

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO

FACULTAD DE QUÍMICA

**“ANÁLISIS COMPARATIVO DE LA RECUPERACIÓN DE
ADN ENDÓGENO DE MOLARES DE MAMUTS
(*Mammuthus columbi*) EXCAVADOS EN SANTA LUCÍA,
ESTADO DE MÉXICO”**

TESIS INDIVIDUAL

QUE PARA OBTENER EL TÍTULO DE

INGENIERA EN BIOTECNOLOGÍA

PRESENTA

ÁNGELES TAVARES GUZMÁN

DIRIGIDA POR

DRA. MARÍA DEL CARMEN ÁVILA ARCOS

SANTIAGO DE QUERÉTARO, QUERÉTARO, 2022.



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO

FACULTAD DE QUÍMICA

**“ANÁLISIS COMPARATIVO DE LA RECUPERACIÓN
DE ADN ENDÓGENO DE MOLARES DE MAMUTS
(*Mammuthus columbi*) EXCAVADOS EN SANTA
LUCÍA, ESTADO DE MÉXICO”**

TESIS INDIVIDUAL

**QUE PARA OBTENER EL TÍTULO DE
INGENIERA EN BIOTECNOLOGÍA**

PRESENTA

ÁNGELES TAVARES GUZMÁN

DIRIGIDA POR

DRA. MARÍA DEL CARMEN ÁVILA ARCOS

SINODALES

DRA. MARÍA DEL CARMEN ÁVILA ARCOS Mary Carmen Ávila Arcos
DIRECTORA

DRA. MONSERRAT ESCAMILLA GARCÍA M. Escamilla
ASESORA INTERNA

DRA. LAURA CRISTINA BERUMEN SEGURA L. Berumen
SINODAL

DR. FEDERICO ANDRÉS SÁNCHEZ QUINTO Federico Sánchez Quinto
SINODAL

AGRADECIMIENTOS

La culminación de este trabajo es también el final de una de las etapas más importantes de mi vida hasta ahora. No hubiera llegado hasta aquí si no hubiera sido por mi familia, mis amigos, mis profesores y mis asesores. Este proyecto lo hice pensando en Ángeles pequeña que le encantaban los dinosaurios y hojeaba el tomo ilustrado de mamíferos de la enciclopedia azul.

Este trabajo forma parte del proyecto de investigación interdisciplinario “La Prehistoria y Paleoambiente del noroeste de la Cuenca de México” registrado ante el INAH. La toma de muestra, extracción y secuenciación de ADN, y análisis genéticos correspondientes como parte de este proyecto paleogenómico, fueron autorizados mediante el permiso correspondiente emitido por parte del Consejo de Paleontología con número de oficio 401.1S.3-2020/098 del 3 de junio de 2020. Agradecemos a todos los miembros del equipo interdisciplinario al igual que al personal del INAH su invaluable ayuda para la realización de este proyecto. Particularmente agradecemos la ayuda del Dr. Joaquín Arroyo Cabrales y el P. Arqueol. Alejandro López Jiménez asociados al INAH, así como al Biol. Omar Bravo Flores de la Universidad Michoacana, por su crucial asistencia durante la toma de muestra orgánica de las muestras paleontológicas.

Quiero agradecer a toda la comunidad del Laboratorio Internacional de Investigación sobre el Genoma Humano (LIIGH), donde realicé este trabajo, a Luis Alberto Aguilar y Jair García del Laboratorio Nacional de Visualización Científica Avanzada (LAVIS) de la Universidad Nacional Autónoma de México. A Carina Uribe y a Alejandra Castillo, las técnicas, por salvar mis experimentos más de una vez. También quiero agradecer al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), que otorgó los fondos para realizar este proyecto. A la Dra. Mashaal Sohail por su apoyo con la secuenciación.

A la Dra. María Ávila, por abrirme las puertas de su laboratorio cuando estaba apenas a la mitad de la carrera, por permitirme conocer el mundo la paleogenómica y por construir un espacio seguro para mí y todos los estudiantes

de su laboratorio. También quiero agradecerle por apoyarme y por haberme ofrecido trabajar en este proyecto, y por siempre haberse preocupado por poner mi salud mental primero. Es de las personas más inspiradoras y trabajadoras que conozco, sin ella no sería la profesional que soy ahora.

Al Dr. Federico Sánchez, por haberme aceptado y por confiar en mí para trabajar en conjunto y desarrollar este proyecto. Por ser tan abierto y tan empático conmigo, y por enseñarme lo que significa la colaboración en la ciencia. Por preocuparse por mis avances y siempre estar pendiente. Sin su apoyo y su compromiso conmigo no habría concluido este trabajo.

A la comunidad del PopEvoGen Lab. A la Dra. Miriam, por enseñarme todo lo que sé de métodos experimentales, por toda la paciencia y los consejos para el registro de mis experimentos, por todas las conversaciones en el laboratorio. A la Mtra. Viridiana Villa y a la Dra. Marcela Sandoval por todo su apoyo para realizar mis experimentos. A Arumi, y a Miguel Alejandro, por hacer mis tardes llevaderas y felices en el laboratorio.

