta linhagem do Brasil para a África, esses eventos de migração representam artefatos da baixa amostragem da diversidade genética viral do continente africano.

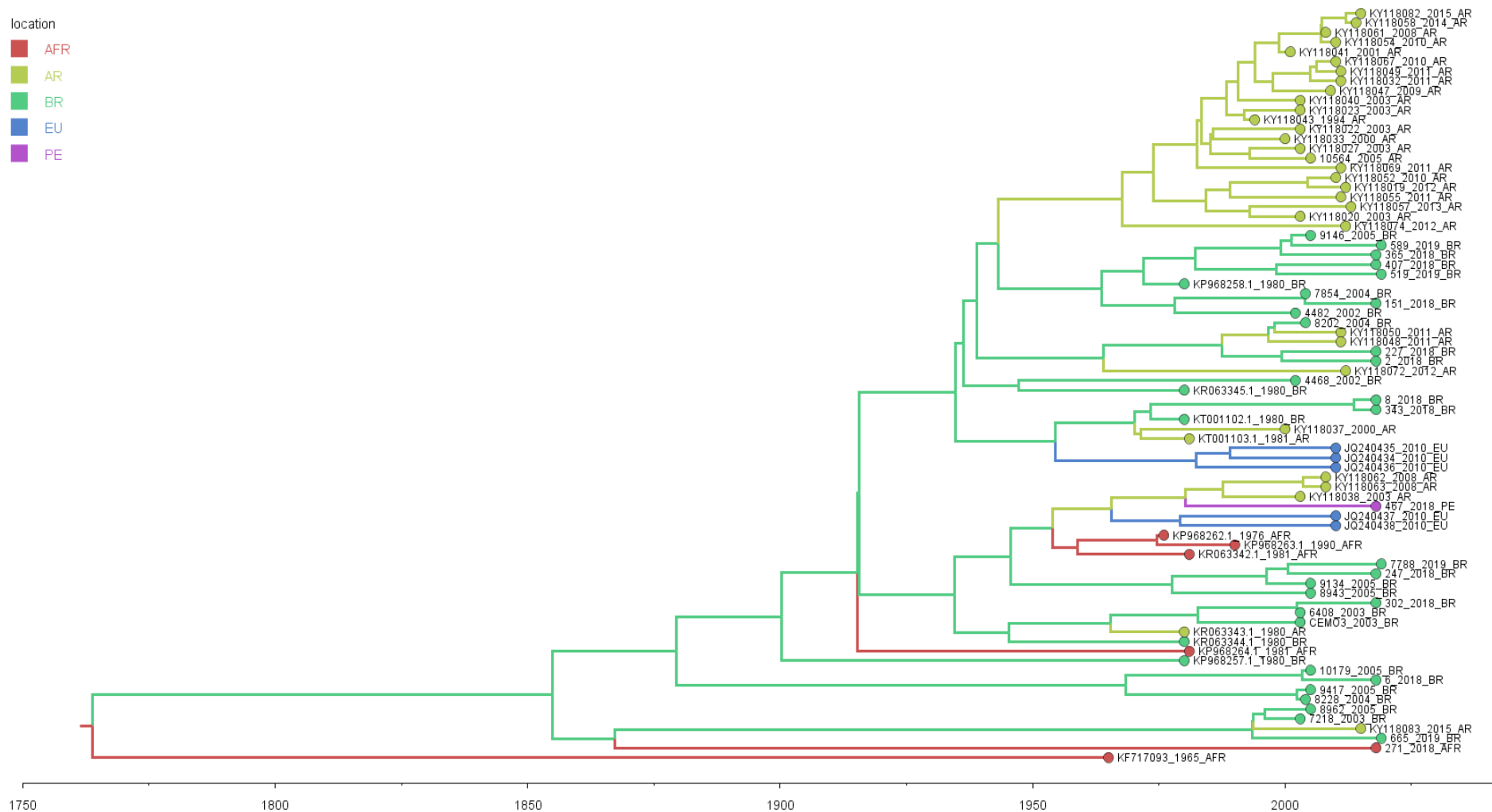


Figura 4.10 – Filogeografia da variante Raji/Argentine circulante na América do Sul. As cores dos ramos representam a localização mais provável de seus nós descendentes, conforme indicado na legenda (canto superior esquerdo). O suporte de ramo é indicado apenas em nós chave (probabilidade de estado posterior / posterior). Todos os comprimentos de ramos horizontais são desenhados em uma escala de anos. AFR: África; EU: Europa; AR: Argentina; BR: Brasil; PE: Peru.



Figura 4.11 – **Esquema demonstrando as migrações virais e origem da linhagem RAJI/ARGENTINE.** Setas vermelhas indicam eventos de migrações virais mais antigos, laranja indica eventos intermediários e amarelos eventos mais recentes.

#### 4.1.3 Associação da variante Tipo1/V3 com as variantes da oncoproteína LMP1

A variante Tipo1/V3 do EBV, pela caracterização do gene *EBNA3C* e o promotor do gene lítico Z (Zp), é sugerida de possuir funcionalmente um alto poder replicativo, além de ter associações com patologias malignas em populações específicas, como no sudeste asiático e em casos endêmicos de linfomas associados à malária na África central (TONG *et al*, 2003; MARTINI *et al*, 2007; BRISTOL *et al*, 2018). No presente trabalho, investigamos essa relação haplotípica com as variantes da oncoproteína LMP1 estudadas. O haplótipo Tipo1/V3 foi detectado em 13,7% de 51 amostras genotipadas para o Zp. De forma inesperada, 85,7% das amostras Tipo1/V3 pertencem à variante Raji/Argentine, sendo 26,1% das Raji/Argentine avaliadas. Por outro lado, o restante das amostras pertenceu à variante Asiática ( $p=0,08$ ) (Tabela 4.6). Dentre as amostras Raji/Argentine + Tipo1/V3, 66,6% são relacionadas a casos malignos. Quando comparados os dois principais clados, Raji *versus* Med, o segundo contendo em sua maioria carreadores assintomáticos e casos benignos, a presença do haplótipo Tipo1/V3 foi significativa na variante Raji, a qual já demonstrava pela análise de polimorfismos da LMP1 possuir um potencial oncogênico aumentado ( $p = 0,027$  - Tabela 4.6 e Figura 4.12).

Tabela 4.6 – Relação da classificação do Zp com as variantes da LMP1 estudadas

Tipo1/V3	N (%)		$p$ - valor
	Ausência	Presença	
Número de sequências	44 (86,3)	7 (13,7)	
<b>Clados pela filogenia</b>			
Med/AG876/Daudi	18 (100)	0 (0,0)	0,08
Raji/Argentina	17 (73,9)	6 (26,1)	
B95-8	4 (100)	0 (0,0)	
Asiática	5 (83,3)	1 (16,1)	
<b>Clados Med x Raji</b>			
Med/AG876/Daudi	18 (100)	0 (0,0)	0,02
Raji/Argentina	17 (73,9)	6 (26,1)	

N: número de amostras; LMP1: proteína latente de membrana 1 (do inglês: *latent membrane protein 1*); pb: pares de base;

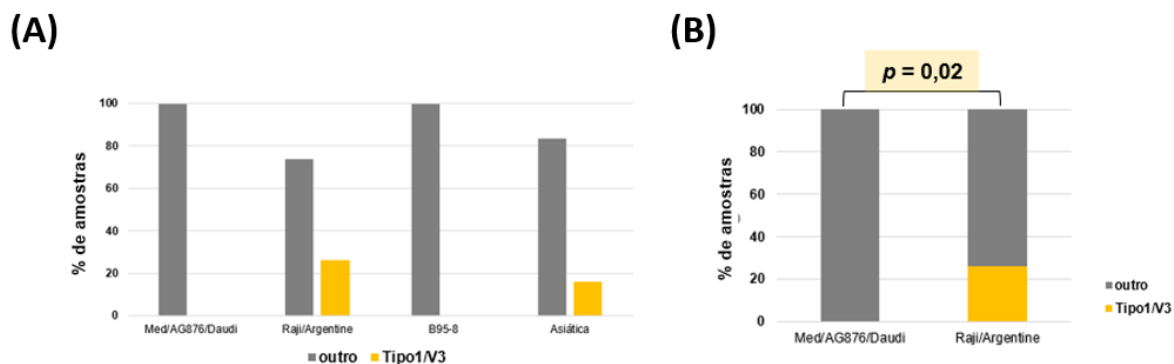


Figura 4.12 - **Identificação do haplótipo Tipo1/V3 nos clados da LMP1 do presente estudo.** (A) Presença do haplótipo Tipo1/V3 (amarelo) nos clados estudados. (B) Presença do haplótipo tipo1/V3 nos dois clados que abarcam mais de 70% da amostragem estudada. Eixo Y identifica a porcentagem da amostragem em cada clado (eixo X).

#### 4.2 Genoma parcial do EBV do subclado South American isolado de um linfoma de Burkitt

Uma amostra representativa da nova sublinhagem da Raji descrita no Brasil, e aqui renomeada como “South American” foi escolhida para ter seu genoma completo elucidado com o objetivo de contribuir com o conhecimento da diversidade do EBV a nível genômico. Desta forma, uma amostra de paciente com linfoma de Burkitt (CEMO3), caracterizada por possuir uma alta carga viral foi submetida ao sequenciamento do genoma completo por NGS.

O genoma parcial do EBV (sem as regiões repetitivas) foi obtido a partir de uma amostra de linfoma de Burkitt de um paciente HIV+ coletada em 2005 no Rio de Janeiro, Brasil. A paciente do sexo feminino tinha de 4 anos de idade e diagnóstico de linfoma de Burkitt associado a AIDS. A amostra corresponde ao tumor retroperitoneal que foi isolado *ex-vivo*, possuindo um cariótipo 46, XX, t(8;14)(q24;q32) (Figura 4.13). O *status* de associação com o EBV foi confirmado por EBER-ISH e PCR para o gene *EBNA3C*, este último indicando que a variante do genoma pertence ao tipo 1 do EBV. O sequenciamento de nova geração gerou 1.826.135 *reads*, das quais 12.488 *reads* correspondentes ao genoma do EBV. Após a montagem, foram gerados 32 *scaffolds* com N50 de 11.683 pb que em seguida originaram uma sequência consenso de 163.115 pb. Algumas regiões não cobertas pelo NGS foram ressequenciadas pelo método de Sanger resultando em uma sequência genômica final de 171,821 kb para o isolado CEMO3 (Figura 4.14).

Em análise preliminar, o isolado CEMO3 foi definido pertencer ao tipo 1 do EBV, abrangendo os principais polimorfismos da LMP1 da que definem a variante South



American. O objetivo do isolamento do genoma inteiro do EBV no presente trabalho, é contribuir com o conhecimento da diversidade do EBV a nível genômico da nossa região geográfica, uma vez que há um número insipiente de genomas brasileiros publicados (n=6), todos oriundos de linfoma de Burkitt da microrregião de São Paulo, onde 83,3% pertencem ao clado da variante-Raji e 16,7% a variante Asiática a nível genômico pela filogenia (LEI *et al*, 2015), fortalecendo mais uma vez a estreita relação entre as cepas do EBV do Brasil e da África. A análise de polimorfismos de outros alvos no presente trabalho indicou que além da identificação da alta frequência do subclado South American (anteriormente classificada como uma variante da cepa Raji em diferentes estudos) a nível genômico em amostras tumorais brasileiras, é observado que a relação haplotípica de oncogenes virais não é homogênea, favorecendo a formação de diferentes haplótipos (Tabela 4.7).

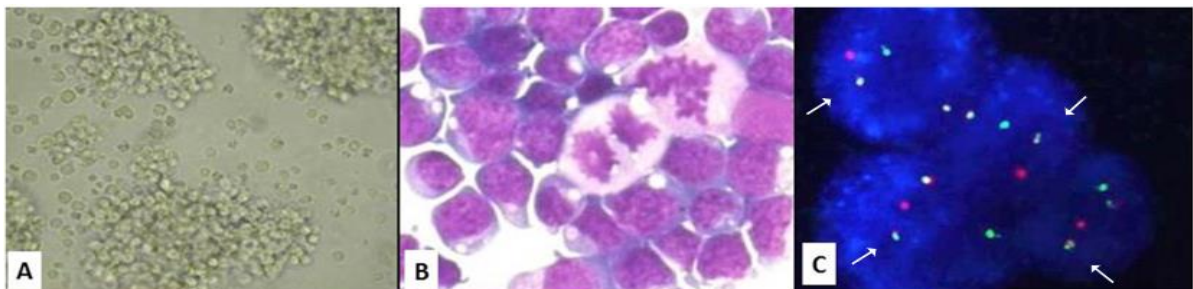


Figura 4.13 - **Células do linfoma de Burkitt do isolado CEMO3.** (A) Cultura de células de LB, com a característica de crescimento em grumos similar às linhagens linfoblásticas EBV+ por causa das moléculas de adesão em sua superfície. (B) Aparência morfológica das células de Burkitt em divisão celular, após *cytospin* e coloração de *May-Grünwald-Giemsa*. (C) Ensaio de FISH revelando a translocação de *MYC* (sinal vermelho) e *IGH* (sinal verde), e seu derivativo (sinal amarelo, indicado pela seta branca) caracterizando a  $t(8;14)(q24;q32)$ . Fotos cedidas por Gustavo Stefanoff e Rocio Hassan.

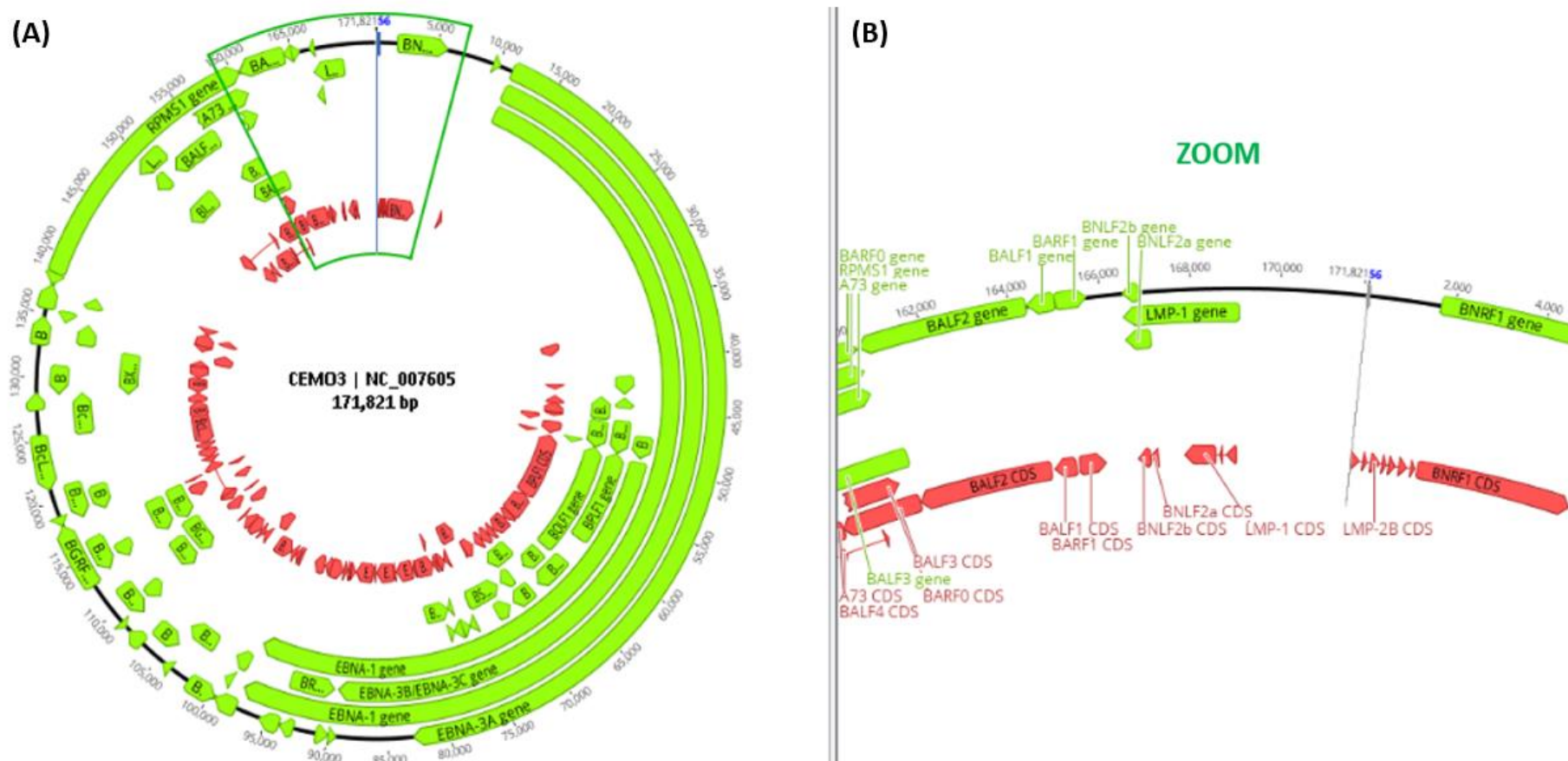


Figura 4.14 - **Esquema do genoma parcial CEMO3.** ORFs (fase de leitura aberta, do inglês: *open reading frame*) são mostradas em verde, e as regiões codificantes em rosa. (A) Esquema em formato de *Circus plot*. (B) Ampliação da área selecionada em verde escuro na parte A da figura, demonstrando a orientação e fase de leitura de alguns oncogenes virais. A anotação do genoma foi realizada utilizando o protótipo B95-8 como referência no programa Geneious®.

