



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO

FACULTAD DE QUÍMICA

**“IDENTIFICACIÓN DE PATÓGENOS VIRALES HUMANOS
EN MUESTRAS ARQUEOLÓGICAS CON INDICIO DE
EPIDEMIA EN MÉXICO”**

TESIS

QUE PARA OBTENER EL TÍTULO DE

INGENIERO EN BIOTECNOLOGÍA

PRESENTA

AXEL ARTURO GUZMÁN SOLÍS

DIRIGIDA POR

Dra. MARÍA DEL CARMEN ÁVILA ARCOS



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO

FACULTAD DE QUÍMICA

**"IDENTIFICACIÓN DE PATÓGENOS VIRALES HUMANOS
EN MUESTRAS ARQUEOLÓGICAS CON INDICIO DE
EPIDEMIA EN MÉXICO"**

TESIS

QUE PARA OBTENER EL TÍTULO DE

INGENIERO EN BIOTECNOLOGÍA

PRESENTA

AXEL ARTURO GUZMÁN SOLÍS

DIRIGIDA POR

Dra. MARÍA DEL CARMEN ÁVILA ARCOS

SINODALES

Dra. MARÍA DEL CARMEN ÁVILA ARCOS *Maria Carmen Avila Arcos*
DIRECTOR

Dr. DANIEL BLANCO MELO
CO-DIRECTOR

Dra. MINERVA RAMOS GÓMEZ
ASESOR INTERNO

Dra. ANGELINA RODRÍGUEZ TORRES
SINODAL

[Handwritten signatures on lines]

SANTIAGO DE QUERÉTARO, QUERÉTARO, 2019.

AGRADECIMIENTOS

El siguiente trabajo de investigación ha sido posible gracias a mis maestros, amigos y familia. A ellos, y no a otros, se debe la finalización de este proceso. Por lo que quiero agradecer y aclarar que todo lo asertivo en este trabajo es culpa de ellos y cualquier error es responsabilidad mía.

Me gustaría agradecer de manera general a todo el equipo del Laboratorio Internacional de Investigación del genoma Humano (LIIGH) y al Laboratorio Nacional de Visualización Científica Avanzada de la Universidad Autónoma de México, donde se realizó este trabajo. Hago énfasis en la dificultad de obtener las instalaciones altamente especializadas para realizar esta investigación, por lo que expreso mi gratitud al Dr. Rafael Palacios y al Dr. Guillermo Dávila por ser los pioneros del proyecto LIIGH.

A la Dra. Lucía Morales quien me abrió las puertas para formar parte de esta prestigiosa institución y por fungir como mentora múltiples veces.

A la Dra. María Ávila y al Dr. Daniel Blanco que fueron vitales para la realización de este trabajo, dirigiendo inteligentemente este proyecto. Además, agradezco su apoyo en mi desarrollo profesional y personal, pues su soporte fue primordial para lograr metas que de otra forma me hubieran sido inalcanzables.

Al equipo de paleogenómica del LIIGH, Viridiana Villa, Miriam Bravo, Camila Tamburrini, Shreya Ramachandran, quienes formaron de manera activa parte del equipo que desarrolló este trabajo.

A mis amigos del programa de biotecnología por su apoyo durante estos años.

A la Universidad Autónoma de Querétaro (UAQ) y a su Facultad de Química por permitirme adquirir el conocimiento y las habilidades para desarrollarme en el área de las ciencias biológicas.

A mis profesores de la UAQ que transmitieron su conocimiento y me dieron las bases para poder entender y desarrollar este proyecto, en especial a la Dra. Rosalía Reynoso, Dra. Laura Berúmen, M. Susana Flores, Dra. Karina Acevedo, Dr. Sergio Romero y al Dr. David García.

A mis sinodales, la Dra. Minerva Ramos y la Dra. Angelina Rodríguez que fueron excelentes profesores y, además, aportaron valiosos comentarios para enriquecer este trabajo.

Finalmente, quiero agradecer profundamente a mis padres y hermana por su apoyo incondicional, ya que sin su ayuda y soporte no hubiera podido concluir este trabajo. A ellos dedico esta tesis con todo mi cariño y admiración.

ÍNDICE GENERAL

Contenido	Página
ÍNDICE GENERAL	i
ÍNDICE DE CUADROS	vi
ÍNDICE DE FIGURAS	viii
RESUMEN	
1.ANTECEDENTES	1
1.1 Intercambio genético post-contacto europeo.	1
1.2 Afrodescendientes durante periodo colonial.	2
1.3 Epidemias del siglo XVI y su impacto en poblaciones indígenas.	3
1.4 Arqueología de una epidemia.	7
1.5 Contexto arqueológico de las muestras.	7
1.5.1 Capilla “La Concepción” (COY).	8
1.5.2 Hospital Real de San José de los Naturales (HSJN).	8
1.6 DNA antiguo (aDNA) y los primeros genomas de patógenos reconstruidos.	11
1.7 Características de aDNA y formas de autenticación.	13

1.8 Prácticas y recomendaciones en investigación con aDNA.	15
1.9 Huesos y dientes para extracción de aDNA.	16
1.10 Procedimientos para extracción y amplificación de aDNA.	18
1.11 Técnicas de secuenciación.	18
1.12 Secuenciación de aDNA para detección de virus.	21
1.13 Virus recuperables a partir de sustrato arqueológico óseo.	23
1.13.1 Parvovirus humano (B19V) y su amplio tropismo.	23
1.13.2 HBV y genotipos en afrodescendientes.	27
1.14 Análisis metagenómico en aDNA.	30
1.15 Enriquecimiento de aDNA.	32
1.16 Herramientas para análisis de datos paleogenómicos.	33
1.16.1 BLAST.	33
1.16.2 MALT.	34
1.16.3 MEGAN.	35
1.16.4 BWA.	36
1.16.5 SAMTools.	37
1.16.6 MapDamage:	37

1.16.7 RAxML.	38
1.16.8 BEAST.	38
2.HIPÓTESIS	40
3.OBJETIVOS	41
3.1 General	41
3.2 Específicos	41
4.METODOLOGÍA	42
4.1 Materiales	42
4.2 Métodos	43
4.2.1 Obtención de muestras.	43
4.2.2 Procesamiento muestras.	43
4.2.3 Extracción de DNA.	43
4.2.4 Cuantificación de DNA por fluorometría.	45
4.2.5 Diseño y obtención de cebadores de PCR para genes de VARV.	45
4.2.6 Síntesis de oligonucleótidos.	46
4.2.7 PCR para genes VARV.	47
4.2.8 Construcción de librería para secuenciación.	48

4.2.8.1 Generación de extremos romos.	48
4.2.8.2 Ligación de adaptadores a librería.	49
4.2.8.3 Reparación de librerías.	49
4.2.8.4 Cuantificación de librerías con qPCR.	50
4.2.8.5 Indexar librerías de DNA.	51
4.2.9 Secuenciación.	52
4.2.10 Identificación de patógenos virales humanos.	52
4.2.11 Diseño de sondas de captura.	54
4.2.12 Enriquecimiento de librerías de secuenciación.	57
4.2.12.1 Preparación de perlas magnéticas.	58
4.2.12.2 Captura y re-suspensión de sondas asociadas a perlas magnéticas.	59
4.2.13 Reconstrucción de genomas virales antiguos.	62
4.2.14 Análisis filogenéticos de B19V.	64
4.2.15 Análisis filogenéticos de HBV.	66
5. RESULTADOS	69
5.1 Obtención de muestras.	69

5.2 Extracción de DNA de dientes.	71
5.3 Amplificación de genes de VARV mediante PCR.	72
5.4 Secuenciación de librerías crudas.	73
5.5 Análisis metagenómico viral.	76
5.5.1 Comparación de análisis metagenómico viral después de ensayo de captura.	77
5.6 Reconstrucción de genomas antiguos virales.	79
5.7 Determinación de patrones de daño de lecturas mapeadas a virus.	85
5.8 Análisis filogenético de B19V.	87
5.9 Análisis filogenético de HBV.	91
6. DISCUSIÓN	94
6.1 Presencia de VARV.	94
6.2 Análisis metagenómico de patógenos virales.	96
6.3 Reconstrucción de genomas virales.	99
6.4 Genoma antiguo de B19V Parvovirus.	101
6.5 Primera evidencia molecular directa de un virus traído de África a América (México) en el periodo colonial.	103
6.5.1 Lesiones osteológicas <i>post-mortem</i> de individuo HSJNC81.	107

6.6 Presencia de secuencias tipo HBV en colección HSJN.	108
6.6.1 Relatos de hepatitis e ictericia en HSJN durante cocoliztli de 1576.	111
8. REFERENCIAS	116
ANEXOS	129

ÍNDICE DE CUADROS

Cuadr		Página
o		
1	Diversidad de población en México durante época colonial.	3
2	Lista de epidemias con sus nombres modernos e indígenas en la Nueva España durante el siglo XVI.	6
3	Genomas antiguos de HBV.	21
4	Genomas antiguos de VARV y B19V.	22
5	Lista de cebadores para detección de VARV mediante PCR tiempo final.	45
6	Lista de secuencias de oligonucleótidos que se utilizaron como controles positivos para detección de VARV mediante PCR.	46
7	Condiciones de amplificación de aDNA mediante PCR.	47
8	Condiciones de amplificación de librerías de secuenciación	50

mediante qPCR.

9	Condiciones de amplificación de librerías de secuenciación mediante PCR tiempo final.	51
10	Lista de genes seleccionados para el diseño de sondas de captura.	53
11	Lista de especies virales incluidas en diseño de captura.	55
12	Lista de especies virales incluidas en diseño de captura (continuación).	56
13	Condiciones de amplificación de librerías de secuenciación mediante qPCR.	59
14	Número de ciclos usados para amplificación de librerías capturadas.	60
15	Lista de genomas de B19V incluidos en mapeo.	62
16	Genomas antiguos de B19V incluidos en análisis filogenéticos.	65
17	Genomas antiguos de HBV incluidos en análisis filogenéticos.	66
18	Concentración de DNA a partir de muestras arqueológicas.	71
19	Resumen de mapeo a B19V por ronda de captura a partir de COYC4.	80
20	Resumen de mapeo a B19V por ronda de captura a partir de	81

	HSJNC81.	
21	Resumen de mapeo a B19V por ronda de captura a partir de HSJN240.	82
22	Resumen de mapeo a HBV a partir de HSJN194 y HSJN275.	81
23	Coberturas de genomas reconstruidos.	100

ÍNDICE DE FIGURAS

Figura		Página
1	Colapso demográfico en México a causa de agentes infecciosos.	1
2	Representaciones de afrodescendientes con españoles.	4
3	Representaciones de cocoliztli.	6
4	Modificaciones dentales típicas de afrodescendientes.	9
5	Entierros de sitio HSJN.	10
6	Patrones de daño en DNA post-mortem.	14
7	Momias representativas de paleogenómica viral.	17
8	Estructura del genoma de B19V.	24
9	Síntomas asociados a infección con B19V.	25
10	Estructura de cápside de HBV.	28
11	Metodología para identificación de patógenos virales humanos.	52
12	Esquema para reconstrucción de genomas virales antiguos.	61
13	Cráneo 81 de colección HSJN.	68
14	Cráneo de individuo 240 de colección HSJN.	69

15	Cráneo de individuo 275 de colección HSJN.	70
16	Cráneo de individuo 194 de colección HSJN.	70
17	PCR con diana a genes de VARV usando DNA de sitio COY.	73
18	PCR con diana a genes de VARV usando DNA de sitio HSJN.	74
19	Número de lecturas obtenidas por secuenciación a baja profundidad.	75
20	Mapa de calor mostrando comparación de abundancias de géneros virales en muestras del sitio COY.	76
21	Mapa de calor mostrando comparación de abundancias de géneros virales en muestras del sitio HSJN.	76
22	Comparación de abundancias normalizadas de genero Erythroparvovirus antes y después de dos rondas de captura.	77
23	Comparación de abundancias normalizadas de genero Orthohepadnavirus antes y después de captura.	78
24	Lecturas usadas para reconstrucción de genomas de B19V de sitio HSJN.	79
25	Lecturas usadas para reconstrucción de genoma B19V de sitio COY.	79

26	Patrones irregulares de cobertura de B19V a partir de aDNA.	83
27	Lecturas usadas para reconstrucción de genoma de HBV de sitio HSJN.	84
28	Patrones de daño típicos de aDNA respecto a genoma de B19V.	85
29	Fenograma basado en matriz de distancias a partir de región codificante de B19V.	86
30	Árbol de máxima verosimilitud basado en región codificante de B19V.	88
31	Árbol de máxima credibilidad de clados usando reloj molecular estricto de B19V.	89
32	Fenograma basado en matriz de distancias a partir de genoma completo de HBV.	90
33	Arbol de máxima verosimilitud a partir de genoma completo de HBV.	91
34	Patrones de daño de HBV de Eurasia.	102
35	Análisis de componentes principales de HSJN (B19V) respecto a panel de referencia 1000 genomas.	105
36	Análisis de componentes principales de componente humano COYC4 respecto a panel de referencia 1000 genomas.	106

RESUMEN

La introducción de patógenos virales durante la colonización de las Américas causó un colapso en la población nativa que sufrió epidemias de patógenos nunca antes vistos. Para entender mejor su impacto tras la colonización, se extrajo y secuenció DNA antiguo (aDNA) de restos esqueléticos pertenecientes a dos sitios arqueológicos de la Ciudad de México (CMX) asociados a epidemias. Se realizó análisis metagenómico para identificar virus de DNA y se diseñó ensayo de captura por hibridación en solución para incrementar la abundancia de DNA de virus con importancia clínica, mismo con el que se logró a partir de 4 individuos distintos (siglo XVI-XVIII) reconstruir 3 genomas de parvovirus humano (B19V) y uno de hepatitis b virus (HBV). Las lecturas utilizadas presentaron patrones típicos de daño (desaminación de citosinas), lo que indica un origen antiguo. Adicionalmente, un individuo con B19V presentó indicadores óseos asociados a anemia, como hiperostosis en bóveda craneal y criba orbitalia, lo cual podría haber sido causado por B19V que infecta precursores de la línea eritroide y ha sido recuperado de individuos con anemia severa. Los 4 genomas recuperados de B19V y HBV son similares a genomas modernos africanos, coherente con la ancestría genética africana inferida de tres de los hospederos. Estos genomas son la primera evidencia directa de dos virus antiguos traídos a América desde África, posiblemente durante el periodo de comercio con personas esclavizadas. Los virus identificados en este trabajo representan los primeros genomas virales antiguos reconstruidos en América, el primer genoma antiguo de B19V genotipo 3 reconstruido, y el primer B19V genotipo 3 encontrado en México hasta ahora.

1. ANTECEDENTES

1.1 Intercambio genético post-contacto europeo.

Desde el siglo XV hasta el XVII hubo exploraciones europeas para la creación de colonias en América. España fue uno de los países más presentes, llegando a colonizar hasta cerca de la mitad del continente (Lindo y col., 2016).

Durante el contacto temprano con europeos en Mesoamérica hubo un intercambio notorio de enfermedades infecciosas, al igual que de especies comerciales de plantas y animales. En el caso de los patógenos, existe un debate sobre los agentes causantes de las primeras enfermedades introducidas al nuevo mundo (Crosby, 1976; Acuña-Soto, 2004).

La introducción de estas enfermedades en poblaciones indígenas nunca antes expuestas causó una alta mortalidad y un colapso demográfico (Figura 1), con estimaciones tan altas como una disminución del 95% de la población indígena (Warinner y col., 2012; Vagene y col., 2018). Aunque esta disminución drástica también se debió al exterminio, explotación y desplazamiento de indígenas por parte de españoles (Mandujano y col., 2003).

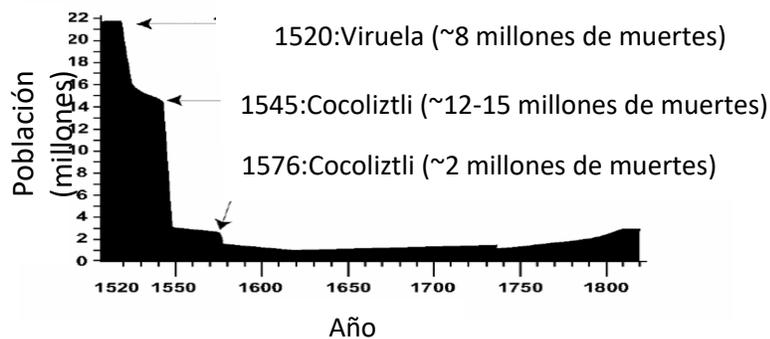


Figura 1. Colapso demográfico en México a causa de agentes infecciosos. Imagen traducida a partir de Acuña y col., 2004. Para explicar el colapso poblacional, se ha propuesto que las poblaciones nativas nunca expuestas a estos antígenos eran más susceptibles (Crosby, 1976).

En el caso de las poblaciones indígenas mexicanas, esta susceptibilidad se vio aumentada por la desnutrición y las pésimas condiciones de vida impuestas sobre este grupo recién conquistado (Meza, 2013). Las infecciones por patógenos traídos del viejo mundo facilitó la conquista española. De forma que, el control español sobre las colonias se expandía después de una disminución catastrófica en la población; sin embargo, hay pocas lesiones *post-mortem* asociadas a estas infecciones, la mayor parte de lo que se sabe es gracias a reportes históricos (Warinner y col., 2012; Vågene y col., 2018).

1.2 Afrodescendientes durante periodo colonial

Después de los indígenas americanos y mestizos, los africanos eran el grupo más numeroso (Cuadro 1) durante el siglo XVI y XVII (Velázquez e Iturralde, 2012), ya que fueron traídos como trabajadores forzados para la construcción de las colonias (Mandujano, 2003; Price y col., 2006).

El transporte de africanos esclavizados se realizó mediante un “comercio triangular” donde los barcos zarpaban de Europa hacia África con mercancías para intercambiar por personas esclavizadas que serían vendidas en América, donde comprarían nuevas mercancías para regresar a Europa (Velázquez y Iturralde, 2012).

El espacio destinado para estas personas en un barco fue limitado (cerca de 183 cm por 40 cm para hombres), y las condiciones eran poco salubres; facilitando el contacto, la transmisión y el transporte de patógenos causantes de epidemias. El

comercio forzado de estos individuos se caracterizó por alta mortandad entre “negros” (>30%) y el personal “blanco”. A pesar que se tomaban medidas de cuarentena, se dio la transmisión de patógenos con periodos de latencia mayores, y se ha descrito disentería, helmintiasis, neumonías, viruela, sarampión y enfermedades venéreas como enfermedades recurrentes durante el periodo colonial (Mandujano, 2003; Velázquez y Iturralde, 2012).

Cuadro 1. Diversidad de población en México durante época colonial (Aguirre Beltrán, 1972).

	Africanos (%)	Europeos (%)	Indígenas (%)	Mestizos (%)
1570	20569 (0.6)	6644 (0.2)	3366860 (98.7)	15939 (0.44)
1646	35089 (2)	13780 (0.8)	1269607 (74.6)	394139 (22.6)
1742	20131 (0.8)	9814 (0.4)	1540256 (62.2)	907076 (36.6)
1793	6100 (0.1)	7904 (0.2)	2319741 (61)	1465816 (38.6)
1810	10000 (0.1)	15000 (0.2)	3676281 (60)	2421073 (39.5)

1.3 Epidemias del siglo XVI y su impacto en poblaciones indígenas.

El traslado forzado de africanos a América tuvo gran importancia durante la colonización en México (Figura 2). Su mano de obra fue fundamental para construir la colonia. También se ha sugerido que con ellos llegaron varios patógenos que causaron epidemias, sobre todo mermando a la población nativa. Algunas de las

enfermedades que se han asociado a introducción desde África son la disentería, helmintiasis, anquilostomiasis, viruela, tétanos, y fiebre nerviosa (quizá tifoidea) (Mandujano y col., 2003). Ejemplo de esto es el relato de Fray Gerónimo de Mendieta que documenta el primer contagio de viruela, supuestamente transmitida por un africano (llamado Francisco de Eguía), quien en 1520 llegó en los navíos de Pánfilo de Narváez a las costas de Veracruz, matando a “menos de la mitad de la gente”, entre ellos Cuitláhuac, hermano del emperador azteca Moctezuma. Los sobrevivientes llamaron a esta epidemia “Hueyzahuatl” (gran pestilencia en náhuatl) (Mandujano y col., 2003).



Figura 2. Representaciones de afrodescendientes con españoles. A) Códice Azcatitlán. Representación de grupo de españoles encabezado por Malintzin seguida de Hernán Cortés (izquierda). B) Códice Durán. Hernán Cortés acompañado por personaje de tez oscura, recibidos en Tlaxcala.

La segunda gran epidemia después del contacto europeo fue en 1531 y también fue traída por españoles que contagiaron a nativos con sarampión (Wesp, 2017). Esta epidemia fue menos mortal para los indígenas y se llamó “Záhuatl tepiton” (peste pequeña). Entre la segunda y tercera gran epidemia, hubo una epidemia en 1538 de viruela y sarampión (Mandujano y col., 2003).

La tercera gran epidemia ocurrió durante 1545-1550 (Wesp, 2017) y es una de las epidemias más devastadoras en la historia del nuevo mundo, calculándose el fallecimiento de millones de personas (Figura 1), en su mayoría indígenas (Mandujano y col., 2003; Acuña y col., 2004). Los españoles lo llamaban “Pujamiento de sangre”, mientras los aztecas en el centro del actual México lo llamaban “Huey cocoliztli” (gran pestilencia en náhuatl). El códice Telleriano-Remensis (Figura 3e) representa la gran mortandad de esta epidemia con cuerpos apilados uno sobre otro y el códice en Cruz (Figura 3A) muestra su sintomatología representativa (sangrado de nariz severo y salpullido) (Warinner y col., 2012; Vagene y col., 2018). El agente causal de la epidemia de 1545 es desconocido, aunque se ha sugerido que esta epidemia fue causada por tifus, neumonía o por un agente viral causante de fiebre hemorrágica (Warinner y col., 2012). Extrañamente, el tifus era muy conocido por los españoles e incluso es recurrentemente abordado en el primer libro de medicina en México (*Opera medicinalia* de Francisco Bravo de 1570), al igual que mencionado en códices indígenas donde lo representaban con manchas parduscas. Esta epidemia es encontrada en varios documentos, literatura y arte de la época (Mandujano y col., 2003).

En 1576-1581 ocurrió otra epidemia también nombrada Cocoliztli (Wesp, 2017), se cree que fue causada por el mismo patógeno, en base a relatos de sobrevivientes españoles (Vagene y col., 2018); sin embargo, existen dudas debido a que también se ha descrito esta enfermedad sin erupciones cutáneas, sólo con abundantes hemorragias nasales y abscesos retroauriculares. Si bien no se conoce el agente

causal debido a la diversidad de síntomas, quizá se trató de una infección causada por varios patógenos. La población más afectada fue la indígena, seguida de los africanos y luego los españoles (Mandujano y col., 2003).

La última epidemia del siglo XVI fue a finales de 1595 y fue causada por sarampión, paperas y tabardillo (tifus exantemático) (Wesp, 2017). Esta tuvo mucho menor impacto que las llamadas cocoliztlis. Asimismo, se reportan otras enfermedades epidémicas menores como rubeola, tos ferina y neumonía (Mandujano y col., 2003) (Cuadro 2).

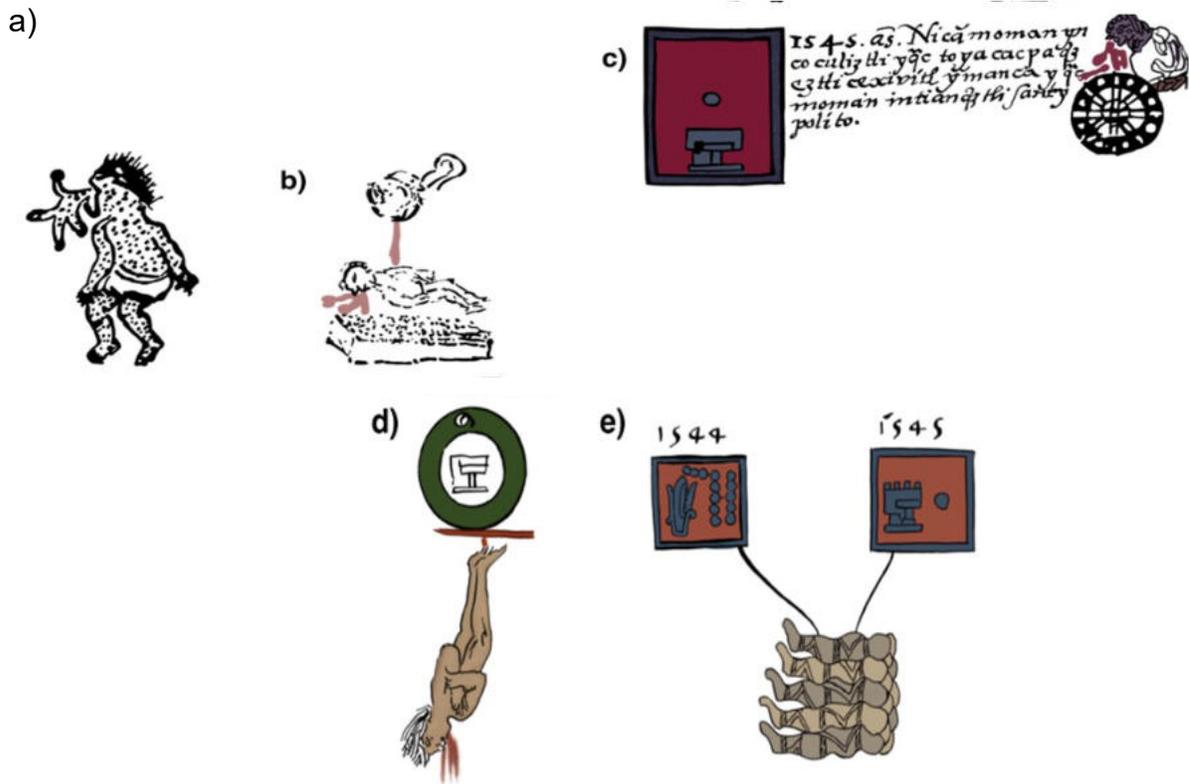


Figura 3. Representaciones de Cocoliztli. Dibujos modernos tomados de Warinner y col., 2012, basados en a) Códice en Cruz; b) Códice Mexicano; c) Códice Aubin; d) Tira de Tepechpan y e) Códice Telleriano-Remendis.

Cuadro 2. Lista de epidemias con sus nombres modernos e indígenas en la Nueva España durante el siglo XVI (Mandujano y col., 2003; Wesp, 2017).

Fecha citada	Enfermedad	Nombre indígena
1520	Viruela	Hueyzáhuatl
1531	Sarampión	Záhuatl tepiton
1538	Viruela/sarampión	Hueyzáhuatl/ Záhuatl
1545	Desconocido	Cocoliztli
1576	Mismo patógeno desconocido	Cocoliztli

1595	Sarampión, paperas y tabardillo (tifus exantemático)	Sin nombre indígena
------	---	---------------------

1.4 Arqueología de una epidemia.

Si bien existen pocas lesiones osteológicas sobre enfermedades infecciosas *post-mortem* en hueso, hay evidencias por el contexto arqueológico que muestran eventos catastróficos en cementerios indígenas (Ortner, 2003). Un claro ejemplo son los entierros masivos de múltiples individuos en una sola fosa, así como su posicionamiento vertical y sucesivo en periodos de tiempo cortos, fuese en una o varias capas (Meza, 2003).

Se han usado restos óseos y reportes históricos con el fin entender la historia de epidemias en poblaciones específicas; sin embargo, muchas de las enfermedades infecciosas no dejan marcas osteológicas debido a periodos de infectividad cortos o por muerte en la fase aguda de la enfermedad antes de dejar marcas evidentes en los huesos (Ortner, 2003). Por otro lado, los reportes históricos existentes se prestan a ser ambiguos en su interpretación, pues hay variación cultural o diferencias en su traducción. Incluso omitiendo estas variaciones, un mismo fenómeno podría tener múltiples causalidades (Warinner y col., 2012).

1.5 Contexto arqueológico de las muestras.

Para este estudio se utilizaron muestras provenientes de entierros con indicios típicos de una epidemia. Uno de estos indicios es el apilamiento indiscriminado de cuerpos en espacios limitados, que sugiere una necesidad por deshacerse rápidamente de los cadáveres, como en el caso de una epidemia (los entierros cristianos tradicionales tienen disposición separada y ordenada) (Wesp, 2017). Otro

signo importante es la forma de descomposición, pues si la degradación del tejido es *in situ*, ciertos huesos permanecen en la misma posición que en el momento del entierro. Por otro lado, exhumaciones y/o inhumaciones posteriores disgregan estos huesos fácilmente por lo que se puede saber si los individuos fueron depositados en un mismo momento cronológico (Meza, 2013). A continuación, se mencionan características específicas para cada sitio involucrado en este estudio.

1.5.1 Capilla “La Concepción” (COY)

La capilla de “La purísima Concepción” se construyó bajo órdenes de Hernán Cortés en el siglo XVII, sobre asentamientos prehispánicos. Es una de las iglesias más antiguas de CMX y sirvió como centro religioso católico durante el periodo colonial, contando con un cementerio del que se obtuvieron restos arqueológicos óseos humanos para este estudio.

Los restos arqueológicos de La Concepción fueron rescatados durante las obras de construcción sobre este sitio en la delegación Coyoacán dentro del barrio de la Sagrada Concepción; mismos que se obtuvieron del osario de la Escuela Nacional de Antropología e Historia (ENAH), y fueron excavados por la arqueóloga María de la Luz Moreno Cabrera. Si bien, el reporte arqueológico aún no se ha publicado, se sabe gracias a los reportes proporcionados por los arqueólogos y antropólogos físicos que estas muestras presentaban características similares a entierros durante periodos de epidemia (apilados de forma vertical, sin cuidado minucioso, uno sobre otro, entre otras). No existe marca ósea específica que defina infección de viruela u otro patógeno; sin embargo, los documentos históricos de la parroquia de la Concepción sugieren epidemia por este virus.

1.5.2 Hospital Real de San José de los Naturales (HSJN)

El segundo grupo de muestras empleadas en este estudio fue otorgado por el Dr. Jorge A. Gómez Valdés, encargado del laboratorio de osteología de posgrado. El proyecto cuenta con la aprobación por el Consejo de Arqueología del INAH (Oficio 401.1S.3-2018/1373). Los individuos de dicha colección fueron encontrados en 1992 con motivo de la construcción de la línea 8 del metro. El sitio donde se encontraron los restos óseos es la calle López esquina con Artículo 123 (Antiguo barrio de Moyotlán). Estas muestras pertenecieron al antiguo Hospital Real de San José de los Naturales (HSJN), que fue el hospital de indios más importante de América (INAH, 2016). Dicho hospital estuvo en funcionamiento desde 1531, año en que se trató epidemia de sarampión y posteriormente innumerables plagas que afectaron la población nativa hasta su clausura en 1822 por falta de fondos (Meza, 2013).

En dicho hospital se estudiaron las primeras autopsias hechas en América por el protomédico Francisco Hernandez quien describe sintomatología de cocoliztli de 1576 (Malvido y Viesca, 2018).

En este sitio arqueológico se encontraron 600 individuos inhumados simultáneamente en fosas comunes (Figura 5), probablemente debido a epidemias del siglo XVI, destacando la presencia de 20 individuos con rasgos morfológicos típicos de afrodescendientes como apertura nasal ancha, prognatismo facial, paladar alargado/elipsoidal y mutilaciones dentales (Figura 4), mientras el resto presentan carácter nativo americano o inclusive peninsular (poco frecuente). Hay que recordar que los afrodescendientes fueron traídos contra su voluntad para trabajos forzados en un transporte con precarias condiciones de higiene, donde eran presa fácil de enfermedades infecciosas como viruela, sarampión, tifo, malaria, sífilis, entre otras (Meza, 2013).



Figura 4. Modificaciones dentales típicas de afrodescendientes. De izquierda a derecha, HSJN150; HSJN214; HSJN296. Fotos amablemente compartidas por la Dra. Julie K. Wesp.

Si bien, el hospital estaba enfocado a los nativos, se piensa que, en momentos de crisis causados por epidemias graves, las instituciones no se negaban a dar atención médica, sin importar el sector o casta al que pertenecieran (Meza, 2013).



c)

a)



b)



Figura 5. Entierros de sitio HSJN. a) Entierro tradicional de usanza católica; b, c) Entierro en fosa común. Fotos tomadas de Dirección de Salvamento Arqueológico (INAH)

A pesar de esta segregación racial institucional, los rituales funerarios eran dictados por la Iglesia católica para indígenas y “negros”. Las personas mejor acomodadas podían ser enterrados en su propio ataúd, mientras el resto, alquilaba o recibía féretros prestados. No es de extrañar que, en caso de una epidemia, los cadáveres fuesen colocados directamente sobre la tierra (Figura 5); como en el caso del Hospital San José de los Naturales, donde lo importante era sepultar a los cientos de cuerpos recién fallecidos (Meza, 2013).

Los cuerpos encontrados en este sitio fueron enterrados según la usanza católica, es decir, colocados boca arriba, extendidos, con manos cruzadas sobre pecho o

cadera, y pies cruzados (Figura 5a). Así mismo, se sabe que los fallecidos fueron puestos directamente sobre la tierra, pues al depositar un individuo en el ataúd se provoca una separación entre los huesos de la cadera y los fémures, creando un espacio vacío, mientras que, al colocar directo sobre el suelo, la tierra toma el lugar de las masas musculares, evitando la dispersión de los huesos. A su vez, hay evidencia de que los individuos en este sitio arqueológico fueron colocados en un sólo momento cronológico, pues los huesos de las manos y de los pies se dispersan fácilmente de su posición original al momento de inhumaciones/exhumaciones posteriores; mientras que, en este sitio arqueológico dichos huesos estaban en su posición original (Meza, 2013).

1.6 DNA antiguo (aDNA) y los primeros genomas de patógenos reconstruidos.

El uso de restos arqueológicos o fósiles provee información valiosa sobre la historia evolutiva en la tierra, la dinámica de poblaciones modernas o antiguas, así como sobre la emergencia de enfermedades infecciosas (Drummond y col., 2012). Rutinariamente, se han usado restos fósiles para la asignación de nodos internos en árboles filogenéticos, basándose únicamente en similitudes morfológicas (Tringe y Rubin, 2005). Recientemente, el desarrollo de tecnologías de secuenciación, su disminución de precio, el uso de metagenómica y los métodos de enriquecimiento de librerías de secuenciación, han permitido obtener millones de lecturas de DNA de restos arqueológicos y fósiles (Warinner y col., 2017).

Históricamente, la identificación de DNA en muestras antiguas se realizaba por reacción en cadena de polimerasa (PCR), que normalmente requiere templados >100 pares de bases (bp), lo cual es raro en la mayoría de fragmentos de DNA antiguo (aDNA) y, además, debido a su baja abundancia se requería un número excesivo de ciclos (>35) para amplificar. Lo anterior hace a este enfoque vulnerable a amplificación de contaminaciones ambientales, sin mencionar errores de

amplificación causados por diferencia de secuencia entre templado y cebador. Debido a estos factores, la replicación de estudios de aDNA era poco rutinaria; sin embargo, el desarrollo de la secuenciación masiva permitió la caracterización de moléculas cortas sin necesidad de tener un blanco específico, con costos cada día más accesibles (Key y col., 2017; Warinner y col., 2017). Estos avances trajeron a su vez nuevos retos para evitar interpretaciones erróneas (Tringe y Rubin, 2005).

