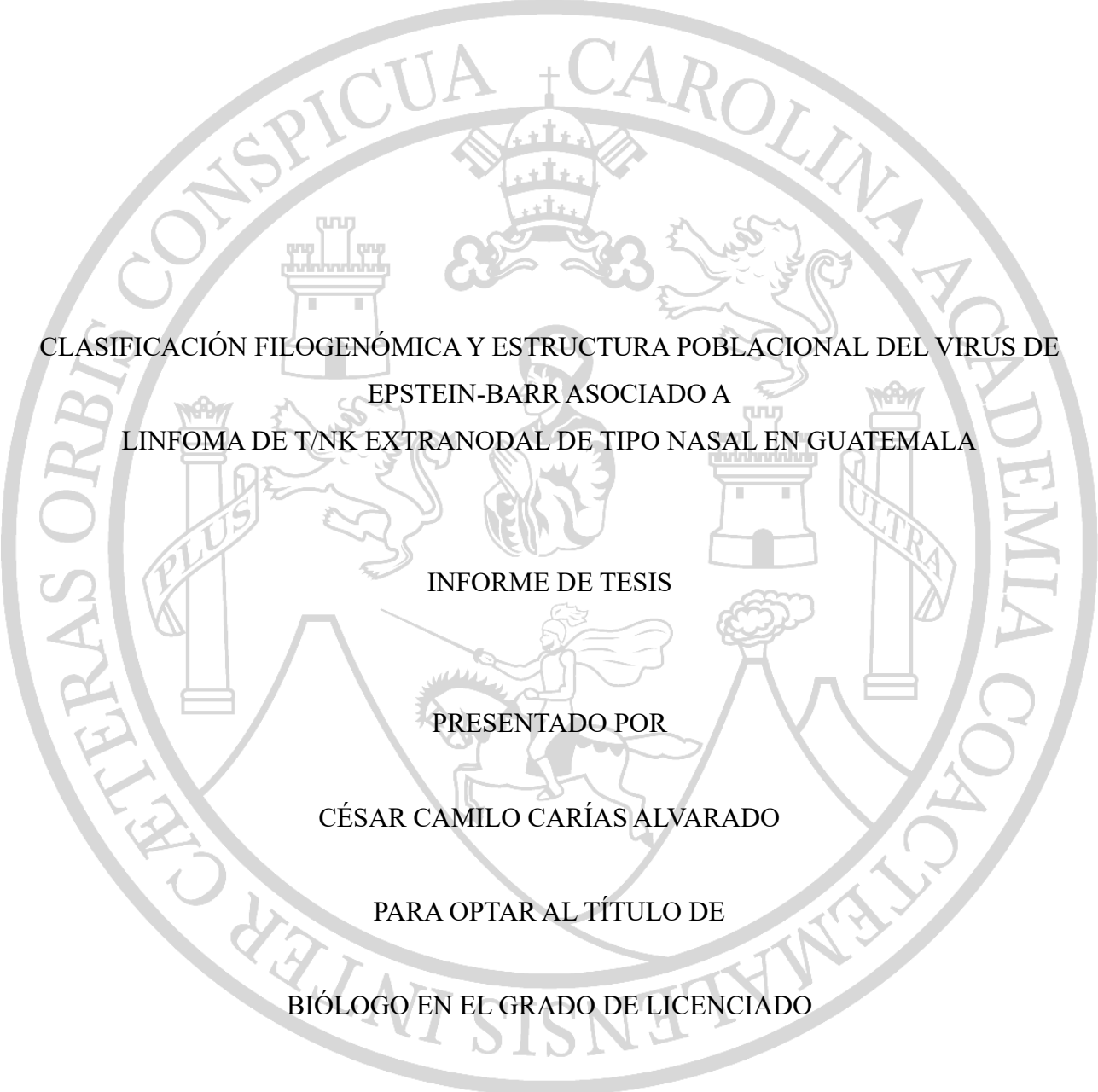


UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE CIENCIAS QUÍMICAS Y FARMACIA

The seal of the University of San Carlos of Guatemala is a large, circular emblem in the background. It features a central shield with a figure on horseback, a crown above, and two towers on either side. The shield is surrounded by a circular border containing the Latin motto "CETERAS ORBIS CONSPICUA CAROLINA ACADEMIA COACTEMMANSIS INTER".

CLASIFICACIÓN FILOGENÓMICA Y ESTRUCTURA POBLACIONAL DEL VIRUS DE
EPSTEIN-BARR ASOCIADO A
LINFOMA DE T/NK EXTRANODAL DE TIPO NASAL EN GUATEMALA

INFORME DE TESIS

PRESENTADO POR

CÉSAR CAMILO CARÍAS ALVARADO

PARA OPTAR AL TÍTULO DE

BIÓLOGO EN EL GRADO DE LICENCIADO

GUATEMALA, MAYO DE 2022

JUNTA DIRECTIVA
FACULTAD DE CIENCIAS QUÍMICAS Y FARMACIA

M. A. Pablo Ernesto Oliva Soto	Decano
Licda. Miriam Roxana Marroquín Leiva	Secretaria Académica
Dr. Juan Francisco Pérez Sabino	Vocal I
Dr. Roberto Flores Arzú	Vocal II
Lic. Carlos Manuel Maldonado Aguilera	Vocal III
Br. Carmen Amalia Rodríguez Ortiz	Vocal IV
Br. Paola Margarita Gaitán Valladares	Vocal V

DEDICATORIA

A las biólogas y los biólogos de Guatemala que dedican sus vidas, conocimientos y fuerzas por el desarrollo de la ciencia en el país.

AGRADECIMIENTOS

A la Universidad de San Carlos de Guatemala por ser mi alma mater.

A los catedráticos de la Facultad de Ciencias Químicas y Farmacia y la Escuela de Biología por brindarme sus conocimientos y contribuir en mi formación profesional en todo momento.

A mis padres y hermanos por apoyarme y darme la oportunidad de estudiar.

A mis familiares que han contribuido en mi formación y crecimiento personal.

A mis amigos y compañeros por sus muestras de amistad y compañerismo.

A todo el equipo de profesionales del Laboratorio de Investigación Biológica en Cáncer y del Centro de Investigaciones Biomédicas por todo el apoyo, conocimientos y oportunidades brindadas.

ÍNDICE

1.	RESUMEN.....	1
2.	INTRODUCCIÓN.....	3
3.	ANTECEDENTES.....	5
3.1	Virus de Epstein-Barr e Infección.....	5
3.2	Ciclo de Vida, Patogénesis y Mecanismos de Oncogénesis del Virus de Epstein-Barr.....	6
3.2.1	Patogénesis y Mecanismos de Oncogénesis del Virus de Epstein-Barr asociado al linfoma de T/NK extranodal de tipo nasal.....	8
3.3	Genoma del Virus de Epstein-Barr.....	8
3.3.1	Genes de latencia.....	11
3.3.1.1	EBNA1.....	11
3.3.1.2	EBNA2.....	12
3.3.1.3	Familia de genes EBNA3.....	12
3.3.1.4	LMP1.....	12
3.3.1.5	LMP2.....	13
3.3.1.6	EBER1 y EBER2.....	13
3.3.1.7	BARTs y BARF1.....	14
3.3.2	Genes líticos.....	14

3.3.2.1	BZLF1 y BRLF1.....	14
3.3.2.2	BNLF2a, BILF1 y BGLF5.....	15
3.4	Diagnóstico de la infección por Virus de Epstein-Barr.....	16
3.4.1	Diagnóstico serológico.....	16
3.4.2	Medición de la carga viral por medio de la amplificación de ácidos nucleicos.....	16
3.4.3	Histopatología.....	17
3.5	Tipos y Clasificación del Virus de Epstein-Barr.....	18
3.6	Neoplasias hematológicas linfoides.....	19
3.6.1	Neoplasias de células T y NK maduras.....	20
3.6.2	Linfoma de Hodgkin.....	21
3.6.3	Neoplasias linfoides de células B pequeñas.....	21
3.6.4	Asociación de la infección del VEB con el desarrollo de neoplasias hematológicas linfoides en Latinoamérica.....	24
3.7	Secuenciación de Nueva Generación y Análisis Bioinformático.....	26
3.7.1	Secuenciación de Nueva Generación.....	26
3.7.2	Bioinformática.....	26
3.7.3	Modelos de sustitución nucleotídica.....	27
3.7.4	Métodos de Máxima Verosimilitud.....	28

3.7.5	Análisis de Estructura Poblacional y Redes de Haplotipos.....	28
4.	JUSTIFICACIÓN.....	30
5.	OBJETIVOS.....	31
5.1	Objetivo General.....	31
5.2	Objetivos Específicos.....	31
6.	HIPÓTESIS.....	32
7.	MATERIALES Y MÉTODOS.....	33
7.1	Tipo de Estudio.....	33
7.2	Universo.....	33
7.2.1	Población.....	33
7.2.2	Muestra.....	33
7.3	Materiales.....	34
7.3.1	Equipo.....	34
7.3.2	Software.....	34
7.3.3	Base de datos.....	35
7.4	Métodos.....	35
7.4.1	Secuenciación.....	35

7.4.2	Preprocesamiento y análisis de datos de secuenciación.....	36
7.4.3	Caracterización del genoma de VEB de Guatemala.....	36
7.4.3.1	Análisis de variantes: SNP/INDEL.....	37
7.4.4	Análisis filogenético.....	37
7.4.5	Análisis de estructura poblacional.....	38
7.4.6	Análisis de recombinación genética.....	38
7.4.7	Análisis de red de haplotipos.....	39
8.	RESULTADOS.....	40
9.	DISCUSIÓN.....	52
10.	CONCLUSIONES.....	62
11.	RECOMENDACIONES.....	63
12.	REFERENCIAS.....	64
13.	ANEXOS.....	94

1. RESUMEN

El Virus de Epstein-Barr (VEB) infecta a más del 95% de la población mundial. El genoma del VEB está compuesto por un ADN linear de doble cadena, con un tamaño aproximado de 172 kb. Se estima que VEB se encuentra asociado a más de 200,000 casos de cáncer al año. Tal es el caso del Linfoma de T/NK Extranodal de Tipo Nasal (ENKTL) asociado a la infección, transformación y proliferación de las células T y NK. Dicha enfermedad tiene una ocurrencia principal en poblaciones asiáticas y de indígenas americanos del centro y sur de América. En Guatemala son poco conocidos muchos aspectos importantes del VEB, entre ellos la caracterización del genoma y sus relaciones filogenómicas. Por ello, es importante caracterizar el genoma del VEB asociado a ENKTL en Guatemala, describir sus relaciones filogenómicas y establecer su estructura poblacional.

Se llevó a cabo el análisis bioinformático de 22 genomas del VEB guatemalteco asociado a ENKTL obtenidos mediante un método de secuenciación de enriquecimiento objetivo. Se llevó a cabo una caracterización del genoma por medio de un análisis de variantes utilizando los paquetes de VCF/BCF de Galaxy y de recombinación genética empleando RDP4. Para llevar a cabo los análisis de relaciones filogenómicas y de estructura poblacional se emplearon 215 del VEB provenientes de otros países y asociados a distintos tipos de cáncer que fueron obtenidos de NCBI. El análisis de las relaciones filogenómicas se realizó con un método de máxima verosimilitud, utilizando los programas raxmlGUI 2.0 y PhyML 3.0. Los análisis de estructura poblacional se realizaron con rhierbaps empleando dos niveles de jerarquía. Los análisis filogenómicos y de estructura poblacional también se llevaron a cabo eliminando regiones recombinantes para evaluar el efecto de la recombinación en la propuesta de clasificación filogenómica del VEB. Finalmete, se realizó una red de haplotipos a partir de la región OriP.

Se obtuvo un genoma de 171,823 pb, del cual se identificaron 30 regiones de recombinación y 1,127 variantes. Las relaciones filogenómicas y la red de haplotipos demuestran una relación entre las cepas guatemaltecas asociadas a ENKTL con cepas de países asiáticos como China y Singapur asociadas a ENKTL, Cáncer Nasofaríngeo y Cáncer Gástrico. Al comparar los resultados obtenidos mediante los análisis con regiones recombinantes y en los que se eliminaron, se obtuvieron resultados coincidentes en las relaciones filogenómicas y filo poblaciones, sin embargo el análisis de estructura poblacional sin regiones recombinantes ha separado a las cepas guatemaltecas de las cepas de China y Singapur asociadas a ENKTL, sugiriendo que la Filo Población Guatemalteca se comporta de manera aislada.

A partir de los resultados obtenidos, acerca del VEB asociado a ENKTL en pacientes del Instituto de Cancerología (INCAN), Guatemala, se puede concluir que posee un genoma de 171,823 pb, del cual se identificaron 30 regiones de recombinación y 1,142 variantes. Además, filogenómicamente el VEB se encuentra estrechamente asociado con aquellas cepas del virus provenientes de países asiáticos, como China y Singapur, y que se encuentran asociadas a ENKTL, cáncer nasofaríngeo y cáncer gástrico. Finalmente, la población de VEB se comporta como una población aislada.

2. INTRODUCCIÓN

El Virus de Epstein-Barr (VEB, *Human gammaherpesvirus 4*, perteneciente al género *Lymphocryptovirus*, sub-familia Gammaherpesviridae, familia *Herpesviridae*), es un gamma-herpes virus que infecta a más del 95% de la población mundial (Peng et al, 2019). El genoma del VEB está compuesto por un ADN linear de doble cadena, con un tamaño aproximado de 172 kilo bases (kb), incluyendo alrededor de 80 regiones codificadoras (CDS), regiones repetitivas y, en el caso de líneas celulares y biopsias de tumores, grandes regiones INDEL (Amon & Farrell, 2005; Zanella et al., 2019). La prevalencia de la infección por el VEB en niños varía entre 10 – 90% y depende de varios factores como lo son la edad, la región geográfica y la etnia o raza de la persona. Así mismo, la mayoría de la población mundial es infectada antes de llegar a la edad adulta (Dunmire, Verghese & Balfour, 2018). Actualmente, se estima que VEB se encuentra asociado a más de 200,000 casos de cáncer al año y un 1.8% de todas las muertes por cáncer se atribuyen a neoplasias malignas causadas por el mismo (Young, Yap & Murray, 2016).

La infección por VEB ha sido asociada con el desarrollo de diversos tipos de cáncer como el linfoma de Burkitt, Linfoma de T/NK Extranodal de Tipo Nasal (ENKTL), carcinoma nasofaríngeo (NPC, por sus siglas en inglés), aproximadamente el 30% de linfoma Hodgkin, el 10% de casos de cáncer gástrico (CG) y casos de linfoma difuso de células B grandes y leiomiomasarcoma (Palser et al., 2015). Así mismo, el VEB puede estar asociado con diferentes desórdenes linfoproliferativos dependiendo el tipo de célula que infecta. Tal es el caso de los desórdenes linfoproliferativos asociados a la infección, transformación y proliferación de las células T y NK por VEB durante el estadio de latencia 2, también conocido como Linfoma de T/NK Extranodal de Tipo Nasal. Dicha enfermedad tiene una ocurrencia principal en poblaciones asiáticas y de indígenas americanos del centro y sur de América (Kim et al., 2019; Kimura et al., 2012; Yamaguchi, Oguchi & SuzReino Unido, 2018). Los ENKTL históricamente se han clasificado en formas indolentes y agresivas o fulminantes y se caracterizan clínicamente con diagnóstico a una edad media entre los 46 a 52 años, lesiones de masas nasales que a su vez causan obstrucción, secreciones nasales y epistaxis. Así mismo, la enfermedad puede diseminarse hacia la piel, el tracto gastrointestinal, testículos y ganglios linfáticos. A comparación con otros linfomas, el 40% de casos de ENKTL presentan síntomas B (Kimura et al., 2012 ; Makita & Tobinai, 2017).

Diversos estudios se han realizado con la finalidad de clasificar el VEB a partir de metodologías distintas. Inicialmente la clasificación del virus se abordó con el análisis del gen EBNA2 es-

tableciendo dos genotipos Tipo 1 (VEB-1 o Tipo A) y Tipo 2 (VEB-2 o Tipo B) (Choi et al., 2018; Neves et al., 2016; Salahuddin et al., 2018). Recientemente, los estudios se han enfocado en clasificar el virus por medio del análisis del genoma, exoma y transcriptoma completo, el tipo de cáncer asociado, el país de procedencia y las características demográficas del paciente, haciendo uso de técnicas de secuenciación de nueva generación (NGS) para obtener datos genómicos (Peng et al., 2019; Xiong et al., 2020; Zanella et al., 2019).

El presente estudio de tesis consistió en determinar la clasificación filogenómica y estructura poblacional del Virus de Epstein-Barr asociado a ENKTL en Guatemala. Para ello se realizó el análisis bioinformático de secuencias de genomas obtenidos por secuenciación de nueva generación del VEB por medio de un método de enriquecimiento objetivo para VEB a partir de 22 muestras de biopsia de tejido tumoral de ENKTL embebido en bloques de parafinas. Las muestras obtenidas en el Instituto de Cancerología (INCAN) fueron recolectadas entre los años 2008 a 2019. Los datos genómicos obtenidos fueron analizados mediante técnicas bioinformáticas y de esa manera se obtuvo la caracterización del genoma del Virus de Epstein-Barr asociado a ENKTL en Guatemala, se describieron las relaciones filogenómicas y se estableció la estructura poblacional entre el Virus de Epstein-Barr asociado a ENKTL en Guatemala con cepas del virus asociadas a otros tipos de cáncer y provenientes de otros países.

3 ANTECEDENTES

3.1 Virus de Epstein-Barr e Infección

El Virus de Epstein-Barr (VEB, *Human gammaherpesvirus 4*, perteneciente al género *Lymphocryptovirus*, sub-familia Gammaherpesviridae, familia *Herpesviridae*), es un gamma-herpes virus que infecta a más del 95% de la población mundial (Peng et al, 2019). Los virus de la familia *Herpesviridae* se caracterizan por ser virus de ADN de cadena doble; causantes de enfermedades como el herpes labial, herpes venéreo, varicela y mononucleosis infecciosa; y por tener la capacidad de permanecer en estados de latencia en el cuerpo del hospedero durante prolongados periodos de tiempo (Prescott, Harley & Klein, 2002). El VEB pertenece a un grupo de herpesvirus carcinogénico el cual está relacionado con el desarrollo de enfermedades como la mononucleosis infecciosa, linfoma de Burkitt (BL), linfoma de Hodgkin (HL), cáncer gástrico, cáncer nasofaríngeo, entre otras (Young & Rickinson, 2004). La estructura de virión del VEB consiste en un núcleo proteico de forma toroidal envuelto del ADN viral dentro de una cápside icosaédrica de 162 capsómeros, un tegumento viral que contiene proteínas que delimitan el espacio entre la nucleocápside, y finalmente la envoltura externa compuesta de diferentes glucoproteínas *spike* que se insertan en la envoltura viral (Bouvard et al., 2009; Smatti et al., 2018).

La prevalencia de la infección por el VEB en niños varía entre 10 – 90% y depende de varios factores como lo son la edad, la región geográfica y la etnia o raza de la persona. Así mismo, la mayoría de la población mundial es infectada antes de llegar a la edad adulta (Dunmire, Verghese & Balfour, 2018). Otro factor importante que determina la infección por el VEB es el desarrollo social del país de residencia pues se conoce que niños de países con mayor desarrollo social y económico se infectan a una edad más tardía que aquellos que viven en países en vías de desarrollo (Young, Yap & Murray, 2016). La edad y la infección primaria por el VEB, el desarrollo y severidad de la enfermedad asociada y el riesgo de desarrollar cáncer poseen una relación muy importante. Por ejemplo, estudios sugieren que en preadolescentes la mononucleosis infecciosa es menos común que en adolescentes y adultos jóvenes (Cohen, 2000; Smatti et al., 2018).

La infección en niños es transmitida en el hogar mediante el contacto con los padres, hermanos y otros miembros de la familia que estén infectados por el virus. Hábitos como el compartir cubiertos, vasos y cepillos de dientes, entre otros, podrían explicar la transmisión del virus en el hogar. En adolescentes y adultos se puede transmitir vía oral a través de besos ya que la carga viral del virus en la boca puede llegar a niveles de 63 mil copias/mL hasta dos meses después de la infección y

es capaz de persistir hasta los cinco meses (Prescott, Harley & Klein, 2002). Así mismo, se ha reportado que las transfusiones de sangre son una vía de transmisión y desarrollo de mononucleosis infecciosa. A pesar que el riesgo de contraer la infección por este medio parece ser bajo, aún no se conocen datos de incidencia. Finalmente, otra vía de transmisión reportada es entre donador y receptor de órganos o tejidos, destacando en los trasplantes de médula ósea (Shapiro et al., 1988).

El periodo de incubación de la mononucleosis infecciosa en adolescentes y adultos varía entre 30 a 50 días, mientras que en niños llega a ser menor. En pacientes pediátricos la infección suele ser asintomática o presentar síntomas similares a la fiebre ocasionada por otros agentes patógenos. Por su parte, en adolescentes y adultos los síntomas son más característicos: se presentan entre la primera y segunda semana de infección, fiebre, dolor de garganta, linfadenopatía, dolor de cabeza, náusea, dolor abdominal y mialgias. Además, los dolores de cabeza y fiebre aumentan gradualmente y se desarrolla esplenomegalia, por la que se puede llegar a requerir atención médica (Jenson, 2011; Young & Rickinson, 2004).

3.2 Ciclo de Vida, Patogénesis y Mecanismos de Oncogénesis del Virus de Epstein-Barr

VEB posee un ciclo de vida bifásico en el que posterior a la infección primaria de los linfocitos B los virus se encuentra en estado lítico expresando genes como *BZFL1*, *BRFL1*, *BNFL2*, *VCA* y *BCRF1*. Seguidamente, el virus entra en una fase de latencia en respuesta a la reacción del sistema inmunológico y el genoma viral queda dentro del núcleo de la célula infectada (episoma), es replicado y se presenta la expresión de varios genes de latencia *EBNA-1*, *EBNA-2*, *EBNA-3A*, *EBNA-3B*, *EBNA-3C*, *EBNA-LP* y *MLP-1* (Cohen, 2000; Tsao et al., 2015; Young, Yap & Murray, 2016). Finalmente, en linfocitos B se produce una fase de reactivación en la cual se estimula el receptor de células B y el virus entra a una fase lítica en la que se producen viriones infecciosos y se expresan genes como *BZFL1*, *BRFL1*, *BNFL2*, *VCA* y *BCRF1* (Amon & Farrell, 2005; Young, Arrand & Murray, 2007). El VEB es capaz de infectar otras células como lo son células epiteliales, linfocitos T, células de plasma, células NK, monocitos, células dendríticas foliculares escamosas, células epiteliales glandulares mioepiteliales y células de músculo liso (Hutt-Fletcher, 2007; Smatti et al., 2018).

Posterior a la entrada del VEB por medio del epitelio de las amígdalas, inicia la fase lítica de la infección que implica su replicación viral. Las células naive B infectadas se tornan en linfoblastos y migran a los folículos de los ganglios linfáticos para iniciar una reacción en el centro germinal del

foliculo utilizando el programa de latencia III en el cual todas las proteínas de crecimiento son expresadas y regulan el crecimiento del virus (Jenson, 2011). Entre las proteínas que se expresan durante este período destacan los antígenos nucleares de VEB (EBNAs 1, 2, 3, 3A, 3B, 3C y LP) y las proteínas latentes de membrana (LMP1, LMP2A y LMP2B) (Taso et al., 2015). El programa de latencia II inicia cuando únicamente EBNA-1, EBERs, BARTs, LMP1 y LMP2A se encuentran expresándose. Es en este momento cuando señales de supervivencia celular llegan al centro germinal e inducen la salida de las células como linfocitos B de memoria. La fase de latencia 0 inicia en los linfocitos B de memoria y cuenta con la principal característica de que las proteínas virales dejan de ser expresadas. Una variante de la fase de latencia 0 es la fase de latencia 1 en la cual únicamente la proteína EBNA1 se encuentra expresándose (Amon & Farrel, 2005; Rowe et al., 1992). Durante dichas fases de latencia los linfocitos T no son capaces de eliminar las células B infectadas y controlar la infección debido a que el virus se mantiene escondido del sistema inmune al no expresar proteínas virales (Smatti et al., 2018; Young, Yap & Murray, 2016). Durante el período de latencia los EBNAs permiten que el virus se mantenga como episoma dentro del núcleo y separado de los cromosomas de las células infectadas. La persistencia del virus se logra gracias a que durante la mitosis los episomas virales migran hacia cada célula hija. Una excepción es el caso de la infección asociada a Linfoma de Burkitt en la que el VEB se integra al cromosoma 8 por medio de una translocación recíproca que involucra al alelo *MYC* (Jenson, 2011; Smatti et al., 2018).

A pesar de que aún no se conoce a totalidad las razones y mecanismos, algunas células B infectadas pasan de la fase de latencia a una fase lítica (o de replicación viral) produciendo, en primer lugar, antígenos tempranos de VEB (EAs) seguido de la replicación de ADN viral, la producción de glucoproteínas estructurales y antígenos virales de la cápside (VCA). Por último, se lleva a cabo una muerte celular que permite la liberación de los viriones los cuales infectarán de manera sistemática a otras células B (Jenson, 2011). La reactivación viral ocurre de manera ocasional en aquellas células B infectadas en fase de latencia mediando un nuevo ciclo viral en el que se llevará a cabo la replicación del virus y la infección de nuevas células. La reactivación en pacientes inmunocompetentes es controlada por la acción de los linfocitos T, mientras que en pacientes inmunosuprimidos debe ser controlada por los riesgos de desarrollo de patologías severas. Entre las causas que estimulan la reactivación viral se encuentran las infecciones de otros agentes patógenos que inducen la división de las células B infectadas y a su vez inducen la reactivación y replicación viral. Agentes químicos, drogas y la inflamación han sido vinculados con la reactivación viral (Amon & Farrell, 2005; Binné, Amon & Farrell, 2002).

3.2.1 Patogénesis y Mecanismos de Oncogénesis del Virus de Epstein-Barr asociado al linfoma de T/NK extranodal de tipo nasal

El VEB asociado al ENKTL se encuentra en estado episomal y corresponde a un tipo 2 de latencia. Se ha documentado que la mayoría de pacientes con la enfermedad posee un subtipo A del virus que cuenta con variantes geográficas. Frecuentemente, el gen LMP1 posee una delección de 30 pb lo cual es asociado a una baja en el reconocimiento inmune y al desarrollo de la enfermedad (García-Cosío et al., 2003; Nagamine et al., 2007). Elevadas copias de ADN del virus reflejan una alta actividad tumoral de la enfermedad y son relacionadas a respuestas pobres al tratamiento y a pronósticos clínicos inferiores. Varios casos de la enfermedad asociada a la infección por VEB suelen presentar configuraciones diferentes en los genes receptores de linfocitos T (TCR), presentándose rearrreglos monoclonales en un 10 – 40% de casos. Así mismo, delecciones en el brazo largo (q) del cromosoma 6 y en las regiones q21 – 23 son frecuentes en los genes supresores de tumores como HACE1, PRMD1, FOXO3 y PTPRK. Aún es desconocido si dichas mutaciones están involucradas en los eventos de patogénesis primarios o como un evento secundario relacionado con la progresión de la enfermedad. Finalmente, perfiles genéticos de la enfermedad han demostrado mutaciones en rutas como JAK3, STAT3 y STAT5B, así mismo en la ARN helicasa DDX3X, el gen supresor de tumores TP53 y genes involucrados en rutas epigenéticas como MLL2 y EP300 (Gualco et al., 2011; Kim et al., 2019; Kwong et al., 2014).

3.3 Genoma del Virus de Epstein-Barr

El genoma del VEB está compuesto por un ADN linear de doble cadena, con un tamaño aproximado de 172 kilobases (kb), incluyendo alrededor de 80 regiones codificadoras (CDS), regiones repetitivas y, en el caso de líneas celulares y biopsias de tumores, grandes regiones INDEL (Amon & Farrell, 2005; Zanella et al., 2019). El genoma se encuentra dividido en una región única corta denominada U1 y una región única larga denominada U2. A su vez, estas se encuentran divididas en cuatro regiones internas repetitivas IR1 – IR4 y en los extremos del genoma cuenta con dos regiones terminales repetitivas (TR). Además, el genoma cuenta con un origen de replicación (Ori-P) y dos orígenes líticos de replicación (Ori-lyt) (Young et al., 2007).

Los genes del VEB se dividen a partir del estado del ciclo de vida en el que son expresados siendo divididos en genes de latencia y genes líticos. El listado de marcos de lectura abiertos (ORF, por sus siglas en inglés) y sus productos (genes o proteínas) caracterizados según el estado en el que

son expresados se presenta en la Tabla 1.

Tabla 1. Marcos de lectura abiertos y sus respectivos productos del genoma del VEB.

Marco de Lectura Abierto (ORF)	Gen o Proteína
Genes de Latencia	
BKRF1	EBNA1
BYRF1	EBNA2
BLRF3/BERF1	EBNA3A (EBNA3)
BERF2a/b	EBNA3B (EBNA4)
BERF3/4	EBNA3C (EBNA6)
BamHI-W (múltiples secciones repetitivas)	EBNA-LP (EBNA5)
BNLF1	LMP1
Regiones Terminales Fusionadas (múltiples secciones)	LMP2A/B
BARTs	A73, RPMS1
EBER1/2	Pequeños ARNs
Genes Líticos Tempranos	
BZLF1	ZEBRA/Zta/EB1
BRLF1	Rta
BRRF1	Factor de transcripción
BORF2	Subunidad grande ribonucleótido reductasa
BaRF1	Subunidad pequeña ribonucleótido reductasa
BXLF1	Timidina quinasa
BGLF5	Exonucleasa alcalina
BLLF3	dUTPasa
BKRF3	Uracil ADN glicosilasa
BALF5	ADN polimerasa
BMRF1	Proteína accesoria de la polimerasa
BALF2	Proteína de unión a ADN
BSLF1	Primasa
BBLF2/3	Proteína accesoria de la primasa
BBLF4	Helicasa
BMLF1	Factor de exportación de ARNm
BSLF2	

BHRF1	Homólogo de bcl-2
BALF1	Antagonista viral de bcl-2
BARF1	Proteína asociada a transformación
BGLF4	Proteína quinasa
BFRF1	Proteína de membrana de 37kDa
BHLF1	Proteína de unión a ADN de cadena simple
BHLF2	Proteína de envoltura
BNLF2a	Evasión inmune
Genes Líticos Tardíos	
BNRF1	Proteína p143 mayor de tegumento
BPLF1	Proteína larga de tegumento
BOLF1	Proteína de tegumento
BVRF1	Proteína de tegumento
BBLF1	Proteína de tegumento
BGLF1	Proteína de tegumento
BSRF1	Proteína de tegumento
BRRF2	Proteína de tegumento
BDLF2	Proteína de tegumento
BKRF4	Proteína de tegumento
BcLF1	Proteína mayor dencápside
BDLF1	Proteína menor de cápside
BFRF3	Proteína p18 de cápside
BLRF2	Proteína p123 de cápside
BdRF1	Proteína p40 de cápside
BBRF1	Proteína de cápside
BVRF2	Proteasa
BGLF2	Proteína 38Kd
BORF1	Ensamblaje de cápside
BLRF1	Glicoproteína gN
BLLF1	Glicoproteína gp350/220
BZLF2	Glicoproteína gp42
BKRF2	Glicoproteína gL (gp25)
BBRF3	Glicoproteína gM
BXLF2	Glicoproteína gH (gp85)
BILF1	Glicoproteína gp60

BILF2	Glicoproteína gp78/55
BALF4	Glicoproteína B (gp110)
BDLF3	Glicoproteína gp150
BMRF2	Proteína de membrana 53/55Kd
BALF3	Glicoproteína de transporte
BCRF1	Citoquina viral IL-10

Fuente: Young et al., 2007.

3.3.1 Genes de latencia

Los genes de latencia poseen la principal función de coordinar la expresión de proteínas responsables para la inmortalización de las células B *in vitro*. Entre los productos de los genes de latencia que se expresan durante el el estado de latencia III o programa de crecimiento se encuentran seis antígenos nucleares de Epstein-Barr (EBNAs 1, 2, 3A, 3B, 3C y EBNA-LP), las proteínas latentes de membrana LMP1 y LMP2 (esta última contando con dos isoformas que son LMP2A y LMP2B), los ARNs no codificantes (EBER1 y EBER2) y otros microARN virales (Palser et al., 2015). Durante los diferentes estados de diferenciación de las células B *in vitro*, el VEB tiene la capacidad de expresar otras alternativas de latencia, siendo los estados de latencia I y latencia II. La expresión de determinado estado de latencia I, II o III durante la transformación de las células B se asocia, y de cierta manera depende, del estado de diferenciación de la célula B del cual cada tumor diferente es derivado. No obstante, actualmente se reconoce que todos los cánceres epiteliales asociados a VEB expresan un programa de latencia II (Rowe et al., 1987; Rowe et al., 1992; Young, Yap & Murray, 2016).

3.3.1.1 EBNA1

Codifica al antígeno nuclear 1 de Epstein-Barr el cual consiste en una proteína de unión a ADN. Su función es esencial para el mantenimiento y la replicación episomal, induce linfomas de células B en ratones lo que sugiere un papel importante en la oncogénesis, incrementa la supervivencia celular, induce la inestabilidad genética, actúa como un transactivador transcripcional que regula positivamente la expresión de varios genes virales, entre ellos, la región promotora del gen LMP1 y el promotor Cp (Knipe et al., 2001).

3.3.1.2 EBNA2

Codifica a la proteína del mismo nombre la cual posee un rol fundamental en el proceso de transformación de las células B. Así mismo, lleva a cabo las funciones de activador transcripcional de genes celulares y virales, regula positivamente la expresión de genes como LMP1 y LMP2 y de antígenos de las células, por ejemplo BCD21 y CD23. EBNA2 también se regula positivamente al promotor Cp observado en la infección temprana de las células B. Finalmente, estudios sugieren que EBNA2 puede funcionalmente reemplazar la región intracelular para las rutas de señalización Notch (Höfelmayr et al., 2001; Kelly et al., 2005).

3.3.1.3 Familia de genes EBNA3

Tanto EBNA3A y EBNA3C son esenciales para la transformación de las células B. Por su parte, EBNA3C puede regular positivamente la expresión de genes celulares (CD21) y virales (LMP1) y regula negativamente la expresión del promotor Cp (Radkov et al., 1997). Entre otras funciones, el gen EBNA3B puede funcionar como un supresor de tumores en el linfoma difuso de células B (Kelly et al., 2005; White et al., 2012).