Tabela 4.7 – Relações haplotípicas de genomas brasileiros classificados como “South American” pela LMP1.

Genoma	Genbank (número de acesso)	País	Haplótipos nas V-Raji				
			<i>EBNA2/3C</i>	<i>EBNA-1</i>	<i>LMP1</i>	<i>Zp</i>	<i>Z</i>
B95-8	V01555	US	Tipo 1	P-ala	WT/ins15/WT	V1	Zta-C
Raji	KF717093	NG	Tipo 1	V-leu Arg	WT/ins15/WT	V1	Zta-C
FNR	KR063345	BR	Tipo 1	P-thr'	I124V/I152L/ins15/del30	V1	Zta-C
MP	KP968258	BR	Tipo 1	V-leu*	I124V/I152L/ins15/del30	V1	Zta-B4
CCH	KP968257	BR	Tipo 1	V-leu Arg	I124V/I152L/ins15/del30	V1	Zta-B4
CEMO3		BR	Tipo 1	V-leu Arg	I124V/I152L/ins15/del30	V1	Zta-B4
RPF	KR063344	BR	Tipo 1	V-leu	I124V/I152L/ins15/del30	V3	Zta-A2
SCL	KP968259	BR	Tipo 1	V-leu Arg	I124V/I152L/ins15/WT	V3	Zta-A2

EBV: vírus Epstein-Barr; V-Raji: variante Raji; EBNA: antígeno nuclear do Epstein-Barr (do inglês: *Epstein-Barr nuclear antigen*); LMP1: proteína latente de membrana 1 (do inglês: *latente membrane protein 1*); Zp: promotor do gene Z; WT: tipo selvagem (do inglês: *wildtype*); US: Estados Unidos da América (do inglês: *United States*); NG: Nigéria; BR: Brasil; Seleção de cor amarela: haplótipo entre LMP1+EBNA1; Seleção de cor vermelha: haplótipos entre LMP1+Zp+Z; Seleção de cor verde: haplótipo entre LMP1+Zp+Z.

## 5. DISCUSSÃO

Com o objetivo de estudar a diversidade do EBV no território brasileiro, o primeiro desafio do nosso trabalho foi obter uma amostragem significativa de carreadores assintomáticos EBV+. Essas amostras foram utilizadas para detecção e avaliação da carga viral e, quando possível, para a genotipagem utilizando três abordagens: pelo gene *EBNA3C*, a sequência codificante nucleotídica completa da oncoproteína LMP1 e a variante do Zp. A diversidade viral presente no grupo dos carreadores assintomáticos foi então comparada àquela detectada na amostragem de indivíduos com doenças linfoproliferativas benignas e malignas diagnosticados no INCA.

Através da pesquisa preliminar de detecção do EBV na saliva de um grande número de carreadores assintomáticos (n=555), obtivemos dados inéditos a respeito da epidemia do EBV no Brasil, como a positividade de 45,8% do EBV na saliva de indivíduos assintomáticos (sem diagnóstico para doenças associadas ao EBV) e a mediana de carga viral de  $3,35 \times 10^3$  EBV/mL de saliva ( $2,0 \times 10^{-1}$  -  $1,82 \times 10^7$ ). Apesar da incipiência de estudos abordando a diversidade e prevalência da infecção pelo EBV na população brasileira assintomática, estudos realizados na Inglaterra, Suécia e Japão têm relatado positividade do EBV variando entre 20 à 50% na saliva de indivíduos assintomáticos (YAO *et al*, 1985; FALK *et al* 1997; IKUTA *et al* 2000; CORREIA *et al* 2017). O EBV do tipo 1 foi detectado em 79,2% da nossa amostragem assintomática e o tipo 2 em 20,8%, valores próximos aos encontrados anteriormente no Brasil, indicando que a prevalência dos diferentes tipos do EBV nos indivíduos assintomáticos não se alterou muito nos últimos 20-30 anos (CHEN *et al*, 1996; HASSAN *et al*, 2006; QUEIROGA *et al*, 2008; MONTEIRO *et al*, 2020).

Por outro lado, observamos um valor baixo de detecção de coinfeção dos tipos 1 e 2 do EBV (2,2%), diferentemente ao reportado pela literatura no Brasil e em outros países, onde a coinfeção ocorre em cerca de 5-11% das amostras analisadas (CORREIA *et al*, 2017; MONTEIRO *et al*, 2020). Limitações da técnica utilizada para a extração do DNA viral pode explicar a baixa frequência de coinfeções observada em nosso estudo. Entretanto outros fatores, como a idade da população analisada, podem também ter contribuído para explicar essas diferenças. A idade foi um fator significativo a influenciar o tipo do EBV encontrado, uma vez que a detecção do tipo 1 do EBV indivíduos  $\leq 18$  anos foi menor em relação aos indivíduos  $\geq 19$  anos (53,8% versus 80,7%, respectivamente, OR=0,278, IC95%=0,089-0,872 –  $p=0,032$ ). Entretanto, quando consideramos somente os casos de linfomas infantis, o tipo 1 do

EBV foi o mais prevalente (HASSAN *et al*, 2008; BARROS *et al*, 2011). De fato o tipo 1 do EBV demonstra ter maior capacidade de proliferação e sobrevivência das células B, o que explicaria sua maior frequência entre estes tipos de linfomas (ARAUJO *et al*, 2006; TZELLOS *et al*, 2014). No entanto, devemos ter cautela ao interpretar esses resultados, uma vez que podemos ter viés da amostragem (amostragem de conveniência), e devemos também considerar que a tipagem é dependente da identificação do vírus em um indivíduo secretor, o que pode influenciar diretamente as prevalências entre os diferentes grupos. Em nosso trabalho, embora a detecção viral tenha sido menor nos indivíduos  $\leq 18$  anos (39%), em comparação aos adultos (46,2%), esta diferença não foi significativa ( $p=0,2$ ).

A dinâmica clássica da secreção do EBV considera que a variação da carga viral ocorre principalmente intraindivíduo do que interindividual, dependendo do período de convalescência da infecção ou reativação viral (HADINOTO *et al*, 2009). Nossos resultados demonstram que a variação da carga do EBV secretada na saliva de pessoas assintomáticas no território brasileiro pode alcançar valores elevados (milhões de cópias de EBV), contudo, não é possível saber através da carga viral quantas partículas infectantes estão presentes, o que denota a necessidade de conduzir estudos específicos. A reativação do EBV pode ocorrer devido a condições de estresses fisiológicos, e já foi demonstrado que a dinâmica da reativação viral, assim como a carga viral, pode influenciar a patogênese das doenças associadas ao EBV (GESER *et al.*, 1982; CHUNG-CHUN *et al*, 2018). Uma vez que nosso estudo teve um caráter transversal, considerando apenas uma avaliação da carga viral, a identificação de uma alta carga viral não pode ser utilizada como valor prognóstico para o processo patogênico, sendo possivelmente o resultado da variação intraindividual comum.

Em relação aos dados etnográficos obtidos da população e a detecção do EBV na saliva, foi observado que indivíduos que autodeclararam ancestralidade materna de origem africana ou ameríndia tiveram maior possibilidade de detecção do EBV na saliva em comparação com indivíduos de ancestralidade materna autodeclarada de origem europeia. Populações ameríndias da América latina estão sujeitas a uma alta incidência de infecções e desenvolvimento de doenças por vírus latentes linfotrópicos, como HTLV-1 e Herpesvírus 8, entretanto, pouco se sabe sobre a dinâmica de reativação do EBV nessas populações (CHABAY *et al*, 2020). No mais, foi observada uma maior possibilidade de secreção do EBV na saliva em indivíduos cuja escolaridade paterna foi baixa ( $\leq 1^{\circ}$  grau completo). Estes resultados apontam que

a reativação viral, seja por estresse fisiológico ou fatores genéticos ainda não bem estabelecidos, está mais presente na população miscigenada e com menor acesso à educação, possivelmente por conta do baixo poder econômico.

De maneira importante, estrangeiros residentes no Brasil, principalmente aqueles de origem africana e que residiram em seus países de origem até os 18 anos, tiveram maior detecção do EBV na saliva em comparação com os indivíduos brasileiros, correspondendo à dados já descritos na literatura sobre a população africana, onde o EBV é detectado na saliva em 79-90% dos indivíduos (MBULAITEYE *et al*, 2006; MATRAJT *et al*, 2016). Um das hipóteses sugeridas para essa alta detecção do EBV na saliva em território africano, é a presença recorrente de outras infecções que poderiam afetar o sistema imune e conseqüentemente a resposta celular específica ao EBV (MBULAITEYE *et al* 2006), no entanto, quando a população de origem africana ou de outros locais são avaliados como imigrantes, temos que considerar outros fatores. Uma vez que em nossa amostragem os imigrantes viveram no mínimo até os 18 anos em seu local de origem, sendo assim, seu sistema imune foi moldado de acordo com os insultos microbianos de sua região de origem, e uma vez sendo imigrante, novos insultos microbianos específicos da nova região geográfica podem favorecer maior ativação imunológica, portanto maior reativação viral, uma vez que o EBV pode ser reativado e iniciar seu ciclo lítico devido à ativação da célula B em que esteja latente, durante a diferenciação desta em célula plasmática (MÜNZ *et al*, 2015). Outro fator a se considerar é o aumento da carga de fatores relacionados a estresses fisiológicos a imigrantes adultos africanos, devido à mudança de cultura e afastamento da família, principalmente quando essa imigração ocorre em idade adulta (BINGHAM *et al*, 2016 HORMENU *et al*, 2020), podendo ser sugestivo a ocorrer a imigrantes de outras origens, de fato estresse fisiológico é associado com uma piora na resposta imune celular e humoral, o que pode levar a reativação de vírus latentes (SEGERSTROM *et al*, 2004).

### 5.1 Diversidade do EBV circulante no Brasil pelo oncogene *LMP1*

Para classificar o EBV circulante na nossa região geográfica pelo *LMP1*, foram utilizadas duas metodologias, uma vez que não existe consenso para o sistema de classificação do *LMP1*. Até o momento, diversos sistemas de classificações já foram descritos com o intuito de identificar variantes associadas ao tumor (SANDVEJ 1997; EDWARDS *et al*, 1999; WALLING *et al*, 1999; LEI *et al*, 2015), sendo o de Edwards e colaboradores (1999), baseado nas assinaturas aminoácidas dos polimorfismos do

domínio C-terminal da proteína LMP1 o mais utilizado atualmente (NEVES *et al*, 2017; CORREIA *et al*, 2017; GANTUZ *et al*, 2017). No trabalho de Edwards (1999) foi proposto as seguintes variantes B95-8, Med+/-, China1 e 2, Alaskan e NC, as quais são utilizadas atualmente como algumas das cepas de referência da proteína LMP1.

Nós classificamos as assinaturas aminoacídicas da oncoproteína LMP1 por toda sua extensão codificante, uma vez que existem restrições geográficas de variantes genéticas virais, e estudos demonstrando que apenas o domínio C-ter pode não ser suficiente para refletir sua variação total, ou até mesmo definir sua classificação (GANTUZ *et al*, 2017; LIAO *et al*, 2018). Por estes motivos, adicionalmente ao método de Edwards e colaboradores (1999), foram realizadas reconstruções filogenéticas de máxima verossimilhança da região codificante completa do gene *LMP1*, utilizando sequências completas do LMP1 provenientes de linhagens celulares EBV+ e de isolados com especificidade geográfica, reconhecidos por caracterizar variantes virais.

Em relação a diversidade da LMP1 no Brasil, estudos anteriores com amostragem brasileira focaram principalmente em polimorfismos específicos no domínio C-terminal da *LMP1* em casos de linfomas, e grupos comparativos de casos benignos (CHEN *et al*, 1996; GUIRETTI *et al*, 2007), sendo a diversidade circulante no Brasil pelo gene *LMP1* caracterizado apenas mais recentemente por Lei e colaboradores (2015) e Liao e colaboradores (2018). Lei *et al* (2015) sugeriram uma nova classificação da LMP1 através de padrões polimórficos principalmente no promotor e no gene do *LMP1*. Através desses padrões foi observada uma forte relação de sequências sul-americanas (Brasil e Argentina) com sequências de origem africana (Gana), todas oriundas de biópsias de linfoma de Burkitt. Os autores puderam associar um padrão de polimorfismos (descrito como Padrão A) observado na maioria das amostras de linfoma de Burkitt dos dois continentes. Ademais, ocorreu uma alta similaridade em relação a alguns polimorfismos identificados no promotor e gene do LMP1 das amostras sul-americanas e de Gana com as da cepa Raji da Nigéria. De maneira importante, cinco das seis sequências brasileiras estudadas se agruparam dentro do clado da cepa africana Raji (LEI *et al*, 2015).

O estudo de Liao e colaboradores (2018), aumentou a amostragem utilizada para a análise do promotor e gene *LMP1* (+15 amostras argentinas, +11 brasileiras e +45 de Gana), assim, Liao *et al*, confirmaram o padrão descrito por Lei *et al*. na maioria das amostras da América do Sul e africanas de linfoma de Burkitt, sugerindo uma associação tumoral do Padrão A proposto (promotor e gene) em relação da menor

frequência deste padrão nas amostras de indivíduos “saudáveis”. Entretanto, o grupo de casos de malignidade utilizado por Liao *et al*, foi composto de amostras argentinas, brasileiras e ganenses, todos oriundos de linfoma de Burkitt, e o grupo comparativo “saudável” foram amostras de sangue periférico de indivíduos de Uganda (LIAO *et al*, 2018). Em adição, a diversidade do *LMP1* circulante na América do Sul, Gantuz e colaboradores classificaram 45 amostras de casos benignos e 33 casos malignos, todos de origem argentina, com base na classificação por Edwards *et al*. e filogenética de máxima verossimilhança. Foi demonstrado que a classificação por Edwards não é suficiente para classificar variantes da *LMP1*, uma vez que a maioria das amostras investigadas foram classificadas como China1 pela classificação por Edwards e pela filogenia, a maioria das amostras são classificadas como variantes da cepa africana Raji, uma vez se agrupavam dentro do clado desta, porém, acarretando polimorfismos de importância oncológica já descritos que as diferem funcionalmente da cepa original Raji, como o I124V/I152L e del30 (ZUERCHER *et al*, 2012; SUEUR *et al*, 2014; LEI *et al*, 2015). Gantuz *et al* sugeriram uma restrição geográfica da variante Raji devido a maioria das sequências argentinas investigadas terem sido classificadas como tal, e a nomearam como Argentine, no entanto, os autores não identificaram associação com malignidades ( $p > 0,05$ ), em concordância com a primeira descrição dessa variante por Zuercher e colaboradores (2012), que a identificaram em uma coorte de pacientes HIV+ da Suíça. Esta variante esteve presente em 24,7% das amostras de sangue de indivíduos sem tumor, e em 12,2% das amostras de sangue de pacientes com linfoma de Hodgkin clássico HIV+, sem a identificação desta em amostras tumorais (ZUERCHER *et al*, 2012).

Nosso trabalho é o primeiro descrever a diversidade viral de carreadores assintomáticos brasileiros e classificar a diversidade do gene inteiro da *LMP1* em um grande número amostras brasileiras oriundas de casos com linfoproliferações benignas e malignas. A classificação através da sequência inteira da oncoproteína *LMP1* das 83 amostras brasileiras investigadas, indicou a presença de cinco clados com alto suporte. A prevalência das cepas consistiu em 40% das amostras definidas como pertencentes ao clado Med/AG876/Daudi, 39% ao clado Raji/Argentine, 6% ao clado B95-8, 6% ao clado Asiático 1 e 1% ao clado Asiático 2. Cerca de 8% das amostras não se agruparam a um clado específico e não puderam ser classificadas por este método.

Não houve diferença estatística entre a frequência das variantes encontradas entre amostras patológicas e de doadores assintomáticos. No entanto, quando



avaliados os dois principais clados que compreenderam cerca de 79% da amostragem estudada, de maneira surpreendente, é possível identificar que as amostras do grupo de doadores assintomáticos e casos benignos compuseram majoritariamente o clado monofilético Med/AG876/Daudi (61%), enquanto as amostras derivadas de casos malignos, em sua maioria, se agruparam no clado monofilético Raji/Argentine (66,7%). Diversos estudos utilizam sequências do EBV oriundas de patologias benignas, a fim de diferenciar variantes presentes em amostras neoplásicas (GUIRETTI *et al*, 2007; LORENZETTI *et al*, 2014; GANTUZ *et al*, 2017).

Adicionalmente, foram avaliados os polimorfismos de importância oncológica já descritos na literatura nos três diferentes domínios da oncoproteína, N-terminal (N-ter), transmembranar (TM) e C-terminal (C-ter), em relação as variantes identificadas na nossa amostragem. O polimorfismo relativo ao domínio N-ter, a perda do sítio de restrição da enzima *Xho*I (c↓tctgag), foi ausente nas variantes Med/AG876/Daudi, Raji/Argentine e B95-8, e presente em 16,7% das amostras classificadas como variante Asiática ( $p=0,009$ ). Esse polimorfismo é comumente encontrado em amostras de carcinoma de nasofaringe no sudeste asiático, sendo sugerida uma associação com este carcinoma (LIN *et al*, 2003; NGUYEN-VAN *et al*, 2008). Recentemente foi identificado em poucas amostras de origem argentina classificadas principalmente como China2 ou Alaskan, distribuídas em casos de mononucleose infecciosa, hiperplasia linfoide reativa e linfomas, sem associação a malignidades ( $p>0,05$ ) (GANTUZ *et al*, 2017). Devido à ausência desse polimorfismo em nossa amostragem, e sua baixa prevalência na Argentina, esse polimorfismo parece ter uma distribuição limitada, sendo raro na América do Sul e estando possivelmente restrito a certas regiões geográficas ou grupos étnicos (NEVES *et al*, 2017).