A la UAQ y a la Facultad de Química, por haberme formado y enseñado lo que la vida universitaria significa. Por haberme brindado el conocimiento para desarrollarme como profesionista. A mis profesores y profesoras, especialmente a la Dra. Talina, David Gustavo, y al Dr. Gracida.

A mis sinodales la Dra. Montserrat Escamilla y la Dra. Laura Berumen, por apoyarme tanto desde que me conocieron en el salón, por sus excelentes clases y por su inmenso apoyo para sacar adelante esta tesis a pesar de la burocracia y la adversidad.

A mis amigos de la fac, porque sin su alegría no hubiera vivido la universidad como lo hice. A Armandito por escuchar mis quejas y por estar en el camino del biotec conmigo desde que salimos de la prepa. En especial quiero agradecer a Luis y a Fritz, porque ellos me motivaron a entrar al LIIGH, y por su ayuda incondicional durante toda la carrera y después de ella. A Miguel, por tanto amor, por tanta escucha, por acompañarme y darme la manita cada que sentía que no

lo lograba. A Paula, mi terapeuta, por ayudarme a mantener la mente en la tierra y por recordarme que todo va a estar bien, aunque parezca que no.

A mis abuelazos Cecilia y Cesáreo, a mis tías Crystal y Zorayda por todo su apoyo. Y lo más importante, quiero agradecer a mi padre, a mi madre y a mi hermana, a quienes dedico con todo mi amor este trabajo. Sin ellos no sería quien soy. Sé que para traerme a donde estoy tuvieron que enfrentar muchas cosas y realizar muchos sacrificios, y sé que muchas veces me pusieron por delante para que pudiera cumplir mis objetivos. Espero que estén orgullosos de mí.

ÍNDICE GENERAL

ÍNDICE GENERAL	i
ÍNDICE DE FIGURAS	iii
ÍNDICE DE CUADROS	iv
RESUMEN	
1. ANTECEDENTES	1
1.1. El género <i>Mammuthus</i> y la paleogenómica	1
1.2. <i>Mammuthus columbi</i> en México	4
1.3. Material osteológico de mamuts	5
1.4. Retos y dificultades del trabajo con ADN	7
1.5. Extracción de ADN	9
1.6. Construcción de librerías genómicas	10
1.7. Secuenciación de ADN y procesamiento de resultados.	11
1.8. Importancia de los mamuts de Santa Lucía en el panorama paleogenómico mexicano	12
2. HIPÓTESIS	13
3. OBJETIVOS	14
3.1. General	14
3.2. Específicos	14
4. METODOLOGÍA	15
4.1. Materiales	15
4.1.1. Muestras osteológicas	15
4.1.2. Oligonucleótidos	15
4.1.3. Reactivos	15
4.2. Métodos	16
4.2.1. Muestreo	16
4.2.2. Extracción de ADN	18
4.2.3. Limpieza de Reservorio Zymo spin	18
4.2.4. Reacción de lisis	18
4.2.5. Extracción de ADN	18

4.2.6. Elución de ADNa	19
4.2.7. Cuantificación de ADNa	19
4.2.8. Construcción de librerías	20
4.2.8.1. Preparación de tampón BST	20
4.2.8.2. Reparación de extremos	20
4.2.8.3. Ligación de adaptadores.	22
4.2.8.4. Rellenado de espacios.	23
4.2.9. Reacción de qPCR	23
4.2.10. Indexado de librerías.	25
4.2.11. Manejo de residuos	26
4.2.12. Secuenciación de nueva generación	27
4.2.13. Cálculo de contenido endógeno	27
5. RESULTADOS	29
5.1. Obtención de muestras	29
6. DISCUSIÓN	49
6.1 Retos para obtener contenido endógeno de muestras antiguas	50
6.2 Extracción de ADN	52
6.3 Construcción de librerías genómicas	53
6.4 Preservación del ADN en dentina	55
6.5 Sexo de los individuos	60
6.6 Condiciones climáticas y del suelo	60
6.7 Preservación de ADNa	64
6.8 Importancia y perspectivas de este estudio	65
7. CONCLUSIONES	66
8. BIBLIOGRAFÍA	67
ANEXOS	80