En 2011 se logró la primera reconstrucción del genoma completo de una bacteria antigua, *Yersinia pestis*, el agente causal de la “Muerte Negra” en Europa (Bos y col., 2011). Posteriormente, se obtuvieron genomas completos de los patógenos causantes de otras enfermedades como *Mycobacterium leprae* (Schuenemann y col., 2013), *Mycobacterium tuberculosis* (Bos y col., 2014), *Tannerella forsythia* (Warinner y col., 2014), *Brucella melitensis* (Kay y col., 2014), *Helicobacter pylori* (Maixner y col., 2016) y *Treponema pallidum* (Schuenemann y col., 2018). Recientemente, se han obtenido genomas completos virales como el del virus causante de la viruela (VARV) (Duggan y col., 2016; Pajer y col., 2017), virus de hepatitis B (HBV) (Kahila y col., 2012; Krause-Kyora y col., 2018; Mühlemann y col., 2018; Ross y col., 2018) y parvovirus humano B19 (B19V) (Mühlemann y col., 2018). Estudios como los anteriores permiten la reconstrucción de rutas evolutivas de especies, sin limitarse a extremos modernos de un árbol filogenético, la calibración objetiva de árboles filogenéticos en base a DNA para la re-estimación de fecha de divergencia del ancestro común más reciente (MRCA, por sus siglas en inglés), modificando notoriamente el entendimiento de la historia evolutiva de patógenos (como VARV, HBV, B19V). Además de su aportación evolutiva, el análisis de genomas antiguos ha permitido dilucidar la adquisición o pérdida de regiones estructurales y funcionales importantes para mecanismos de patogenicidad, al comparar con genomas modernos de la misma especie (Warinner y col., 2014).

1.7 Características de aDNA y formas de autenticación.

Uno de los retos generados a partir de la secuenciación masiva de muestras antiguas es probar la autenticidad del material genético encontrado y descartar que éste provenga de contaminación con DNA moderno. Posterior a la muerte de un organismo, el daño molecular en el DNA se acumula de forma recurrente. Los decaimientos más comunes son la depuración de bases, cortes en DNA y la desaminación de citosinas; estos eventos limitan la cantidad de DNA a analizar, pero dan patrones evaluables para su autenticación. La ruptura hidrolítica de grupos amino ocurre en adeninas, citosinas, 5-metil-citosinas y guaninas, dando como resultado hipoxantina, uracilo, timina y cantina, respectivamente (Pääbo y col., 2004). El caso mejor estudiado es la desaminación de citosinas en uracilo (C>U), que se produce en los extremos de los fragmentos de aDNA (Briggs y col., 2007), generando así una lectura de timina (análogo de uracilo) al ser secuenciado. Este fenómeno es casi simétrico en la cadena complementaria (guanina transformada en adenina, G>A) (Figura 6). Estas modificaciones correlacionan con la edad de la muestra (aunque no de forma sistemática) y se han encontrado en todas las muestras antiguas analizadas a la fecha, sin importar el organismo o la parte muestreada; inclusive, se han detectado dichos patrones incluso en muestras con alto grado de conservación como las obtenidas en el permafrost o hielos perpetuos (suelo permanentemente congelado en regiones peri-glaciares) (Warinner y col., 2017).

Tras la muerte de un organismo, los componentes celulares se rompen liberando enzimas catalíticas, como nucleasas lisosomales que escinden cadenas de DNA. También existe fragmentación no enzimática de enlaces fosfodiéster y de esqueleto fosfato-azúcar (Pääbo y col., 2004). El aDNA es fragmentado en un proceso donde primero se depurina el esqueleto de DNA para después ser hidrolizado; esto es notorio al analizar lecturas por secuenciación masiva, ya que hay cortes

enriquecidos en purinas (A y G) (Warinner y col., 2017). Además, los enlaces glucosídicos entre bases nitrogenadas y esqueleto de azúcar pueden ser igualmente hidrolizados, resultando en un sitio abásico que promueve la ruptura de las cadenas (Figura 6) (Pääbo y col., 2004). Debido a esta fragmentación, el aDNA usualmente presenta tamaños de fragmentos de 40-80 bp y, en algunos grupos de datos, se presenta periodicidad de 10 bp en la distribución de tamaño. Este patrón se ha explicado mediante la posible protección del DNA alrededor de los nucleosomas, ya que se han observado regiones de mayor cobertura que coinciden sobre un mapa de nucleosomas. En cada vuelta de 10 bp, el DNA envuelve un nucleosoma donde sólo 1 nt es expuesto a hidrólisis, este fenómeno podría explicar la periodicidad de tamaño en algunos fragmentos de aDNA (Pedersen y col., 2014). En el caso de aDNA de patógenos, se espera una cobertura similar de todo el genoma ya que su genoma no se compacta con nucleosomas (al menos en bacterias y virus) (Key y col., 2017).

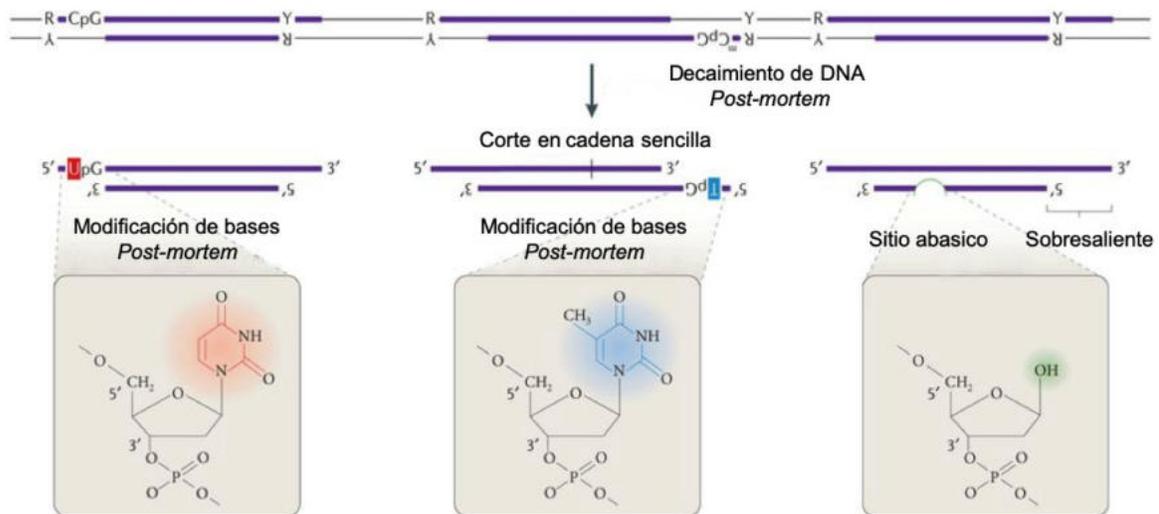


Figura 6. Patrones de Daño en DNA *post-mortem*. R, purinas; Y, pirimidinas.
(Figura modificada de Orlando y col., 2015).

Aún si el aDNA presenta patrones característicos de daño, debe considerarse que durante la descomposición de un organismo diversos microorganismos invadirán el cadáver y depositarán su material genético hasta que éste decaiga de manera similar al organismo hospedero. Lo anterior complica la identificación de patógenos muy relacionados a organismos comunes ambientales (como suelo, si la muestra es de un entierro, por ejemplo). Un ejemplo es el caso de la bacteria *Mycobacterium smegmatis*, cuyo DNA se encuentra de manera cotidiana en el ambiente y puede confundirse con el de *Mycobacterium leprae* y *Mycobacterium tuberculosis* (patógenos de humanos), mostrando patrones de daño característicos. A su vez, *Mycobacterium leprae* presenta un decaimiento de DNA reducido (quizá por los componentes químicos de su pared celular), por lo que deben tenerse controles adecuados para evitar interpretaciones equívocas (Warinner y col., 2017).

1.8 Prácticas y recomendaciones en investigación con aDNA

Debido a las múltiples complicaciones existentes al trabajar con muestras antiguas deben tomarse medidas de higiene muy estrictas. Es necesario trabajar en un laboratorio con filtración de aire independiente, radiación UV cotidiana, esterilización con NaOCl, así como usar protección personal (trajes de cuerpo completo, doble par de guantes, cubre bocas, y lentes). Para poder evaluar posible contaminación, deben usarse siempre controles negativos en todos los experimentos, se recomienda un análisis riguroso de todos los reactivos antes de usarlos con muestras antiguas (se ha mostrado que algunas enzimas, dNTPS, oligonucleótidos y amortiguadores de pH comerciales cuentan con contaminaciones de DNA microbiano, quizá proveniente de vectores de expresión), así como llevar un flujo unidireccional. Específicamente para el estudio de microorganismos antiguos, estos análisis nunca se llevan a cabo en laboratorios donde se cultiven, amplifiquen o procesen muestras de microorganismos vivos o viables (virus) (Warinner y col., 2017).

1.9 Huesos y dientes para extracción de aDNA

Para el análisis de aDNA se han utilizado rutinariamente huesos y dientes como sustrato para extracción de material genético. La parte más densa y mejor protegida del cráneo mamífero se encuentra en la parte interna del hueso petroso, siendo éste el lugar donde hay más cantidad de moléculas endógenas de DNA y en mejor estado de conservación, incluso en ambientes cálidos como los de África u Oceanía (Hansen y col., 2017). Sin embargo, actualmente su extracción es altamente destructiva y vuelve problemático su uso en muestras preciosas, ya que sólo hay dos huesos petrosos por cráneo (Key y col., 2017). Además, su uso es limitado para estudios de microbioma o detección de patógenos, ya que es poco accesible a microorganismos por lo que presenta poca o nula diversidad de patógenos (Margaryan y col., 2018).

La conservación de patógenos en muestras óseas depende de su tropismo y del estado de infección en que el individuo haya muerto. Debido a que muchas enfermedades no producen cambios osteológicos o producen muerte en su fase aguda de infección, se ha analizado material dental para detección de patógenos. La cámara interna del diente se encuentra muy vascularizada durante toda la vida de los mamíferos, permitiendo así, la obtención DNA de patógenos que estuvieron presentes en altas concentraciones en el torrente sanguíneo mientras que el hospedero vivía (Key y col., 2017). Además del tejido óseo, se han utilizado tejidos antiguos momificados (Figura 7) para la detección de patógenos (Key y col., 2017), como en el caso de muestras con posible Papillomavirus (VPH) (Fornaciari y col., 2003); VARV (Duggan y col., 2016) y HBV (Kahila y col., 2012; Ross y col., 2018). De la misma forma, se han analizado muestras de museo con VARV (Pajer y col., 2017) o con “Monkeypoxviurs” (MPX) (Tiee y col., 2018) y tejido congelado o fijado en parafina para detección de influenza (Reid y col., 1999).

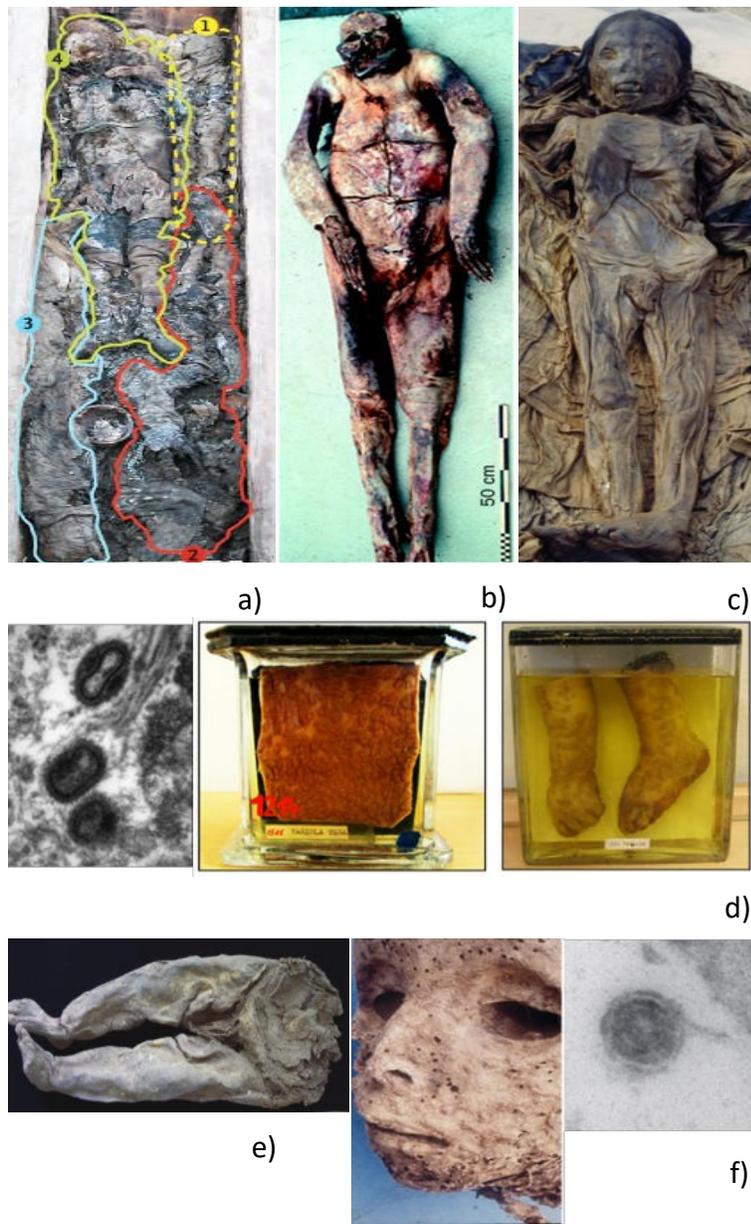


Figura 7. Momias representativas en paleogenómica viral. a) Momia de Siberia (siglo XVIII); c) momia de Nápoles (siglo XVI); momia coreana (siglo XVI); d) muestras de museo de Praga (siglo XX y XIX), y visualización de partículas de

VARV (izquierda); e) momia de Lituania (siglo XVII); f) momia de Palermo (siglo XVI) y visualización de partículas de HBV (derecha).

1.10 Procedimientos para extracción y amplificación de aDNA

Para la extracción de aDNA deben evitarse tratamientos agresivos como temperaturas altas o uso de detergentes fuertes, también debe considerarse que inhibidores de PCR son co-purificados con DNA de muestras de huesos y dientes (Rohland y Hofreiter, 2007). De igual forma, el DNA a amplificar puede presentar lesiones que inhiban PCR, muchas de estas causadas por radicales libres como el anión superóxido (O_2^-), el peróxido de hidrógeno (H_2O_2) y los radicales hidroxilos (OH^-), en ocasiones inducidos por radiación ambiental. El ataque oxidativo puede ocurrir en enlaces dobles de pirimidinas y purinas, fragmentando el anillo en su estructura; de manera similar, el enlace glucosídico entre desoxirribosas puede fragmentarse por oxidación. Muestras con cantidades elevadas de pirimidinas oxidadas (5-hidroxi-5metilhydantoin y 5-hidroxihydantoin) no pueden amplificarse por PCR con Taq DNA polimerasa (Pääbo y col., 2004). El protocolo de Rohland y col. (2007) permite extracción de DNA con SiO_2 de manera isotérmica, evitando degradar aún mas el material genético de la muestra y remueve considerablemente inhibidores de PCR; a su vez, este protocolo hace uso de una concentración alta de EDTA que permite la disolución de la matriz de hidroxiapatita de huesos y dientes (Rohland y col., 2007).

En la amplificación de aDNA, las polimerasas comunes no extenderían la cadena al encontrarse con residuo de uracilo (producido por desaminación de citosinas). Sin embargo, se han modificado enzimas comerciales para amplificar DNA con uracilo, lo cual permite amplificación de DNA con patrones típicos de daño, aunque también se incrementan las mutaciones G/C a A/T. Para esto se utilizará una enzima comercial modificada para incorporar uracilo, “Phusion U Hot Start DNA Polymerase” (Thermofisher, 2018).

1.11 Técnicas de secuenciación

Con la llegada de la primera plataforma de secuenciación de alto rendimiento en 2005 con el secuenciador Roche 454, seguida de la comercialización de tecnologías Illumina, hubo una disminución de 50,000 veces en los costos de secuenciación, y un aumento en su capacidad de generación de lecturas de 100 a 1,000 veces, tomando como punto de comparación el Proyecto del Genoma Humano (Goodwin y col., 2016). Comúnmente, se les conoce a estas tecnologías en conjunto como secuenciación de nueva generación (NGS, por sus siglas en inglés). Sin embargo, ninguna tecnología viene sin limitaciones y, comparativamente con la secuenciación por Sanger, se tienen tasas de error mayores (0.1-15%) con longitudes de lecturas generalmente más cortas (35-700 bp para métodos de lecturas cortas). Si bien, existen tecnologías que permiten la secuenciación de lecturas largas, su costo continúa siendo alto, su tasa de error es mayor (Goodwin y col., 2016), además no se esperan lecturas largas en estudios de aDNA.

Las técnicas de NGS más usadas actualmente son las de síntesis (Illumina) que usan una polimerasa y una señal a detectar por el secuenciador (fluoróforo) con el fin de detectar la incorporación de un nucleótido. La amplificación se realiza en una superficie sólida, aumentando el número de copias de DNA en un área específica (Goodwin y col., 2016).

Para la generación de las librerías de secuenciación, el primer paso es fragmentar el DNA genómico en longitudes menores; esto se puede lograr mecánicamente por sonicación o enzimáticamente. Posteriormente, el segmento se liga a un adaptador común para amplificar y secuenciar. La fragmentación de DNA puede omitirse cuando se trabaja con aDNA pues las cadenas ya se encuentran en fragmentos pequeños.

Para la secuenciación por síntesis se inmovilizan templados a una celda con variado número de líneas (según el modelo), creando una fase sólida para la amplificación de hasta 1000 copias idénticas en un diámetro de una micra o menos. Primero, el DNA ligado a adaptadores se une a adaptadores inmovilizados en la celda, que están en dirección 5' y 3', creando así un fragmento de cadena doble de DNA en forma de puente que servirá como cebador a la DNA polimerasa para comenzar su síntesis. En cada ciclo se agregan deoxinucleotidos (dNTPs) bloqueados reversiblemente en grupo hidroxilo en posición 3' de ribosa (similar a secuenciación por Sanger), que a su vez están marcados con un fluoróforo excitable con láser. Posteriormente, se remueven los dNTPs, se emite un láser que excita al fluoróforo, se captura y procesa la imagen. Una vez procesada la señal, el fluoróforo es escindido enzimáticamente para permitir la incorporación del próximo nucleótido. Esta especie de amplificación se hace de manera isotérmica a 60 °C usando soluciones amortiguadoras que median la desnaturalización de cadena doble de DNA (Buermans y Dunnen, 2014; Goodwin y col., 2016).

Illumina es la plataforma comercial con más auge en los últimos años, dominando la industria de secuenciación de lecturas cortas, y es un sistema de terminación cíclica reversible (síntesis). Cuenta con secuenciadores para muy diversos usos, capaces de leer diferentes tamaños de lecturas y con diferentes rendimientos. La identificación de dNTPs en plataformas Illumina se realiza con microscopía interna de fluorescencia, usando dos o cuatro canales láser. En este último, cada fluoróforo se une a un tipo distinto de base. Las plataformas NextSeq y Mini-Seq (bajo rendimiento) usan dos fluoróforos (Goodwin y col., 2016) y pueden generar un máximo de 400 y 25 millones de lecturas por corridas de entre 12-30 horas y 4-24 horas, respectivamente (Illumina, 2019). Illumina también cuenta con secuenciadores de alto rendimiento, como el HiSeq 400 capaz de secuenciar hasta

5 mil millones de lecturas por corridas de menos de 4 días, y el NovaSeq 6000 que es capaz de secuenciar hasta 20 mil millones de lecturas por corridas de hasta 44 horas; mismo que es el secuenciador de mayor rendimiento disponible en la actualidad (Illumina, 2019).

1.12 Secuenciación de aDNA para detección de virus

En el caso de patógenos virales antiguos se tienen pocas probabilidades de encontrarles en substrato arqueológico, ya que sus genomas son órdenes de magnitud más pequeños que el genoma humano o el de muchas bacterias (Key y col. 2017) y los virus de RNA o cadena sencilla de DNA (ssDNA) presentan poca o nula preservación de material genético a través de los años (Key y col., 2017). Existen contados estudios donde se han reconstruido genomas virales de DNA (Cuadro 3 y 4).

Cuadro 3. Genomas antiguos de HBV. Datación reportada como media a partir de calibración con ¹⁴C; ND: No disponible.

Virus/ Clasificación de Baltimore	HBV / VII (RT-DNA)			
	Número de muestras totales	1	1	53
Genomas virales (>80%)	1	1	3	11
Sustrato arqueológico	Tejido	Tejido y hueso	Restos óseos	Restos óseos
Datación	1682 D.C.	1569 D.C.	2989 A.C., 1203 A.C., y 1068 D.C.	Desde 2403 A.C. hasta 1262 D.C.

Cobertura	ND	163.8x	104x, 47x y 46x	4.3x-89.2x
Referencia	Kahila y col., 2012	Ross y col., 2018	Krause-Kyora y col., 2018	Mühlemann y col., 2018

Actualmente, HBV es el agente viral del que se han reconstruido más genomas antiguos. Se ha obtenido su genoma a partir de momias del siglo XVI encontradas en Corea (Kahila y col., 2012) y en Italia (Ross y col., 2018) o de muestras óseas (Krause-Kyora y col., 2018; Mühlemann y col., 2018) de hasta más de 4000 años de antigüedad (Cuadro 3).

Cuadro 4. Genomas antiguos de VARV y B19V. Datación reportada como media a partir de calibración con ^{14}C ; * : Racemización D, L-acido aspártico.

Virus	VARV		B19V	
Clasificación de Baltimore	1 (dsDNA)		II (ssDNA)	
Número de muestras totales	1	2	1578	
Genomas virales (>70%)	1	2	8	
Sustrato arqueológico	Tejido momificado	Pies/piel de museo	Restos óseos	
Datación	1654 D.C.	1857* D.C. y 1957* D.C.	4777 A.C. hasta 1541 D.C.	
Cobertura	18x	1700x y 6900x	1.6x-7.8x	
Referencia	Duggan y col., 2016	Pajer y col., 2017	Mühlemann y col., 2018	

En el caso de VARV (Cuadro 4) se ha identificado DNA en una momia siberiana de 300 años de antigüedad mediante PCR (Biaggini y col., 2012); y se ha reconstruido un genoma completo a partir de una momia de Lituania del siglo XVII (Duggan y col., 2016), y de muestras de museo en Praga con datación de 60 y 160 años de antigüedad (Pajer y col., 2017). Toppinen y col. (2015) lograron la amplificación mediante PCR tiempo real (qPCR) de productos de B19V en 43 muestras óseas de 106 individuos muertos durante segunda guerra mundial; y más recientemente, se logró la obtención de 10 genomas completos/parciales de parvovirus B19V a partir de muestras óseas de Eurasia (Mühlemann y col., 2018) (Cuadro 4). Fornaciari y col. (2003) reportaron amplificación de PCR con diana a genes de VPH en momia del siglo XVI; sin embargo, es el único grupo que lo ha conseguido y no se han obtenido genomas completos.

En tanto a virus de RNA que infecten humanos, sólo se ha podido recuperar influenza (Taubenberger y col., 2005) y HIV-I (Worobey y col., 2016) de tejido congelado del siglo XX.

1.13 Virus recuperables a partir de sustrato arqueológico óseo.

Actualmente sólo ha sido posible recuperar genomas completos de HBV y B19V a partir de dientes y huesos pertenecientes a restos arqueológicos. Por lo que se describen de manera general en este apartado.

1.13.1 Parvovirus humano (B19V) y su amplio tropismo.

La familia *Parvoviridae* está compuesta por dos subfamilias, *Densovirinae* y *Parvovirinae*. Dentro de la última se encuentra el género *Erythroparvovirus* que incluye 6 especies virales (ICTV, 2019); de las que el parvovirus de primates 1 (B19V) y 2 (SPV) son los únicos patógenos a humanos perteneciente a este

género; aunque existen otros géneros con parvovirus patógenos a humanos como *Bocaparvovirus*, *Protoparvovirus* y *Tetraparvovirus* (Qiu y col., 2017).

B19V es transmitido por vías respiratorias, por contacto sanguíneo o vía vertical de madre a hijo (Broliden y col., 2006). B19V pertenece a la clasificación de Baltimore II con un genoma de DNA de cadena sencilla y una longitud de cerca de 5500 nucleótidos (Figura 8). En los extremos terminales, el genoma está flanqueado por dos secuencias repetidas (ITR), idénticas e invertidas, lo cual crea una estructura de DNA de doble cadena (dsDNA) al formarse palíndromos imperfectos de aproximadamente 400 nt en cada extremo (Figura 8) (Luo y Qiu, 2015).



Figura 8. Estructura del genoma de B19V. Modificado de Luo y Qiu (2015).

B19V infecta células humanas progenitoras de la línea eritroide (hEPCs) en médula ósea y/o hígado, y es causante de desórdenes hematológicos desde moderados hasta severos (Luo y Qiu, 2015). Estas complicaciones clínicas van desde hidropesía fetal, *Erythrema infectiosum* (quinta enfermedad) en infantes (Figura 9), así como artropatía en adultos sanos (en particular en mujeres de mediana edad) (Janovitz y col., 2017; Qiu y col., 2017).

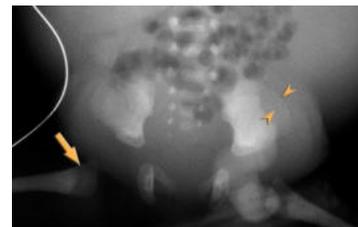
B19V se ha encontrado recurrentemente en individuos con anemias severas (Broliden y col., 2016). Cuando se produce una infección de B19V en pacientes inmunosuprimidos, con anemia hemolítica, talasemia, anemia falciforme, deficiencia de hierro, hemorragias o co-infección con malaria, se han reportado casos de crisis aplásica transitoria resultando en fatiga, dificultad para respirar, agrandamiento de

corazón, insuficiencia cardíaca, e infecciones frecuentes (Broliden y col., 2006; Qiu y col., 2017).

Además de las manifestaciones típicas mencionadas, B19V se ha asociado a enfermedades cardiovasculares, como miocarditis en adultos y niños, así como con enfermedades de hígado, como hepatitis aguda/crónica, falla hepática fulminante y hepatitis autoinmune (Luo y Qiu, 2015; Qiu y col., 2017). Aunque se ha calculado una alta tasa de seropositividad (60-90%) en adultos (Broliden y col., 2006; Janovitz y col., 2017), actualmente, no existen vacunas o antivirales para combatir este virus (Bonvicini y col., 2019). Para la infección de B19V se requiere cierto nivel de diferenciación, incluida la expresión del marcador de superficie CD36 en plaquetas, monocitos, macrófagos y células endoteliales originadas a partir de células madre hematopoyéticas (HSCs) CD34+ (hEPCs y células endoteliales) (Abbas y col., 2015; Luo y Qiu, 2015). El receptor primario de B19V es el antígeno P de eritrocitos (globósido) y se ha sugerido la integrina $\alpha 5\beta 1$ y KU80 como receptores secundarios que se encuentran en precursores de línea eritroide y en algunas células no eritroides como miocitos fetales, placenta, megacariocitos y células endoteliales. Adicionalmente, se ha reportado la infección de B19V en líneas celulares monocíticas y endotelio de miocardio mediante entrada facilitada por anticuerpos (con replicación abortiva), lo anterior podría explicar la infección de células de líneas no eritroides y las diversas manifestaciones en distintos tejidos (Luo y Qiu, 2015).



© MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH. ALL RIGHTS RESERVED.



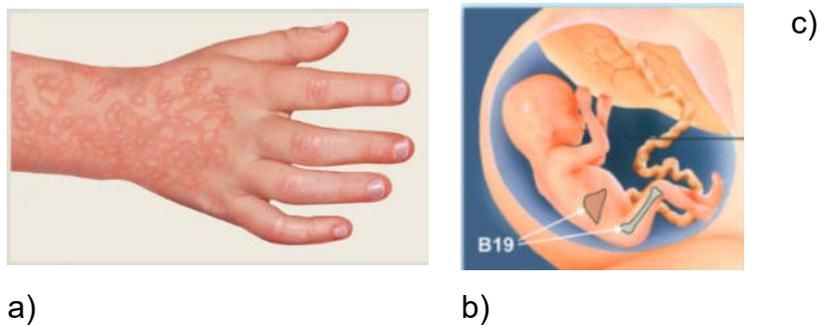


Figura 9. Síntomas asociados a infección con B19V. a) *Erythema infectiosum* (Clinica Mayo, 2019) b) hidropesía fetal (Broliden y col., 2006); c) Diferencia en densidad ósea de huesos largos asociada a infección de B19V (Cantey y col., 2013).

B19V es capaz de integrarse en el genoma del hospedero humano mediante la proteína viral NS1. Esta proteína participa en la resolución terminal del genoma de B19V, y se ha sugerido su capacidad de escindir DNA cromosómico del hospedero humano para la integración del genoma viral, con mayor afinidad a regiones intrónicas y sitios H3K9me3 (Janovitz y col., 2017). Los sitios de integración en el humano contienen una secuencia consenso de 41bp con sitio de corte potencial para NS1, así como un motivo central de 6-10 nt con homología al promotor P6 de B19V (Janovitz y col., 2017).

Después de la infección de B19V su DNA es detectable en diversos tejidos, como medula ósea, hueso, corazón, cerebro, hígado, riñones, nódulos linfáticos, glándulas salivarias, amígdalas, tejido sinovial y testículos (Broliden y col., 2006; Janovitz y col., 2017; Qiu y col., 2017).

En el caso de infección congénita con B19V se han encontrado lesiones en huesos largos de infantes (Figura 9c) con hidropesía fetal después de 10 semanas de nacidos (Cantey y col., 2013); sin embargo, no hay marcas específicas *post-mortem*

asociadas a infección de B19V. B19V se ha encontrado mediante PCR en restos óseos de humanos con 70 años después de su fallecimiento (Toppinen y col., 2015) y recientemente se obtuvieron 10 genomas (enteros/parciales) a partir de restos óseos de hasta 6.9 mil años mediante NGS (Mühlemann y col., 2018).

B19V cuenta con un sólo serotipo que incluye a los 3 genotipos. La mayoría de las infecciones son actualmente causadas por el genotipo 1, que se encuentra en todo el mundo. El genotipo 2 se considera casi extinto y se ha encontrado principalmente en individuos nacidos antes de 1970 en el norte de Europa (Jain y Kant, 2018). El genotipo 3 es endémico de Ghana, Brasil e India (Pyoria y col., 2017) y cuenta con la mayor diversidad intragenotípica, lo cual sugiere una historia evolutiva más antigua. Este genotipo se divide en dos subgenotipos; el 3A es encontrado mayormente en África oriental (Ghana); mientras que, el 3B se ha encontrado ocasionalmente fuera de África, como en Brasil e India, y en extranjeros muestreados en Alemania, provenientes de Marruecos, Egipto y Turquía (Hübschen y col., 2009).

En México se ha detectado B19V mediante serología en mujeres mayas con padecimientos reumatológicos, donde además se detectaron amplicones del genotipo 1 mediante qPCR (Valencia-Pacheco y col., 2017). No se ha reportado a la fecha genotipos 2 o 3 en México, que es coherente con su distribución en el norte de Europa y Africa, respectivamente (Hübschen y col, 2009; Pyoria y col., 2017).

1.13.2 HBV y genotipos en afrodescendientes.

El virus de hepatitis B es uno de los patógenos más prevalentes en la actualidad, cerca del 3.5 % de la población mundial (240 millones) se encuentra infectada de manera crónica (Tu y col., 2017; Yuen y col., 2019), con una tasa de muerte global cerca de 1 millón de personas siendo el agente causal más conocido de la cirrosis

hepática y cáncer de hígado (hepatocarcinoma) (Tu y col., 2017). Su transmisión es horizontal mediante fluidos sanguíneos y genitales, aunque también se transmite verticalmente de madre a hijo (Inoue y Tanaka., 2016).

HBV se ha observado casi exclusivamente en hepatocitos, donde se encuentra su receptor principal (NTCP) (Watashi y Wakita, 2015). Además, HBV es capaz de integrarse en el genoma humano, con preferencia a sitios de corte de cadena doble (Tu y col., 2017). HBV tiene un genoma circular relajado de cadena doble de DNA (dsDNA) (Figura 10) con aproximadamente 3.2 kb y una región de RNA de cadena sencilla en el extremo 3'. Dicho genoma codifica 4 marcos de lectura superpuestos entre ellos (Tu y col., 2017).

A pesar de ser una especie viral amplia y profundamente estudiada, su historia evolutiva aún es incierta y controversial debido a la complejidad de integrar todos los datos existentes (Krause-Kyora y col. 2018; Ross y col., 2018; Yuen y col., 2018). HBV es capaz de infectar primates no humanos, como chimpancés y gorilas; estas cepas son genéticamente similares a cepas humanas, por lo que se ha propuesto que el origen de la especie fue en África (Krause-Kyora y col., 2018).

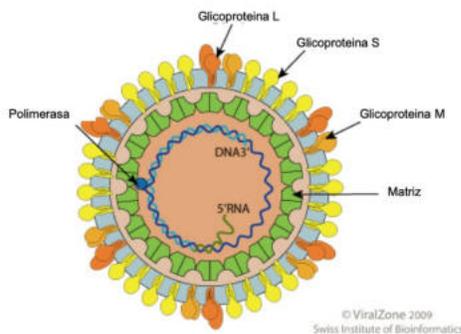


Figura 10. Estructura de una partícula de HBV con genoma dsDNA circular relajado (Swiss Institute of bioinformatics, 2019).

HBV es clasificado en mínimo 10 genotipos (HBV/A-J) (Kahila y col., 2012; Ross y col, 2018; Yuen y col., 2018) con distribuciones geográficas distintivas, por lo que se han propuesto modelos para explicar su origen. El primero propone un origen americano posterior al contacto con europeos, debido al posicionamiento basal de HBV/F y HBV/H, que son encontrados de forma endémica únicamente en América; sin embargo, es inconsistente con la diversidad encontrada en África (sugiere historia evolutiva larga), y su amplia distribución en primates no humanos. Otro modelo propone que HBV co-evolucionó con humanos a partir de su salida de África hace 60-100 mil años, debido a su cercanía filogenética respecto a primates africanos, mas es incoherente con el posicionamiento basal de HBV/F (encontrado en nativo americanos), sin mencionar que bajo este modelo los primates deberían tener una diversidad más alta de HBV, al tener una historia evolutiva más larga que los humanos (Locarni y col., 2013). Un tercer modelo usando métodos bayesianos propone origen de HBV hace 33600 años, usando calibración con fechas de migración de poblaciones nativo americanas (con HBV/F y HBV/H), expansión polinesia en el pacífico (con HBV/D4) y migración de África a Haití (con HBV/A5), donde se propone múltiples eventos de infecciones inter especies de humanos a primates no humanos; sin embargo, la diversificación de HBV a Asia, Europa y Sudamérica en solo unos pocos miles de años ha sido motivo de discusión (Paraskevis y col., 2013). Los estudios más recientes con aDNA han propuesto un origen de la especie hace cerca de 20.9 mil años (Mühlemann y col., 2018), basados en genomas completos reconstruidos a partir de restos arqueológicos. Sin embargo, la estimación de una tasa de mutación es compleja debido a que sus regiones codificantes se encuentran sobrepuestas y a que la tasa de mutación varía dependiendo de la fase de infección (Mühlemann y col., 2018).