Em relação aos polimorfismos na região TM, dois principais, I124V/I152L e F144I/D150A/L151I, foram descritos por caracterizar um aumento da atividade do NF- $\kappa$ B *in vitro* e por estarem associados a amostras tumorais (ZUERCHER *et al*, 2012). O primeiro foi identificado quase que exclusivamente nas variantes do clado Raji/Argentine e em 100% dos casos malignos desta variante ( $p<0,0001$ ). Esse polimorfismo também esteve presente na maioria das amostras brasileiras oriundas de linfoma de Burkitt identificadas por Lei *et al*. (2015). No entanto, a tríade polimórfica F144I/D150A/L151I, só esteve presente nas variantes Asiáticas em grande prevalência. Este achado concorda com o descrito por Zuercher e colaboradores (2012), quando se referem que estes dois conjuntos polimórficos estão presentes em grupos filogenéticos distintos, apesar da convergência de funcionalidade. Ademais, foi

observada uma associação do polimorfismo L151I com as sequências do clado Raji/Argentine, enquanto em menor frequência nas sequências do clado Med/AG876/Daudi, e ausente no clado B95-8 ( $p > 0,0001$ ), em concordância com as amostras reportadas por Lei *et al* (2015).

Os polimorfismos de importância oncológica relacionados ao domínio C-ter, são conhecidos por definir variantes da LMP1, devido à capacidade de alterar a atividade de sinalização (EDWARDS *et al*, 1999; EDWARDS *et al*, 2004). Dentre estes polimorfismos, destaca-se a inserção de 15 pb, que codifica um motivo de sinalização para Janus Kinase 3 (JAK3), e vem sendo associada a amostras tumorais, principalmente em estudos do Brasil e Argentina (GUIRETTI *et al.*, 2007; LORENZETTI *et al.*, 2012). A nossa amostragem confirmou essa associação de casos de malignidades em comparação com o grupo assintomático e de casos benignos, assim como nas amostras malignas do clado Raji/Argentine (81,3%) quando comparado com o grupo assintomático e casos benignos do clado Med/AG876/Daudi (4,8%).

Não foi possível confirmar na nossa amostragem a associação direta da del30 com casos de linfomas conforme reportado anteriormente por Guiretti *et al* (2007). No entanto, quando avaliado os casos de linfoma agrupados ao clado Raji/Argentine em comparação aos casos assintomáticos e benignos do clado Med/AG876/Daudi, é possível observar a associação da del30 com os casos malignos. Esses resultados refletem dados já reportados na literatura que descreveram a del30 em menor número associada a doadores sadios na Argentina (CORREA *et al.*, 2004). Surpreendentemente, foi observado pela primeira vez na América do Sul a del69, presente exclusivamente em casos de carreadores assintomáticos. A del30 é um polimorfismo associado a linfomas e linfomas associados a AIDS, e também presente na cepa CAO da LMP1, isolada de um carcinoma de nasofaringe e descrita por ter um potencial aumentado de transformação *in vitro* e capaz de causar tumores *in vivo* (HU *et al.*, 1993; GUIRETTI *et al.*, 2007; DOLCETTI *et al.*, 2011; SPINA *et al.*, 2011; LORENZETTI *et al.*, 2012), ao contrário da baixa frequência da del69 que apresenta uma baixa frequência em tumores (KNECHT *et al.*, 1993; HADHRI-GUIGA, *et al.*, 2006). Sueur e colaboradores (2014) demonstraram que a LMP1 tem capacidade de induzir diversas citocinas pró-inflamatórias, como TNF-alfa, TNF-beta, IL-6 e INF-gama, no entanto, a variante da LMP1 com del30 induziu a expressão de citocinas em um nível menor que outras variantes, especialmente IFN-gama, enquanto a LMP1 com del69 estimulou uma maior expressão de citocinas, esses resultados sugerem

um papel de escape imune da resposta antiviral, onde, essa característica em conjunto com outros polimorfismos de importância oncogênica poderia favorecer a seleção positiva de variantes oncogênicas.

De maneira geral, a quantidade de repetições de 33pb (rep33pb) não se associou aos diferentes clados identificados. Entretanto o clado Med/AG876/Daudi possui principalmente 4 ou menos rep33 pb em comparação com a Raji/Argentine, que possui principalmente 5 ou mais. Apesar do número de repetições estarem, por vezes, associadas a amostras tumorais, assim como já reportado em estudos com amostras coletadas em países sul-americanos, ainda não é claro sua funcionalidade, sendo seu alto número sugestivo de estar relacionado ao processo linfomagênico (GUIRETTI *et al.*, 2007; LORENZETTI *et al.*, 2012; GANTUZ *et al.*, 2017).

Nossos resultados confirmam em parte os estudos de GUIRETTI *et al.* (2007) e LORENZETTI *et al.* (2012), onde foi demonstrado uma associação das del30 com um alto número de rep33 e a presença da ins15 em casos neoplásicos sul-americanos. Nossos resultados entretanto são contrários ao de Gantuz *et al.* (2017), que apesar da alta frequência da variante Argentine na população estudada, não encontrou associação da variante e nem dos polimorfismos com casos tumorais. A presença da associação de uma cepa com casos malignos no Brasil, poderia ser explicada, em parte, pela constituição imune da nossa população, sugerindo um escape imune desta variante Raji em nossa região. Apesar do EBV possuir um genoma de dsDNA considerado estável, mecanismos genéticos que contribuem para a variabilidade podem ocorrer, os principais que contribuem para a variabilidade da LMP1 são mutação pontual, deleção e duplicação e recombinação homóloga intra/intercepta (WALLING *et al.*, 1999). De fato o EBV fornece um exemplo bem documentado da ocorrência natural de mutantes de escape de linfócitos T citotóxicos em um vírus geneticamente estável (DE CAMPOS-LIMA *et al.*, 1993).

É possível que baixa descrição dessa variante Raji na literatura em estudos prévios seja devido à utilização de sequências da região C-ter que, como demonstrados por Gantuz *et al.* (2017) e este presente trabalho, não são suficientes para separar essa variante. De fato um trabalho prévio da nossa equipe, em um grupo diferente de amostras de linfomas, utilizando apenas polimorfismos do domínio C-ter, não detectou a presença desta variante (GUIRETTI *et al.*, 2007).

A análise de sequências da LMP1 da nossa região geográfica demonstra similaridade com o que já foi reportado para polimorfismos e variantes do EBV circulantes no mundo. No entanto, a pouca representação isolados de EBV sul-

americanos na literatura, em especial oriundos do Brasil nas classificações mais utilizadas, seja a nível gênico ou genômico (SANDVEJ *et al*, 1997; EDWARDS *et al* 1999; WALLING *et al*, 1999; LORENZETTI *et al*, 2012, LORENZETTI *et al*, 2014; PALSER *et al* 2015; Lei *et al*, 2015; CHIARA *et al*, 2016; CORREIA *et al*, 2017; GANTUZ *et al*, 2017; LIAO *et al*, 2018; BERENSTEIN *et al*, 2018; BLAZQUEZ *et al*, 2021), talvez seja motivo pelo qual o sistema de classificação mais utilizado atualmente não descreve com clareza a diversidade encontrada na nossa região, que poderiam ser explicados, pelo menos em parte, por efeitos fundadores e pressões seletivas.

Para tentar contornar as limitações do estudo da diversidade do EBV através da análise dos polimorfismos genéticos isoladamente, geramos um genoma quase completo (sem as regiões repetitivas) de uma representante do clado Raji/Argentine (cujas análises genéticas e evolutivas serão aprofundadas no próximo tópico, sendo redefinida como “South American”). Observamos que quando analisamos diversos alvos genéticos simultaneamente, a relação haplotípica de oncogenes virais não é homogênea, levantando a necessidade de mais estudos, no intuito de entender essas diferentes relações haplotípicas em amostras tumorais brasileiras. Recentemente, foi observado que análises filogenéticas de diferentes oncogenes isoladamente podem gerar topologias filogenéticas diferentes, geralmente cladogramas parafiléticos que dificultam a definição de cladogramas para uma classificação adequada, justificando a utilização de sistemas de classificação que considerem o genoma completo do EBV (ZANELLA *et al*, 2019). Análises com os genomas brasileiros são necessárias para entendermos melhor a dispersão das cepas de EBV na nossa área geográfica, e de que forma as características genéticas e etnográficas intrínsecas do Brasil permitiram a adaptação e persistência de uma cepa, que a nível genético e genômico está associada a amostras tumorais.

## 5.2 Reavaliando a origem da variante Raji Sul-Americana

Recentemente, Gantuz *et al*, (2017) através análise do gene inteiro da oncoproteína LMP1, definiu uma sublinhagem presente dentro do clado da cepa Raji como uma nova variante que foi denominada como Argentine. Como dito anteriormente, essa nova variante é caracterizada principalmente pelos polimorfismos I124V/152L e a del30, e se agrupam filogeneticamente como um subclado monofilético dentro do clado da cepa africana Raji. Gantuz *et al*, foi o primeiro trabalho a demonstrar a circulação dessa variante na América do Sul, no entanto, apesar da

alta frequência na população argentina em patologias benignas e malignas (46%), a variante não teve associação com amostras tumorais ( $p > 0,05$ ). Parte da amostragem de Gantuz *et al.* conteve amostras que foram previamente classificadas pelo domínio C-ter e definidas como variantes da cepa China1 (LORENZETTI *et al.*, 2012), demonstrando a importância da análise do gene inteiro para classificar variantes da LMP1. Os autores sugeriram que a alta prevalência dessa variante na população argentina poderia ser devida a distribuição geográfica restrita, onde condições genéticas hospedeiro favoreceram a seleção e adaptação da nova cepa, devido a provável introdução da cepa africana Raji original no século XVII-XIX devido ao tráfico de escravos, uma vez que a Argentina também foi um país de entrada da população africana em diáspora (CHIARA *et al.*, 2016; SANTOS, 2017). No entanto, é importante salientar que não foram incluídas sequências brasileiras da LMP1 disponíveis em banco de dados e além disso não foi realizada uma análise filogeográfica formal.

No presente trabalho, 39% das amostras brasileiras, em sua maioria provenientes de linfomas, se agruparam dentro do clado da cepa Raji na análise de classificação de máxima verossimilhança (MV), com 93,8% possuindo os polimorfismos I124V/152L e 71,9% a del30 pb. Em seguida, foram adicionadas 252 sequências da LMP1 disponíveis publicamente nas análises de MV. As cepas brasileiras se agruparam com outras da América do Sul, caracterizando os clados Raji/Argentine e Med como as principais em circulação na América do Sul, reunindo 72,3% das sequências sul-americanas. Curiosamente, ainda que 44% das amostras representativas da América do Sul sejam da Argentina e 55% do Brasil, as duas principais variantes circulantes no continente, os clados Med e Raji/Argentine refletem o espectro das principais variantes circulantes no Brasil detectadas no presente trabalho (79%). Em consequência do clado Raji/Argentine se destacar como a variante mais frequente em circulação na América do Sul (60,7%), a restrição desta a amostras tumorais brasileiras, e principalmente, o fato do Brasil ter sido principal porta de entrada sul-americana da diáspora africana durante os séculos XVI-XIX, tendo recebido por volta de 4 milhões de escravos a mais que a Argentina (SANTOS, 2017), resolvemos testar a hipótese da origem geográfica da variante Argentine. Para tanto, reconstruímos a dinâmica das migrações virais do clado Raji/Argentine através de análises filogeográficas bayesianas. A análise bayesiana indicou uma taxa de evolução do gene *LMP1* por volta de  $4.8 \times 10^{-5}$  ( $2.2 \times 10^{-5}$  -  $7.6 \times 10^{-5}$ ) substituições por sítio por ano, em concordância com o observado para outros vírus de dsDNA (FIRTH *et al.*, 2010). Assim, a mediana da data da introdução do ancestral africano no

Brasil foi estimada em 1812 (1665 - 1899). E a mediana da data do surgimento do ancestral no BR foi em 1868 (1767 - 1935). Desta forma, nossos achados contrariam a hipótese de Gantuz *et al* (2017), sugerindo que a origem da linhagem “Argentine” na verdade ocorreu no Brasil entre os séculos XVII e XIX. Nossos dados sugerem que além do Brasil ter sido a origem da introdução desta linhagem, ele também atuou como a principal fonte de disseminação da linhagem Sul-Americana derivada da cepa Raji no continente e para além dele. Desta forma, propomos que esta nova variante proposta por Gantuz *et al* (2017) seja renomeada para “South-American”, dada sua importância para origem e disseminação da variante.

Em conjunto, nossos resultados demonstram a grande diversidade genética do EBV circulante no Brasil, e a importância do monitoramento e avaliação das mutações no potencial patogênico das diferentes cepas. Adicionalmente, outro fator que poderia explicar sua oncogenicidade da Raji no Brasil, é a associação haplotípica com outros genes do genoma viral, como os polimorfismos do Tipo1/V3, pelo gene *EBNA3C* e o promotor do gene lítico *Z* (*Zp*), este último sendo um dos transcritos iniciais produzidos na troca do perfil latente para o lítico. O haplótipo Tipo1/V3 é sugerido de possuir funcionalmente um alto poder replicativo (BRISTOL *et al*, 2018), e tem sido reportado em populações específicas, como em casos de carcinoma de nasofaringe no sudeste asiático (TONG *et al*, 2003) e em linfomas associados a AIDS (MARTINI *et al*, 2007), assim como a del30 pb da LMP1 como discutido anteriormente. Em análise preliminar em nosso presente trabalho, analisamos estes dois genótipos em 51 amostras com variantes LMP1 classificadas, onde 7 amostras foram positivas para o haplótipo Tipo1/V3 (13,7%), 85,6% (n=6) são variantes dentro do clado Raji, e 14,4% (n=1) Asiático 2 ( $p=0,08$ , para todas as variantes). Em relação, ao clado Med/AG876/Daudi, associado ao grupo de carreadores assintomáticos e casos benignos, onde esse haplótipo Tipo1/V3 esteve ausente, a comparação com as variantes Raji ( $p=0,027$ ). Estes resultados em conjunto com a análise preliminar do genoma, onde avaliamos alguns polimorfismos de genes de importância oncológica, e observamos diferentes haplótipos dentro de uma mesma variante da LMP1, esses dados reforçam a investigação de outros genes. Zanella *et al* (2019) demonstraram que através de um novo sistema de classificação baseado em análises filogenéticas do genoma inteiro, a utilização de um gene não é o suficiente para a classificação do EBV. Em conjunto, isso demonstra que o espectro de oncogenicidade do EBV pode ir muito além de um único gene, tornando-se importante avaliar diferentes haplótipos virais.

Por fim, o potencial oncogênico viral, pode não ser a única explicação da maior frequência da variante Raji a amostras tumorais em nossa população, uma vez que os polimorfismos não estão 100% presentes em todos os casos malignos, como a del30 pb (68,8%) e os genótipos Tipo1/V3 (26,1%). Assim, é importante levar em conta fatores genéticos do hospedeiro que poderiam favorecer a transmissão de cepas virais com potencial oncogênico, devido à constituição imune da específica de uma determinada população, uma vez que o câncer é uma doença complexa e multifatorial. Em suma, torna-se importante a avaliação e monitoramento de diferentes alvos virais, assim como a imunogenética do hospedeiro que poderiam explicar, em parte, a diferença do potencial oncogênico da variante Raji e suas sublinhagem - nesse caso a recém proposta variante "South-American" - em populações geográficas diferentes, como na Argentina, em que a prevalência da cepa é alta, mas a associação com o tumor, baixa. Dessa forma, as perspectivas do presente trabalho, é entender de que forma a variante Raji se estabeleceu principalmente em amostras tumorais brasileiras, a fim de testar a hipótese da baixa imunogenicidade dessa variante, onde o escape imune a HLAs específicos da nossa população poderiam favorecer a sobrevivência e persistência viral, podendo levar ao desenvolvimento do câncer.

## **6 CONCLUSÕES**

- A população brasileira apresenta uma positividade elevada de secreção do EBV na saliva com uma prevalência dos tipos 1 e 2 que aparentemente tem se mantido constante nos últimos 30 anos.
- Encontramos uma associação significativa entre o clado Med e carreadores assintomáticos e benignos, e do clado Raji e casos malignos.
- A classificação clássica sugerida por Edwards *et al* e as análises filogenéticas considerando o gene inteiro apresentam divergências, sugerindo a existência de linhagens não contempladas dentre as referências sugeridas por Edwards ou a existência de possíveis recombinantes, principalmente no EBV presente nos casos malignos.
- Os polimorfismos de importância oncológica estavam majoritariamente presentes no subclado sul-americano do clado Raji, e ausente no clado Med. Dentre as sequências

presentes nesse subclado, as provenientes dos casos malignos apresentaram uma frequência maior dos polimorfismos com maior potencial oncogênico.

- Existem duas linhagens principais do EBV circulando na América do Sul, correspondendo aos clados agrupando as linhagens referência Raji e Med, com uma prevalência estimada em >70%.

- O subclado sul-americano da linhagem Raji está presente no Brasil, Argentina e Peru. Este subclado se originou através de uma única introdução viral proveniente do continente Africano para o Brasil entre os séculos XVII e XIX. Posteriormente, o Brasil (e em menor intensidade a Argentina) atuou como uma fonte de exportação desta linhagem para outros países do continente americano (Argentina e Peru) e para a Europa (Suíça).