3.3.1.4 LMP1

La proteína LMP1 por el gen es la principal proteína transformadora del VEB que posee un comportamiento de oncogén clásico y es esencial para la transformación de las células B *in vitro* (Knipe et al., 2001). Además, LMP1 posee efectos pleiotrópicos cuando es expresado en las células dando como resultado la inducción de moléculas de adhesión a la superficie celular activación de antígenos, regulación positiva de la producción de proteínas antiapoptóticas y la estimulación de la producción de citoquinas. LMP1 también se encuentra relacionado con la activación de rutas de señalización formando parte de la súper familia de receptores del factor de necrosis tumoral (TNFR) (Dawson, Port & Young, 2012; Kilger et al., 1998).

La proteína LMP1 consiste en una proteína integral de membrana con un peso de 63 kD y que se encuentra conformada por tres dominios: 1) una cola citoplasmática N-terminal que une la proteína LMP1 a la membrana plasmática; 2) seis bucles transmembrana hidrofóbicos que partici-

pan en la autoagregación y oligomerización; y 3) una región citoplasmática C-terminal que lleva a cabo la mayor parte de la actividad de señalización de la proteína (Gires et al., 1997; Young et al., 2007). Además, la proteína cuenta con dos dominios funcionales denominados regiones 1 y 2 de activación C-terminal (CTAR1 y CTAR2) los cuales son importantes para la activación de la vía del factor de transcripción NF- κ B (factor nuclear potenciador de las cadenas ligeras kappa de las células B activadas); lo cual contribuye a las muchas consecuencias fenotípicas de la expresión de LMP1, entre ellas: la inducción de varios genes antiapoptóticos y las citoquinas. Finalmente, la proteína LMP1 también es capaz de activar la cascada MAP quinasa resultando en la estimulación de las vías JAK/STAT (Eliopoulos et al., 1999).

3.3.1.5 LMP2

El gen LMP2 codifica a las proteínas LMP2A y LMP2B. Ambas proteínas cuentan con 12 dominios transmembrana y 27 aminoácidos citoplásmicos C-terminal; la diferencia entre ellas reside en que la proteína LMP2A posee un dominio amino terminal de 119 aminoácidos citoplásmicos. Ninguna de las dos proteínas codificadas por LMP2 juegan un papel esencial en la transformación de las células B. No obstante, se sugiere que las proteínas participan en la modificación del desarrollo normal de las células B para favorecer el ciclo de latencia del VEB y prevenir la activación en un momento inapropiado del ciclo lítico del virus (Knipe et al., 2001). Por su parte, LMP2A posee un motivo en el dominio amino terminal denominado motivo de activación inmunorreceptor basado en tirosina (ITAM) el cual al estar fosforilado en el receptor de células B tiene la capacidad de mediar la proliferación y diferenciación de linfocitos (Longnecker, 2000).

3.3.1.6 EBER1 y EBER2

Los EBER1 y EBER2 son dos pequeños ARN no poliadenilados y no codificantes son expresados en el estado de latencia del virus. Así mismo, EBER1 y EBER2 no son esenciales para que se lleve a cabo la transformación primaria de los linfocitos B (Laing et al., 2002). Los EBER se unen en forma de partículas de ribonucleoproteínas (complejos formados por un ácido ribonucleico y una proteína de unión con ARN) estables con el auto antígeno La y la proteína ribosómica L22, y a su vez, se unen con la proteína quinasa PKR (Takada, 2012). PKR al ser una quinasa que media los efectos antivirales de los interferones, se ha sugerido que su inhibición mediada por EBER podría tener un papel importante en la persistencia viral. Ejemplo de ello es que en líneas celulares de

linfoma de Burkitt que expresan EBER se obtiene un aumento en la tumorigenicidad, promueve la supervivencia celular e induce la expresión de la interleucina-10 (Kitagawa et al., 2000).

3.3.1.7 BARTs y BARF1

BamHIA rightward transcripts (BARTs) son un grupo de transcritos codificados en la región BamHIA del genoma del VEB y su expresión ha sido identificada en diferentes tipos de cáncer como el cáncer nasofaríngeo, el linfoma de Burkitt, linfoma Hodgkin y el linfoma nasal de células T, como también han sido identificados en sangre periférica de pacientes sanos (Chen et al., 1999). Los BARTs codifican marcos de lectura abiertos como BARF0, RK-BARF0, A73 y RPMS1. Así mismo, se dividen en dos grupos de ARN no codificante 1) un grupo de microARNs (miARNs) y 2) una familia compleja de empale alternativo de ARNs poliadenilados (Verhoeven et al., 2019).

BAMHI-A fragment rightward reading frame 1 (BARF1) es un marco de lectura abierto que codifica a una proteína de 33 kDa. Cuenta con la capacidad de inducir la transformación de células B en líneas celulares humanas y en líneas celulares de fibroblastos de roedores. Es considerado un gen que se expresa de manera temprana en la etapa lítica del ciclo de vida del virus. Se expresa en cáncer nasofaríngeo y cáncer gástrico positivo a la infección por el VEB cuando hay una ausencia de expresión de los genes de la etapa lítica del virus (Decaussin et al., 2000; Sheng et al., 2001).

3.3.2 Genes líticos

En la fase lítica el VEB lleva a cabo, principalmente, su replicación. Estímulos de diversos factores como lo son el 12-O-tetradecanoilforbol-13-acetato y el Butirato de sodio y la infección de otros agentes patógenos tienen la capacidad de inducir el cambio desde la fase de latencia a la fase lítica en el virus (Temple et al., 2014). No obstante, el principal responsable de llevar a cabo el cambio de fase en el ciclo de vida del VEB es el gen BLZF1 (Sun & Thorley-Lawson, 2007). El proceso de activación de la fase lítica del VEB se encuentra estrechamente relacionado con la diferenciación de las células B y células epiteliales infectadas. No obstante, aún es totalmente desconocido todos los mecanismos y las condiciones que lo activan (Young, Yap & Murray, 2016).

3.3.2.1 BZLF1 y BRLF1

Sinérgicamente los genes BZLF1 y BRLF1 tienen la capacidad de inducir la expresión de múltiples genes de la fase lítica temprana. Entre estos genes se encuentran 6 proteínas virales nece-

sarias para ensamblar el mecanismo de replicación, entre ellas (Fixman, Hayward & Hayward, 1992; Wen et al., 2007):

- BALF5 correspondiente a la ADN polimerasa.
- BALF2 correspondiente al homólogo de la proteína de unión al ADN monocatenario.
- BMRF1 correspondiente al factor de procesividad de la ADN polimerasa.
- BSLF1 correspondiente al homólogo de la primasa.
- BBLF4 correspondiente al homólogo de la helicasa.
- BBL2/3 correspondiente a un potencial homólogo del tercer componente del complejo helicasa-primasa. Es formado a partir del empalme alternativo entre los marcos de lectura abiertos BBLF2 y BBLF3.

Además de su función como transactivador, la proteína BZLF1 posee interacciones directas con el origen de replicación lítico (oriLyt) y con las demás proteínas de replicación, las cuales son esenciales para la replicación del ADN del VEB. Así mismo, BZLF1 tiene la capacidad de inducir inestabilidad genómica (Wen et al., 2007; Wu et al., 2010).

3.3.2.2 BNLF2a, BILF1 y BGLF5

Los genes líticos BNLF2a, BILF1 y BGLF5 codifican a proteínas de evasión inmune. BNLF2a codifica a una proteína que inhibe el transportador del procesamiento de antígenos, BILF1 codifica a una proteína que induce la internalización y degradación del complejo mayor de histocompatibilidad de clase I (CMH I), y BGLF5 codifica una proteína que media la regulación negativa de la expresión genética de la célula huésped favoreciendo la evasión de las respuestas inmunitarias antivirales. Así mismo, BGLF5 y BGLF4 inducen inestabilidad genómica (Hislop et al., 2007; Rowe et al., 2007; Zuo, 2009).

Finalmente, estudios recientes de secuenciación de ARN han demostrado niveles altos en la expresión de genes líticos en líneas celulares y biopsias de tumores, entre los cuales se encuentran: LF1, LF2, LF3, BILF1, BALF4 y BHLF1; no obstante, aún no está claro la expresión de dichos genes en una fase de latencia (Lin et al., 2010).

3.4 Diagnóstico de la infección por Virus de Epstein-Barr

3.4.1 Diagnóstico serológico

El diagnóstico serológico de la infección se lleva a cabo de manera rutinaria a partir de distintas metodologías como lo es el análisis de inmunofluorescencia, enzima-inmuno análisis, ELISA y el inmuno análisis de quimioluminiscencia (De Paschale & Clerici, 2012). El diagnóstico serológico de la infección por VEB no es recomendado en pacientes inmunosuprimidos debido a la confusión que puede obtenerse entre el resultado con la incapacidad del paciente para generar una cantidad apropiada de anticuerpos. Así mismo, el diagnóstico de la infección por serología no es recomendado en pacientes que cuentan con deficiencias en la producción de anticuerpos congénitas o adquiridas (Nowalk & Green, 2016). Los diagnósticos clásicos de serología a partir de técnicas de inmunofluorescencia se basan en detectar VCA-IgM, VCA-IgG y el anti-EBNA-IgG, junto al anti-EA-IgG. Dicho diagnóstico cuenta con la ventaja de ser fácilmente reproducibles (Niller & Bauer, 2017).

El uso de anticuerpos heterófilos ha sido ampliamente usado en el diagnóstico de la infección por VEB. Dichos anticuerpos se caracterizan por ser tipo IgM que aglutinan las hemáticas de especies de mamíferos diferentes a las de la fuente. La prueba cualitativa rápida de eritrocitos de caballo es el método más utilizado para detectar los anticuerpos heterófilos. La prueba es capaz de detectar los anticuerpos heterófilos en el 90% de casos de mononucleosis infecciosa en niños mayores a los 4 años de edad y en adultos. Actualmente no se recomienda su uso para diagnosticar la mononucleosis infecciosa en niños menores de 4 años debido a la baja producción de anticuerpos (Jenson, 2011).

3.4.2 Medición de la carga viral por medio de la amplificación de ácidos nucleicos

La medición de la carga viral en sangre periférica por medio de qPCR (Reacción en Cadena de la Polimerasa Cuantitativa o PCR en tiempo real), se ha convertido en un importante método diagnóstico del VEB. Dicha metodología es ampliamente usada y recomendada para el diagnóstico y el monitoreo de la infección y la enfermedad (en pacientes inmunocompetentes como inmunosuprimidos), para la detección de recaídas y para el diagnóstico de la infección en pacientes receptores de órganos. Así mismo, la medición de la carga viral ha sido propuesta para guiar intervenciones preventivas contra el desarrollo de la infección (Gärtner & Preiksaitis, 2010; Niller & Bauer, 2017). Una desventaja del método es que aún no ha sido totalmente estandarizado entre laboratorios resultando en inconsistencias entre laboratorios y datos discordantes y de esa manera limitando las capa-

ciudades de extrapolar datos publicados y definir un umbral de la carga viral que determine el diagnóstico y tratamiento de la infección (Preiksaitis et al., 2009).

Otro aspecto a considerar es la sensibilidad de la carga viral para el diagnóstico de la infección. Hecho que es evidente en aquellos pacientes inmunosuprimidos que, aún cuando se ha confirmado la infección por otro método, poseen una carga viral baja en sangre periférica. Se han encontrado diferencias entre la mayoría de pacientes que desarrollan una elevada carga viral anterior a presentar síntomas con aquellos pacientes que presentan una progresión rápida de la enfermedad y aparición rápida de síntomas antes de realizar la medición de la carga viral. En dichos pacientes se recomienda realizar medición periódicamente posterior a la primera medición (Nowalk & Green, 2016). Así mismo, aún se encuentra en debate e investigación a partir de qué fracción de la sangre periférica se deberían realizar las mediciones: linfocitos de sangre periférica y células mononucleadas de la sangre contienen las células B infectadas por el VEB, mientras que a partir del suero y el plasma se puede medir la presencia de ADN viral. Por su parte, el uso de sangre completa proporciona resultados similares a los obtenidos a partir de linfocitos de sangre periférica y células mononucleadas (Tsai et al., 2008; Wadovsky et al., 2003).

3.4.3 Histopatología

El diagnóstico definitivo (estándar de oro) de la infección y enfermedad por VEB se realiza a partir de la obtención de biopsia de las lesiones o tejidos afectados y es llevado a cabo por un médico patólogo. Así mismo, la histopatología permite clasificar y categorizar la enfermedad asociada y orientar el tratamiento y seguimiento del paciente. Se llevan a cabo estudios de inmunohistoquímica utilizando marcadores específicos como el CD20 y la proteína latente de membrana (LMP1) y también se lleva a cabo la hibridización *in situ* (ISH) para ARN (EBER1) codificado por el virus lo cual permite caracterizar el tipo de célula que se encuentra involucrada en la infección, ya sea célula T o célula B (Young et al., 1989; Nowalk & Green, 2016). Actualmente, la metodología manual de ISH para EBER1 es la más sensible para la detección del virus alcanzando una sensibilidad del 100%. En años recientes se han introducido metodologías automatizadas para llevar a cabo ISH las cuales cuentan con un desempeño de 94% sensibilidad, 69% especificidad y valores predictivos positivos y negativos de 79% y 90%, respectivamente, al ser comparada con la metodología manual (Fanaian et al., 2009).

3.5 Tipos y Clasificación del Virus de Epstein-Barr

Diversos estudios se han realizado con la finalidad de clasificar el VEB a partir de metodologías distintas. Inicialmente la clasificación del virus se abordó con el análisis del gen EBNA2 estableciendo dos genotipos Tipo 1 (VEB-1 o Tipo A) y Tipo 2 (VEB-2 o Tipo B) (Choi et al., 2018; Neves et al., 2016; Salahuddin et al., 2018). El gen EBNA2 posee una gran variación genética en comparación con los demás genes del VEB los cuales no llegan a alcanzar más del 5% de variación, siendo la homología del gen EBNA2 de un 54% de entre ambos tipos, demostrando una gran divergencia y variación genética (Smatti et al., 2018; Tzellos & Farrell, 2012). Los Tipo 1 y Tipo 2 VEB demuestran diferencias en sus capacidades de transformación de las células que infectan siendo el Tipo 1 el que cuenta con una mayor capacidad transformante de las células B. Ambos tipos 1 y 2 pueden subdividirse en cepas diferentes del virus según la etapa del ciclo de vida en que se encuentre. Ejemplo de ello es el paso de los linfocitos infectados por el VEB por el centro germinal del nodo linfático el cual posee un ambiente que propicia un aumento en las tasas de mutaciones del virus y de esa manera errores de replicación que dan lugar a variaciones genéticas (Neves et al., 2016; Walling et al., 2003). Finalmente, el estudio de genes como BZLF1, EBNA1, 2, 3, 3B, 3C, LMP1 y LMP2A han demostrado variación genética relacionada a las diferentes localidades geográficas de origen y a las enfermedades asociadas (Görzer et al., 2006; Wang et al., 2010).

Además del gen EBNA2 la proteína LMP1 presenta un alto nivel polimórfico que ha permitido realizar una clasificación del virus en base a las mutaciones y variaciones genéticas del gen. De tal manera, se realizó una clasificación que refleja el origen geográfico de las cepas (Mainou & Raab-Traub, 2006):

- Alaska (Ala)
- China (Ch1 y Ch2)
- Mediterráneo (Med+ y Med-)
- Carolina del Norte (CN)

Posteriormente, otros grupos de cepas fueron reportados y descritos: SEA1 y SEA2 provenientes de Tailandia (Smatti et al., 2018; Tzellos & Farrell, 2012). Así mismo, se ha reportado que las variantes del gen LMP1 correlacionan con las enfermedades asociadas como las siguientes: deleción de 30 pb se han encontrado en el cáncer nasofaríngeo y deleciones de 69 pb se han reportado en el Linfoma de Burkitt. Finalmente, estudios recientes sugieren una correlación entre el origen geográfico y el porcentaje de pacientes sanos con las deleciones de 30 pb siendo los resultados de un 62.5% en taiwaneses, 55.6% en polacos, 0% en árabes y 28% en argentinos (See et al., 2008).

Estudios han abordado la clasificación genómica del virus a partir del establecimiento de tres fa-

ses de latencia basadas en la expresión de los genes de latencia: tipo I de latencia (expresando *EBNA1* y *EBER*), tipo II de latencia (expresando *EBNA-LP*, *LMP1*, *LMP2A*, y *LMP2B*) y tipo III de latencia (expresando *EBNA2*, *EBNA3A*, *EBNA3B*, y *EBNA3C*) (Choi et al., 2018).

Recientemente, los estudios se han enfocado en clasificar el virus por medio del análisis del genoma, exoma y transcriptoma completo, el tipo de cáncer asociado, el país de procedencia y las características demográficas del paciente. Zanella y colaboradores (2019) proponen la clasificación del virus a partir del análisis del genoma completo obteniendo 12 poblaciones relacionadas a la localización geográfica y, tres de ellas, relacionadas con el tipo de cáncer: EBV-p1/Asia/CG, EBV-p2/Asia II/Tumores y EBV-p4/China/NPC (Zanella et al., 2019). Peng y colaboradores (2019) proponen una clasificación en la que los virus asociados con ENKTL se encuentran agrupados en China y Singapur, los asociados con NPC se encuentran agrupados en China, Japón y Hong Kong, y los asociados con linfoma de Burkitt agrupados en países africanos y americanos (Peng et al., 2019). Así mismo, Xiong y colaboradores (2020) proponen una clasificación del virus a partir de los mismos parámetros obteniendo grupos de cepas asiáticas asociadas a cáncer nasofaríngeo, cepas de China asociadas a ENKTL, cepas africanas asociadas a líneas celulares de linfoma espontáneo y cepas europeas asociadas a linfoma Hodgkin (Xiong et al., 2020).

3.6 Neoplasias hematológicas linfoides

Las neoplasias hematológicas linfoides consisten en un grupo de enfermedades que surgen de la expansión clonal de células hematopoyéticas. Las neoplasias de células B y T/NK son tumores clonales de células B maduras o inmaduras, células T o NK (Natural Killer o asesina natural) en diferentes estados de diferenciación. Las neoplasias de células T y NK se consideran juntas dentro de la misma clase por sus características funcionales e inmunofenotípicas similares (Morton et al., 2006). El sistema inmune cuenta con una respuesta innata y una respuesta adaptativa, las cuales se diferencian en la naturaleza del objetivo y el tipo de la respuesta inmune. Las células del sistema inmune innato representan la primera línea de defensa e incluye células como las NK, CD3+ CD56+ y células NK similares a célula T. Dichas células actúan como barrera defensiva en la inmunidad cutánea y de mucosas. La respuesta innata no requiere de un encuentro con el antígeno para iniciar la respuesta inmune. En cambio, el sistema inmune adaptativo, en el que participan células B y T, es específico para un patógeno determinado, siendo la memoria y la especificidad sus principales características (Campo et al., 2008).

Las neoplasias de células B se caracterizan por presentar un desarrollo y diferenciación normal de las células B. Inicialmente, la diferenciación de las células B implica que un precursor de células B sufra la recombinación de los genes *VDJ* de las inmunoglobulinas y posteriormente se lleve a cabo la maduración a célula B naive (también conocida virgen o inexperta). Posteriormente, al tener contacto con un antígeno que sea complementario a su receptor Ig, la célula B naive sufrirá una transformación, proliferación y, finalmente, una maduración hacia la secreción de anticuerpos y células B de memoria. Así mismo, las células B naive pueden madurar directamente produciendo los anticuerpos IgM tempranos (Hummel et al, 1994; Inghirami et al., 1991; Kipps, 1989).

3.6.1 Neoplasias de células T y NK maduras

Actualmente, se estima que VEB se encuentra asociado a más de 200,000 casos de cáncer al año y un 1.8% de todas las muertes por cáncer se atribuyen a neoplasias malignas causadas por el mismo (Young, Yap & Murray, 2016). La infección por VEB ha sido asociada con el desarrollo de diversos tipos de cáncer como el linfoma de Burkitt, linfoma T/NK, carcinoma nasofaríngeo (NPC, por sus siglas en inglés), aproximadamente el 30% de linfoma Hodgkin, el 10% de casos de cáncer gástrico (CG) y casos de linfoma difuso de células B grandes y leiomiomasarcoma (Palser et al., 2015). VEB puede estar asociado con diferentes desórdenes linfoproliferativos dependiendo el tipo de célula que infecta. Tal es el caso de los desórdenes linfoproliferativos asociados a la infección, transformación y proliferación de las células T y NK por VEB durante el estadio de latencia 2, también conocido como linfoma T/NK. Dicho tipo de linfomas tienen una ocurrencia principal en poblaciones asiáticas y de indígenas americanos del centro y sur de América (Kim et al., 2019; Kimura et al., 2012; Yamaguchi, Oguchi & SuzReino Unidoi, 2018). Los linfomas T/NK históricamente se han clasificado en formas indolentes y agresivas o fulminantes y se caracterizan clínicamente con diagnóstico a una edad media entre los 46 a 52 años, lesiones de masas nasales que a su vez causan obstrucción, secreciones nasales y epistaxis. Así mismo, la enfermedad puede diseminarse hacia la piel, el tracto gastrointestinal, testículos y ganglios linfáticos. A comparación con otros linfomas, el 40% de casos de linfoma T/NK presentan síntomas B (Kimura et al., 2012 ; Makita & Tobinai, 2017).

El linfoma de T/NK extranodal de tipo nasal (ENKTL) consiste en un linfoma extranodal caracterizado por el daño y destrucción vascular, la necrosis, fenotipo citotóxico y la asociación con la infección del VEB en el 100% de los casos. La mayoría de casos consisten en neoplasias de células NK pero algunos casos presentan un fenotipo citotóxico de células T (Campo et al., 2008). El ENKTL posee una prevalencia mayor en poblaciones asiáticas y de indígenas mexicanos, de Centro

América y Sur América (Au et al., 2005). La asociación con el VEB sugiere un rol patogénico del virus, el cual se presenta como un episoma clonal en el programa de latencia II, con una delección de 30 pb en el gen LMP1 y que, en la mayoría de veces, es del tipo 1 (Peh, Kim & Poppema, 2002). El ENKTL presenta una supervivencia entre el 10 – 40%, siendo una enfermedad con una alta mortalidad cuando se presentan estadios avanzados, mientras que los estadios tempranos de la enfermedad poseen un buen pronóstico (Hong et al., 2019; Yang et al., 2021).

3.6.2 Linfoma de Hodgkin

Los linfomas de Hodgkin (HL) presenta las siguientes características: se origina en los nodos linfáticos, la mayoría de ellos se manifiesta clínicamente en adultos jóvenes, los tejidos neoplásicos usualmente contienen un número pequeño de grandes células tumorales mono y multinucleadas y las células tumorales a menudo se encuentran anilladas por linfocitos T a manera de roseta (Anagnostopoulos et al., 2000). Los HL comprenden alrededor del 30% de todos los linfomas (Poppema, 2021).

Actualmente los HL se dividen en dos enfermedades distintas: Linfoma de Hodgkin con predominio linfocitario nodular y el Linfoma de Hodgkin Clásico. Ambos tipos de enfermedades se diferencian en sus características clínicas, en su composición celular, morfología, inmunofenotipo y en la preservación o extinción de la expresión de genes de células B (Anagnostopoulos et al., 2000). A su vez, el Linfoma de Hodgkin Clásico se divide en cuatro subtipos que son de esclerosis nodular, celularidad mixta, rico en linfocitos y de depleción linfocítica. Los cuatro subtipos se diferencian por los sitios en los que se desarrollan, las características clínicas, el patrón de crecimiento, la presencia de fibrosis, la composición celular, el número o grado de atipia de las células tumorales y la frecuencia de la infección por VEB (Nogová et al., 2008).

3.6.3 Neoplasias linfoides de células B pequeñas

Los linfomas de células B pequeñas se componen principalmente de linfocitos pequeños, a los cuales se les conoce como linfomas de células B de "bajo grado". La clasificación de la OMS intencionalmente no divide los linfomas por grado, sino que el nombre preferido que se usa es “linfomas de células B pequeñas” (Quintanilla-Martinez, 2017). Entre ellos se encuentran: el linfoma difuso de células grandes B, linfoma folicular, linfoma de células del manto, linfoma de Burkitt (BL), entre otros (Jaffe et al., 2008).

El linfoma Difuso de Células Grandes B (DLBCL, por sus siglas en inglés) consiste en una neoplasia de células B maduras que se caracteriza por células linfoides B con el tamaño nuclear igual o mayor a los núcleos de macrófagos normales o más del doble del tamaño de un linfocito normal y que posee un patrón de crecimiento difuso (Park et al., 2007). El DLBCL ha sido clasificado y subdividido en base a sus características morfológicas, moleculares e inmunofenotipo. 1) Los Subtipos del DLBCL son: linfoma de células B grandes rico en células T / histiocitos, DLBCL primario del sistema nervioso central, entre otros; 2) Otros DLBCL: LBCL ALK-positivo, linfoma plasmablastico, entre otros; 3) Categorías limítrofes que consisten en linfomas de células B no clasificables con características intermedias entre el DLBCL y el BL y el HL. No obstante, todos aquellos casos en los que no existe un criterio claro para ser subdivididos son clasificados como DLBCL, NOS (no está especificado de otra manera) (Campo et al., 2008). El DLBCL, NOS es una enfermedad que usualmente surge *de novo* pero también puede ser producto de la progresión o transformación de otros linfomas menos agresivos como linfoma folicular, entre otros. Los pacientes que sufren de dicha enfermedad la presentan en sitios ganglionares o extraganglionares, siendo un 40% de ellos los que se limitan a sitios extraganglionares (Campbell et al., 2007; Chung et al., 2007). Las variantes morfológicas más importantes del DLBCL, NOS son: variante centroblástica, variante inmunoblástica, la variante anaplásica y otras variantes morfológicas raras (Engelhard et al., 1997; Haralambieva et al., 2000; Harris et al., 1994). Así mismo, la enfermedad se divide por subgrupos moleculares (de centro germinal similar a célula B y similar a una célula B activada) y subgrupos inmunohistoquímicos (DLBCL CD5 positivo, centro germinal similar a una célula B y centro no germinal similar a una célula B) (Campo et al., 2008).

El linfoma folicular (LF) consiste en una enfermedad clínica y morfológicamente heterogénea, con anomalías citogenéticas y moleculares complejas. (Quintanilla-Martinez et al., 2016). El LF está precedido por una fase preclínica asintomática en la que las células B premalignas portadoras de la translocación t(14;18) acumulan alteraciones genéticas adicionales, no obstante no todas las células afectadas progresan a una fase tumoral. Los LF se caracterizan por presentar patrones foliculares (nodulares) de crecimiento, así como por diversos grados de crecimiento neoplásico difuso. (Carbone et al., 2019). Actualmente, el LF representa ~5% de todas las neoplasias hematológicas, siendo el segundo subtipo más común de los Linfoma no Hodgkin ya que representa entre el 20 a 25% de los nuevos diagnósticos en los países occidentales (Teras et al., 2016). El LF se clasifica por grados según el número de centroblastos por campo de microscopía de alta potencia. Siendo los grados 1, 2, 3A y 3B los reconocidos por la OMS (Campo et al., 2008). Mutaciones en los genes como KMT2D, IgHV, IgLV, RB1, CDK4, BCL6, H1-2, H1-4, MEF2B, EP300, BCL2 y EPHA7,

entre otras, se encuentran en más del 10% de los casos de FL (Carbone et al., 2019). Otras categorías han sido adoptadas por la OMS como la Neoplasia Folicular *in situ*, el LF tipo Pediátrico y el LF tipo Duodenal (Swerdlow et al., 2017).

El linfoma de células del manto (LCM) es un subtipo poco común de linfoma no Hodgkin con características clínicas, biológicas y moleculares distintivas. El LCM representa ~3 – 6% de los linfomas no Hodgkin, posee una incidencia anual de 0.5 por 100,000 habitantes en los países occidentales y la mediana de edad al diagnóstico es de 68 años, mostrándose un predominio de casos masculinos (3:1) (Cheah et al., 2016). El LCM se caracteriza por presentarse en estadios avanzados y de progresión rápida, siendo considerado una enfermedad agresiva. Actualmente se reconocen dos formas patogénicas, siendo el LCM clásico y el LCM no ganglionar leucémico. El LCM clásico se origina a partir de células B con IGHV no mutado y expresión de SOX11. Por su parte, el LCM no ganglionar leucémico se desarrolla a partir de células B SOX11-B mutadas por IGHV. Este último involucra principalmente el tejido óseo primario, la médula ósea y el bazo. Además, se han reconocido mutaciones secundarias en los genes TP53 y NOTCH 1/2 en una proporción pequeña de casos, las cuales dan lugar a un comportamiento agresivo de la enfermedad (Quintanilla-Martinez et al., 2016; Swerdlow et al., 2017).

El linfoma de Burkitt (BL) es un linfoma no Hodgkin de células B altamente agresivo. Se deriva de células B maduras germinales o post germinales del centro, las cuales presentan reordenamientos del oncogén MYC que contribuyen a la sobre expresión de c-myc (Saleh et al., 2020). Según la clasificación de la OMS, se han identificado tres entidades diferentes de BL con morfología, inmunofenotipo y características genéticas similares: endémica, esporádica (variante predominante diagnosticada en áreas no palúdicas) y relacionada con inmunodeficiencia (Jaffe et al., 2008). En general, el BL representa menos del 5% de los casos de linfoma en adultos; sin embargo, representa el 40 % de todos los linfomas no Hodgkin infantiles (Saleh et al., 2020). El tipo endémico ocurre principalmente en áreas de malaria en el continente africano y en la mayoría de casos se encuentra con la infección del VEB. El BL endémico presenta una prevalencia alta y una incidencia estimada en 4-5 por cada 100, 000 niños menores de 18 años por año en África ecuatorial (Magrath, 2012). Por su parte, la variante esporádica es predominantemente en Europa occidental, América del Norte y Asia oriental. Así mismo, se asocia con menos frecuencia (10-20%) a la infección por VEB. El BL esporádico representa del ~2% de los linfomas en adultos y del 30 – 40 % de los linfomas no Hodgkin pediátricos en América del Norte y Europa Occidental, presentando una incidencia anual de 2 a 3 casos por millón cada año (Mbulaiteye et al., 2009). Finalmente, el BL relacionado con inmunodeficiencia se observa en pacientes positivos a VIH (Saleh et al., 2020). Actualmente, la clasifica-

ción de la OMS distinguió tres neoplasias agresivas de células B que imitan al BL, estas son: 1) linfoma tipo Burkitt con aberración 11q, 2) linfoma de células B de alto grado con reordenamiento de MYC y BCL2 y/o BCL6 y 3) linfoma de células B de alto grado, no especificado de otra manera (Swerdlow et al., 2017).

3.6.4 Asociación de la infección del VEB con el desarrollo de neoplasias hematológicas linfoides en Latinoamérica

Recientemente un estudio llevado a cabo por Chabay y colaboradores (2020) presenta la asociación entre la infección del VEB con el desarrollo de neoplasias hematológicas linfoides en países latinoamericanos, haciendo un principal énfasis en el HL, linfoma de Burkitt (BL), linfoma Difuso de Células Grandes B y procesos linfoproliferativos de células T y NK. La asociación del VEB con el HL ha sido ampliamente reportada y descrita en la región latinoamericana. En la mayoría de países en desarrollo los casos de HL (con edades de 14 años o menores) presentan un alto grado de asociación con la infección por el VEB. Así mismo, en grupos menores a los cinco años hay un predominio de infección de hombres sobre mujeres mostrando una relación de 5:1, respectivamente. No obstante, en adolescentes y adultos jóvenes con HL la asociación con la infección de VEB es menor (Howlader et al., 2016). Ejemplo de ello es que datos de Argentina indican una infección del ~50% en pacientes pediátricos, mientras que en adultos la infección se presenta en un 31%. Así mismo, en Sao Paulo, Brasil, un estudio que revisó 817 casos de HL por 54 años reportó que los casos con infección positiva descendieron desde un 87% a un 46% durante el estudio y que en pacientes adultos la infección también descendió en adultos jóvenes desde un 85% a un 32% (Campos et al., 2018; De Matteo et al., 2003). Por el contrario, un estudio llevado a cabo en el norte de Brasil reportó un porcentaje de asociación con la infección de 87%. Dichos datos apoyan la idea que el nivel de desarrollo socio-económico es un factor que determina la presencia de la infección (Araujo et al., 2006). Un estudio mexicano con una cohorte de 42 casos pediátricos con una media de edad de 5 años reportó una asociación de la infección de un 76.1% (Palma et al., 2013). Estudios en México que incluyeron principalmente adultos jóvenes y de mayor edad han presentado porcentajes de la infección de un 61% y 77%, respectivamente (Dirnhofer et al., 1999;). Así mismo, estudios llevados a cabo en Perú, Colombia y Honduras confirman la asociación de la infección del virus con pacientes pediátricos de HL siendo los porcentajes de infección de 94%, 84% y 100%, respectivamente (Chang et al., 1993; Quijano et al., 2003; Quijano et al., 2004). Para Centro América y el Caribe se han realizado estudios en Puerto Rico, Costa Rica y la República Dominicana los cua-

les han reportado frecuencias del 50%, 40% y 64%, respectivamente; no obstante, dichos estudios cuentan con cohortes reducidas por lo que un mayor número de participantes podría presentar resultados diferentes (Aybar & Krause, 2005; Córdova-Pérez et al., 2003; Gulley et al., 1994; Monterroso et al., 1998; Sánchez-Giler et al., 2018; Weinreb et al., 1996).

El BL afecta principalmente a niños y es más predominante en hombres, teniendo una asociación con la infección por VEB del 100% para BL endémico y alrededor del 20% de casos esporádicos en Norte América y Europa (Crawford, 2001). En Argentina la mayor incidencia de casos VEB positivos se encuentra en niños menores de 5 años (Lara et al., 2014). Estudios realizados en Brasil apoyan la idea que los casos positivos de VEB son mayores en regiones en desarrollo social y económico (Klumb et al., 2004). Así mismo, un estudio llevado a cabo en Brasil reportó que la incidencia de la infección fue elevada para aquellos pacientes pediátricos con edades entre 1-4 años (de Oliveira Ferreira et al., 2012).