Las distribuciones de los genotipos de HBV se han correlacionado con patrones de migraciones humanas (Kahila y col., 2012; Mühlemann y col., 2018; Yuen y col., 2019). Ejemplos de esto son, el aislamiento de HBV/F y HBV/H en América (Locarni y col., 2013), HBV/C4 que sólo está presente en nativos de Australia (Yuen y col., 2019), y HBV/A que se ha asociado a migraciones forzadas de africanos (Facholi y col., 2008; Anderchach y col., 2009; Brichlet y col., 2013; Lago y col., 2014; Soares y col., 2018).

Actualmente en África el genotipo mas común es el HBV/E (Senegal hasta Angola), seguido del HBV/A (Pigmeos y Bantúes) (Olinger y col., 2006). HBV/E presenta una diversidad genética baja, lo cual refleja una historia evolutiva reciente; se ha propuesto su aparición dentro de los últimos 150-200 años (McMahon, 2009; Brichler y col., 2013) después del término de comercio de africanos esclavizados (Andernach y col., 2009). Mientras que, HBV/A tiene una mayor diversidad en África (Andernach y col., 2009) por lo que se ha propuesto un origen africano (Kramvis y col., 2013) donde aún es frecuente; así como en países con historia demográfica relacionada a África. En India y países latinoamericanos se ha propuesto la introducción de este genotipo con el comercio de africanos esclavizados (Brichler y col., 2013; Lago y col., 2014; Soares y col., 2018). Los subgenotipos HBV/A3 y el HBV/A4 se han identificado sólo en África y afrodescendientes, mientras el HBV/A2 se ha encontrado en el sur de África, noroeste de Europa y norte de América (Soares y col., 2018).

Los subgenotipos HBV/A1 y HBV/A2 son los más esparcidos entre afrodescendientes fuera de África y se ha asociado al traslado forzado de personas esclavizadas a islas de Cabo Verde (Soares y col., 2018), Brasil (Facholi y col., 2008; Lago y col., 2014), Haití (Anderchach y col., 2009) y Martinica (Brichlet y col., 2013). Además, el genotipo A se ha encontrado con mayor frecuencia en poblaciones afro-venezolanas que en el resto de la población de Venezuela

(Quintero y col., 2002), así como en Nahuas de México (Roman y col., 2010) donde de igual forma se ha asociado a la introducción de este genotipo relacionada con origen africano (Quintero y col., 2002; Roman y col., 2010).

1.14 Análisis metagenómico en aDNA.

Existen diferentes programas utilizados para identificación de taxas en muestras complejas de DNA que hacen uso de distintos enfoques. En el caso de análisis metagenómicos en aDNA la herramienta canónica es BLAST (Altschul y col., 1990), aunque también se han utilizado otros enfoques como Kraken (Wood y Salzberg, 2014), MetaPhlAn (Segata y col., 2012), MIDAS (Nayfach y col., 2016), QIIME (Kuczynski y col., 2011) y mas recientemente MALT (Herbig y col., 2016).

BLAST representa una herramienta robusta y flexible; sin embargo, al haber sido desarrollada previo a la época de datos genómicos masivos, su tiempo de ejecución es muy lento y costoso computacionalmente, lo cual lo vuelve prohibitivo para el análisis con millones de lecturas (Wood y Salzberf, 2014), como es el caso de este estudio.

Para evitar esta limitante, una serie de programas se han utilizado para la asignación taxonómica con datos de metagenómica. Algunos realizan esta tarea con un panel de genes informativos, como QIIME/UCLUST, MetaPhlAn y MIDAS. Otros usan k-meros a partir de genomas completos de referencia, como DIAMOND y Kraken; incluso se han usado programas híbridos de k-meros con alineamientos tipo “semilla extensible”, como MALT. Todas estas metodologías han sido usadas en el análisis de aDNA y DNA moderno.

QIIME/UCLUST utiliza la región 16 rRNA; este enfoque es inútil para la identificación de patógenos virales pues no tienen ribosomas, ni genes conservados en todas las especies descritas, ya que no se consideran entidades

monofiléticas (Moreira y col., 2009). Por su parte, MetaPhlAn y MIDAS utilizan un panel de referencia de genes en una sola copia (Velsko y col., 2018); sin embargo, es de esperar que en una muestra compleja como la nuestra (ambiental), la abundancia de DNA de patógenos sea baja (Huebler y col., 2019), haciendo que este enfoque carezca de sensibilidad para análisis metagenómico con aDNA por lo que muy posiblemente los genes dianas no se recuperen en análisis metagenómico (Huebler y col., 2019). DIAMOND (Buchfink y col., 2015) asigna taxones usando el paradigma de “semilla extensible”, utilizando solo una parte de una semilla larga (lectura de interés) que ordena en una matriz asociativa respecto a lista de semillas de base de referencia, reduciendo notablemente el tiempo de corrida (20,000 veces comparado con BLASTx) (Buchfink y col., 2015) con una base de datos de referencia de proteínas. Kraken utiliza k-meros a partir de genoma completo y ha demostrado ser altamente eficaz en análisis de aDNA siendo de los programas más rápidos a la fecha (Huebler y col., 2019). Finalmente, MALT hace uso del aproximamiento de semilla extensible (similar a BLAST), para alinear secuencias contra una matriz asociativa que contiene semillas poco frecuentes en la base de datos de referencia, volviendo el proceso 120 veces más rápido que BLAST (Herbig y col., 2016); sin embargo, es más lento que Kraken (Huebler y col., 2019).

Huebler y col. (2019) muestran que MALT es capaz de asignar taxonómicamente lecturas en muy baja abundancia (0.00001 %), con resultados más confiables (en base a lecturas generadas *in silico*) que Kraken. Además, Kraken presenta alta tasa de falsos positivos para taxas en baja abundancia (Huebler y col., 2019), como es de esperar al trabajar con aDNA viral.

MALT se ha utilizado exitosamente para el análisis de datos metagenómicos antiguos (Krause-Kyora y col., 2018; Vagene y col., 2018), y está automatizado para producir un archivo de salida que puede ser directamente visualizado en MEGA (Huson y col., 2016), donde se puede evaluar la distribución de lecturas

para descartar falsos positivos. Por lo anterior, MALT fue escogido para realizar la asignación taxonómica viral.

1.15 Enriquecimiento de aDNA

Además de la estrategia de metagenómica tradicional a partir de datos de secuenciación de librerías crudas de DNA, mencionado previamente, puede realizarse el enriquecimiento de DNA específico sobre las librerías de secuenciación. A este proceso se le conoce como “captura” y actualmente se realiza en fase sólida (con micro-arreglos) o en solución; el objetivo es enriquecer DNA presente en baja abundancia, como el de ciertos patógenos (Warinner y col., 2017).

En estudios de metagenómica con DNA moderno proveniente de muestras complejas (suero humano), se ha encontrado dificultad de obtener suficiente material genético para ensamblaje de genomas completos incluso a partir de secuenciación a razonable profundidad (>4.5 millones de lecturas), como en el caso de brotes de virus de Zika que presenta bajos títulos virales (Metsky y col., 2019) y un tamaño de genoma relativamente pequeño (11 kb). Para aumentar la sensibilidad de estudios de metagenómica, Metsky y col. diseñaron ensayos de captura con sondas de ssDNA (75 nt) o ssRNA (100 nt) en solución, consiguiendo un enriquecimiento en órdenes de magnitud de lecturas asignadas a especies virales, volviendo posible el ensamblaje de genomas parciales o enteros de especies en baja abundancia a partir de muestras complejas (Metsky y col., 2019).

Para realizar la captura en solución, se diseñan “cebos” que hibriden con moléculas de DNA de un organismo diana en específico (Warinner y col., 2017). La producción de estos “cebos” se hace a partir de la transcripción *in vitro* de RNA moderno complementario a diana de interés bajo promotor de la RNA Polimerasa (T4) que a su vez incorpora UTPs biotinizados. Una vez generados los productos

de RNA, estos son hibridados a la librería de aDNA para su posterior captura con perlas magnéticas asociadas a estreptavidina. El sobrenadante es descartado; mientras que, las moléculas capturadas son eluidas, amplificadas y secuenciadas (Ávila-Arcos y col., 2015).

Los métodos de captura en solución se han utilizado ampliamente en el área de paleogenómica en donde el aDNA está en abundancias aún más bajas. Su uso en captura de patógenos virales es primordial y ha permitido obtención de genomas completos de virus como VARV (Duggan y col., 2016), HBV (Mühlemann y col., 2018; Ross y col., 2018) y B19V (Mühlemann y col., 2018).

1.16 Herramientas para análisis de datos paleogenómicos.

A continuación se describen las herramientas más importantes que se utilizaron en este proyecto para identificación, reconstrucción, autenticación y análisis filogenéticos de genomas antiguos y que, además, han sido ampliamente utilizadas en paleogenómica.

1.16.1 BLAST.

Publicado por Altschul y col. (2009), es un programa usado para alinear secuencias respecto a una base de datos conocida.

Para la alineación de las secuencias de interés, el primer paso es usar una palabra de cierto tamaño (semilla), determinada por el usuario. Se crea una lista de semillas que se compara con la base de datos a usar. Una vez que se identifica un alineamiento perfecto de la semilla con una secuencia de la base de datos, BLAST extiende el alineamiento en ambas regiones vecinas (Lobo y col., 2008).

Existen ciertos valores para filtrar las secuencias alineadas, entre los que más importan son el “e-value” y el “bitscore”, que determinan si la identidad entre dos

secuencias es estadísticamente significativa y, por tanto, si las secuencias están relacionadas. El “bitscore” indica el tamaño que una base de datos debe tener para encontrar el alineamiento por puro hazar, por lo que es usado para comparar alineamientos en bases de datos que aumentan constantemente. Mientras más grande es el “bitscore” mejor es el alineamiento. El “e-value” es una corrección del “bitscore” respecto al tamaño de la base de datos usada, mientras más pequeña la base de datos el “e-value” de un alineamiento es mejor en comparación con una base de datos mayor. Mientras más pequeño es el “e-value” mejor es el alineamiento (Pearson, 2013).

Existen distintos tipos de alineamientos dentro de la familia BLAST: BLASTn compara nucleótidos contra base de datos de nucleótidos, BLASTp compara proteínas contra base de datos de proteínas, y BLASTx compara nucleótidos traducidos contra base de datos de proteínas (Lobo y col., 2008).

Actualmente, existen programas más rápidos que BLAST usados en metagenómica (Kraken, MALT, DIAMOND, MetaPhlAn, entre otros); sin embargo, BLAST se considera el programa canónico para el análisis de secuencias (Herbig y col., 2016).

1.16.2 MALT.

Por sus siglas en inglés “Herramienta de Alineamiento de MEGAN” fue desarrollado por Herbig y col. (2016). Realiza alineamientos rápidos y análisis metagenómicos a partir de datos de secuenciación de DNA. Realiza dos procesos comunes en análisis metagenómicos. El primero es el perfilamiento taxonómico, que en base a las lecturas de secuenciación estima el contenido microbiano sin intentar asignar a un organismo específico. Después realiza el agrupamiento taxonómico, que estima la estructura de la comunidad, y asigna el mayor número de lecturas a un taxa específico. Para esto, MALT usa el algoritmo LCA (último ancestro común) que a su

vez ha sido usado en MEGAN. Los archivos generados en MALT (formato rma6) son leídos automáticamente en MEGAN.

MALT ha mostrado ser notoriamente más rápido que el canónico BLAST, por lo que su uso es justificado en muestras de metagenómica (millones de lecturas). Ambos están basados en el paradigma semilla-extensible y realizan alineamientos locales (Algoritmo Smith-Waterman). Además, MALT puede hacer alineamientos semi-globales (Algoritmo Needleman-Wunsch), en donde las lecturas son alineadas en extremos. MALT consiste en dos programas, “malt-build” y “malt-run”, el primero realiza un índice utilizando una base de datos de referencia con secuencias a evaluar, busca ocurrencias de semillas y las coloca en una tabla asociativa, conservando sólo las que no ocurran frecuentemente. El segundo programa alinea lecturas a analizar contra la base de datos de referencia, creando una lista de semillas y extendiéndolas hasta obtener un alineamiento entero, se estima así el “bitscore” y el “e-value” los cuales pueden ser usados para conservar o descartar asignaciones (Herbig y col., 2016).

1.16.3 MEGAN.

Programa interactivo gráfico desarrollado por Huson y col. (2007) para el análisis de lecturas de secuenciación masiva en un contexto de metagenómica. Debido al reto computacional que implica analizar microbiomas (que pueden generar billones de lecturas) se desarrolló esta herramienta para la asignación taxonómica (con referencia a la taxonomía de NCBI) y funcional (SEED, eggNOG, KEEG entre otros), así como para la generación de análisis de coordenadas principales (PCoA), por mencionar algunas.

A grandes rasgos, las lecturas de secuenciación previamente alineadas (por herramientas como BLAST o MALT) son comparadas contra una base de datos para realizar una asignación taxonómica con el algoritmo LCA (último ancestro

común, por sus siglas en inglés). Las lecturas que alinean a genes muy conservados son asignadas a un nivel taxonómico alto (filo, familia, entre otros); mientras que, si lo hacen a genes específicos de un organismo, se asignan a un nivel taxonómico bajo (género o especie). MEGAN puede reconocer una etiqueta en el encabezado de una secuencia de referencia, para así asignar a un organismo específico, también puede obtenerlos de un archivo con índice (Huson y col., 2016).

1.16.4 BWA.

Publicado originalmente por Li y Durbin en 2009, es un paquete de software nombrado por el tipo de algoritmo (Burrows-Wheeler) que se usa para alineación de lecturas. Este programa permite mapear secuencias poco divergentes contra un genoma largo de referencia. El mapeo de las lecturas se realiza mediante alineamientos semi-globales. Específicamente, BWA cuenta con tres algoritmos, “BWA-MEM”, para lecturas de DNA provenientes de plataformas Illumina, Ion Torrent o Sanger con un tamaño de 70 bp o mayores; “BWA-backtrack” para secuencias cortas; y “BWA-SW” para mapeo de secuencias con espacios vacíos respecto al genoma de referencia. (Li y Durbin, 2009).

En el caso de aDNA se deben tener en cuenta parámetros especiales para evitar ser muy estricto en el análisis y, a su vez, lograr discriminar falsos positivos. Los parámetros que tienen mayor impacto para secuenciación mediante plataforma Illumina fueron evaluados por Schubert y col. (2012), siendo la desactivación de la semilla y la distancia máxima de edición los más importantes (Schubert y col., 2012). BWA usa 32 nt como semilla en extremo terminal permitiendo hasta 2 diferencias para extender y alinear lectura, lo cual es una limitante para la identificación de aDNA (dañado en extremos del fragmento) que es divergente a genomas actuales respecto al paso del tiempo. Al desactivar esta opción se incrementan lecturas mapeadas; sin embargo, se duplica el tiempo de análisis. A su

vez, BWA permite modificar la distancia máxima de edición, que es el número de cambios permitidos para asignar una lectura a una referencia. Automáticamente se permiten 3 diferencias para lecturas menores a 38 nucleótidos, 4 para 64 nucleótidos o más; sin embargo, para aDNA la desaminación de citosinas limita la asignación adecuada. El ser muy laxo en este valor permitiría asignaciones erróneas. Schuebert y col. (2012) analizaron el efecto de la eliminación de los extremos dañados de las lecturas de secuenciación, que resultó en un aumento mínimo (0.1%) del número de lecturas mapeadas respecto al mapeo inicial; a su vez, se incrementó el tiempo de análisis (Schubert y col., 2012).

1.16.5 SAMTools.

Publicado por Li y col. (2009), es un programa que permite gestionar archivos en formato SAM (mapa de alineamiento de secuencia, por sus siglas en inglés). Dicho formato sirve para almacenar alineamientos de secuencias nucleotídicas contra un genoma de referencia. Este formato compacta la información, evita cargar el alineamiento completo en memoria, y permite indexar posiciones genómicas de manera eficiente para asignar lecturas alineadas a un locus. Los archivos generados en formato SAM a partir de secuenciación masiva pueden ser leídos, escritos, editados, indexados o vistos mediante una serie de utilidades llamada SAMTools (Li y col., 2009).

1.16.6 MapDamage:

Publicado por Ginolhac y col. (2011), este programa permite computar los daños de nucleótidos recurrentes en aDNA (desaminación de citosinas) y los patrones de fragmentación en archivos con lecturas de secuenciación masiva (en formato SAM/BAM) después de ser mapeados contra un genoma de referencia (con BWA, por ejemplo). A su vez, los resultados son procesados automáticamente con un programa en R para obtener un resultado gráfico (Ginolhac y col., 2011). En 2013

se mejoró el programa al incorporar un modelo estadístico Bayesiano, donde los eventos de daño dependen sólo de la desaminación de citosinas (asumiéndolo como el patrón de daño más importante en base a muestras experimentales), dependiente de su posición y ruptura de cadenas, proporcionado así una probabilidad bayesiana de daño. El programa muta las bases siguiendo la matriz canónica de transición Hasegawa, Kishino y Yano (HKY) y luego agrega patrón de daño *post-mortem* a estas bases, para ser evaluadas mediante cadenas de Markov (Jónsson y col., 2013).

1.16.7 RAxML.

Programa desarrollado por Stamatakis en 2006, permite la estimación de árboles filogenéticos de máxima verosimilitud de manera rápida, permitiendo análisis de grandes conjuntos de datos. Su mayor ventaja es que cuenta con algoritmo que permite una búsqueda rápida del árbol con mejores valores de máxima verosimilitud (Stamatakis, 2014).

1.16.8 BEAST.

Es un software desarrollado por Drummond y col. (2012), este permite hacer un análisis evolutivo bayesiano implementando algoritmos (Cadena Markov Monte Carlo) para realización de inferencia filogenética bayesiana, datación de tiempo de divergencia, análisis de coalescencia, filogeografía, (mapeo con interfaz a GoogleEarth), relojes moleculares (estrictos o laxos), entre otros. Los parámetros de este programa son altamente flexibles y pueden agregarse mediante código XML, además, existe una herramienta gráfica integrada en el mismo paquete llamada BEAUti (Drummond y col., 2012) que realiza esta tarea sin necesidad de programar en XML. A su vez, se incluyen herramientas para visualización e integración de análisis. La última versión de BEAST hace uso de la librería BEAGLE, que permite el uso de procesadores múltiples (GPU, en caso de estar

disponibles), con el fin de correr procesos de forma paralela, disminuyendo su tiempo de procesamiento (Drummond y col., 2012).

2.HIPÓTESIS

Los individuos encontrados en entierros típicos de epidemia en La Concepción y/o Hospital San José de los Naturales fallecieron por complicaciones asociadas a infección por un agente viral, por lo cual es posible identificar dicho patógeno mediante análisis de DNA antiguo.

3.OBJETIVOS

3.1 General

Evaluar la presencia de aDNA de especies virales patógenas humanas en restos arqueológicos humanos de Ciudad de México mediante técnicas de secuenciación, metagenómica y biología molecular.

3.2 Específicos

- Evaluar la presencia de VARV mediante PCR.
- Evaluar la presencia de patógenos virales humanos mediante técnicas de secuenciación de última generación y metagenómica.
- Diseñar sondas de captura para enriquecimiento de aDNA viral.
- Realizar captura con sondas personalizadas para analizar la presencia de DNA viral mediante secuenciación de ultima generación y metagenómica.
- Reconstruir genomas antiguos de patógenos virales para asignación filogenética y determinar la fecha de divergencia del ancestro común más reciente.

4.METODOLOGÍA

4.1 Materiales

- Laboratorio de DNA antiguo: aislado de laboratorios modernos con radiación UV y con cuartos separados unidireccionales.
- Cuarto para procesar muestras.
- Traje de cuerpo completo personal.
- Cubre zapatos, mangas, cubre bocas y guantes.
- Caja aislada para corte de hueso.
- Cortador de hueso.
- Cámara UV.
- Cuarto para preparación de PCR.
- Laboratorio moderno.
- Termociclador qPCR (StepOnePlus, Applied Biosystems).
- Termociclador PCR (Veriti 96 Well, Applied Biosystems)
- Material para PCR.
- Cebadores personalizados (diseño en 4.2.7).
- Material básico para biología molecular.

- Magneto DynaMag-2 (Invitrogen).
- Clúster computacional (Laboratorio Nacional de Visualización Científica “LAVIS”).
- Paquete comercial “NEBNext” (New England Biolabs Inc.).
- Equipo de electroforesis capilar, Bioanalyzer 2100 (Agilent).
- Paquete comercial “Min Elute” y “QIAquick” para purificar PCR (Qiagen).
- Secuenciador Illumina NextSeq 500 (Laboratorio Nacional de Genómica para la Biodiversidad “LANGEBIO”).
- Fluorómetro, Qubit 3.0 (Invitrogen).

4.2 Métodos

4.2.1 Obtención de muestras:

Se recurrió al osario en la Escuela Nacional de Antropología e Historia (ENAH) en Ciudad de México (CMX) para la obtención de muestras humanas de sitios arqueológicos (Capilla “La Concepción” en Coyoacán y Hospital Real San José de los Naturales en el antiguo barrio de Moyotlán) con evidencia histórica-arqueológica de posible epidemia por viruela. Se tomaron los premolares y se seleccionaron en base al estado de conservación de las piezas. Éstas se transportaron al laboratorio de DNA antiguo del Laboratorio Internacional de Investigación sobre el Genoma Humano (LIIGH) en Querétaro para su tratamiento posterior. Los dientes o sus fragmentos nunca estuvieron en contacto con el laboratorio moderno del LIIGH.

4.2.2 Procesamiento muestras

La superficie de las muestras se limpió arduamente con hipoclorito de sodio (al 70%) y etanol (70%), seguida de exposición a UV durante 1.5 min. Las piezas dentales se cortaron con disco metálico, y se separó la raíz de la corona. Las coronas se guardaron; mientras que, las raíces fragmentadas se separaron en dos tubos, cada uno con cerca de 200 mg de muestra.

4.2.3 Extracción de DNA

Se usó el protocolo de Rohland y col. (2007) optimizado para 200 mg de muestra.

Se mezclaron 4.8 g de dióxido de silicio (SiO_2) en 40 ml de agua estéril, y se asentó durante 1 h. Después se transfirieron 39 ml de la mezcla a un nuevo tubo de 50 ml y se dejó sedimentar por 4 h. Subsecuentemente, se desecharon 35 ml del tubo, luego, se agregaron 50 μl de HCl (37%) al sedimento. Se prepararon alícuotas de 500 μl y se envolvieron cada una en papel aluminio para guardarlas.

Después, las muestras se pre-digirieron con 1 ml de amortiguador de extracción (0.45 M EDTA, 1 % amortiguador TE, 1 % proteinasa K y 0.5 % N-laurilsarcosina) por cada una y se incubó durante 15 min a 37 °C, seguido por centrifugación a 1000 g durante 1 min con remoción del sobrenadante.

Las muestras se digirieron para disolver matriz de hidroxiapatita de calcio con 4 ml de amortiguador de extracción y se incubaron toda la noche a 37 °C.

Se purificaron ácidos nucleicos con solución sódica (500 ml amortiguador PB, 500 μl rojo fenol, 2.5 ml de 5 M NaCl, 9 ml de 5 M acetato de sodio, pH 4-4.5). Las muestras digeridas se centrifugaron a 2000 g durante 2 min y se recuperaron los sobrenadantes.

En los tubos con muestra se agregaron 50 μ l del preparado de sílice y 40 ml de amortiguador de unión (5 M Gu-HCl, 30% isopropanol, 0.1 M acetato de sodio y 0.01 de solución sódica). Posteriormente, se ajustó el pH a 4 (no menor) con HCl (37%). Los tubos se cubrieron con papel aluminio y se rotaron durante 1 h a temperatura ambiente (RT). Las muestras se centrifugaron a 2000 g por 5 min y se desecharon los sobrenadantes.

La pastilla se re-suspendió en el fondo con 1 ml de etanol (80%) y se centrifugó a 2000 g por 2 min, el sobrenadante se desechó. Se repitió el último paso, y antes de centrifugar se transfirió a un tubo nuevo. Se centrifugó por 10000 g durante 2 min y el etanol se removió lo más que se pudo sin tocar la pastilla del fondo. La pastilla se secó durante 15 min.

La pastilla de sílice se re-suspendió por completo en 60 μ l de amortiguador EBT y se incubó a 37 °C por 15 min. Las muestras se centrifugaron a 13000 g por 1 min y el sobrenadante (con el DNA) se transfirió a un tubo nuevo.

4.2.4 Cuantificación de DNA por fluorometría

Se midió concentración de DNA en fluorómetro (Qubit 3.0 Invitrogen), ya que su límite de cuantificación (0.01 ng/ μ l) permite cuantificar muestras antiguas donde se presupone tendrán concentraciones bajas de DNA. Se prepararon soluciones para calibrar con 10 μ l de cada reactivo estándar en un volumen total de 200 μ l, aforado con amortiguador de disolución. Para la determinación de la concentración de DNA en las muestras, se prepararon soluciones con 1 μ l de la muestra aforadas a 200 μ l con amortiguador de disolución. Finalmente, las muestras se agitaron en mezclador de vórtice e incubaron 2 min a temperatura ambiente. Se utilizó programa de alta sensibilidad (dsDNA) de Qubit 3.0.

4.2.5 Diseño y obtención de cebadores de PCR para genes de VARV

Se buscó cumplir con los siguientes parámetros:

- Minimizar la formación de dímeros o auto hibridación.
- Secuencia específicos para genes de VARV.
- Cebadores con temperatura de hibridación teórica similar.
- Contenido de GC de 40-60%.

Se sintetizaron los cebadores que se indican en el Cuadro 5 con empresa comercial Oligo T4.

Una vez obtenidos, los cebadores se hidrataron en solución amortiguadora (10 mM Tris-HCl, pH 8.0-8.5) y se prepararon alícuotas de 10 uM.

El nombre del cebador corresponde al nombre del gen de VARV donde se encuentra el segmento a amplificar, la última letra indica si el cebador es en sentido 5' (F) o 3' (R).

Cuadro 5. Lista de cebadores para detección de VARV mediante PCR tiempo final.

Nombre de cebador	Secuencia (5'->3')	Tamaño (bp)	% GC	Producto (bp)	T Alineamiento (°C)
D8LF	CTTTCACCGCA	22	45	70	66
	TAGTTGTTTGC				
D8LR	CGCCCAGATTG	20	55	70	70
	GTCGGAAAT				
E3LF	TATGGTGTACA	20	55	71	64
	GCTCCGACG				
E3LR	TAGCATCCGGC	19	58	67	67
	TTATCCGC				
E9LF	AGAAAATGCTC	20	50	70	68
	CGTCGCAGA				
E9LR	TAGTGGCGTTT	22	45	67	67
	TGTATGTTGCG				
B7RF	TGGAATCCCGT	20	50	72	66
	ACTCCCAAT				
B7RR	ATCATCGGGAC	20	55	69	69
	CATCCTCCA				

4.2.6 Síntesis de oligonucleótidos

Debido a la dificultad de conseguir controles positivos para VARV en muestras antiguas, se usó producto de PCR *in silico* (70 bp) para la síntesis de controles positivos en base a genes de referencia de VARV. Se sintetizaron 10 bp extras al producto de PCR (Cuadro 6) en cada extremo y se tomó en cuenta que el DNA puede estar fragmentado, por lo que se diseñaron amplicones cortos.

Se hidrataron oligonucleótidos en solución amortiguada (10 mM Tris-HCl, pH 8.0-8.5) y se diluyeron hasta conseguir alícuotas a una concentración <1 ng/μl.

Cuadro 6. Lista de secuencias de oligonucleótidos que se utilizaron como controles positivos para detección de VARV mediante PCR. El nombre de los oligonucleótidos corresponde al gen de VARV donde se encuentra la diana de PCR in silico.

Gen de VARV	Producto + 10 bp en extremos 5' y 3'
D8L	CCTTATTAAC TTT CACCGCATAGTTGTTTGCAAATACGGTTAATCCTT TGACCTCGTCGATTTCCGACCAATCTGGGCGTATAATGAAT
E3L	AACGTAGTGCTATGGTGTACAGCTCCGACGATATTCCTCCTCGTTG GTTTATGACAACTGAGGCGGATAAGCCGGATGCTATGACTATGG
E9L	TATTAAGATAGAAAATGCTCCGTCGCAGATATGTGGTTAATAGAAG AGCCAAAGAAACGCAACATACAAAACGCCACTATGGATGAATT
B7R	CGACGGTAAATGGAATCCCGTACTCCCAATATGTATACGATCTAAC GAAGAATTTGATCCAGTGGAGGATGGTCCCGATGATGAGACAGATC

4.2.7 PCR para genes VARV

Se mezcló para reacciones de 15 μl:

- 7.5 μl 2X KAPA HiFi Hot Start Uracil + Ready Mix
- 0.45 μl cebador 5' (10 μM)
- 0.45 μl cebador 3' (10 μM)
- Aforado a 15 μl con H₂O

Se usó una concentración de DNA lo más cercano a 10 ng para muestras arqueológicas, siempre y cuando no representara más de 6 μl de muestra por

51

reacción; esto para evitar terminar con el extracto de DNA, que es irremplazable. Para los controles positivos se añadió 1 ng de DNA (Cuadro 6) a la reacción de PCR.

Se transfirió un volumen de mezcla de PCR a cada tubo de PCR, se mezcló y centrifugó brevemente. Se emplearon las condiciones de amplificación específicas mostradas en el Cuadro 7.

Cuadro 7. Condiciones de amplificación de aDNA mediante PCR.

Paso	Temperatura (°C)	Duración	Número ciclos
Desnaturalización inicial	95	3 min	1
Desnaturalización	98	20 s	
Hibridación	Varía según par de cebador (Ver Cuadro 5)	15 s	35
Extensión	72	30 s	
Extensión final	72	30 s	1

4.2.8 Construcción de librería para secuenciación

Se usó el paquete comercial NEBNext (New England Biolabs Inc.) y se siguió el protocolo del proveedor con algunas modificaciones como se describen a continuación.

4.2.8.1 Generación de extremos romos

Se realizó la siguiente mezcla maestra en tubos de PCR para reacciones con 50 µl totales:

- 30 μ l Extracto de DNA
- 12.5 μ l H₂O
- 5.0 μ l NEBNext 10x amortiguador de reparación de extremos
- 2.5 μ l Mezcla NEBNext de enzima para reparación de extremos

Se incubó la mezcla durante 20 min a 12 °C, seguido de una incubación de 15 min a 37 °C y se purificó en columna Min Elute (Qiagen). Se usaron 250 μ l de amortiguador PB (5x) para mezclar reacción maestra. Se lavó con 750 μ l de PE y se eluyó en 32 μ l EB. La columna se incubó 15 min a 37 °C y el DNA se eluyó.

Se consideró un control negativo que se secuenció a la par que las muestras, donde en lugar de extracto de DNA se usó H₂O de la misma alícuota usada con las muestras de interés.

4.2.8.2 Ligación de adaptadores a librería

Se mezclaron los componentes siguientes en tubos de PCR separados, para obtener reacciones de 50 μ l totales.

- 30 μ l DNA con extremos romos
- 10 μ l Amortiguador de ligación NEBNext 5x (concentración final 1x)
- 5 μ l Mezcla de adaptador 2.5 μ M (Adaptador romo) (concentración final 0.25 μ M)
- 5 μ l Ligasa T4

Se mezcló con pipeta y se agitó usando un mezclador de vórtice, se centrifugó unos segundos e incubó durante 20 min a 20 °C.

Se purificó el DNA con columna QiaQuick (Qiagen); en donde, la reacción de ligación se mezcló en 250 μ l de amortiguador PB 5x, se añadieron 750 μ l de

amortiguador PE y se incubó columna durante 37° C por 15 min; se eluyó con 42 µl de amortiguador EB y se colectó el DNA.

4.2.8.3 Reparación de librerías

Se mezclaron los siguientes componentes en tubos de PCR para reacciones de 50 µl totales:

- 42 µl DNA con adaptadores ligados
- 5 µl Amortiguador de reacción para llenado de adaptadores NEBNext
- 3 µl DNA polimerasa Bst

Se incubó la mezcla a 65 °C durante 20 min seguido de una incubación de 80 °C durante 20 min y se guardaron las muestras en congelación.

4.2.8.4 Cuantificación de librerías con qPCR

Se realizó la mezcla maestra con los siguientes componentes para reacciones de 20 µl totales:

- 10 µl KAPA SYBR FAST MIX
- 0.4 µl Cebador sentido (10 µM) PE 1.0
- 0.4 µl Cebador anti sentido (10 µM) (índice variable)
- 0.4 µl ROX 50X (alto)
- 7.8 µl H₂O
- 1 µl DNA con adaptadores

Se tomó en consideración agregar un control negativo que se secuenció a la par que las muestras, donde se agregó H₂O en lugar de DNA ligado a adaptadores.

Para esta reacción se emplearon las condiciones mostradas en el Cuadro 8.

Cuadro 8. Condiciones de amplificación de librerías de secuenciación mediante qPCR.

Paso	Temperatura	Duración	Ciclos
Activación de enzima	95 °C	10 min	1
Desnaturalización	95 °C	15 s	1
Extensión	60 °C	20 s	40
	95 °C	15 s	1
Curva de fusión	60 °C	1 min	1
	95 °C	15 s	1

4.2.8.5 Indexar librerías de DNA

Se usó la enzima KAPA HiFi Hot Start Uracil y se preparó una mezcla maestra para reacciones de 15 μ l:

- 1 μ l BSA
- 12.5 μ l 2X KAPA HiFi HotStart Uracil + ReadyMix
- 0.75 μ l Cebador sentido PE (10 μ M)
- 0.75 μ l Cebador anti-sentido (índice) (10 μ M)
- 10 μ l Librería de DNA

Se amplificó DNA con índice utilizando las condiciones mostradas en el Cuadro 9.

Los productos de PCR se purificaron con columnas QIAquick (Qiagen) y se eluyeron en 32 μ l de amortiguador EB. Se incubaron por 10 min a 37 °C, y se centrifugaron brevemente.

Cuadro 9. Condiciones de amplificación de librerías de secuenciación mediante PCR punto final.