- A sublinhagem da Raji, identificada em grande parte das amostras do presente trabalho, definida como “Argentine” em 2017, foi definida nesse trabalho como “South-American”. Essa sublinhagem foi associada a casos malignos, possuindo polimorfismos potencialmente oncogênicos na LMP1, e haplótipos no genoma de importância oncológica que favorecem a replicação viral.

- A análise de haplótipos do genoma completo, corrobora com Zanella et al (2019), onde confirma que a utilização de um gene não é suficiente para a classificação de variantes EBV.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

ABDEL-HAMID, M. *et al.* EBV strain variation: geographical distribution and relation to disease state. **Virology**, v. 190, n. 1, p. 168-75, 1992.

ABREU, E. S. *et al.* Doença de Hodgkin infanto-juvenil no estado do Ceará e sua relação com o vírus de Epstein-Barr: parâmetros clínicos e análises morfológicas imuno-histoquímica e por hibridização *in situ*. **Jornal Brasileiro de Patologia**, v. 33, n. 4, p. 178-84, 1997.

ALBA, M. M. *et al.* Genome wide function conservation and phylogeny in the Herpesviridae. **Genome Research**, v. 11, n. 1, p. 43-54, 2001.

AMBINDER, R. F. *et al.* Epstein -Barr virus and childhood Hodgkin's disease in Honduras and United States. **Blood**, v. 81, p. 462-7, 1993.



AMBINDER, R. F., MULLEN, M., CHANG, Y-N., HAYWARD, G. S., and HAYWARD, S. D. (1991). Functional domains of Epstein-Barr virus nuclear antigen EBNA 1.1. **Virology** 65, 1466-1478.

ANDREWS S. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>, 2010.

ARAUJO, I. *et al.* The high frequency of EBV infection in pediatric Hodgkin lymphoma is related to the classical type in Bahia, Brazil. **Virchows Archives**, v. 449, n. 3, p. 315-9, 2006.

AVIEL, S. *et al.* Degradation of the Epstein-Barr virus latent membrane protein 1 (LMP1) by the ubiquitin-proteasome pathway. Targeting via ubiquitination of the N-terminal residue. **Journal of Biological Chemistry**. v. 275, n. 31, p. 23491-9, 2000.

BABCOCK, G. J. *et al.* EBV persistence in memory B cells in vivo. **Immunity**, v. 9, n. 3, p. 395-404, 1998.

BAER, R. *et al.* DNA sequence and expression of the B95-8 Epstein-Barr virus genome. **Nature**, v. 310, n. 5974, p. 207-11, 1984.

BANKIER, A.T. *et al.* DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-Barr virus containing the terminal repeat sequences. **Molecular Biology & Medicine** 1: 425-445, 1983.

BARROS, M. H. M. *et al.* Cell cycle characteristics and Epstein-Barr virus are differentially associated with aggressive and non-aggressive subsets of Hodgkin lymphoma in pediatric patients. **Leukemia & lymphoma**, v. 51, n. 8, p. 1513–1522, ago. 2010.

BARROS, M. H.; HASSAN, R.; NIEDOBITEK, G. Disease patterns in pediatric classical Hodgkin lymphoma: a report from a developing area in Brazil. **Hematological Oncology**, v. 29, n. 4, p. 190-5, 2011.

BARROS, M. H. M.; VERA-LOZADA, G.; SOARES, F.A.; NIEDOBITEK, G.; HASSAN, R. Tumor microenvironment composition in pediatric classical Hodgkin lymphoma is modulated by age and Epstein-Barr virus infection. **International Journal of Cancer** 131:1142–52, 2012.

BARROS, M. H. M.; VERA-LOZADA, M. G.; SEGGES, P.; HASSAN, R.; NIEDOBITEK, G. Revisiting the Tissue Microenvironment of Infectious Mononucleosis: Identification of EBV Infection in T Cells and Deep Characterization of Immune Profiles. **Front Immunol**. 2019.

BINGHAM, BA. *et al.* The Association between Stress Measured by Allostatic Load Score and Physiologic Dysregulation in African Immigrants: The Africans in America Study. **Front Public Health**. Nov 25;4:265. 2016.

BENTZ, GL. *et al.* Epstein-Barr virus latent membrane protein 1 (LMP1) C-terminal-activating region 3 contributes to LMP1-mediated cellular migration via its interaction with Ubc9. **J Virol**, 85(19):10144–53, 2011.

BENTZ, GL. *et al.* Epstein-Barr virus latent membrane protein 1 regulates the function of interferon regulatory factor 7 by inducing its sumoylation. **J Virol**, 86(22):12251–61, 2012.

BERENSTEIN AJ. *et al.* Recombination rates along the entire Epstein Barr virus genome display a highly heterogeneous landscape. **Infect Genet Evol.** Nov;65:96-103. 2018

BERGER, C. *et al.* Carboxy terminal variants of Epstein-Barr virus-encoded latent membrane protein 1 during long-term human immunodeficiency virus infection: reliable markers for individual strain identification. **Journal of Infectious Diseases**, v. 179, n. 1, p. 240-4, 1999.

BIGGAR, R.J.; HENLE, G; BÖCKER, J; LENNETTE, E.T; FLEISHER, G; HENLE, W. Primary EBV infections in African infants. II. Clinical and serological observations during seroconversion. **International Journal of Cancer**, 22: 244-250, 1978.

BLAZQUEZ, AC. *et al.* Comprehensive Evolutionary Analysis of Complete Epstein–Barr Virus Genomes from Argentina and Other Geographies. **Viruses**, 13(6), 1172, 2021.

BORNKAMM, G. W.; HAMMERSCHMIDT, W. Molecular virology of Epstein Barr virus. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 356, n. 1408, p. 437-59, 2001.

BRISTOL, JA. A cancer-associated Epstein-Barr virus BZLF1 promoter variant enhances lytic infection. **Plos Pathog** 14(7): e1007179. 2018

Brodeur, SR. *et al.* Localization of the major NF-kappa B-activating site and the sole TRAF3 binding site of LMP-1 defines two distinct signaling motifs. **Journal of Biological Chemistry**, 272(32):19777–19784, 1997.

BURKITT, D. A children's cancer dependent on climatic factors. **Nature**, 194:232–4, 1962.

BURKITT, D. A sarcoma involving the jaws in African children. **British Journal of Surgery**, 46 (197): 218 –23, 1958.

BURKITT, D. A tumour syndrome affecting children in tropical Africa. **Postgraduate Medical Journal**, 38:71–9, 1962.

BURKITT, D. Determining the climatic limitations of a children's cancer common in Africa. **British Journal of Surgery**, 2(5311):1019–23, 1962.

CALCIAN, L. *et al.* C-terminal region of EBNA-2 determines the superior transforming ability of type 1 Epstein-Barr virus by enhanced gene regulation of LMP-1 and CXCR7. **Plos Pathog** Jul;7(7):e1002164. 2011.

CARBONE, A.; GLOGHINI, A.; DOTTI, G. EBV-associated lymphoproliferative disorders: classification and treatment. **Oncologist**, v. 13, n. 5, p. 577-85, May 2008.

CHABAY, P.; LENS, D.; HASSAN, R. *et al.* Lymphotropic Viruses EBV, KSHV and HTLV in Latin America: Epidemiology and Associated Malignancies. A Literature-Based Study by the RIAL-CYTED, **Cancers**, 2020.

CHAN, K. H. *et al.* Epstein-Barr virus (EBV) infection in infancy. **Journal of Clinical Virology**, v. 21, n. 1, p. 57-62, 2001.

CHANG, K. L. *et al.* High prevalence of Epstein-Barr virus in the Reed-Sternberg cells of Hodgkin's disease occurring in Peru. **Blood**, v. 81, p. 496-501, 1993.

CHANG, C. M. *et al.* The extent of genetic diversity of Epstein-Barr virus and its geographic and disease patterns: a need for reappraisal. **Virus Res**, v. 143, n. 2, p. 209-21, 2009.

CHEN, M. L. *et al.* Cloning and characterization of the latent membrane protein (LMP) of a specific Epstein-Barr virus variant derived from the nasopharyngeal carcinoma in the Taiwanese population. **Oncogene**, v. 7, n. 11, p. 2131-40, 1992.

CHEN, WG. *et al.* Genotyping of Epstein-Barr virus in Brazilian Burkitt's lymphoma and reactive lymphoid tissue - Type A with a high prevalence of deletions within the latent membrane protein gene. **American J of Pathol.** 148, n. 1, p. 17-23, 1996

CHIARA M. *et al.* Geographic Population Structure in Epstein-Barr Virus Revealed by Comparative Genomics. **Genome Biol Evol Dec** 14;8(11):3284-3291. 2016

CHUNG, P. J. *et al.* Negative regulation of Epstein-Barr virus latent membrane protein 1-mediated functions by the bone morphogenetic protein receptor IA binding protein, BRAM1. **Journal of Biological Chemistry**, v. 277, n. 42, p. 39850-7, 2002.

CHUNG-CHUN, W. *et al.* Perspective: Contribution of Epstein-Barr virus (EBV) Reactivation to the Carcinogenicity of Nasopharyngeal Cancer Cells. **Cancers** Apr; 10(4): 120. 2018 .

CORREA, RM. *et al.* Epstein-barr virus (EBV) in healthy carriers: Distribution of genotypes and 30 bp deletion in latent membrane protein-1 (LMP-1) oncogene. **J Med Virol** Aug;73(4):583-8. 2004

CORREIA, S. *et al.* Sequence variation of Epstein-Barr virus: viral types, geography, codon usage, and diseases. **J Virol.** 92:18, 2018.

CORREIA, S.; PALSER, A. *et al.* Natural Variation of *Epstein-Barr Virus* Genes, Proteins, and Primary MicroRNA. **J Virol.** 12;91(15):e00375-17, 2017.

CRAWFORD, D. H. Biology and disease associations of Epstein-Barr virus. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 356, n. 1408, p. 461-73, 2001.

CRAWFORD, D. H. *et al.* Sexual history and Epstein-Barr virus infection. **J Infect Dis.** 186(6):731-6, 2002.

DAMBAUGH, T.; HENNESSY, K.; CHAMNANKIT, L.; KIEFF, E. U2 region of Epstein-Barr virus DNA may encode Epstein-Barr nuclear antigen 2. **Proc Natl Acad Sci U S A**, 81(23):7632–6, 1984.

DE MARTEL, *et al.* Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. **Lancet Glob Health**, Feb;8(2):e180-e190, 2020.

DEVERGNE, O. *et al.* Association of TRAF1, TRAF2, and TRAF3 with an Epstein-Barr virus LMP1 domain important for B-lymphocyte transformation: role in NF-kappaB activation. **Molecular and Cellular Biology**, v. 16, n. 12, p. 7098-108, 1996.

DEVERGNE, O. *et al.* Role of the TRAF binding site and NF-kappaB activation in Epstein-Barr virus latent membrane protein 1-induced cell gene expression. **Journal of Virology**, v. 72, n. 10, p. 7900-8, 1998.

DOLAN, A. *et al.* The genome of Epstein-Barr virus type 2 strain AG876. **Virology** 350:164–170, 2006.

EDWARDS, R. H.; SEILLIER-MOISEWITSCH, F.; RAAB-TRAUB, N. Signature amino acid changes in latent membrane protein 1 distinguish Epstein-Barr virus strains. **Virology**, v. 261, n. 1, p. 79-95, 1999.

EDWARDS, R. H. *et al.* Potential selection of LMP1 variants in nasopharyngeal carcinoma. **Journal of Virology**, v. 78, n. 2, p. 868-81, 2004.

EHLERS, B. *et al.* Lymphocryptovirus phylogeny and the origins of Epstein-Barr virus **Journal of General Virology**, v. 91, n. Pt 3, p. 630-42, 2010.

ELIOPOULOS, A. G.; YOUNG, L. S. LMP1 structure and signal transduction. **Seminars in Cancer Biology**, v. 11, n. 6, p. 435-44, 2001.

ENGEL, M. *et al.* Improved prognosis of Epstein-Barr virus associated childhood Hodgkin's lymphoma: study of 47 South African cases. **Journal of Clinical Pathology**, v. 53, n. 3, p. 182-6, 2000.

EPSTEIN, M.A.; ACHONG, B.G.; BARR, Y.M. Virus Particles in cultured Lymphoblasts from Burkitt's Lymphoma. **Lancet**, 1(7335):702–3, 1964.

FALK, KI. *et al.* Direct identification by PCY of EBV types and variants in clinical samples. **J Med Virol**. 1997

FARRELL, P. J. *et al.* Epstein-Barr virus BZLF1 trans-activator specifically binds to a consensus AP-1 site and is related to c-fos. **EMBO J**, 8(1):127–32, 1989.

FARRELL, P. J. Epstein–Barr virus and cancer. **Ann Rev Pathol**, v.14:29–53, 2019.

FENNEWALD, S.; VAN SANTEN, V.; KIEFF, E. Nucleotide sequence of an mRNA transcribed in latent growth transforming virus infection indicates that it may encode a membrane protein. **J. Virol**. 51:411-419, 1984.

FERREIRA, J. M. *et al.* Lymphoma subtype incidence rates in children and adolescents: first report from Brazil. **Cancer Epidemiology**, v. 36, n. 4, p. e221-6, 2012.

FERRÉS, M. *et al.* Seroprevalencia de infección por virus de Epstein-Barr em población sana de Santiago de Chile. **Ver Med Chile**, 123 (12): 1447-1452, 1995.

FIGUEIRA-SILVA, C.M.; PEREIRA, F.E.L.. Prevalence of Epstein-Barr virus antibodies in healthy children and adolescents in Vitória, State of Espírito Santo, Brazil. **Revista da Sociedade Brasileira de Medicina Tropical**, Uberaba, v. 37, n. 5, p. 409-412, 2004.

FLEISHER, G. *et al.* Primary infection with Epstein-Barr virus in infants in the United States: clinical and serologic observations. **Journal of Infectious Diseases**, v. 139, n. 5, p. 553-8, 1979.

FLOETTMANN JE. Epstein-Barr virus latent membrane protein-1 (LMP1) C-terminus activation region 2 (CTAR2) maps to the far C-terminus and requires oligomerization for NF-kappaB activation. **Oncogene**, 15(15):1851–8, 1997.

GANTUZ, M.; LORENZETTI, M. A.; CHABAY, P.A. A novel recombinant variant of latent membrane protein 1 from Epstein Barr virus in Argentina denotes phylogeographical association. **PLoS One**, v.12: e0174221, 2017.

GESER, A; DE THÉ, G. *et al.* Final case reporting from the ugandan prospective study of the relationship between ebv and burkitt's lymphoma. **International Journal of Cancer**, v. 29, n. 4, 1982.

GIRES, F. *et al.* Latent membrane protein 1 of Epstein-Barr virus interacts with JAK3 and activates STAT proteins. **EMBO J**, 18(11):3064–73, 1999.

GLASER, S. L. *et al.* Hodgkin lymphoma incidence in California Hispanics: influence of nativity and tumor Epstein-Barr virus. **Cancer Causes Control**, v. 25(6):709-25, 2014.

GRIVENNIKOV, S. I.; GRETEN, F. R.; KARIN, M. Immunity, Inflammation, and Cancer. **Cell**, v. 140, n. 6, p. 883–899, mar. 2010.

GUINDON, S.; GASCUELO. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. **Systematic Biology**, 52(5):696-704, 2003.

GUIRETTI, D.M. *et al.* Structural variability of the carboxy-terminus of Epstein–Barr virus encoded latent membrane protein 1 gene in Hodgkin's lymphomas. **Journal of Medical Virology** Volume 79, Issue 11, 2007.

GUVEN-MAIOROV, E.; TSAI, C. J.; NUSSINOV, R. Structural host-microbiota interaction networks. **PLoS Comput Biol**, 13: e1005579, 2017.

HADHRI-GUIGA, B. *et al.* Various 30 and 69 bp deletion variants of the Epstein-Barr virus LMP1 may arise by homologous recombination in nasopharyngeal carcinoma of Tunisian patients. **Virus Res** 115(1):24–30. 2006.

HADINOTO, V. *et al.* On the dynamics of acute EBV infection and the pathogenesis of infectious mononucleosis. **Blood**, v. 111, n. 3, p. 1420-7, 2008.

HADINOTO, V. *et al.* The dynamics of EBV shedding implicate a central role for epithelial cells in amplifying viral output. **PLOS Pathogens**, 5(7): e1000496, 2009.

HANAHAN, D. WEINBERG, RA. Hallmarks of cancer: the next generation. **Cell** 2011 Mar 4;144(5):646-74

HASSAN, R. *et al.* Epstein-Barr virus (EBV) detection and typing by PCR: a contribution to diagnostic screening of EBV-positive Burkitt's lymphoma. **Diagn Pathol**, v. 1, p. 17, 2006.

HASSAN, R. *et al.* Clinical and demographic characteristics of Epstein-Barr virus-associated childhood Burkitt's lymphoma in Southeastern Brazil: epidemiological insights from an intermediate risk region. **Haematol** May;93(5):780-3, 2008.

HASSANPOUR, HS. MOHAMMADAMIN, D. Review of cancer from perspective of molecular. **J Can Res and Pract** 2017

HENLE, W.; DIEHL, V.; KOHN, G.; ZUR HAUSEN, H.; HENLE, G. Herpes-type virus and chromosome marker in normal leukocytes after growth with irradiated Burkitt cells. **Science**, v. 157, 1064–1065, 1967.