El Linfoma Difuso de Células Grandes B (DLBCL, por sus siglas en inglés) es una neoplasia altamente agresiva, rara en pacientes pediátricos y adultos jóvenes, asociada a la inmunosupresión y que se presenta en una edad avanzada (Campo et al., 2008). El DLBCL positivo a VEB a una edad de 50 años fue reconocido provisionalmente como un subtipo del DLBCL, teniendo una representación del 5-11% en países del Este de Asia y menor a 5% en poblaciones occidentales (Hoeller et al., 2010; Shimoyama et al., 2008). Posteriormente, la clasificación de Neoplasias Linfoides de la WHO 2016 eliminó el factor de la edad (Swerdlow et al., 2017). En Latinoamérica se presenta una asociación de la enfermedad con la infección ligeramente mayor. Reportes de países como México, Brasil, Argentina y Perú reportan una asociación de 7, 9, 13 y 14-28%, respectivamente (Beltran et al., 2011; Cohen et al., 2014; Cohen et al., 2017; Hofscheier et al., 2011).

La infección crónica activa por el VEB también es reconocida por la clasificación de Neoplasias Linfoides de la WHO 2016 para los desórdenes linfoproliferativos de células T y NK, entre ellos: ENKTL y el linfoma nodal de células T/NK nodal VEB-positivo (Kim et al., 2019). El mecanismo por el que el VEB infecta las células T y NK aún sigue sin entenderse del todo, no obstante existen reportes que sugieren una preferencia de infección hacia las células T de parte del subtipo VEB-2 (Coleman et al., 2018). El ENKTL consiste en un linfoma VEB positivo el cual ha sido descrito principalmente en el Este de Asia y en Latinoamérica, siendo la etnia de indígenas de América la que cuenta con la mayor proporción de casos. Países como México, Perú, Chile y Guatemala han reportado casos de la enfermedad en los que la infección ha estado presente en todos (Barrionuevo et al., 2007; Elenitoba-Johnson et al., 1998; Schwartz et al., 2008). Un estudio reciente en Brasil re-

portó una cohorte de 122 casos (mayoría pacientes adultos) de ENKTL, reportando características clínicas y patológicas similares a las reportadas en pacientes provenientes del Este de Asia y otros países de América. No obstante, datos étnicos no fueron evaluados en dicho estudio (Gualco et al., 2011). Finalmente, un estudio que incluyó 16 pacientes con linfoma de origen de células T/NK y 9 de origen de células B, no presentó diferencias significativas entre la etnia del paciente (Altemani et al., 2002).

3.7 Secuenciación de Nueva Generación y Análisis Bioinformático

3.7.1 Secuenciación de Nueva Generación

La secuenciación de nueva generación (NGS, por sus siglas en inglés) hace referencia a las tecnologías de secuenciación que permiten obtener las secuencias completas de genomas a costos menores. La NGS también es conocida como secuenciación de segunda generación al recopilar todos los métodos de secuenciación posteriores a la secuenciación de Sanger. Las metodologías de NGS pueden ser agrupadas en dos categorías principales: secuenciación por hibridación y secuenciación por síntesis (Slatko, Gardner & Ausubel, 2018; Yohe & Thyagarajan, 2017). La NGS cuenta con un gran potencial de aplicación en la práctica clínica, microbiología y en estudios filogenéticos. Entre las aplicaciones se pueden mencionar: la obtención de mutaciones y variantes a lo largo de un genoma o exoma completo, conocer la relación de las mutaciones y variantes con una enfermedad determinada, evaluar el efecto de la metilación, identificación y caracterización del genoma de patógenos, definir las relaciones filogenéticas de organismos, entre otras (Behjati & Tarpey, 2013; Harris et al., 2013). Los métodos de secuenciación con enriquecimiento objetivo son una alternativa que consisten en la captura de genes, regiones del genoma, exoma o genoma completo personalizado, con el objetivo de obtener una mayor resolución de lo secuenciado y para que la profundidad (cobertura) de la secuenciación obtenida sea mucho mayor. Así mismo, dichos métodos permiten al investigador identificar con mayor precisión SNPs e INDEL, así como otras mutaciones como variantes en el número de copias (Ballester et al., 2016; Kozarewa et al., 2015)

3.7.2 Bioinformática

La bioinformática es la ciencia que desarrolla y aplica herramientas de computación para el manejo de datos biológicos (Luscombe, Greenbaum & Gerstein, 2001). Actualmente la bioinformá-

tica es una ciencia aplicada que surge de la unión entre la biología y las ciencias de la computación con la finalidad de procesar y ordenar datos biológicos (ADN, ARN, proteínas, entre otros) y utiliza métodos y algoritmos matemáticos y estadísticos para analizar e interpretar dichos datos. La bioinformática es una disciplina con la que es posible realizar inferencias, hacer conexiones, asociaciones y realizar predicciones usando los datos biológicos disponibles (Lesk, 2019). La función proteica, que viene determinada por la estructura de las proteínas, y esta a su vez determinada por la secuencia de aminoácidos codificada por la secuencia de nucleótidos (ADN) permite a los métodos bioinformáticos realizar una cantidad diversa de estudios y análisis empleando datos biológicos. De tal manera, la bioinformática trabaja con secuencias de ADN (nucleótidos), secuencias de proteínas (aminoácidos), estructuras y motivos proteico y sus correspondientes mutaciones y variantes (Baxevanis, 2001).

3.7.3 Modelos de sustitución nucleotídica

Los modelos de sustitución nucleotídica se encargan de describir los procesos de evolución molecular que se llevan a cabo al estudiar un fenómeno evolutivo. Los modelos de sustitución nucleotídica se basan en la idea fundamental de que ciertas sustituciones de nucleótidos, codones o aminoácidos, se presentan de forma más frecuente que otras. De tal manera, los diferentes modelos de sustitución nucleotídica pueden suponer que los nucleótidos no se encuentran en la misma proporción en las secuencias, que las transiciones y transversiones ocurren con diferentes frecuencias o que la tasa de sustitución varía entre los diferentes sitios de las secuencias (Brinkamn & Leipe, 2001). Existe un gran número de modelos de sustitución nucleotídica que pueden ser asignados a set de datos dependiendo el análisis a realizar. Entre ellos, el JC que indica que transiciones y transversiones ocurren a iguales tasas; el modelo de dos parámetros de Kimura (1980; K2P) establece que la tasa de transiciones es mayor a la de transversiones: $\alpha + 2\beta$, donde α son transiciones y 2β transversiones (Kimura, 1979); el modelo de Felsenstein (1981; F81) que permite que las frecuencias relativas de los nucleótidos de las secuencias varían y que por lo tanto ciertas sustituciones tienen mayor probabilidad que otras (Felsenstein, 1981); el modelo GTR, uno de los más afinados, que permite frecuencias desiguales de nucleótidos en diferentes posiciones y que cada uno de los seis posibles cambios posee una tasa diferente: $A \leftrightarrow C$, $A \leftrightarrow G$, $A \leftrightarrow T$, $C \leftrightarrow G$, $C \leftrightarrow T$ y $G \leftrightarrow T$ (Rodríguez et al., 1990).

3.7.4 Métodos de Máxima Verosimilitud

Los análisis de relaciones filogenéticas parten del uso de métodos que permiten estimar y realizar inferencias filogenéticas a partir de un set de datos de los grupos de estudio. Los métodos más utilizados y que poseen un mejor rendimiento son los basados en los caracteres los cuales optimizan la distribución de los patrones de los datos reales para cada uno de los caracteres (nucleótidos o aminoácidos). En la actualidad los mejores métodos para la obtención de inferencias filogenéticas son los de Máxima Parsimonia, Neighbor Joining (NJ), métodos Bayesianos y de Máxima Verosimilitud (ML, por sus siglas en inglés) (Yoshida & Masatoshi, 2016). Los métodos de Máxima Verosimilitud emplean la especificación completa de los datos y un modelo probabilístico que describa los mismos. De tal manera, la probabilidad de observar los datos dado el modelo que se ha tomado cambia en función de los valores de parámetros escogidos en el modelo (White, 1982). La máxima verosimilitud busca la obtención de filogenias que maximicen la probabilidad de observar el set de datos dado un modelo de sustitución y la misma filogenia obtenida. Además, la máxima verosimilitud no trata las variables del modelo como aleatorias (Beerli, 2006).

Para llevar a cabo métodos de máxima verosimilitud se emplean cuatro pasos principales: 1) elección de un modelo evolutivo que permita explicar las variables de respuesta observadas en función de uno o varios predictivos; 2) estimación de los parámetros por máxima verosimilitud, lo cual permite maximizar la probabilidad de que todos y cada uno de los valores observados para las variables de respuesta se encuentren dentro de una función de distribución de errores definida a priori y centrada en el valor predicho por el modelo previamente seleccionado; 3) comparación de modelos, frecuentemente se utiliza el criterio de información de Akaike (AIC, por sus siglas en inglés) y 4) evaluación del mejor o mejores modelos (Gómez-Aparicio, Ávila & Cayuela, 2013).

3.7.5 Análisis de Estructura Poblacional y Redes de Haplotipos

El análisis de estructura poblacional permite conocer si una determinada población se encuentra dividida y de esa manera permite identificar subpoblaciones de la misma. Una de las principales importancias de estudiar la estructura poblacional es el hecho que las subpoblaciones pueden llegar a tener la capacidad de evolucionar independientemente confiriendo características diferentes a los individuos que conforman cada una de las subpoblaciones (Andam et al., 2017). Diversos procesos evolutivos moldean la estructura poblacional de los organismos, procesos como la recombinación genética y las mutaciones parecen estar estrechamente relacionados con la estructura poblacio-

nal. Las metodologías de NGS permiten la obtención de un gran número de genomas y a partir de su análisis se puede obtener información sobre la estructura poblacional. Los análisis de estructura poblacional permite evaluar la posible variación que presentan los grupos de alelos (o haplotipos) y las distintas frecuencias alélicas con las distribuciones geográficas (Chiara et al., 2016; Hubisz et al., 2009). Los estudios de estructura poblacional basados en datos genómicos proveen de una mayor sensibilidad para la detección de genotipos raros y poseen una mayor resolución de las relaciones filogenéticas (Meier-Kolthoff et al., 2013). Programas como STRUCTURE y BAPS (y sus modificaciones como hierBAPS) permiten realizar inferencias acerca del número de subpoblaciones presentes en una muestra de poblaciones bacterianas o virales. Dichos métodos permiten realizar clústeres jerárquicos a partir de información genómica o de secuencias específicas de ADN (Corander & Marttinen, 2006; Falush, Stephens & Pritchard, 2007).

Las redes de haplotipos son análisis que cuentan con la ventaja de poder describir las relaciones intraespecíficas en una red y las relaciones reticuladas entre haplotipos, no desde un punto de vista jerárquico. Además, a partir de las redes de haplotipos se obtiene gráficamente la frecuencia de cada uno de los haplotipos estudiados permitiendo inferir aquellos haplotipos faltantes (Posada & Crandall, 2001; Templeton, Crandall & Sing, 1992). Entre las desventajas y limitantes que presentan las redes de haplotipos destaca el no poder visualizar distancias genéticas de forma gráfica, no se describen relaciones evolutivas y los límites de confianza son sensibles al número de muestras y sus respectivas distribuciones geográficas (Excoffier et al., 2007; Muwakina et al., 2003; Vázquez-Domínguez et al., 2009). Las conexiones entre haplotipos resultantes representan eventos de coalescencia y a partir de ellos se pueden realizar las siguientes predicciones: 1) haplotipos de mayor frecuencia tienen mayor probabilidad de ser los ancestrales, 2) haplotipos con múltiples conexiones probablemente son los ancestrales, 3) los haplotipos ancestrales se encuentran en el interior de una red y los haplotipos nuevos en la periferia, 4) se espera que los haplotipos ancestrales presenten una distribución geográfica amplia, y 5) los haplotipos que poseen una única conexión estarán unidos a haplotipos de la misma población (Clement, Posada & Crandall, 2000). Las redes de haplotipos se pueden clasificar a partir del método de construcción (distancias, parsimonia, máxima verosimilitud o inferencia bayesiana) y por el objetivo principal por el que se realiza la red: observar procesos evolutivos como la hibridación, transferencia horizontal de genes o recombinación, entre otros (Paradis, 2018).

4. JUSTIFICACIÓN

El VEB es uno de los principales factores asociados al desarrollo de cáncer de distintos tipos a nivel mundial, entre ellos el ENKTL. Actualmente, los estudios de VEB buscan conocer cuáles son las relaciones entre sus características genéticas y las características demográficas de los pacientes con el desarrollo de diferentes procesos linfoproliferativos. Así mismo, los estudios sobre la clasificación del virus se enfocan a partir del análisis del genoma. En Guatemala son poco conocidos muchos aspectos importantes sobre la relación entre la infección por VEB y el desarrollo de procesos linfoproliferativos de células T y NK, entre ellos la caracterización del genoma y sus relaciones filogenómicas. Por ello, es importante caracterizar el genoma del VEB asociado a ENKTL en Guatemala, establecer la estructura poblacional del VEB en Guatemala y describir sus relaciones filogenómicas con otras cepas del mundo en base a los sistemas de clasificación del virus más recientes que han sido publicados a nivel mundial.

5. OBJETIVOS

5.1 Objetivo General

Determinar la clasificación filogenómica y estructura poblacional del Virus de Epstein-Barr asociado a linfoma T/NK extranodal de tipo nasal (ENKTL) en pacientes del Instituto de Cancerología (INCAN), Guatemala.

5.2 Objetivos Específicos

5.2.1 Caracterizar el genoma del Virus de Epstein-Barr asociado a linfoma T/NK extranodal de tipo nasal (ENKTL) en pacientes del Instituto de Cancerología (INCAN), Guatemala

5.2.2 Describir las relaciones filogenómicas entre el Virus de Epstein-Barr asociado a linfoma T/NK extranodal de tipo nasal (ENKTL) en pacientes del Instituto de Cancerología (INCAN), Guatemala, con cepas del virus asociadas a otros tipos de cáncer y provenientes de otros países.

5.2.3 Establecer la estructura poblacional entre el Virus de Epstein-Barr asociado a linfoma T/NK extranodal de tipo nasal (ENKTL) en pacientes del Instituto de Cancerología (INCAN), Guatemala, y compararla con cepas del virus asociadas a otros tipos de cáncer y provenientes de otros países.

6. HIPÓTESIS

Al ser una investigación descriptiva no se considera la formulación de una hipótesis de investigación.

7. MATERIALES Y MÉTODOS

7.1 Tipo de Estudio

Estudio de tipo descriptivo, transversal, observacional y retrospectivo.

7.2 Universo

7.2.1 Población

Pacientes con Diagnóstico de Linfoma de T/NK Extranodal de Tipo Nasal (ENKTL) de Guatemala.

7.2.2 Muestra

Pacientes con Diagnóstico de Linfoma de T/NK Extranodal de Tipo Nasal (ENKLT) de Guatemala que acudieron al Instituto de Cancerología (INCAN).

22 muestras de biopsia de tejido tumoral de linfoma T/NK Extranodal de Tipo Nasal (ENKLT) embebido en bloques de parafinas que fueron obtenidas en el Instituto de Cancerología (INCAN) fueron recolectadas entre los años 2008 a 2019. En la Tabla 2 se detallan las características de las muestras.

El muestreo se realizó por conveniencia debido a los costos de secuenciación y a la accesibilidad de muestras con tejido en óptimo estado.

Tabla 2. Características demográficas de pacientes con diagnóstico de linfoma de T/NK extranodal de tipo nasal (ENKTL).

No. de Estudio	Edad al Diagnóstico	Sexo	Etnia	Departamento de Procedencia
EBV00079	53	Femenino	Mestizo	Totonicapán
EBV00080	59	Masculino	Maya	Jutiapa
EBV00081	60	Masculino	Mestizo	Guatemala
EBV00082	Sin Dato	Sin Dato	Sin Dato	Sin Dato
EBV00083	24	Femenino	Mestizo	Quetzaltenango
EBV00084	58	Masculino	Maya	Totonicapán
EBV00085	43	Masculino	Sin Dato	Guatemala
EBV00086	26	Masculino	Maya	Guatemala
EBV00087	61	Femenino	Mestizo	Retalhuleu
EBV00088	53	Femenino	Maya	San Marcos
EBV00089	38	Masculino	Mestizo	Guatemala

EBV00090	53	Femenino	Maya	Quiché
EBV00105	67	Masculino	Sin Dato	Quetzaltenango
EBV00106	36	Masculino	Maya	Alta Verapaz
EBV00107	38	Masculino	Mestizo	El Progreso
EBV00108	24	Masculino	Mestizo	Huehuetenango
EBV00109	19	Femenino	Maya	Sololá
EBV00110	Sin Dato	Sin Dato	Sin Dato	Sin Dato
EBV00111	75	Masculino	Maya	Baja Verapaz
EBV00112	37	Masculino	Maya	Chimaltenango
EBV00113	74	Masculino	Maya	Chiquimula
EBV00114	64	Femenino	Maya	Guatemala

Fuente: Instituto de Cancerología (INCAN), 2021.

7.3 Materiales

7.3.1 Equipo

- Computadora personal Dell Inspiron 15 5000 Series
- Computadora personal Dell Latitude E 7240
- Material de oficina

7.3.2 Software

- FastQC Version 0.11.9
- Trimmomatic Galaxy Version 0.38.0
- Bowtie2 – map reads against reference genome Galaxy Version 2.3.4.3+galaxy0
- CollectInsertSizeMetrics plots distribution of insert sizes Galaxy Version 2.18.2.1
- Map with BWA-MEM Galaxy Version 0.7.17.1
- Unipro UGENE version 36.0 64-bit version
- bcftools mpileup (Galaxy Version 1.4.0.0)
- bcftools call (Galaxy Version 1.4.0)
- bcftools counts (Galaxy Version 1.4.0)
- GenomeVx
- Atom version 1.54.0 x64
- Sublime Text Version 3.2.2, Build 3211

- MAFFT version 7
- MEGA-X (Molecular Evolutionary Genetic Analysis) version 10.1.8
- Gblocks Version 0.9.1
- jMOTest 2.1.10 v20160303
- SMS Smart Model Selection in PhyML
- raxmlGUI 2.0
- PhyML 3.0
- FigTree v1.4.4
- iTOL 5.7
- Paquete rhierbaps
- R version 4.0.2
- DnaSP (DNA Sequence Polymorphism) Version 6.12
- PopART version 1.7
- RDP4 (Recombination Detection Program Version 4)
- LibreOffice 7.0

7.3.3 Base de datos

- National Center for Biotechnology Information

7.4 Métodos

7.4.1 Secuenciación

22 muestras de biopsia de tejido tumoral de ENKTL embebido en bloques de parafinas fueron enviadas a Novogene Corporation Inc., Ohio, Estados Unidos, en donde se realizó la extracción de ADN utilizando el kit de QIAamp DNA FFPE Tissue Kit (Catálogo 56404). Se llevó a cabo la secuenciación de nueva generación del VEB por medio de un método de enriquecimiento objetivo para VEB. A partir de un total de 1.0 µg de ADN genómico se prepararon las librerías de secuenciación. Las librerías se generaron usando el kit de Agilent SureSelect Human All Exon kit (Agilent Technologies, CA, USA), la fragmentación se llevó a cabo mediante un sistema de cizallamiento hidrodinámico (Covaris, Massachusetts, USA), se realizó la adenilización de los extremos 3' de los fragmentos de ADN y se unieron adaptadores. Seguidamente se llevó a cabo un paso de enriquecimiento selectivo para Virus de Epstein-Barr por medio de una reacción de PCR para preparar el paso de la hibridización. Los productos fueron purificados usando AMPure XP system (Beckam Coulter, Beverly, USA) y cuantificados con el sistema Agilent Bioanalyzer 2100.

Las librerías fueron secuenciadas en una plataforma Illumina generando lecturas paired-end con un tamaño máximo de 150 pb.

7.4.2 Preprocesamiento y análisis de datos de secuenciación

Las lecturas paired-end obtenidas pasaron por un flujo de trabajo para el procesamiento y análisis de datos (Bao et al., 2014; Gallo et al., 2016). Se evaluó la calidad de las secuencias obtenidas por medio de un control de calidad usando el programa FastQC Version 0.11.9 (Andrews, 2012; <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). La calidad de las secuencias se evaluó por medio de los factores de: 1) calidad de las bases por secuencia, 2) calidad de sector por secuencia, 3) calidad de las puntuaciones por secuencia, 4) contenido de bases por secuencia, 5) contenido de CG por secuencia, 6) contenido de N por base, 7) distribución de la longitud de las secuencias, 8) nivel de duplicación de las secuencias, 9) secuencias sobrerrepresentadas y 10) contenido de adaptadores. El preprocesamiento de las secuencias se realizó con el programa Trimmomatic Galaxy Version 0.38.0 mediante el cual se eliminaron las lecturas y secuencias de baja calidad menores a 30 (Afgan et al., 2016; Bolger, Lohse & Usadel, 2014; <https://usegalaxy.org/>). Se utilizó el programa Bowtie2 – map reads against reference genome Galaxy Version 2.3.4.3+galaxy0 (Langmead & Salzberg, 2012) para obtener un archivo de salida BAM el cual fue utilizado en el programa CollectInsertSizeMetrics plots distribution of insert sizes Galaxy Version 2.18.2.1 (Broad Institute, 2020) con la finalidad de obtener el parámetro de la media de las secuencias paired-end. Se usó el programa Map with BWA-MEM Galaxy Version 0.7.17.1 (Li & Durbin, 2009) para obtener un ensamblaje de novo a partir de las lecturas paired-end por cada muestra, utilizando el genoma de referencia de VEB Human gammaherpesvirus 4, complete genome (GenBank NC_007605.1; https://www.ncbi.nlm.nih.gov/nuccore/NC_007605) y el parámetro de la media de las secuencias paired-end, previamente obtenido. El ensamblaje de novo obtenido fue visualizado con el software Unipro UGENE version 36.0 64-bit version (Okonechnikov et al., 2012).

7.4.3 Caracterización del genoma de VEB de Guatemala

Para caracterizar el genoma de VEB de Guatemala se llevó a cabo una representación por medio del mapeo de las regiones del mismo usando el software GenomeVx (Conan & Wolfe, 2008; <http://wolfe.ucd.ie/GenomeVx/>) y Unipro UGENE version 36.0 64-bit version. Se usó de guía el genoma de referencia Human gammaherpesvirus 4, complete genome (GenBank NC_007605.1; https://www.ncbi.nlm.nih.gov/nuccore/NC_007605).

7.4.3.1 Análisis de variantes: SNP/INDEL

Para llevar a cabo el análisis de variantes: SNP/INDEL, se utilizaron los paquetes de VCF/BCF: bcftools mpileup (Galaxy Version 1.4.0.0), bcftools call (Galaxy Version 1.4.0), bcftools counts (Galaxy Version 1.4.0), bcftools stats (Galaxy Versión 1.10) y VCFfilter (Galaxy Version 1.0.0). El paquete de bcftools mpileup se utilizó para generar probabilidades de genotipo que contengan VCF (formato de llamada de variantes) o BCF (formato de variantes binarias) para uno o varios archivos de alineación (BAM o CRAM); el paquete bcftools call se utilizó para obtener los SNP (Polimorfismo de Nucleótido Único) y los INDEL (Inserción / Delección) a partir de los VCF o BCF generados previamente; VCFfilter fue utilizado para filtrar aquellos SNP e INDEL con valores de calidad menores a 220; y bcftools stats se utilizó para obtener el número de transiciones y tranversiones. Finalmente el paquete bcftools counts se utilizó para obtener un resumen de los SNP, INDEL y sitios totales analizados (Danecel et al., 2021; Garrison et al., 2021; Li, 2011).

7.4.4 Análisis filogenético

Para llevar a cabo un análisis de la clasificación filogenómica de VEB se realizaron análisis filogenéticos haciendo uso del genoma y tomando en cuenta el tipo de cáncer asociado y país de procedencia. Las secuencias del genoma de VEB asociado a ENKTL de Guatemala junto a secuencias obtenidas en National Center for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov/>) correspondientes a otros países y regiones del mundo y otros tipos de cáncer (ver Anexo 1) fueron alineadas a un genoma de referencia Human gammaherpesvirus 4, complete genome (GenBank NC_007605.1; https://www.ncbi.nlm.nih.gov/nuccore/NC_007605) utilizando el software MAFFT version 7 (Katoh & Standley, 2013; <https://mafft.cbrc.jp/alignment/software/>). Los nombres y cabezales de las secuencias fueron visualizados y editados a conveniencia en los programas Atom version 1.54.0 x64 (Atom, 2021) y Sublime Text Version 3.2.2 (Sublime, 2019). Los alineamientos generados fueron visualizados en el software MEGA-X (Molecular Evolutionary Genetic Analysis) version 10.1.8 (Kumar et al., 2018). Posteriormente, se eliminaron los sitios pobremente alineados usando el software Gblocks Version 0.91b (Castresana, 2000; molevol.cmima.csic.es/castresana/Gblocks.html). Se calculó el mejor modelo de sustitución nucleotídica para el alineamiento generado usando los programas jModeltest 2.1.10 v20160303 (<https://github.com/ddarriba/jmodeltest2/releases/tag/v2.1.10r20160303>) y SMS Smart Model Selection in PhyML (Lefort, Longueville & Gascuel, 2017). Se llevó a cabo el análisis de inferencia filogenética mediante el método de Maxima Verosimilitud (ML, por sus siglas en inglés) usando dos software. El primer software raxmlGUI 2.0 fue utilizado con un análisis de ML + rapid

bootstrap, parámetros de 1000 réplicas de Bootstrap, con el modelo de sustitución nucleotídica obtenido anteriormente y con los demás valores pre definidos. El segundo software PhyML 3.0 (Guindon et al., 2010) fue utilizado con un análisis de aLRT SH-like, con el modelo de sustitución nucleotídica obtenido anteriormente y con los demás valores pre definidos. El uso de ambos métodos permitió confirmar los resultados obtenidos. Las filogenias obtenidas fueron visualizadas en FigTree v1.4.4 (Rambaut & Drummond, 2018; <http://tree.bio.ed.ac.uk/software/figtree/>) y editadas para su mejor visualización y presentación con iTOL (Letunic & Bork, 2016; <https://itol.embl.de/>). Usando iTOL se generaron los sets de datos de país y enfermedad asociada para ser visualizados en cada filogenia y se visualizaron los valores de Bootstrap en los nodos de las filogenias obtenidas. Para las filogenias obtenidas a partir de raxmlGUI 2.0 se visualizaron los valores de Bootstrap ≥ 70 , mientras que en las filogenias obtenidas por PhyML 3.0 se visualizaron los valores de Bootstrap en una escala de 0 a 1.

7.4.5 Análisis de estructura poblacional

Se llevó a cabo un análisis de estructura poblacional con el paquete rhierbaps usando el software R version 4.0.2 (Cheng et al., 2013; <https://cran.r-project.org/web/packages/rhierbaps/vignettes/introduction.html>; R Core Team, 2020; Tonkin-Hill et al., 2018). También se utilizaron los paquetes de ggtree (Yu et al., 2016), phytools (Revell, 2012) y ape (Paradis & Shliep, 2019). El análisis se llevó a cabo a partir del alineamiento de genomas completos (tomando en cuenta el tipo de cáncer asociado y país de procedencia) y ambos resultados de los análisis filogenéticos previamente obtenidos. El análisis de estructura poblacional se llevó a cabo con dos niveles de jerarquía para la identificación de filo poblaciones y sub-filo poblaciones del VEB. El script de R utilizado para llevar a cabo el análisis de estructura poblacional con rhierbaps se puede observar en el Anexo 2. Los resultados del análisis de estructura poblacional fueron editados para una mejor visualización en el software iTOL.

7.4.6 Análisis de recombinación genética

Se realizó un análisis para identificar los eventos de recombinación presentes en los genomas del VEB de Guatemala y aquellos que se encuentren distribuidos a lo largo de los demás genomas del VEB de otros países utilizados. Se usó el software RDP4 (Recombination Detection Program Version 4) (Martin et al., 2015). El análisis de recombinación se llevó a cabo a partir de 5 metodologías:

- RDP4 (Martin et al., 2015)

- Chimaera (Posada & Crandall, 2001)
- GENECONV (Padidam, Sawyer & Fauquet, 1999)
- MaxChi (Smith, 1992)
- SisScan (Gibbs, Armstrong & Gibbs, 2000)
- 3Seq (Boni, Posada & Feldman, 2007)

Posterior a la identificación de los eventos de recombinación se obtuvo un nuevo alineamiento en el cual fueron eliminadas las regiones recombinantes. A partir de dicho alineamiento se llevó a cabo un nuevo análisis filogenético por raxmlGUI 2.0 y PhyML 3.0 y el posterior análisis de estructura poblacional con la finalidad de evaluar si los eventos de recombinación poseen un mayor efecto en la clasificación filogenómica del VEB obtenida. Se usaron ambos métodos de inferencia filogenética para confirmar resultados.

7.4.7 Análisis de red de haplotipos

Se realizó un análisis de red de haplotipos a partir de las frecuencias de los haplotipos de la región OriP del genoma viral utilizando el software PopART version 1.7 (Leigh & Bryant, 2015). Las frecuencias de la región OriP fueron obtenidas en el software DnaSP (DNA Sequence Polymorphism) Version 6.12 (Rozas et al., 2017). Así mismo, se realizó una prueba de neutralidad para evaluar si la región OriP se encontraba sujeta a selección. Para ello se utilizó el software DnaSP (DNA Sequence Polymorphism) Version 6.12.

Usando LibreOffice Calc se generó una matriz a partir de las frecuencias de los haplotipos obtenidos, el país de origen y la enfermedad asociada. Con dicha matriz se llevó a cabo el análisis de red de haplotipos en PopART usando el análisis de Minimum Spanning Network. Los resultados fueron visualizados e interpretados en el mismo programa.

8. RESULTADOS

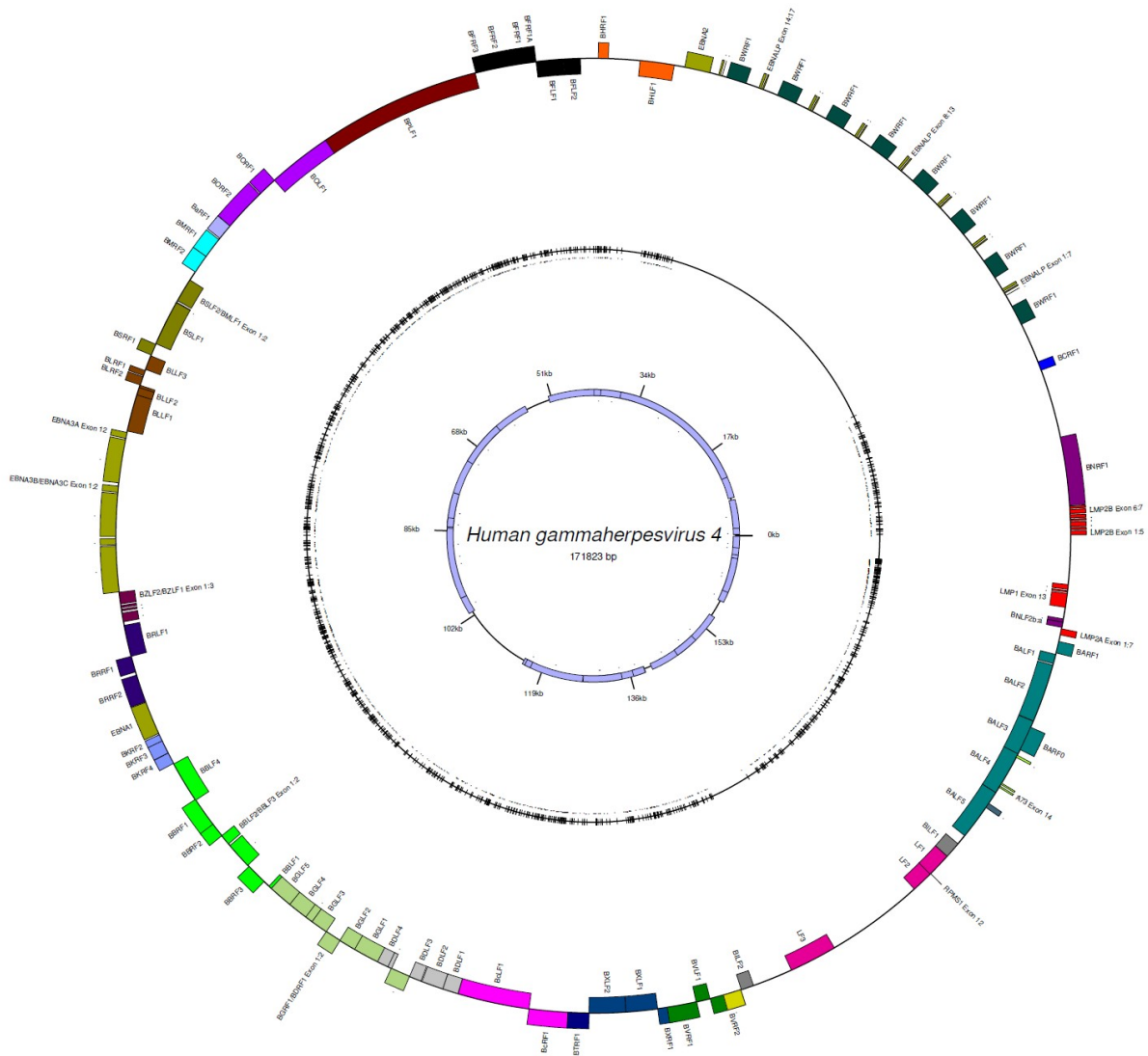


Figura 1. Representación del genoma del VEB asociado a ENKTL. El anillo exterior muestra genes y exones. El anillo medio muestra las variantes identificadas en color negro. El anillo interior muestra los eventos de recombinación identificados en color celeste.

En la Figura 1 se presenta una representación del genoma del VEB asociado a ENKTL en Guatemala. Se obtuvo un genoma de 171,823 pb producto del ensamblaje de novo a partir de las lecturas paired-end por cada muestra. Se identificaron 81 genes y 56 exones. Además, se identificaron 30 regiones de recombinación a lo largo del genoma del virus. 1,127 variantes que consisten en sitios SNP e INDEL fueron obtenidas a partir del análisis de variantes VCF/BCF.

La Tabla 3 presenta el resumen del análisis de variantes realizado. Un total de 1,492 variantes fueron identificadas al no tomar ningún valor de filtro de calidad. De las 1,492 variantes, 1,432 representaron sitios SNP y 62 sitios INDEL. Al aplicar filtros de calidad al análisis de variantes con valores de 200 y 220 el número de variantes disminuyó, observándose un mayor número de SNPs y sitios INDEL con un valor de 200 que con un valor de 220.