Paso	Temperatura	Tiempo	Ciclos
Desnaturalización inicial	98 °C	45 s	1
Desnaturalización	98 °C	15 s	
Hibridación	63 °C	30 s	
Extensión	72 °C	30 s	11-13
Extensión final	72 °C	1 min	1
Guardar	4 °C	---	1

4.2.9 Secuenciación

Se recurrió a los servicios del “Laboratorio Nacional de Genómica para la Biodiversidad” (LANGEBIO) para secuenciar las librerías en Illumina NextSeq 500, con formato 2x75 y rendimiento medio (lecturas en ambos sentidos, de tamaño aproximado a 75 bp), para obtener ~110,000,000 lecturas en cada sentido.

4.2.10 Identificación de patógenos virales humanos

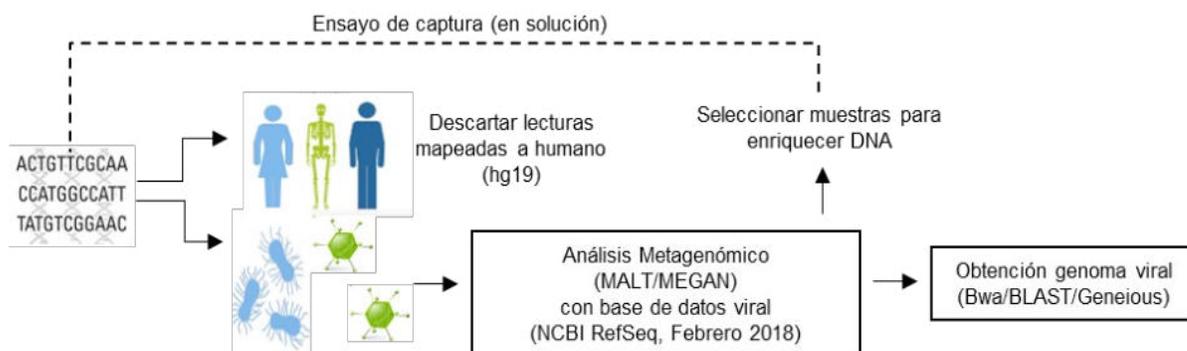
Las lecturas de secuenciación fueron mapeadas (BWA 0.7.13) a genoma humano (hg19) y solo se utilizaron lecturas no mapeadas (Anexo I: comando 1).

Se descargó una base de datos con 7,530 genomas virales contenidos en NCBI RefSeq en febrero de 2018 (Anexo I: comando 2).

La asignación taxonómica se realizó con el programa MALT 0.4.0 que consiste en dos sub-programas, “malt-build” y “malt-run”, el primero se usó para formatear base

de datos y crear un índice de semillas poco frecuentes (Anexo I: comando 3). Para alinear lecturas respecto a base de datos viral se corrió programa “malt-run”, usando secuencias en formato fastq, con el algoritmo BLASTn de manera semi-global y filtrando con un porcentaje de identidad mayor al 90%, generando un archivo con terminación .rma (Anexo 1: comando 4)

A partir del archivo .rma generado en MALT se realizó asignación taxonómica viral en MEGAN modificando el algoritmo LCA como hizo Maixner y col. (2014) para análisis de metagenomas antiguos. De forma que solo se retuvieron los alineamientos con valores de “bitscore” mayores a 50 (“min-score”), que tuvieran mínimo 10% del mejor valor asignado en la muestra (“top-percent”), con mas de 95% de identidad a referencia (“min-percent-identity”) y que taxón estuvieran presentes mínimo 0.001% de las lecturas totales asignadas (“min-support-percent”),



los demás parámetros se conservaron con valores automáticos. El procedimiento anterior se resume en Figura 11.

Figura 11. Metodología para identificación de patógenos virales humanos (A partir de Key y col., 2017; Mühlemann y col., 2018).

Las asignaciones taxonómicas fueron normalizadas automáticamente en MEGAN (número de lecturas asignadas dividido entre tamaño total de muestra, y multiplicado por tamaño total de la muestra mas pequeña), para asignarles un valor de z respecto a abundancia a nivel de genero viral. De esta forma se seleccionaron

muestras con mayor abundancia de secuencias virales para realizar un ensayo de captura.

4.2.11 Diseño de sondas de captura

Para el diseño de sondas de genes de VARV se usaron todos los genomas completos de VARV disponibles en NCBI, incluidos genomas antiguos (Anexo II). Usando como templatado la secuencia de referencia de genes de VACV (Cuadro 10) se encontró con BLAST las coordenadas de los genes de interés (replicación, estructurales o reguladores de sistema inmune, mostrados en Cuadro 10) en los genomas de VARV descargadas. Luego, se crearon archivos FASTA para cada gen de cada cepa, cuidando que todas las secuencias comenzaran con codón de metionina y terminaran en codón de paro. Adicionalmente, se agregaron 40 bp previos al codón de inicio, y 40 bp después de codón de paro de todos los genes de interés respecto a genoma de referencia de VARV (NC_001611.1) para asegurar cobertura homogénea de gen completo (desde codón de inicio hasta codón de paro) en caso de muestras positivas a VARV.

Cuadro 10. Lista de genes seleccionados para el diseño de sondas de captura. El nombre del gen es respecto a VACV.

Función	Gene
Replicación (poco variables)	J6R, A24R, A29L, E4L, A50R, A5R, D7R, H4L, E9L.
Estructurales (variables, posibles dianas de anticuerpos)	A27L, A25, D8L, H3L, L1R, A33R, B5R, A16L.
Reguladores de sistema inmune (interacción hospedero-virus)	B18R, A46R, B15R, K7R, N1L, M2L, E3L, H1L, B8R, D9R, D10, K3L.

Se alinearon las secuencias encontradas en los 50 genomas disponibles y se obtuvo secuencia consenso. Se consideró secuencia consenso cuando una base estuvo presente mínimo 26/50 veces, se revisó manualmente el archivo generado para conservar solo las secuencias más frecuentes en los 50 genomas usados. Se generó un archivo FASTA con todas las secuencias consenso cuidando que el tamaño no rebasara el 20 % del genoma de referencia de VARV (NC_001611.1). De lo contrario se necesitaría autorización especial de la Organización Mundial de la Salud, pues actualmente el virus está extinto y su manipulación requiere condiciones de bioseguridad 4 (OMS, 1999).

Adicional a los genes de interés de VARV (Cuadro 10), se incluyeron genes o genomas de otros patógenos virales de DNA con importancia clínica para humanos, mostrados en Cuadro 11 y 12. Se incluyó el genoma de HBV (consenso) y B19V, pues son las únicas especies virales que se han podido recuperar a la fecha a partir de restos arqueológicos óseos (Ross y col., 2017; Krause-Kyora y col., 2018; Mühleman y col., 2018).

Para HBV se realizó un alineamiento de genoma completo de secuencias pertenecientes a genotipos A, B, C, D, E, F, G, H y secuencia antigua RISE254 obtenida por Mühlemann y col. (2018). Se realizó secuencia consenso en Jalview 2 (Waterhouse y col., 2009) y se agregó a archivo FASTA para diseño de sondas.

En el caso de parvovirus humanos se incluyeron genomas completos de diferentes especies virales, incluidas B19V y B19V-V9 (Cuadro 11 y 12).

Para el resto de los virus se incluyeron genomas completos de las especies representativas de cada familia, y se seleccionaron genes importantes (mostrados en Anexo III) para 3 especies virales pertenecientes a familia *Herpesviridae* (Cuadro 11 y 12). Con base al archivo Fasta con las secuencias de interés (Cuadro 11 y 12), se diseñaron y sintetizaron sondas magnéticas con la empresa comercial

Arbor BioSciences. Se consideró una longitud de 80 nt por sonda, una diferencia de 20 nt entre cada una, y que no tuvieran similitud con el genoma humano (hg38) mediante BLAST. El diseño resultó en un total de 19147 sondas de ssRNA asociadas a partículas magnéticas.

Familia de virus	Virus	Secuencia	Referencia	Tamaño (kp)
<i>Poxviridae</i>	VARV	30 genes	Consenso	36.44
<i>Hepadnaviridae</i>	HBV	Genoma	Consenso	3.26
	Virus adenoasociado 2	Genoma	NC_001401.2	4.68
<i>Parvoviridae</i>	Bocaviruses	Genoma	Consenso	5.54
	B19V	Genoma	NC_000883.2	5.6
	B19V V9	Genoma	NC_004295.1	5.03
	Herpes virus simple 1	6 genes	NC_001806.1	17.54
<i>Herpesviridae</i>	Cytomegalovirus	6 genes	NC_006273.1	14.85
	Epstein bar virus	7 genes	NC_009333.1	27.03
<i>Retroviridae</i>	Lentivirus	Genoma	Consenso	11.05
	Foamy humano	Genoma	KX087159.1	11.95
<i>Papillomaviridae</i>	Papiloma humano 16	Genoma	NC_001526.4	7.91
	Papiloma humano 5	Genoma	NC_001531.1	7.75
	Papiloma humano 4	Genoma	NC_001457.1	7.35

	Papiloma humano 1	Genoma	NC_001356.1	7.81
	Papiloma humano 41	Genoma	NC_001354.1	7.61

Cuadro 11. Lista de especies virales incluidas en diseño de captura.

Familia de virus	Virus	Secuencia	Referencia	Tamaño (kp)
<i>Polyomaviridae</i>	Poliomavirus humano 9	Genoma	NC_015150.1	5.03
	Poliomavirus JC	Genoma	NC_001699.1	5.13
	Poliomavirus BK	Genoma	NC_001538.1	5.15
	Poliomavirus humano 6	Genoma	NC_014406.1	4.93
<i>Circoviridae</i>	Ciclovirus 1	Genoma	KR902499.1	16.14
	Ciclovirus PK5034	Genoma	GQ404845.1	
	Ciclovirus NG12	Genoma	GQ404854.1	
	Ciclovirus NG14	Genoma	GQ404855.1	
	Ciclovirus VN	Genoma	KF031466.1	
	Ciclovirus VS	Genoma	KC771281.1	
	Ciclovirus PK5222	Genoma	GQ404846.1	
	Ciclovirus 10	Genoma	KF726984.2	
	Ciclovirus SL	Genoma	KJ831064.1	
Longitud total				217.77

--	--

Cuadro 12. Lista de especies virales incluidas en diseño de captura (continuación).

4.2.12 Enriquecimiento de librerías de secuenciación

Se descongelaron y homogeneizaron reactivos incluidos en paquete comercial proveído por Arbor Biosciences, y se realizó mezcla (HYB) con los volúmenes mostrados a continuación por reacción de captura de 20 μ l.

- 9.25 μ l HybN
- 3.5 μ l HybD
- 0.5 μ l HybS
- 1.25 μ l HybR
- 5.5 μ l Sondas ssRNA

Se incubó la mezcla de hibridación (HYB) durante 10 min a 60 °C, con eventuales centrifugaciones para evitar evaporación. Se dejó reposar durante 5 min a temperatura ambiente.

Se concentraron librerías de secuenciación con índice, mediante centrifugación a temperatura ambiente, durante aproximadamente 10 min.

Se realizó mezcla maestra de bloqueadores de librerías por cada reacción de captura como se muestra a continuación.

- 0.5 μ l Bloqueador A
- 2.5 μ l Bloqueador C
- 2.5 μ l Bloqueador O

Se tomaron 5 μ l de la mezcla y se les adicionó cerca de 6 μ l de librería previamente concentrada, esta mezcla se nombró LIB.

Dentro de termociclador, se colocaron las reacciones LIB durante 5 min a 95 °C, posteriormente y sin retirar las reacciones LIB, se colocaron las reacciones HYB a 60 °C durante 5 min. Una vez terminado el tiempo, se transfirieron 18 μ l de cada reacción HYB a cada reacción LIB. La mezcla final (HYB/LIB) se colocó en el termociclador para hibridación a 60 °C durante 48 h.

4.2.12.1 Preparación de perlas magnéticas

Una vez pasadas 46 h de hibridación, se preparó amortiguador de lavado para 17 reacciones. Se mezclaron 100 μ l HYB S, 9.9 mL H₂O y 2.5 ml de amortiguador de lavado (parte de paquete comercial), se mezcló vigorosamente en agitador de vórtice y se calentó a 60 °C durante 30 min antes de usarse.

Se realizaron alícuotas de 30 μ l de perlas magnéticas, y se sedimentaron con magneto comercial (DynaMag-2) durante 1-2 min, retirando el sobrenadante. Se agregaron 200 μ l de amortiguador de unión a cada alícuota, y se mezcló con agitador en vórtice para re-suspender, con breve centrifugación posterior. Se recolectó nuevamente el sedimento con magneto, y se desechó el sobrenadante. Se realizó un total de tres lavados, repitiendo pasos anteriores, y se re-suspendió finalmente en 70 μ l de amortiguador de unión.

4.2.12.2 Captura y re-suspensión de sondas asociadas a perlas magnéticas.

Se calentaron alícuotas de perlas magnéticas (previamente lavadas) a 60 °C durante mas de 2 min, y se colocó una reacción de hibridación (HYB/LIB) por alícuota de perlas magnéticas, se mezcló con pipeta suavemente. Se colocó a 60 °C durante 5 min, con agitación y breve centrifugación cada 2.5 min.

Se sedimentaron perlas con magneto hasta que la solución fue transparente y se desechó el sobrenadante. Se agregaron 375 µl de amortiguador de lavado previamente preparado, y se mezcló brevemente con agitador en vórtice. Se centrifugó brevemente y se incubó 5 min a 60 °C, con agitación y breve centrifugación cada 2.5 min. Se repitió lavado hasta tener tres lavados totales, y en el último se removió todo el líquido posible sin tocar el sedimento de perlas magnéticas.

Se preparó amortiguador de re-suspensión con 10 mM Tris-Cl, 0.05 % Tween-20 y pH 8-8.5. Se agregaron 30 µl de este amortiguador a cada alícuota de perlas magnéticas unidas a sondas. Se incubó durante 5 min a 95 °C, e inmediatamente después se sedimentaron perlas con magneto, se conservó solo el sobrenadante.

4.2.12.3 Amplificación de librerías de secuenciación enriquecidas

Se realizó qPCR para determinar el número de ciclos ideales para amplificar librerías capturadas, se utilizaron cebadores para adaptadores de librerías y se consideró un control negativo donde en lugar de librerías se usó H₂O. Se preparó mezcla como se muestra a continuación para reacciones de 20 µl .

- 10 µl KAPA Sybr Fast Mix
- 0.4 µl Cebador sentido 10 µM (PE)
- 0.4 µl Cebador anti-sentido 10 µM (PE)
- 0.4 µl ROX 50X (high)
- 7.8 µl H₂O
- 1 µl Librería capturada

Se utilizaron los siguientes parámetros de amplificación.

Cuadro 13: Condiciones de amplificación de librerías de secuenciación mediante qPCR.

Paso	Temperatura (°C)	Duración	Número ciclos
Desnaturalización inicial	98	3 min	1
Desnaturalización	98	20 s	
Hibridación	60	30 s	40
Extensión	72	30 s	
Extensión Final	72	5 min	1

Una vez obtenidas las curvas de amplificación se determinó el número de ciclos en que la librería capturada alcanzó fase exponencial.

Se realizó PCR de tiempo final para amplificación de librerías capturadas como se muestra a continuación para reacciones de 20 μ l.

- 10 μ l 2x Phusion U Hot Master Mix
- 1 μ l Cebador sentido 10 μ M (PE)
- 1 μ l Cebador anti-sentido 10 μ M (PE)
- 8 μ l Librería capturada

Usando los mismos parámetros mostrados en Cuadro 13, con excepción del número de ciclos donde se utilizaron los mostrados en Cuadro 14, que corresponden a los ciclos en que cada librería capturada amplificó en qPCR.

Se determinó la concentración de las librerías capturadas y amplificadas por fluorometría en Qubit 3.0 y por electroforesis capilar en Bioanalyzer (Agilent), para su posterior secuenciación (como en 4.2.9).

Cuadro 14. Número de ciclos usados para amplificación de librerías capturadas. Algunas librerías fueron capturadas dos veces, por lo que se indican dos números de ciclos

Muestra	Número de ciclos
HSJNC4	18, 20
HSJNC81	18, 20
HSJNC177	17
HSJN194	20
HSJN240	20, 20
HSJN275	20
COYC4	17, 20
COYC9	18

4.2.13 Reconstrucción
antiguos

de genomas virales

Para reconstrucción de genomas de B19V se siguió metodología propuesta por Mühlemann y col. (2018), aumentando astringencia de filtrado de alineamientos, mostrado en Figura 12.

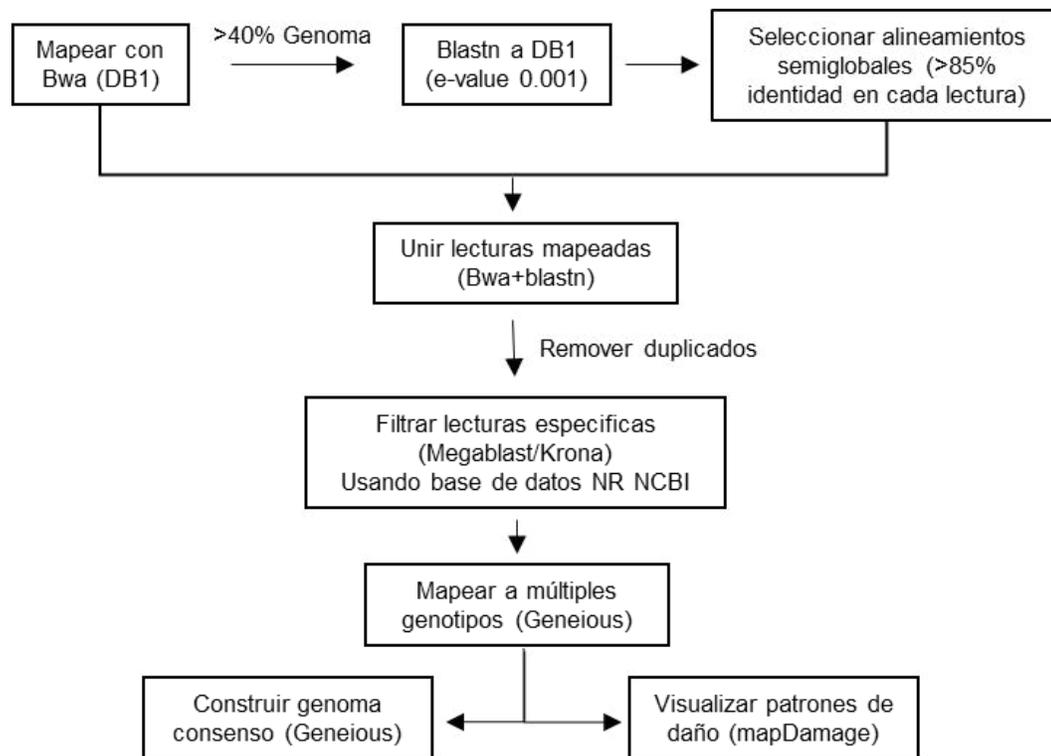


Figura 12. Esquema para reconstrucción de genomas antiguos de B19V. Procedimiento modificado de Mühleman y col. (2018).

Primero se mapearon competitivamente las lecturas (a archivo multifasta DB1, Cuadro 15) obtenidas mediante BWA 0.7.13 usando el algoritmo “aln”, desactivando la semilla (-l 1500) con una calidad de mapeo de 25. Las lecturas mapeadas se extrajeron en formato fasta (Anexo 1: comando 5).

En caso que el mapeo cubriera mas del 40% del genoma de referencia se utilizó BLAST+ 2.7.1 para alinear lecturas contra base de datos personalizada “DB1” (mostrada en Cuadro 15, para B19V) con diversos genomas de parvovirus humanos y Parvovirus bovino, esto con el fin de obtener la mayor diversidad de lecturas con similitud a alguna referencia de parvovirus (Anexo 1: comando 6).

Cuadro 15. Lista de genomas de B19V incluidos en mapeo (DB1).

Genotipo	Procedencia	Identificador	Tipo secuencia
1	América del sur	KT268312	Genoma
	Europa del norte	AY504945	Genoma
	América del norte	FJ591158	Genoma
2	América del norte	EF216869	CDS
	Europa del norte	AY064476	CDS
	Europa occidental	DQ333427	CDS
	NA	AB550331	Genoma
3	África occidental	AY582124	CDS
	Europa occidental	DQ408305	CDS
	América del norte	FJ265736	CDS
	Europa occidental	AJ249437.1	Genoma
NA	NA	NC_004295.1	Genoma RefSeq
	NA	NC_000883.2	Genoma RefSeq
	NA	NC_001540	Genoma Bovino

Posteriormente, se seleccionaron las lecturas únicas que tuvieran similitud a algún parvovirus de DB1, y que además cubrieran >85% de la lectura de secuenciación, (Anexo 1: comando 7) para desechar alineamientos cortos locales.

Se unieron lecturas mapeadas por BLASTn y por BWA. Debido a que se amplificó librería mediante PCR es de esperarse alta presencia de duplicados, por lo que se removieron estas secuencias (Anexo 1: comando 8).

Las lecturas restantes se alinearon con MegaBLAST contra la base de datos no redundante (NR) de NCBI, usando parámetros de herramienta en línea, descargados en octubre de 2018 (Anexo 1: comando 9).

El archivo generado por MegaBLAST se visualizó con Krona 2.7 usando parámetros automáticos (Anexo 1: comando 10).

Solo las lecturas específicas a Parvovirus humano, con un “e value” menor a 0.001, se consideraron para ensamblaje en GeneiousPrime 2019.0.4, donde se especificó sensibilidad mediana/rápida y hasta 5 veces de iteración, usando genotipos 1, 2 y 3 independientemente (Mühlemann y col., 2018). Se conservaron ensamblajes con mayor número de bases cubiertas, incluso si tuviesen mayor similitud a otro genotipo.

Para las muestras positivas a HBV, se realizó mapeo competitivo respecto a archivo fasta con múltiples secuencias de HBV (Anexo V) usando BWA, como se menciona previamente. Se seleccionó la referencia con más lecturas mapeadas para mapear únicamente a esta secuencia. El genoma consenso se obtuvo de manera manual con “samtools” y “vcftools” (Anexo 1: comando 11).

Las lecturas resultantes fueron alineadas con MegaBLAST y visualizadas con Krona, como se menciona previamente.

La visualización de los patrones de daño de las lecturas mapeadas a referencia viral (B19V o HBV) se realizó en mapDamage 2.0 (Anexo 1: comando 12) a partir de archivo bam.

4.2.14 Análisis filogenéticos de B19V.

Una vez obtenido un genoma de B19V, se determinó región codificante de 5080 bp (CDS) respecto a 112 genomas descargadas de NCBI (Anexo IV) pertenecientes a 71

los 3 genotipos reportados de B19V. Se alinearon cuidando que todas las secuencias empezaran con codón de inicio y terminaran con codón de paro, dentro de este conjunto de genomas parciales se incluyeron 10 de los genomas antiguos mejor cubiertos reportados por Mühlemann y col. (2018), mostrados en Cuadro 16, de los cuales 5 secuencias son basales para el genotipo 1 y 5 para el genotipo 2; sin embargo, no existen genotipos 3 antiguos reportados a la fecha (Mühleman y col., 2018).

Este alineamiento se analizó mediante el programa jModelTest 2.1.10, para evaluar 88 modelos evolutivos, y posteriormente se hizo prueba de criterio de información Akaike corregida (AICc) y bayesiana (BIC), dando cien por ciento de confiabilidad en uso de modelos HKY+G y GTR+G.

Se utilizó GeneiousPrime 2019.0.4 con un método de distancias pareadas usando modelo HKT para calcular un fenograma con 1000 repeticiones, y filtrando topologías presentes >70%. El árbol generado fue visualizado y formateado en FigTree 1.4.3.

Posteriormente se construyó un árbol filogenético de máxima verosimilitud, haciendo uso del modelo GTRGamma con 1000 replicas, mediante el programa RAxML 8.2.10, (Anexo 1: comando 13). El árbol generado se exportó en formato Newick, se visualizó y editó en FigTree 1.4.3.

Con el fin de calcular fecha de divergencia de MRCA, se utilizó el paquete BEAST 2.5.1, que incluye los programas mencionados a continuación. En el programa “BEAUti” se establecieron parámetros para estimar dos árboles, uno con reloj molecular estricto y otro relajado. Para ambos se especificó un modelo GTRGamma, con una tasa de mutación a priori $1 \times 10^{-4} - 1 \times 10^{-9}$ (Mühleman y col., 2018) y una cadena de Markov de 50 y 150 millones, respectivamente, con muestreo cada 5000 pasos. Los archivos xml generados se analizaron con BEAST,

y usando Tracer 1.6 se corroboró que tuviera un tamaño de muestreo efectivo (ESS) mayor a 200. Utilizando TreeAnnotator 1.8.4 se estimó un árbol con la máxima credibilidad de clados eliminando el 25% de los primeros arboles muestreados, y conservando solo las secuencias con mas de 80% de densidad posterior. El árbol generado se exportó en formato Newick, se visualizo y editó en FigTree 1.4.3.

Cuadro 16. Genomas antiguos de B19V incluidos en análisis filogenéticos.
(Mühleman y col., 2018).

Genotipo	Identificador	Fecha (años antes de 2018)	Lugar de obtención	Cobertura
1	DA251-user	6863	Asia oriental (Rusia)	82.8 % (1.3x)
	DA336	4147	Asia oriental (Rusia)	63.9 % (1.3x)
	DA337n	3968	Asia oriental (Rusia)	98.4 % (13.5x)
	DA66	1519	Asia central (Kirguistán)	99.7 % (7.8x)
	NEO105	545	Asia oriental (Rusia)	84.5 % (2.4x)
2	RISE569	1305	Europa oriental (Republica Checa)	84.1 % (3.4x)
	VK143	1078	Europa del norte (Reino Unido)	67.3 % (1.6x)
	VK154	1000	Europa oriental (Polonia)	71.4 % (1.6x)
	VK477	1000	Europa del norte (Suecia)	87.7% (3.1x)
	VK6	1000	América del norte (Groenlandia)	82.5 % (4x)

4.2.15 Análisis filogenéticos de HBV.

Para el análisis filogenético de HBV se utilizó el genoma completo de 124 cepas (basándose en el árbol reportado por Mühlemann y col., 2018), incluidas secuencias pertenecientes a los genotipos humanos A-H, secuencias de chimpancé, gorila, y gibón, así como todos los genomas antiguos de HBV reportados a la fecha, mostrados en Cuadro 17.

Cuadro 17. Genomas antiguos de HBV incluidos en análisis filogenéticos.
Reportados por *: Kahila y col., 2012; **: Ross y col., 2018; +: Krause-Kyora y col.,
2018 y ++: Mühleman y col., 2018.

Genotipo	Identificador	Fecha (años antes de 2018)	Lugar de obtención	Cobertura %
A	DA119++	1567	Europa oriental (Eslovaquia)	99 (53x)
	DA195++	2647	Europa oriental (Hungría)	99 (29x)
	RISE386++	4188	Asia oriental (Rusia)	98 (7x)
B	DA45++	2120	Asia oriental (Mongolia)	87 (4x)
C	JN315779 *	337	Asia oriental (Corea)	ND
D	DA29++	822	Asia central (Kazajistán)	87 (5)
	DA222++	1169	Asia central (Kazajistán)	100 (89x)
	DA27++	1610	Asia central (Kazajistán)	90 (14x)
	DA51++	2299	Asia central (Kirguistán)	99 (14x)
	MG585269**	450	Sur de Europa (Italia)	100 (164x)
	Petersberg+	951	Europa occidental (Alemania)	88 (46x)
Genotipo antiguo (Primates no humanos)	RISE254++	4009	Europa oriental (Hungría)	99 (37x)
	RISE563++	4488	Europa occidental (Alemania)	100 (79x)
	RISE154++	3851	Europa oriental (Polonia)	57 (2x)
	RISE387++	4282	Asia oriental (Rusia)	87 (6x)
	Karsdorf**	5008	Europa occidental (Alemania)	97 (104x)
	Sorsum**	3221	Europa occidental (Alemania)	100 (47x)

Se alinearon 124 genomas y se revisó manualmente que tuvieran misma longitud, para su posterior análisis en jModelTest 2.1.10, donde se evaluaron 88 modelos evolutivos, y posteriormente se hizo prueba de criterio de información Akaike corregida (AICc) y bayesiana (BIC), dando cien por ciento de confiabilidad en uso de modelos GTR+G+I y HKNY+G+I.

Se utilizó GeneiousPrime 2019.0.4 con un método de distancias pareadas usando modelo HKT para calcular un fenograma con 1000 repeticiones, y filtrando topologías presentes >85%. El árbol generado fue visualizado y formateado en FigTree 1.4.3.

Posteriormente se construyó un árbol filogenético de máxima verosimilitud, haciendo uso del modelo GTR+G+I con 1000 replicas, mediante el programa RAxML 8.2.10, (Anexo 1: comando 13). El árbol generado fue exportado en formato Newick para ser visualizado y formateado en FigTree 1.4.3.

5. RESULTADOS

5.1 Obtención de muestras

A continuación, se muestran las fotografías de los cráneos de los individuos a partir de los cuales se obtuvieron genomas virales.

Las notas osteológicas relevantes al proyecto, así como las fotos mostradas a continuación (Figura 13-16) fueron realizadas y compartidas por la Dra. Julie K. Wesp; quien, además, seleccionó los individuos con morfología típica de afrodescendientes de la colección HSJN con ayuda de la Dra. Marcela Sandoval Velasco.

Para el cráneo 81 (HSJNC81) se tomó el primer molar maxilar izquierdo (Figura 13). El cráneo 81 de HSJN, presentó hiperostosis porótica (ligera a moderada) en bóveda craneal y criba orbitalia en cuencas oculares. Marcas típicas asociadas a deficiencia de hierro, posiblemente por anemia (Mansegosa y col., 2018).





Figura 13. Cráneo 81 de colección HSJN. Fotografías tomadas por Julie K. Wesp.



Figura 14. Cráneo de individuo 240 de colección HSJN. Fotografías tomadas por Julie K. Wesp.

Figura 15. Cráneo de individuo 275 de colección HSJN. Fotografías de cráneo y mandíbula tomadas por Julie K. Wesp, de cuerpo completo (derecha) tomada de la fototeca de dirección de salvamento arqueológico (Secretaría de cultura, INAH SINAFO).

Para los individuos 240 (HSJN240) y 275 (HSJN275) se muestreó el primer premolar maxilar derecho, e izquierdo, respectivamente (Figura 14 y 15).

Se muestreó el primer molar maxilar izquierdo del individuo 194 (HSJN194). Mostrado en Figura 16.

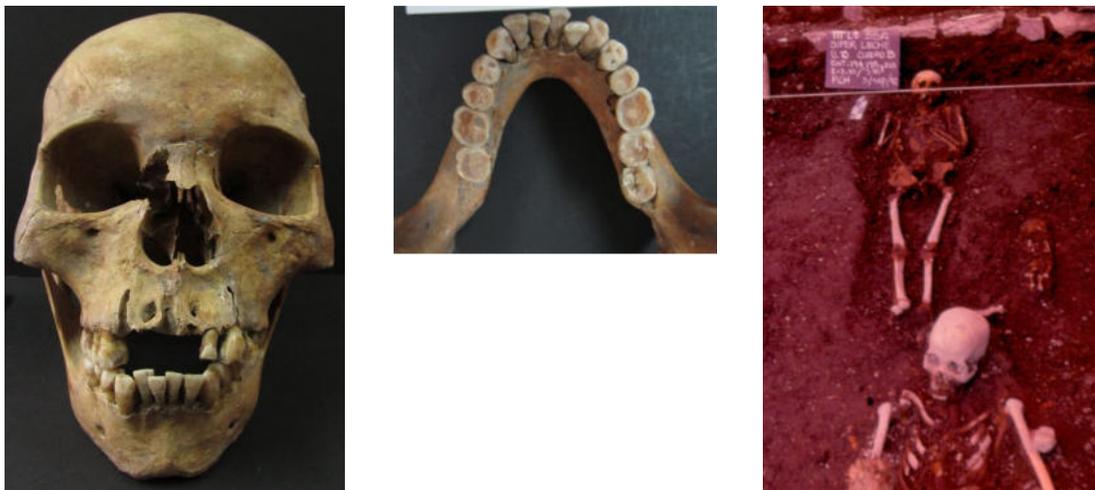


Figura 16. Cráneo de individuo 194 de colección HSJN. Fotografía de cráneo y mandíbula tomada por Julie K. West, de cuerpo completo (derecha) tomada de la fototeca de dirección de salvamento arqueológico (Secretaría de cultura, INAH SINAFO).

5.2 Extracción de DNA de dientes

Se extrajo DNA de 28 muestras arqueológicas en tres grupos independientes, cada uno con un control negativo (H₂O en lugar de sustrato arqueológico). Se cuantificó el DNA mediante fluorometría con Qubit 3; si bien, la concentración de DNA fue mínima para la mayoría de las muestras (Cuadro 18), todas las muestras presentaron una concentración mayor al límite de cuantificación (0.01 ng/μl) del equipo, exceptuando los controles negativos que tuvieron niveles indetectables de DNA, lo cual indica nula o mínima contaminación al momento de extracción (reactivos, contaminación ambiental o humana) de DNA. Los asteriscos en el Cuadro 18 muestran las 11 muestras con mayor concentración de DNA, que fueron usadas para análisis por PCR posteriores. Todas las muestras fueron consideradas para construcción de librerías de secuenciación posteriores, incluidos controles negativos.

Cuadro 18. Concentración de DNA a partir de muestras arqueológicas.

Primer grupo		Segundo grupo		Tercer grupo	
Muestra	ng/μl	Muestra	ng/μl	Muestra	ng/μl
Control 1	No Detectable	Control 2	No Detectable	Control 3	No Detectable
HSJN88*	1.510	HSJN417	0.564	COYC11	0.48
HSJN275	0.200	HSJNC59 A	0.216	COYC20 *	2.06
HSJNC177*	2.420	HSJN214	0.280	COYC4*	2.44
HSJNC143	0.588	HSJN318	0.164	COYC5	0.21
HSJN296*	1.170	HSJN71*	0.936	COYC9*	6.12
HSJN366*	0.896	HSJNC4A	0.686	COYM2*	0.946
HSJN194*	1.410	HSJN300	0.434	COYM3*	1.16

HSJN150	0.560	HSJN214	0.554
HSJNC81	0.570	HSJN197	0.500
HSJN148*	1.400		
HSJN350	0.260		
HSJN284	0.700		

5.3 Amplificación de genes de VARV mediante PCR

Se usaron extractos de DNA de las colecciones La Concepción (COY) y Hospital San José de los Naturales (HSJN), para análisis de PCR con diana a genes de VARV.

Las reacciones de PCR se prepararon en el laboratorio de DNA antiguo del LIIGH, donde nunca se ha amplificado ningún producto de DNA para evitar posibles contaminaciones.

Como control positivo se utilizó un fragmento de genes dianas (E3L, E9L, D8L, B7R) sintetizado *de novo* a partir de genoma de referencia de VARV (Cuadro 6), se respetó que oligonucleótidos tuvieran tamaño menor a 500 bp, por lo que no fue necesario la obtención de permisos adicionales para su síntesis (WHO, 2010).