HERBST, H. *et al.* Epstein-Barr virus latent membrane protein expression in Hodgkin and Reed-Sternberg cells. **Proceedings of the National Academy of Sciences of the United States of America**, v. 88, n. 11, p. 4766-70, 1991.

HIGUCHI, E. *et al.* The Epstein-Barr virus latent membrane protein 1 putative Janus kinase 3 (JAK3) binding domain does not mediate JAK3 association or activation in B-lymphoma or lymphoblastoid cell lines. **J Virol**, 76(1):455–9, 2002.

HISLOP, A. D. *et al.* Cellular responses to viral infection in humans: lessons from Epstein-Barr virus. **Annual Review of Immunology**, v. 25, p. 587-617, 2007.

HJALGRIM, H. *et al.* Characteristics of Hodgkin's lymphoma after infectious mononucleosis. **New England Journal of Medicine**, v. 349, n. 14, p. 1324-32, 2003.

HJALGRIM, H. *et al.* Infectious mononucleosis, childhood social environment, and risk of Hodgkin lymphoma. **Cancer Research**, v. 67, n. 5, p. 2382-8, 2007.

HJALGRIM, H. *et al.* HLA-A alleles and infectious mononucleosis suggest a critical role for cytotoxic T-cell response in EBV-related Hodgkin lymphoma. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 14, p. 6400-5, 2010.

HJALGRIM, H. *et al.* Risk of Hodgkin's disease and other cancers after infectious mononucleosis. **Journal of the National Cancer Institute**, v. 92, n. 18, p. 1522-8, 2000.

HOCHBERG, D. *et al.* Demonstration of the Burkitt's lymphoma Epstein-Barr virus phenotype in dividing latently infected memory cells in vivo. **Proc Natl Acad Sci U S A**, v. 101, n. 1, p. 239-44, 2004.

HORMENU, T. *et al.* Stress Measured by Allostatic Load Varies by Reason for Immigration, Age at Immigration, and Number of Children: The Africans in America Study. **Int J of Environ Res and Public Health**, Vol. 17, Ed. 12, 2020.

HU, L. F. *et al.* Clonability and tumorigenicity of human epithelial cells expressing the EBV encoded membrane protein LMP1. **Oncogene**, v. 8, n. 6, p. 1575-83, 1993.

HUANG, X. *et al.* HLA-A\*02:07 is a protective allele for EBV negative and a susceptibility allele for EBV positive classical Hodgkin lymphoma in China. **PLoS One**, v. 7, n. 2, p. e31865, 2012.

HUMANS, I. W. G. O. T. E. O. C. R. T.; CANCER, T. I. A. F. R. O. Epstein-Barr Virus and Kaposi's Sarcoma Herpesvirus/Human Herpesvirus 8: **IARC Working Group on the Evaluation of Carcinogenic Risks to Humans**, Lyon, 1997.

IKUTA, K. *et al.* Detection of Epstein-Barr virus in salivas and throat washings in healthy adults and children. **Microbes Infect.** 2000.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Estimativa 2021: incidência de câncer no Brasil**. [s.l.] Coordenação de Prevenção e Vigilância, 2021.

IZUMI, K. M.; KAYE, K. M.; KIEFF, E. Epstein-Barr virus recombinant molecular genetic analysis of the LMP1 amino-terminal cytoplasmic domain reveals a probable structural role, with no component essential for primary B lymphocyte growth transformation. **J Virol**, v. 68, n. 7, p. 4369-76, 1994.

JARRET, RF. MACKENZIE, J. Epstein-Barr virus and other candidate viruses in the pathogenesis of Hodgkin's disease.

JOHANNSEN, E. *et al.* Proteins of purified Epstein-Barr virus. **Proceedings of the National Academy of Sciences of the United States of America**, 101(46):16286–91, 2004

JONES, MD. *et al.* The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion similar to that observed in a non-transforming strain (P3HR-1) of the virus. **EMBO J.** 3, 813–21. 1984.

JOSEPH, A. M.; BABCOCK, G. J.; THORLEY-LAWSON, D. A. Cells expressing the Epstein-Barr virus growth program are present in and restricted to the naive B-cell subset of healthy tonsils. **J Virol**, v. 74, n. 21, p. 9964-71, 2000.

KANDA, T.; YAJIMA, M.; IKUTA, K. Epstein-Barr virus strain variation and cancer. **Cancer Sci.** 110:1132–9, 2019.

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

- KAYE, K. M.; IZUMI, K. M.; KIEFF, E. Epstein-Barr virus latent membrane protein 1 is essential for B-lymphocyte growth transformation. **Proc Natl Acad Sci U S A**, v. 90, n. 19, p. 9150-4, 1993.
- KAYKAS, A.; WORRINGER, K.; SUGDEN, B. LMP-1's transmembrane domains encode multiple functions required for LMP-1's efficient signaling. **J Virol**, v. 76, n. 22, p. 11551-60, 2002.
- KNECHT, H. et al. Deletions within the LMP1 oncogene of Epstein-Barr virus are clustered in Hodgkin's disease and identical to those observed in nasopharyngeal carcinoma. **Blood** 82(10):2937–2942, 1993.
- KIEFF, E.; RICKINSON, A.B. Epstein-Barr Virus and its replication. In: Knipe D, Howley P, editors. *Fields Virology*, 5th edition. 5th ed. **Philadelphia: Lippincott**. pp. 2603-2654, 2007.
- KLUMB, C.E.; HASSAN, R. et al. Geographic variation in Epstein-Barr virus-associated Burkitt's lymphoma in children from Brazil. **Int J Cancer**, 108:66–70, 2004.
- KNECHT, H. et al. Mutational hot spots within the carboxy terminal region of the LMP1 oncogene of Epstein-Barr virus are frequent in lymphoproliferative disorders. **Oncogene**, v. 10, n. 3, p. 523-8, 1995.
- KORSMAN S.; VAN ZYL, G.; PREISER, W.; ANDERSSON, L. N. M. *Virology: An Illustrated Colour Text*. **Churchill Livingstone**, 2012.
- KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. **Mol Biol Evol** 2016, 33, 1870–1874.
- KUPPERS, R. B cells under influence: transformation of B cells by Epstein-Barr virus. **Nat Rev Immunol**, v. 3, n. 10, p. 801-12, 2003.
- KUPPERS, R. The biology of Hodgkin's lymphoma. **Nat Rev Cancer**, v. 9, n. 1, p. 15-27, 2009.
- KWOK, H. et al. Genomic sequencing and comparative analysis of Epstein-Barr virus genome isolated from primary nasopharyngeal carcinoma biopsy. **PLoS One**, 7(5):e36939, 2012.
- KWOK, H. et al. Genomic Diversity of Epstein-Barr Virus Genomes Isolated from Primary Nasopharyngeal Carcinoma Biopsy Samples. **J Virol**. Sep; 88(18): 10662–10672, 2014.
- LACOSTE, V. et al. Genetic diversity and molecular evolution of human and non-human primate Gammaherpesvirinae. **Infect Genet Evol**, v. 10, n. 1, p. 1- 13, 2010.
- LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nat Methods**. 9: 357-359, 2012.



LANGMEAD, B.; TRAPNELL, C.; POP, M.; SALZBERG, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome Biol.** 2009.

LEFKOWITZ, E. J. *et al.* Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV) **Nucleic Acids Res.** 2017

LEI, H. *et al.* Epstein–Barr virus from Burkitt Lymphoma biopsies from Africa and South America share novel LMP-1 promoter and gene variations. **Sci. Rep.** 5, 16706 (2015).

LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics**, 25(14), 1754–1760, 2009.

LIAO, HM. *et al.* Frequency of EBV LMP-1 Promoter and Coding Variations in Burkitt Lymphoma Samples in Africa and South America and Peripheral Blood in Uganda. **Cancers** Jun 2;10(6):177. 2018.

LIN, S.X.; ZONG, Y.S.; WU, Q.L.; HAN, A. J.; LIANG, Y. J. Loss of an XhoI-site within N-terminal region of Epstein-Barr virus LMP1 gene in nasopharyngeal carcinoma. **Ai Zheng** 22 (11), 1147–1151, 2003.

LIN, Z. *et al.* Whole-genome sequencing of the Akata and Mutu Epstein–Barr virus strains. **J. Virol.** 87, 1172–1182, 2013.

LIU, P. *et al.* Direct sequencing and characterization of a clinical isolate of Epstein–Barr virus from nasopharyngeal carcinoma tissue by using next generation sequencing technology. **J. Virol.** 85, 11291–11299, 2011.

LO, AKF. *et al.* The Role of EBV-Encoded LMP1 in the NPC Tumor Microenvironment From Function to Therapy. **Front Oncol** Feb 25;11:640207. 2021.

LOLE, K. S. *et al.* Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. **J. Virol.** 73:152–160, 1990.

LORENZETTI, M. A. *et al.* Distinctive Epstein-Barr virus variants associated with benign and malignant pediatric pathologies: LMP1 sequence characterization and linkage with other viral gene polymorphisms. **J Clin Microbiol** 2012

LORENZETTI, M. A. *et al.* Epstein-Barr virus BZLF1 gene polymorphisms: malignancy related or geographically distributed variants? **Clin Microbiol Infect**, 20:O861–O864, 2014.

LUFTIG, T. Epstein-Barr virus latent infection membrane protein 1 TRAF-binding site induces NIK/IKK alpha-dependent noncanonical NF-kappaB activation. **Proc Natl Acad Sci U S A**, 101(1):141–6, 2004.

ROWE, M. *et al.* Distinction between Epstein-Barr virus type A (EBNA 2A) and type B (EBNA 2B) isolates extends to the EBNA 3 family of nuclear proteins. **J Virol**, 63(3):1031–9, 1989.

MARTINEZ, Z. R. *et al.* Evaluation of IgG-IgM antibodies against the Epstein-Barr virus nuclear antigen (EBNA-1) using ELISA. **Enfermedades Infecciosas y Microbiología Clínica**, 10:281-283, 1992.

MARTINI, M. *et al.* Characterization of variants in the promoter of EBV gene BZLF1 in normal donors, HIV-positive patients and in AIDS-related lymphomas. **J Infect**, Pages 298-306, 2007.

MASSINI, G. *et al.* EBV in Hodgkin Lymphoma. **Mediterranean Journal of Hematology and Infectious Diseases**, 2009.

MCGEOCH, D. J. *et al.* Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. **J Mol Biol**, v. 247, n. 3, p. 443-58, 1995.

MCGEOCH, D. J.; RIXON, F. J.; DAVISON, A. J. Topics in herpesvirus genomics and evolution. **Virus Res**, v. 117, n. 1, p. 90-104, 2006.

MENDOZA, H. R. *et al.* Prevalencia de anticuerpos contra el virus de Epstein-Barr en niños de 9m a 16 años. **Archivos Dominicanos de Pediatría**, 5: 17-18, 1989

MIDDELDORP, J. M.; PEGTEL, D. M. Multiple roles of LMP1 in Epstein-Barr virus induced immune escape. **Semin Cancer Biol**, v. 18, n. 6, p. 388-96, 2008.

MILLER G., Lipman M. Release of infectious Epstein-Barr virus by transformed marmoset leukocytes. **Proc. Natl. Acad. Sci. U.S.A.** 70:190-194, 1973.

MILLER G., Shope T., Lisco H., Stitt D., Lipman M. Epstein-Barr virus: transformation, cytopathic changes, and viral antigens in squirrel monkey and marmoset leukocytes. **Proc. Natl. Acad. Sci. U.S.A.** 69:383-387, 1972.

MILLER, WE. The NPC derived C15 LMP1 protein confers enhanced activation of NF-kappa B and induction of the EGFR in epithelial cells. **Oncogene**, 16(14):1869-77, 1998.

MONTEIRO, TAF. *et al.* Genotypes of Epstein-Barr virus (EBV1/EBV2) in individuals with infectious mononucleosis in the metropolitan area of Belém, Brazil, between 2005 and 2016. **Braz J Infect Dis**. Jul-Aug 2020

MOORE, P. S.; CHANG, Y. Why do viruses cause cancer? Highlights of the first century of human tumour virology. **Nature Rev. Cancer**, v.10, 878-889, 2010.

MÜNZ, C. Epstein-Barr virus, vol 2: one herpes virus: many diseases, Vol: 391, Pages: 61-117, ISSN: 0070-217X. 2015.

NEVES, M.; MARINHO-DIAS, J.; RIBEIRO, J.; SOUSA, H. Epstein-barr virus strains and variations: Geographic or disease-specific variants? **Journal of medical virology**, 2016.

NGUYEN-VAN, D. *et al.* Epstein-Barr virus genetic variation in Vietnamese patients with nasopharyngeal carcinoma: full-length analysis of LMP1. **Virus Genes**. Oct;37(2):273-81. 2008.

NICHOLLS JM. *et al.* The association of squamous cell carcinomas of the nasopharynx with Epstein-Barr virus shows geographical variation reminiscent of Burkitt's lymphoma. **J Pathol.** 183:164 – 8, 1997.

NIENS, M. *et al.* HLA-A\*02 is associated with a reduced risk and HLA-A\*01 with an increased risk of developing EBV+ Hodgkin lymphoma. **Blood**, v. 110, n. 9, p. 3310-5, 2007.

NIENS, M. *et al.* The human leukocyte antigen class I region is associated with EBV-positive Hodgkin's lymphoma: HLA-A and HLA complex group 9 are putative candidate genes. **Cancer Epidemiol Biomarkers Prev**, v. 15, n. 11, p. 2280-4, 2006.

ODUMADE, O. A. *et al.* Progress and problems in understanding and managing primary Epstein-Barr virus infections. **Clin Microbiol Rev** 24:193–209, 2011.

PALSER, A. L. *et al.* Genome diversity of Epstein–Barr virus from multiple tumor types and normal infection. **J. Virol.** 89, 5222–5237. 2015.

PANNUTI, C. S. *et al.* A prospective clinical study of the mononucleosis syndrome in a developing country. **Int J Epidemiol**, v. 9, n. 4, p. 349-53, 1980.

PARKER, B. D.; BANKIER, A.; SATCHWELL, A.; BARRELL, B.; FARRELL, P. J. Sequence and transcription of Raji Epstein-Barr virus DNA spanning the B95-8 deletion region. **Virology.** 179:339, 1990.

PARKIN, D. M. The global health burden of infection-associated cancers in the year 2002. **Int J Cancer** 118: 3030-3044, 2006.

PAYNE, R. *et al.* A New Leukocyte Isoantigen System in Man. **Cold Spring Harb Symp Quant Biol**, v. 29, p. 285-95, 1964.

PEARSON, G.R. *et al.* Identification of an Epstein-Barr virus early gene encoding a second component of the restricted early antigen complex. **Virology** 160: 151-161, 1987.

PILERI, S. A. *et al.* Hodgkin's lymphoma: the pathologist's viewpoint. **J Clin Pathol**, v. 55, n. 3, p. 162-76, Mar 2002.

PIZZO P.A. *et al.* A new tumour-derived transforming strain of Epstein-Barr virus. **Nature**, 272:629-631, 1978.

PLUMMER, M. *et al.* Global burden of cancers attributable to infections in 2012: a synthetic analysis. **Lancet Glob Health**, v.4: e609–16, 2016

POPE, J. H.; HOME, M. K.; SCOTT, W. Transformation of foetal human leukocytes *in vitro* by filtrates of a human leukaemic cell line containing herpes-like virus **Int. J. Cancer**, v. 3, 857, 1968.

QUEIROGA, EM. *et al.* Burkitt lymphoma in Brazil is characterized by geographically distinct clinicopathologic features. **Am J Clin Pathol.** 2008

- RAAB-TRAUB, N. Epstein-Barr virus in the pathogenesis of NPC. **Semin Cancer Biol.** 2002
- RICKINSON, A. B., L. S. YOUNG, AND M. ROWE. Influence of the Epstein-Barr virus nuclear antigen EBNA 2 on the growth phenotype of virus-transformed B cells. **J. Virol.** 61:1310-1317, 1987.
- RICKINSON, A. Epstein-Barr virus. **Virus Res**, v. 82, n. 1-2, p. 109-13, 2002.
- ROTHENBERGER, S.; ZUERCHER, E. Epstein-Barr Virus Latent Membrane Protein 1: Genetic Diversity, Classification, and Implication for NF- $\kappa$ B Activation. **Advances in Virus Research.** PRESS, I. 2012.
- ROUS. P. 1911. A sarcoma of the fowl transmissible by an agent separable from the tumor cells. **J. Exp. Med.** 13:397–411.
- SAITO, N. *et al.* Two carboxyl-terminal activation regions of Epstein-Barr virus latent membrane protein 1 activate NF-kappaB through distinct signaling pathways in fibroblast cell lines. **J Biol Chem**, 278(47):46565–75, 2003.
- SAMBROOK, J.; RUSSELL, D. W. Molecular Cloning: A Laboratory Manual. **Cold Spring Harbor Laboratory Press**, 2001.
- SAMPLE, J. *et al.* Epstein-Barr virus types 1 and 2 differ in their EBNA-3A, EBNA-3B, and EBNA-3C genes. **J Virol**, 64(9):4084–92, 1990.
- SANDVEJ, K. *et al.* Analysis of the Epstein-Barr virus (EBV) latent membrane protein 1 (LMP-1) gene and promoter in Hodgkin's disease isolates: selection against EBV variants with mutations in the LMP-1 promoter ATF-1/CREB-1 binding site. **Mol Pathol**, v. 53, n. 5, p. 280-8, 2000.
- SANDVEJ, K. *et al.* Sequence analysis of the Epstein-Barr virus (EBV) latent membrane protein-1 gene and promoter region: identification of four variants among wild-type EBV isolates. **Blood**, v. 90, n. 1, p. 323-30, 1997.
- SEGERSTROM, AC. *et al.* Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. **Psychol Bull.** Jul;130(4):601-30. 2004,
- SIMBIRI, K. O. *et al.* Epstein–Barr Virus Genetic Variation in Lymphoblastoid Cell Lines Derived from Kenyan Pediatric Population. **PLoS ONE** 10(5): e0125420, 2015.
- SCHNEIDER, J. *et al.* The viral oncoprotein LMP1 exploits TRADD for signaling by masking its apoptotic activity. **PLoS Biol**, 6(1):e8, 2008.
- SANTOS, YL. História da África e do Brasil afrodescendente. 1. ED. RIO DE JANEIRO: PALLAS, 2017.
- SOUZA, T. A. *et al.* Peripheral B cells latently infected with Epstein-Barr virus display molecular hallmarks of classical antigen-selected memory B cells. **Proc Natl Acad Sci U S A**, v. 102, n. 50, p. 18093-8, Dec 13 2005.