Tabla 3. Resumen del Análisis de variantes: Sitios SNP e INDEL.

Valor de filtro de variantes	Sitios SNP	Sitios INDEL	Total de Sitios
Sin valor de filtro	1432	62	1494
200	1353	26	1379
220	1127	15	1142

El resumen de los tipos de sustituciones obtenidas se presenta en la Tabla 4. De los 1,127 SNP obtenidos a partir del análisis de variantes realizado con un filtro de 220 se obtuvo un mayor número de transiciones que de transversiones. El detalle de las 1,142 variantes obtenidas se encuentra en el Anexo 3. Los genes con mayor número de variantes fueron EBNA1, EBNA3B3C, EBNA3A, LMP2A, BOLF1, RPMS1 y BPLF1. El listado completo se presenta en el Anexo 4.

Tabla 4. Tipos de sustituciones.

Tipo de sustitución	Sitios SNP
A>C	58
A>G	167
A>T	21
C>A	68
C>G	40
C>T	166
G>A	196
G>C	34
G>T	80
T>A	24
T>C	193
T>G	80
Transiciones	722
Transversiones	405

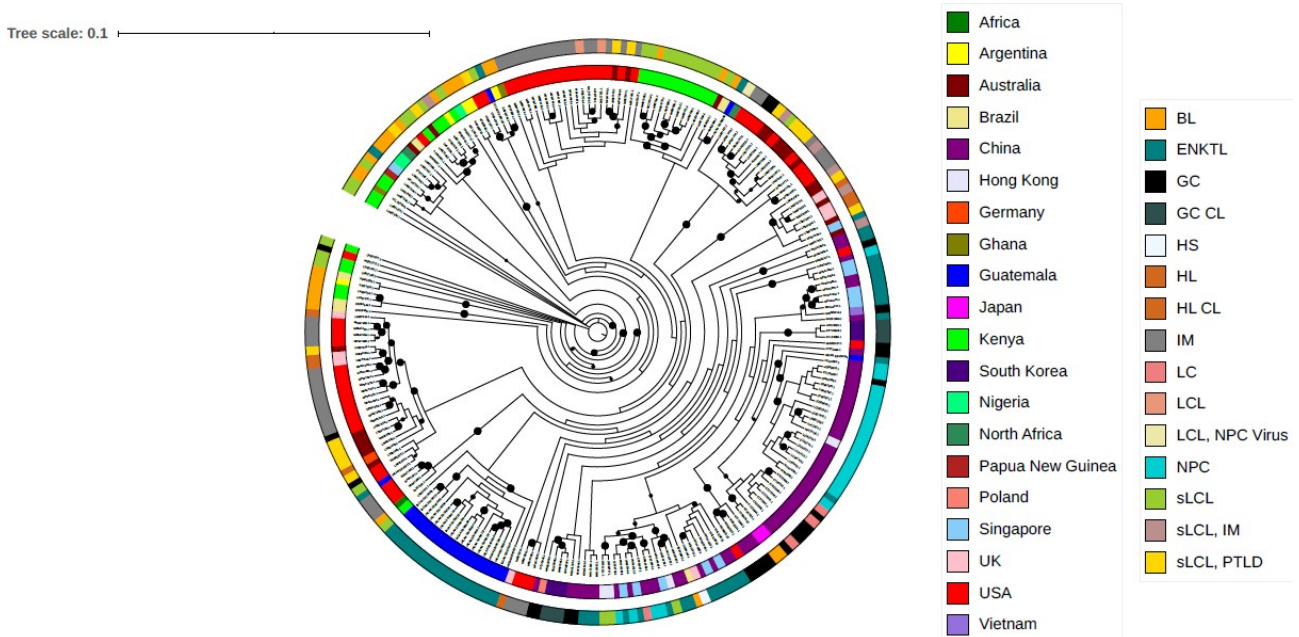


Figura 2. Relaciones filogenómicas del Virus de Epstein-Barr asociado a diferentes tipos de cáncer y procedente de diferentes países del mundo. El anillo interior indica el país de procedencia y el anillo exterior el tipo de cáncer asociado. Se indican los valores de bootstrap >70 (escala 0-100), correspondientes a las ramas. Resultado obtenido a partir de raxmlGUI 2.0. (Ver anexo 2 para abreviaturas).

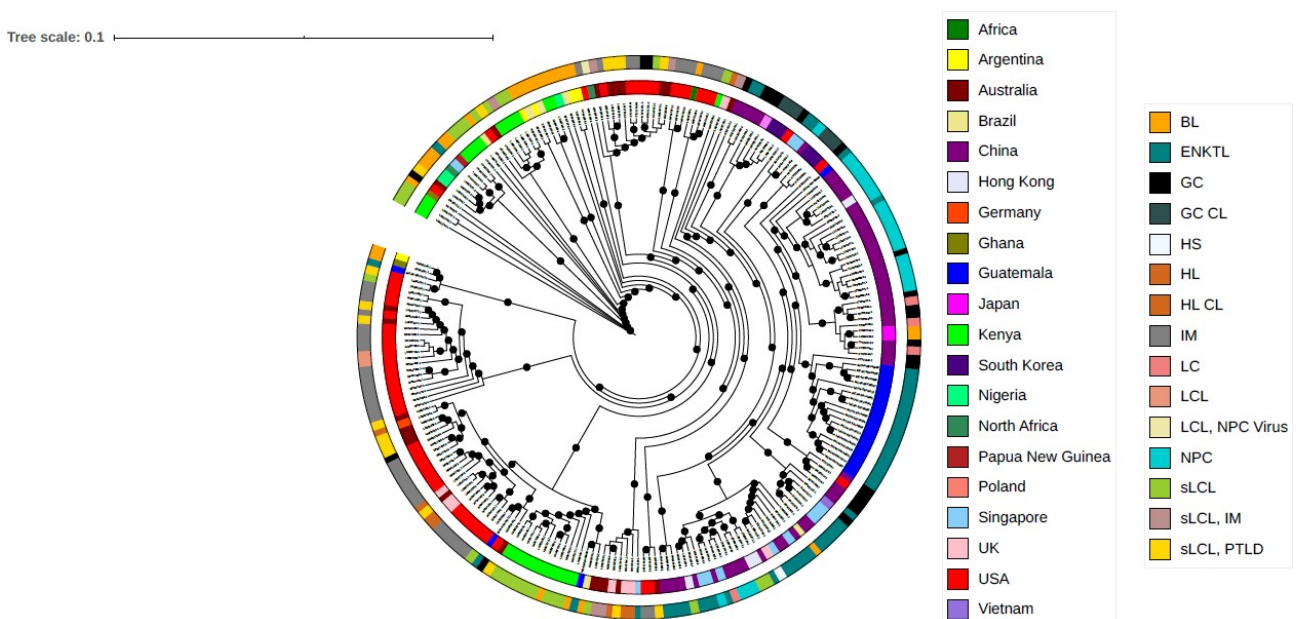


Figura 3. Relaciones filogenómicas del Virus de Epstein-Barr asociado a diferentes tipos de cáncer y procedente de diferentes países del mundo. El anillo interior indica el país de procedencia y el anillo exterior el tipo de cáncer asociado. Se indican los valores de bootstrap de 1 (escala 0-1),

correspondientes a las ramas. Resultado obtenido a partir de PhyML 3.0. (Ver anexo 2 para abreviaturas).

Las relaciones filogenómicas del VEB presentadas en las Figuras 2 y 3, demuestran una relación entre las cepas guatemaltecas asociadas a ENKTL con cepas de países asiáticos como China y Singapur asociadas a ENKTL, NPC y CG, lo que sugiere que dichas cepas poseen características genéticas asociadas al desarrollo de las enfermedades mencionadas. De la misma manera, se obtuvieron relaciones filogenéticas con otras cepas de Estados Unidos o Kenia. Por otro lado, cepas asociadas a BL o IM y provenientes de otras regiones se encuentran relacionadas filogenéticamente entre ellas. Los resultados obtenidos son consistentes a partir de ambas metodologías empleadas. Es importante destacar que a pesar de que algunas ramas no poseen un buen soporte en el resultado obtenido con raxmlGUI 2.0 (valores <70 no graficados), la mayoría de las relaciones filogenómicas obtenidas con PhyML 3.0 poseen un soporte de ramas aceptable en la escala de 0 a 1 empleada.

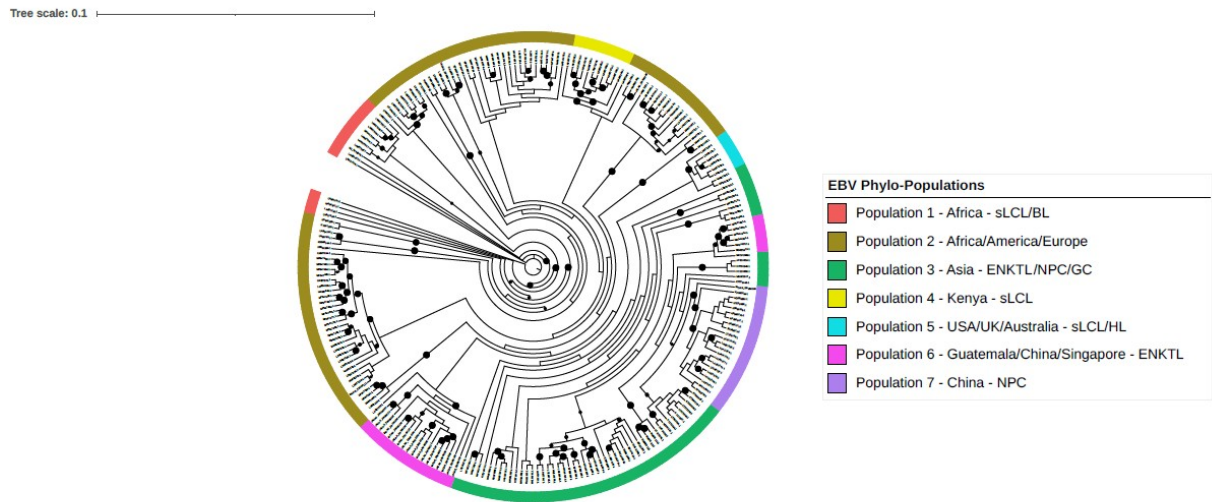


Figura 4. Filo poblaciones del VEB con un nivel de jerarquía. Resultado obtenido a partir de raxmlGUI 2.0. Se indican los valores de bootstrap >70 (escala 0-100), correspondientes a las ramas. El anillo indica cada una de las siete Filo Poblaciones identificadas.

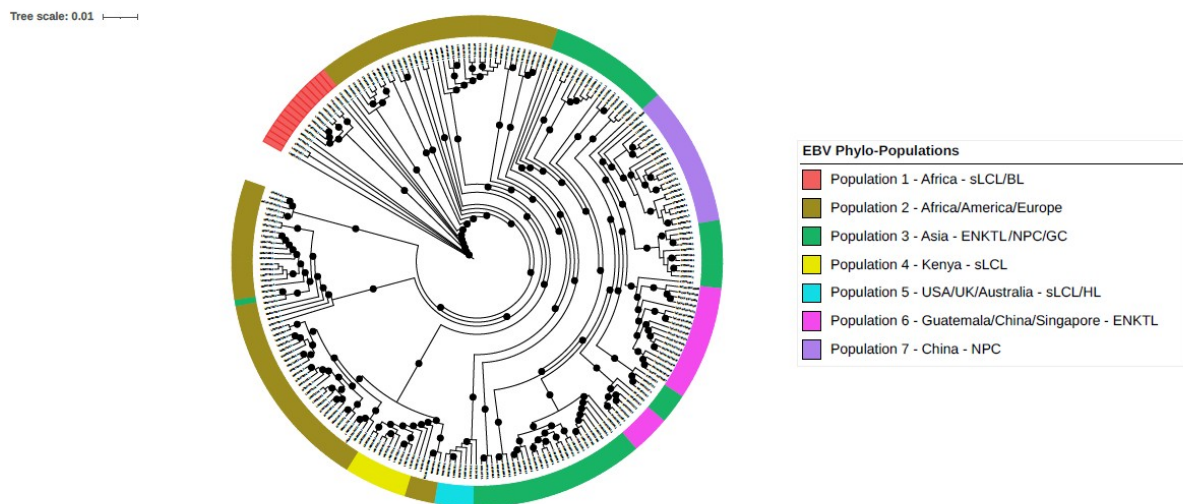


Figura 5. Filo poblaciones del VEB con un nivel de jerarquía. Resultado obtenido a partir de PhyML 3.0. Se indican los valores de bootstrap de 1 (escala 0-1), correspondientes a las ramas. El anillo indica cada una de las siete Filo Poblaciones identificadas.

En las Figuras 4 y 5 se presentan las 7 Filo Poblaciones del VEB propuestas a partir del análisis de estructura poblacional con un nivel de jerarquía. La Filo Población 6 se encuentra formada por cepas de Guatemala, China y Singapur asociadas a ENKTL, lo que sugiere que dichas cepas poseen características genéticas asociadas al desarrollo de ENKTL. Resultados consistentes fueron

obtenidos por el análisis de estructura poblacional a partir del análisis filogenómico obtenido por ambos software.

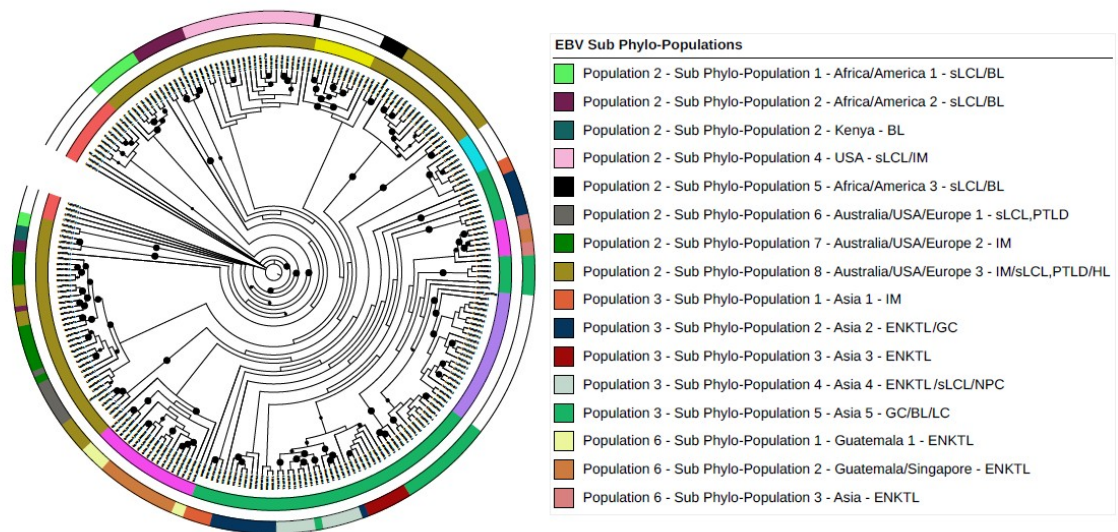


Figura 6. Sub-filo poblaciones del VEB con dos niveles de jerarquía. Resultado obtenido a partir de raxmlGUI 2.0. Se indican los valores de bootstrap >70 (escala 0-100), correspondientes a las ramas. El anillo interior indica el primer nivel y el exterior el segundo nivel de jerarquía.

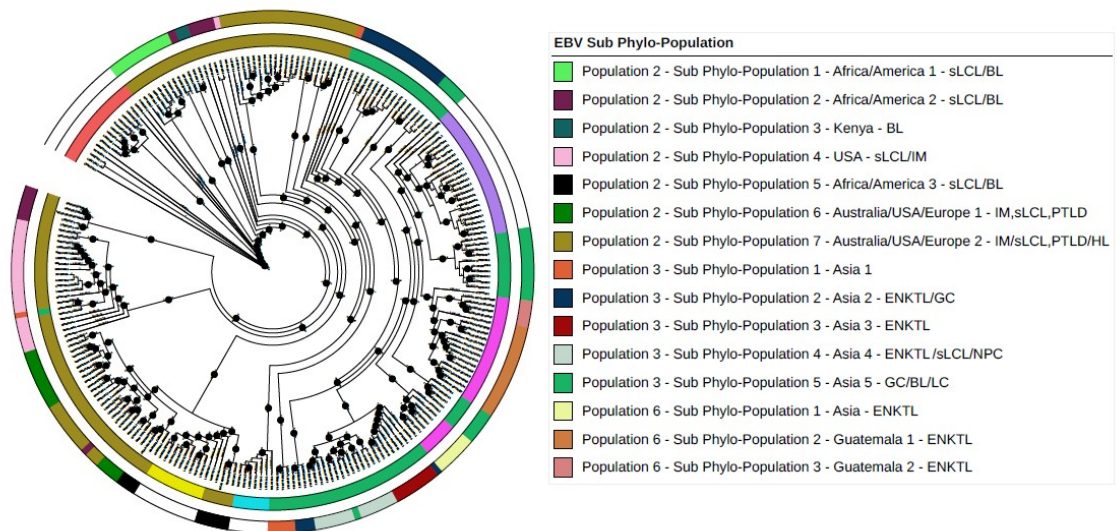


Figura 7. Sub-filo poblaciones del VEB con dos niveles de jerarquía. Resultado obtenido a partir de PhyML 3.0. Se indican los valores de bootstrap de 1 (escala 0-1), correspondientes a las ramas. El anillo interior indica el primer nivel y el exterior el segundo nivel de jerarquía.

El análisis de estructura poblacional con dos niveles de jerarquía presentado en las Figuras 6 y 7,

divide a las Filo Poblaciones 2, 3 y 6 en 8, 5 y 3 Sub-filo Poblaciones, respectivamente; excepto en el resultado obtenido por PhyML 3.0 que únicamente obtuvo 7 Sub-filo Poblaciones para la Filo Población 2. La Filo Población 6 obtenida del análisis filogenómico con raxmlGUI 2.0 fue dividida de manera que la Sub-filo Población 2 agrupó cepas guatemaltecas con cepas de Singapur, situación que no ocurre en el resultado obtenido a partir de PhyML 3.0.

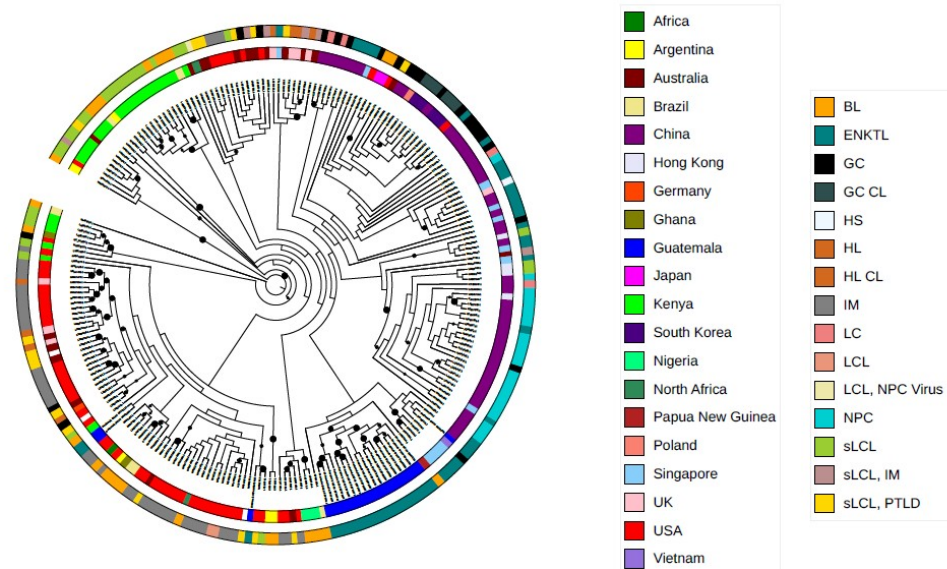


Figura 8. Relaciones filogenómicas del Virus de Epstein-Barr asociado a diferentes tipos de cáncer y procedente de diferentes países del mundo. El anillo interior indica el país de procedencia y el anillo exterior el tipo de cáncer asociado. Se indican los valores de bootstrap >70 (escala 0-100), correspondientes a las ramas. Resultado obtenido a partir de un alineamiento sin regiones recombinantes y haciendo uso de raxmlGUI 2.0 (Ver anexo 2 para abreviaturas).

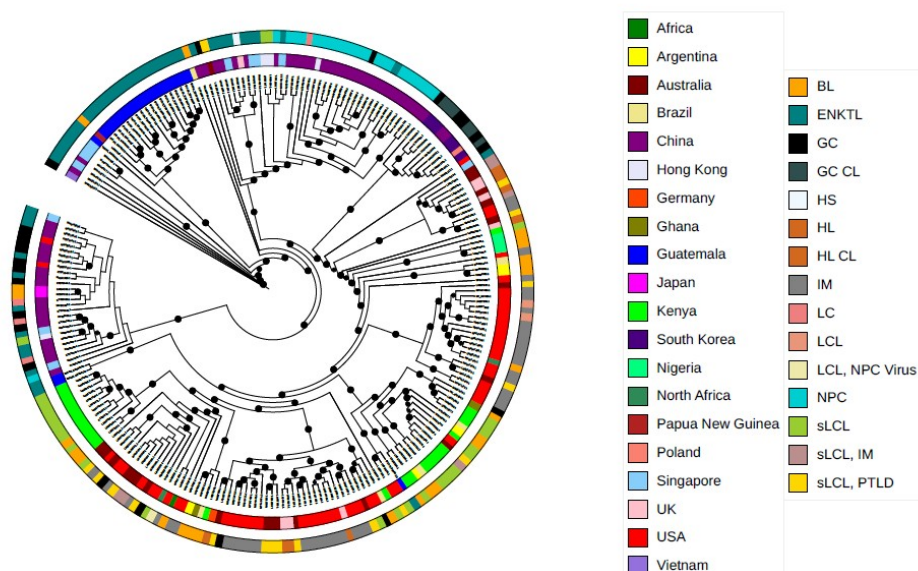


Figura 9. Relaciones filogenómicas del Virus de Epstein-Barr asociado a diferentes tipos de cáncer y procedente de diferentes países del mundo. El anillo interior indica el país de procedencia y el anillo exterior el tipo de cáncer asociado. Se indican los valores de bootstrap de 1 (escala 0-1), correspondientes a las ramas. Resultado obtenido a partir de un alineamiento sin regiones recombinantes y haciendo uso de PhyML 3.0. (Ver anexo 2 para abreviaturas).

En las figuras 8 y 9 se presentan las relaciones filogenómicas del VEB obtenidas a partir del análisis sin regiones recombinantes. Las relaciones filogenómicas fueron consistentes con los resultados obtenidos a partir del análisis de todo el alineamiento (Figuras 2 y 3). En ambos casos, existen estrechas relaciones filogenómicas entre las cepas guatemaltecas asociadas a ENKTL con cepas de países asiáticos como China y Singapur asociadas a ENKTL, NPC y CG, lo que sugiere que dichas cepas poseen características genéticas asociadas al desarrollo de las enfermedades mencionadas. Por otro lado, también se obtuvieron relaciones filogenéticas entre las cepas guatemaltecas con otras cepas de Estados Unidos o Kenia. Finalmente, cepas asociadas a BL o IM y provenientes de otras regiones se encuentran relacionadas filogenéticamente entre ellas. Los resultados obtenidos son consistentes a partir de ambas metodologías empleadas. Es importante destacar que a pesar de que algunas ramas no poseen un buen soporte en el resultado obtenido con raxmlGUI 2.0 (valores <70 no graficados), la mayoría de las relaciones filogenómicas obtenidas con PhyML 3.0 poseen un soporte de ramas aceptable en la escala de 0 a 1 empleada.

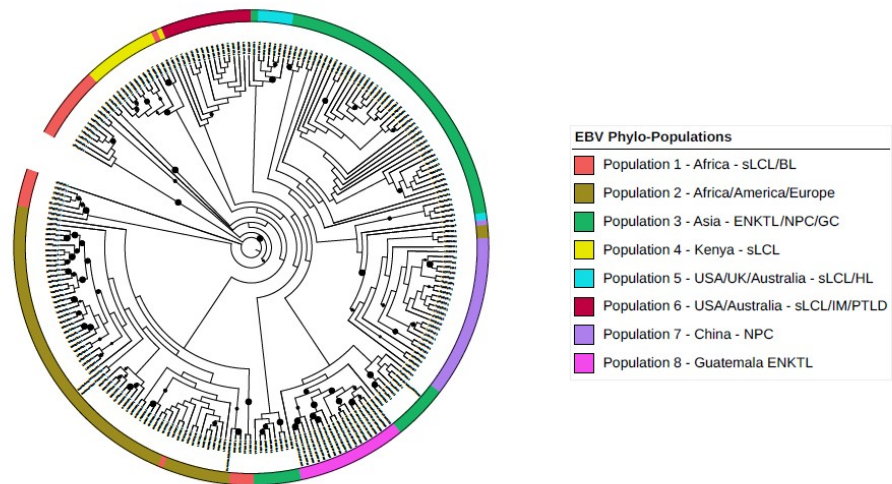


Figura 10. Filo poblaciones del VEB con un nivel de jerarquía. Resultado obtenido a partir de un alineamiento sin regiones recombinantes y haciendo uso de raxmlGUI 2.0. Se indican los valores de bootstrap >70 (escala 0-100), correspondientes a las ramas. El anillo indica cada una de las ocho Filo Poblaciones identificadas.

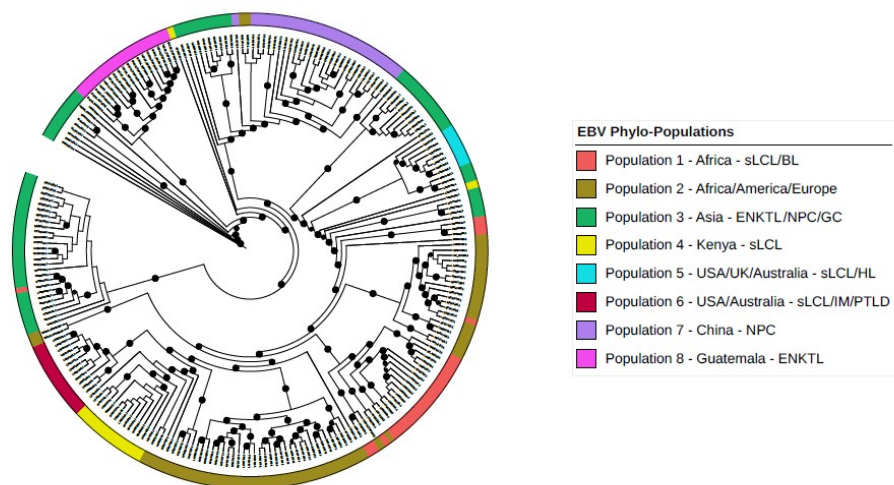


Figura 11. Filo poblaciones del VEB con un nivel de jerarquía. Resultado obtenido a partir de un alineamiento sin regiones recombinantes y haciendo uso de PhyML 3.0. Se indican los valores de bootstrap de 1 (escala 0-1), correspondientes a las ramas. El anillo indica cada una de las ocho Filo Poblaciones identificadas.

Se proponen 8 Filo Poblaciones del VEB a partir del análisis de estructura poblacional con un nivel de jerarquía a partir del análisis sin recombinación (Figuras 10 y 11). A diferencia de los resultados

obtenidos con recombinación, las cepas guatemaltecas asociadas a ENKTL formaron una Filo Población aislada, mientras que las cepas asiáticas asociadas a ENKTL, NPC y CG se encontraron en las Filo Poblaciones 3 y 7. Además, la Filo Población 5 se dividió en dos, estando la Filo Población 6 formada por cepas de USA y Australia asociadas a sLCL, IM y PTLD.

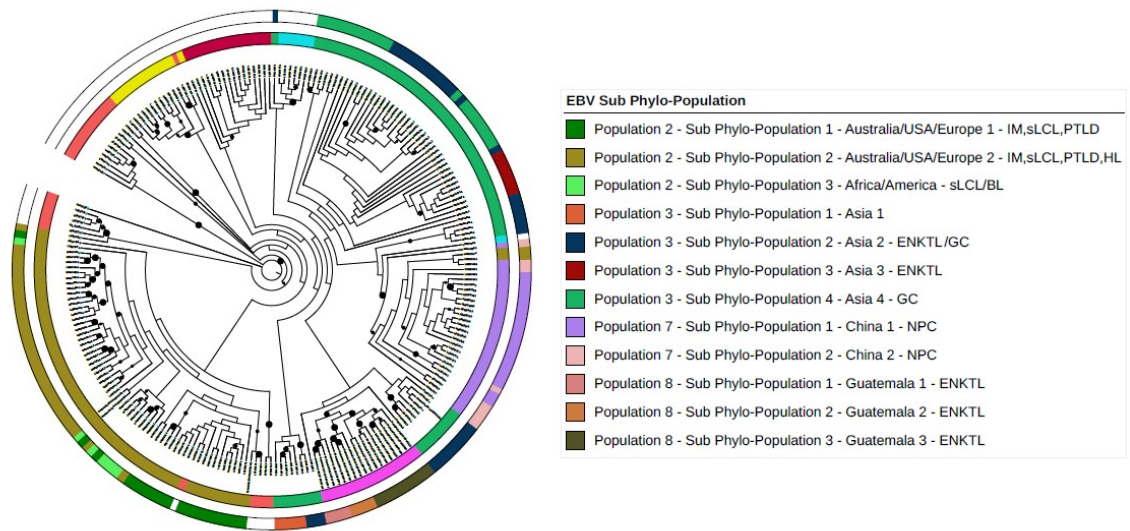


Figura 12. Sub-filo poblaciones del VEB con dos niveles de jerarquía. Resultado obtenido a partir de un alineamiento sin regiones recombinantes y haciendo uso de raxmlGUI 2.0. Se indican los valores de bootstrap >70 (escala 0-100), correspondientes a las ramas. El anillo interior indica el primer nivel y el exterior el segundo nivel de jerarquía.

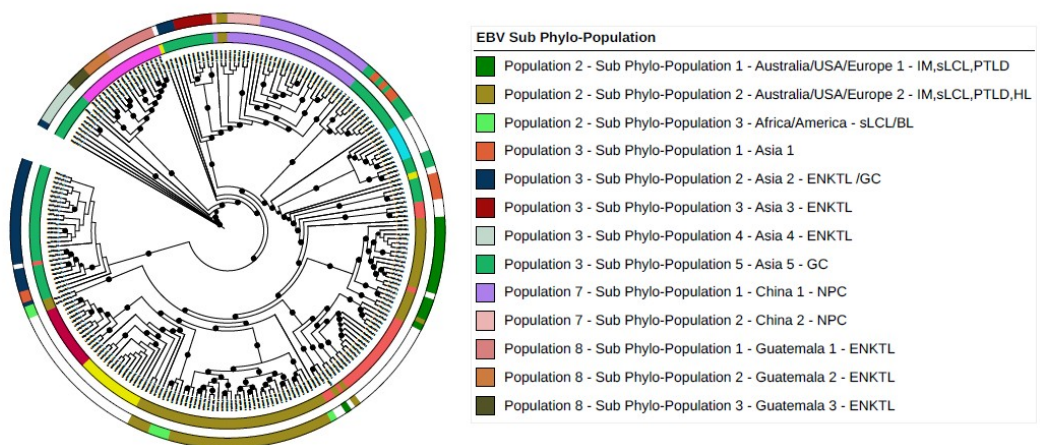


Figura 13. Sub-filo poblaciones del VEB con dos niveles de jerarquía. Resultado obtenido a partir de un alineamiento sin regiones recombinantes y haciendo uso de PhyML 3.0. Se indican los

valores de bootstrap de 1 (escala 0-1), correspondientes a las ramas. El anillo interior indica el primer nivel y el exterior el segundo nivel de jerarquía.

En las Figuras 12 y 13 se presentan las Sub-Filo Poblaciones obtenidas a partir del análisis sin recombinación. Los resultados son diferentes a los obtenidos tomando en cuenta la recombinación genética. De tal manera, las Filo Poblaciones 2, 3, 7 y 8 se dividieron en 3, 4, 2 y 3 Sub-Filo Poblaciones, respectivamente; excepto que la Población 3 que presentó una Sub-Filo Población 5 en el resultado obtenido por PhyML 3.0. Los resultados obtenidos son consistentes entre ambos software utilizados. La Filo Población de cepas guatemaltecas se dividió en 3 Sub-Filo Poblaciones, no obstante ninguna razón biológica o clínica se ha encontrado para ello (ver Tabla 2), por lo que se considera como un artefacto del análisis.