Posteriormente, se añadió 1 ng/μl del DNA sintetizado *de novo* a los controles positivos en el laboratorio moderno del LIIGH. Los demás tubos permanecieron cerrados hasta terminados los ciclos de amplificación, por lo que las reacciones con aDNA y el control negativo nunca estuvieron en contacto con el control positivo, logrando eliminar la posibilidad de contaminación cruzada.

A manera de control negativo se incluyó H₂O (de la misma alícuota usada para la extracción) en lugar de DNA. Los controles negativos no presentaron amplificación en ninguno de los casos. Mientras, los controles positivos fueron amplificados en

todos los casos. Ningún extracto de DNA a partir de muestras arqueológicas presento amplificación de los genes de interés al visualizarse en gel de agarosa (Figura 17 y 18), por lo que se concluye que todas las muestras analizadas por PCR fueron negativas a amplificación de genes de VARV.

5.4 Secuenciación de librerías crudas

Se construyeron librerías de secuenciación para todas las muestras obtenidas de las colecciones COY y HSJN (incluidos controles negativos de extracción), mismas que fueron secuenciadas a baja profundidad. Se descartaron secuencias mapeadas a humano y se usaron sólo secuencias no mapeadas a genoma de referencia (hg19), el número de secuencias generadas es mostrado en Figura 19.

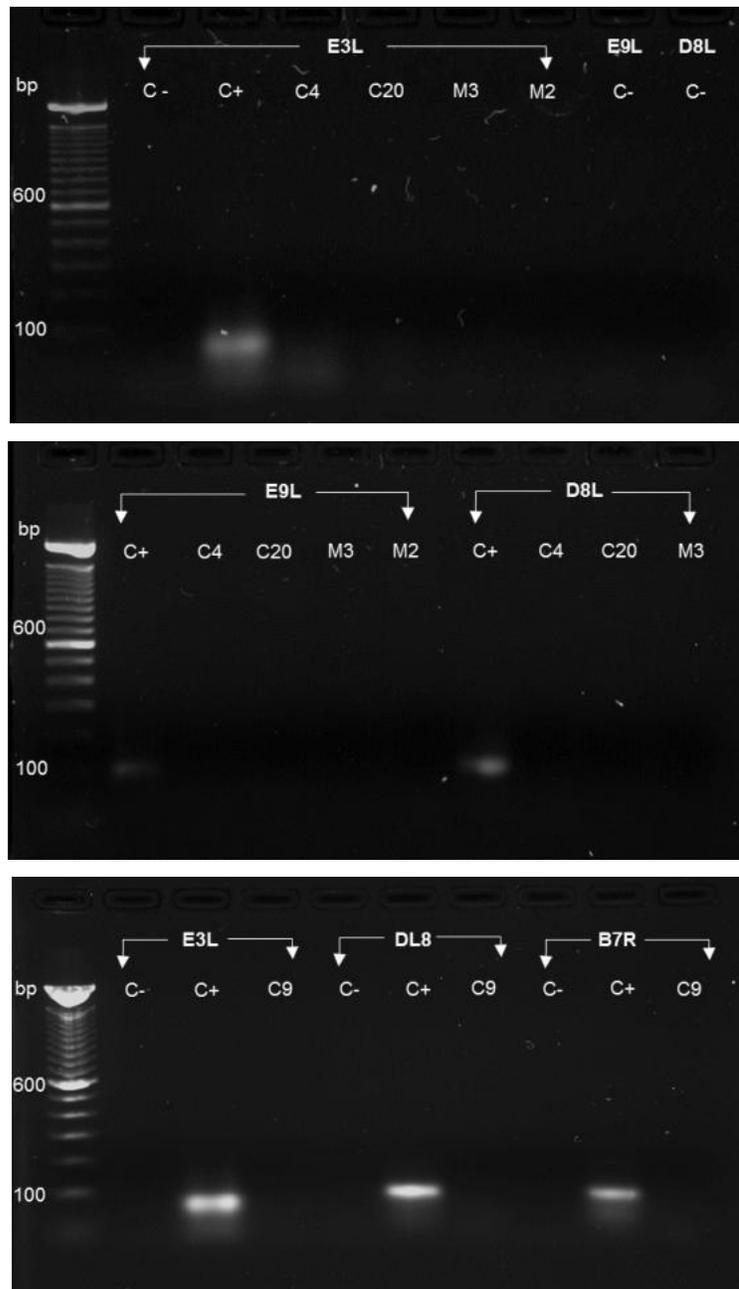


Figura 17. PCR con diana a genes de VARV usando DNA de sito COY. Producto de PCR, visualizado en gel de agarosa (4%) con SybrSafe. E3L, E9L, D8L: Genes de VARV a amplificar (90bp); C-: Control negativo de PCR; C+: Control positivo de PCR; C(M)N: Número de cráneo (mandíbula) usado.

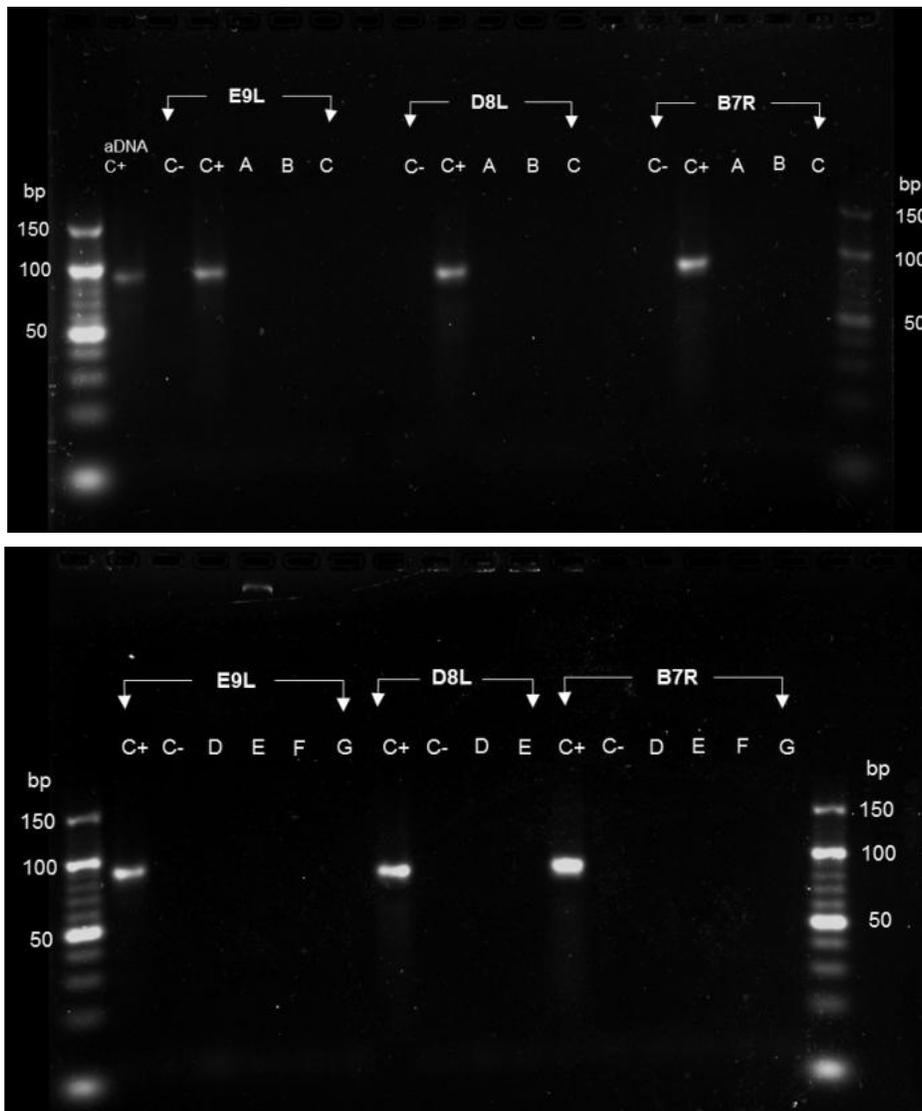
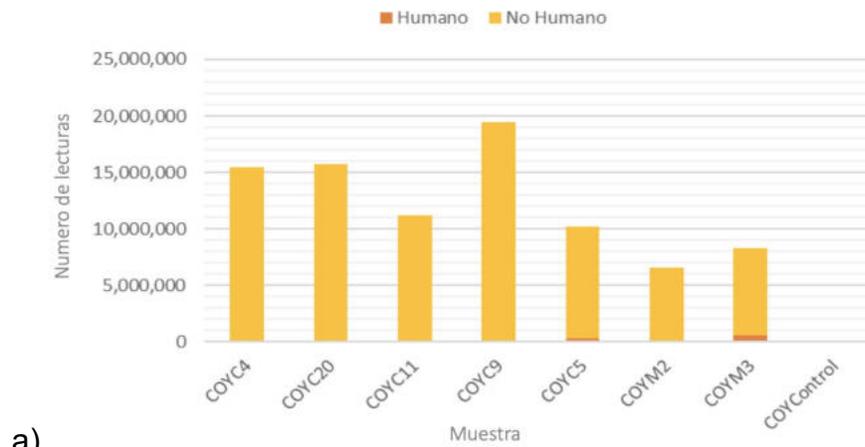
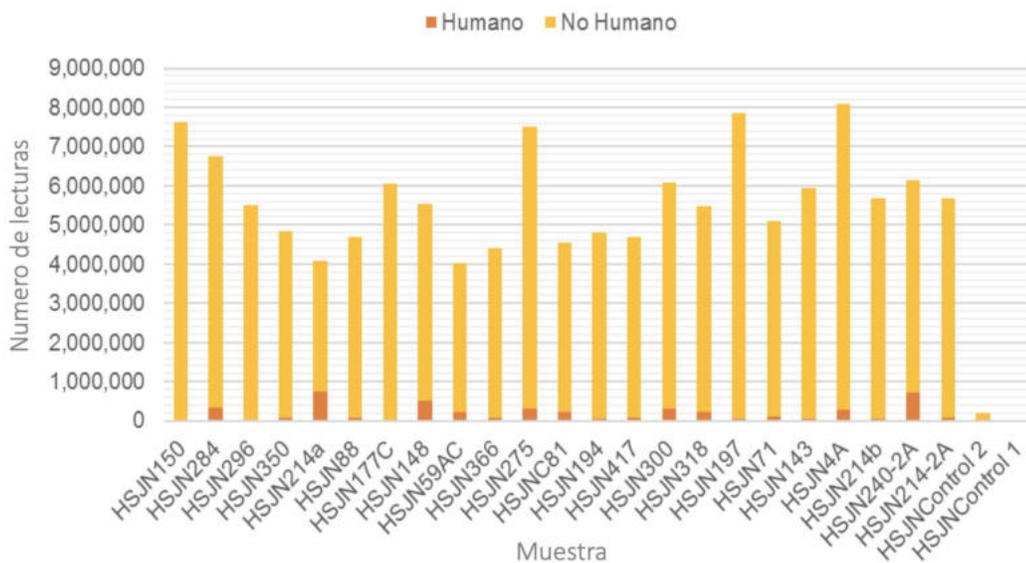


Figura 18. PCR con diana a genes de VARV usando DNA de sito HSJN. Producto de PCR, visualizado en gel de agarosa (4%) con SybrSafe. E3L, E9L, D8L: Genes de VARV a amplificar (90bp); C-: Control negativo de PCR; C+: Control positivo de PCR; A: Cráneo 177; B: Cráneo 194; C: Cráneo 148; D: Cráneo 88; E: Cráneo 296; F: Cráneo 366; G: Cráneo 71.



a)



b)

Figura 19. Número de lecturas obtenidas por secuenciación a baja profundidad. a) Colección COY; b) Colección HSJN.

5.5 Análisis metagenómico viral

Las lecturas restantes se alinearon con MALT 0.4.0 respecto a base de datos viral (NCBI RefSeq) y se analizaron en MEGAN para realizar la asignación taxonómica a

nivel de género. Las muestras de una misma colección se compararon normalizando el número de lecturas asignadas a un género viral específico. Se construyó un mapa de calor (en MEGAN) con base a diferencias en valor de z respecto a cada género viral. La comparación de los géneros virales de nuestro interés (previamente obtenidos a partir de sustratos arqueológicos) se muestran en Figura 20 y 21, para los sitios COY y HSJN, respectivamente. El gráfico comparando todos los géneros virales asignados puede observarse en Anexo VI.

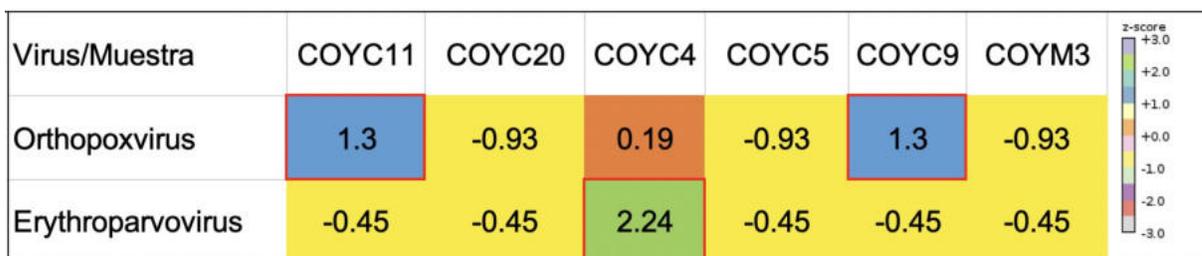


Figura 20. Mapa de calor mostrando comparación de abundancias de géneros virales en muestras del sitio COY. Las abundancias normalizadas se calcularon en MEGAN a partir de la asignación taxonómica con MALT. Las muestras resaltadas en rectángulos rojos se seleccionaron para ensayo de enriquecimiento.

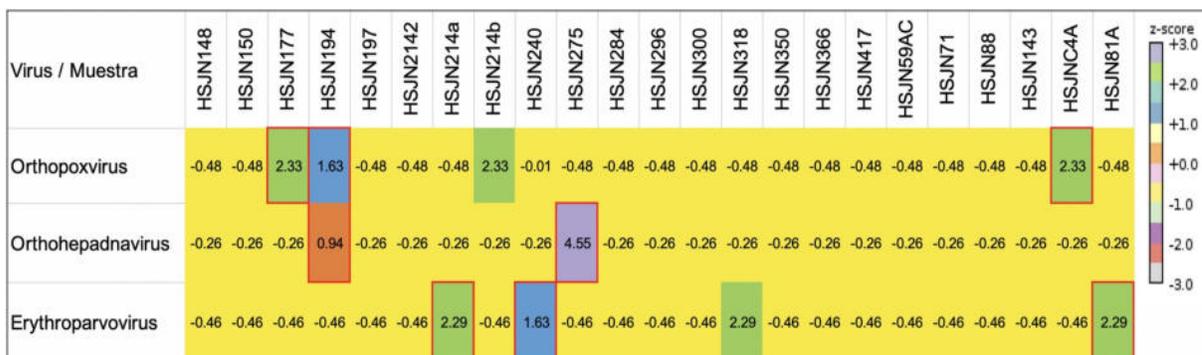


Figura 21. Mapa de calor mostrando comparación de abundancias de géneros virales en muestras del sitio HSJN. Las muestras resaltadas en rectángulos rojos se seleccionaron para ensayo de enriquecimiento.

5.5.1 Comparación de análisis metagenómico viral después de ensayo de captura.

A partir de las mismas librerías utilizadas, se realizó un ensayo de captura y se secuenció nuevamente a baja profundidad. Se realizó la asignación taxonómica (MALT 0.4.0) de la misma manera que para las lecturas provenientes de las librerías crudas (sin capturar). La cantidad de lecturas asignadas a nivel de género fue normalizada para cada individuo, se compararon lecturas asignadas a patógenos virales (B19V/EBV) con base a sus respectivos análisis metagenómicos a partir de librerías sin capturar. En Figura 22 se muestra el incremento de abundancias normalizadas del género Eritroparvovirus antes y después de ensayos de captura, respecto de su respectiva librería cruda.

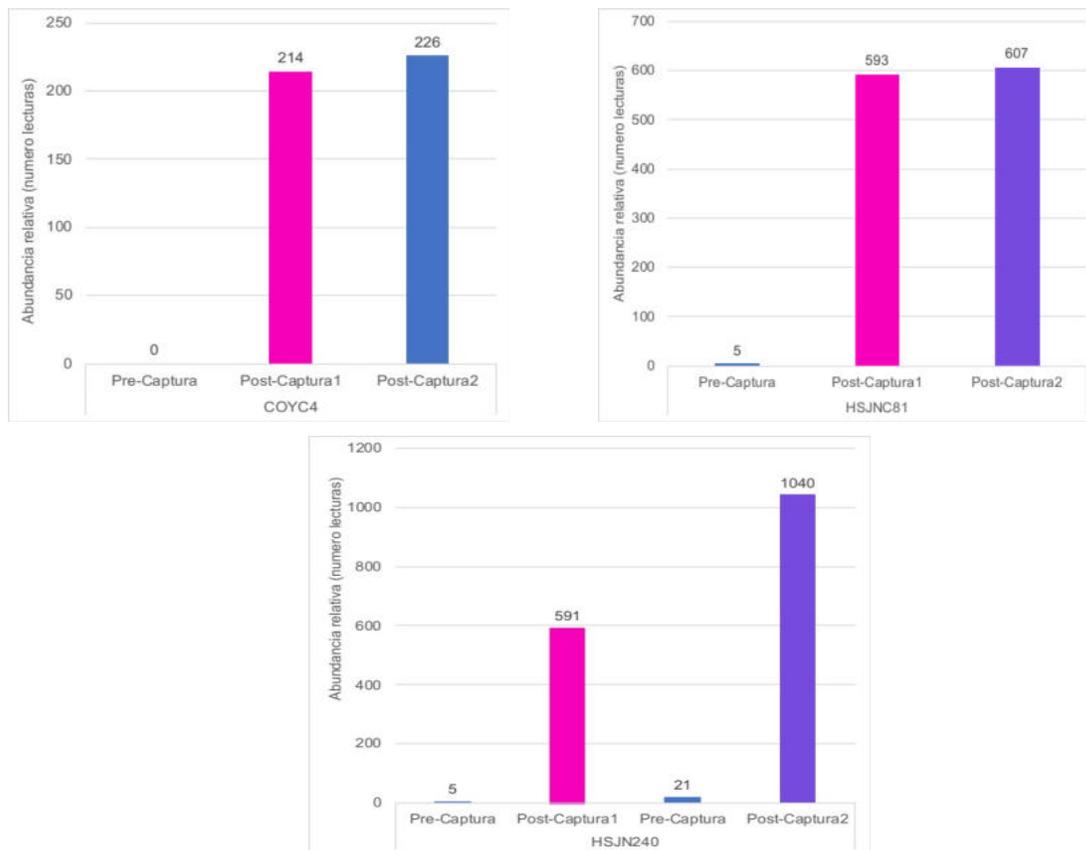


Figura 22. Comparación de abundancias normalizadas de género *Erythroparvovirus* antes y después de dos rondas de captura. Se realizaron dos librerías independientes a partir de individuo HSJN240, por lo que se compararon por separado. Eje X: experimento; Eje Y: Número de lecturas normalizadas.

En el caso de las muestras positivas para género *Orthohepadnavirus*, la comparación normalizada del género se muestra en la Figura 23.

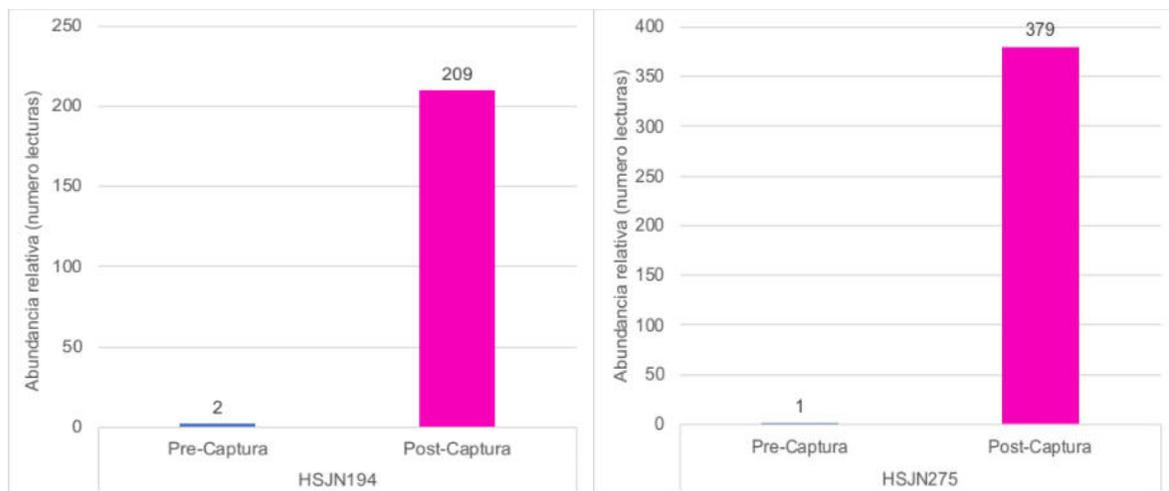


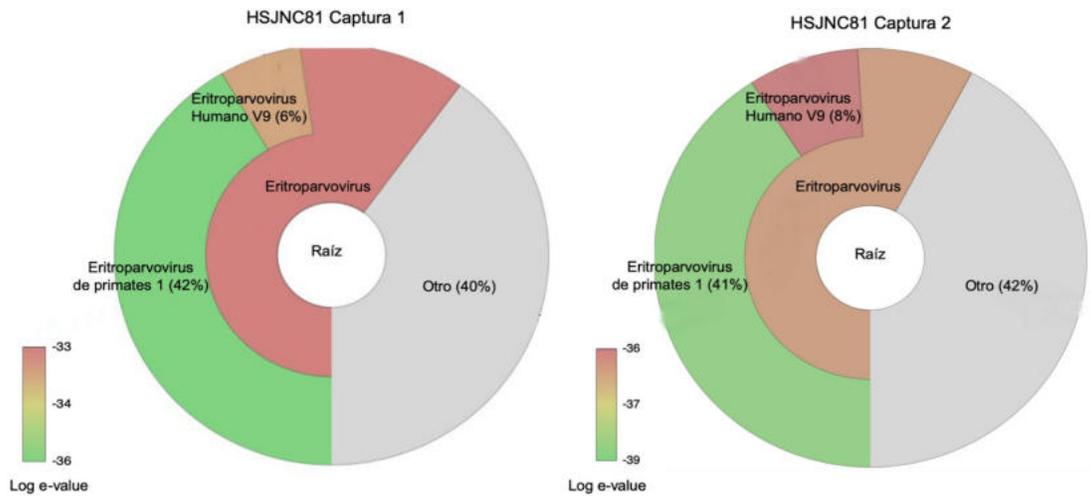
Figura 23. Comparación de abundancias normalizadas de género *Orthohepadnavirus* antes y después de captura. Eje X: experimento; Eje Y: Número de lecturas normalizadas.

La comparación de abundancias normalizadas para todos los géneros virales detectados respecto a sus respectivas librerías crudas se muestra en el Anexo VII.

5.6 Reconstrucción de genomas antiguos virales.

Se hizo el análisis descrito en 4.2.13. Se visualizó la asignación taxonómica de las lecturas filtradas respecto a base de datos NR (no redundante, NCBI), hecha con 89

MegaBLAST y Krona, mostrada en las Figuras 24 y 25. Para el ensamblaje del genoma viral, sólo se consideraron las lecturas asignadas específicamente para especie viral, Figuras 24-25 y Figura 27 para *Erythroparvovirus* y *Orthohepadnavirus*, respectivamente.



Figur

a 24. Lecturas usadas para reconstrucción de genomas de B19V de sitio HSJN. Diagrama de pastel mostrando asignación taxonómica hecha por MegaBLAST respecto a base de datos NR (NCBI), las lecturas con incertidumbre (gris) se desecharon.

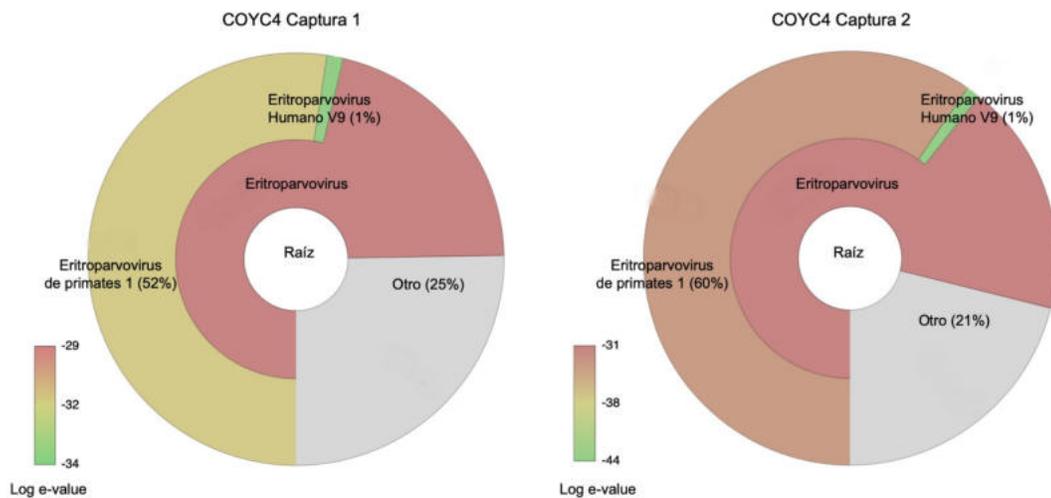


Figura 25. Lecturas usadas para reconstrucción de genoma B19V de sitio COY. Diagrama de pastel mostrando asignación taxonómica hecha por MegaBLAST respecto a base de datos NR (NCBI), las lecturas con incertidumbre (gris) se desecharon.

Las lecturas filtradas obtenidas por cada ronda de captura se mapearon al genoma de referencia, de manera independiente y en conjunto (Cuadro 19-22). Se seleccionó el ensamblaje que cubriera más posiciones para generar un genoma consenso, mostrado en los Cuadros 19-21 y Cuadro 22, para las muestras positivas a B19V y para HBV, respectivamente. Para el mapeo de B19V se utilizaron las referencias FJ591158.1, AB550331.1, NC_004295.1 (disponibles en NCBI), que corresponden a genomas completos del genotipo 1, 2 y 3, respectivamente.

Muestra	COYC4								
	Captura 1			Captura 2			Captura 1 + 2		
Experimento									
Lecturas filtradas totales	240			509			749		
Genotipo de referencia	1	2	3	1	2	3	1	2	3
Lecturas mapeadas	233	239	240	504	508	509	743	747	749
Posiciones cubiertas (≥ 1)	3453	3902	3523	4480	4818	4257	5103	5125	4566
Cobertura (%)	63.8	69.8	70.1	82.8	86.2	84.7	94.3	91.7	90.8
Identidad a referencia (%)	92.9	94.2	94.4	91.5	92.5	93.3	92.6	93.2	93.7

Cuadro 19. Resumen de mapeo a B19V por ronda de captura a partir de COYC4.

En la Figura 26 (a, b, c) se muestra mapeo respecto a Genotipo 2, usando unión de lecturas de captura 1 y 2, para individuos positivos a B19V. Se comparó respecto a lecturas usadas para ensamblaje de genomas antiguos de B19V en Eurasia (Mühlemann y col., 2018), mostrados en Figura 26 (d, e, f). Cabe mencionar que en dicho estudio no se realizó ensayo de captura, sino que se secuenció a mayor profundidad. Posiblemente el ensayo de captura causó una distribución menos homogénea de las lecturas mapeadas de nuestro estudio (Figura 26).

Cuadro 20. Resumen de mapeo a B19V por ronda de captura a partir de HSJNC81.

Muestra	HSJNC81								
	Captura 1			Captura 2			Captura 1 + 2		
Experimento									
Lecturas filtradas totales	2112			11152			13265		
Genotipo de referencia	1	2	3	1	2	3	1	2	3
Lecturas mapeadas	2102	2111	2112	11120	11152	11152	13264	13264	13264
Posiciones cubiertas (≥ 1)	3881	4281	3881	4747	4941	4381	5097	5253	4693
Cobertura (%)	71.7	76.6	75.8	87.7	88.4	87.1	94.2	94	93.3
Identidad a referencia (%)	93.9	94.1	94.1	92.6	92.6	92.6	92.8	92.8	92.8

En el caso de los individuos HSJN194 y HSJN275 se realizó el mapeo respecto a HBV. Tomando como referencia archivo fasta con múltiples secuencias (Anexo V) para hacer mapeo competitivo, con el fin de determinar la secuencia moderna más

similar a la encontrada en HSJN194. Se seleccionó el genotipo A, a partir del cual se determinaron los estadísticos básicos mostrados en Cuadro 22.

Con las lecturas obtenidas a partir de mapeo a referencia escogida se realizó alineamiento con MEGABLAST respecto a base de datos NR (no redundante, NCBI), visualizado en Krona, con el fin de corroborar identidad de las lecturas mapeadas. La asignación realizada a partir de lecturas mapeadas a HBV de HSJN194 y HSJN275 se muestran en Figura 27.

Cuadro 21. Resumen de mapeo a B19V por ronda de captura a partir de HSJN240.

Muestra	HSJN240								
Experimento	Captura 1			Captura 2			Captura 1 + 2		
Lecturas filtradas totales	3689			4239			7928		
Genotipo de referencia	1	2	3	1	2	3	1	2	3
Lecturas mapeadas	3688	3689	3689	4232	4239	4239	7928	7928	7928
Posiciones cubiertas (≥1)	5204	5345	4785	5091	5393	4822	5355	5488	4928
Cobertura (%)	96.2	95.7	95.2	94.1	96.5	95.9	98.9	98.2	98
Identidad a referencia (%)	92.3	92.5	92.5	92.4	92.6	92.6	92.5	92.5	92.6

Cuadro 22. Resumen de mapeo a HBV a partir de HSJN194 y HSJN275.

Muestra	HSJN194	HSJN275
Experimento	Captura 1	Captura 1
Lecturas mapeadas totales	10768	1447
Genotipo de referencia	A (GQ331046)	A2 (HE974383)

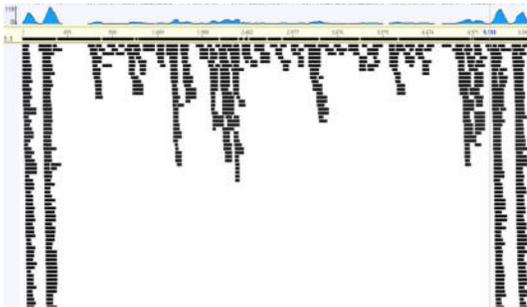
Lecturas mapeadas a sitios únicos sin duplicados	1215	106
Posiciones cubiertas (≥ 1)	2896	1615
Cobertura (%)	89	50
Profundidad	30.82x	4.22x

a)

COYC4

b)

HSJNC81

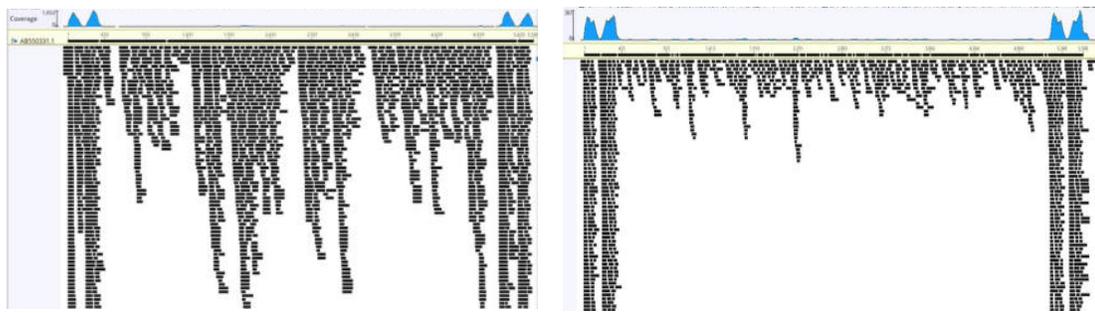


c)

HSJNC240

d)

RISE569



e)

VK477 f)

DA337

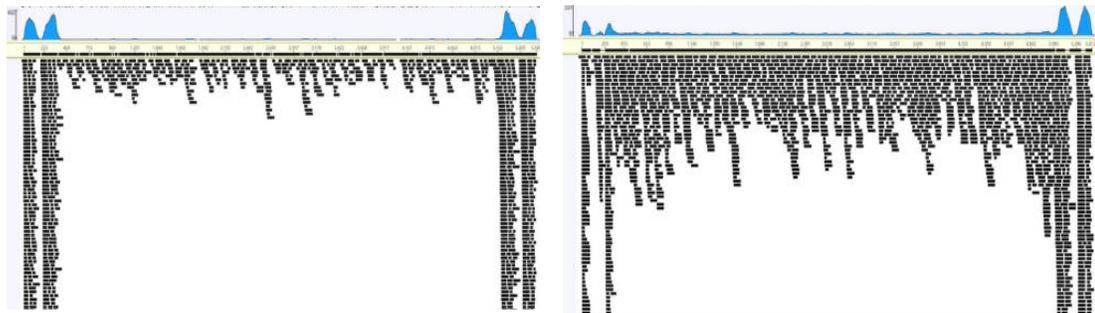


Figura 26. Patrones irregulares de cobertura de B19V a partir de aDNA. Captura de pantalla de mapeo en Geneious respecto a genoma de referencia (B19V), para individuos de las colecciones de CMX (a, b, c) y de Eurasia (d, e, f; datos re-analizados a partir de Mühlemann y col., 2018).

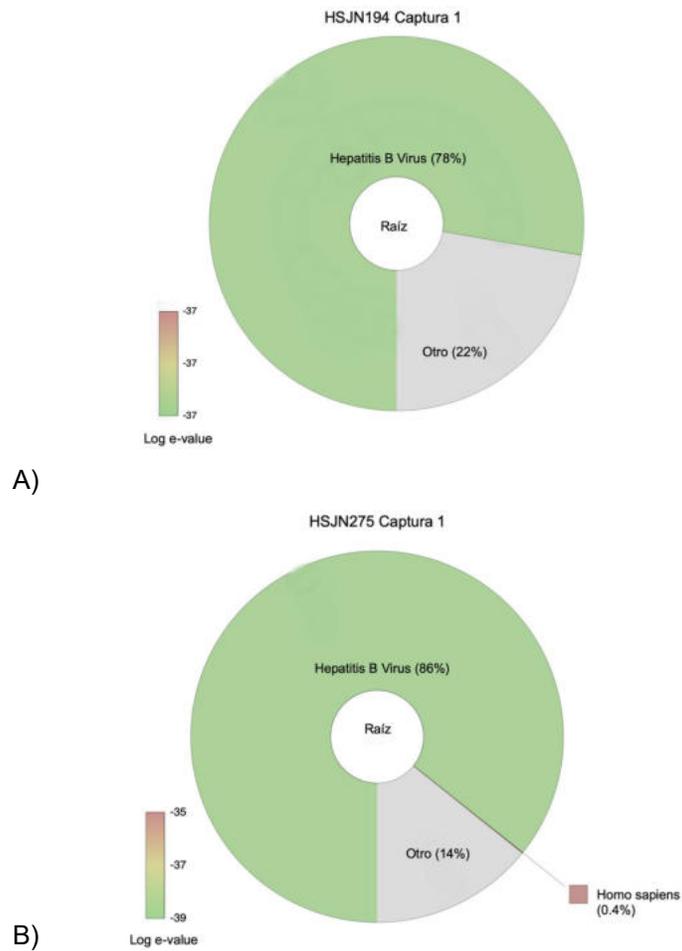
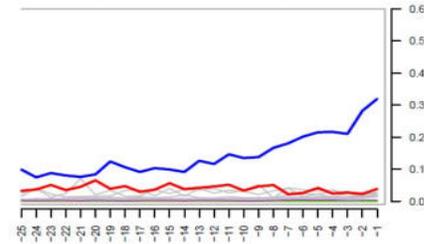
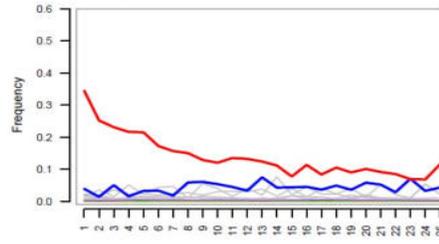


Figura 27. Lecturas usadas para reconstrucción de genoma de HBV de sitio HSJN. Diagrama de pastel mostrando asignación taxonómica hecha por MegaBLAST respecto a base de datos NR (NCBI) para HSJN194 (A) y HSJN275 (B).