STEFANOFF, C.G. *et al.* Laboratory strategies for efficient handling of paraffin-embedded tissues for molecular detection of clonality in non-Hodgkin lymphomas. **Diagn Mol Pathol**, 12, 79-87, 2003.

SUEUR, C. *et al.* Difference in cytokine production and cell cycle progression induced by Epstein-Barr virus Lmp1 deletion variants in Kmh2, a Hodgkin lymphoma cell line. **Virology Journal** 11:94, 2014.

SUN L. *et al.* Sequence analysis of Epstein-Barr virus (EBV) early genes BARP1 and BHRF1 in NK/T cell lymphoma from Northern China. **Viol J.** 12:135, 2015.

SWERDLOW SH. International Agency for Research on Cancer., World Health Organization. WHO classification of tumours of haematopoietic and lymphoid tissues. Lyon, France: **International Agency for Research on Cancer**, 2016.

TAKADA K. Epstein-Barr virus and gastric carcinoma. **Mol Pathol** Oct; 53(5): 255–261. 2000.

THACKER, E. L.; MIRZAEI, F.; ASCHERIO, A. Infectious mononucleosis and risk for multiple sclerosis: a meta-analysis. **Ann Neurol**, v. 59, n. 3, p. 499-503, 2006.

THOMPSON, M. P.; KURZROCK, R. Epstein-Barr virus and cancer. **Clin Cancer Res**, v. 10, n. 3, p. 803-21, 2004.

THORNBURG, NJ and RAAB-TRAUB, N. Induction of epidermal growth factor receptor expression by Epstein-Barr virus latent membrane protein 1 C-terminal-activating region 1 is mediated by NF-kappaB p50 homodimer/Bcl-3 complexes. **J Virol**, 81(23):12954–61,2007.

THORLEY-LAWSON, D. A. Epstein-Barr virus: exploiting the immune system. **Nat Rev Immunol**, v. 1, n. 1, p. 75-82, 2001.

THORLEY-LAWSON, D. A.; GROSS, A. Persistence of the Epstein-Barr virus and the origins of associated lymphomas. **N Engl J Med**, v. 350, n. 13, p. 1328- 37, 2004.

THUAN, L. D. *et al.* Novel Patterns of the Epstein-Barr Nuclear Antigen (EBNA-1) V-Val Subtype in EBV-associated Nasopharyngeal Carcinoma from Vietnam Balkan **J Med Genet.** 28;22(1):61-68, 2019.

TIERNEY, R. J. *et al.* Epstein-Barr virus latency in blood mononuclear cells: analysis of viral gene transcription during primary infection and in the carrier state. **J Virol**, v. 68, n. 11, p. 7374-85, Nov 1994.

TOMKINSON, B. *et al.* Second-site homologous recombination in Epstein-Barr virus: insertion of type 1 EBNA 3 genes in place of type 2 has no effect on in vitro infection. **J Virol**, 66(2):780–9, 1992.

TONG, JHM. *Et al.* Discrete Alterations in the BZLF1 Promoter in Tumor and Non-Tumor-Associated Epstein–Barr Virus. **Journal of the National Cancer Institute**, Volume 95, Issue 13, 2 July 2003

TRIVEDI, P. *et al.* Epstein-Barr virus (EBV)-encoded membrane protein LMP1 from a nasopharyngeal carcinoma is non-immunogenic in a murine model system, in contrast to a B cell-derived homologue. **Eur J Cancer**, v. 30A, n. 1, p. 84-8, 1994.

TURNER, D.M. *et al.* An investigation of polymorphism in the interleukin-10 gene promoter. **Eur J Immunogenet**, v. 24, p. 1-8, 1997.

TZELLOS, S.; FARRELL, P.J. Epstein-barr virus sequence variation-biology and disease. **Pathogens**, 1(2):156–74, 2012.

TZELLOS, S. *et al.* A single amino acid in EBNA-2 determines superior B lymphoblastoid cell line growth maintenance by Epstein-Barr virus type 1 EBNA-2. **J Virol** Aug;88(16):8743-53. 2014

UCHIDA, J. *et al.* Mimicry of CD40 signals by Epstein-Barr virus LMP1 in B lymphocyte responses. **Science**, v. 286, n. 5438, p. 300-3, 1999.

VETSIKA, E. K.; CALLAN, M. Infectious mononucleosis and Epstein-Barr virus. **Expert Rev Mol Med**, v. 6, n. 23, p. 1-16, 2004.

WALLING, D. M. *et al.* The molecular epidemiology and evolution of Epstein Barr virus: sequence variation and genetic recombination in the latent membrane protein -1 gene. **J Infect Dis**, v. 179, n. 4, p. 763-74, 1999.

WANG, D.; LIEBOWITZ, D.; KIEFF, E. An EBV membrane protein expressed in immortalized lymphocytes transforms established rodent cells. **Cell**, v. 43, n. 3 Pt 2, p. 831-40, 1985.

WANG, F. *et al.* Simian homologues of Epstein-Barr virus. **Philos Trans R Soc Lond B Biol Sci**, v. 356, n. 1408, p. 489-97, 2001.

YAO, QY. *et al.* A re-examination of the Epstein-Barr virus carrier state in healthy seropositive individuals. **Int J Cancer**. 1985

YOUNG, L. S.; RICKINSON, A. B. Epstein-Barr virus: 40 years on. **Nat Rev Cancer**, v. 4, n. 10, p. 757-68, Oct 2004.

YOUNG, LS. *Et al.* Epstein-Barr virus and nasopharyngeal carcinoma. **Chin J Cancer**. 2014

ZANELLA, L. *et al.* A reliable Epstein-Barr Virus classification based on phylogenomic and population analyses. **Scien Rep** 9:9829. 2019.

ZENG, M. S. *et al.* Genomic sequence analysis of Epstein–Barr virus strain GD1 from a nasopharyngeal carcinoma patient. **J. Virol.**, 79 (2005), pp. 15323-15330

ZUERCHER E. *et al.* Genetic diversity of EBV encoded LMP1 in the Swiss HIV Cohort Study and implication for NF-Kappab activation. **PLoS One**, v.7: e32168, 2012.

ZUR HAUSEN, H. Papillomaviruses in the causation of human cancers - a brief historical account. **Virology**, 384 (2) (2009), pp. 260-265

## ANEXO I. TABELAS SUPLEMENTARES

Tabela S1: Características sociodemográficas dos carreadores assintomáticos

Variável	N (%)
<b>Total</b>	<b>524 (100)</b>
<b>Criança x Adulto</b>	
≤18 anos	33 (6,3)
≥19 anos	491 (93,7)
<b>Sexo</b>	
Feminino	289 (55,1)
Masculino	235 (44,8)
<b>Cor autodeclarada</b>	
Branca	285 (54,4)
Não branca	234 (44,7)
Não declarado	5 (1,0)
<b>Residiu com Pai até 18 anos</b>	
Sim	382 (72,9)
Não	127 (24,2)
Não declarado	15 (2,9)
<b>Residiu com Mãe até 18 anos</b>	
Sim	476 (90,8)
Não	33 (6,3)
Não declarado	15 (2,9)
<b>Residiu com Avó até 18 anos</b>	
Sim	128 (24,4)
Não	381 (72,7)
Não declarado	15 (2,9)
<b>Estado civil</b>	
Solteiro(a)	310 (59,2)
Mora junto/casado(a)/viúvo(a)	212 (40,5)
Não declarado	2 (0,3)
<b>Possui filhos</b>	
Sim	159 (30,3)
Não	362 (69,5)
Não declarado	3 (0,6)
<b>Ato de beijar filho(s) e/ou pais na boca</b>	
Sim	72 (13,7)
Não	438 (83,6)
Não declarado	14 (2,7)
<b>Origem dos ancestrais Maternos autodeclarados</b>	
Africano	46 (8,8)
Europeu	162 (30,9)
Ameríndio	58 (11,1)
Outro(s)	3 (0,6)
Diversos	77 (14,7)
Não sei	155 (29,6)

Não declarado	23 (4,4)
<b>Origem dos ancestrais Paternos autodeclarados</b>	
Africano	44 (8,4)
Europeu	163 (31,1)
Ameríndio	49 (9,4)
Outro(s)	5 (1,0)
Diversos	54 (10,3)
Não sei	171 (32,6)
Não declarado	38 (7,3)
<b>Brasileiro x Estrangeiro</b>	
Sim	496 (94,7)
Não	28 (5,3)
<b>Origem até 18 anos - Continente</b>	
América Latina	509 (97,1)
África	15 (2,9)
<b>País até 18 anos</b>	
Brasil	495 (94,5)
Angola	14 (2,7)
Guiné-Bissau	1 (0,2)
Argentina	3 (0,6)
Venezuela	5 (1,0)
Uruguai	1 (0,2)
Colômbia	1 (0,2)
Cuba	1 (0,2)
Peru	2 (0,4)
Diversos	1 (0,2)
<b>Origem até 18 anos - Região brasileira</b>	
Norte	10 (1,9)
Nordeste	27 (5,2)
Centro-oeste	13 (2,5)
Sudeste	423 (80,7)
Sul	17 (3,2)
Diversas	5 (1,0)
Não se aplica	29 (5,5)
<b>Origem até 18 anos - Estado do Rio de Janeiro</b>	
Sim	373 (71,2)
Não	134 (25,6)
>= 2 estados (RJ incluso)	17 (3,2)
<b>Origem até 18 anos - Cidade do Rio de Janeiro</b>	
Sim	265 (50,6)
Não	247 (47,1)
>= 2 cidades (RJ incluso)	12 (2,3)
<b>Reside atualmente - Região brasileira</b>	
Norte	5 (1,0)
Nordeste	11 (2,1)
Centro-oeste	5 (1,0)
Sudeste	497 (94,8)
Sul	6 (1,1)



<b>Reside atualmente - Estado do Rio de Janeiro</b>	
Sim	489 (93,3)
Não	33 (6,3)
Não declarado	2 (0,4)
<b>Reside atualmente - Cidade do Rio de Janeiro</b>	
Sim	369 (70,4)
Não	153 (29,2)
Não declarado	2 (0,4)
<b>Escolaridade</b>	
Não se aplica =< 5 anos	2 (0,4)
Baixa (= < 1º grau completo)	43 (8,2)
Média e alta (>1º grau completo)	474 (90,5)
Não declarado	5 (1,0)
<b>Escolaridade materna</b>	
Baixa (= < 1º grau completo)	128 (24,4)
Média e alta (>1º grau completo)	371 (70,8)
Não declarado ou não sabe	25 (4,8)
<b>Escolaridade paterna</b>	
Baixa (= < 1º grau completo)	143 (27,3)
Média e alta (>1º grau completo)	336 (64,1)
Não declarado ou não sabe	45 (8,6)
<b>Escolaridade responsável (exceto pai e mãe)</b>	
Baixa (= < 1º grau completo)	31 (45,6)
Média e alta (>1º grau completo)	37 (54,4)
Não se aplica	456 (87,0)
<b>Renda familiar</b>	
Até 1 salário mínimo (até R\$ 954,00)	28 (5,3)
De 1 a 2 salários mínimos	66 (12,6)
De 2 a 5 salários mínimos	159 (30,3)
De 5 a 10 salários mínimos	155 (29,6)
De 10 a 30 salários mínimos	102 (19,5)
Acima de 30 salários mínimos	8 (1,5)
Não declarado	6 (1,1)
<b>Número de pessoas na residência atual (doador incluso)</b>	
Moro sozinho(a)	46 (8,8)
Duas	162 (30,9)
Três	143 (27,3)
Quatro	118 (22,5)
Cinco	31 (5,9)
Seis ou mais	21 (4,0)
Não declarado	3 (0,6)
<b>Diagnóstico de mononucleose infecciosa</b>	
Sim	20 (3,8)
Não	309 (59,0)
Não declarado ou não sabe	195 (37,2)
<b>Cirurgia de retirada das amígdalas</b>	
Sim	44 (8,4)

Não	470 (89,7)
Não declarado ou não sabe	10 (1,9)
<b>Câncer prévio</b>	
Sim	17 (3,2)
Não	506 (96,6)
Não declarado	1 (0,2)
<b>Doença crônica</b>	
Sim	90 (17,2)
Não	401 (76,5)
Não declarado	33 (6,3)
<b>Doença autoimune</b>	
Sim	11 (2,1)
Não	513 (97,9)
<b>Relacionamentos afetivos &gt;2 meses</b>	
Até 3	255 (48,7)
Entre 3 e 10	184 (35,1)
Acima de 11	36 (6,9)
Nenhum	40 (7,6)
Não declarado	9 (1,7)

Tabela S2: Carga viral dos casos de patologias hematológicas EBV+

Patologia	EBV/1x10 <sup>6</sup> de células MN		
	mediana	min-máx	N
HR	4,55E+04	–	1/3
PTLD	4,25E+08	4,25E+06	1/1
LB	5,57E+08	1,64E+08 - 8,95E+10	3/7
LHc	3,26E+05	1,84E+03 - 1,87E+07	26/26

EBV: vírus Epstein-Bar; MN: mononuclear; MNI: mononucleose infecciosa; HR: hiperplasia reativa; L. T/NK: linfoma de células T/K; DLBCL: linfoma difuso de grandes células B (do inglês: *diffuse large B-cell lymphoma*). LB: linfoma de Burkitt; LHc: linfoma de Hodgkin clássico.

Tabela S3: Associação de fatores sociodemográficos com a detecção do EBV na 1ª amostra coletada da população sadia.

Variável	Status EBV (1ª amostra)		OR	IC95%	p- value
	Positivo	Negativo			
	N (%)	N (%)			
<b>Total</b>	240 (45,8)	284 (54,2)			
<b>Criança x Adulto</b>					
≤18 anos	13 (39,4)	20 (60,6)	0,756	0,368- 1,554	0,281

≥19 anos	227 (46,2)	264 (53,8)			
<b>Sexo</b>					
Feminino	131 (45,3)	158 (54,7)	1,043	0,739- 1,474	0,439
Masculino	109 (46,4)	126 (53,6)			
<b>Cor autodeclarada</b>					
Branca	126 (44,2)	159 (55,8)	0,878	0,621- 1,243	0,259
Não branca	111 (47,4)	123 (52,6)			
<b>Residiu com Pai até 18 anos</b>					
Sim	172 (45,0)	210 (55,0)	0,974	0,651- 1,458	0,490
Não	58 (45,7)	69 (54,3)			
<b>Residiu com Mãe até 18 anos</b>					
Sim	213 (44,7)	263 (55,3)	0,762	0,76- 1,545	0,282
Não	17 (51,5)	16 (48,5)			
<b>Residiu com Avó até 18 anos</b>					
Sim	64 (50,0)	64 (50,0)	1,295	0,867- 1,935	0,123
Não	166 (43,6)	215 (56,4)			
<b>Estado civil</b>					
Solteiro(a)	137 (44,2)	173 (55,8)	0,854	0,601- 1,212	0,214
Mora junto com companheiro(a) /casado(a)/divorciado(a)/viúvo(a)	102 (48,1)	110 (51,8)			
<b>Possui filhos</b>					
Sim	76 (47,8)	83 (52,2)	1,130	0,778- 1,643	0,292
Não	162 (44,8)	200 (55,2)			
<b>Ato de beijar filho(s) e/ou pais na boca</b>					
Sim	32 (44,4)	40 (55,6)	0,943	0,571- 1,557	0,461
Não	201 (45,9)	237 (54,1)			
<b>Ancestrais Maternos Africanos x Europeus</b>					
Africanos	27 (58,7)	19 (41,3)	2,067	1,063- 4,021	0,023
Europeus	66 (40,7)	96 (59,3)			
<b>Ancestrais Maternos Ameríndios x Europeus</b>					