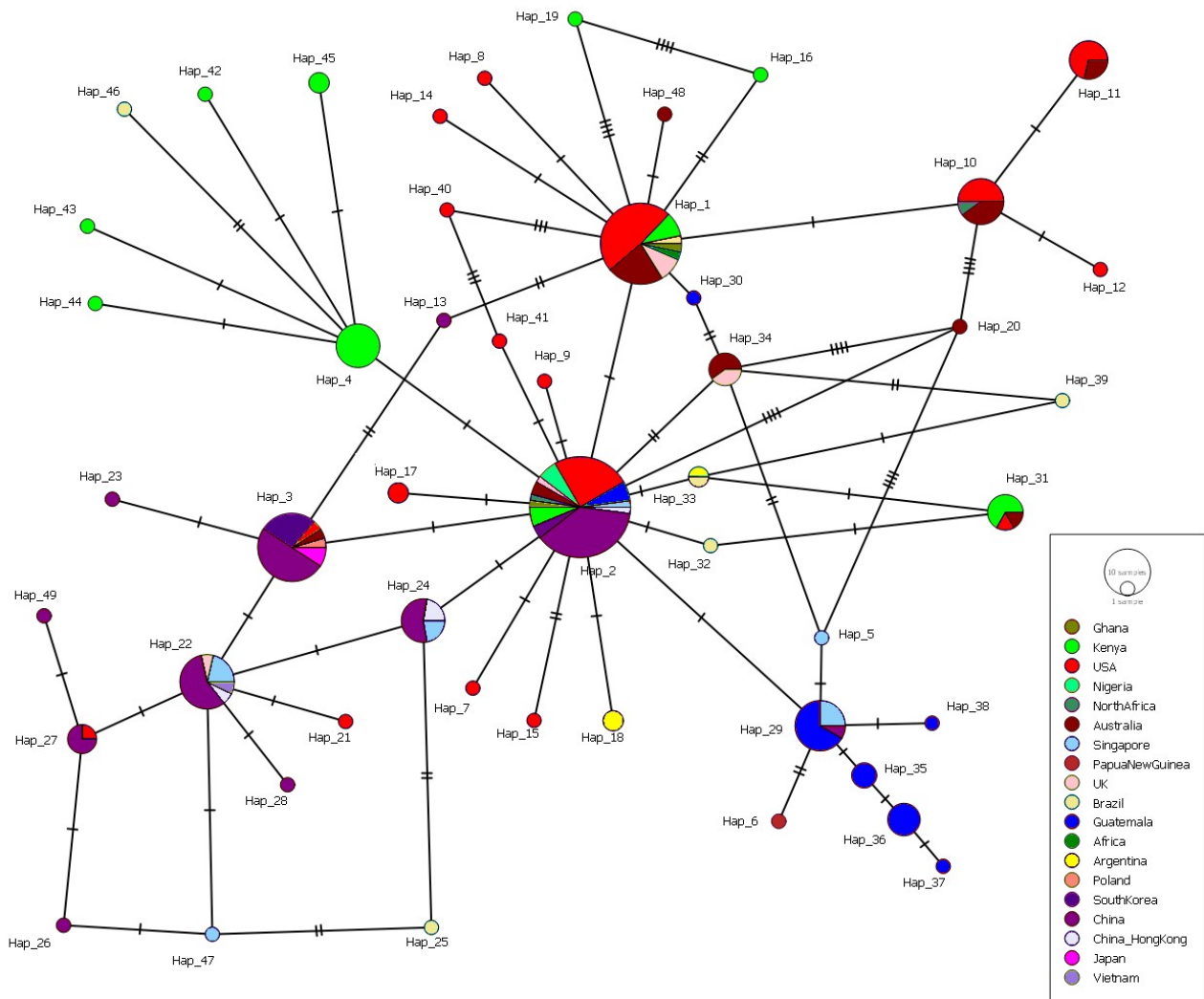


Figura 14. Red de haplotipos basada en el análisis de la región OriP. Cada país se encuentra representado por un color diferente. Las marcas entre haplotipos indican el número de mutaciones.

Se identificaron 49 haplotipos al analizar la región OriP de los genomas del VEB. Los haplotipos 2, 1 y 3 se presentaron como los más frecuentes, con mayores conexiones y de distribuciones geográficas amplias (distribuidos en América, Europa, Asia y Oceanía), por lo que son considerados como los haplotipos ancestrales. La mayor parte de los haplotipos guatemaltecos se presentan como periféricos y con conexiones a haplotipos asiáticos, principalmente de China y Singapur. Dichas conexiones entre los haplotipos guatemaltecos con los de China y Singapur coinciden con las relaciones filogenómicas descritas anteriormente. La prueba *D* de Tajima (*D*) realizada a la región OriP evidencia la existencia de una expansión poblacional o un barrido selectivo producto de selección natural ($D = -1.91987$, $p < 0.05$). Además, se identificaron 41 sitios variables y una diversidad de haplotipos de 0.9132. Demás información sobre la diversidad genética de la región OriP analizada se detalla en el Anexo 5.

9. DISCUSIÓN

Se ha reportado que el VEB posee un genoma que se encuentra compuesto por ADN linear de doble cadena, con un tamaño aproximado de 172 kilo bases (kb), incluyendo alrededor de 80 regiones codificadoras (CDS), regiones repetitivas y grandes regiones INDEL (Amon & Farrell, 2005; Zanella et al., 2019); siendo la cepa B95-8 con la cual se publicó el genoma por primera vez. Posteriormente, junto a la cepa Raji se construyó el genoma de referencia actual (Young et al., 2016). El tamaño del genoma publicado y sus características son consistentes con el genoma obtenido a partir de la secuenciación por medio de un método de enriquecimiento. El método de enriquecimiento que se utilizó para la obtención del genoma es similar al utilizado en publicaciones anteriores por otros autores como Xu y colaboradores (2019). Al contrario, autores como Peng y colaboradores (2019) utilizaron una metodología estándar de *Whole Genome Sequence* (WGS). Los resultados del análisis de variantes realizado a las secuencias del genoma del VEB guatemalteco fue llevado a cabo a partir del flujo de trabajo VCF/BCF de la plataforma Galaxy, obteniéndose un total de 1142 variantes (1,127 SNP y 15 INDEL). Mediante el uso de scripts de Phyton *in-house*, previamente se han descrito un total de 1,152 SNPs a lo largo del genoma del VEB asociado a ENKTL, siendo los genes BPLF1, BRLF1, BBLF2-BBLF3, BOLF1, BBLF4, BORF2, BZLF1, BALF5, BNRF1, BPLF1, EBNA2, BFRF2, EBNA3B los que presentaron el mayor número de variantes. 31 Variantes en dichos genes fueron asociadas como factor de riesgo a ENKTL (Peng et al., 2019). En número de variantes obtenidas, los resultados del estudio de Peng y colaboradores son similares a los obtenidos en el presente trabajo de tesis, no obstante no se identificó ninguna de las 31 variantes descritas como factores de riesgo a ENKTL. Por otro lado, es importante mencionar que se identificaron variantes en todos los genes anteriormente mencionados. En contraste, el estudio realizado por Xu y colaboradores (2019) describió un total de 8,469 variantes (8,015 SNPs y 454 INDEL) al analizar 215 genomas de VEB, esto haciendo uso del flujo de trabajo de Genome Analysis Toolkit (GATK). Dicho estudio describió que los genes EBNA-2, -3A, -3B y -3C, LMP-2A y -2B fueron los más polimórficos (Xu et al., 2019). Al comparar entre la metodología VCF/BCF empleada para el análisis de variantes con otras utilizadas en trabajos previos (los anteriormente mencionados scripts de Phyton *in-house* y GATK), resalta que al llevar a cabo un filtro en el análisis, se tomaron en cuenta únicamente aquellas variantes que poseían una calificación igual o mayor al mismo, descartando aquellas variantes que, en su mayoría, consisten en falsas positivas. De tal manera, al utilizar el valor más alto posible en el análisis de variantes realizado, se redujo la probabilidad de obtener variantes falsas positivas y, de esa manera, el nivel de confianza y precisión en los resultados obtenidos es mayor (Carson et al., 2014; Pedersen et al.,

2021).

A la fecha se han reportado variantes con importancia oncogénica del virus asociadas al ENKTL y otras enfermedades como NPC y CG. Un estudio reciente analizó el genoma del VEB en un paciente chino diagnosticado con NPC, en dicho estudio se reportaron un total de 1,413 SNP y 87 sitios INDEL (siendo 43 deleciones y 44 inserciones), de las variantes encontradas en los genes LMP1 y EBNA1 se han encontrado en pacientes asiáticos con frecuencias desde 80-100% en cohortes de 54 pacientes (Zeng et al., 2005). Entre las variantes importantes encontradas en VEB asociado a NPC se han reportado las deleciones de 30 pb del gen LMP1 (encontrada en frecuencias de 80-100% en las cohortes estudiadas) y de la región Xhol. Recientes estudios han sugerido que dichas deleciones representan variantes geográficas por lo que no poseen una asociación con el desarrollo del NPC o alguna otra enfermedad (Zhang et al., 2002). En el estudio de Peng y colaboradores (2019) se llevó a cabo un análisis de variantes del VEB asociado a ENKTL en el que se identificaron a los genes BPLF1, BRLF1, EBNA2, BBLF2, BOLF1, entre otros; como genes en los que la frecuencia de mutaciones fue mayor al comparar con NPC y CG. Dichos genes presentaron un número considerable de variantes en las secuencias analizadas de VEB asociado a ENKTL guatemalteco. No obstante, se constató que genes como EBNA1, EBNA3A, EBNA3B3C y LMP2A presentaron mayor número de variantes en el presente estudio. Así mismo, en el estudio destaca que mutaciones en EBNA2 fueron menores que las encontradas en el VEB asociado a NPC.

En el presente proyecto de tesis se obtuvieron 15 sitios INDEL a diferencia de otros estudios en los que se llegan a reportar >40 sitios INDEL (Peng et al., 2019). De manera interesante, al utilizar un menor valor de filtro para las variantes se llega a obtener >60 sitios INDEL. Diferencias en los algoritmos y herramientas bioinformáticas pueden ser una razón para la obtención de resultados dispares. Además, la falta de información para países latinoamericanas impide llevar a cabo una comparación más adecuada. Otra diferencia interesante que se ha obtenido es que no se ha identificado la deleción de 30 pb del gen LMP1 (del-LMP1) en las secuencias del VEB guatemalteco, la cual fue reportada en 21/27 muestras obtenidas de pacientes chinos con diagnóstico de ENKTL (Peng et al., 2019). Una de las razones que explicaría el por qué no se ha encontrado la deleción es la propuesta de autores que sostienen que la deleción consiste en un polimorfismo regional y no se encuentra asociada a una enfermedad, no obstante, la falta de estudios en población sana y que la mayoría de estudios llevados a cabo han sido realizados en población asiática, impiden tener un marco más grande de comparación (Neves et al., 2016). A partir del análisis realizado a las secuencias del VEB de Guatemala, destaca la variante SNP

162476T>C, la cual previamente ha sido reportada como un riesgo para el desarrollo de Cáncer Nasofaríngeo (NPC) en un metaanálisis que tomó en cuenta 270 genomas del virus provenientes de China (Xu et al., 2019). Además, la variante 162476T>C ha empezado a ser utilizada en ensayos de genotipificación de cepas de VEB por su asociación a NPC (Miller et al., 2021). No obstante, en el presente estudio de tesis no se ha realizado un análisis de validación de la variante descrita por medio del uso de un modelo de regresión logística el cual permitiría conocer la asociación entre 162476T>C y el riesgo a ENKTL.

El ENKTL tiene una distribución geográfica restringida con una ocurrencia principal en poblaciones asiáticas y nativas de Centro y Sudamérica (Montes-Mojarro et al., 2019). En los países asiáticos donde más se ha estudiado esta patología, China se encuentra con una frecuencia relativa del 14.9% con los demás linfomas no Hodgkin (Sun et al., 2012; Yang et al., 2011). En países como Japón, Corea del Sur y Tailandia se reportan frecuencias de 2.7%, 4.8% y 2.7%, respectivamente (Intragumtornchai et al., 2018). Por su parte, en países de América del Sur se han reportado frecuencias de 2.6% en Chile, 2.4% para Perú, 0.5% en Argentina y 1% en Brasil (Cabrera et al., 2012; Laurini et al., 2012). Adicionalmente, en países latinoamericanos con una gran población indígena, la incidencia de ENKTL parece incrementada, como es el caso de Guatemala, donde el porcentaje de población indígena supera el 40% (Avilés, 2015; Castellanos, 2015). En Guatemala, ENKTL representa el 7.8% de los linfomas no Hodgkin, siendo la población maya la más afectada en Guatemala (90.6% de los casos) (Sánchez-Romero et al., 2019), mientras que en Estados Unidos y Europa son menores al 0.05%. (Laurini et al., 2012). Por lo tanto, la epidemiología del ENKTL sugiere que su desarrollo se encuentra fuertemente asociado a la genética y/o al origen geográfico de la persona (etnia). Las relaciones filogenómicas obtenidas sugieren una estrecha asociación entre el VEB asociado a ENKTL de Guatemala con aquellas cepas del virus provenientes de países asiáticos y que se encuentran asociadas a ENKTL, NPC (Cáncer Nasofaríngeo) y CG (Cáncer Gástrico), principalmente. Al haberse realizado el análisis filogenómico a partir de dos software distintos, se pudo confirmar el resultado obtenido con raxmlGUI 2.0, el cual representa un análisis más robusto, con el obtenido a partir de PhyML 3.0, que es un software de ejecución más rápida, esto debido a la coincidencia en las relaciones filogenómicas entre ambos métodos. Es importante destacar que a pesar de que algunas ramas no poseen un buen soporte en el resultado obtenido con raxmlGUI 2.0 (valores <70 no graficados), la mayoría de las relaciones filogenómicas obtenidas con PhyML 3.0 poseen un soporte de ramas aceptable, lo cual, sumado a la gran coincidencia obtenida, permite la confirmación de resultados entre ambos métodos y representa una solución al tiempo computacional requerido para el análisis filogenético de grandes sets de datos como el

utilizado en el presente trabajo de tesis. Las relaciones filogenómicas presentadas en el presente trabajo son similares a las reportadas por varios autores que han abordado el tema en años anteriores. De tal manera, investigaciones previas han tratado de explicar las razones del por qué la ocurrencia es mayor en las regiones mencionadas utilizando un enfoque filogenético. Al igual como se ha realizado en el presente trabajo, estudios filogenómicos han evaluado la relación entre el genoma del virus con el tipo de enfermedad al cual se encuentra asociada la infección y el país o región de origen. En el reciente estudio de Xiong y colaboradores (2020) se presentaron relaciones filogenómicas muy similares a las obtenidos en el presente trabajo de tesis, destacando, que las cepas del VEB asociadas a ENKTL provenientes de Asia se encontraron estrechamente relacionadas con otras cepas asiáticas y que se encuentran asociadas a NPC y CG, de la misma manera en la que se han presentando las relaciones filogenómicas de las cepas guatemaltecas. Autores sostienen la idea que dichos resultados filogenómicos consisten en evidencia genética que explicaría el por qué de la gran prevalencia del virus asociado con el desarrollo de cáncer en el área nasal y paranasal en poblaciones asiáticas (Xiong & Zhao, 2018), la cual podría ser la razón por la cual las poblaciones con ascendencia maya presentan más del 90% de los casos de ENKTL en el país (Sánchez-Romero et al., 2019). En el 2019, Peng y colaboradores obtuvieron una filogenia en la que las cepas del VEB provenientes de países asiáticos asociadas a ENKTL formaban un grupo definido, mientras que cepas del resto de países analizados y asociadas a otras enfermedades se agruparon en otros grupos. Interesantemente, cepas asociadas a NPC y CG también se encontraron estrechamente relacionadas con las cepas de ENKTL, tal como se indicó anteriormente (Peng et al., 2019).

Estudios anteriores que han analizado las relaciones filogenéticas de cepas asociadas a NPC y CG, han encontrado que se encuentran estrechamente relacionadas, siendo una de las principales razones la delección de 30 pb del gen LMP1 (Borozan et al., 2018). No obstante, como se ha comentado anteriormente, se sostiene la idea que dicha delección consiste en un polimorfismo regional y no se encuentra asociada a una enfermedad (Neves et al., 2016). Por otro lado, han habido estudios que han reportado sus relaciones filogenómicas al realizar los análisis del VEB tomando en cuenta únicamente el tipo de enfermedad. En dichos estudios se han obtenido resultados que no explican completamente la relación que posee el virus con el tipo de enfermedad y cuál es el papel que posee el país o región de origen del cual ha sido aislado el virus. Dichos enfoques no permiten conocer la relación del origen geográfico con las características genómicas del virus (Choi et al., 2018). En cambio, según los resultados obtenidos en el presente trabajo de tesis se puede afirmar que la enfermedad asociada es tan importante como el origen geográfico para entender las relaciones filogenómicas del virus y proponer una clasificación filogenómica tomando

en cuenta ambos aspectos mencionados.

Para motivos del presente trabajo se emplearon únicamente dos niveles de jerarquía para establecer la estructura poblacional del VEB. La consideración de emplear solamente dos niveles de jerarquía se debe a que al emplear más niveles, los algoritmos dividen los grupos previamente formados de manera que ya no se aporta información importante respecto a las filo poblaciones propuestas. Por último, emplear dos niveles de jerarquía suele ser lo recomendado por los desarrolladores y otros autores que han empleado el paquete RhierBAPS utilizado (Cheng et al., 2013; Tonkin-Hill et al., 2018). Los resultados obtenidos identifican 7 Filo Poblaciones del VEB al tomarse en cuenta un nivel de jerarquía. Con respecto al VEB asociado a ENKTL guatemalteco, este se encuentra en la Filo Población 6, la cual está también formada por cepas de China y Singapur asociadas a ENKTL, lo que sugiere, al igual que en las relaciones filogenómicas, que dichas cepas poseen características genéticas asociadas al desarrollo de ENKTL (Xiong & Zhao, 2018). Al igual que con las relaciones filogenómicas, los resultados obtenidos fueron consistentes a partir del análisis filogenómico obtenido tanto por raxmlGUI 2.0 y por PhyML 3.0, lo cual apoya la metodología empleada en el presente trabajo que fue la de utilizar una metodología rápida como PhyML 3.0 para confirmar los resultados.

Uno de los aspectos importantes a mencionar sobre los resultados de relaciones filogenómicas y estructura poblacional obtenidos es el haber realizado cada uno de ellos a partir de los datos con recombinación y sin recombinación. Resultados similares han sido obtenidos a partir del análisis tomando en cuenta la recombinación con aquellos en los que las regiones recombinantes fueron eliminadas, sugiriendo que existen regiones únicas en el genoma del VEB que se encuentran asociadas geográficamente o a una enfermedad en específico. De tal manera, se ha obtenido una coincidencia en la mayor parte de las relaciones filogenómicas y las filo poblaciones, salvo el análisis de estructura poblacional que ha identificado 8 Filo Poblaciones al analizar los resultados sin recombinación y ha separado a las cepas guatemaltecas de las cepas de China y Singapur asociadas a ENKTL, sugiriendo que la población Filo Población Guatemalteca se comporta de manera aislada. Trabajos anteriores no han llevado a cabo en sus metodologías un análisis de recombinación para eliminar las regiones recombinantes y así obtener tanto las relaciones filogenómicas y/o resultados de estructura poblacional del virus (Peng et al., 2019; Xiong et al., 2020), salvo el trabajo de Zanella y colaboradores (2019) en el que emplearon los algoritmos de RDP4, Chimaera, GENECONV, MaxChi, SisScan y 3Seqe, del software RDP4 (mismos que se han utilizado para detectar las regiones recombinantes, eliminarlas y generar un alineamiento sin

secuencias recombinantes en este trabajo). Los resultados de estructura poblacional obtenidos en el presente trabajo de tesis son consistentes con los reportados en el estudio (Zanella et al., 2019). El trabajo de Zanella y colaboradores (2019) propone 12 Filo Poblaciones las cuales se encuentran separadas por continente (origen geográfico) y por tipo de enfermedad asociada. No obstante, al comparar los resultados obtenidos a partir del análisis de un alineamiento con recombinación con los resultados a partir de un alineamiento sin recombinación, se han obtenido resultados más consistentes a partir del primero al ser comparado con lo publicado con la demás literatura. Finalmente, otra consideración para utilizar y no descartar las regiones recombinantes en los análisis consiste en que el análisis de recombinación ha identificado eventos y regiones de recombinación a lo largo de todo el alineamiento de secuencias analizado, los cuales representan una gran proporción de las secuencias. Estudios previos han evaluado el impacto y la contribución de la recombinación genética en la reconstrucción de filogenias. Autores sostienen que la recombinación contribuye considerablemente a la obtención de las filogenias y que los eventos de recombinación poseen efectos sobre las reconstrucciones filogenéticas (Posada & Crandall, 2002; Posada, Crandall y Holmes, 2002). De tal manera, se considera que los resultados obtenidos sin eliminar la recombinación describen las relaciones filogenómicas del VEB de manera acertada, sin embargo el análisis de estructura poblacional sin recombinación genética permite identificar a la población de VEB guatemalteco asociado a ENKTL como una población aislada. La cual es considerada como la correcta debido a la eliminación de la recombinación. La coincidencia de resultados al eliminar la recombinación del alineamiento sugeriría que dentro del genoma de las distintas cepas del VEB se encuentran secuencias o regiones únicas asociadas geográficamente o a una enfermedad en específico, tal como lo ha sugerido autores en estudios previos (Xiong & Zhao, 2018). Por último, existen pequeñas diferencias entre el trabajo de Zanella y colaboradores (2019) y el presente trabajo de tesis. Entre ellas se puede mencionar el grupo *EBV-p12 / Kenya II*, el cual se encuentra formado únicamente por dos cepas y que se encuentra estrechamente relacionado con el grupo *EBV-p5 / Africa-America I* que cuenta con cepas provenientes de Kenia.

La región OriP utilizada para generar la red de haplotipos se encarga de la sincronización entre el ADN episomal del virus y la división celular del hospedero durante la fase de latencia de la infección. Así mismo, regula la replicación del virus en la fase de latencia (Young et al., 2007; Yates, Camiolo & Bashaw, 2000). Los resultados obtenidos sugieren una conexión principal entre los haplotipos de Guatemala con los de China y Singapur. Resultados que coinciden con los resultados filogenómicos y de estructura poblacional mencionados anteriormente. De esa manera, las cepas asociadas a ENKTL en China y Singapur fueron las que presentaron una relación

filogenómica más estrecha, forman una filo población y también presentan conexiones directas de haplotipos con las cepas de Guatemala. Existen limitantes que se presentan al realizar análisis de red de haplotipos, entre ellos se pueden mencionar que no se describen las relaciones evolutivas y que los límites de confianza son sensibles al número de muestras y sus respectivas distribuciones geográficas (Excoffier et al., 2007; Muwakina et al., 2003; Vázquez-Domínguez et al., 2009). Con respecto a lo primero, al contarse con los resultados de las relaciones filogenómicas del VEB, la red de haplotipos es útil para visualizar de otra manera resultados que han sido coincidentes, destacando los haplotipos guatemaltecos y sus conexiones con haplotipos de China y Singapur. Por otro lado, se considera que el número de cepas utilizadas y su distribución no es una limitante en el presente proyecto. No obstante, un muestreo homogéneo alrededor del mundo que permita el acceso a más genomas del VEB proveniente de otras regiones y que se encuentre asociado a otras enfermedades sigue siendo un reto y una limitante en este tipo de estudios. Finalmente, la expansión poblacional o el barrido selectivo, producto de la selección natural positiva, a la que está sometida la región OriP propiciaría la abundancia de alelos raros y/o la fijación de ciertos alelos (Sibley & Ajioka, 2008).

En los últimos años también se han realizado estudios que han utilizado otros enfoques para evaluar las características genéticas del VEB y su distribución geográfica. Tal es el caso del estudio de Santpere y colaboradores (2014) en el cual al analizar las variantes de genomas completos del virus por medio de un análisis de componentes principales (ACP), obtuvieron que el componente principal 2 (CP2) separó las cepas provenientes de Asia de las cepas africanas. Es importante destacar que el CP1 separó las cepas del VEB en los dos tipos de genotipos: Tipo 1 (VEB-1 o Tipo A) y Tipo 2 (VEB-2 o Tipo B), enfoque mencionado previamente pero que no se ha considerado de utilidad en el presente trabajo de tesis, al emplear datos genómicos que han explicado de manera satisfactoria la clasificación, las relaciones filogenómicas y la estructura poblacional del virus (Xiong et al., 2020; Zanella et al., 2019). Otros estudios han intentado comparar las filogenias obtenidas a partir del análisis de genoma completo del virus con el análisis de genes específicos. Tal es el caso de un estudio que evaluó cepas de VEB asociadas a cáncer nasofaríngeo, en él se llevaron a cabo análisis filogenéticos de los genes LMP1, BLLF1 y BPLF1, y se obtuvieron resultados coincidentes con los obtenidos por análisis de genoma completo (Kwok et al., 2014). Dichos resultados confirman la ventaja de utilizar datos genómicos los cuales proveen mayor información para las construcciones filogenéticas y los análisis de estructura poblacional. Por otro lado, metaanálisis que han evaluado los estados mutacionales de genes como lo son las deleciones de 30 pb del gen LMP1 (del-LMP1) y de la región Xhol, han concluido que dichas mutaciones se encuentran fuertemente asociadas a diferentes enfermedades. No obstante, la falta de estudios en

población sana no permite descartar la idea que dichas relaciones filogenéticas del virus están relacionadas geográficamente y no a una enfermedad específica (Neves et al., 2016). Por otro lado, estudios basados en el análisis mutacional del gen EBNA1 han reportado que la variante V-val EBNA1 ha sido encontrada tanto en población sana como en pacientes con NPC y HL de Asia (Ai et al., 2012), sugiriendo que la variante es regional y no relacionada a una enfermedad específica. No obstante, más estudios que evalúen otras poblaciones alrededor del mundo (incluyendo población sana) aún son necesarios para la obtención de resultados que permitan afirmar conclusiones al respecto.

Finalmente, estudios recientes han logrado explicar la relación del VEB con el ENKTL a partir del análisis de expresión genética (transcriptoma) del virus. Peng y colaboradores (2019) evaluaron el perfil de expresión del VEB a partir del análisis de transcriptoma completo, encontrando que transcritos de genes de latencia como EBNA1, LMP2, A73, BARF0 y RPMS1, presentaban una expresión significativamente menor en ENKTL que en NPC y GC; así mismo, transcritos de genes líticos de las familias BALF y LF presentaron una expresión menor en ENKTL. Al contrario, los transcritos de genes líticos como BBFF3, BLRF2 y BSFR1 presentaron una sobreexpresión en ENKTL al ser comparado con otros tipos de cáncer. Futuros estudios en Guatemala que analicen el transcriptoma del virus asociado a ENKTL y que pueda ser comparado con otros perfiles de expresión reportados para otras enfermedades (NPC y/o CG) permitirán obtener más respuestas sobre la relación entre el virus y el desarrollo de la enfermedad en la población guatemalteca. Ejemplo de ello es el estudio de Borozan y colaboradores (2018) en el que al analizar el transcriptoma de 8 muestras de VEB asociado a CG identificaron que los transcritos de genes de proteínas de latencia se encontraron en niveles de expresión consistentes (EBNA1, LMP1 y LMP2A). Así mismo, otros genes se encontraron altamente activados (BALF3, BALF4, BALF5, BILF1, LF1, LF2 y BNLF2a). Los resultados son similares a los reportados previamente para NPC y linfoma de Burkitt, excepto por los transcritos de BALF5 que no son identificados en el linfoma de Burkitt (Borozan et al., 2018; Abate et al., 2015; Hu et al., 2016). Dicho perfil de expresión genética no coincidió con el perfil lítico de infección del virus, debido que genes tempranos y tardíos se encuentran expresados, sugiriendo que los patrones de expresión genética y los mecanismos de regulación de genes específicos del VEB cambian respecto a la enfermedad asociada. Por otro lado, el análisis genético de ENKTL (genoma, exoma, transcriptoma completos, variantes, etc.) ha permitido conocer más sobre la enfermedad y su desarrollo. De tal manera, en estudios llevados a cabo en los países asiáticos de China, Corea del Sur y Japón, se han reportado diferentes frecuencias de mutaciones de los genes JAK3, STAT3, STAT5B, DDX3X, BCOR y TP53

en ENKTL, obteniendo mayores frecuencias para los genes JAK3, STAT3 y TP53, mientras que DDX3X se encuentra mutado entre 0-20% de pacientes estudiados, BCOR entre 0-32%, estando restringida únicamente a ENKTL dentro de las demás neoplasias linfoides (Küçük et al., 2015; Lee et al., 2015; Jiang et al., 2015; Dobashi et al., 2016). Por otro lado, estudios llevados a cabo en otros tipos de linfoma, como el linfoma de Burkitt, han reportado bajas frecuencias de mutaciones en genes como MYC, ID2, TCF3 y TP53; mientras que mutaciones en ARID1A poseen frecuencias mayores (Abate et al., 2015). Dicho esto, se sugiere que el ENKTL guatemalteco posee frecuencias de mutaciones específicas que son determinantes para el desarrollo de la enfermedad. Estudios similares en pacientes guatemaltecos serían útiles para conocer más sobre la enfermedad y relacionar la genética de la misma con la del VEB.

Aunque se sabe que existen variaciones del genoma del VEB entre diferentes tipos de tumores en todo el mundo (Peng et al., 2018), aún queda mucho por conocer sobre la importancia de la variación del genoma del VEB con el ENKTL y la variación genética presente en las poblaciones con una mayor prevalencia de la enfermedad. A partir de los resultados filogenómicos y de estructura poblacional obtenidos, en los que destaca relaciones filogenómicas estrechas entre el VEB asociado a ENKTL proveniente de poblaciones asiáticas con el VEB de Guatemala, se sugiere que la variación del genoma del VEB es un factor importante para el desarrollo del ENKTL en Guatemala. De igual manera, se sugiere la importancia de la genética de la población de ascendencia maya de Guatemala. Esto debido a que en Guatemala el 90.6% de los casos ENKTL se presentan en personas con ascendencia maya (Sánchez-Romero et al., 2019).

A pesar de que varios estudios llevados a cabo en las últimas décadas han sugerido la asociación de las diferentes cepas de VEB con el desarrollo de diferentes tipos de cáncer, aún existen hipótesis que plantean que las cepas de VEB solamente están relacionadas étnica y geográficamente, siendo polimorfismos regionales (Neves et al., 2016). De tal manera, el estudio de genomas del VEB (incluyendo el análisis de sus variantes y polimorfismos), obtenidos de diferentes países y regiones, aún sigue siendo materia de estudio. A partir de los resultados presentados en el presente trabajo de tesis, siendo los más relevantes las estrechas relaciones filogenómicas, la identificación de una Filo Población y la conexión entre los haplotipos de las cepas del VEB guatemalteco asociadas a ENKTL con las de países asiáticos como China y Singapur asociadas a ENKTL, NPC y CG, se puede apoyar la idea de que existe una relación tanto con el desarrollo de la enfermedad como con la etnia y origen geográfico del hospedero. Además de que se puede sugerir que dichas cepas poseen características genéticas específicas que se asocian al desarrollo de

ENKTL. No obstante, se cuentan con limitaciones como el haber tenido acceso únicamente a genomas del virus asociados a ENKTL en Guatemala. Así mismo, el reducido número de muestras analizadas representa una limitante en el estudio. El acceso a más genomas del virus asociados a otros linfomas o tipos de cáncer y en población sana en el país, sería valioso para evaluar si la relación en Guatemala del VEB es con la etnia y geográficamente o con el desarrollo de las enfermedades, o de ambas partes.

10. CONCLUSIONES

1. Se caracterizó el genoma del VEB asociado a linfoma T/NK extranodal de tipo nasal en pacientes del Instituto de Cancerología (INCAN), Guatemala. Se observó una longitud de 171,823 pb, y se identificaron 30 regiones de recombinación y 1,142 variantes. De las 1,142 variantes, 1,127 corresponden a sitios SNP y 15 a INDEL. De los 1,127 SNP obtenidos 722 representaron transiciones y 405 transversiones.
2. Filogenómicamente el VEB asociado a linfoma T/NK extranodal de tipo nasal en pacientes del Instituto de Cancerología (INCAN), Guatemala, se encuentra estrechamente asociado con aquellas cepas del virus provenientes de países asiáticos y que se encuentran asociadas a linfoma T/NK extranodal de tipo nasal, cáncer nasofaríngeo y cáncer gástrico.
3. El análisis de estructura poblacional establece a la población de VEB asociado a linfoma T/NK extranodal de tipo nasal en pacientes del Instituto de Cancerología (INCAN), Guatemala, como una población aislada.

11. RECOMENDACIONES

1. Realizar análisis filogenómicos y de estructura poblacional a partir de la secuenciación del genoma completo del VEB asociado a personas sanas y a pacientes con otros tipos de cáncer, entre ellos: otros tipos de linfomas (linfoma de Hodgkin, linfoma difuso de células B grandes y linfoma de Burkitt), cáncer nasofaríngeo y cáncer gástrico, en Guatemala.
2. Llevar a cabo más estudios multiómicos como lo son estudios de genoma, exoma y transcriptoma completos en el VEB asociado a personas sanas, a linfoma T/NK extranodal de tipo nasal y a otros tipos de enfermedades y cáncer en Guatemala, a partir de otros métodos y plataformas de secuenciación de nueva generación (NGS).
3. Llevar a cabo estudios de asociación entre las variantes (SNP e INDEL) y el riesgo al desarrollo de linfoma T/NK extranodal de tipo nasal y a otros tipos de enfermedades y cáncer en Guatemala.

12. REFERENCIAS

- Abate, F., Ambrosio, M. R., Mundo, L., Laginestra, M. A., Fuligni, F., Rossi, M., Zairis, S., Gazaneo, S., De Falco, G., Lazzi, S., Bellan, C., Rocca, B. J., Amato, T., Marasco, E., Etebari, M., Ogwang, M., Calbi, V., Ndede, I., Patel, K., Chumba, D., ... Rabadan, R. (2015). Distinct Viral and Mutational Spectrum of Endemic Burkitt Lymphoma. *PLoS pathogens*, *11*(10), e1005158. <https://doi.org/10.1371/journal.ppat.1005158>
- Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., & Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research*, *44*(W1), W3–W10. <https://doi.org/10.1093/nar/gkw343>
- Ai, J., Xie, Z., Liu, C., Huang, Z., & Xu, J. (2012). Analysis of EBNA-1 and LMP-1 variants in diseases associated with EBV infection in Chinese children. *Virology journal*, *9*, 13. <https://doi.org/10.1186/1743-422X-9-13>
- Altemani, A., Barbosa, A. C., Kulka, M., Takahashi, T., Endo, L., Vassallo, J., & Lorand-Metze, I. (2002). Characteristics of nasal T/NK-cell lymphoma among Brazilians. *Neoplasma*, *49*(1), 55-60.
- Amon, W., & Farrell, P. J. (2005). Reactivation of Epstein-Barr virus from latency. *Reviews in medical virology*, *15*(3), 149-156. <https://doi.org/10.1002/rmv.456>
- Anagnostopoulos, I., Hansmann, M. L., Franssila, K., Harris, M., Harris, N. L., Jaffe, E. S., Han, J., van Krieken, J. M., Poppema, S., Marafioti, T., Franklin, J., Sextro, M., Diehl, V., & Stein, H. (2000). European Task Force on Lymphoma project on lymphocyte predominance Hodgkin disease: histologic and immunohistologic analysis of submitted cases reveals 2 types of Hodgkin disease with a nodular growth pattern and abundant lymphocytes. *Blood*, *96*(5), 1889–1899.
- Andam, C. P., Challagundla, L., Azarian, T., Hanage, W. P., & Robinson, D. A. (2017). Population

Structure of Pathogenic Bacteria. En M. Tibayrenc. (Ed) *Genetics and Evolution of Infectious Diseases*. (pp. 51-70). Estados Unidos: Elsevier.

Andrews, S. (2012). FastQC, Cambridge, Reino Unido: Babraham Bioinformatics.

Araujo, I., Bittencourt, A. L., Barbosa, H. S., Netto, E. M., Mendonça, N., Foss, H. D., ... & Stein, H. (2006). The high frequency of EBV infection in pediatric Hodgkin lymphoma is related to the classical type in Bahia, Brasil. *Virchows Archiv*, 449(3), 315-319.