ÍNDICE DE FIGURAS

Figura		Página
1	Esquema de molar de mamut	7
2	Dificultades del trabajo con ADN	8
3	Nomenclatura para muestras procesadas	29
4	Fotografías de restos fósiles muestreados en Santa Lucía	30
5	Fotografías de restos fósiles muestreados en Santa Lucía (continuación)	31
6	Gráfico de caja sobre la relación entre contenido endógeno y tipo de material	39
7	Gráfico de caja del contenido endógeno en relación al sexo molecular	40
8	Diagrama de dispersión de la relación entre el porcentaje de contenido endógeno y la profundidad mínima de excavación	41
9	Gráfico de residuales para la relación entre porcentaje de contenido endógeno y profundidad mínima de excavación	42
10	Diagrama de dispersión de la relación entre el porcentaje de contenido endógeno y la profundidad máxima de excavación	43
11	Gráfico de residuales para la relación entre porcentaje de contenido endógeno y profundidad máxima de excavación	44
12	Porcentaje de contenido endógeno en relación a la concentración de ADN obtenida mediante Qubit	45
13	Gráfico de residuales para la relación entre porcentaje de contenido endógeno y el ADN cuantificado mediante Qubit	46
14	Porcentaje de contenido endógeno en relación a la concentración de ADN obtenida mediante BioAnalyzer	47
15	Gráfico de residuales para la relación entre porcentaje de contenido endógeno y el ADN cuantificado mediante BioAnalyzer	48
16	Patrones de daño <i>post-mortem</i> en ADN	50
17	Proceso de mineralización del tejido óseo	56

ÍNDICE DE CUADROS

Cuadro		Página
1	Oligonucleótidos para construcción de librerías	15
2	Componentes de reacción de reparación de extremos	21
3	Componentes de reacción de ligación de adaptadores	22
4	Componentes de reacción de rellenado de espacios	23
5	Componentes de reacción de qPCR	24
6	Programa de reacción de qPCR	25
7	Componentes de reacción de indexado de librerías	25
8	Programa de PCR para indexado de librerías	26
9	Concentración de ADN obtenido de las muestras	32
10	Estadísticas de secuenciación de las muestras	34
11	Estadísticas de secuenciación de las muestras (continuación)	35
12	Resultados del sexamiento de los individuos muestreados	36
13	Profundidades de excavación de muestras de mamut	37

RESUMEN

Mammuthus es un género extinto de elefantes que aparecieron en África hace aproximadamente 5 millones de años. En México, el registro de *Mammuthus columbi* existe en todo el territorio, excepto en la Península de Yucatán. Excavaciones en el área de la Base Aérea Militar de Santa Lucía, en el Estado de México, han revelado la existencia de restos óseos de al menos 80 mamuts, que representan una fuente de conocimiento invaluable. El estudio de los mamuts en México ha sido amplio en el campo paleontológico, pero existen cuestiones sobre su historia evolutiva y diversidad genética que no han sido resueltas. Para obtener datos paleogenómicos de calidad, es crítico que los métodos experimentales se realicen de manera precisa y que se conozca el rendimiento y la calidad del ADN extraído de las muestras, es decir, el porcentaje de material endógeno. En este trabajo se comparó el rendimiento del porcentaje de contenido endógeno de las extracciones de ADN de mamut de acuerdo a distintas variables (sexo del individuo, profundidad de excavación, tipo de material). No se observó una correlación determinante con ninguna de estas variables, por lo que este trabajo destaca la importancia de aumentar el conocimiento sobre los factores que influyen en la preservación del ADN en los molares.

1. ANTECEDENTES

1.1. El género *Mammuthus* y la paleogenómica

Dentro de la diversidad de megafauna en el registro fósil, los mamuts destacan como los más icónicos. *Mammuthus* es un género extinto de elefantes, pertenecen a la familia Elephantidae y al orden Proboscidea. Aparecieron en África hace aproximadamente 5 millones de años (Ma) y abarcaron un amplio rango geográfico durante el Pleistoceno, expandiéndose desde Eurasia hasta Norteamérica e inclusive algunas partes de Centroamérica (Bravo-Cuevas y col., 2020; van der Valk y col., 2021). De acuerdo al registro fósil, se ha sugerido que aparecieron durante el Mioceno tardío en África, y posteriormente migraron en dirección a Eurasia, y eventualmente Norteamérica entre el Plioceno medio y el Pleistoceno temprano (Chang y col., 2017; Lister, 1996; Lister y Sher, 2015).

La evolución de este género durante el Pleistoceno suele describirse como una sucesión de especies que se sobrelapan temporalmente: *Mammuthus meridionalis* (mamut del sur), *M. trogontherii* (mamut de las estepas), *M. primigenius* (mamut lanudo) y *M. columbi* (mamut colombino) (Chang y col., 2017; Lister y col., 2005; Lister y Sher, 2015).

En cuanto a su presencia en Norteamérica, Lister y Sher sugieren que los primeros mamuts en el continente fueron producto de migraciones de *M. trogontherii* hace aproximadamente 1Ma (Lister y Sher, 2015). Por otro lado, el registro fósil sugiere que *M. primigenius* llegó mucho después a Norteamérica, hace ~125K años antes del presente (AP). Los mamuts desaparecieron del registro fósil hace alrededor de 15-11K años AP, al final de la última glaciación (Lister y Sher, 2015), excepto aquellos de la isla de Wrangel (una subpoblación de *M. primigenius*), en el océano Ártico, que desaparecieron del registro hace aproximadamente 4K años.