Ameríndios	33 (56,9)	25 (43,1)	1,920	1,047- 3,523	0,025
Europeus	66 (40,7)	96 (59,3)			
<b>Ancestrais Paternos Africanos x Europeus</b>					
Africanos	26 (59,1)	18 (40,9)	1,737	0,884- 3,414	0,074
Europeus	74 (45,4)	89 (54,9)			
<b>Ancestrais Paternos Ameríndios x Europeus</b>					
Ameríndios	19 (38,8)	30 (61,2)	0,762	0,397- 1,462	0,257
Europeus	74 (45,4)	89 (54,6)			
<b>Brasileiro x Estrangeiro</b>					
BR	221 (44,6)	275 (55,4)	0,381	0,169- 858	0,013
Estrangeiro	19 (67,9)	9 (32,1)			
<b>Origem até 18 anos - Continente</b>					
América Latina	227 (44,6)	282 (55,4)	0,124	0,028- 0,554	0,001
África	13 (86,7)	2 (13,3)			
<b>Origem até 18 anos - País</b>					
Brasil	221 (44,6)	274 (55,4)			0,084
Angola	12 (85,7)	2 (14,3)			
Guiné-Bissau	1 (100,0)	0 (0,0)			
Argentina	2 (66,7)	1 (33,3)			
Venezuela	3 (60,0)	2 (40,0)			
Uruguai	0 (0,0)	1 (100,0)			
Colômbia	0 (0,0)	1 (100,0)			
Cuba	0 (0,0)	1 (100,0)			
Peru	1 (50,0)	1 (50,0)			
<b>Origem até 18 anos Brasil x Angola</b>					
Brasil	221 (44,6)	274 (55,4)	0,134	0,030- 0,607	0,002
Angola	12 (85,7)	2 (14,3)			
<b>Origem até 18 anos - Região brasileira</b>					
Norte	6 (60,0)	4 (40,0)			0,426

Nordeste	15 (55,6)	12 (44,4)			
Centro-oeste	4 (30,8)	9 (69,2)			
Sudeste	189 (44,7)	234 (55,3)			
Sul	6 (35,3)	11 (64,7)			
<b>Origem até 18 anos</b>					
<b>Região Nordeste x Sudeste (Brasil)</b>					
Região nordeste	15 (55,6)	12 (44,4)	1,584	0,707- 3,386	0,184
Região sudeste	189 (44,7)	234 (55,3)			
<b>Origem até 18 anos</b>					
<b>Estado do Rio de Janeiro</b>					
Sim	169 (45,3)	204 (54,7)	0,934	0,628- 1,387	0,405
Não	63 (47,0)	71 (53,0)			
<b>Origem até 18 anos</b>					
<b>Cidade do Rio de Janeiro</b>					
Sim	109 (41,1)	156 (58,9)	0,719	0,505- 1,015	0,037
Não	122 (49,4)	125 (50,6)			
<b>Residência atual</b>					
<b>Região brasileira</b>					
Norte	2 (40,0)	3 (60,0)			0,756
Nordeste	4 (36,4)	7 (63,6)			
Centro-oeste	1 (20,0)	4 (80,0)			
Sudeste	230 (46,3)	267 (53,7)			
Sul	3 (50,0)	3 (50,0)			
<b>Residência atual</b>					
<b>Região Nordeste x Sudeste (Brasil)</b>					
Região nordeste	4 (36,4)	7 (63,6)	0,663	0,192- 2,295	0,368
Região sudeste	230 (46,3)	267 (53,7)			
<b>Residência</b>					
<b>Estado do Rio de Janeiro</b>					
Sim	229 (46,8)	260 (53,2)	1,762	0,836- 3,711	0,920
Não	11 (33,3)	22 (66,7)			
<b>Residência atual</b>					
<b>Cidade do Rio de Janeiro</b>					
Sim	171 (46,3)	198 (53,7)	1,051	0,720- 1,535	0,436

Não	69 (45,1)	84 (54,9)			
<b>Escolaridade (só adultos)</b>					
Baixa (= < 1º grau completo)	9 (33,3)	18 (66,7)	0,543	0,280- 1,054	0,116
Média e alta (>= 2º grau completo)	223 (47,0)	251 (53,0)			
<b>Escolaridade materna</b>					
Baixa (= < 1º grau completo)	64 (50,0)	64 (50,0)	1,290	0,862- 1,930	0,128
Média e alta (>= 2º grau completo)	162 (43,7)	209 (56,3)			
<b>Escolaridade paterna</b>					
Baixa (= < 1º grau completo)	74 (51,7)	69 (48,3)	1,430	0,965- 2,118	0,046
Média e alta (>= 2º grau completo)	144 (42,9)	192 (57,1)			
<b>Escolaridade responsável (exceto pai e mãe)</b>					
Baixa (= < 1º grau completo)	16 (51,6)	15 (48,4)	1,126	0,433- 2,926	0,500
Média e alta (>= 2º grau completo)	18 (48,6)	19 (51,4)			
<b>Renda familiar</b>					
Até 1 salário mínimo (até R\$ 954,00)	12 (42,9)	16 (57,1)			0,795
De 1 a 2 salários mínimos	29 (43,9)	37 (56,1)			
De 2 a 5 salários mínimos	76 (47,8)	83 (52,2)			
De 5 a 10 salários mínimos	66 (42,6)	89 (57,4)			
De 10 a 30 salários mínimos	48 (47,1)	54 (52,9)			
Acima de 30 salários mínimos	5 (62,5)	3 (37,5)			
<b>Nº de pessoas na residência atual (doador incluso)</b>					
Moro sozinho (a)	25 (54,3)	21 (45,7)			0,315
Duas	79 (48,8)	83 (51,2)			
Três	64 (44,8)	79 (55,2)			
Quatro	46 (39,0)	72 (61,0)			
Cinco	12 (38,7)	19 (61,3)			
Seis ou mais	12 (57,1)	9 (42,9)			
<b>Diagnóstico de mononucleose infecciosa</b>					

Sim	6 (30,0)	14 (70,0)	0,485	0,182-1,294	0,106
Não	145 (46,9)	164 (53,1)			
<b>Cirurgia de retirada das amígdalas</b>					
Sim	21 (47,7)	23 (52,3)	1,130	0,609-2,099	0,408
Não	210 (44,7)	260 (55,3)			
<b>Câncer prévio</b>					
Sim	6 (35,3)	11 (64,7)	0,639	0,233-1,755	0,267
Não	233 (46,0)	273 (54,0)			
<b>Doença crônica</b>					
Sim	41 (45,6)	49 (54,4)	0,987	0,624-1,562	0,525
Não	184 (45,9)	217 (54,1)			
<b>Doença autoimune</b>					
Sim	6 (54,5)	5 (45,5)	1,431	0,431-4,748	0,387
Não	234 (45,6)	279 (54,4)			
<b>Relacionamentos afetivos &gt;2 meses</b>					
Até 3	118 (46,3)	137 (53,7)			0,230
Entre 3 e 10	86 (46,7)	98 (53,3)			
Acima de 11	20 (55,6)	16 (44,4)			
Nenhum	13 (32,5)	27 (67,5)			

Tabela S4: Associação de fatores sociodemográficos com o tipo 1 ou 2 do EBV na população sadia.

Variável	Tipo do EBV (1ª amostra)		OR	IC95%	p-value
	Tipo 1 N (%)	Tipo 2 N (%)			
<b>Total</b>	183 (79,2)	48 (20,8)			
<b>Criança x Adulto</b>					
≤18 anos	7 (53,8)	6 (46,2)	0,278	0,089-0,872	0,032
≥19 anos	176 (80,7)	42 (19,3)			
<b>Sexo</b>					

Masculino	87 (82,9)	18 (17,1)	1,510	0,787- 2,900	0,140
Feminino	96 (76,2)	30 (23,8)			
<b>Cor autodeclarada</b>					
Branca	99 (81,1)	23 (18,9)	1,260	0,663- 2,395	0,294
Não branca	82 (77,4)	24 (22,6)			
<b>Estado civil</b>					
Solteiro(a)	105 (80,2)	26 (19,8)	1,154	0,609- 2,187	0,390
mora junto/casado(a)/viúvo(a)	77 (77,8)	22 (22,2)			
<b>Possui filhos</b>					
Sim	56 (77,8)	16 (22,2)	0,896	0,455- 1,765	0,438
Não	125 (79,6)	32 (20,4)			
<b>Ato de beijar filho(s) e/ou pais na boca</b>					
Sim	26 (83,9)	5 (16,1)	1,446	0,523- 3,996	0,327
Não	151 (78,2)	42 (21,8)			
<b>Ancestrais Maternos Africanos x Europeus</b>					
Africanos	19 (82,6)	4 (17,4)	0,968	0,275- 3,405	0,594
Europeus	54 (83,1)	11 (16,9)			
<b>Ancestrais Maternos Ameríndios x Europeus</b>					
Ameríndios	22 (68,8)	10 (31,2)	0,448	0,167- 1,205	0,090
Europeus	54 (83,1)	11 (16,9)			
<b>Ancestrais Paternos Africanos x Europeus</b>					
Africanos	18 (78,3)	5 (21,7)	0,964	0,307- 3,024	0,580
Europeus	56 (78,9)	15 (21,1)			
<b>Ancestrais Paternos Ameríndios x Europeus</b>					
Ameríndios	14 (77,8)	4 (22,2)	0,938	0,269- 3,268	0,572
Europeus	56 (78,9)	15 (21,1)			
<b>Brasileiro x Estrangeiro</b>					
BR	170 (79,4)	44 (20,6)	1,189	0,369- 3,825	0,487
Estrangeiro	13 (76,5)	4 (23,5)			
<b>Origem até 18 anos - Continente</b>					
América Latina	174 (79,1)	46 (20,9)	0,841	0,176- 4,025	0,591



África	9 (81,8)	2 (18,2)			
<b>Origem até 18 anos - País</b>					
Brasil	170 (79,4)	44 (20,6)			0,429
Angola	8 (80,0)	2 (20,0)			
Guiné-Bissau	1 (100,0)	0 (0,0)			
Argentina	2 (100,0)	0 (0,0)			
Venezuela	2 (66,7)	1 (33,3)			
Peru	0 (0,0)	1 (100,0)			
<b>Origem até 18 anos - Região brasileira</b>					
Norte	4 (66,7)	2 (33,3)			0,521
Nordeste	11 (78,6)	3 (21,4)			
Centro-oeste	4 (100,0)	0 (0,0)			
Sudeste	145 (79,2)	38 (20,8)			
Sul	6 (100,0)	0 (0,0)			
<b>Origem até 18 anos Estado do Rio de Janeiro</b>					
Sim	130 (79,8)	46 (76,7)	1,199	0,590-2,438	0,370
Não	33 (20,2)	14 (23,3)			
<b>Origem até 18 anos Cidade do Rio de Janeiro</b>					
Sim	88 (82,2)	19 (17,8)	1,491	0,775-2,866	0,150
Não	87 (75,7)	28 (24,3)			
<b>Residência atual - Região brasileira</b>					
Norte	1 (50,0)	1 (50,0)			0,667
Nordeste	2 (66,7)	1 (33,3)			
Centro-oeste	1 (100,0)	0 (0,0)			
Sudeste	176 (79,3)	46 (20,7)			
Sul	3 (100,0)	0 (0,0)			
<b>Residência atual Estado do Rio de Janeiro</b>					
Sim	175 (79,2)	46 (20,8)	0,951	0,195-4,632	0,655
Não	8 (80,0)	2 (20,0)			
<b>Residência atual Cidade do Rio de Janeiro</b>					
Sim	134 (81,2)	31 (18,8)	1,500	0,763-2,948	0,159
Não	49 (74,2)	17 (25,8)			
<b>Escolaridade (só adulta)</b>					
Baixa (= < 1º grau completo)	6 (75,0)	3 (25,0)	0,754	0,147-3,869	0,507

Média e alta (>= 2º grau completo)	171 (79,9)	43 (20,1)			
<b>Escolaridade materna</b>					
Baixa (= < 1º grau completo)	48 (77,4)	14 (22,6)	0,885	0,435-1,801	0,433
Média e alta (>= 2º grau completo)	124 (79,5)	32 (20,5)			
<b>Escolaridade paterna</b>					
Baixa (= < 1º grau completo)	58 (81,7)	13 (18,3)	1,228	0,596-2,534	0,358
Média e alta (>= 2º grau completo)	109 (78,4)	30 (21,6)			
<b>Escolaridade responsável (exceto pai e mãe)</b>					
Baixa (= < 1º grau completo)	13 (81,2)	3 (18,8)	0,578	0,083-4,009	0,470
Média e alta (>= 2º grau completo)	15 (88,2)	2 (11,8)			
<b>Renda familiar</b>					
Até 1 salário mínimo (até R\$ 954,00)	9 (81,8)	2 (18,2)			0,719
De 1 a 2 salários mínimos	22 (78,6)	6 (21,4)			
De 2 a 5 salários mínimos	62 (86,1)	10 (13,9)			
De 5 a 10 salários mínimos	48 (76,2)	15 (23,8)			
De 10 a 30 salários mínimos	35 (72,9)	13 (27,1)			
Acima de 30 salários mínimos	4 (80,0)	1 (20,0)			
<b>Número de pessoas na residência atual (doador incluso)</b>					
Moro sozinho (a)	18 (78,3)	5 (21,7)			0,469
Duas	60 (78,9)	16 (21,1)			
Três	50 (79,4)	13 (20,6)			
Quatro	31 (72,1)	12 (27,9)			
Cinco	10 (83,3)	2 (16,7)			
Seis ou mais	12 (100,0)	0 (0,0)			
<b>Diagnóstico de mononucleose infecciosa</b>					
Sim	6 (100,0)	0 (0,0)	0,944	0,902-0,989	0,187
Não	102 (75,0)	34 (25,0)			
<b>Cirurgia de retirada das amígdalas</b>					
Sim	16 (80,0)	4 (20,0)	1,114	0,354-3,502	0,557
Não	158 (78,2)	44 (21,8)			
<b>Câncer prévio</b>					

Sim	6 (100,0)	0 (0,0)	0,967	0,941- 0,993	0,241
Não	176 (78,6)	48 (21,4)			
<b>Doença crônica</b>					
Sim	28 (70,0)	12 (30,0)	0,535	0,246- 1,160	0,086
Não	144 (81,4)	33 (18,6)			
<b>Doença autoimune</b>					
Sim	3 (50,0)	3 (50,0)	0,250	0,049- 1,280	0,106
Não	180 (80,0)	45 (20,0)			
<b>Relacionamentos afetivos &gt;2 meses</b>					
Até 3	86 (76,8)	26 (23,2)			0,361
Entre 3 e 10	66 (78,6)	18 (21,4)			
Acima de 11	18 (94,7)	1 (5,3)			
Nenhum	10 (76,9)	3 (23,1)			

Nota: Os cinco casos de coinfeção pelo tipo 1 e 2 foram excluídos das análises estatísticas acima.

Tabela S5 - Clados do gene *LMP1* definidos pela filogenia de Máxima verossimilhança.

<b>Clado</b>	<b>N (%)</b>	<b>Apoio (aLRT)</b>
Med/AG876/Daudi	33 (40)	0.95
Raji/Argentine	32 (39)	0.92
B95-8	5 (6)	0.96
China1-Cao	5 (6)	0.96
China2	1 (1)	0.81
Não classificadas	7 (8)	NA

NA: não se aplica; N: número; aLRT: teste da razão de verossimilhança média

Tabela S6 - Relação entre os polimorfismos e as variantes da oncoproteína LMP1

Polimorfismos da LMP1	N (%)				<i>p</i> - valor
	Med/AG876/Daudi	Raji/Argentine	B95-8	Asiática	
<b>Polimorfismo do N-terminal</b>					
Sítio de restrição da <i>Xho</i> I (c↓tcgag)	33 (100)	32 (100)	5 (100)	5 (83,3)	<i>p</i> = 0,009
Ausência do sítio de restrição da <i>Xho</i> I	0 (0,0)	0 (0,0)	0 (0,0)	1 (16,7)	
<b>Polimorfismos da Transmembrana</b>					
Posição 169055 wt	33 (100)	32 (100)	5 (100)	1 (16,7)	<i>p</i> <0,0001
Posição 169055 C>T	0 (0,0)	0 (0,0)	0 (0,0)	5 (83,3)	
wt/wt	33 (100)	2 (6,2)	5 (100)	6 (100)	<i>p</i> <0,0001
I124V/I152L	0 (0,0)	30 (93,8)	0 (0,0)	0 (0,0)	
wt/wt/wt	21 (63,6)	0 (0,0)	5 (100)	0 (0,0)	<i>p</i> <0,0001
wt/wt/L151I	12 (36,4)	32 (100)	0 (0,0)	0 (0,0)	
F144I/wt/L151I	0 (0,0)	0 (0,0)	0 (0,0)	1 (16,7)	
F144I/D150A/L151I	0 (0,0)	0 (0,0)	0 (0,0)	5 (83,3)	
<b>Polimorfismos do C-terminal</b>					
N° de repetição 33 pb					
x1	1 (3,0)	0 (0,0)	0 (0,0)	0 (0,0)	<i>p</i> = 0,061
x2	0 (0,0)	0 (0,0)	0 (0,0)	1 (16,7)	
x3	4 (12,1)	0 (0,0)	0 (0,0)	0 (0,0)	
x4	16 (48,5)	9 (28,1)	3 (60,0)	1 (16,7)	
x5	7 (21,2)	19 (59,4)	2 (40,0)	4 (66,7)	
x6	3 (9,1)	4 (12,5)	0 (0,0)	0 (0,0)	
x7	1 (3,0)	0 (0,0)	0 (0,0)	0 (0,0)	
x8	1 (3,0)	0 (0,0)	0 (0,0)	0 (0,0)	
wt	29 (87,9)	4 (12,5)	0 (0,0)	1 (16,7)	<i>p</i> <0,0001
Ins15 pb	4 (12,1)	28 (87,5)	5 (100)	5 (83,3)	
	24 (72,7)	9 (28,1)	0 (0,0)	5 (83,3)	<u><i>p</i> &lt;0,0001</u>

wt				
del30	6 (18,2)	23 (71,9)	5 (100)	1 (16,7)
del69	3 (9,1)	0 (0,0)	0 (0,0)	0 (0,0)

---

LMP1: proteína latente de membrana 1 (do inglês: *latent membrane protein 1*); DA: doador assintomático; CB: casos benignos; CM: casos malignos; *Xho*I: endonuclease do tipo 2, de *Xanthomonas holcicola*; wt: tipo selvagem (do inglês: *wildtype*); C: citosina; T: timina; I: isoleucina; V: valina; L: leucina; F: fenilalanina; D: ácido aspártico; A: alanina; rep: repetição; pb: pares de base; ins15: inserção de 15 pb; del30: deleção de 30 pb; del69: deleção de 69 pb;

Tabela S7 - Relação dos polimorfismos da oncoproteína LMP1 com as variantes Med/AG876/Daudi e Raji/Argentine

Polimorfismos da LMP1	N (%)		p - valor
	Med	Raji	
<b>Polimorfismo do N-terminal</b>			
Sítio de restrição da <i>Xho</i> I (c↓tctgag)	33 (100)	32 (100)	p*
Ausência do sítio de restrição da <i>Xho</i> I	0 (0,0)	0 (0,0)	
<b>Polimorfismos da Transmembrana</b>			
Posição 169055 wt	33 (100)	32 (100)	p*
Posição 169055 C>T	0 (0,0)	0 (0,0)	
wt/wt	33 (100)	2 (6,2)	p < 0,0001
I124V/I152L	0 (0,0)	30 (93,8)	
wt/wt/wt	21 (63,6)	0 (0,0)	p < 0,0001
wt/wt/L151I	12 (36,4)	32 (100)	
F144I/wt/L151I	0 (0,0)	0 (0,0)	
F144I/D150A/L151I	0 (0,0)	0 (0,0)	
<b>Polimorfismos do C-terminal</b>			
N° de repetição 33 pb			
x1	1 (3,0)	0 (0,0)	p = 0,023
x2	0 (0,0)	0 (0,0)	
x3	4 (12,1)	0 (0,0)	
x4	16 (48,5)	9 (28,1)	
x5	7 (21,2)	19 (59,4)	
x6	3 (9,1)	4 (12,5)	
x7	1 (3,0)	0 (0,0)	
x8	1 (3,0)	0 (0,0)	
wt	29 (87,9)	4 (12,5)	p < 0,0001
Ins15 pb	4 (12,1)	28 (87,5)	
wt	24 (72,7)	9 (28,1)	p < 0,0001
del30	6 (18,2)	23 (71,9)	
del69	3 (9,1)	0 (0,0)	

p\*: não há cálculo estatístico, visto que a variável é constante; LMP1: ; Med : Med/AG876/Daudi; Raji : Raji/Argentine;

Tabela S8 - Frequência das variantes do EBV pelo oncogene *LMP1* pelas classificações por Edwards et al. e filogenética.