Atom. (2021). Atom version 1.54.0 x64. Recuperado de <https://www.atom.io>

Au, W. Y., Ma, S. Y., Chim, C. S., Choy, C., Loong, F., Lie, A. K., Lam, C. C., Leung, A. Y., Tse, E., Yau, C. C., Liang, R., & Kwong, Y. L. (2005). Clinicopathologic features and treatment outcome of mature T-cell and natural killer-cell lymphomas diagnosed according to the World Health Organization classification scheme: a single center experience of 10 years. *Annals of oncology : official journal of the European Society for Medical Oncology*, 16(2), 206–214. <https://doi.org/10.1093/annonc/mdi037>

Avilés A. (2015). Nasal NK/T-Cell Lymphoma. A Comparative Analysis of a Mexican Population with the Other Populations of Latin-America. *Mediterranean journal of hematology and infectious diseases*, 7(1), e2015052. <https://doi.org/10.4084/MJHID.2015.052>

Aybar, A., & Krause, J. R. (2005). Coexpression of major histocompatibility complex class I and LMP1 in Epstein-Barr virus-positive Hodgkin's Lymphoma in a pediatric age group. *Microscopy research and technique*, 68(3-4), 247-249.

Ballester, L. Y., Luthra, R., Kanagal-Shamanna, R., & Singh, R. R. (2016). Advances in clinical next-generation sequencing: target enrichment and sequencing technologies. *Expert review of molecular diagnostics*, 16(3), 357–372. <https://doi.org/10.1586/14737159.2016.1133298>

Bao, R., Huang, L., Andrade, J., Tan, W., Kibbe, W. A., Jiang, H., & Feng, G. (2014). Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. *Cancer informatics*, 13(S2), 67–82. <https://doi.org/10.4137/CIN.S13779>

- Barrionuevo, C., Zaharia, M., Martinez, M. T., Taxa, L., Misad, O., Moscol, A., ... & Zevallos-Giampietri, E. A. (2007). Extranodal NK/T-cell lymphoma, nasal type: study of clinicopathologic and prognosis factors in a series of 78 cases from Peru. *Applied Immunohistochemistry & Molecular Morphology*, 15(1), 38-44.
- Baxevanis A. D. (2001). Predictive methods using DNA sequences. *Methods of biochemical analysis*, 43, 233–252. <https://doi.org/10.1002/0471223921.ch10>
- Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing?. Archives of disease in childhood. *Education and practice edition*, 98(6), 236–238. <https://doi.org/10.1136/archdischild-2013-304340>
- Beltran, B. E., Castillo, J. J., Morales, D., de Mendoza, F. H., Quinones, P., Miranda, R. N., ... & Sotomayor, E. M. (2011). EBV-positive diffuse large B-cell lymphoma of the elderly: a case series from Peru. *American journal of hematology*, 86(8), 663-667.
- Beerli P. (2006). Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics (Oxford, England)*, 22(3), 341–345. <https://doi.org/10.1093/bioinformatics/bti803>
- Binné, U. K., Amon, W., & Farrell, P. J. (2002). Promoter sequences required for reactivation of Epstein-Barr virus from latency. *Journal of virology*, 76(20), 10282–10289. <https://doi.org/10.1128/jvi.76.20.10282-10289.2002>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Boni, M. F., Posada, D., & Feldman, M. W. (2007). An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics*, 176(2), 1035–1047. <https://doi.org/10.1534/genetics.106.068874>
- Borozan, I., Zapatka, M., Frappier, L., & Ferretti, V. (2018). Analysis of Epstein-Barr Virus Genomes and Expression Profiles in Gastric Adenocarcinoma. *Journal of virology*, 92(2),

e01239-17. <https://doi.org/10.1128/JVI.01239-17>

- Brinkman, F. S., & Leipe, D. D. (2001). Phylogenetic analysis. *Methods of biochemical analysis*, 43, 323–358. <https://doi.org/10.1002/0471223921.ch14>
- Broad Institute. (2020). Picard Tools. Recuperado de <http://broadinstitute.github.io/picard/>
- Bouvard, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., Benbrahim-Tallaa, L., Guha, N., Freeman, C., Galichet, L., Cogliano, V., & WHO International Agency for Research on Cancer Monograph Working Group (2009). A review of human carcinogens-- Part B: biological agents. *The Lancet. Oncology*, 10(4), 321–322. [https://doi.org/10.1016/s1470-2045\(09\)70096-8](https://doi.org/10.1016/s1470-2045(09)70096-8)
- Cabrera, M. E., Martinez, V., Nathwani, B. N., Muller-Hermelink, H. K., Diebold, J., Maclennan, K. A., Armitage, J., & Weisenburger, D. D. (2012). Non-Hodgkin lymphoma in Chile: a review of 207 consecutive adult cases by a panel of five expert hematopathologists. *Leukemia & lymphoma*, 53(7), 1311–1317. <https://doi.org/10.3109/10428194.2012.654471>
- Campbell, J., Seymour, J. F., Matthews, J., Wolf, M., Stone, J., & Juneja, S. (2006). The prognostic impact of bone marrow involvement in patients with diffuse large cell lymphoma varies according to the degree of infiltration and presence of discordant marrow involvement. *European journal of haematology*, 76(6), 473–480. <https://doi.org/10.1111/j.1600-0609.2006.00644.x>
- Campo, E., Harris, N. L., Jaffe, E. S., Pileri, S. A., Stein, H., Thiele, J., & Vardiman, J. W. (2008). WHO classification of tumours of haematopoietic and lymphoid tissues (Vol. 2, p. 439). S. H. Swerdlow (Ed.). Lyon, France: International agency for research on cancer.
- Campos, A. H. J. F. M., Moreira, A., Ribeiro, K. B., Paes, R. P., Zerbini, M. C., Aldred, V., ... & Vassallo, J. (2018). Frequency of EBV associated classical Hodgkin lymphoma decreases over a 54-year period in a Brazilian population. *Scientific reports*, 8(1), 1-8.
- Carbone, A., Roulland, S., Gloghini, A., Younes, A., von Keudell, G., López-Guillermo, A., & Fitzgibbon, J. (2019). Follicular lymphoma. *Nature Reviews Disease Primers*, 5(1), 1-20.

<https://doi.org/10.1038/s41572-019-0132-x>

- Carson, A. R., Smith, E. N., Matsui, H., Brækkan, S. K., Jepsen, K., Hansen, J. B., & Frazer, K. A. (2014). Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC bioinformatics*, *15*, 125. <https://doi.org/10.1186/1471-2105-15-125>
- Castellanos, G. D. (2015). Guatemala: una nación, dos países. *Sociedad y Discurso*, (27), 84-100. ISSN 1601-1686
- Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, *17*(4), 540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>
- Chabay, P., Lens, D., Hassan, R., Rodríguez Pinilla, S. M., Valvert Gamboa, F., Rivera, I., ... & Fuentes-Pananá, E. M. (2020). Lymphotropic Viruses EBV, KSHV and HTLV in Latin America: Epidemiology and Associated Malignancies. A Literature-Based Study by the RIAL-CYTED. *Cancers*, *12*(8), 2166.
- Chang, K. L., Albújar, P. F., Chen, Y. Y., Johnson, R. M., & Weiss, L. M. (1993). High prevalence of Epstein-Barr virus in the Reed-Sternberg cells of Hodgkin's disease occurring in Peru. *Blood*, *81*(2), 496-501.
- Cheah, C. Y., Seymour, J. F., & Wang, M. L. (2016). Mantle cell lymphoma. *Journal of clinical oncology*, *34*(11), 1256-1269. <http://dx.doi.org/10.1200/JCO.2015.63.5904>
- Chen, H., Smith, P., Ambinder, R. F., & Hayward, S. D. (1999). Expression of Epstein-Barr virus BamHI-A rightward transcripts in latently infected B cells from peripheral blood. *Blood*, *93*(9), 3026–3032.
- Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M., & Corander, J. (2013). Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Molecular biology and evolution*, *30*(5), 1224–1228. <https://doi.org/10.1093/molbev/mst028>

- Chiara, M., Manzari, C., Lionetti, C., Mechelli, R., Anastasiadou, E., Chiara Buscarinu, M., Ristori, G., Salvetti, M., Picardi, E., D'Erchia, A. M., Pesole, G., & Horner, D. S. (2016). Geographic Population Structure in Epstein-Barr Virus Revealed by Comparative Genomics. *Genome biology and evolution*, 8(11), 3284–3291. <https://doi.org/10.1093/gbe/evw226>
- Choi, S. J., Jung, S. W., Huh, S., Cho, H., & Kang, H. (2018). Phylogenetic comparison of Epstein-Barr virus genomes. *Journal of microbiology (Seoul, Corea Del Sur)*, 56(8), 525–533. <https://doi.org/10.1007/s12275-018-8039-x>
- Chung, R., Lai, R., Wei, P., Lee, J., Hanson, J., Belch, A. R., Turner, A. R., & Reiman, T. (2007). Concordant but not discordant bone marrow involvement in diffuse large B-cell lymphoma predicts a poor clinical outcome independent of the International Prognostic Index. *Blood*, 110(4), 1278–1282. <https://doi.org/10.1182/blood-2007-01-070300>
- Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular ecology*, 9(10), 1657–1659. <https://doi.org/10.1046/j.1365-294x.2000.01020.x>
- Cohen, J. I. (2000). Epstein–Barr virus infection. *New England Journal of Medicine*, 343(7), 481–492. DOI: 10.1056/NEJM200008173430707
- Cohen, M., Narbaitz, M., Metrebian, F., De Matteo, E., Preciado, M. V., & Chabay, P. A. (2014). Epstein-Barr virus-positive diffuse large B-cell lymphoma association is not only restricted to elderly patients. *International journal of cancer*, 135(12), 2816–2824.
- Cohen, M., Vistarop, A. G., Huaman, F., Narbaitz, M., Metrebian, F., De Matteo, E., ... & Chabay, P. A. (2017). Cytotoxic response against Epstein Barr virus coexists with diffuse large B-cell lymphoma tolerogenic microenvironment: clinical features and survival impact. *Scientific reports*, 7(1), 1–10.
- Coleman, C. B., Lang, J., Sweet, L. A., Smith, N. A., Freed, B. M., Pan, Z., ... & Rochford, R. (2018). Epstein-Barr virus type 2 infects T cells and induces B cell lymphomagenesis in humanized mice. *Journal of virology*, 92(21).

- Conant, G. C., & Wolfe, K. H. (2008). GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics*, *24*(6), 861-862.
- Corander, J., & Marttinen, P. (2006). Bayesian identification of admixture events using multilocus molecular markers. *Molecular ecology*, *15*(10), 2833–2843. <https://doi.org/10.1111/j.1365-294X.2006.02994.x>
- Córdova-Pérez, F. J., González-Keelan, C. I., & Vélez, R. (2003). Epstein-Barr virus in biopsies from patients with Hodgkin and non-Hodgkin lymphoma at the University of Puerto Rico immunohistochemistry laboratory. *Puerto Rico health sciences journal*, *22*(2), 125-129.
- Crawford, D. H. (2001). Biology and disease associations of Epstein–Barr virus. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1408), 461-473.
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., ... & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008.
- Dawson, C. W., Port, R. J., & Young, L. S. (2012). The role of the EBV-encoded latent membrane proteins LMP1 and LMP2 in the pathogenesis of nasopharyngeal carcinoma (NPC). *In Seminars in cancer biology*, *22*(2), 144-153.
- Decaussin, G., Sbih-Lammali, F., de Turenne-Tessier, M., Bouguermouh, A., & Ooka, T. (2000). Expression of BARF1 gene encoded by Epstein-Barr virus in nasopharyngeal carcinoma biopsies. *Cancer research*, *60*(19), 5584–5588.
- De Matteo, E., Barón, A., Chabay, P., Porta, J., Dragosky, M., & Preciado, M. V. (2003). Comparison of Epstein-Barr virus presence in Hodgkin lymphoma in pediatric versus adult Argentine patients. *Archives of pathology & laboratory medicine*, *127*(10), 1325-1329.
- de Oliveira Ferreira, J. M., Klumb, C. E., de Souza Reis, R., de Oliveira Santos, M., Oliveira, J. F. P., de Camargo, B., & Pombo-de-Oliveira, M. S. (2012). Lymphoma subtype incidence rates in children and adolescents: first report from Brasil. *Cancer epidemiology*, *36*(4), e221-e226.

- De Paschale, M., & Clerici, P. (2012). Serological diagnosis of Epstein-Barr virus infection: Problems and solutions. *World journal of virology*, *1*(1), 31–43. <https://doi.org/10.5501/wjv.v1.i1.31>
- Dirnhofer, S., Angeles-Angeles, A., Ortiz-Hidalgo, C., Reyes, E., Gredler, E., Krugmann, J., ... & Quintanilla-Martinez, L. (1999). High prevalence of a 30-base pair deletion in the Epstein-Barr virus (EBV) latent membrane protein 1 gene and of strain type B EBV in Mexican classical Hodgkin's disease and reactive lymphoid tissue. *Human pathology*, *30*(7), 781–787.
- Dobashi, A., Tsuyama, N., Asaka, R., Togashi, Y., Ueda, K., Sakata, S., Baba, S., Sakamoto, K., Hatake, K., & Takeuchi, K. (2016). Frequent BCOR aberrations in extranodal NK/T-Cell lymphoma, nasal type. *Genes, chromosomes & cancer*, *55*(5), 460–471. <https://doi.org/10.1002/gcc.22348>
- Dunmire, S. K., Verghese, P. S., & Balfour, H. H., Jr (2018). Primary Epstein-Barr virus infection. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, *102*, 84–92. <https://doi.org/10.1016/j.jcv.2018.03.001>
- Elenitoba-Johnson, K. S., Zarate-Osorno, A., Meneses, A., Krenacs, L., Kingma, D. W., Raffeld, M., & Jaffe, E. S. (1998). Cytotoxic granular protein expression, Epstein-Barr virus strain type, and latent membrane protein-1 oncogene deletions in nasal T-lymphocyte/natural killer cell lymphomas from Mexico. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc*, *11*(8), 754–761.
- Eliopoulos, A. G., Gallagher, N. J., Blake, S. M., Dawson, C. W., & Young, L. S. (1999). Activation of the p38 mitogen-activated protein kinase pathway by Epstein-Barr virus-encoded latent membrane protein 1 coregulates interleukin-6 and interleukin-8 production. *The Journal of biological chemistry*, *274*(23), 16085–16096. <https://doi.org/10.1074/jbc.274.23.16085>
- Engelhard, M., Brittinger, G., Huhn, D., Gerhartz, H. H., Meusers, P., Siegert, W., Thiel, E., Wilmanns, W., Aydemir, U., Bierwolf, S., Griesser, H., Tiemann, M., & Lennert, K. (1997). Subclassification of diffuse large B-cell lymphomas according to the Kiel classification:

distinction of centroblastic and immunoblastic lymphomas is a significant prognostic risk factor. *Blood*, 89(7), 2291–2297.

- Excoffier, L., Laval, G., & Schneider, S. (2007). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics online*, 1, 47–50.
- Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular ecology notes*, 7(4), 574–578. <https://doi.org/10.1111/j.1471-8286.2007.01758.x>
- Fanaian, N. K., Cohen, C., Waldrop, S., Wang, J., & Shehata, B. M. (2009). Epstein-Barr virus (EBV)-encoded RNA: automated in-situ hybridization (ISH) compared with manual ISH and immunohistochemistry for detection of EBV in pediatric lymphoproliferative disorders. *Pediatric and developmental pathology : the official journal of the Society for Pediatric Pathology and the Paediatric Pathology Society*, 12(3), 195–199. <https://doi.org/10.2350/07-07-0316.1>
- Felsenstein J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of molecular evolution*, 17(6), 368–376. <https://doi.org/10.1007/BF01734359>
- Fixman, E. D., Hayward, G. S., & Hayward, S. D. (1992). trans-acting requirements for replication of Epstein-Barr virus ori-Lyt. *Journal of virology*, 66(8), 5030–5039. <https://doi.org/10.1128/JVI.66.8.5030-5039.1992>
- Gallo, T., Frampton, D., Sugar, R., Charles, E., Kozlakidis, Z & Kellam, P. (2016). High-throughput pipeline for the de novo viral genome assembly and the identification of minority variants from Next-Generation Sequencing of residual diagnostic samples. *BioRxiv*. doi: <https://doi.org/10.1101/035154>
- García-Cosío, M., Santón, A., Méndez, M. C., Rivas, C., Martín, C., & Bellas, C. (2003). Nasopharyngeal/nasal type T/NK lymphomas: analysis of 14 cases and review of the literature. *Tumori*, 89(3), 278–284.
- Gärtner, B., & Preiksaitis, J. K. (2010). EBV viral load detection in clinical virology. *Journal of*

clinical virology: the official publication of the Pan American Society for Clinical Virology, 48(2), 82–90. <https://doi.org/10.1016/j.jcv.2010.03.016>

- Garrison, E., Kronenberg, Z. N., Dawson, E. T., Pedersen, B. S., & Prins, P. (2021). Vcflib and tools for processing the VCF variant call format. *bioRxiv*. <https://doi.org/10.1101/2021.05.21.445151>
- Gibbs, M. J., Armstrong, J. S., & Gibbs, A. J. (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics (Oxford, England)*, 16(7), 573–582. <https://doi.org/10.1093/bioinformatics/16.7.573>
- Gires, O., Zimmer-Strobl, U., Gonnella, R., Ueffing, M., Marschall, G., Zeidler, R., Pich, D., & Hammerschmidt, W. (1997). Latent membrane protein 1 of Epstein-Barr virus mimics a constitutively active receptor molecule. *The EMBO journal*, 16(20), 6131–6140. <https://doi.org/10.1093/emboj/16.20.6131>
- Gómez-Aparicio, L., Ávila, J. M. & Cayuela, L. (2013). Métodos de máxima verosimilitud en ecología y su aplicación en modelos de vecindad. *Ecosistemas*, 22(3), 12-20. [10.7818/ECOS.2013.22-3.03](https://doi.org/10.7818/ECOS.2013.22-3.03)
- Görzer, I., Niesters, H. G., Cornelissen, J. J., & Puchhammer-Stöckl, E. (2006). Characterization of Epstein-Barr virus Type I variants based on linked polymorphism among EBNA3A, -3B, and -3C genes. *Virus research*, 118(1-2), 105–114. <https://doi.org/10.1016/j.virusres.2005.11.020>
- Gualco, G., Domeny-Duarte, P., Chioato, L., Barber, G., Natkunam, Y., & Bacchi, C. E. (2011). Clinicopathologic and molecular features of 122 Brazilian cases of nodal and extranodal NK/T-cell lymphoma, nasal type, with EBV subtyping analysis. *The American journal of surgical pathology*, 35(8), 1195-1203.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), 307–321. <https://doi.org/10.1093/sysbio/syq010>

- Gulley, M. L., Eagan, P. A., Quintanilla-Martinez, L., Picado, A. L., Smir, B. N., Childs, C., ... & Banks, P. M. (1994). Epstein-Barr virus DNA is abundant and monoclonal in the Reed-Sternberg cells of Hodgkin's disease: association with mixed cellularity subtype and Hispanic American ethnicity. *Blood* 1994, 83, 1595-1602.
- Haralambieva, E., Pulford, K. A., Lamant, L., Pileri, S., Roncador, G., Gatter, K. C., Delsol, G., & Mason, D. Y. (2000). Anaplastic large-cell lymphomas of B-cell phenotype are anaplastic lymphoma kinase (ALK) negative and belong to the spectrum of diffuse large B-cell lymphomas. *British journal of haematology*, 109(3), 584–591.
<https://doi.org/10.1046/j.1365-2141.2000.02045.x>
- Harris, N. L., Jaffe, E. S., Stein, H., Banks, P. M., Chan, J. K., Cleary, M. L., Delsol, G., De Wolf-Peeters, C., Falini, B., & Gatter, K. C. (1994). A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. *Blood*, 84(5), 1361–1392.
- Harris, S. R., Cartwright, E. J., Török, M. E., Holden, M. T., Brown, N. M., Ogilvy-Stuart, A. L., Ellington, M. J., Quail, M. A., Bentley, S. D., Parkhill, J., & Peacock, S. J. (2013). Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. *The Lancet. Infectious diseases*, 13(2), 130–136.
[https://doi.org/10.1016/S1473-3099\(12\)70268-2](https://doi.org/10.1016/S1473-3099(12)70268-2)
- Hislop, A. D., Rensing, M. E., van Leeuwen, D., Pudney, V. A., Horst, D., Koppers-Lalic, D., Croft, N. P., Neefjes, J. J., Rickinson, A. B., & Wiertz, E. J. (2007). A CD8+ T cell immune evasion protein specific to Epstein-Barr virus and its close relatives in Old World primates. *The Journal of experimental medicine*, 204(8), 1863–1873.
<https://doi.org/10.1084/jem.20070256>
- Hoeller, S., Tzankov, A., Pileri, S. A., Went, P., & Dirnhofer, S. (2010). Epstein-Barr virus–positive diffuse large B-cell lymphoma in elderly patients is rare in Western populations. *Human pathology*, 41(3), 352-357.
- Höfelmayr, H., Strobl, L. J., Marschall, G., Bornkamm, G. W., & Zimmer-Strobl, U. (2001). Activated Notch1 can transiently substitute for EBNA2 in the maintenance of proliferation

of LMP1-expressing immortalized B cells. *Journal of virology*, 75(5), 2033-2040.

- Hofscheier, A., Ponciano, A., Bonzheim, I., Adam, P., Lome-Maldonado, C., Vela, T., ... & Quintanilla-Martinez, L. (2011). Geographic variation in the prevalence of Epstein–Barr virus-positive diffuse large B-cell lymphoma of the elderly: a comparative analysis of a Mexican and a German population. *Modern Pathology*, 24(8), 1046-1054.
- Hong, H., Huang, H., Fang, X., Wang, Z., Ye, S., Zhang, H., Huang, Y., Guo, H., Chen, X., Liang, C., Pu, X., Cao, Y., Lin, S., Li, X., Ren, Q., Liu, Q., & Lin, T. (2019). A prognostic index for nasal-type early-stage extranodal natural killer/T-cell lymphoma: A multicenter study. *American journal of hematology*, 94(5), E122–E124. <https://doi.org/10.1002/ajh.25426>
- Howlader, N., Noone, A. M., Krapcho, M., Miller, D., Brest, A., Yu, M., ... & Lewis, D. R. (2016). *SEER Cancer Statistics Review, 1975–2016*. USA: National Cancer Institute.
- Hu, L., Lin, Z., Wu, Y., Dong, J., Zhao, B., Cheng, Y., Huang, P., Xu, L., Xia, T., Xiong, D., Wang, H., Li, M., Guo, L., Kieff, E., Zeng, Y., Zhong, Q., & Zeng, M. (2016). Comprehensive profiling of EBV gene expression in nasopharyngeal carcinoma through paired-end transcriptome sequencing. *Frontiers of medicine*, 10(1), 61–75. <https://doi.org/10.1007/s11684-016-0436-0>
- Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular ecology resources*, 9(5), 1322–1332. <https://doi.org/10.1111/j.1755-0998.2009.02591.x>
- Hummel, M., Tamaru, J., Kalvelage, B., & Stein, H. (1994). Mantle cell (previously centrocytic) lymphomas express VH genes with no or very little somatic mutations like the physiologic cells of the follicle mantle. *Blood*, 84(2), 403–407.
- Hutt-Fletcher L. M. (2007). Epstein-Barr virus entry. *Journal of virology*, 81(15), 7825–7832. <https://doi.org/10.1128/JVI.00445-07>
- Inghirami, G., Foitl, D. R., Sabichi, A., Zhu, B. Y., & Knowles, D. M. (1991). Autoantibody-associated cross-reactive idiotype-bearing human B lymphocytes: distribution and

characterization, including Ig VH gene and CD5 antigen expression. *Blood*, 78(6), 1503–1515.

Intragumtornchai, T., Bunworasate, U., Wudhikarn, K., Lekhakula, A., Julamanee, J., Chansung, K., Sirijerachai, C., Norasetthada, L., Nawarawong, W., Khuhapinant, A., Siritanaratanakul, N., Numbenjapon, T., Prayongratana, K., Chuncharunee, S., Niparuck, P., Suwanban, T., Kanitsap, N., Wongkhantee, S., Pornvipavee, R., Wong, P., ... Warnnissorn, N. (2018). Non-Hodgkin lymphoma in South East Asia: An analysis of the histopathology, clinical features, and survival from Thailand. *Hematological oncology*, 36(1), 28–36. <https://doi.org/10.1002/hon.2392>

Jaffe, E. S., Harris, N. L., Stein, H., & Isaacson, P. G. (2008). Classification of lymphoid neoplasms: the microscope as a tool for disease discovery. *Blood*, 112(12), 4384–4399. <https://doi.org/10.1182/blood-2008-07-077982>

Jenson H. B. (2011). Epstein-Barr virus. *Pediatrics in review*, 32(9), 375–384. <https://doi.org/10.1542/pir.32-9-375>

Jiang, L., Gu, Z. H., Yan, Z. X., Zhao, X., Xie, Y. Y., Zhang, Z. G., Pan, C. M., Hu, Y., Cai, C. P., Dong, Y., Huang, J. Y., Wang, L., Shen, Y., Meng, G., Zhou, J. F., Hu, J. D., Wang, J. F., Liu, Y. H., Yang, L. H., Zhang, F., ... Chen, S. J. (2015). Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma. *Nature genetics*, 47(9), 1061–1066. <https://doi.org/10.1038/ng.3358>

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>

Kelly, G. L., Milner, A. E., Tierney, R. J., Croom-Carter, D. S., Altmann, M., Hammerschmidt, W., ... & Rickinson, A. B. (2005). Epstein-Barr virus nuclear antigen 2 (EBNA2) gene deletion is consistently linked with EBNA3A,-3B, and-3C expression in Burkitt's lymphoma cells and with increased resistance to apoptosis. *Journal of virology*, 79(16), 10709-10717.

Kilger, E., Kieser, A., Baumann, M., & Hammerschmidt, W. (1998). Epstein-Barr virus-mediated

B-cell proliferation is dependent upon latent membrane protein 1, which simulates an activated CD40 receptor. *The EMBO journal*, 17(6), 1700-1709.

Kim, W. Y., Montes-Mojarro, I. A., Fend, F., & Quintanilla-Martinez, L. (2019). Epstein-Barr Virus-Associated T and NK-Cell Lymphoproliferative Diseases. *Frontiers in pediatrics*, 7(71), 1-18. <https://doi.org/10.3389/fped.2019.00071>

Kimura M. (1979). Model of effectively neutral mutations in which selective constraint is incorporated. *Proceedings of the National Academy of Sciences of the United States of America*, 76(7), 3440–3444. <https://doi.org/10.1073/pnas.76.7.3440>

Kimura, H., Ito, Y., Kawabe, S., Gotoh, K., Takahashi, Y., Kojima, S., Naoe, T., Esaki, S., Kikuta, A., Sawada, A., Kawa, K., Ohshima, K., & Nakamura, S. (2012). EBV-associated T/NK-cell lymphoproliferative diseases in nonimmunocompromised hosts: prospective analysis of 108 cases. *Blood*, 119(3), 673–686. <https://doi.org/10.1182/blood-2011-10-381921>

Kipps T. J. (1989). The CD5 B cell. *Advances in immunology*, 47, 117–185. [https://doi.org/10.1016/s0065-2776\(08\)60663-x](https://doi.org/10.1016/s0065-2776(08)60663-x)

Kitagawa, N., Goto, M., Kurozumi, K., Maruo, S., FReino Unidoayama, M., Naoe, T., YasReino Unidoawa, M., Hino, K., SuzReino Unidoi, T., Todo, S., & Takada, K. (2000). Epstein-Barr virus-encoded poly(A)(-) RNA supports Burkitt's lymphoma growth through interleReino Unidoi-10 induction. *The EMBO journal*, 19(24), 6742–6750. <https://doi.org/10.1093/emboj/19.24.6742>

Klumb, C. E., Hassan, R., De Oliveira, D. E., De Resende, L. M. M., Carriço, M. K., De Almeida Dobbin, J., ... & Maia, R. C. (2004). Geographic variation in Epstein-Barr virus-associated Burkitt's lymphoma in children from Brasil. *International journal of cancer*, 108(1), 66-70.

Knipe D. M., Howley P. M., Kieff, E., and Rickinson, A. B. (2001). Epstein-Barr virus and its replication. In (eds.), *Fields Virology Philadelphia: Lippincott Williams & Wilkins*, pp. 2511–2574.

- Kozarewa, I., Armisen, J., Gardner, A. F., Slatko, B. E., & Hendrickson, C. L. (2015). Overview of Target Enrichment Strategies. *Current protocols in molecular biology*, *112*, 7.21.1–7.21.23. <https://doi.org/10.1002/0471142727.mb0721s112>
- Küçük, C., Jiang, B., Hu, X., Zhang, W., Chan, J. K., Xiao, W., Lack, N., Alkan, C., Williams, J. C., Avery, K. N., Kavak, P., Scuto, A., Sen, E., Gaulard, P., Staudt, L., Iqbal, J., Zhang, W., Cornish, A., Gong, Q., Yang, Q., ... Chan, W. C. (2015). Activating mutations of STAT5B and STAT3 in lymphomas derived from $\gamma\delta$ -T or NK cells. *Nature communications*, *6*, 6025. <https://doi.org/10.1038/ncomms7025>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular biology and evolution*, *35*(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Kwok, H., Wu, C. W., Palser, A. L., Kellam, P., Sham, P. C., Kwong, D. L., & Chiang, A. K. (2014). Genomic diversity of Epstein-Barr virus genomes isolated from primary nasopharyngeal carcinoma biopsy samples. *Journal of virology*, *88*(18), 10662–10672. <https://doi.org/10.1128/JVI.01665-14>
- Kwong, Y. L., Pang, A. W., Leung, A. Y., Chim, C. S., & Tse, E. (2014). Quantification of circulating Epstein-Barr virus DNA in NK/T-cell lymphoma treated with the SMILE protocol: diagnostic and prognostic significance. *LeReino Unidoemia*, *28*(4), 865–870. <https://doi.org/10.1038/leu.2013.212>
- Laing, K. G., Elia, A., Jeffrey, I., Matys, V., Tilleray, V. J., Souberbielle, B., & Clemens, M. J. (2002). In vivo effects of the Epstein-Barr virus small RNA EBER-1 on protein synthesis and cell growth regulation. *Virology*, *297*(2), 253–269. <https://doi.org/10.1006/viro.2002.1354>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357. <https://doi.org/10.1038/nmeth.1923>
- Lara, J., Cohen, M., De Matteo, E., Aversa, L., Preciado, M. V., & Chabay, P. (2014). Epstein–Barr virus (EBV) association and latency profile in pediatric Burkitt's lymphoma: Experience of a

single institution in Argentina. *Journal of medical virology*, 86(5), 845-850.

- Laurini, J. A., Perry, A. M., Boilesen, E., Diebold, J., MacLennan, K. A., Müller-Hermelink, H. K., Nathwani, B. N., Armitage, J. O., & Weisenburger, D. D. (2012). Classification of non-Hodgkin lymphoma in Central and South America: a review of 1028 cases. *Blood*, 120(24), 4795–4801. <https://doi.org/10.1182/blood-2012-07-440073>
- Lee, S., Park, H. Y., Kang, S. Y., Kim, S. J., Hwang, J., Lee, S., Kwak, S. H., Park, K. S., Yoo, H. Y., Kim, W. S., Kim, J. I., & Ko, Y. H. (2015). Genetic alterations of JAK/STAT cascade and histone modification in extranodal NK/T-cell lymphoma nasal type. *Oncotarget*, 6(19), 17764–17776. <https://doi.org/10.18632/oncotarget.3776>
- Lefort, V., Longueville, J. E., & Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. *Molecular biology and evolution*, 34(9), 2422–2424. <https://doi.org/10.1093/molbev/msx149>
- Leigh, J. W., & Bryant, D. (2015). POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9), 1110-1116.
- Lesk, A. (2019). *Introduction to bioinformatics*. Reino Unido: Oxford university press.
- Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research*, 44(W1), W242-W245. <https://doi.org/10.1093/nar/gkw290>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987-2993.
- Lin, Z., Xu, G., Deng, N., Taylor, C., Zhu, D., & Flemington, E. K. (2010). Quantitative and

qualitative RNA-Seq-based evaluation of Epstein-Barr virus transcription in type I latency Burkitt's lymphoma cells. *Journal of virology*, 84(24), 13053–13058. <https://doi.org/10.1128/JVI.01521-10>

Longnecker, R. (2000). Epstein-Barr virus latency: LMP2, a regulator or means for Epstein–Barr virus persistence? *Advances in Cancer Research*, (79),175–200.

Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of information in medicine*, 40(04), 346-358.