Las causas de la extinción de este género han sido objeto de debate, abarcando el cambio climático y la sobrecaza por parte de los humanos, entre otros factores, sin embargo, no se ha llegado a una resolución clara. Así pues, en conjunto con

otras disciplinas, el área del ADN antiguo (ADNa) ha tenido gran interés en el estudio de la historia evolutiva y las posibles causas de la extinción de los mamuts. Además, cabe destacar que *M. primigenius* fue el primer organismo antiguo del cual se obtuvo el genoma nuclear completo (Miller y col., 2008), suceso que representó la transición del ADNa hacia la paleogenómica.

En un estudio de 2018, Palkopoulou y col. analizaron los genomas de 14 proboscídeos extintos y modernos, los cuales incluyeron cuatro elefantes africanos; dos del bosque (*Loxodonta cyclotis*) y dos de la sabana (*Loxodonta africana*), dos elefantes asiáticos (*Elephas maximus*), un elefante extinto “de defensas rectas” (*Palaeoloxodon antiquus*), cuatro mamuts lanudos, un mamut colombino y dos mastodontes americanos (*Mammuth americanum*). En este estudio se sugiere que durante la evolución de la familia Elephantidae la hibridación interespecie pudo haber sido recurrente, contrario a lo que se había modelado antes, pues sólo se utilizaban árboles bifurcantes. Además de esto, se muestra evidencia de que hubo mestizaje entre *M. primigenius* y *M. columbi*, probablemente siguiendo un gradiente latitudinal a través de Norteamérica (Palkopoulou y col., 2018).

En el estudio realizado por Chang y col. (2017), se analizaron los mitogenomas de 143 mamuts obtenidos de fósiles muestreados a lo largo de su hábitat en la región holártica, región que comprende Europa, el norte de Asia y África, América del Norte y Groenlandia (Morrone, 2008). Los análisis filogenéticos realizados revelaron que la población de mamuts durante el Pleistoceno tardío estaba compuesta por tres linajes mitocondriales geográficamente estructurados a nivel continental, los cuales divergieron hace aproximadamente 1-2 millones AP. Además, se sugiere que los patrones de variación morfológica entre mamuts tienen que ver más con el flujo génico dependiente de los machos más que con expansiones geográficas a gran escala de ambos sexos. Esto debido a que la mayoría de las hembras presentaban una tendencia a la filopatría, al contrario de los machos, quienes solían expandirse geográficamente.

Un estudio realizado por (Pečnerová y col., 2017) es otro ejemplo de cómo los datos paleogenómicos pueden utilizarse para conocer mejor este género extinto: en éste se determinó el sexo cromosómico de 98 *M. primigenius* de Siberia, incluyendo la isla de Wrangel, para elucidar características del comportamiento y estructura social de los mamuts durante sus últimos 60 mil años de existencia. En este estudio se encontró que los machos estaban sobrerrepresentados (69% vs 31%), lo cual se interpretó como que los machos eran más propensos a morir en trampas naturales, probablemente debido a la estructura social de los proboscídeos, caracterizado por jerarquías matriarcales en las que los machos suelen separarse de las manadas. Este comportamiento sigue vigente en proboscídeos actuales (Lee y Moss, 1999; Vidya y Sukumar, 2005), por lo que los autores sugieren que esta característica es un rasgo ancestral de su comportamiento.

Recientemente, en un estudio publicado por (van der Valk y col., 2021) se reportó la recuperación de los genomas de tres mamuts presentes en Siberia del este durante el Pleistoceno temprano e intermedio, denominados Adycha, Krestovka y Chukochya. Los primeros dos fueron datados con más de 1 Ma de antigüedad, y fueron clasificados como *M. trogontherii*, mientras que Chukochya se clasificó como *M. primigenius* de aproximadamente 650,000 años de antigüedad. Gracias a los análisis filogenéticos realizados, los autores pudieron reportar que durante el Pleistoceno temprano existían dos linajes de mamuts, uno que dio origen a *M. primigenius*, y otro que no había sido reportado hasta el momento, representado por el espécimen Krestovka. Este último, además, es ancestral a *M. columbi*. De igual forma, infirieron que el origen de *M. columbi* es el producto de la hibridación entre *M. primigenius* y el linaje de Krestovka, confirmando el modelo que implicaba que los primeros mamuts de Norteamérica derivan de un ancestro similar a *M. trogontherii* (Lister y Sher, 2015) y no de una expansión de *M. meridionalis* al continente.