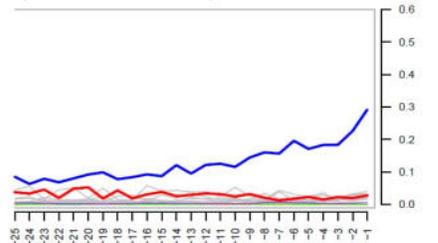
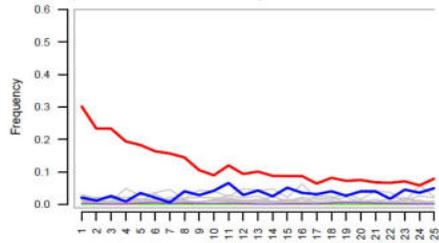
5.7 Determinación de patrones de daño de lecturas mapeadas a virus.

A partir del mapeo respecto a referencia viral se extrajo el archivo de alineamiento en formato BAM. Se determinaron patrones de daño con mapDamage 2.0 para B19V y HBV, mostrados en Figura 28 y Anexo IX (HSJN275).

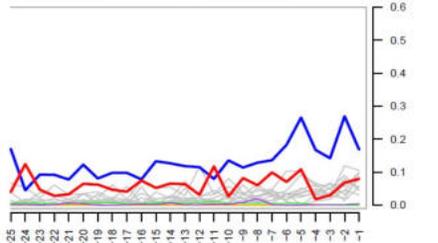
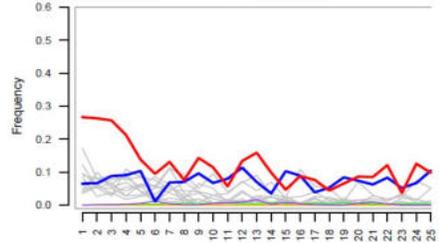
a) HSJN240
(7928
lecturas)



b) HSJNC81
(13265
lecturas)



c) COYC4 (749
lecturas)



d) HSJN214
(1215
lecturas)

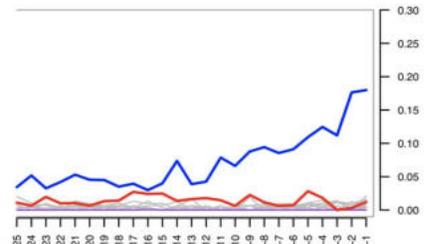
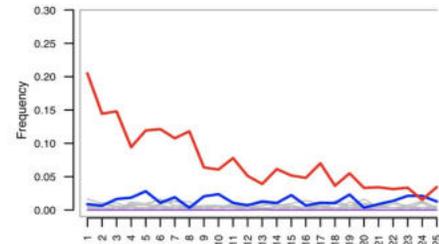


Figura 28. Patrones de daño típicos de aDNA respecto a especie viral. Se tomó como referencia genoma de B19V (a, b, c) o HBV (d). El eje X muestra nucleótidos respecto a extremo de lectura. El eje Y muestra la presencia de daño (C>T: rojo; G>A: azul; Otros: gris). Los gráficos representan extremo 5´(izquierda) y 3´(derecha).

Para el alineamiento de lecturas pertenecientes a HSJN275 respecto a HBV no se obtuvieron patrones típicos de daño, posiblemente por cantidad mínima (106) de lecturas mapeadas, mostrado en Anexo IX.

5.8 Análisis filogenético de B19V

Se alinearon 112 genomas de B19V y se recortaron para conservar sólo región codificante (CDS). Este alineamiento se analizó en jModelTest y se seleccionaron los modelos evolutivos (HKY, GTR+G) con intervalo de confianza de 100%.

Se construyó un fenograma con 1000 pseudoréplicas usando modelo HKY en Geneious a partir del alineamiento de 112 regiones codificantes de B19V, mostrado en la Figura 29.

Figura 29. Fenograma basado en matriz de distancias a partir de región codificante de B19V. Los genomas antiguos de Eurasia son mostrados en azul (genotipo 1) y rosa (genotipo 2) (Mühleman y col., 2018). Los genomas antiguos de época Colonial mexicana en rojo (genotipo 3)

Los genomas antiguos de Eurasia son mostrados en azul (genotipo 1) y rosa (genotipo 2) (Mühleman y col., 2018). Los genomas antiguos de época Colonial de CMX en rojo (genotipo 3), y el formato de nombre se respetó en los tres análisis filogenéticos mostrados en esta sección (Genotipo_Muestra_Antigüedad_País_Region_Hospedero).

Con base al fenograma basado en distancias (Figura 29) se pueden asignar a los tres genomas antiguos reconstruidos al genotipo 3 de B19V. El proveniente de HSJNC81 es similar a V9 Parvovirus, que en este estudio se considerara como genotipo 3A por su cercanía a este subgenotipo, mientras los provenientes de HSJN240 y COYC4 son similares a genotipo 3B; para todos los casos se consiguió un soporte >99% de las pseudorélicas (1000).

Se construyó un árbol de máxima verosimilitud con 1000 pseudorélicas usando modelo GTR+Gamma en RAxML, con alineamiento de 112 regiones codificantes de B19V, mostrado en Figura 30.

En base al árbol de máxima verosimilitud (Figura 30) se logró distinguir a los genomas reconstruidos de B19V en genotipo 3A (HSJNC81) y genotipo 3B (HSJN240 y COYC4) en >90% de las pseudorélicas (1000). Las secuencias del genotipo 3B son similares a secuencias etiquetadas con origen europeo; sin embargo, los individuos de los que se obtuvo la muestra eran extranjeros africanos viviendo en Alemania (Hübschen y col., 2009).

Se construyó un árbol de máxima credibilidad de clados en BEAST, utilizando un reloj molecular estricto (Figura 31) y relajado (Anexo VIII). Se estimó una tasa de mutación de 1.13×10^{-5} s/s/y y 1.75×10^{-5} s/s/y, y una fecha de divergencia de MRCA de 10.9 mil años (kya) y 8.8 kya, respectivamente.

La fecha de divergencia de los genotipos 1, 2 y 3, se estimó de 7.2, 2.2 y 3.1 kya, respectivamente, considerando un reloj molecular estricto; y de 7.4, 1.9, y 1.9 kya, respectivamente, considerando para un reloj molecular relajado.

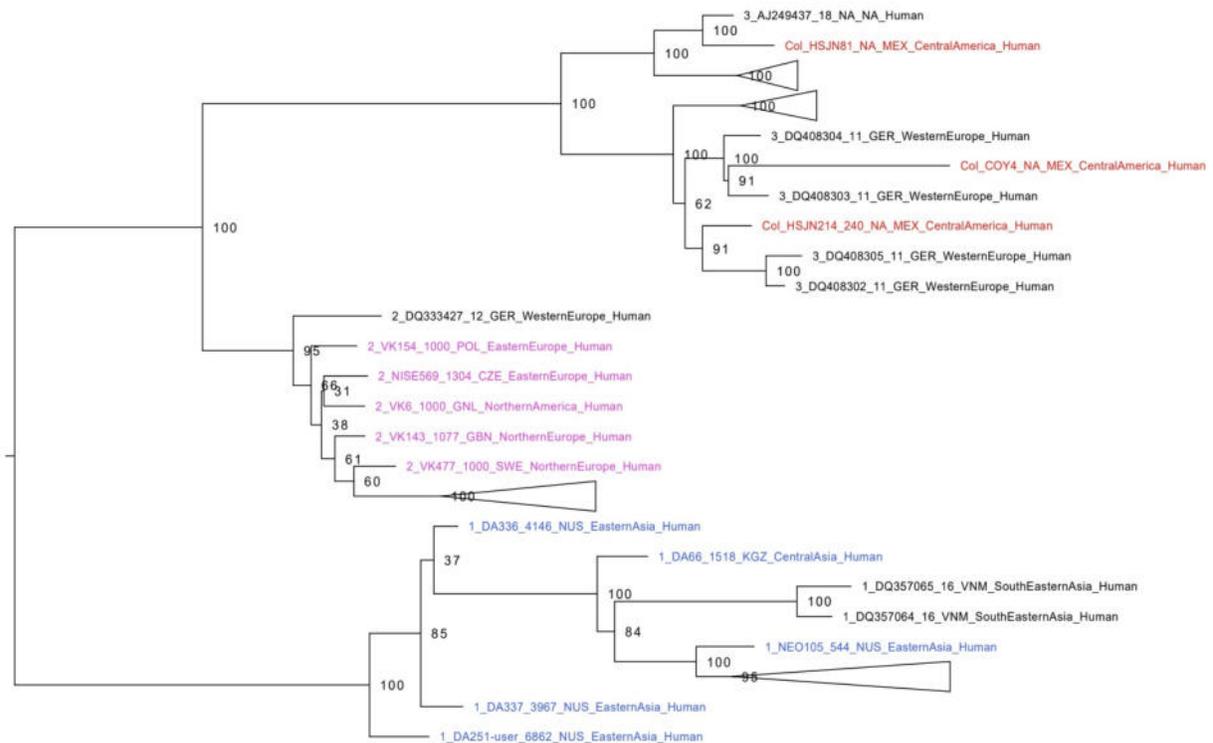


Figura 30. Árbol de máxima verosimilitud basado en región codificante de B19V. Se colapsaron clados que no tuvieron relación directa con genomas antiguos.

Los valores obtenidos son similares a los obtenidos previamente con un reloj molecular estricto y relajado, donde se reportaron tasas de mutación 1.22×10^{-5} y 1.67×10^{-5} s/s/y y una fecha de divergencia del MRCA de 10.1 kya y 8.4 kya, respectivamente (Mühlemann y col., 2018). La adición de tres genomas antiguos mexicanos tuvo un efecto sutil en la tasa de mutación y, por tanto, en las fechas de divergencia de la especie pues Mühlemann y col. (2018) incluyeron decenas de genomas antiguos con dataciones notoriamente mayores (más de 6000 años de antigüedad) en sus análisis.

El árbol de máxima credibilidad de clados (Figura 31) soporta la topología obtenida por el fenograma de distancias (Figura 29) y el árbol de máxima verosimilitud (Figura 3), por lo que podemos con certeza (mediante 3 métodos filogenéticos) asignar a los genomas antiguos de B19V obtenidos de sitios del periodo colonial de México al genotipo 3A (HSJNC81) y genotipo 3B (HSJN240 y COYC4). No existe otro reporte de reconstrucción de genotipo 3 antiguo de B19V.

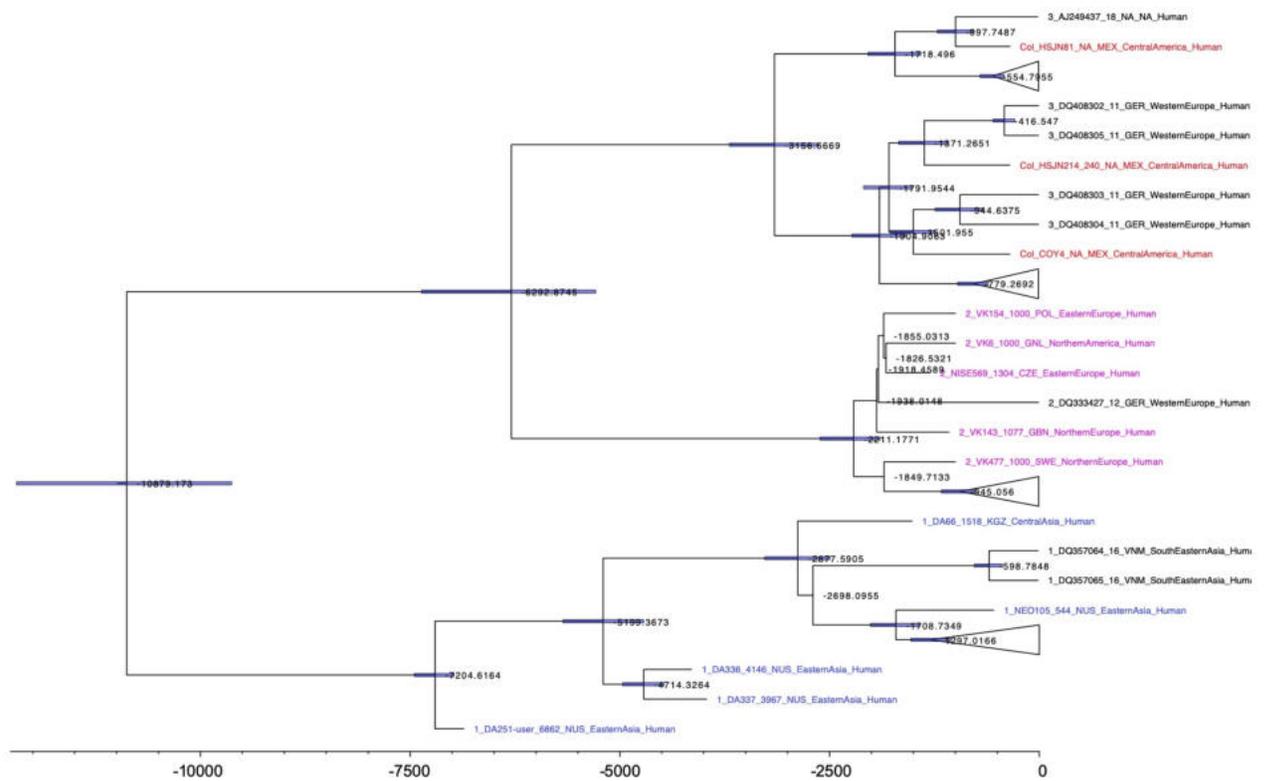


Figura 31. Árbol de máxima credibilidad de clados usando reloj molecular estricto. Las barras azules muestran intervalo con 95% de densidad posterior (HPD) y el eje x muestra años a partir del presente.

5.9 Análisis filogenético de HBV.

Se alinearon 126 genomas de HBV, cuidando que todos tuvieran misma longitud. Este alineamiento fue analizado en jModelTest y se escogieron modelos evolutivos (GTR+G+I, HKY+G+I) con intervalo de confianza de 100%. Se construyó un fenograma con 1000 pseudoréplicas usando modelo HKY en Geneious, a partir de alineamiento, mostrado en Figura 32.

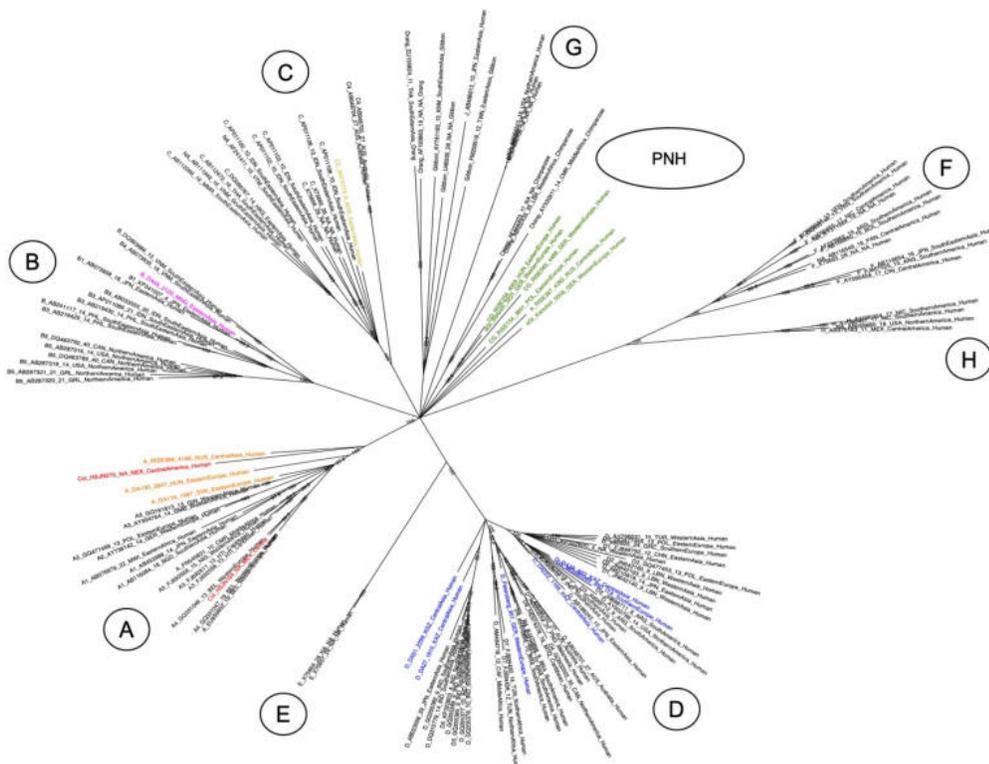


Figura 32. Fenograma basado en matriz de distancias a partir de genoma completo de HBV. Los genomas antiguos son mostrados en color, los modernos en negro y los genomas antiguos de época Colonial mexicana en rojo (HBV/A). PNH: primates no humanos.

Los genomas antiguos son mostrados en color, mientras los modernos en negro. Los genomas antiguos de la época Colonial de CMX se muestran en rojo (genotipo A), mientras los genotipos HBV/A, B, C, D y similar a primates no humanos se muestran en naranja, rosa, amarillo, azul, y verde, respectivamente. El formato de nombre corresponde a Genotipo_Muestra_Antigüedad_País_Región_Hospedero.

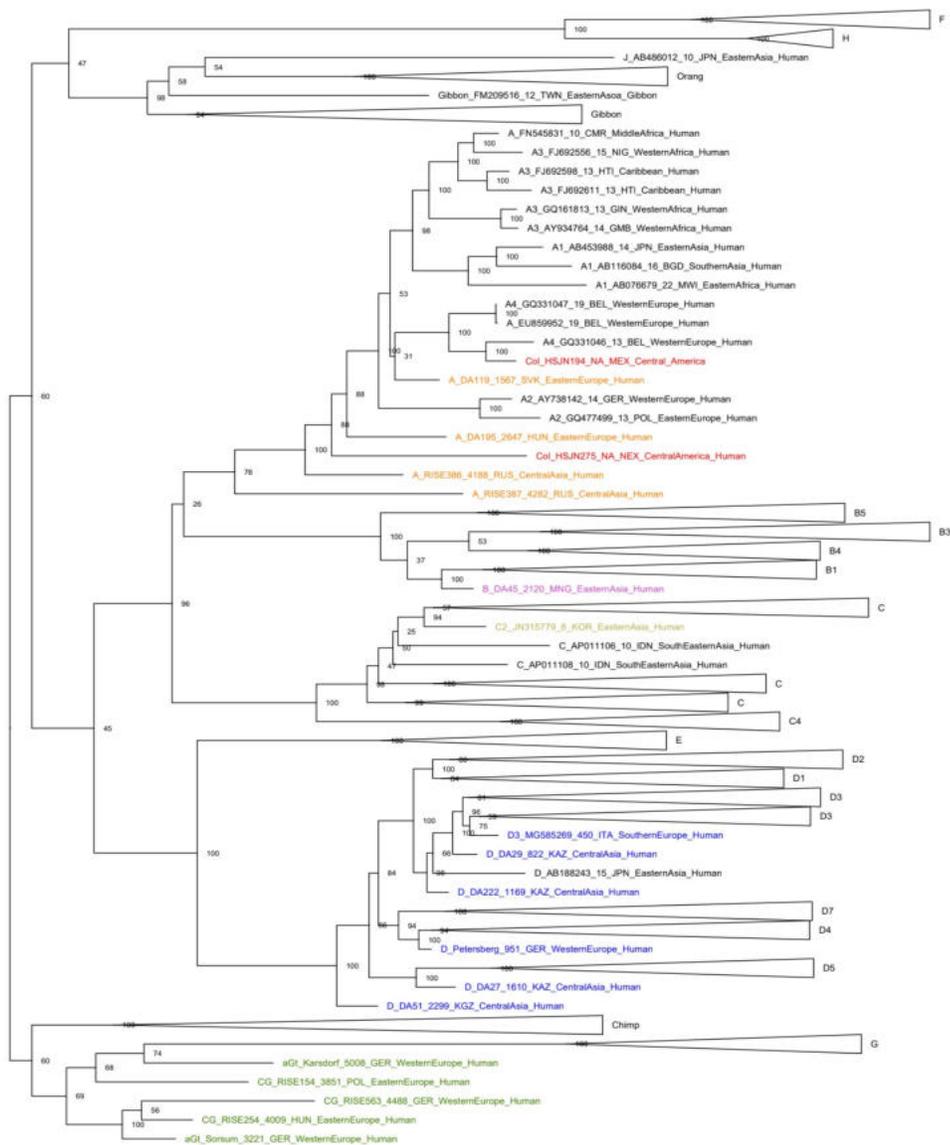


Figura 33. Árbol de máxima verosimilitud a partir de genoma completo de HBV. Los genomas antiguos son mostrados en color, los modernos en negro y los genomas antiguos de época Colonial mexicana en rojo (genotipo A).

A partir del fenograma de distancias (Figura 32) se pudo observar que los genomas reconstruidos de HBV son similares a HBV/A, con un soporte de 100% de las pseudoréplicas.

Se construyó árbol de máxima verosimilitud con 1000 pseudoréplicas usando modelo GTR+G+I en RAxML, con alineamiento de 126 genomas de HBV, mostrado en Figura 33; los genomas pertenecientes al HBV/A no se encuentran colapsados. A partir de dicho árbol (Figura 33) se observa similitud de genomas reconstruidos a HBV/A.

El genoma de HBV a partir de HSJN275 es similar a genomas antiguos previamente reportados (Mühlemann y col., 2018) de Rusia y Hungría, que podría sugerir origen antiguo; mientras que, el genoma obtenido de HSJN194 es similar a genomas obtenidos en Bélgica a partir de africanos (Pourkaim y col., 2010).

6. DISCUSIÓN

6.1 Presencia de VARV

Con base a fuentes históricas, se sabe que ciertos virus fueron causantes de múltiples epidemias durante la época colonial de México, como la viruela (VARV), el sarampión y las paperas (Mandujano y col., 2003). De los anteriores, VARV es el único virus con genoma de dsDNA, y el único recuperado a partir de restos arqueológicos (Duggan y col., 2017; Pajer y col., 2017). Al tratarse de uno de los virus más importantes para la historia de la humanidad, se incluyó un ensayo enfocado en su identificación mediante PCR.

El sitio HSJN representa un sitio típico de epidemia del siglo XVI (Meza, 2013). Aunque existen evidencias óseas de infecciones por sífilis y tuberculosis (Wesp, 2017) se desconocen agente(s) patógeno(s) causante(s) de las epidemias ocurridas en este hospital. VARV nunca se ha descartado como el agente causal (INAH, 2016), ya que durante la colonia fue común (Mandujano y col, 2003) y presenta una tasa de mortandad de hasta 30% de los infectados (Théves y col, 2014). Por ello, VARV es un patógeno capaz de causar epidemias masivas como la encontradas en el HSJN. Además, la presencia de osteomielitis en 34 de los individuos de la colección (Wesp, 2017), pudiera indicar una infección de VARV; sin embargo, la osteomielitis no es un signo osteológico único por la infección de este virus, ya que es causado por otros patógenos, incluso bacterianos (Castillo-Chavez, 2000; Eid, 2003).

Se extrajo DNA bajo estrictas condiciones de aDNA y se obtuvieron concentraciones desde 0.2 ng/μl hasta 6.12 ng/μl (Cuadro 18), rango que entra en el límite de cuantificación (0.01-100 ng/μl) de fluorómetro Qubit. Adicionalmente, fue

indetectable la presencia de DNA en controles negativos lo cual indica nula o mínima contaminación en fase de extracción de DNA.

El primer aproximamiento llevado a cabo se enfocó en identificación de VARV mediante PCR en tiempo final. El análisis por PCR con aDNA fue históricamente el primer acercamiento para el estudio de muestras antiguas (Paabo, 1989) y ha demostrado ser una metodología con recurrentes contaminaciones modernas (ambientales, humanas), por lo que debe llevarse a cabo con rigurosas medidas preventivas y en un laboratorio aislado al laboratorio moderno (Hofreiter y col., 2001; Key y col., 2017). Los oligonucleótidos usados como control positivo para VARV jamás estuvieron en contacto con extractos de aDNA ni con controles negativos (los oligonucleótidos se agregaron al control positivo en el laboratorio moderno y el resto de las reacciones sólo fueron abiertas en el laboratorio de aDNA, donde nunca se ha amplificado ningún producto de PCR). Se utilizó la misma alícuota de H₂O en todos los experimentos para descartar presencia de contaminantes en solvente de extracción. Con lo anterior se consiguieron controles positivos y negativos correctos y es posible determinar que el total de las muestras analizadas por esta metodología resultaron negativas para los genes diana a amplificar (Figura 17 y 18).

El análisis por PCR presenta múltiples desventajas. La primera es que, en el caso de muestras complejas (no purificadas y sin enriquecer), se recomienda usar >10 ng/μl de extracto de DNA para su amplificación; dichas concentraciones representan en ocasiones todo el DNA obtenido de las muestras de nuestro estudio, por lo que sólo se usaron las muestras con las concentraciones más altas de DNA (asteriscos en Cuadro 18).

En el caso de DNA en baja abundancia (como el viral) se espera que sea órdenes de magnitud menor que el DNA humano o bacteriano (Key y col., 2017), por lo que

su amplificación requeriría muchos ciclos, que a su vez aumenta la posibilidad de amplificación de falsos positivos (Lorenz, 2012).

El análisis por PCR es secuencia específico, donde inclusive con la presencia de aDNA (fragmentado) autentico de VARV, la DNA polimerasa no amplificaría sin la presencia de cadena doble molde hibridada entre cebador y aDNA. Con el fin de disminuir este efecto se diseñaron amplicones cortos (<100 bp) de un tamaño esperado para fragmentos de aDNA (Velsko y col., 2018); sin embargo, la limitante sólo fue atenuada.

Los patógenos virales evolucionan mas rápido que muchas de las bacterias u eucariotas (Holmes, 2011), por lo que es de esperar que las cepas antiguas sean altamente distintas a las secuencias modernas (Krause-Kyora y col., 2018; Mühlemann y col., 2019), disminuyendo o evitando hibridación de cebadores con aDNA.

El análisis por PCR es una metodología rápida y barata, mas no representa la mejor opción para identificación de patógenos virales. De todos los estudios de patógenos virales antiguos hechos a la fecha, la amplificación por PCR sólo ha sido efectiva para identificar DNA viral a partir de piel de momias, como en el caso de VPH (Fornaciari y col., 2003), VARV (Biaginni y col, 2012), HBV (Kahila y col., 2012) o a partir de piel de muestras en museo para VARV (Pajer y col, 2017).

En los últimos años, el análisis metagenómico a partir de datos de secuenciación masiva ha ganado auge en el área de estudios de aDNA (Key y col., 2018), por lo que se utilizó como metodología imparcial para el escrutinio de aDNA viral.

6.2 Análisis metagenómico de patógenos virales.

Con la cantidad de lecturas obtenidas a partir de secuenciación a baja profundidad (<20 millones de lecturas) de librerías crudas, el número de lecturas asignadas a cada especie viral fue mínimo, similar a lo reportado en análisis metagenómicos a partir de muestras humanas complejas (suero) donde patógenos virales se encuentran en baja abundancia (Metsky y col., 2019), imposibilitando la construcción de ensamblajes o genomas completos; por lo que se diseñó un ensayo de captura que incluyó patógenos virales con importancia clínica (Cuadro 11 y 12), se incluyeron especies virales previamente capturadas a partir de muestras arqueológicas (VARV, HBV, B19V), al igual que otras especies virales con genomas de DNA. Se omitieron especies virales con genomas de RNA pues no existe evidencia que sugiera la posibilidad de recuperar RNA a partir de muestras óseas arqueológicas; además, sería inútil debido a que nuestra metodología contempla únicamente secuenciación de DNA, sin paso previo de retro-transcripción; sin embargo, se incluyó una secuencia consenso de retrovirus (HIV1/SIV) en el diseño pues se sabe que esta familia de virus es capaz de integrarse al genoma humano en forma de dsDNA (Lesbats y col., 2016).

Metsky y col. (2019) utilizaron sondas de ssRNA y ssDNA para hacer captura de patógenos virales en muestras complejas humanas, sin encontrar diferencia aparente en eficiencia de captura, a pesar de que la interacción RNA:DNA es más estable que la de DNA:DNA (Lesnik y Freier, 1995). Sin embargo, no se ha comparado eficiencia de captura usando sondas de ssDNA en muestras antiguas, por lo que se utilizaron sondas de ssRNA.

Después de comparar las abundancias normalizadas de todas las especies virales detectadas en cada sitio arqueológico (Anexo VI), se seleccionaron muestras con mayor abundancia de algún género viral humano previamente recuperado de restos

arqueológicos (*Orthopoxvirus*, *Orthohepadnavirus*, *Erythroparvovirus*) para enriquecer mediante ensayo de captura (rectángulos rojos en Figura 20 y 21).

Las librerías capturadas fueron secuenciadas y se logró enriquecimiento de DNA pertenecientes al género *Erythroparvovirus*. Se identificaron dos “especies”, el Parvovirus de primates 1 (B19V) y el Parvovirus humano V9. V9 fue considerado como una nueva especie viral, distinta a B19V cuando fue descubierto (Nguyen y col., 1999), debido a que presenta mayor variabilidad (11%) respecto a cepas pertenecientes al genotipo 3 (variación intragenotipo <4.5%) (Hübschen y col., 2009). Sin embargo, muchos autores lo consideran parte del genotipo 3 (Parsyan y col., 2007; Hübschen y col., 2009), de igual manera que en este trabajo.

Se logró el enriquecimiento de tres muestras arqueológicas, el cráneo 4 de la Concepción (COYC4), el cráneo 81A (HSJNC81), y el cráneo 240 de HSJN (HSJN240). Este enriquecimiento es esperado, tomando en cuenta que estas muestras presentaron la mayor abundancia del género *Erythroparvovirus* al realizar la comparación del análisis metagenómico con lecturas crudas (mostrado en Figura 20 y 21). En las muestras enriquecidas para B19V ninguna otra especie viral fue enriquecida, a pesar de que secuencias de otras especies virales fueron incluidas en el diseño de sondas de captura (Cuadros 11 y 12), por lo que se considera que el enriquecimiento fue específico a DNA de B19V.

Adicionalmente, se enriqueció DNA de HBV a partir de ensayo de captura de librerías de los individuos HSJN194 y HSJN275, que es de esperarse considerando el análisis metagenómico de las librerías crudas, donde estos individuos tienen la mayor abundancia del género *Orthohepadnavirus* al compararse con los demás individuos de la colección HSJN (Figura 21). Sólo las muestras con lecturas asignadas a HBV presentaron enriquecimiento de esta especie viral después del ensayo de captura; a pesar que se usaron las mismas sondas de ssRNA en otras

muestras (como las enriquecidas con B19V), por lo que el enriquecimiento fue específico a la presencia de HBV o B19V (Figuras 20 y 21).

Si bien varias muestras (HSJN177, HSJN194, HSJNC4) presentaron una abundancia alta de lecturas (Figura 21) asignadas al género *Orthopoxvirus*, no se obtuvo enriquecimiento de este género en ninguna de las capturas realizadas. Las lecturas asignadas a *Poxviridae* en su mayoría fueron asignadas a BeAn 58058, un virus encontrado recientemente en roedores de Brazil (Wanzeller y col., 2018); además, todas las lecturas encontradas en nuestro estudio se agrupan en una sola región del genoma, lo cual sugiere falso positivo (Warinner y col., 2017). En análisis metagenómico independiente de muestras fecales se ha encontrado el mismo patrón, donde existen lecturas que mapean a una región (8200bp a 8700bp) de BeAn 58058 (Tithi y col., 2018). En nuestro estudio las lecturas se agrupan exactamente a la misma región por lo que muy posiblemente se trata de un falso positivo. Tithi y col. (2018) analizaron más a profundidad esta misma región y reportaron una alta similitud con elementos repetidos Alu de humanos.

Con lo anterior queda concluir que no se encontró aDNA de VARV en individuos del periodo colonial de México. Lo que sugiere que los individuos analizados del HSJN no fueron infectados por VARV, o que el DNA de esta especie no se conserva en sustrato óseo antiguo. Cabe mencionar, que el HSJN atendió enfermos de las epidemias de sarampión (-ssRNA) (Meza, 2013), mismo que sería imposible de recuperar con nuestra metodología; además que, no existe evidencia que proponga que el RNA puede permanecer viable durante periodos tan largos de tiempo en sustrato arqueológico sin congelar.

6.3 Reconstrucción de genomas virales

Para la reconstrucción de genomas de B19V se utilizaron las lecturas obtenidas de cada ronda de secuenciación post-captura en conjunto y se siguió la metodología

modificada de Mühleman y col. (2018). BWA localiza una semilla (32 nt) en el extremo de la lectura que alinee con sólo 2 nt distintos a una referencia, después se extiende un alineamiento global con cierto número de diferencias (-n); sin embargo, en el caso de aDNA, los extremos terminales presentan alto contenido de cambios C>T, lo que evitaría conservar lecturas dañadas, incluso si fuesen auténticas. Por lo anterior, en el mapeo con BWA se deshabilitó la extensión de semilla (-l) pues análisis *in silico* ha demostrado que permite mapear un mayor número de lecturas auténticas, incluso con patrones típicos de daño en extremos terminales (Schubert y col., 2012).

Se consideraron muestras positivas a B19V cuando se logró cubrir, al menos una vez, más del 40% de la región codificante de B19V (Mühleman y col., 2018) o del genoma completo de HBV.

En el caso de muestras positivas a B19V se adicionó un paso al mapeo por BWA donde se identificaron las lecturas similares (mediante BLASTn) a algún tipo de parvovirus (Mühlemann y col., 2018). Este aproximamiento realiza alineamientos locales cortos, por lo que sólo se conservaron los alineamientos que cubrieran más del 85% de la lectura, para evitar alineamientos azarosos de dominios comunes.

Se unieron las lecturas mapeadas a algún tipo de parvovirus, provenientes de mapeo con BWA y alineamientos con BLASTn.