Variável	N (%)			p-valor
	Total	DA+CB	CM	
<b>C-terminal (por Edwards et al, 1999)</b>				
Med	33 (39,8)	22 (44,9)	11 (32,4)	
Raji	32 (38,6)	16 (32,7)	16 (47,1)	
B95-8	3 (3,6)	0 (0,0)	3 (8,8)	
B95-8/Med	3 (3,7)	3 (6,1)	0 (0,0)	
B95-8/Raji	1 (1,2)	0 (0,0)	1 (2,9)	$p = 0,09$
China1	8 (9,6)	6 (12,2)	2 (5,9)	
Raji/China1	1 (1,2)	0 (0,0)	1 (2,9)	
Alaskan	1 (1,2)	1 (2,0)	0 (0,0)	
China2/Alaskan	1 (1,2)	1 (2,0)	0 (0,0)	
<b>Região codificante inteira (por filogenia)</b>				
Med/AG876/Daudi	33 (39,8)	25 (51,0)	8 (23,5)	
Raji/Argentina	32 (38,6)	16 (32,7)	16 (47,1)	
B95-8	5 (6,0)	2 (4,1)	3 (8,8)	$p = 0,07$
Asiática	6 (7,2)	4 (8,2)	2 (5,9)	
Não classificadas	7 (8,4)	2 (4,1)	5 (14,7)	

DA: doador assintomático; PE: patologia EBV+; LMP1: proteína latente de membrana 1 (do inglês: *latente membrane protein 1*);

Tabela S9 - Comparação das classificações da LMP1 por Edwards et al. e filogenética

Classificação (N)	Filogenia					N total	Resultado	
	Edwards et al, 1999	Med/AG876/Daudi	Raji/Argentine	B95-8	Asiática			Não classificada
Med		28	0	0	0	5	33	divergente
Raji		0	32	0	0	0	32	ok
B95-8		0	0	3	0	0	3	ok
China1		4	0	0	3	1	8	divergente
Alaskan		0	0	0	1	0	1	ok
B95-8/Med		1	0	2	0	0	3	divergente
B95-8/Raji		0	0	0	0	1	1	ok*
China2/Alaskan		0	0	0	1	0	1	ok

Em verde: classificações iguais pelos métodos de Edwards et al e filogenética (83%); Em vermelho claro e escuro: classificações divergentes pelos dois métodos (8,5 %); Em vermelho escuro: classificação China1 por Edwards et al é a maioria entre as divergentes; Em amarelo: não classificadas pelo método filogenético (8,5 %). N: número de amostragem; LMP1: ; ok\*: dita recombinante pela classificação por Edwards et al., e sugestiva de recombinante pela filogenia.



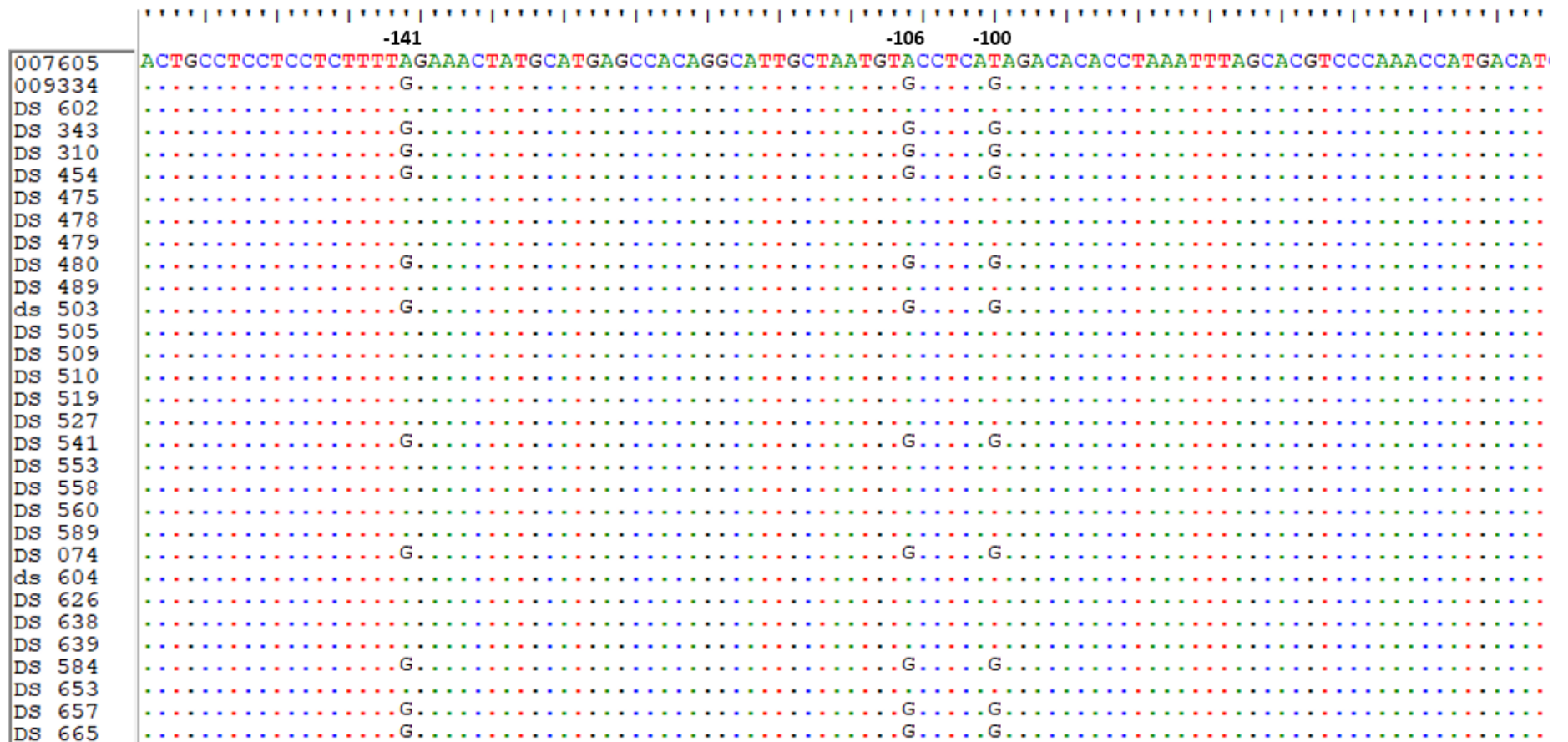


Figura S1 - **Representação da classificação dos polimorfismos do Zp.** A primeira sequência identificada como 007605 é o protótipo do EBV-1 (sequência selvagem) e a segunda identificada como 009334 é o protótipo do EBV-2, sequência referência para o polimorfismo V3 do Zp nas posições indicadas: -141 (A>G), -106 (A>G), -100 (T>G). As demais sequências são amostragens.

## ANEXO II. PARECER DO COMITÊ DE ÉTICA EM PESQUISA



INSTITUTO NACIONAL DE  
CÂNCER JOSÉ ALENCAR  
GOMES DA SILVA - INCA



### PARECER CONSUBSTANCIADO DO CEP

#### DADOS DO PROJETO DE PESQUISA

**Título da Pesquisa:** Investigação do papel de fatores genéticos e do vírus Epstein-Barr nos linfomas: na busca de um modelo multifatorial de susceptibilidade genética e resposta imune no linfoma de Hodgkin clássico

**Pesquisador:** Claudia Esther Alicia Rocio Hassan

**Área Temática:** Genética Humana:

(Trata-se de pesquisa envolvendo Genética Humana que não necessita de análise ética por parte da CONEP);

**Versão:** 3

**CAAE:** 56998916.5.0000.5274

**Instituição Proponente:** Centro de Transplante de Medula Óssea

**Patrocinador Principal:** Financiamento Próprio

#### DADOS DO PARECER

**Número do Parecer:** 1.752.286

#### **Apresentação do Projeto:**

Conforme descrito no Parecer Consubstanciado do CEP-INCA número 1.632.168, datado de 12 de julho de 2016.

#### **Objetivo da Pesquisa:**

Conforme descrito no Parecer Consubstanciado do CEP-INCA número 1.632.168, datado de 12 de julho de 2016.

#### **Avaliação dos Riscos e Benefícios:**

Conforme descrito no Parecer Consubstanciado do CEP-INCA número 1.632.168, datado de 12 de julho de 2016.

#### **Considerações Finais a critério do CEP:**

Diante do exposto, o Comitê de Ética em Pesquisa do Instituto Nacional de Câncer (CEP-INCA), de acordo com as atribuições definidas na Resolução CNS 466/12 e na Norma Operacional Nº 001/2013 do CNS, manifesta-se pela aprovação do projeto de pesquisa proposto.

Ressalto o(a) pesquisador(a) responsável deverá apresentar relatórios semestrais a respeito do seu estudo.

## ANEXO III. QUESTIONÁRIO SOCIOECONÔMICO



CAAE: 36999916.3.0000.5274 – Aprovado em 26 de setembro de 2016 – Versão 2



### QUESTIONÁRIO SOCIOECONÔMICO

Prezado participante, este questionário tem como objetivo conhecer o perfil socioeconômico daqueles doadores a quem o Projeto de Pesquisa: “Investigação do papel de fatores genéticos e do vírus Epstein-Barr nos linfomas: na busca de um modelo multifatorial de susceptibilidade genética e resposta imune no linfoma de Hodgkin clássico” se dirige.

Desde já, agradecemos sua valiosa colaboração.

1. Qual o seu sexo?

- (A) Feminino.
- (B) Masculino.

2. Qual a sua idade?

R.: \_\_\_\_\_.

3. Como você se considera:

- (A) Branco (a).
- (B) Pardo (a).
- (C) Preto (a).
- (D) Amarelo (a).
- (E) Indígena.
- (F) Outro \_\_\_\_\_.

4. Qual (s) etnia (s) que você saiba é (são) dos seus ancestrais familiares?

Família materna	Família paterna
(A) Não sei.	(A) Não sei.
(B) Africano:	(B) Africano:
(C) Europeu:	(C) Europeu:
(D) Ameríndio:	(D) Ameríndio:
(E) Outra (s):	(E) Outra (s):

5. Qual seu estado civil?

- (A) Solteiro (a).
- (B) Casado (a) / mora com um (a) companheiro (a).
- (C) Separado(a) / divorciado(a) / desquitado(a).
- (D) Viúvo(a).

6. Quantos (as) filhos (as) você tem?

- (A) Um (a).
- (B) Dois (duas).
- (C) Três.
- (D) Quatro ou mais.
- (E) Não tenho filhos (as).

7. Costuma beijar seus filhos/pais na boca?

- (A) Sim, sempre.
- (B) Sim, às vezes.
- (C) Não.

8. Qual (s) cidade(s) residiu até os 18 anos?

Se for cidade do Rio de Janeiro, indique o bairro.

R.: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

9. Qual cidade e bairro que você reside atualmente? Se possível, indique o CEP (código de endereçamento postal).

R.: \_\_\_\_\_

\_\_\_\_\_

10. Com quantos parentes morou até os 18 anos?

R.: \_\_\_\_\_

11. Com qual (s) parente(s) morou até os 18 anos?

R.: \_\_\_\_\_

12. Até quando seu pai estudou?

- (A) Não estudou.
- (B) Da 1ª à 4ª série do ensino fundamental (antigo primário).
- (C) Da 5ª à 8ª série do ensino fundamental (antigo ginásio).
- (D) Ensino médio (antigo 2º grau) incompleto.
- (E) Ensino médio completo.
- (F) Ensino superior incompleto.
- (G) Ensino superior completo.
- (H) Pós-graduação.
- (I) Não sei.

Rubrica do doador ou responsável: \_\_\_\_\_

Rubrica do responsável pelo questionário: \_\_\_\_\_

## ANEXO III. QUESTIONÁRIO SOCIOECONÔMICO (continuação)



CAAE: 36999916.3.0000.5274 – Aprovado em 26 de setembro de 2016 – Versão 2



13. Até quando sua mãe estudou?

- (A) Não estudou.
- (B) Da 1ª à 4ª série do ensino fundamental.
- (C) Da 5ª à 8ª série do ensino fundamental.
- (D) Ensino médio incompleto.
- (E) Ensino médio completo.
- (F) Ensino superior incompleto.
- (G) Ensino superior completo.
- (H) Pós-graduação.
- (I) Não sei.

14. Se não os pais, até quando a pessoa que te criou estudou?

- (A) Não estudou.
- (B) Da 1ª à 4ª série do ensino fundamental.
- (C) Da 5ª à 8ª série do ensino fundamental.
- (D) Ensino médio incompleto.
- (E) Ensino médio completo.
- (F) Ensino superior incompleto.
- (G) Ensino superior completo.
- (H) Pós-graduação.
- (I) Não sei.

15. Qual o seu grau de escolaridade?

- (A) Não estudei.
- (B) Da 1ª à 4ª série do ensino fundamental.
- (C) Da 5ª à 8ª série do ensino fundamental.
- (D) Ensino médio incompleto.
- (E) Ensino médio completo.
- (F) Ensino superior incompleto.
- (G) Ensino superior completo.
- (H) Pós-graduação.

16. Quantas pessoas moram em sua casa?

(Contando com você, seus pais, irmãos ou outros parentes/amigos que moram em uma mesma casa).

- (A) Duas pessoas.
- (B) Três pessoas.
- (C) Quatro pessoas.
- (D) Cinco pessoas.
- (E) Seis pessoas.
- (F) Mais de seis pessoas.
- (G) Moro sozinho (a).

17. Somando a sua renda com a renda das pessoas que moram com você, quanto é, aproximadamente, a renda familiar?

(Considere a renda de todos que moram na sua casa.)

- (A) Até 1 salário mínimo (até R\$ 954,00).
- (B) De 1 a 2 salários mínimos.
- (C) De 2 a 5 salários mínimos.
- (D) De 5 a 10 salários mínimos.
- (E) De 10 a 30 salários mínimos.
- (F) Acima de 30 salários mínimos.
- (H) Nenhuma renda.

18. Teve mononucleose infecciosa na infância?

- (A) Não sei o que é isso.
- (B) Não.
- (C) Não sei ou não lembro.
- (D) Sim, com menos de 10 anos.
- (E) Sim, acima de 10 anos.

19. Já realizou a cirurgia de retirada de amígdalas ou tonsilas?

- (A) Não.
- (B) Sim.
- (C) Não sei.

20. Já teve câncer?

- (A) Sim. Tipo: \_\_\_\_\_
- (B) Não.

21. Possui alguma doença crônica como diabetes, artrite, artrose, hipertensão arterial, etc.

R.: \_\_\_\_\_

22. Quantos relacionamentos afetivos você teve durante sua vida que duraram acima de 2 meses?

- (A) Até 3.
- (B) Entre 3 e 10
- (C) Acima de 11.
- (D) Nenhum.

Nome completo e assinatura do doador ou responsável:

\_\_\_\_\_ Data: \_\_\_/\_\_\_/\_\_\_

Nome completo e assinatura do responsável pelo questionário:

\_\_\_\_\_ Data: \_\_\_/\_\_\_/\_\_\_