1.2. *Mammuthus columbi* en México

Durante el Pleistoceno tardío, tres familias de los proboscídeos cohabitaron el territorio mexicano: Gomphoteriidae, Mammutidae y Elephantidae (Pérez-Crespo y col., 2012). Esta última familia fue representada por el género *Mammuthus* con el mamut colombino (*M. columbi*), considerado endémico del continente americano (Enk y col., 2016). Esta especie estuvo distribuida a lo largo del territorio mexicano, pues se cuenta con su registro en casi toda la República, excepto en Tabasco y la Península de Yucatán (Bravo-Cuevas y col., 2020), siendo la cuenca de México el sitio con mayor abundancia (Arroyo-Cabrales y col., 2007).

Excavaciones desde 1980, en conjunto con las obras de reconversión de la Base Aérea Militar de Santa Lucía (Aeropuerto Internacional Felipe Ángeles AIFA), en el Estado de México, han revelado la existencia de restos óseos de al menos 80 mamuts (Graham, 2020), con más de 7,000 huesos. Los restos pertenecen a los niveles estratigráficos llamados Santa Lucía I y Santa Lucía II, fechados con ^{14}C a partir de materia orgánica en sedimentos entre 23.9K-26.3K años AP respectivamente (Pérez-Crespo y col., 2012). Por otro lado, en la localidad de San Antonio Xahuento, perteneciente al municipio de Tultepec se han encontrado 1,300 restos óseos de mamuts, los cuales representan al menos 16 individuos de un nivel estratigráfico inferior (Tultepec II). Del nivel estratigráfico superior (Tultepec I) no se han recuperado restos.

A pesar de la abundancia de restos de mamuts en México, la mayoría de los estudios se han concentrado en la descripción e identificación de especímenes, así como la caracterización morfológica de los restos que se han encontrado en los diversos sitios de excavación (Bravo-Cuevas y col., 2020). Asimismo, existen estudios de su caracterización taxonómica y su relación espacio-temporal con otros proboscídeos durante el Pleistoceno (Arroyo-Cabrales y col., 2007). Por otro lado, también se han intentado responder preguntas sobre su ecología (Arroyo-Cabrales y col., 2001) y se ha estudiado la dieta de estos animales por medio de

estudios con isótopos estables y análisis de microdesgaste (Bravo-Cuevas y col., 2020; Pérez-Crespo y col., 2012).

Aunque el estudio de los mamuts ha sido muy amplio en el aspecto paleontológico, existen cuestiones sobre su historia evolutiva que no han sido abordadas todavía. Realizando los análisis paleogenómicos adecuados, será posible determinar, por ejemplo, las afinidades genéticas de los mamuts, así como la estructura social de las poblaciones, y evidenciar la diversidad genética de las poblaciones de dichos animales en el Noreste de la Cuenca de México (NOCMex). Además de esto, cabe destacar que el origen de la especie *M. columbi* no ha sido del todo aclarado (van der Valk y col., 2021), por lo que los datos paleogenómicos, en conjunto con el conocimiento generado a través de los análisis paleontológicos anteriormente mencionados serían de gran utilidad para proponer nuevos modelos sobre su origen y evolución en el continente americano, en especial en la NOCMex, así como su relación genética con los proboscídeos contemporáneos.

La paleogenómica es un área del conocimiento que apenas está sentando sus bases en México, por lo que es difícil encontrar instalaciones adecuadas para realizar los experimentos necesarios. Afortunadamente, en el Laboratorio Internacional de Investigación sobre el Genoma Humano (LIIGH) se cuenta con la infraestructura y la experiencia necesarias para realizar los procedimientos experimentales que la paleogenómica requiere.

1.3. Material osteológico de mamuts

El ADN puede obtenerse de una variedad de fuentes, por ejemplo, huesos, dientes, pelo, semillas, cálculo dental, tejido suave conservado, sedimentos, entre otros (Orlando y col., 2021). No obstante, el material mineralizado suele estar mejor preservado que los tejidos suaves, además de que es más abundante, por lo que los estudios de ADN se han enfocado en dicho material, especialmente en algunos huesos y dientes (Orlando y col., 2021). En cuanto a los últimos, la

dentina suele preferirse para análisis genéticos, debido a que, además de que permite recuperar el ADN del individuo, también puede preservar el ADN de los patógenos que estuvieron presentes en el torrente sanguíneo (Orlando y col., 2021).

Los dientes de los vertebrados, aunque forman parte del esqueleto difieren del resto de los huesos por sus patrones de desarrollo, modificación a través de su uso y su eventual pérdida. Todos los tejidos del diente (dentina, esmalte y cemento) se forman por aposición, en la cual se añade capa tras capa en la localización inicial de la mineralización, con poca o nula resorción subsecuente (Fisher, 2018). La resorción se refiere al proceso mediante el cual los osteoclastos eliminan la matriz del hueso y liberan dichos minerales (Quiles y col., 2020). La dentina se añade en capas hacia la cavidad de la pulpa y la raíz, formando la corona y las raíces respectivamente (Fisher, 2018), dicha estructura se puede observar en la Figura 1. En cuanto al esmalte, este se añade en capas hacia afuera desde el primer locus de formación de dentina, y hacia abajo de los flancos de la corona. A excepción de las raíces, cada diente tiene un tamaño definitivo. Una vez alcanzado este tamaño, las modificaciones se dan por atrición (choque entre los dientes por la masticación), abrasión (con interacción con la comida, arena ingerida o sustratos externos), por ruptura, o por resorción de las raíces (Fisher, 2018).