Magrath I. (2012). Epidemiology: clues to the pathogenesis of Burkitt lymphoma. *British journal of haematology*, 156(6), 744–756. <https://doi.org/10.1111/j.1365-2141.2011.09013.x>

Mainou, B. A., & Raab-Traub, N. (2006). LMP1 strain variants: biological and molecular properties. *Journal of virology*, 80(13), 6458–6468. <https://doi.org/10.1128/JVI.00135-06>

Makita, S., & Tobinai, K. (2017). Clinical Features and Current Optimal Management of Natural Killer/T-Cell Lymphoma. *Hematology/oncology clinics of North America*, 31(2), 239–253. <https://doi.org/10.1016/j.hoc.2016.11.007>

Martin, C., Carrasco, C., Sotomayor, C., & Oth, C. (2010). Prevalence of Epstein–Barr virus associated with nasal lymphoma in patients attending the regional hospital of Valdivia, Chile, between 1987 and 2005. *Journal of medical virology*, 82(5), 825-828.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus evolution*, 1(1), vev003. <https://doi.org/10.1093/ve/vev003>

Mbulaiteye, S. M., Biggar, R. J., Bhatia, K., Linet, M. S., & Devesa, S. S. (2009). Sporadic childhood Burkitt lymphoma incidence in the United States during 1992-2005. *Pediatric blood & cancer*, 53(3), 366–370. <https://doi.org/10.1002/pbc.22047>

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., & Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC*

bioinformatics, 14, 60. <https://doi.org/10.1186/1471-2105-14-60>

- Miller, J. A., Sahoo, M. K., Huang, C., Wang, H., Le, Q. T., & Pinsky, B. (2021). Multiplex Epstein-Barr Virus Genotyping PCR Detects High-Risk Variants in Plasma for Population-Level Nasopharyngeal Carcinoma Screening. *International Journal of Radiation Oncology, Biology, Physics*, 111(3), e386. <https://doi.org/10.1016/j.ijrobp.2021.07.1129>
- Monterroso, V., Zhou, Y., Koo, S., Glackin, C., Bujan, W., & Medeiros, L. J. (1998). Hodgkin's disease in Costa Rica: a report of 40 cases analyzed for Epstein-Barr virus. *American journal of clinical pathology*, 109(5), 618-624.
- Montes-Mojarro, I. A., Chen, B. J., Ramirez-Ibarguen, A. F., Quezada-Fiallos, C. M., Pérez-Báez, W. B., Dueñas, D., Casavilca-Zambrano, S., Ortiz-Mayor, M., Rojas-Bilbao, E., García-Rivello, H., Metrebian, M. F., Narbaitz, M., Barrionuevo, C., Lome-Maldonado, C., Bonzheim, I., Fend, F., Steinhilber, J., & Quintanilla-Martinez, L. (2020). Mutational profile and EBV strains of extranodal NK/T-cell lymphoma, nasal type in Latin America. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc*, 33(5), 781–791. <https://doi.org/10.1038/s41379-019-0415-5>
- Morton, L. M., Wang, S. S., Devesa, S. S., Hartge, P., Weisenburger, D. D., & Linet, M. S. (2006). Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001. *Blood*, 107(1), 265–276. <https://doi.org/10.1182/blood-2005-06-2508>
- Muwanika, V. B., Nyakaana, S., Siegismund, H. R., & Arctander, P. (2003). Phylogeography and population structure of the common warthog (*Phacochoerus africanus*) inferred from variation in mitochondrial DNA sequences and microsatellite loci. *Heredity*, 91(4), 361–372. <https://doi.org/10.1038/sj.hdy.6800341>
- Nagamine, M., Takahara, M., Kishibe, K., Nagato, T., Ishii, H., Bandoh, N., Ogino, T., & Harabuchi, Y. (2007). Sequence variations of Epstein-Barr virus LMP1 gene in nasal NK/T-cell lymphoma. *Virus genes*, 34(1), 47–54. <https://doi.org/10.1007/s11262-006-0008-5>
- Neves, M., Marinho-Dias, J., Ribeiro, J., & Sousa, H. (2016). Epstein-Barr virus strains and variations: Geographic or disease-specific variants?. *Journal of medical virology*, 89(3),

373–387. <https://doi.org/10.1002/jmv.24633>

Niller, H. H., & Bauer, G. (2017). Epstein-Barr Virus: Clinical Diagnostics. *Methods in molecular biology (Clifton, N.J.)*, 1532, 33–55. https://doi.org/10.1007/978-1-4939-6655-4_2

Nogová, L., Reineke, T., Brillant, C., Sieniawski, M., Rüdiger, T., Josting, A., Bredenfeld, H., Skripnitchenko, R., Müller, R. P., Müller-Hermelink, H. K., Diehl, V., Engert, A., & German Hodgkin Study Group (2008). Lymphocyte-predominant and classical Hodgkin's lymphoma: a comprehensive analysis from the German Hodgkin Study Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 26(3), 434–439. <https://doi.org/10.1200/JCO.2007.11.8869>

Nowalk, A., & Green, M. (2016). Epstein-Barr Virus. *Microbiology spectrum*, 4(3), 10.1128/microbiolspec.DMIH2-0011-2015. <https://doi.org/10.1128/microbiolspec.DMIH2-0011-2015>

Okonechnikov, K., Golosova, O., Fursov, M., & UGENE team (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics (Oxford, England)*, 28(8), 1166–1167. <https://doi.org/10.1093/bioinformatics/bts091>

Padidam, M., Sawyer, S., & Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology*, 265(2), 218–225. <https://doi.org/10.1006/viro.1999.0056>

Palma, I., Sánchez, A. E., Jiménez-Hernández, E., Alvarez-Rodríguez, F., Nava-Frias, M., Valencia-Mayoral, P., ... & Arellano-Galindo, J. (2013). Detection of Epstein-Barr virus and genotyping based on EBNA2 protein in Mexican patients with Hodgkin lymphoma: a comparative study in children and adults. *Clinical Lymphoma Myeloma and LeReino Unidoemia*, 13(3), 266-272.

Palser, A. L., Grayson, N. E., White, R. E., Corton, C., Correia, S., Watson, S. J., ... & Rickinson, A. B. (2015). Genome diversity of Epstein-Barr virus from multiple tumor types and normal infection. *Journal of virology*, 89(10), 5222-5237. DOI: 10.1128/JVI.03614-14

Paradis, E. (2018). Analysis of haplotype networks: The randomized minimum spanning tree

method. *Methods in Ecology and Evolution*, 9(5), 1308-1317. <https://doi.org/10.1111/2041-210X.12969>

- Paradis, E. & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526-528.
- Park, S., Lee, J., Ko, Y. H., Han, A., Jun, H. J., Lee, S. C., Hwang, I. G., Park, Y. H., Ahn, J. S., Jung, C. W., Kim, K., Ahn, Y. C., Kang, W. K., Park, K., & Kim, W. S. (2007). The impact of Epstein-Barr virus status on clinical outcome in diffuse large B-cell lymphoma. *Blood*, 110(3), 972–978. <https://doi.org/10.1182/blood-2007-01-067769>
- Pedersen, B. S., Brown, J. M., Dashnow, H., Wallace, A. D., Velinder, M., Tristani-Firouzi, M., Schiffman, J. D., Tvrđik, T., Mao, R., Best, D. H., Bayrak-Toydemir, P., & Quinlan, A. R. (2021). Effective variant filtering and expected candidate variant yield in studies of rare human disease. *NPJ genomic medicine*, 6(1), 60. <https://doi.org/10.1038/s41525-021-00227-3>
- Peh, S. C., Kim, L. H., & Poppema, S. (2002). Frequent presence of subtype A virus in Epstein-Barr virus-associated malignancies. *Pathology*, 34(5), 446–450. <https://doi.org/10.1080/0031302021000009379>
- Peng, R. J., Han, B. W., Cai, Q. Q., Zuo, X. Y., Xia, T., Chen, J. R., ... & Guo, Y. M. (2019). Genomic and transcriptomic landscapes of Epstein-Barr virus in extranodal natural killer T-cell lymphoma. *LeReino Unidoemia*, 33(6), 1451-1462. <https://doi.org/10.1038/s41375-018-0324-5>
- Poppema S. (2021). Lymphocyte predominant Hodgkin lymphoma, antigen-driven after all?. *The Journal of pathology*, 253(1), 1–10. <https://doi.org/10.1002/path.5567>
- Posada, D., & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 98(24), 13757–13762. <https://doi.org/10.1073/pnas.241370698>

- Posada, D., & Crandall, K. A. (2001). Intraspecific gene genealogies: trees grafting into networks. *Trends in ecology & evolution*, *16*(1), 37–45. [https://doi.org/10.1016/s0169-5347\(00\)02026-7](https://doi.org/10.1016/s0169-5347(00)02026-7)
- Posada, D., & Crandall, K. A. (2002). The effect of recombination on the accuracy of phylogeny estimation. *Journal of molecular evolution*, *54*(3), 396–402. <https://doi.org/10.1007/s00239-001-0034-9>
- Posada, D., Crandall, K. A., & Holmes, E. C. (2002). Recombination in evolutionary genomics. *Annual review of genetics*, *36*, 75–97. <https://doi.org/10.1146/annurev.genet.36.040202.111115>
- Preiksaitis, J. K., Pang, X. L., Fox, J. D., Fenton, J. M., Caliendo, A. M., Miller, G. G., & American Society of Transplantation Infectious Diseases Community of Practice (2009). Interlaboratory comparison of epstein-barr virus viral load assays. *American journal of transplantation: official journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, *9*(2), 269–279. <https://doi.org/10.1111/j.1600-6143.2008.02514.x>
- Prescott, L., Harley, J. & Klein, D. (2002). *Microbiología*. Madrid, España: McGraw-Hill-Interamericana.
- Quijano, S. M., Saavedra, C., Bravo, M. M., Fiorentino, S., & Orozco, O. (2003). Expression of heat shock proteins HSP72 and HSP73 in Colombian patients with Hodgkin lymphoma positive and negative for Epstein Barr virus. *Revista medica de Chile*, *131*(12), 1375-1381.
- Quijano, S., Saavedra, C., Fiorentino, S., Orozco, O., & Bravo, M. M. (2004). Epstein-Barr virus presence in Colombian Hodgkin lymphoma cases and its relation to treatment response. *Biomedica*, *24*(2), 163-173.
- Quintanilla-Martinez, L. (2017). The 2016 updated WHO classification of lymphoid neoplasias. *Hematological oncology*, *35*, 37-45. <https://doi.org/10.1002/hon.2399>
- Quintanilla-Martinez, L., Sander, B., Chan, J. K., Xerri, L., Ott, G., Campo, E., & Swerdlow, S. H.

(2016). Indolent lymphomas in the pediatric population: follicular lymphoma, IRF4/MUM1+ lymphoma, nodal marginal zone lymphoma and chronic lymphocytic leukemia. *Virchows Archiv*, 468(2), 141-157. <https://doi.org/10.1007/s00428-015-1855-z>

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Recuperado de <https://www.R-project.org/>.

Radkov, S. A., Bain, M., Farrell, P. J., West, M., Rowe, M., & Allday, M. J. (1997). Epstein-Barr virus EBNA3C represses Cp, the major promoter for EBNA expression, but has no effect on the promoter of the cell gene CD21. *Journal of virology*, 71(11), 8552-8562.

Rambaut, A., & Drummond, A. J. (2018). FigTree v1. 4.4. Institute of Evolutionary Biology. University of Edinburgh, Edinburgh. Recuperado de <http://tree.bio.ed.ac.uk/software/figtree/>

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.*, 3, 217-223. doi:10.1111/j.2041-210X.2011.00169.x

Rodríguez, F., Oliver, J. L., Marín, A., & Medina, J. R. (1990). The general stochastic model of nucleotide substitution. *Journal of theoretical biology*, 142(4), 485–501. [https://doi.org/10.1016/s0022-5193\(05\)80104-3](https://doi.org/10.1016/s0022-5193(05)80104-3)

Rowe, M., Glaunsinger, B., van Leeuwen, D., Zuo, J., Sweetman, D., Ganem, D., Middeldorp, J., Wiertz, E. J., & Rensing, M. E. (2007). Host shutoff during productive Epstein-Barr virus infection is mediated by BGLF5 and may contribute to immune evasion. *Proceedings of the National Academy of Sciences of the United States of America*, 104(9), 3366–3371. <https://doi.org/10.1073/pnas.0611128104>

Rowe, M., Lear, A. L., Croom-Carter, D., Davies, A. H., & Rickinson, A. B. (1992). Three pathways of Epstein-Barr virus gene activation from EBNA1-positive latency in B lymphocytes. *Journal of virology*, 66(1), 122–131. <https://doi.org/10.1128/JVI.66.1.122-131.1992>

Rowe, M., Rowe, D. T., Gregory, C. D., Young, L. S., Farrell, P. J., Rupani, H., & Rickinson, A. B.

(1987). Differences in B cell growth phenotype reflect novel patterns of Epstein-Barr virus latent gene expression in Burkitt's lymphoma cells. *The EMBO journal*, 6(9), 2743–2751.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular biology and evolution*, 34(12), 3299–3302. <https://doi.org/10.1093/molbev/msx248>

Salahuddin, S., Khan, J., Azhar, J., B. Whitehurst, C., Qadri, I., Shackelford, J., Pagano, J. S., Muhammad, D., & Richards, K. L. (2018). Prevalence of Epstein–Barr Virus Genotypes in Pakistani Lymphoma Patients. *Asian Pacific journal of cancer prevention : APJCP*, 19(11), 3153–3159. <https://doi.org/10.31557/APJCP.2018.19.11.3153>

Saleh, K., Michot, J. M., Camara-Clayette, V., Vassetsky, Y., & Ribrag, V. (2020). Burkitt and Burkitt-like lymphomas: a systematic review. *Current oncology reports*, 22(4), 1-9. <https://doi.org/10.1007/s11912-020-0898-8>

Sánchez-Giler, S., Herrera-Vásquez, A., Castillo-Zambrano, C., Solórzano-Alava, L., Zambrano-Castro, D., Kuon-Yeng, C., ... & Cucalón-González, G. (2018). Detección del Virus de Epstein-Barr en linfoma mediante qPCR//Detection of Epstein-Barr Virus (EBV) in lymphoma through qPCR. *Ciencia Unemi*, 11(26), 126-133.

Sánchez-Romero, C., Paes de Almeida, O., Rendón Henao, J., & Carlos, R. (2019). Extranodal NK/T-Cell Lymphoma, Nasal Type in Guatemala: An 86-Case Series Emphasizing Clinical Presentation and Microscopic Characteristics. *Head and neck pathology*, 13(4), 624–634. <https://doi.org/10.1007/s12105-019-01027-z>

Schwartz, E. J., Molina-Kirsch, H., Zhao, S., Marinelli, R. J., Warnke, R. A., & Natkunam, Y. (2008). Immunohistochemical characterization of nasal-type extranodal NK/T-cell lymphoma using a tissue microarray: an analysis of 84 cases. *American journal of clinical pathology*, 130(3), 343-351.

See, H. S., Yap, Y. Y., Yip, W. K., & Seow, H. F. (2008). Epstein-Barr virus latent membrane protein-1 (LMP-1) 30-bp deletion and Xho I-loss is associated with type III nasopharyngeal

carcinoma in Malaysia. *World journal of surgical oncology*, 6, 18.
<https://doi.org/10.1186/1477-7819-6-18>

Shapiro, R. S., McClain, K., Frizzera, G., Gajl-Peczalska, K. J., Kersey, J. H., Blazar, B. R., Arthur, D. C., Patton, D. F., Greenberg, J. S., & Burke, B. (1988). Epstein-Barr virus associated B cell lymphoproliferative disorders following bone marrow transplantation. *Blood*, 71(5), 1234–1243.

Sheng, W., Decaussin, G., Sumner, S., & Ooka, T. (2001). N-terminal domain of BARP1 gene encoded by Epstein-Barr virus is essential for malignant transformation of rodent fibroblasts and activation of BCL-2. *Oncogene*, 20(10), 1176–1185.
<https://doi.org/10.1038/sj.onc.1204217>

Shimoyama, Y., Yamamoto, K., Asano, N., Oyama, T., Kinoshita, T., & Nakamura, S. (2008). Age-related Epstein–Barr virus-associated B-cell lymphoproliferative disorders: special references to lymphomas surrounding this newly recognized clinicopathologic disease. *Cancer science*, 99(6), 1085-1091.

Sibley, L. D., & Ajioka, J. W. (2008). Population structure of *Toxoplasma gondii*: clonal expansion driven by infrequent recombination and selective sweeps. *Annual review of microbiology*, 62, 329–351. <https://doi.org/10.1146/annurev.micro.62.081307.162925>

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current protocols in molecular biology*, 122(1), e59.
<https://doi.org/10.1002/cpmb.59>

Smatti, M. K., Al-Sadeq, D. W., Ali, N. H., Pintus, G., Abou-Saleh, H., & Nasrallah, G. K. (2018). Epstein-Barr Virus Epidemiology, Serology, and Genetic Variability of LMP-1 Oncogene Among Healthy Population: An Update. *Frontiers in oncology*, 8, 211.
<https://doi.org/10.3389/fonc.2018.00211>

Smith J. M. (1992). Analyzing the mosaic structure of genes. *Journal of molecular evolution*, 34(2), 126–129. <https://doi.org/10.1007/BF00182389>

Sublime. (2019). Sublime Text Version 3.2.2. Recuperado de <https://www.sublimetext.com>

Sun, C. C., & Thorley-Lawson, D. A. (2007). Plasma cell-specific transcription factor XBP-1s binds to and transactivates the Epstein-Barr virus BZLF1 promoter. *Journal of virology*, *81*(24), 13566–13577. <https://doi.org/10.1128/JVI.01055-07>

Sun, J., Yang, Q., Lu, Z., He, M., Gao, L., Zhu, M., Sun, L., Wei, L., Li, M., Liu, C., Zheng, J., Liu, W., Li, G., & Chen, J. (2012). Distribution of lymphoid neoplasms in China: analysis of 4,638 cases according to the World Health Organization classification. *American journal of clinical pathology*, *138*(3), 429–434. <https://doi.org/10.1309/AJCP7YLTQPUSDQ5C>

Swerdlow, S., Campo, E., Harris, N. L., Jaffe, E. S., Pileri, S. A., Stein, H. & Thiele, J. (2017). WHO classification of tumours of haematopoietic and lymphoid tissues (Vol. 2, p. 439). Lyon, France: International agency for research on cancer.

Takada K. (2012). Role of EBER and BARF1 in nasopharyngeal carcinoma (NPC) tumorigenesis. *Seminars in cancer biology*, *22*(2), 162–165. <https://doi.org/10.1016/j.semcancer.2011.12.007>

Temple, R. M., Zhu, J., Budgeon, L., Christensen, N. D., Meyers, C., & Sample, C. E. (2014). Efficient replication of Epstein-Barr virus in stratified epithelium in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(46), 16544–16549. <https://doi.org/10.1073/pnas.1400818111>

Templeton, A. R., Crandall, K. A., & Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, *132*(2), 619–633.

Teras, L. R., DeSantis, C. E., Cerhan, J. R., Morton, L. M., Jemal, A., & Flowers, C. R. (2016). 2016 US lymphoid malignancy statistics by World Health Organization subtypes. *CA: a cancer journal for clinicians*, *66*(6), 443-459. <https://doi.org/10.3322/caac.21357>

Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S., & Corander, J. (2018). RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. *Wellcome open research*,

3, 93. <https://doi.org/10.12688/wellcomeopenres.14694.1>

- Tsai, D. E., Douglas, L., Andreadis, C., Vogl, D. T., Arnoldi, S., Kotloff, R., Svoboda, J., Bloom, R. D., Olthoff, K. M., Brozena, S. C., Schuster, S. J., Stadtmauer, E. A., Robertson, E. S., Wasik, M. A., & Ahya, V. N. (2008). EBV PCR in the diagnosis and monitoring of posttransplant lymphoproliferative disorder: results of a two-arm prospective trial. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, 8(5), 1016–1024. <https://doi.org/10.1111/j.1600-6143.2008.02183.x>
- Tsao, S. W., Tsang, C. M., To, K. F., & Lo, K. W. (2015). The role of Epstein–Barr virus in epithelial malignancies. *The Journal of pathology*, 235(2), 323–333. <https://doi.org/10.1002/path.4448>
- Tzellos, S., & Farrell, P. J. (2012). Epstein-barr virus sequence variation-biology and disease. *Pathogens*, 1(2), 156–174. <https://doi.org/10.3390/pathogens1020156>
- Vázquez-Domínguez, Ella, Castañeda-Rico, Susette, Garrido-Garduño, Tania, & Gutiérrez-García, Tania A. (2009). Avances metodológicos para el estudio conjunto de la información genética, genealógica y geográfica en análisis evolutivos y de distribución. *Revista chilena de historia natural*, 82(2), 277–297. <https://dx.doi.org/10.4067/S0716-078X2009000200009>
- Verhoeven, R. J., Tong, S., Mok, B. W. Y., Liu, J., He, S., Zong, J., ... & Chen, H. (2019). Epstein-barr virus BART long non-coding RNAs function as epigenetic modulators in nasopharyngeal carcinoma. *Frontiers in oncology*, 9, 1120.
- Wadowsky, R. M., Laus, S., Green, M., Webber, S. A., & Rowe, D. (2003). Measurement of Epstein-Barr virus DNA loads in whole blood and plasma by TaqMan PCR and in peripheral blood lymphocytes by competitive PCR. *Journal of clinical microbiology*, 41(11), 5245–5249. <https://doi.org/10.1128/jcm.41.11.5245-5249.2003>
- Walling, D. M., Brown, A. L., Etienne, W., Keitel, W. A., & Ling, P. D. (2003). Multiple Epstein-Barr virus infections in healthy individuals. *Journal of virology*, 77(11), 6546–6550. <https://doi.org/10.1128/jvi.77.11.6546-6550.2003>

- Wang, X., Liu, X., Jia, Y., Chao, Y., Xing, X., Wang, Y., & Luo, B. (2010). Widespread sequence variation in the Epstein-Barr virus latent membrane protein 2A gene among northern Chinese isolates. *The Journal of general virology*, *91*(10), 2564–2573. <https://doi.org/10.1099/vir.0.021881-0>
- Weinreb, M., Day, P. J., Niggli, F., Powell, J. E., Raafat, F., Hesseling, P. B., ... & Mann, J. R. (1996). The role of Epstein-Barr virus in Hodgkin's disease from different geographical areas. *Archives of disease in childhood*, *74*(1), 27-31.
- Wen, W., Iwakiri, D., Yamamoto, K., Maruo, S., Kanda, T., & Takada, K. (2007). Epstein-Barr virus BZLF1 gene, a switch from latency to lytic infection, is expressed as an immediate-early gene after primary infection of B lymphocytes. *Journal of virology*, *81*(2), 1037–1042. <https://doi.org/10.1128/JVI.01416-06>
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, 1-25. DOI: 10.2307/1912526
- White, R. E., Rämer, P. C., Naresh, K. N., Meixlsperger, S., Pinaud, L., Rooney, C., ... & Allday, M. J. (2012). EBNA3B-deficient EBV promotes B cell lymphomagenesis in humanized mice and is found in human tumors. *The Journal of clinical investigation*, *122*(4).
- Wu, C. C., Liu, M. T., Chang, Y. T., Fang, C. Y., Chou, S. P., Liao, H. W., Kuo, K. L., Hsu, S. L., Chen, Y. R., Wang, P. W., Chen, Y. L., Chuang, H. Y., Lee, C. H., Chen, M., Wayne Chang, W. S., & Chen, J. Y. (2010). Epstein-Barr virus DNase (BGLF5) induces genomic instability in human epithelial cells. *Nucleic acids research*, *38*(6), 1932–1949. <https://doi.org/10.1093/nar/gkp1169>
- Xiong, J., Cui, B. W., Wang, N., Dai, Y. T., Zhang, H., Wang, C. F., ... & Zhang, X. (2020). Genomic and Transcriptomic Characterization of Natural Killer T Cell Lymphoma. *Cancer Cell*, *37*(3), 403-419. <https://doi.org/10.1016/j.ccell.2020.02.005>
- Xiong, J., & Zhao, W. L. (2018). Advances in multiple omics of natural-killer/T cell lymphoma. *Journal of hematology & oncology*, *11*(1), 134. <https://doi.org/10.1186/s13045-018-0678-1>

- Xu, M., Yao, Y., Chen, H., Zhang, S., Cao, S. M., Zhang, Z., Luo, B., Liu, Z., Li, Z., Xiang, T., He, G., Feng, Q. S., Chen, L. Z., Guo, X., Jia, W. H., Chen, M. Y., Zhang, X., Xie, S. H., Peng, R., Chang, E. T., ... Liu, J. (2019). Genome sequencing analysis identifies Epstein-Barr virus subtypes associated with high risk of nasopharyngeal carcinoma. *Nature genetics*, *51*(7), 1131–1136. <https://doi.org/10.1038/s41588-019-0436-5>
- Yamaguchi, M., Oguchi, M., & SuzReino Unidoi, R. (2018). Extranodal NK/T-cell lymphoma: Updates in biology and management strategies. Best practice & research. *Clinical haematology*, *31*(3), 315–321. <https://doi.org/10.1016/j.beha.2018.07.002>
- Yang, Q. P., Zhang, W. Y., Yu, J. B., Zhao, S., Xu, H., Wang, W. Y., Bi, C. F., Zuo, Z., Wang, X. Q., Huang, J., Dai, L., & Liu, W. P. (2011). Subtype distribution of lymphomas in Southwest China: analysis of 6,382 cases using WHO classification in a single institution. *Diagnostic pathology*, *6*, 77. <https://doi.org/10.1186/1746-1596-6-77>
- Yang, Y., Wang, Y., Liu, X., He, X., Zhang, L. L., Wu, G., Qu, B. L., Qian, L. T., Hou, X. R., Zhang, F. Q., Qiao, X. Y., Wang, H., Li, G. F., Zhu, Y., Cao, J. Z., Wu, J. X., Wu, T., Zhu, S. Y., Shi, M., Xu, L. M., ... Li, Y. X. (2021). Progression-free survival at 24 months and subsequent survival of patients with extranodal NK/T-cell lymphoma: a China Lymphoma Collaborative Group (CLCG) study. *Leukemia*, *35*(6), 1671–1682. <https://doi.org/10.1038/s41375-020-01042-y>
- Yates, J. L., Camiolo, S. M., & Bashaw, J. M. (2000). The minimal replicator of Epstein-Barr virus oriP. *Journal of virology*, *74*(10), 4512–4522. <https://doi.org/10.1128/jvi.74.10.4512-4522.2000>
- Yohe, S., & Thyagarajan, B. (2017). Review of Clinical Next-Generation Sequencing. *Archives of pathology & laboratory medicine*, *141*(11), 1544–1557. <https://doi.org/10.5858/arpa.2016-0501-RA>
- Yoshida, R., & Nei, M. (2016). Efficiencies of the NJp, Maximum Likelihood, and Bayesian Methods of Phylogenetic Construction for Compositional and Noncompositional Genes. *Molecular biology and evolution*, *33*(6), 1618–1624. <https://doi.org/10.1093/molbev/msw042>

- Young, L., Alfieri, C., Hennessy, K., Evans, H., O'Hara, C., Anderson, K. C., Ritz, J., Shapiro, R. S., Rickinson, A., & Kieff, E. (1989). Expression of Epstein-Barr virus transformation-associated genes in tissues of patients with EBV lymphoproliferative disease. *The New England journal of medicine*, *321*(16), 1080–1085. <https://doi.org/10.1056/NEJM198910193211604>
- Young, L. S., Arrand, J. R., & Murray, P. G. (2007). *"EBV gene expression and regulation."* *Human herpesviruses: Biology, therapy, and immunoprophylaxis*. Ingleterra: Cambridge University Press.
- Young, L. S., & Rickinson, A. B. (2004). Epstein-Barr virus: 40 years on. *Nature reviews. Cancer*, *4*(10), 757–768. <https://doi.org/10.1038/nrc1452>
- Young, L. S., Yap, L. F., & Murray, P. G. (2016). Epstein–Barr virus: more than 50 years old and still providing surprises. *Nature Reviews Cancer*, *16*(12), 789-802. <https://doi.org/10.1038/nrc.2016.92>
- Yu, G., Smith, D., Zhu, H., Guan, Y. & Lam, T. (2017). “ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data.” *Methods in Ecology and Evolution*, *8*, 28-36. doi: 10.1111/2041-210X.12628
- Zanella, L., Riquelme, I., Buchegger, K., Abanto, M., Ili, C., & Brebi, P. (2019). A reliable Epstein-Barr Virus classification based on phylogenomic and population analyses. *Scientific reports*, *9*(1), 1-11. <https://doi.org/10.1038/s41598-019-45986-3>
- Zhang, X. S., Song, K. H., Mai, H. Q., Jia, W. H., Feng, B. J., Xia, J. C., Zhang, R. H., Huang, L. X., Yu, X. J., Feng, Q. S., Huang, P., Chen, J. J., & Zeng, Y. X. (2002). The 30-bp deletion variant: a polymorphism of latent membrane protein 1 prevalent in endemic and non-endemic areas of nasopharyngeal carcinomas in China. *Cancer letters*, *176*(1), 65–73. [https://doi.org/10.1016/s0304-3835\(01\)00733-9](https://doi.org/10.1016/s0304-3835(01)00733-9)
- Zeng, M. S., Li, D. J., Liu, Q. L., Song, L. B., Li, M. Z., Zhang, R. H., Yu, X. J., Wang, H. M., Ernberg, I., & Zeng, Y. X. (2005). Genomic sequence analysis of Epstein-Barr virus strain GD1 from a nasopharyngeal carcinoma patient. *Journal of virology*, *79*(24), 15323–15330.

<https://doi.org/10.1128/JVI.79.24.15323-15330.2005>

Zuo, J., Currin, A., Griffin, B. D., Shannon-Lowe, C., Thomas, W. A., Rensing, M. E., Wiertz, E. J., & Rowe, M. (2009). The Epstein-Barr virus G-protein-coupled receptor contributes to immune evasion by targeting MHC class I molecules for degradation. *PLoS pathogens*, 5(1), e1000255. <https://doi.org/10.1371/journal.ppat.1000255>

12. ANEXOS

Anexo 1. Listado de genomas analizados.

No.	Isolado	Referencia	País	Enfermedad
1	B95.8Raji	NC_007605	Estados Unidos	IM
2	AG876	NC_009334	Ghana	BL
3	YCCEL1	AP015016	Corea del Sur	CG Cell Line
4	M81	KF373730	Hong Kong	NPC
5	Cheptages	LN827556	Kenia	BL
6	sLCL-IS1.20	LN827576	Australia	sLCL, PTLD
7	sLCL-2.16	LN827580	Kenia	sLCL
8	sLCL-IM1.02	LN827596	Australia	sLCL, IM
9	sLCL-2.14	LN827560	Kenia	sLCL
10	sLCL-2.21	LN827587	Kenia	sLCL
11	sLCL-IS1.13	LN827578	Australia	sLCL, PTLD
12	Jjjoy	LN827800	Nigeria	BL
13	sLCL-IS1.14	LN827575	Australia	sLCL, PTLD
14	sLCL-2.15	LN827591	Kenia	sLCL
15	sLCL-IM1.05	LN827590	Australia	sLCL, IM
16	sLCL-2.22	LN831023	Kenia	sLCL
17	Daudi	LN827545	Kenia	BL
18	P3HR1 c16	LN827548	Nigeria	BL
19	sLCL-IM1.09	LN827567	Australia	sLCL, IM
20	sLCL-IM1.17	LN827583	Australia	sLCL, IM
21	sLCL-IS2.01	LN827589	Australia	sLCL, PTLD
22	HKN14	LN824209	Hong Kong	sLCL
23	SLCL-IS1.19	LN827588	Australia	sLCL, PTLD
24	HL11	LN827524	Reino Unido	HL
25	Wewak1	LN827544	Papúa Nueva Guinea	BL
26	B136	LN827557	Norte de África	BL
27	HKN15	LN827547	Hong Kong	sLCL
28	HKN19	LN824224	Hong Kong	sLCL
29	sLCL-IS1.01	LN827570	Australia	sLCL, PTLD
30	sLCL-IS1.11	LN827569	Australia	sLCL, PTLD
31	sLCL-IS1.10	LN827592	Australia	sLCL, PTLD

32	sLCL-IS1.18	LN827572	Australia	sLCL, PTLD
33	sLCL-IS1.12	LN827593	Australia	sLCL, PTLD
34	EBVaGC1	KT273942	China	CG
35	M-ABA	LN827527	Norte de África	LCL, NPC virus
36	HL04	LN827564	Reino Unido	HL
37	sLCL-1.12	LN824205	Australia	sLCL
38	SLCL-IS1.07	LN827594	Australia	sLCL, PTLD
39	HL09	LN827522	Reino Unido	HL
40	D3201.2	LN827549	China	NPC biopsy
41	YCCEL1	LN827561	Corea del Sur	CG Cell Line
42	sLCL-BL1.03	LN827582	Kenia	sLCL
43	B137	LN827526	África	BL
44	sLCL-1.19	LN827562	Kenia	sLCL
45	sLCL-IS1.15	LN827586	Australia	sLCL, PTLD
46	C666-1 resequence	LN827525	China	NPC Cell Line
47	sLCL-IM1.05	LN827581	Kenia	sLCL, IM
48	sLCL-1.11	LN827550	Kenia	sLCL
49	sLCL-1.18	LN827563	Kenia	sLCL
50	sLCL-1.06	LN827566	Kenia	sLCL
51	sLCL-BL1.20	LN827571	Kenia	sLCL
52	sLCL-1.04	LN827585	Kenia	sLCL
53	sLCL-IM1.16	LN827799	Australia	sLCL, IM
54	sLCL-1.17	LN827577	Kenia	sLCL
55	sLCL-1.08	LN827552	Kenia	sLCL
56	sLCL-1.24	LN827568	Kenia	sLCL
57	HL02	LN827546	Reino Unido	HL
58	HL08	LN824225	Reino Unido	HL
59	sLCL-1.10	LN827573	Kenia	sLCL
60	LCLB95-8-del-EBER2 re sequence	LN827739	Estados Unidos	LCL
61	sLCL-1.07	LN827565	Kenia	sLCL
62	RPF	KR063344	Brasil	BL
63	sLCL-1.02	LN827558	Kenia	sLCL
64	sLCL-IS1.08	LN827553	Australia	sLCL, PTLD
65	EBVaGC9	KT273949	China	CG
66	PLCL-TRL595	LN827559	Estados Unidos	sLCL

67	K4413-Mi	KC440852	Estados Unidos	sLCL
68	Mak1 duplicate	LN824203	Kenia	BL
69	sLCL-IS1.06	LN827584	Australia	sLCL, PTLD
70	FNR	KR063345	Brasil	BL
71	H03753A	KR063342	Argentina	BL
72	H002213	KP968264	Argentina	BL
73	B95.8Raji	AJ507799	Estados Unidos	IM
74	H058015C	KP968263	Argentina	BL
75	CV-ARG	KR063343	Argentina	BL
76	SCL	KP968259	Brasil	BL
77	X50-7	LN827555	Estados Unidos	LCL
78	SG	KT001103	Argentina	BL
79	HU11393	KP968261	Ghana	BL
80	HL01	LN824226	Reino Unido	HL
81	H018436D	KP968262	Argentina	BL
82	K4123-Mi	KC440851	Estados Unidos	sLCL
83	sLCL-1.13	LN827579	Kenia	sLCL
84	sLCL-IS1.03	LN827595	Australia	sLCL, PTLD
85	HL05	LN824204	Reino Unido	HL
86	HNNPC5	LC150338	China	NPC
87	sLCL-1.09	LN827574	Kenia	sLCL
88	CCH	KP968257	Brasil	BL
89	sLCL-IS1.04	LN827597	Australia	sLCL, PTLD
90	EBVaGC2	KT273943	China	CG
91	EBVaGC5	KT273945	China	CG
92	Akata resequence	LN824208	Japan	BL
93	LC3	KT823508	China	LC
94	Saliva1	LN824142	Reino Unido	Healthy saliva
95	Makau	LN827551	Kenia	BL
96	PLCL-TRL1-post	LN824206	Estados Unidos	sLCL
97	MP	KP968258	Brasil	BL
98	Mutu	KC207814	Kenia	BL
99	EBVaGC7	KT273947	China	CG
100	VGO	KP968260	Brasil	BL
101	EBVaGC8	KT273948	China	CG
102	EBVaGC4	KT273944	China	CG