Debido a que las librerías capturadas fueron amplificadas por PCR, se esperaba presencia de secuencias duplicadas (Metsky y col., 2019) por lo que fueron removidas. El total de secuencias restantes se alineó (con MegaBLAST) respecto a la base de datos no redundante (NR) de NCBI, que incluye mamíferos, arqueas, hongos, protozoarios, virus y otros organismos, obteniéndose la asignación taxonómica mostrada en las Figuras 24 y 25. Sólo las lecturas específicas para *Erythroparvovirus* fueron utilizadas para construcción de genoma consenso en

Geneious, donde de manera independiente se mapeó respecto a genomas de referencia de los tres genotipos reportados (B19V), y se conservó genoma consenso donde se cubriesen más sitios (Cuadro 19-21).

Al obtener mapeo respecto a referencia de B19V se encontró cobertura desigual en los 400 nt de ambos extremos terminales (Figura 26a-c). Esto tiene sentido si se considera que el genoma de B19V es de ssDNA, excepto en extremos terminales, donde es de dsDNA (Figura 8); es bien aceptado que los procesos de depuración y desaminación ocurren ms rápidamente en ssDNA que en dsDNA (Mühlemann y col., 2018), por lo que la región ssDNA del genoma B19V es más propensa a procesos de degradación *post-mortem*. Se volvieron a mapear lecturas asignadas a B19V antiguo reportadas por Mühlemann y col. (2018) y todas presentaron patrón similar de cobertura, tres de ellas son mostradas en Figura 26c-e.

En el caso de HBV, se obtuvieron directamente las lecturas mapeadas a genotipo seleccionado (4.2.13) obtenidas con BWA. Se corroboró su similitud a HBV mediante MegaBLAST respecto a una base de datos no redundante, para asegurar que las lecturas no tuvieran un origen distinto a esta especie viral.

Se consiguió reconstruir tres genomas de B19V y dos de HBV. El Cuadro 23 muestra la cobertura final obtenida.

Cuadro 23. Coberturas de genomas reconstruidos. (*): respecto a región codificante.

Especie viral	Muestra	Cobertura de referencia (%)	Profundidad promedio
B19V*	HSJNC81	92.7	7.9x
	HSJN240	98.7	25.2x
	COYC4	93.5	6.7x

	HSJN194	89.9	30.8x
HBV	HSJN275	86.4	23.1x

6.4 Genoma antiguo de B19V Parvovirus

B19V es un virus común en poblaciones modernas con tasas de seropositividad tan altas como 85% en adultos mayores (Broliden y col., 2006), y se transmite principalmente por vías aéreas, por lo que se determinó su autenticidad para descartar origen moderno.

Para estudios de aDNA, rigurosamente se requiere demostrar que los fragmentos utilizados para reconstruir un genoma presentan daños en extremos terminales, sobre todo desaminación de citosinas, que es el fenómeno más estudiado en muestras antiguas (Lindahl y col, 1993; Hofreiter y col., 2001; Dabney y col., 2013). Los patrones de daño para las muestras HSJNC81, HSJN240 y COYC4 se muestran en Figura 28 respecto a referencia B19V (genotipo 2), y respecto a HBV (genotipo A) para HSJN194; los de HSJN275 se muestran en Anexo IX.

En el caso del mapeo de lecturas modernas, se esperaría una baja frecuencia de diferencias respecto a la referencia (moderna), distribuidas de manera homogénea a lo largo de las lecturas y sin sesgo por cambio de C>T. En el caso del mapeo a una especie muy divergente (distinta), las lecturas presentarían alta frecuencia de diferencias con respecto a referencia, mismas que estarían distribuidas a lo largo del genoma, y sin sesgo por cambio de C>T.

Por el contrario, en nuestros análisis (Figura 28) se observa que la cantidad de diferencias respecto a la secuencia de referencia son bajas comparadas con los cambios C>T que alcanzan valores máximos (25%) en los extremos terminales de las lecturas. Este patrón es típico de aDNA (Warinner y col., 2017). La muestra

COYC4 (Figura 28) no presenta patrones limpios de daño; sin embargo, se observa una tendencia similar. Esto es atribuido a la baja cantidad de lecturas mapeadas a B19V (menos de 10% de las mapeadas a HSJNC81). El efecto del número de lecturas mapeadas respecto a patrones de daño es mostrado por Mühlemann y col. (2018) (Figura 34), quienes realizan sub-muestreo de lecturas mapeadas a HBV para construir patrones de daño que son más claros en función del número de lecturas utilizadas.

La alta frecuencia de C>T (G>A complementaria) en extremos terminales de lecturas, así como el tamaño corto de los fragmentos de DNA, corrobora la presencia de patrones de daño típicos de aDNA en las lecturas utilizadas para reconstrucción de los tres genomas de B19V.

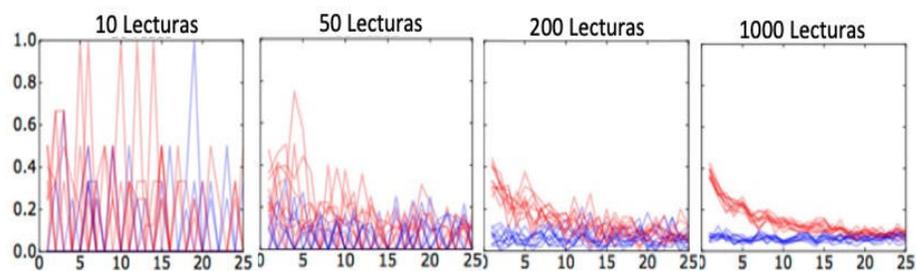


Figura 34. Patrones de daño de HBV antiguos de Eurasia. Patrones de daño a partir de sub-muestreo de número de lecturas mapeadas a HBV (Mühlemann y col., 2018).

6.5 Primera evidencia molecular directa de un virus traído de África a América (México) en el periodo colonial.

Previamente Mühlemann y col. (2018) reportaron un conjunto de genomas parciales o completos de B19V obtenidos de diversas partes de Eurasia; sin embargo, no reportan ningún genotipo 3. En su artículo no descartan la posibilidad de reconstruirlo a partir de restos arqueológicos óseos, pues no hay diferencias

evidentes en estructura de este genotipo respecto al 1 ó 2 que evitara su conservación durante siglos (Mühlemann y col., 2018). En el mismo estudio se identifica además un genoma parcial (VK6) de B19V genotipo 2 obtenido de un entierro múltiple del siglo X D.C. en Groenlandia (Igaliku Kujalleq). Geográficamente, Groenlandia pertenece a América; sin embargo, tiene una historia poblacional distinta a la América continental. Groenlandia fue colonizada por los vikingos en 985 D.C. (Park y col., 2008), más de 500 años antes de que Cristóbal Colón llegara a América (1492 D.C.). La muestra con B19V (VK6) estudiada por Mühlemann y col. fue obtenida a partir de un entierro múltiple, del que se hizo análisis de isótopos de $^{87}\text{Sr}/^{86}\text{Sr}$ de algunos de los individuos, lo cual permitió determinar que los restos óseos pertenecían a inmigrantes de Islandia (Mühlemann y col., 2018). Este virus es similar a secuencias modernas del genotipo 2 (Figura 29-31), que en base a su distribución moderna es raramente encontrado fuera del Norte de Europa (Janovitz y col., 2017), por lo que posiblemente fue traído a América (Groenlandia) por migraciones vikingas provenientes de Islandia. Es sabido que hubo exploraciones vikingas a Canadá Ártica (Park y col., 2008); mas no existe evidencia que sugiera que B19V fue llevado a América continental desde Groenlandia, o que B19V existiera en América previamente a contacto europeo.

En tanto a los genomas obtenidos en zonas arqueológicas de México, éstos presentan similitud con secuencias del genotipo 3, que es endémico de Ghana, Brasil e India, con infecciones esporádicas en otros países del mundo (Hübschen y col., 2009). Actualmente, se conocen dos subgenotipos de este virus, el 3A es mayormente encontrado en Ghana y el 3B que se ha encontrado fuera de África como en Brasil y Alemania (extranjeros de Marruecos, Egipto y Turquía) (Pourkarim y col., 2010).

Se ha propuesto que el origen de B19V es africano debido a que el genotipo 3 cuenta con la mayor diversidad de este virus, reflejando su larga historia evolutiva (Hübschen y col., 2009); mientras que, su escasa presencia fuera de Africa se ha observado en países con historias relacionadas con Africa (como Brasil e India).

Independientemente, se infirió la ancestría genética del hospedero al comparar la fracción humana (no usada en metagenómica) con panel de referencia que incluyó individuos de América (México y Perú), Europa (España y residentes de Utah con ancestría europea), Asia (China) y África (Nigeria, Gambia y Sierra Leona) obtenidos de panel de referencia “1000 genomes” (1000 Genomes Project Consortium, 2015). El análisis genómico de la fracción humana perteneciente a la colección HSJN fue realizado por la Dra. María C. Ávila Arcos y es mostrado en un análisis de componentes principales (PCA) en la Figura 35, donde permite identificar a los individuos 81 y 240 del HSJN como afrodescendientes.

La ancestría genética/morfológica de los individuos 81 y 240 de HSJN, en conjunto con la similitud del genoma ensamblado de B19V con secuencias africanas, nos permite asociar a los genomas reconstruidos de B19V genotipo 3 con un origen africano. Lo cual no es sorprendente si se considera que este genotipo es endémico de países del Oeste africano (Ghana) (Hübschen y col., 2009) que fueron puntos importantes para la obtención de personas esclavizadas con destino a las colonias americanas (Velázquez e Iturralde, 2012). En un estudio independiente, algunos de los individuos de esta colección (HSJN150, HSJN214, HSJN296) fueron sometidos a análisis por isótopos de $^{87}\text{Sr}/^{86}\text{Sr}$ y ^{14}C , corroborando su origen africano y fecha de muerte durante época colonial (Barquera y col., 2018), por lo que análisis similares en los individuos HSJNC81, HSJN240, HSJN275 permitirán identificar si los individuos HSJN240 y HSJNC81 nacieron en África o en México.

Se sabe que durante el contacto de nativo americanos con europeos hubo gran intercambio genético, incluyendo patógenos y existe evidencia histórica que asocia el traslado forzado de africanos con la introducción de enfermedades a América, como es el caso de la viruela (Mandujano y col., 2003). Después de una búsqueda minuciosa en la literatura reportada hasta ahora, no existen evidencias moleculares directas de patógenos virales traídos de África durante la época colonial de México.

En el caso del individuo COYC4 el genoma reconstruido de B19V presentó mayor similitud a genotipo 3; sin embargo, la ancestría genética inferida de este individuo no es africana, sino una mezcla entre componente americano moderno y asiático moderno, mostrado en Figura 36. Históricamente se han reportado infecciones por interacciones entre individuos del viejo mundo con nativo americanos (Mandujano y col., 2003), y debido a la alta presencia de africanos en el México colonial (Cuadro 1) (Velázquez y Iturralde, 2012), el individuo COYC4 pudo haber interactuado con parvovirus de genotipo 3, proveniente de África. Cabe mencionar que no se encontró ningún reporte de la presencia del genotipo 3 de B19V en México, después de un análisis exhaustivo de la literatura.

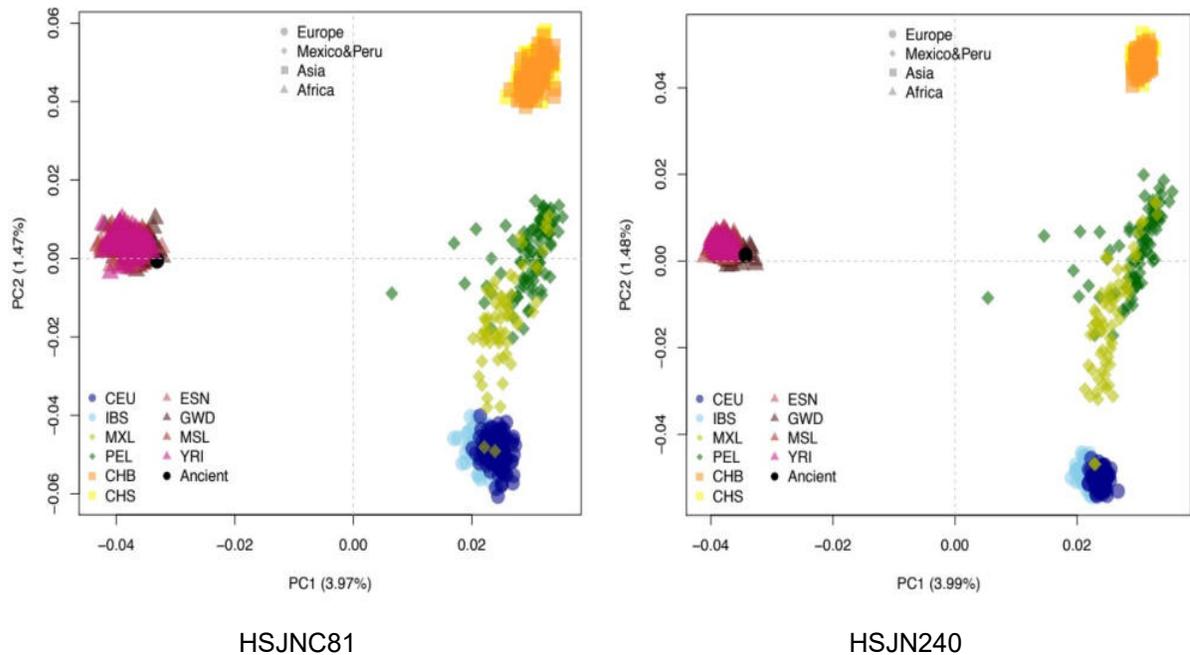


Figura 35. Análisis de componentes principales de HSJN respecto a panel de referencia 1000 genomas. Los individuos antiguos son representados con un punto negro. CEU: residentes de Utah; IBS: ibéricos de España; MXL: mexicanos en L.A.; PEL: peruanos de Lima; CHB: Han de Beijing; CHS: Han de sur de China; ESN: essan de Nigeria; GWD: gambianos de Gambia; MSL: mende de Sierra leona; YRI: yoruba de Nigeria.

Si bien, actualmente el genotipo 3 de B19V es endémico de Brasil, el origen de los genomas antiguos del HSJN en América es discrepante con la evidencia morfológica y genética de la ancestría de los hospederos humanos, que soporta un origen africano.

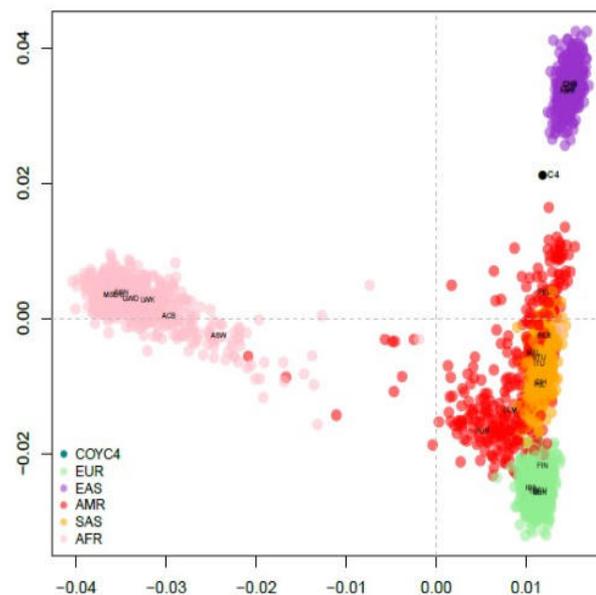


Figura 36. Análisis de componentes principales de componente humano COYC4 respecto a panel de referencia 1000 genomas. El individuo antiguo esta representado con etiqueta C4. 5 súper poblaciones incluidas; EUR: europeos; EAS: asiáticos; AMR: americanos; SAS: asiáticos del sur; AFR: africanos.

6.5.1 Lesiones osteológicas *post-mortem* de individuo HSJNC81

La secuencia ensamblada a partir de individuo 81 del HSJN tiene mayor cercanía con Parvovirus V9 (AJ249437) (Figura 29-31). V9 fue aislado a partir de un niño en Francia que presentó fiebre, anemia severa, insuficiente hemoglobina en eritroblastos y deficiencia de hierro (tinción con azul de Prusia), sintomatología que se asoció a la infección con V9 (Nguyen y col., 1999). Además de este individuo, los 3 genotipos de B19V se han encontrado comúnmente en individuos con anemia y/o hepatitis (Broliden y col., 2006). El cráneo 81 del HSJN presenta ligera a moderada hiperostosis porótica en la bóveda craneal, así como criba orbitalia en las cuentas oculares (identificadas por Julie K. Wesp, sin publicar). Estas lesiones

osteológicas se han asociado tradicionalmente a anemias genéticas (falciforme, talasemia); así como a posibles déficits nutricionales (falta de hierro) e infecciones parasitarias como malaria (Mansegosa y col., 2018).

De manera típica, las infecciones agudas transmitidas por vías sanguíneas no producen marcas osteológicas que sirvan como diagnóstico *post-mortem*; sin embargo, lesiones esqueléticas se han observado en individuos con enfermedades crónicas (Spyrou y col., 2019). No es posible determinar si la infección con B19V fue el agente causal de los signos de anemia presentes en el cráneo del individuo HSJNC81; sin embargo, si el individuo presentó anemia por otra causa, el efecto sinérgico de infección con B19V pudo ser severo para el individuo 81, ya que la infección de B19V (en precursores de la línea eritroide) tiene efecto potencializado (crisis aplásica por ejemplo) al presentarse a la par con padecimientos hematológicos como talasemia, anemia falciforme, anemia hemolítica, malaria o deficiencia de hierro (Broliden y col., 2006), por mencionar algunos.

6.6 Presencia de secuencias tipo HBV en colección HSJN

Se logró reconstruir dos genomas parciales de HBV (Cuadro 23) a partir de dientes de los individuos HSJN194 y HSJN275.

Todos los controles de secuenciación resultaron negativos para HBV, y sólo se pudo reconstruir genoma de HBV de dos individuos, a pesar que se capturaron mas librerías con las mismas sondas; por lo que, se descarta que su presencia se deba a contaminación ambiental o captura no específica.

La alta frecuencia de C>T (G>A complementaria) en extremos terminales de las lecturas usadas para reconstruir genoma de HBV a partir del individuo HSJN191, así como el tamaño corto de los fragmentos de DNA, corrobora la presencia de patrones de daño típicos de aDNA; sin embargo, no se identificó patrón similar en

genoma reconstruido a partir de HSJN275, en el que se puede observar alta frecuencia de diferencias respecto a genoma de referencia, mostrado en Anexo IX. Lo anterior sugiere que el mapeo de lecturas pertenecientes a HSJN275 se realizó a un genoma de referencia muy divergente o que simplemente no se obtuvieron suficientes lecturas pertenecientes a HBV.

La distribución de lecturas respecto al genoma de referencia no es uniforme, en ambas muestras positivas para HBV. Existen regiones sin cubrir, mientras otras regiones son cubiertas >10x (mostrado en Anexo IX). En el caso de secuenciación imparcial (sin enriquecimiento), se ha sugerido que las lecturas no pertenecen a especie mapeada (HBV), sino a regiones similares de otra especie (Warinner y col., 2017); sin embargo, en nuestro caso realizamos un ensayo de captura que pudo haber hibridado de manera heterogénea con aDNA de HBV, por lo que este patrón de cobertura puede ser un artefacto del ensayo de captura. Lo más similar (analizado con MegaBLAST/Krona) en la base de datos NR (NCBI febrero 2019) a las lecturas utilizadas para reconstrucción del genoma fue HBV (Figura 27). Los espacios vacíos respecto a la referencia coinciden con regiones de baja complejidad (mostrado en Anexo IX). Las regiones con muy bajo/alto porcentaje de GC necesitan temperaturas más altas/bajas para deshibridar (Hurst y Merchant, 2001) y permitir hibridación posterior con sondas de captura de ssRNA. Una vez hibridado el DNA con ssRNA es posible recuperar DNA con perlas magnéticas, por lo que una hibridación ineficiente de regiones de baja complejidad pudo haber evitado su recuperación con perlas magnéticas.

Al analizar los genomas reconstruidos de HBV respecto a genomas modernos y antiguos (Kahila y col. 2012; Ross y col., 2017; Krause-Kyora y col., 2018; Mühlemann y col., 2018) de HBV con herramientas filogenéticas (método de distancias y máxima verosimilitud), el genoma de HSJN275 se posiciona basal a secuencias modernas del genotipo A (Figura 32 y 33) y en medio de las secuencias

antiguas DA195 (Hungaria, 2479 años) y RISE387 (Rusia, 4215 años) del mismo genotipo (A), obtenidas por Mühlemann y col. (2018). Su posicionamiento basal al genotipo A (común en afrodescendientes) (Quintero y col., 2002; Anderchach y col., 2009; Roman y col., 2010; Brichlet y col., 2013; Facholi y col., 2008; Kramvis y col., 2013; Lago y col., 2014), en conjunto con la ancestría africana del hospedero humano (HSJN275), mostrada en Figura 37, sugiere origen antiguo. A pesar de esto, los métodos tradicionales (mapDamage) de determinación de patrones de daño no permiten corroborar su origen antiguo. No se probaron otros métodos menos citados para determinar el daño de las lecturas; como el método computacional reportado por Weiss y col. (2015), o el paquete de R “aRchaic” propuesto por Al-Asadi y col. (2018). que pudieran corroborar el origen antiguo del genoma reconstruido de HBV.

El genoma de HBV de HSJN194 es similar a secuencias del genotipo A4 de afrodescendientes (Figuras 32 y 33), y las lecturas utilizadas para su obtención presentan patrones de daño típicos de aDNA (Figuras 32 y 33).

Previamente se ha propuesto que el genotipo A fue introducido a América durante el traslado de personas esclavizadas africanas (Quintero y col., 2002; Facholi y col., 2008; Anderchach y col., 2009; Roman y col., 2010; Brichlet y col., 2013; Kramvis y col., 2013; Lago y col., 2014). Este modelo es coherente con la ancestría genética africana del hospedero (Figura 37) y con el origen de las secuencias más cercanas filogenéticamente (Figura 33) (GQ331047 y GQ331046), las cuales se obtuvieron a partir de individuos africanos (Congo y Rwanda) en Bélgica (Pourkarim y col., 2010).

El genoma de HBV obtenido a partir del individuo HSJN194 provee evidencia directa sobre la introducción del genotipo A a América con un origen muy cercano al periodo de traslado de personas esclavizadas africanas. Análisis adicionales de

$^{87}\text{Sr}/^{86}\text{Sr}$ y ^{14}C podrán determinar el lugar de nacimiento y fecha de muerte del individuo y en su caso, fortalecer nuestra hipótesis.

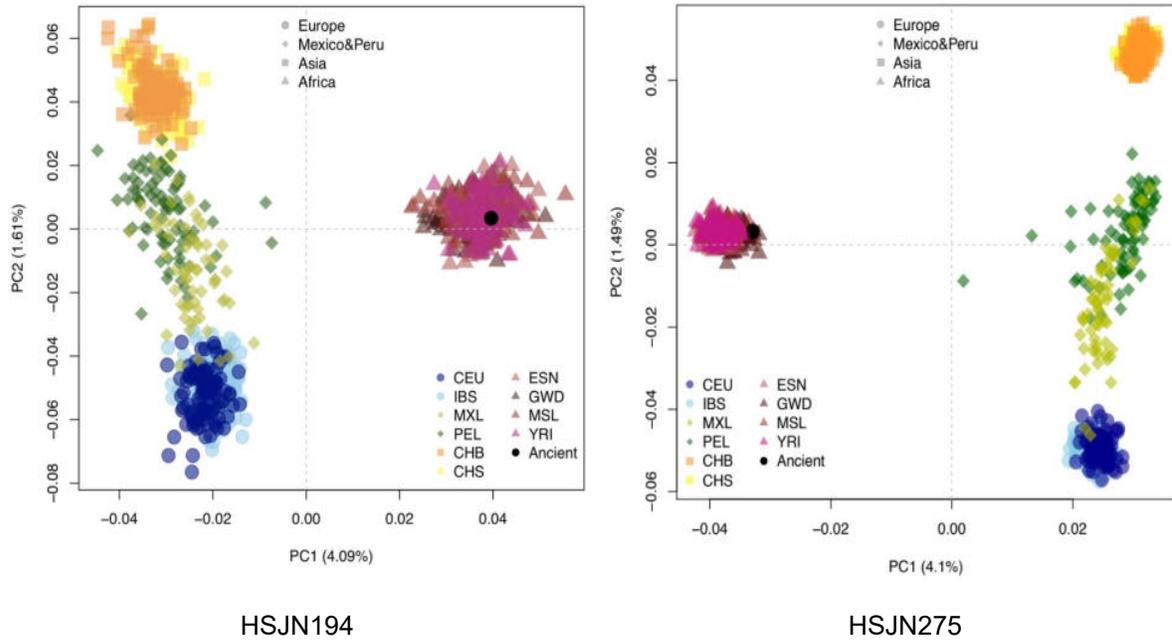


Figura 37. Análisis de componentes principales de HSJN respecto a panel de referencia 1000 genomas. Los individuos antiguos son representados con un punto negro y se usaron mismas etiquetas que en Figura 35. Analisis realizado por Dra. María C. Ávila Arcos.

6.6.1 Relatos de hepatitis e ictericia en HSJN durante cocoliztli de 1576

El HSJN fue un importante centro de atención medica para indígenas durante el periodo colonial de CMX (Meza, 2003). Existen relatos históricos directos sobre la epidemia de 1576, realizados por médicos de la época en este hospital. Uno de los mas importantes fue el explorador y protomédico Francisco Hernández, quien llevó

a cabo directamente una autopsia sobre individuo muerto por el cocoliztli de 1576 en el HSJN (Martínez-Hernandez, 2008), que se muestra a continuación:

“orinaban con micción abundante y pálida, partidos (en la autopsia) mostraban el hígado hinchado, el corazón negro, manado de un líquido pálido (amarillo) y después, sangre negra, el bazo y el pulmón negros y semi putrefactos; la atrabilis podía ser contemplada en su vasija, en vientre seco, y el resto del cuerpo, por cualquier parte que fuere cortado, palidísimo.”

Según Maldivo y Viesca (1982), Francisco Hernández relata:

“...pulsos frecuentes y rápidos, más pequeños y débiles; de vez en cuando hasta nulos. Los ojos y todo el cuerpo amarillo. Seguía delirio y convulsión, postemas detrás de una o ambas orejas, y tumor duro y doloroso, dolor de corazón...”

En el mismo hospital trabajó Alonso López de Hinojoso como cirujano y mayordomo del hospital, quien presencié y describió de la misma manera síntomas de hepatitis e ictericia de la misma epidemia de cocoliztli (Maldivo y Viesca, 1982), según Martínez-Hernández, 2008:

“...tenían los enfermos el hígado acirrado y muy duro, que se les paraba tan deforme que parecía hígado de toro y alzaba las costillas hacia arriba y hacía el pecho muy deforme, porque con su grandeza y tumor hacía monstruosidad. Los bofes o livianos tenían azules y secos; la hiel apostemada y opilada y muy grande; la cólera que dentro estaba se pudría y la cólera que quedaba fuera no podía estar dentro. Por esta causa se paraban los heridos de este mal muy amarillos y atiriciados...”

Martínez-Hernández (2008) describe a su vez el relato de Fray Agustín Dávila Padilla sobre una autopsia realizada por Juan de la Fuente en 1576, médico de la Inquisición y primer catedrático de medicina de la universidad mexicana:

“...hizo anatomía de vn indio en el Hospital Real de México: y aunque se le halló el hígado inflamado y con corrupción venosa de sangre, y advirtieron de allí en adelante con singularissimo cuidado, para poner el remedio adonde conocieron el daño...”

Si bien el cocoliztli corresponde a un cuadro clínico muy complejo, que no corresponde a un solo patógeno conocido a la fecha, se ha propuesto que era un efecto sinérgico causado por varios patógenos, entre los que se ha propuesto tifo exantémico (Mandujano y col., 2003), virus hemorrágicos, salmonela entérica (Vagene y col., 2018), e incluso Maldivo y Viesca (1982) han propuesto hepatitis epidémica. Respecto a esta última propuesta, parece que dentro de los relatos de sintomatología de cocoliztli del HSJN (mostrados previamente) es común la descripción de individuos ictericos y con hepatitis. La ictericia es causada por exceso de bilirubina, que es un producto de la degradación de hemoglobina. La bilirubina es transportada al hígado para unirse con la bilirrubina y ser deshechada. Si bien las causas de la ictericia son diversas, algunas son hepatitis inflamatorias (virales, inmunes), así como enfermedades que interfieren flujo de bilirrubina, o que ocasionan destrucción de glóbulos rojos (hemólisis), lo que aumenta cantidad de bilirubina en hígado (Herrine, 2016). La ictericia puede desencadenar cuadros clínicos graves, como cambios de estado mental (encefalopatía), hipertensión de venas que llevan la sangre al hígado (sangrado de esófago y a veces estomago) y/o coagulopatía (como hematomas, hemorragia digestiva, sangrados espontaneos) (Herrine, 2016), que concuerda con parte del cuadro clínico (hemorragias, delirios, hepatitis, ictericia) descrito para cocoliztli de 1576, más no por completo.

Aunque quizá nunca se conozcan los patógenos causantes de las epidemias del HSJN con certeza, la presencia de genomas antiguos africanos de HBV y B19V permite entender mejor la diversidad de patógenos virales a que los habitantes del México colonial estaban expuestos. Su importancia aumenta al tomar en consideración que las infecciones por HBV y B19V pueden tener cuadros de hepatitis, ictericia y problemas cardiacos (B19V) (Broliden y col., 2006); mismos cuadros que han sido descritos durante periodo colonial del HSJN (Martínez-Hernández, 2008).

7. CONCLUSIONES

No se encontró DNA perteneciente a VARV mediante PCR, metagenómica, o enriquecimiento con captura; ni de algún patógeno viral capaz de causar directamente las epidemias en los sitios del periodo colonial de CMX.

Se enriqueció DNA de HBV y se reconstruyeron dos genomas parciales pertenecientes al genotipo A (común en África), que es coherente con la ancestría africana (genética y morfológica) de los hospederos (HSJN194 y HSJN275) y la epidemiología moderna de HBV; sin embargo, sólo el de HSJN194 presenta patrones típicos de daño.

Se logró enriquecer DNA perteneciente a B19V en 3 individuos distintos del periodo colonial de CMX, a partir de los que se reconstruyeron tres genomas con patrones típicos de daño, lo que indica un origen antiguo.

Los genomas antiguos de B19V son similares al genotipo 3 (endémico en África) que es coherente con la ancestría africana (genética y morfológica) de dos de los hospederos humanos (HSJN).

Se reconstruyó un genoma antiguo de B19V a partir de un individuo mestizo (COY4), similar al genotipo 3, lo que sugiere infección entre individuos mestizos y africanos.

Se reconstruyeron los primeros genomas virales antiguos a partir de restos arqueológicos de América.

Se reconstruyó el primer genotipo 3 antiguo de B19V y representa el primer reporte de su presencia (moderna o antigua) en México.

Los genomas antiguos de B19V y HBV representan la primera evidencia molecular directa de patógenos virales traídos a América desde África, posiblemente durante el periodo de comercio de personas esclavizadas.

8. REFERENCIAS

1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. **2015**;526:68-74.

Acuna-Soto R, Stahle DW, Therrell MD, Griffin RD, Cleaveland MK. When half of the population died: the epidemic of hemorrhagic fevers of 1576 in México. *FEMS Microbiol Lett*. **2004**;240:1–5.

Aguirre-Beltrán G. La población negra en México: Estudio etnohistórico. 2a. edición. México: Fondo de Cultura Económica. **1972**:374.

Al-Asadi H, Dey KK, Novembre J, Stephens M. Inference and visualization of DNA damage patterns using a grade of membership model. *Bioinformatics*. **2018**:1-7.

Ali S, Baldanta S, Fernández-Escobaar M, Guerra S. Use of reporter genes in the generation of vaccinia virus-derived vectors. *Viruses*. **2016**;8:134.

Altschul SF, Gish W, Miller W, Myers E, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. **1990**;215:403-410.

Andernach IE, Nolte C, Pape JW, Muller CP. Slave trade and hepatitis B virus genotypes and subgenotypes in Haiti and Africa. *Emerg Infect Dis*. **2009**;15:1222-1228.

Ávila-Arcos MC, Sandoval-Velasco M, Schroeder H, Carpenter ML, Malaspinas AS, Wales N, Peñaliza F, Bustamante C, Gilbert T. Comparative performance of two whole-genome capture methodologies on ancient DNA Illumina libraries. *Methods Ecol Evol*. **2015**;6:725-734.

Barquera R, Bernal-Felipe N, Hernandez-Zaragoza D, Lamnidis T, Nelson E, Acuña-Alonso V, Marqués-Morfin L, Krause J. The genetic makeup of enslaved africans from early colonial México City. *Research Gate [serie en internet]* 2019 [consultado Abril 2019]; [aprox. 2 pp]. Disponible en: www.researchgate.net/publication/327816852_The_genetic_makeup_of_enslaved_Africans_from_early_Colonial_Mexico_City.

Biagini P, Thèves C, Balaesque P, Géraut A, Cannet C, Keyser K, Nikolaeva D, Gérard P, Chechesne S, Orlando L, Willerslev E, Ludes B. Variola virus in a 300-year-old siberian mummy. *N Engl J Med.* **2012**;367:2057-2059.

Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest S, Bryant J, Harris S, Shuenemann, Campbell T, Guichon R, Steadman W, Cook D, Behr M, Bastida R, Nieselt K, Parkhill J, Gagneux S, Stone A, Krause J. Pre-columbian mycobacterial genomes reveal seals as a source of new world human tuberculosis. *Nature.* **2014**;514:494–497.

Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coomber B, McPhee J, Meyer M, Schmedes S, Wood J, Herring D, Bauer P, Poinar H, Krause J. A draft genome of *Yersinia pestis* from victims of the black death. *Nature.* **2011**;478:506–510.

Brichler S, Lagathu G, Chekaraou MA, et al. African, Amerindian and European hepatitis B virus strains circulate on the Caribbean Island of Martinique. *J Gen Virol.* **2013**;94:2318-2329.

Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüffer K, Meyer M, Krause J, Ronan MT, Lachmann M, Pääbo S. Patterns of damage in genomic DNA sequences from a neandertal. *Proc Natl Acad Sci U S A.* **2007**;104:14616-14621.

Broliden K, Tolfvenstam T, Norbeck O. Clinical aspects of parvovirus B19 infection. *J Intern Med.* **2006**;260:285-304.

Buchfink B, Xie C, Huson D. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* **2015**;12:59-60.

Cantey JB, Pritchard MA, Sanchez PJ. Bone lesions in an infant with congenital parvovirus B19 infection. *Pediatrics.* **2013**;131:1659–63.

Castillo-Chavez O. Condiciones de vida y salud de una muestra poblacional de la ciudad de México en la época colonial. *Esc Nac Antropol e Hist.* **2000**:150-190.

Crosby AW. Virgin soil epidemics as a factor in the aboriginal depopulation in America. *William Mary Q.* **1976**;33:289–299.

Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* **2012**;29:1969-1973.

Duggan AT, Perdomo MF, Piombino-Mascoli D, Marciniak S, Poinar D, Emery M, Buchman J, Duchene S, Humphreys M, Southon J, Devault A, Sahl J, Dutour O, Smith G, Holmes E, Poinar H. 17th century variola virus reveals the recent history of smallpox. *Curr Biol.* **2016**;26:3407-3412.

Eid A. Osteomyelitis: historical review. *Pan Arab J Orthop Trauma.* **2003**; 7:95-109.
Eritema Infeccioso. Mayo Clinic [serie en internet] 2018 [consultado 2019 Mayo 19]: [aprox. 5 pp]. Disponible en: www.mayoclinic.org/es-es/diseases-conditions/parvovirus-infection/symptoms-causes/syc-20376085

Esposito JJ, Sammons SA, Frace AM, Osborne J, Olsen-Rasmussen, Zhang M, Govil D, Damon I, Laker M, Li Y, Smith G, Meyer H, Wohlhueter. Genome sequence diversity and clues to the evolution of VARV virus. *Science.* **2006**;322:807-812.

Facholi GB, Martins RMB, Motta-Castro ARC, Araujo, NM, Niel, C, Mello FCA. Molecular epidemiology of hepatitis B virus in an isolated afro-brazilian community. *Arch Virol.* **2008**;153:2197-2205.

Fornaciari G, Zavaglia K, Laura G, Claudia V, Rosalba C. Human papillomavirus in a 16th century mummy. *Lancet.* **2003**;362:1160.

García-Lascuráin AR. Códice Azcatitlan. *Arqueol Mex.* **2014**;54:28-31.

Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. mapDamage: Testing for damage patterns in ancient DNA sequences. *Bioinformatics.* **2011**; 27:2153-2155.

Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* **2016**;17:333-351.

Hansen HB, Damgaard PB, Margaryan A, Lynneruo, Allentoft M. Comparing ancient DNA preservation in petrous bone and tooth cementum. *PLoS One.* **2017**; 12:1-18.

Hendrickson RC, Wang C, Hatcher EL, Lefkowitz EJ. Orthopoxvirus genome evolution: the role of gene loss. *Viruses*. **2010**;2:1933-1967.

Herbig A, Maixner F, Bos KI, Zink A, Krause J, Huson DH. MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the tyrolean iceman. *bioRxiv*. **2016**:50559.

Ictericia. Herrine S [serie en internet] 2016 [consultado 2019 Mayo 19]_ [aprox. 7 pp]. Disponible en: <https://www.msmanuals.com/es-mx/professional/trastornos-hepáticos-y-biliares/aproximación-al-paciente-con-hepatopat%C3%ADas/ictericia>

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. Ancient DNA. *Nat Rev Genet*. **2001**;2:353-359.

Holmes E. What does virus evolution tell us about virus origins?. *J Virol*. **2011**;85:5247–5251.

Hübschen JM, Mihneva Z, Mentis AF, Schneider F, Aboudy, Y, Grossman Z, Sarv I, Tarnagda Z, Chitadze N, Forbi JC, Otegbayo JA, Charpentier E, Muller CP. Phylogenetic analysis of human parvovirus B19 sequences from eleven different countries confirms the predominance of genotype 1 and suggests the spread of genotype 3b. *J Clin Microbiol*. **2009**;47:3735-3738.

Huebler R, Key FM, Christina Warinner C, Boss KI, Krause J, Herbig A. HOPS: Automated detection and authentication of pathogen DNA in archaeological remains. *bioRxiv*. **2019**;14:36.

Hughes AL, Irausquin S, Friedman R. The evolutionary biology of poxviruses. *Infect Genet Evol*. **2010**;10:50-59.

Hughes AL. Origin and evolution of viral interleukin-10 and other DNA virus genes with vertebrate homologues. *J Mol Evol*. **2002**;54:90-101.

Hurst LD, Merchant AR. High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes. *Proc Biol Sci*. **2001**;268:493–497.

Huson DH, Beier S, Flade I, Górska A, Hadidi ME, Mitra S, Ruscheweyh HJ, Tappu R. MEGAN community edition, interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput Biol. **2016**;12:1-12.

Illumina Sequencing Platforms. Illumina [serie en internet] 2019 [consultado 2019 Mayo 19]: [aprox. 3 pp]. Disponible en: www.illumina.com/systems/sequencing-platforms.html

INAH. Asoman vestigios de desaparecido barrio indígena en el centro de la CDMX. Dir medios Comun. **2016**;186:3.

Inoue T, Tanaka Y. Hepatitis B virus and its sexually transmitted infection - an update. Microb Cell. **2016**;3:420-437.

Jahrling PB, Hensley LE, Martinez MJ, LeDuc J, Rubins K, Relman D, Huggins J. Exploring the potential of variola virus infection of cynomolgus macaques as a model for human smallpox. Proc Natl Acad Sci U S A. **2004**;101:15196-15200.

Jain A, Kant R. Genotypes of erythrovirus B19, their geographical distribution & circulation in cases with various clinical manifestations. Indian J Med Res. **2018**;147:239-247.

Janovitz T, Wong S, Young NS, Oliveira T, Falck-Pedersen E. Parvovirus B19 integration into human CD36+ erythroid progenitor cells. Virology. **2017**;511:40-48.

Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. MapDamage2.0: fast approximate bayesian estimates of ancient DNA damage parameters. Bioinformatics. **2013**;29:1682-1684.

Kahila GBG, Kim MJ, Klein A, Shin D, Oh Ch, Wan-Kim J, Kim T, Kim S, Grant P, Pappo O, Shouval D. Tracing hepatitis B virus to the 16th century in a korean mummy. Hepatology. **2012**;56:1671-1680.

Kay GL, Sergeant MJ, Giuffra V, Bandiera P, Milanese M, Bramanti B, Bianucci R, Pallen M. Recovery of a medieval *Brucella melitensis* genome using shotgun metagenomics. mBio. **2014**;5:1337-1414.

Key FM, Posth C, Krause J, Herbig A, Bos KI. Mining metagenomic data sets for ancient DNA: Recommended Protocols for Authentication. *Trends Genet.* **2017**; 10:1-13.

Kircher M. Analysis of high-throughput ancient DNA sequencing data. *Methods Mol Bio.* **2014**;840:197-228.

Kramvis A, Paraskevis D. Subgenotype A1 of HBV - Tracing human migrations in and out of Africa. *Antivir Ther.* **2013**;18:513-521.

Krause-Kyora B, Susat J, Key FM, Kühnert D, Bosse E, Immel A, Rinne C, Kornell SC, Yepes D, Franzenburg S, Heyne HO, Meier T, Lösch S, Meller H, Friedrich S, Nicklisch N, Alt K, Schreiber S, Tholey A, Herbig A, Nebel A, Krause J. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *eLife.* **2018**;7,1-15.

Kuczynski J, Stombaugh J, Walters WA, Gonzalez A, Caporaso JG, Knight R. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. *Curr Protoc Bioinformatics.* **2011**;10:7.

Lago B V., Mello FC, Kramvis A, Niel C, Gomes SA. Hepatitis B virus subgenotype A1: Evolutionary relationships between Brazilian, African and Asian isolates. *PLoS One.* **2014**;9.

Lesbats P, Engelman AN, Cherepanov P. Retroviral DNA Integration. *Chem Rev.* **2016**;116:12730-12757.

Lesnik EA, Freier SM. Relative thermodynamic stability of DNA, RNA, and DNA: RNA hybrid duplexes: Relationship with base composition and structure. *Biochem.* **1995**;34;10807-10815.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics.* **2009**;25:2078-2079.

Li H. and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* **2009**;25:1754-1760.

Li H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*. **2014**;30:2843-2851.

Li Y, Carroll DS, Gardner SN, Walsh MC, Vitalis EA, Damon IK. On the origin of smallpox: correlating variola phylogenics with historical smallpox records. *Proc Natl Acad Sci U S A*. **2007**;40:15787-15792.

Lindahl T. Instability and decay of the primary structure of DNA. *Nature*. **1993**;362:709-715.

Lindo J, Huerta-Sánchez E, Nakagome S, Rasmussen M, Mitchel JB, Cybulski JS, Willerslev E, DeGiorgio M. A time transect of exomes from a native american population before and after european contact. *Nat Commun*. **2016**;7:13175.

Lobo, I. Basic local alignment search tool (BLAST). *Nature Education*. **2008**;1(1):215,

Locarnini S, Littlejohn M, Aziz MN, Yuen L. Possible origins and evolution of the hepatitis B virus (HBV). *Semin. Cancer Biol*. **2013**;23:561–575.

Lorenz T. Polymerase Chain Reaction: Basic protocol plus troubleshooting and optimization strategies. *J Vis Exp*. **2012**;63:3998.

Luo Y, Qiu J. Human parvovirus B19: A mechanistic overview of infection and DNA replication. *Future Virol*. **2015**;10:155-167.

Maixner F, Krause-Kyora B, Turaev D, Herbig A, Hoopmann MR, Hallows J, Kusebauch U, Megraud F, Cipollini G, Coia V, Samadelli M, Linz B, Moritz RL, Frimm Rm Krause K, Nene A, Moodley Y, Rattei T, Zink A. The 5300-year-old *Helicobacter pylori* genome of the iceman. *Science*. **2016**;351:162–65.

Maldiva E, Viesca C. INAH [serie en internet] 1982. [consultado 2019 Mayo 20]: [5 pp] Disponible en: https://www.estudioshistoricos.inah.gob.mx/revistaHistorias/wp-content/uploads/historias_11_24-33.pdf
www.estudioshistoricos.unah.gob.mx/revistaHistorias/wp-content/uploads/historias_11_24_33.pdf

Mandujano S A, Camarillo S L, Mandujano MA. Historia de las epidemias en el México antiguo. *Casa del tiempo* [serie en internet]. **2003**. [consultado 2017

noviembre 30]:9-21. Disponible en:
<http://www.uam.mx/difusion/revista/abr2003/mandujano.pdf>.

Mansegosa DA, Chiavazza H, Giannotti PS. Estrés metabólico y entierros secundarios: nuevas evidencias para el estudio de las poblaciones coloniales del centro-oeste argentino. *Rev Argentina Antropol Biológica*. **2018**;20:5.

Margaryan A, Hansen HB, Rasmussen S, et al. Ancient pathogen DNA in human teeth and petrous bones. *Ecol Evol*. **2018**:1-9.

McMahon BJ. The influence of hepatitis B virus genotype and subgenotype on the natural history of chronic hepatitis B. *Hepatology*. **2009**;3:334-342.

Metsky HC, Siddle KJ, Gladden-Young A, Qu J, Yang QD, Brehio P, Goldfarb A, Piantadosi A, Wohl S, Lin Bozza F, Souza T, Isern S, Michael SF, Tully D, Garcia K, Parham LA, Odia I, Eromon P, Goba A, Simon-Loriere E, Balmaseda A, Kwon D, Allen T, Runstadler JA, Lorenzana I, Gehrke L, Ebel G, Happi C, Gnirke A, Sabeto CP, Matranga CB. Capturing diverse microbial sequence with comprehensive and scalable probe design. *Nat Biotechnol*. **2019**;37:160-168.

Meza A. Las raíces africanas de México. *Arqueol Mex*. **2013**;119.

Moreira D, López-García P. Ten reasons to exclude viruses from the tree of life. *Nat Rev Microb*. **2009**;7:306-311.

Mühlemann B, Jones TC, Damgaard PB, Allentoft ME, Shevnina I, Logvin A, Boldgiv B, Bazartseren T, Tashbaeva K, Merz V, Irina P, Price TD, Hansen AJ, Orlando L, Vinner L, Smith DJ, Glebe D, Fouchier RAM, Kristiansen K, Willerslev E. Christian ancient hepatitis B viruses from the bronze age to the medieval period. *Nature*. **2018**;557:418-423.

Mühlemann B, Margaryan A, Barros P De, Allentoft ME, Vinner L, Hansen AJ, Weber A, Molak M, Bogdanowicz W, Sablin M, Sten S, Tashbaeva K, Sikora M, Smith DJ, Fouchier RAM, Drosten C, Kristiansen K, Willerslev E, Jones TC. Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans evolution. *Proc Natl Acad Sci U S A*. **2018**;115:7557-7562.

Nayfach S, Rodriguez-Mueller B, Garud N, Pollard K. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Res.* **2016**;26:1612–1625.

Nguyen QT, Sifer C, Schneider V, Allaume X, Servant A, Bernaudin F, Auguste V, Garbarg-Chenon A. Novel human erythrovirus associated with transient aplastic anemia. *J Clin Microbiol.* **1999**;37:2483-2487.

Olinger CM, Venard V, Njyou M, et al. Phylogenetic analysis of the precore/core gene of hepatitis B virus genotypes E and A in West Africa: New subtypes, mixed infections and recombinations. *J Gen Virol.* **2006**;87:1163-1173.

Orlando L, Gilbert MTP, Willerslev E. Reconstructing ancient genomes and epigenomes. *Nat Rev Genet.* **2015**;16.

Ortner DJ. Identification of pathological conditions in human skeletal remains. *Sci Dir.* **2003**;2:250-300.

Osiowy C, Gordon D, Borlang J; Giles E, Villeneuve JP. Hepatitis B virus genotype G epidemiology and co-infection with genotype A in Canada. *J Gen Virol.* **2008**;89:3009-3015.

Paabo S, Poinar H, Serre D, Deprés V, Hebler J, Rohland N, Kuch M, Krause J, Hofreiter M. Genetic analyses from ancient DNA. *Annu Rev Genet.* **2004**;38:645-679.

Paabo S. Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci U S A.* **1989**;86:1939-1943.

Pajer P, Dresler J, Kabíckova H, Pisa L, Aganov P, Fucik K, Elleder D, Hron R, Velemínský P, Fucikova A, Pejchal J, Benes V, Rausch T, Dundr P, Pilin A, Cabala R, Stribny J, Meyer H. Characterization of two historic smallpox specimens from a Czech museum. *Viruses.* **2017**;9:1-14.

Paraskevis D, Magiorkinis G, Magiorkinis E, Ho SY, Belshaw R, Hatzakis A. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology.* **2013**;57:908-916.

Park RW. Contact between the Norse Vikings and the Dorset culture in Arctic Canada. *Antiquity*. **2008**;82:189-198.

Parsyan A, Szmaragd C, Allain JP, Candotti D. Identification and genetic diversity of two human parvovirus B19 genotype 3 subtypes. *J Gen Virol*. **2007**;88:428-431.

Pedersen JS, Valen E, Velazquez A, Parker B, Rasmussen M, Lindgreen S, Lilje B, Tobin D, Kelly T, Vang S, Andersson R, Jones P, Hoover C, Tikhonov A, Rubin E, Sandellin A, Gilbert T, Krough, Orlando L. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res*. **2014**;24:454–466.

Pourkarim MR, Lemey P, Amini-Bavil-Olyae S, Maes P, Van Ranst M. Novel hepatitis B virus subgenotype A6 in african-belgian patients. *J. Clin. Virol*. **2010**:47;93–96.

Price TD, Tiesler V, Burton JH. Early african diaspora in colonial Campeche, Mexico: Strontium isotopic evidence. *Am J Phys Anthropol*. **2006**;490:485-490.

Pyoria L, Toppinen M, Mäntylä E, et al. Extinct type of human parvovirus B19 persists in tonsillar B cells. *Nat Commun*. **2017**;8.

Qiu J, Söderlund-Venermo M, Young NS. Human Parvoviruses. *Clin. Microbiol. Rev*. **2017**;30:43–113.

Quintero A, Martínez D, Alarcón De Noya B, et al. Molecular epidemiology of hepatitis B virus in afro-venezuelan populations. *Arch Virol*. **2002**; 147:1829-1836.

Roche. KAPA HiFi Uracil+ Kit. [serie en internet]. **2018**. [consultado 2018 junio 30]; Disponible en: <https://sequencing.roche.com/en/products-solutions/by-category/library-preparation/library-amplification/kapa-hifi-uracil-kit/specifications.html>

Rohland N, Hofreiter M. Ancient DNA extraction from bones and teeth. *Nat Protoc*. **2007**;2:1756-1762.

Roman S, Tanaka Y, Khan A, Kurbanov F, Kato H, Mizokami M, Panduro A. Occult hepatitis B in the genotype H-infected Nahuas and Huichol native Mexican population. *J Med Virol.* **2010**;82:1527-36.

Ross ZP, Klunk J, Fornaciari G, Giuddra V, Duchene S, Duggan A, Poinar D, Eden S, Holmes E, Poinar H. The paradox of HBV evolution as revealed from a 16th century mummy. *PLOS Pathog.* **2018**;14:1-25

Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Rasheid KAS, Willerslev E, Krogh A, Orlando L. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics.* **2012**;13:178.

Schuenemann VJ, Kumar Lankapalli A, Barquera R, et al. Historic *Treponema pallidum* genomes from colonial Mexico retrieved from archaeological remains. *PLoS Negl Trop Dis.* **2018**;12.

Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jager G, Bos K, Herbig A, Benjak A, Busso P, Boldsen J, Stewart G, Taylor M, Bauer T, Lee O, Tucker K, Roffey S, Sow S, Cole S, Nieselt K, Krause J. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science.* **2013**;341:179–83.

Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods.* **2012**;9:811-814.

Smith O, Clapham A, Rose P, Liu Y, Wang J, Allaby RG. A complete ancient RNA genome: Identification, reconstruction and evolutionary history of archaeological Barley stripe mosaic virus. *Sci Rep.* **2014**;4:1-6.

Smithson C, Imbery J, Upton C. Re-assembly and analysis of an ancient variola virus genome. *Viruses.* **2017**;9:253.

Soares CC, de Pina-Araujo IIM, Gomes SA, Spitz N, Lago B V., Niel C. Hepatitis B virus genotypes A1, A2 and E in Cape Verde: Unequal distribution through the islands and association with human flows. *PLoS One.* **2018**;13.

Spyrou MA, Bos KI, Herbig A, Krause J. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat Rev Genet.* **2019**;1.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* **2014**;30:1312-1313.

Swiss Institute of Bioinformatics. *Hepadnaviridae*. ViralZone [serie en internet] 2009. [consultado 2019 Junio 1]: [2 pp] Disponible en: https://www.estudioshistoricos.inah.gob.mx/revistaHistorias/wp-content/uploads/historias_11_24-33.pdf www.viralzone.expasy.org/9

Taubenberger JK. Characterization of the 1918 influenza virus polymerase genes. *Nature.* **2005**;437,889–893.

Thèves C, Biagini P, Crubézy E. The rediscovery of smallpox. *Clin Microbiol Infect.* **2014**;20:210-218.

Tiee MS, Harrigan RJ, Thomassen HA, Smith TB. Ghosts of infections past: using archival samples to understand a century of monkeypox virus prevalence among host communities across space and time. *R Soc Open Sci.* **2018**;5:171089.

Tithi S, Frank O, Aylward F, Jensen RV, Zhang L. FastViromeExplorer: a pipeline for virus and phage identification and abundance profiling in metagenomics data. *PeerJ.* **2018**;6:4227.

Toppinen M, Perdomo MF, Palo JU, Simons P, Lycett S, Venermo M, Hedman K. Bones hold the key to DNA virus history and epidemiology. *Sci Rep.* **2015**;5:1-8.

Tringe S, Rubin E, Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet.* **2005**;6:805-815.

Tu T, Budzinska MA, Shackel NA, Urban S. HBV DNA integration: Molecular mechanisms and clinical implications. *Viruses.* **2017**;9.

Vagene AJ, Herbig A, Campana MG, et al. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat Ecol Evol.* **2018**;2:1-9.

Valencia-Pacheco G., Nakazawa Ueji YE, Angulo-Ramírez AV, López-Villanueva RF, Quintal-Ortiz IG, Rosado Paredes EP. Serological and molecular analysis of

parvovirus B19 infection in Mayan women with systemic lupus erythematosus in Mexico. *Colom Med.* **2017**;48:105-113.

Valero-García AR. *Codice Azcatitlan. Arqueo Mex.* **2014**;52:28-31.

Velázquez ME, Iturralde-Nieto G. *Afrodescendientes en México. 1ª Edición. Ciudad de México: Editorial Conaculta,* **2012**:37-47.

Velsko IM, Frantz LAF, Herbig A, Larson G, Warinner C. Selection of appropriate metagenome taxonomic classifiers for ancient microbiome research. *mSystems.* **2018**;3:4

Wanzeller A, Souza AL, Azebedo R, Filho LC, Oliveira R, Junior JV, Vasconcelos PV. Complete genome sequence of the BeAn 58058 virus isolated from *Oryzomys sp.* rodents in the amazon region of Brazil. *Genome Announc.* **2017**;5:1575-16.

Warinner C, García NR, Spores R, Tuross N. Disease, demography, and diet in early colonial New Spain: Investigation of a sixteenth-century mixtec cemetery at Teposcolula Yucundaa. *Lat Am Antiq.* **2012**;23:467-489.

Warinner C, Herbig A, Mann A, Yates J, Weiss C, Burbano H, Orlando L, Krause J. A Robust framework for microbial archaeology. *Annu Rev Genomics Hum Genet.* **2017**;18.

Warinner, C, Matias-Rodriguez J, Vyas R, Traschel C, Shved N, Radini A, Tito RI, Speller C, Hendy J, Charlton S, Luder HU, Eppler E, Seiler R, Teoh KY, Olsen JV, Nanni P, Willerslev E, Gilbert T, Capellini E. Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **2014**;46:336–344.

Watashi K y Wakita T. Hepatitis B virus and hepatitis D virus entry, species specificity, and tissue tropism. *Cold Spring Harb Perspect Med.* **2015**;5:1-10.

Waterhouse A, Procter JB, Martin DM, Clamp M, Barton G. Jalview Version 2, a multiple sequence alignment editor and analysis workbench. *Bioinform.* **2009**; 25:1189-1191.

Weiss CL, Dannemann M, Prüfer K, Burbano HA. Contesting the presence of wheat in the British Isles 8,000 years ago by assessing ancient DNA authenticity from low-coverage data. *Elife*. **2015**;4:1-10.

Wesp JK. Caring for bodies or simply saving souls: the emergence of institutional care in spanish colonial America. *New Dev Bioarchaeology Care*. **2017**:253-276.

Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol*. **2014**;15:46.

World Health Organization. Assessment of future scientific needs for live variola virus. The National Academies Press [serie en internet] 1999 [consultado 2017 diciembre 1]:46-61. Disponible en: <https://doi.org/10.17226/6445>.

Worobey M, Watts T, McKay R, Granade T, Teuwen D, Koblin B, Lemey P, Jaffe H. 1970 and 'patient 0' HIV-1 genomes illuminate early HIV/AIDS history in north America. *Nature*. **2016**;539:98–101.

Yuen LW, Littlejohn M, Duchêne S, et al. Tracing ancient human migrations into Sahul using hepatitis B virus genomes. *Mol Biol Evol*. **2019**; 21.

Yuen M, Chen DS, Dusheiko GM, Janssen H, Lau DT, Locarnini S, Peters M, Lai C. Hepatitis B virus infection. *Nat. Rev*. **2018**;392:2313–2324.

ANEXOS

Anexo I. Lista de principales programas y comandos utilizados

1. samtools view -bh -f 4 Archivo.mapeado.bam > Archivo.nomapeado.bam
2. for next in \$(cat Indice_NCBI.txt); do wget -P Genomas_virales "\$next"/*genomic.fna.gz; done
3. malt-build -i Genomas_virales -s DNA -d Genomas_virales_formateado -c Taxonomy
4. malt-run -m BlastN -at SemiGlobal -i Archivo.nomapeado.bam.fastq -d Genomas_virales_formateado -o . -id 90
5. samtools view Archivo.mapeado.bam | cut -f1,10 | sed 's/^>/' | sed 's/\t/n/' > Lecturasmapeadas_bwa.fasta
6. blastn -db DS1.fa -query Muestra.fasta -out Muestra.blastn -evaluate 0.001 -outfmt '6 qaccver saccver pident length mismatch gapopen qstart qend sstart send evaluate bitscore qcovs pident'
7. awk '{ if (\$13 >= 85) { print } }' Muestra.blastn | awk '{print\$1}' | uniq
8. sh RmDup.sh Muestra_Bwa_Blastn_DS1.fa > Muestra_Bwa_Blastn_DS1.rmdup.fa. Disponible en gist.github.com/ShaiberAlon/ddcaf32b63f7072670d74206541f1131.
9. blastn -db nr -query Muestra_Bwa_Blastn_DS1.rmdup.fa -remote -out Lecturas.especificas.blastn -outfmt 6 -import_search_strategy Estrategia.asn
10. ktlImportBLAST Lecturas.especificas.blastn
11. samtools mpileup -uf Referencia.fa Archivo.rescaled.bam | bcftools call -c | vcfutils.pl vcf2fq > Consensus_rescaled.fastq
12. mapDamage -i LecturasMapeadas.filtradas.bam -r GenomaViral.fasta --rescale --rescale-out LecturasMapeadas.filtradas.rescaled.bam

```
13.raxmlHPC -s Alineamiento.phy -m GTRGAMMAI -# 1000 -x 12345 -n  
NombreArbol.tree -k -f a -p 12345
```

Anexo II. Lista de genomas disponibles en NCBI de VARV.

Los siguientes números de acceso se incluyeron para obtener genes de interés y alinearlos con el fin de construir genes consenso incluidos en el diseño de sondas para ensayo de captura.

BK010317, DQ437580, DQ437581, DQ437582, DQ437583, DQ437584,
DQ437585, DQ437586, DQ437587, DQ437588, DQ437589, DQ437590,
DQ437591, DQ437592, DQ441416, DQ441417, DQ441418, DQ441419,
DQ441420, DQ441421, DQ441423, DQ441424, DQ441425, DQ441426,
DQ441427, DQ441428, DQ441429, DQ441430, DQ441431, DQ441432,
DQ441433, DQ441434, DQ441435, DQ441436, DQ441437, DQ441438,
DQ441439, DQ441440, DQ441441, DQ441442, DQ441443, DQ441444,
DQ441445, DQ441446, DQ441447, DQ441448, L22579, LT706528, LT706529,
NC_001611, X69198, Y16780.

Anexo III. Lista de genes de interés de familia *Herpesviridae*.

Los siguientes genes fueron incluidos en diseño de ensayo de captura.

Virus	Gene	Tamaño
Herpes virus simple 1	UL30	3747
	UL31	1003
	UL19	5506
	UL27	2857
	US6	2860
	UL10	1565
Cytomegalovirus	UL54	3729
	UL53	2253
	UL86	4113
	UL115	837
	UL75	2229
	UL83	1686
Epstein bar virus	ORF9	5698
	ORF69	971
	ORF25	5971
	ORF47	2571
	ORF8	8482
	vIRF2	2571
	K5	771

Anexo IV. Lista de genomas usados (DB2) para generar CDS de B19V incluidos en análisis filogenéticos.

AB030673, AB030693, AB030694, AB126262, AB126263, AB126264, AB126265, AB126266, AB126267, AB126268, AB126269, AB126270, AB126271, AB550331, AF113323, AF162273, AJ249437, AJ717293, AJ781038, AY028237, AY044266, AY064475, AY064476, AY083234, AY386330, AY504945, AY582124, AY582125, AY647977, AY903437, DQ225148, DQ225149, DQ225150, DQ225151, DQ234769, DQ234771, DQ234775, DQ234778, DQ234779, DQ293995, DQ333426, DQ333428, DQ357064, DQ357065, DQ408301, DQ408302, DQ408303, DQ408304, DQ408305, EF216869, FJ265736, FJ591158, FN598217, FN598218, KC013305, KC013308, KC013312, KC013313, KC013314, KC013316, KC013321, KC013322, KC013324, KC013325, KC013327, KC013329, KC013331, KC013332, KC013333, KC013338, KC013340, KC013343, KC013344, KC013346, KC013351, KF724386, KF724387, KM065414, KM065415, KM393163, KM393164, KM393165, KM393166, KM393167, KM393168, KM393169, KR005640, KR005641, KR005642, KR005643, KR005644, KT268312, KT310174, KX752821, M13178, M24682, NC_000883, NC_004295, Z68146, Z70528, Z70560, Z70599.

10 genomas antiguos (Mühlemann y col., 2018), disponibles en:

<https://github.com/acorg/parvo-2018/tree/master/consensuses>

DA251-user, DA336, DA337n, DA66, NEO105, RISE569, VK143, VK154, VK477, VK6.

Anexo V. Genomas utilizados para análisis de HBV.

DB1: Genomas utilizados para mapeo competitivo de HBV.

HE974381, HE974383, AY934764, GQ331046, AB602818, AB033555, AB073835, AB287316, AB241117, AB111946, X75656, AB048704, AF241411, AP011102, AP011106, AP011108, AB644287, FJ899792, GQ477453, JN688710, HE974373, FJ904430, AB033559, HE974384, AY090458, AB116654, AY090455, DQ899144, HE974369, AB116549, AF223962, AB166850, AP007264, AB516395, AB486012, LT992443, LT992459, AF046996.

DB2: Genomas utilizados para análisis filogenéticos.

a) Genotipos modernos A-H.

AB076679, AB116084, AB453988, AY738142, GQ477499, AY934764, FJ692556, FJ692598, FJ692611, GQ161813, GQ331046, AB073858, AB033555, AB219429, AB219430, AP011089, AB073835, AB287316, AB287318, AB287320, AB287321, DQ463789, DQ463792, AB241117, DQ993686, AB111946, AB112066, AB112472, DQ089767, X75656, X75665, AB048704, AB048705, AF241411, AP011100, AP011102, AP011103, AP011106, AP011108, FJ899792, JN642140, GQ477453, GQ477455, JN642160, JN642163, JN688710, JN688711, GQ922005, HE974378, KJ470893, KJ470896, KJ470898, FJ904430, FJ904436, AB033559, AB048701, AB048702, AB188243, AB210818, AM494716, AY796031, AY902768, DQ315779, X80925, X75657, X75664, AY090458, AB116654, FJ657525, AY090455, AY311369, DQ899144, DQ899146, AB116549, X75663, AF223962, AB166850, AB056513, AB064312, AF405706, AB059660, AB375163, AY090454, AY090457, AB486012.

b) 17 genomas antiguos reportados por:
149

Kahila y col., 2012:

JN315779

Ross y col., 2018:

MG585269

Mühlemann y col., 2018:

LT992439, LT992442, LT992444, LT992440, LT992441, LT992454, LT992455,
LT992459, LT992448, LT992447, LT992443

Krause-Kyora y col. 2018 (compartidos directamente por autores):

Petersberg, Sorsum, Karsdorf.

c) 8 genomas de HBV de primates no humanos (gibon, gorila y chimapance):

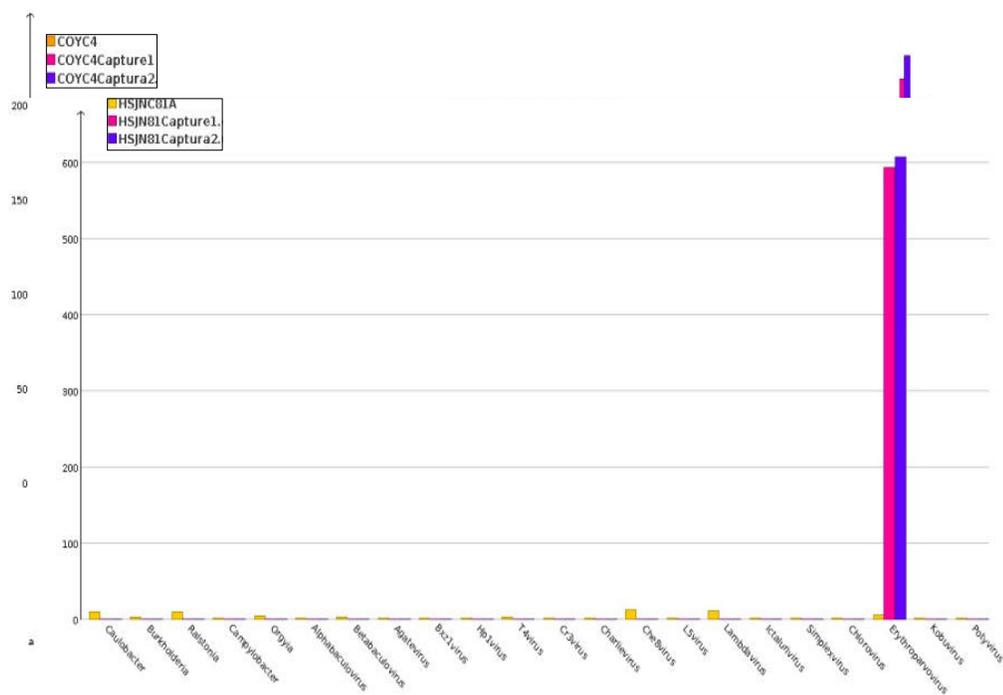
AF222323, AB032433, AY330911, FM209516, U46935, AY781180, AF193863,
EU155824

Anexo VI. Análisis metagenómico viral.

A continuación se muestra análisis metagenómico viral de colecciones COY y HSJN, respectivamente. La comparación de abundancias se realizó en base al número de lecturas asignadas a cada genero y normalizadas en MEGAN.

Anexo VII. Comparación de abundancias virales antes y después de captura.

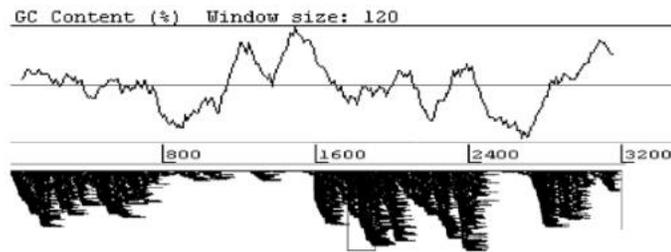
Se normalizaron abundancias por genero viral en MEGAN, y se compararon sus asignaciones taxonómicas antes y después de captura para cada individuo.



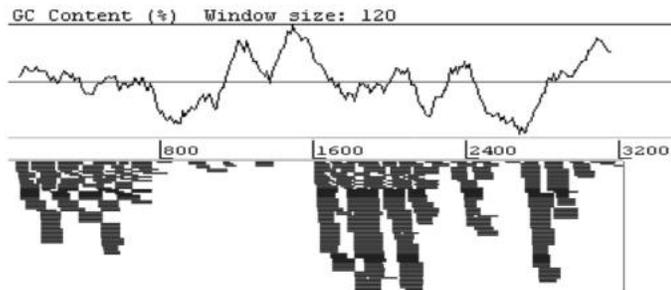
Comparación de abundancias relativas virales para HSJNC81

Anexo IX. Cobertura de genoma de HBV y patrones de daño

En la siguiente figura se muestra la captura de pantalla del mapeo a HBV usando lecturas filtradas de HSJN194 y HSJN275 (respectivamente) en Tablet (cada rectángulo negro representa una lectura de secuenciación). Se incluyó contenido de GC calculado con Artemis.



Distribución de cobertura respecto a HBV de HSJN194



Distribución de cobertura respecto a HBV de HSJN275

A continuación se muestra el patrón de daño de lecturas obtenidas a partir del individuo HSJN275 y filtradas respecto a genoma de HBV.