En cuanto a los proboscídeos, de acuerdo a la nomenclatura de Saunders (McDaniel y Jefferson, 2006), la dentición molariforme consiste en tres premolares deciduos (Dp2, Dp3, Dp4),, seguidos de tres molares (M1, M2 y M3) en cada cuadrante (izquierdo, derecho, superior e inferior) de la cavidad oral (Fisher, 2018).

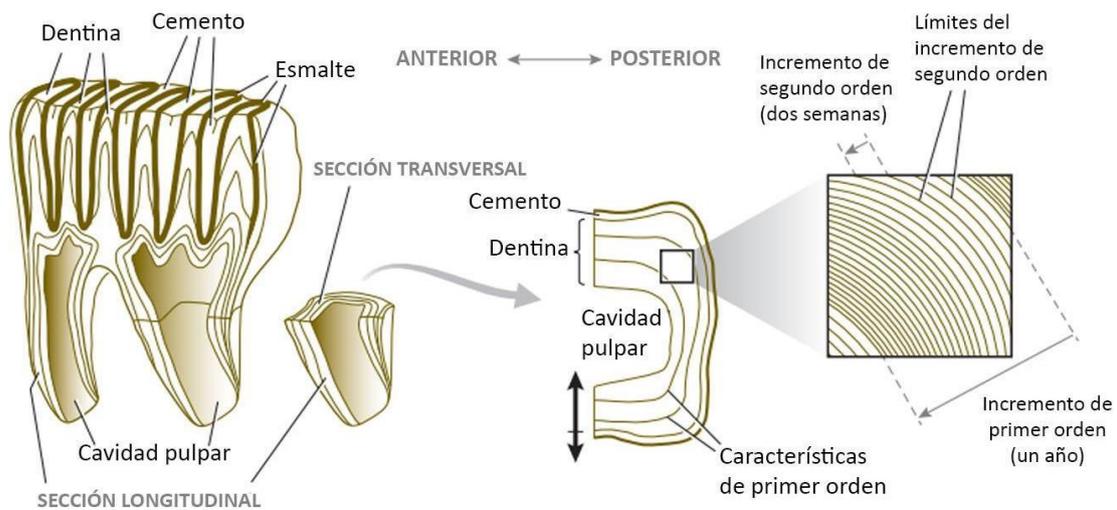


Figura 1. Representación esquemática de premolar dP3 de proboscídeo, de la familia Elephantidae. Las flechas en color negro señalan la dirección de aposición de dentina y cemento. Este diente tiene una corona compuesta de esmalte (en color sepia sólido), y crestas de dentina cubiertas por cemento. Figura adaptada de Fisher, 2018.

1.4. Retos y dificultades del trabajo con ADN

Los estudios con ADN permiten conocer el pasado genético de los organismos para esclarecer sus historias evolutivas, lo que no sería posible usando ADN moderno o estudios paleontológicos únicamente. No obstante, el manejo del ADN implica ciertas dificultades.

Después de la muerte de un organismo, el material genético comienza a degradarse y dañarse debido, en parte, a que las vías moleculares de reparación del ADN dejan de funcionar. El ADN que logra preservarse suele estar altamente fragmentado, esto debido a la depuración hidrolítica, seguida por reacciones de β -eliminación que provoca la ruptura del esqueleto del ADN (Orlando y col., 2021). Dependiendo de las condiciones del ambiente (i.e. temperatura, humedad), los fragmentos de ADN pueden estar entre 36-300 pb. Esto es relevante, ya que en zonas tropicales, donde hay mayor temperatura y humedad, la degradación del

ADN será mayor, al contrario de lo que sucede en lugares con clima frío (Shapiro y Hofreiter, 2012).

Además de esto, las hebras de ADN suelen tener varias modificaciones químicas. Una de ellas es que el daño hidrolítico provoca roturas de cadena sencilla, especialmente en los extremos (Hofreiter y col., 2001). La hidrólisis también puede provocar otro tipo de daño, por ejemplo, la desaminación de citosina a uracilo (Shapiro y col., 2019), esto provoca que cuando las moléculas de ADN se procesan para generar librerías para su secuenciación, la ADN polimerasa lea el uracilo como timina en lugar de la citosina original (Hofreiter y col., 2001). Ya que el cambio de base de citosina a timina es el más común y detectable, se utiliza como un estándar de autenticación del ADN.

Finalmente, otra de las características del ADN es que la proporción de moléculas de ADN pertenecientes al organismo (endógeno) suele ser baja en comparación con el ADN ambiental que se recupera de la muestra (Damgaard y col., 2015).