103	GD1	AY961628	China	NPC saliva
104	LC2	KT823507	China	LC
105	HNNPC2	LC149491	China	NPC
106	EBVaGC3	KT254013	China	CG
107	EBVaGC6	KT273946	China	CG
108	LC4	KT823509	China	LC
109	L591	LN827523	Alemania	HL Cell Line
110	pLCL-TRL1-pre	LN824207	Estados Unidos	sLCL, PTLD
111	LC1	KT823506	China	LC
112	HKNPC3	KF992565	China	NPC Tumor
113	HKNPC6	KF992568	China	NPC Tumor
114	HKNPC1	JQ009376	China	NPC Tumor
115	HNNPC6	LC150741	China	NPC Tumor
116	HNNPC4	LC150337	China	NPC Tumor
117	HKNPC8	KF992570	China	NPC Tumor
118	HNNPC1	LC137018	China	NPC Tumor
119	HKNPC5	KF992567	China	NPC Tumor
120	Akata	KC207813	Japan	BL
121	C666-1	KC617875	China	NPC
122	HKNPC4	KF992566	China	NPC Tumor
123	HNNPC7	LC150742	China	NPC Tumor
124	HNNPC3	LC150327	China	NPC Tumor
125	C666-1	KJ411974	China	NPC Cell Line
126	HKNPC9	KF992571	China	NPC Tumor
127	GC1	KP735248	Corea del Sur	CG Cell Line
128	HKNPC7	KF992569	China	NPC Tumor
129	SNU-719	KX125052	Corea del Sur	CG Cell Line
130	HKNPC2	KF992564	China	NPC Tumor
131	HNNPC8	LC150743	China	NPC Tumor
132	GC-EBV1	KX125050	China	CG
133	SNU719	AP015015	Corea del Sur	CG Cell Line
134	YCCEL1	KX125053	Corea del Sur	CG Cell Line
135	Raji	KF717093	Nigeria	BL
136	GD2	HQ020558	China	NPC-Tumor
137	E1583_BCv1	MF547453	Estados Unidos	IM
138	E1583_OWv1	MF547454	Estados Unidos	IM

139	E1583_BCv7	MF547455	Estados Unidos	IM
140	E1583_OWv7	MF547456	Estados Unidos	IM
141	E1587_BCv1	MF547457	Estados Unidos	IM
142	E1587_OWv1	MF547458	Estados Unidos	IM
143	E1587_OWv7	MF547460	Estados Unidos	IM
144	E1536_BCv1	MF547461	Estados Unidos	IM
145	E1536_OWv1	MF547462	Estados Unidos	IM
146	E1563_BCv1	MF547463	Estados Unidos	IM
147	E1536_OWv7	MF547464	Estados Unidos	IM
148	E1536_BCv7	MF547465	Estados Unidos	IM
149	E1548_BCv1	MF547466	Estados Unidos	IM
150	E1548_OWv1	MF547467	Estados Unidos	IM
151	E1548_BCv7	MF547468	Estados Unidos	IM
152	E1548_OWv7	MF547469	Estados Unidos	IM
153	E1563_OWv1	MF547470	Estados Unidos	IM
154	E1563_OWv7	MF547471	Estados Unidos	IM
155	E1563_BCV7	MF547472	Estados Unidos	IM
156	E1590_BCv1	MF547473	Estados Unidos	IM
157	E1590_OWv1	MF547474	Estados Unidos	IM
158	E1590_BCv7	MF547475	Estados Unidos	IM
159	E1590_OWv7	MF547476	Estados Unidos	IM
160	E1492_BCv1	MF547477	Estados Unidos	IM
161	E1492_OWv1	MF547478	Estados Unidos	IM
162	E1492_BCv7	MF547479	Estados Unidos	IM
163	E1492_OWv7	MF547480	Estados Unidos	IM
164	E1503_BCv1	MF547481	Estados Unidos	IM
165	E1503_OWv1	MF547482	Estados Unidos	IM
166	E1503_BCv7	MF547483	Estados Unidos	IM
167	E1503_OWv7	MF547484	Estados Unidos	IM
168	E1578_BCv1	MF547485	Estados Unidos	IM
169	E1578_BCv7	MF547486	Estados Unidos	IM
170	E1578_OWv7	MF547487	Estados Unidos	IM
171	E1578_OWv1	MF547488	Estados Unidos	IM
172	E1577_BCv1	MF547489	Estados Unidos	IM
173	E1577_OWv1	MF547490	Estados Unidos	IM
174	E1577_OWv7	MF547491	Estados Unidos	IM

175	E1577_BCv7	MF547492	Estados Unidos	IM
176	EBVaGC8-2	MG021314	China	CG
177	YCCEL1-GC1	MG021305	China	CG
178	EBVaGC8-3	MG021315	China	CG
179	EBVaGC5-1	MG021317	Estados Unidos	CG
180	Mutu-GC1	MG021308	Estados Unidos	CG
181	Akata-GC1	MG021307	Estados Unidos	CG
182	AG876-GC1	MG021312	Estados Unidos	CG
183	HKNPC6-GC1	MG021316	Vietnam	CG
184	Mutu-GC3	MG021310	Estados Unidos	CG
185	Mutu-GC4	MG021311	Estados Unidos	CG
186	Mutu-GC2	MG021309	Estados Unidos	CG
187	EBVaGC8-1	MG021313	Estados Unidos	CG
188	YCCEL1-GC2	MG021306	Polonia	CG
189	SC01	MH101957	China	ENKTL
190	SC02	MH101958	China	ENKTL
191	SC03	MH101959	China	ENKTL
192	SC04	MH101960	China	ENKTL
193	SC05	MH101961	China	ENKTL
194	SC06	MH101962	China	ENKTL
195	SC07	MH101963	China	ENKTL
196	SC08	MH101964	China	ENKTL
197	SC09	MH101965	China	ENKTL
198	SC10	MH101966	China	ENKTL
199	SC11	MH101967	China	ENKTL
200	SC12	MH101968	China	ENKTL
201	SC13	MH101969	China	ENKTL
202	SC14	MH101970	China	ENKTL
203	SC15	MH101971	China	ENKTL
204	SG01	MH144212	Singapor	ENKTL
205	SG02	MH144213	Singapor	ENKTL
206	SG03	MH144214	Singapor	ENKTL
207	SG04	MH144215	Singapor	ENKTL
208	SG05	MH144216	Singapor	ENKTL
209	SG06	MH144217	Singapor	ENKTL
210	SG07	MH144218	Singapor	ENKTL

211	SG08	MH144219	Singapor	ENKTL
212	SG09	MH144220	Singapor	ENKTL
213	SG10	MH144221	Singapor	ENKTL
214	SG11	MH144222	Singapor	ENKTL
215	SG12	MH144223	Singapor	ENKTL
216	Guatemala_INCAN_EBV00079	No Referencia	Guatemala	ENKTL
217	Guatemala_INCAN_EBV00080	No Referencia	Guatemala	ENKTL
218	Guatemala_INCAN_EBV00081	No Referencia	Guatemala	ENKTL
219	Guatemala_INCAN_EBV00082	No Referencia	Guatemala	ENKTL
220	Guatemala_INCAN_EBV00083	No Referencia	Guatemala	ENKTL
221	Guatemala_INCAN_EBV00084	No Referencia	Guatemala	ENKTL
222	Guatemala_INCAN_EBV00085	No Referencia	Guatemala	ENKTL
223	Guatemala_INCAN_EBV00086	No Referencia	Guatemala	ENKTL
224	Guatemala_INCAN_EBV00087	No Referencia	Guatemala	ENKTL
225	Guatemala_INCAN_EBV00088	No Referencia	Guatemala	ENKTL
226	Guatemala_INCAN_EBV00089	No Referencia	Guatemala	ENKTL
227	Guatemala_INCAN_EBV00090	No Referencia	Guatemala	ENKTL
228	Guatemala_INCAN_EBV00105	No Referencia	Guatemala	ENKTL
229	Guatemala_INCAN_EBV00106	No Referencia	Guatemala	ENKTL
230	Guatemala_INCAN_EBV00107	No Referencia	Guatemala	ENKTL
231	Guatemala_INCAN_EBV00108	No Referencia	Guatemala	ENKTL
232	Guatemala_INCAN_EBV00109	No Referencia	Guatemala	ENKTL
233	Guatemala_INCAN_EBV00110	No Referencia	Guatemala	ENKTL

234	Guatemala_INCAN_EBV00111	No Referencia	Guatemala	ENKTL
235	Guatemala_INCAN_EBV00112	No Referencia	Guatemala	ENKTL
236	Guatemala_INCAN_EBV00113	No Referencia	Guatemala	ENKTL
237	Guatemala_INCAN_EBV00114	No Referencia	Guatemala	ENKTL

Fuente: National Center for Biotechnology Information.

IM	Mononucleosis Infecciosa (Control)	PTLD	Enfermedad linfoproliferativa postrasplante
BL	Linfoma de Burkitt	HL	Linfoma de Hodgkin
CG	Cáncer Gástrico	LCL	Línea Celular Linfoblastoide
NPC	Cáncer Nasofaríngeo	ENKTL	Linfoma de T/NK Extranodal de Tipo Nasal
sLCL	Línea Celular Linfoblastoide Espontánea		

Anexo 2. Script de rhierbaps para llevar a cabo análisis de estructura poblacional.

```
-----  
rhierbaps  
-----  
-----
```

```
Seleccionar directorio  
-----  
-----
```

```
Cargar paquetes
```

```
library(rhierbaps)
```

```
library(ggtree)
```

```
library(phytools)
```

```
library(ape)
```

```
set.seed(1234)  
-----  
-----
```

```
Cargando archivo
```

```
fasta.file.name <- "Nombre_Alineamiento.fasta"
```

```
fasta.file.name  
-----  
-----
```

```
Creando la Matriz
```

```
snp.matrix <- load_fasta(fasta.file.name)
```

```
write.table(snp.matrix, "matriz_1.txt", sep = "\t", quote = T, row.names = T)  
-----  
-----
```

```
Corriendo hierBAPS
```

```
hb.results <- hierBAPS(snp.matrix, max.depth = 2, n.pops = 47, quiet = TRUE)
```

```
head(hb.results$partition.df)
```


 Visualizando el resultado de primer nivel de jerarquía

```
newick.file.name <- "Nombre.tree"
iqtree <- phytools::read.newick(newick.file.name)

gg <- ggtree(iqtree, layout = "circular")
gg <- gg %+ %+ hb.results$partition.df
gg <- gg + geom_tippoint(aes(color = factor(`level 1`)))
gg

gg <- ggtree(iqtree, layout = "circular", branch.length = "none")
gg <- gg %+ %+ hb.results$partition.df
gg <- gg + geom_tippoint(aes(color = factor(`level 1`)))
gg <- gg + theme(legend.position = "right")
gg <- gg + geom_tiplab(aes(label = `level 1`), size = 1, offset = 1)
gg
```


 Visualizando los resultados de segundo nivel de jerarquía

```
plot_sub_cluster(hb.results, iqtree, level = 1, sub.cluster = 1)
plot_sub_cluster(hb.results, iqtree, level = 1, sub.cluster = 2)
plot_sub_cluster(hb.results, iqtree, level = 1, sub.cluster = 3)
plot_sub_cluster(hb.results, iqtree, level = 1, sub.cluster = 4)
plot_sub_cluster(hb.results, iqtree, level = 1, sub.cluster = 5)
plot_sub_cluster(hb.results, iqtree, level = 1, sub.cluster = 6)
plot_sub_cluster(hb.results, iqtree, level = 1, sub.cluster = 7)
plot_sub_cluster(hb.results, iqtree, level = 1, sub.cluster = 8)
```

Anexo 3. Listado de Sitios SNP e INDEL del genoma del VEB asociado a ENKTL guatemalteco.

Posición	Referencia	Variante
100	C	G
159	C	G
391	G	A
410	C	A
571	G	A
613	C	G
629	G	C
632	C	G
1073	A	G
1105	C	A
1350	C	A
1374	A	C
1385	G	C
1392	T	C
1401	C	T
1438	T	A
1617	C	T
1741	A	G
1841	G	C
2186	A	C
2710	G	A
2777	G	A
2830	C	T
3064	G	A
3091	C	A
3101	G	A
3112	T	C
3224	A	G
3568	G	A
3694	G	T
3934	C	T
3942	G	T
4075	G	A
4125	A	G
4135	C	T
4246	C	A
4303	A	G
4381	T	C
4576	C	T
4690	A	C
4823	T	C
4855	G	A
4978	G	T
5064	T	C
5098	C	T
5323	C	T
5452	T	C

5746	T	G
5800	ACCTCCC	ACCC
5808	CCCA	C
6182	T	C
6255	A	G
6370	T	G
6618	T	A
6808	T	A
6866	G	A
7123	A	G
7447	A	C
7448	T	G
7474	C	T
7479	T	C
7499	A	G
8019	T	C
8028	C	T
8033	G	C
8122	T	C
8223	T	C
8243	G	A
8275	T	G
8304	G	A
8385	T	G
8568	A	T
8573	C	A
8594	C	A
8744	C	T
8832	T	C
8926	T	C
9019	A	G
9062	A	G
9154	C	T
9509	C	T
9518	G	A
9593	C	T
9690	G	A
9741	G	A
9747	G	A
9980	C	A
10566	G	A
10726	T	C
10751	T	C
10926	G	T
10935	T	C
11039	A	C
11089	C	T
11747	G	A
35337	G	C
35339	T	C

35624	A	G
35759	C	T
35857	T	G
36046	G	A
36068	T	C
36227	C	T
36471	C	T
36480	T	C
36497	C	T
36668	T	G
36711	C	G
36769	A	G
36793	C	T
36818	C	T
36848	A	ACTC
36853	T	C
36858	C	A
37050	C	G
37054	C	A
37288	G	A
37325	T	A
37466	G	A
37642	A	G
37672	T	C
37683	C	A
37908	G	A
37955	G	T
38034	G	C
38035	G	A
38083	T	G
38301	T	A
40218	G	A
40665	T	C
41354	C	A
41370	A	G
41403	A	G
41672	G	A
41704	G	A
41715	T	G
41835	T	C
41840	G	C
41847	T	C
41888	A	T
41893	C	T
41904	T	C
41952	T	C
42157	A	G
42200	T	C
42319	C	A
42322	G	T

42327	T	C
42446	C	A
42553	G	A
42567	A	C
42712	C	T
42932	T	G
43281	A	G
43675	T	C
43820	C	T
43844	A	G
44123	T	G
44450	T	G
44617	C	A
44627	A	C
44804	C	T
45089	C	T
45095	A	C
45101	C	T
45224	G	A
45489	A	T
45740	A	C
46015	C	T
46433	G	A
46916	A	G
47004	C	T
47239	G	A
47325	C	T
47408	T	C
47490	C	T
47510	T	C
47568	T	G
47620	A	C
47638	G	A
47888	C	T
47907	G	T
48025	A	C
48575	C	T
48725	A	G
48954	C	T
49353	A	G
49374	A	G
49453	A	C
49545	G	T
49586	C	T
49653	T	G
49713	C	A
49725	A	G
50241	C	T
50316	C	G
50524	C	T

50534	T	C
50588	C	G
50635	A	G
50672	A	C
50691	G	A
50882	T	C
51153	C	G
51228	C	A
51270	A	G
51394	T	C
51515	A	G
51518	G	A
51589	G	T
51766	T	G
51836	C	T
51848	G	A
51905	A	G
52007	A	G
52071	C	T
52082	G	T
52139	G	T
52172	T	A
52187	G	A
52244	T	C
52314	C	T
52388	C	T
52403	T	C
52504	T	C
52550	A	G
52697	C	T
52705	C	T
52709	C	T
52740	C	T
52808	T	A
53030	G	C
53297	G	T
53771	C	A
53873	T	C
53879	G	A
53948	A	G
54095	C	T
54444	G	A
54584	C	G
54635	A	C
54801	G	A
54809	T	C
54957	T	C
55040	C	T
55297	T	C
55523	G	C

55583	G	A
55592	G	T
55610	T	C
55748	T	C
55850	T	G
56150	T	C
56266	C	T
56432	T	C
56514	G	C
56701	T	C
56854	A	C
57104	G	T
57209	C	A
57361	C	T
57538	G	C
57553	T	G
57560	C	A
57627	G	A
57697	T	C
57748	G	A
57798	T	C
57889	C	A
57891	T	G
58024	C	T
58027	T	C
58184	G	A
58191	G	A
58193	G	A
58208	G	A
58223	G	A
58599	G	A
58655	G	A
58832	G	A
59024	G	A
59105	C	T
59119	G	A
59206	T	G
59243	C	T
59253	G	A
59335	T	C
59391	T	G
59436	G	A
		TCTCCTCCTCCTCCTCCT
59509	TCTCCTCCTCCTCC	CC
59515	CCTCCT	CCTCCTTCTCCT
59518	CCTCC	CCTCCTTCTCC
59521	CC	CCTCCTTC
59522	C	CTCCTCT
59589	G	A
59819	C	T

59826	G	A
59966	T	G
60028	G	A
60049	A	G
60240	A	G
60527	T	C
60981	G	A
61168	T	C
61260	T	C
61278	T	C
61290	G	C
61296	T	G
61480	G	A
61785	C	T
61845	C	T
61905	G	A
61935	A	G
61997	A	G
62037	A	G
62142	T	C
62163	ACC	ACCGCC
62501	C	T
62541	G	A
62592	T	G
62750	G	A
62769	C	T
62892	T	C
62924	T	C
62967	C	T
63055	A	G
63390	A	G
63510	T	C
63597	C	T
63621	G	T
63689	T	G
63864	T	C
63897	A	G
64061	CAGGAGGAGGAGG	CAGGAGGAGG
64132	G	A
64191	T	G
64217	C	A
64862	T	C
64872	A	C
64883	T	C
65123	A	G
65214	T	C
65435	A	T
65507	C	A
65573	T	C
65624	C	A

66068	T	C
66221	T	C
66245	C	T
66314	C	T
66320	T	C
66571	C	G
66750	T	C
66776	G	A
66962	A	C
67055	T	C
67746	C	A
68305	T	C
68448	C	A
68453	G	A
68567	A	C
68886	C	A
69712	C	A
69771	C	A
69778	G	T
69911	A	C
69993	C	A
70001	G	A
70015	T	C
70155	T	G
70253	G	A
70257	G	A
70488	T	G
70506	A	G
70860	C	T
70875	A	G
71010	C	T
71238	G	T
71277	C	G
71436	C	T
71780	G	T
71944	G	A
72689	G	A
72936	G	A
73063	C	G
73090	A	G
73212	G	A
73265	T	C
73351	G	C
73395	C	T
73726	T	C
73814	G	T
73829	C	T
73861	C	T
73889	T	C
73924	G	T

74189	C	T
74222	T	C
74228	A	G
74364	T	C
74466	G	A
74528	G	C
74587	C	T
74854	T	C
74998	T	C
75031	C	T
75043	G	A
75430	T	C
75448	T	C
76019	A	G
76094	C	T
76499	A	G
76604	T	C
76909	T	G
77131	T	C
77211	A	G
77213	A	C
77319	C	T
77452	C	T
77606	G	T
77703	T	C
77851	T	A
77896	A	C
77920	G	T
77941	T	G
78029	C	T
78245	C	T
78307	G	C
78332	G	C
78336	G	T
78341	A	G
78383	A	G
78456	G	T
78663	C	T
79299	C	T
80062	T	C
80350	C	G
80616	A	G
80698	T	C
80920	G	A
81039	A	C
81095	A	G
81213	C	T
81239	A	G
81299	G	A
81458	A	C

81500	T	C
81517	T	C
81723	A	T
81900	C	A
82084	T	C
82214	G	A
82235	T	C
82240	A	G
82252	C	A
82424	T	C
82439	C	T
82462	C	A
82473	A	C
82483	G	T
82487	T	C
82606	G	A
82672	A	G
82772	A	G
82783	G	A
82786	C	A
82791	A	C
82815	G	A
82838	C	T
82920	TAA	TA
83063	T	A
83178	G	T
83184	T	G
83293	C	G
83319	C	T
83331	T	A
83343	C	G
84156	C	T
84258	G	A
84525	A	T
84539	T	C
84767	T	A
84800	C	T
84888	G	A
84918	C	T
85023	C	A
85129	A	G
85170	T	C
85197	G	C
85201	C	T
85228	A	G
85272	A	G
85331	C	T
85336	G	A
85337	T	C
85342	C	A

85343	G	T
85586	A	G
85837	G	A
85841	A	G
86392	A	G
86402	C	T
86424	T	C
86577	A	G
86675	A	G
86812	T	G
86850	G	T
87134	T	C
87223	C	T
87240	G	T
87266	T	C
87961	T	C
87996	G	C
88025	A	G
88170	G	A
88217	C	T
88257	G	C
88458	C	G
88495	C	T
88555	C	T
88604	T	C
88640	G	T
88880	T	C
89009	T	C
89694	T	A
90073	G	A
90108	G	A
90126	G	C
90128	A	G
90159	T	A
90294	A	G
90303	A	G
90321	A	T
90355	G	T
90380	T	A
90390	G	T
90394	G	A
90413	A	G
90431	A	G
90488	G	T
90498	T	G
90554	T	A
90574	G	T
90642	A	G
90661	A	T
90666	T	C

90685	T	G
90748	C	T
91180	G	T
91271	C	T
91366	T	C
91431	G	T
91784	G	T
91885	A	G
91950	T	C
91975	A	G
92033	G	C
92046	C	T
92071	G	A
92078	C	A
92119	G	C
92233	A	G
92239	G	A
92623	A	G
93016	C	T
93226	G	A
93263	C	T
93368	A	C
93469	G	A
93523	C	G
93742	C	T
93792	T	C
93889	A	T
93894	G	T
94165	C	A
94332	C	T
94564	T	C
94601	C	T
94733	C	T
94736	T	G
94784	A	G
94841	G	A
94866	C	A
94901	T	G
94951	A	G
94980	C	A
94987	C	T
94997	T	C
95012	C	T
95157	C	T
95182	T	C
95187	A	C
95197	A	G
95198	A	G
95202	T	C
95265	T	C

95302	T	C
95380	T	C
95401	A	C
95435	C	T
95678	C	A
95707	G	C
95714	G	A
95719	A	T
95769	T	G
95814	C	T
95843	G	A
95914	A	G
95946	G	A
96273	A	C
96357	A	C
96372	A	C
96618	A	C
96705	A	G
96751	G	A
96792	A	T
96890	G	C
96894	A	T
96914	A	T
96976	G	A
97140	C	G
97158	C	G
97166	C	A
97221	A	C
97232	C	T
97243	A	G
97246	C	T
97405	C	T
97413	G	A
97442	G	A
97698	T	C
97793	T	C
97973	G	A
98439	C	T
98796	T	G
98842	C	A
99351	G	T
99356	C	A
99851	T	C
100304	G	A
100433	T	G
100589	C	T
100640	G	A
100652	G	T
100658	C	A
100748	C	A

100841	G	A
100856	A	G
100886	A	G
100957	C	T
101159	G	A
101291	A	G
101306	G	A
101629	C	T
101692	T	A
101716	G	A
101850	T	A
101862	C	T
101921	C	T
101978	T	G
101987	G	A
102017	G	T
102068	C	T
102502	A	G
102587	T	C
102830	A	G
102989	A	G
102995	A	G
103206	G	A
103334	A	G
103640	G	A
103694	C	T
103695	A	G
103825	A	G
103885	G	A
103986	G	A
104127	T	C
104536	G	A
104550	A	C
104555	T	C
104673	G	A
104705	A	G
104781	G	A
104820	C	A
104824	T	C
104904	A	G
105036	A	G
105096	A	G
105107	T	A
105221	T	C
105332	T	C
105596	T	C
106025	A	G
106058	G	A
106273	T	C
106399	T	G

106622	G	A
106688	A	G
106715	G	C
106891	G	T
107013	T	G
107142	C	T
107223	A	G
107349	C	T
107793	G	A
108118	G	T
108192	T	G
108220	A	G
108256	G	T
108283	A	G
108333	C	A
109340	T	C
109358	G	A
109466	G	A
109917	A	T
109940	T	C
110463	C	A
110688	A	C
111243	T	C
111506	T	C
112041	T	C
112088	T	C
112100	C	T
112112	G	A
112202	T	A
112349	A	C
112439	G	C
112664	G	A
112763	A	G
112919	T	C
112970	C	T
112994	A	C
113264	A	G
113309	A	G
113558	A	C
113719	A	G
113774	G	T
114680	C	T
114719	A	C
114758	C	G
114763	G	A
114864	ATAGTAG	ATAG
115296	A	G
115313	C	A
115364	G	A
115372	G	T

115478	C	T
115691	A	C
115712	C	G
115832	A	C
116305	C	G
116808	A	G
116902	C	A
117490	C	T
117698	A	G
117892	G	A
117916	C	T
117949	T	G
118356	A	C
118373	T	C
118379	A	T
118426	G	A
118531	A	G
118562	T	G
118569	T	C
118587	A	G
118603	G	T
118635	C	T
118674	C	A
118710	A	G
118745	T	G
118746	A	G
119296	T	A
119400	G	A
119415	A	G
119623	C	G
119787	G	T
119805	T	C
120436	T	C
120607	C	T
120673	T	C
120679	G	T
120913	G	A
120958	C	T
121161	G	T
121165	G	A
121231	G	T
121426	A	G
121474	T	C
121762	A	G
121777	G	T
121957	A	G
122209	A	G
122482	A	G
122491	A	G
122611	G	T

122725	T	C
123012	G	T
123031	G	A
123228	A	C
123313	G	A
123376	G	A
123472	A	G
123555	G	A
123601	G	A
123655	A	G
123673	A	C
123877	T	G
123925	G	A
123988	G	T
124048	G	A
124108	T	G
124273	A	G
124576	G	A
124642	G	A
124758	G	A
124903	C	T
125082	A	G
125089	A	G
125212	C	A
125273	A	G
125275	A	G
125476	T	C
125574	A	G
125662	G	A
125719	T	G
125785	C	T
125851	G	A
125890	T	C
125894	C	G
125911	G	A
125935	C	T
126024	A	G
126066	C	T
126094	T	C
126098	T	G
126125	T	G
126136	G	T
126232	A	G
126271	A	G
126679	T	C
126819	G	A
126882	A	G
127031	G	A
127409	T	C
127841	G	A

127992	G	A
128037	A	G
128269	C	A
129101	A	G
129200	C	A
129234	G	A
129338	T	C
129560	G	T
129716	T	C
129740	A	G
130159	C	T
130451	G	A
131154	G	A
132049	G	A
132221	G	A
132246	C	T
132393	G	T
132399	G	A
132429	A	G
132474	G	A
132585	G	A
132606	A	G
132633	T	C
132673	C	A
132717	G	A
132756	A	G
133241	C	A
133649	A	C
133948	C	T
134139	G	T
134372	A	G
134386	T	C
134425	C	T
134719	C	G
134730	T	C
134761	G	A
134767	A	C
134789	C	T
134824	G	T
134875	A	G
134887	C	G
135061	A	G
135079	T	G
135103	G	A
135109	A	C
135118	G	A
135178	G	T
135357	C	T
135403	T	C
135517	G	A

135967	C	T
136078	T	C
136194	G	A
136252	A	C
136351	C	A
136646	G	T
136915	G	A
137143	G	A
137347	G	A
137481	C	A
137697	C	G
137799	G	C
138283	T	C
138577	T	G
138580	A	C
138699	G	T
138723	T	C
138813	T	G
138987	C	A
139004	T	G
139070	T	C
139600	A	G
139628	G	A
139654	A	C
139657	T	G
140002	T	G
140381	G	T
140464	C	T
140470	C	A
140601	T	G
144457	G	A
145493	A	G
145539	T	C
145589	C	A
145696	A	G
146049	GAAA	GAA
146116	G	A
146221	T	C
146242	A	G
146300	G	A
146317	G	T
146399	G	T
146594	C	G
147126	T	TG
147144	A	T
147146	C	A
147275	G	A
147688	CTT	CTTT
147784	T	C
147850	G	T

147883	T	C
147900	C	G
148284	T	C
148429	T	A
148432	T	C
148484	T	C
148628	T	A
149635	T	G
149794	G	A
149836	G	A
150471	A	G
150675	A	G
150754	G	A
150778	G	A
150876	G	A
151005	A	G
151078	A	G
151140	G	T
151277	C	G
151338	C	T
151353	G	A
151644	T	C
151671	C	A
151701	A	G
151822	A	G
151943	A	G
152024	A	G
152245	T	G
152612	G	A
152625	G	A
152767	G	A
152962	C	T
153045	C	T
153166	T	C
153562	C	T
153601	G	A
153937	C	T
154366	G	A
154387	T	C
154621	C	T
154906	C	T
154972	G	T
155275	C	T
155920	A	G
156013	C	T
156133	G	A
156154	C	T
156184	T	G
156696	A	C
156798	G	A

156810	T	G
156881	G	T
157125	G	A
157230	A	G
157359	T	G
157419	G	T
157500	A	G
157535	T	C
157557	G	A
157568	C	T
157578	A	G
157598	G	A
157619	A	G
157626	T	G
158144	T	G
158408	T	C
158481	G	C
158808	A	G
159178	G	A
159208	T	C
159960	C	G
159997	G	A
160087	C	T
160099	C	T
160249	G	C
160297	C	T
160621	A	G
160804	C	T
160842	T	C
161028	C	T
161119	T	C
161637	C	T
162147	G	A
162195	A	C
162464	G	A
162476	T	C
162675	T	C
162852	G	T
162963	A	G
163377	A	G
163458	T	C
163482	T	G
164277	G	T
164883	C	T
165131	T	C
165310	T	C
165339	T	C
165486	C	T
165492	T	A
165737	T	C

165870	T	G
165922	C	A
165939	G	C
165965	T	C
166169	T	G
166282	T	C
166321	C	T
166347	A	C
166508	T	A
166625	A	C
166632	G	T
166754	T	C
166797	G	C
166810	T	C
166873	A	G
166991	T	C
167051	T	C
167202	A	G
167286	A	T
167310	G	T
167378	G	T
167392	G	C
167412	G	T
167458	C	T
167466	T	G
167470	T	C
167489	T	G
167612	T	C
167620	T	A
167628	G	A
167767	A	T
167802	C	T
167808	T	C
167832	C	T
167837	T	A
167850	A	G
167851	A	G
167872	C	G
167903	A	C
167920	C	G
167964	T	G
167985	T	C
168137	T	G
168167	G	T
168229	C	T
168236	A	G
168288	C	G
168326	C	G
168401	G	A
168412	G	T

168433	A	T
168467	T	C
168476	C	A
168485	T	G
168499	T	G
168507	A	G
168544	G	A
168560	T	G
168594	C	G
168595	C	G
168611	C	G
168622	A	T
168707	A	C
168724	G	T
168764	T	G
168767	A	C
168772	G	C
168942	T	C
169120	T	G
169127	T	G
169128	A	G
169164	T	G
169170	T	C
169171	A	G
169215	G	A
169220	G	A
169240	T	C
169262	C	T
169264	G	A
169329	G	C
169341	C	A
169343	T	C
169355	A	T
169371	T	C
169385	C	T
169386	G	A
169413	G	T
169429	C	T
169495	C	A
169504	C	G
169511	T	G
169520	G	C
169543	T	C
169572	A	C
169584	T	C
169593	T	C

Anexo 4. Número de variantes identificadas en los genes del VEB asociado a ENKTL guatemalteco.

Gen	Número de Variantes
EBNA1	536
EBNA3B3C	430
EBNA3A	376
LMP2A	142
BOLF1	131
RPMS1	113
BPLF1	103
LMP2B	67
LMP-1	60
BRLF1	49
BCLF1	38
BNRF1	33
A73	33
BCRF1	32
BORF1	32
BGRF1/BDRF1	31
BSLF1	31
BALF3	31
EBNA2	28
BALF5	27
EBNA-LP	27
BDLF1	26
BRRF2	26
BFRF1A	25
BBRF1	24
BXLF1	23
BZLF1	23
BILF1	23
BBLF2/BBLF3	22
BBLF4	21
BDLF3.5	21
BALF4	20
BHRF1	20

BDLF2	20
BLLF1	20
BXRF1	20
BORF2	20
BDLF4	19
BGLF1	18
BGLF3	18
BFRF2	17
BVLF1	17
LF1	16
BARF1	15
BALF2	14
BFLF1	14
BDLF3	14
BVRF1	14
BMRF1	10
BSLF2/BMLF1	10
BVRF2	10
BILF2	9
BGLF3.5	9
BNLF2a	9
BARF0	8
BGLF4	8
BKRF2	8
BFRF3	8
BRRF1	8
BXLF2	7
BBRF3	7
BFLF2	6
BMRF2	6
BNLF2B	6
BLLF2	6
BBRF2	6
BdRF1	6
BGLF2	5
BGLF5	5

BHLF1	5
BARF1	5
BKRF3	5
BCRF1	4
BSRF1	4
BTRF1	4
LF2	3
BLLF3	3
BLRF1	3
BALF1	2
BLRF2	2
BZLF2	1
LF3	1

Anexo 5. Diversidad nucleotídica de la región OriP.

π	θ por secuencia	θ por sitio	k
0.00386	6.84860	0.01145	2.30561

π = Diversidad nucleotídica

θ = Diversidad nucleotídica

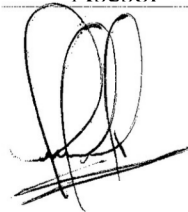
k = número de diferencias nucleotídicas



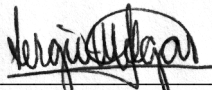
f) _____
César Camilo Carías Alvarado
Tesisista



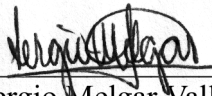
f) _____
Lic. Francisco Josué López Hun
Asesor



f) _____
Dra. Fabiola Valvert Gamboa
Asesora



f) _____
Dr. Sergio Melgar Valladares
Revisor



f) _____
Dr. Sergio Melgar Valladares
Director Escuela de Biología



f) _____
M.A. Pablo Ernesto Oliva Soto
Decano