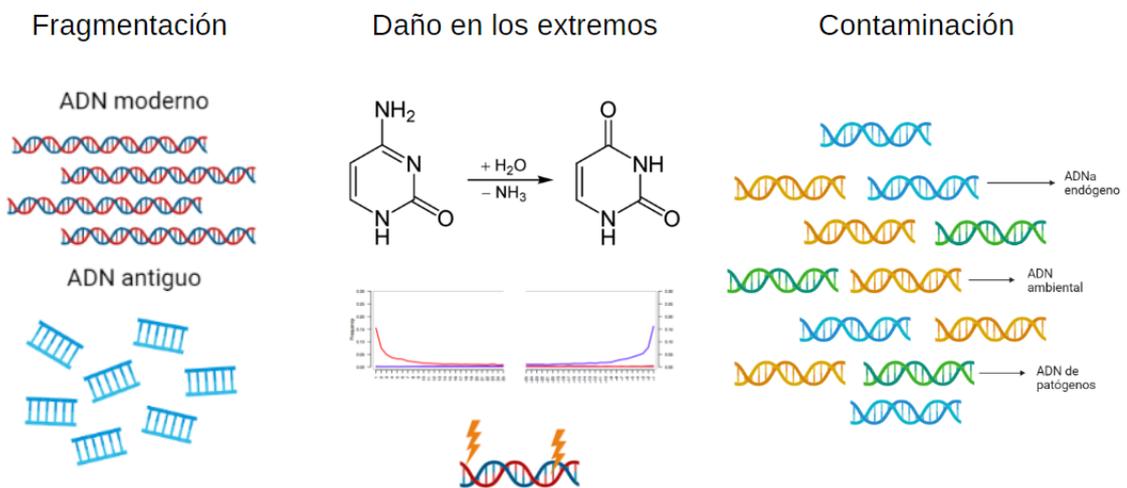


Figura 2. Los retos de trabajar con ADN antiguo incluyen la fragmentación, la característica desaminación de purinas que provoca que las citosinas se lean como timinas en los extremos, y la contaminación, que puede provenir de distintas fuentes. Imagen realizada en BioRender.com.

1.5. Extracción de ADN

La extracción de ADN se refiere al proceso de recuperación de esta molécula, al separarlo del resto de los componentes celulares, que pueden ser membranas, proteínas, u otros (Kobun, 2021). Tradicionalmente, los procesos de extracción de ADN involucran tres pasos básicos: lisis de las células, liberación del ADN de las proteínas unidas a éste y de los compartimientos celulares, y por último, la remoción de sales por medio de su precipitación con etanol o isopropanol (Green y Sambrook, 2018).

Por otro lado, los protocolos de extracción de ADN se han enfrentado a una serie de dificultades debido a la naturaleza de los especímenes. En primer lugar, estos ya no cuentan con estructuras celulares, y las modificaciones químicas del ADN puede dificultar que esté en solución (Rohland y col., 2018). También, debido a lo dañado de este material se deben evitar tratamientos con temperaturas demasiado altas o el uso de detergentes agresivos (Rohland y col., 2018), además de que los contaminantes presentes en los extractos pueden ser inhibidores de la Taq polimerasa (Höss y Pääbo, 1993).

En la mayoría de los protocolos de extracción de ADN, la muestra debe reducirse a polvo por medios mecánicos y después el ADN es liberado de la matriz mineral al incubar la muestra en tampones que descalcifican dicha matriz, además de que degradan proteínas y lípidos, y rompen las interacciones del ADN con otros compuestos orgánicos e inorgánicos en un proceso llamado digestión (Orlando y col., 2021). Posteriormente se realiza el paso de purificación, el cual tiene un número de variantes. Aunque se han desarrollado métodos basados en la extracción con fenol-cloroformo (Wales y col., 2014; Wales y Kistler, 2019), métodos no destructivos (Harney y col., 2021; Hofreiter, 2012; Rohland y col., 2004), y también se han utilizado membranas de ultrafiltración como complemento (Schwarz y col., 2009), la mayoría utiliza la capacidad adsorbente de la sílica para llevar a cabo la purificación. En estos protocolos, la sílica puede estar en solución (Allentoft y col., 2015; Höss y Pääbo, 1993; Rohland y Hofreiter, 2007), en un

tampón de unión caotrópico, o en perlas magnéticas recubiertas de sílica (Glocke y Meyer, 2017; Rohland y col., 2018), así como en columnas de centrifugación (Boessenkool y col., 2017; Dabney y col., 2013; Gamba y col., 2016; Yang y col., 1998).

Gracias a estos avances metodológicos, ha sido posible recuperar fragmentos de ADN de por lo menos 35 pb (Dabney y col., 2013), y hasta 25 pb (Glocke y Meyer, 2017), esto es importante ya que si no se recuperaran dichos fragmentos, se perdería información y, en consecuencia, los análisis paleogenómicos perderían resolución en los resultados.

Para extraer ADN de mamuts, se ha reportado el uso del método fenol-cloroformo (Enk y col., 2016; Poinar y col., 2006; Schwarz y col., 2009), también han sido utilizados los métodos que requieren sílica en columnas de centrifugación (Brace y col., 2012; Chang y col., 2017; Miller y col., 2008; Palkopoulou y col., 2015; van der Valk y col., 2021), en especial los métodos de Yang y col. (1998), Dabney y col. (2013) y Rohland y Hofreiter (2007). Es importante mencionar que no siempre los métodos publicados por los autores son seguidos al pie de la letra, muchas veces son modificados tomando en cuenta la muestra o las necesidades de la investigación, por ejemplo, se pueden modificar los tampones de lisis y unión, el tipo de columna que se utiliza, etc (Brace y col., 2012; Palkopoulou y col., 2015).

1.6. Construcción de librerías genómicas

Actualmente existen dos metodologías disponibles para secuenciación por síntesis, que es la principal manera de NGS (secuenciación de nueva generación por sus siglas en inglés) de ADN: 1) preparación de librerías de doble cadena y 2) de cadena sencilla.

Para el método de cadena doble, los extremos de las moléculas de aDNA son reparados y después adaptadores son ligados a dichos extremos. Por otro lado, en el método de cadena sencilla, las moléculas de aDNA y los adaptadores son

ligados en estado de desnaturalización, por lo que dan como resultado librerías de cadena sencilla (Orlando y col., 2021). Posterior a esto, se debe llevar a cabo el indexado de dichas librerías, que consiste en ligar fragmentos cortos de secuencias conocidas de ADN a los extremos de cada molécula, lo cual permite la secuenciación simultánea de múltiples muestras (Orlando y col., 2021).

1.7. Secuenciación de ADN y procesamiento de resultados.

A lo largo de los años se han utilizado distintas tecnologías de NGS para la obtención de datos de ADN a partir de librerías genómicas. El primero en utilizarse fue el Roche 454 en 2006, que permaneció como el más popular hasta 2010. Actualmente la mayoría de los datos de secuenciación de ADN son obtenidos mediante las plataformas Illumina, debido a su alto rendimiento, precio y su margen de error relativamente bajo, además de que la plataforma es adecuada para lecturas relativamente cortas de ADN (<300 pb), lo que la hace ideal para la secuenciación de ADN (Goodwin y col., 2016).

Tras la secuenciación, el análisis de datos incluye el procesamiento de los datos crudos, alineamiento a genomas de referencia y la autenticación de los datos, además del análisis del margen de error.

El primer paso se denomina *demultiplexing*, el cual consiste en identificar las lecturas (la secuencia de bases perteneciente a un fragmento de ADN) que pertenecen a cada librería secuenciada tomando en cuenta los índices que les fueron añadidos previos a la secuenciación. Después de esto, se realiza el corte bioinformático de los adaptadores y/o bases terminales de baja calidad, para finalmente colapsar las lecturas que se superpongan en secuencias consenso (Orlando y col., 2021).

Luego de esto las secuencias suelen ser mapeadas contra el genoma de referencia utilizando BWA o Bowtie2. Este proceso puede dar distintos resultados no sólo dependiendo del software que se utilice, sino también del procedimiento

experimental que se haya llevado a cabo para la construcción de librerías (Orlando y col., 2021).

1.8. Importancia de los mamuts de Santa Lucía en el panorama paleogenómico mexicano

Como se ha mencionado anteriormente, los ejemplares encontrados en Santa Lucía y Tultepec representan una fuente de conocimiento invaluable, y la paleogenómica es la herramienta adecuada para acceder a él. No obstante, es la primera vez que se realizan estudios de esta índole con mamuts en México. Ya que la extracción de ADN es el primer paso para obtener datos paleogenómicos, es crítico no sólo que se realice de manera correcta, sino que también se conozca el rendimiento y la calidad del ADN extraído de las muestras, es decir, cuál es el porcentaje de material endógeno que se obtiene. De esta manera, la justificación de este trabajo yace en que, comparar el material endógeno obtenido de las extracciones de ADN, permitirá identificar las características de las muestras que den un mejor rendimiento en el material endógeno, y así será posible priorizar el muestreo y optimizar el procesamiento de la gran cantidad de muestras de mamuts que se han recuperado. Gracias a esto la obtención de la información genética será de mejor calidad y en consecuencia se tendrá mejor información sobre la megafauna que anteriormente existió en Santa Lucía y Tultepec.

2. HIPÓTESIS

Las muestras mejor conservadas y que fueron excavadas a menor profundidad darán un mejor rendimiento en el material endógeno, por lo que en el futuro el muestreo de molares con características similares debería ser priorizado.

3. OBJETIVOS

3.1. General

Analizar y comparar la recuperación de ADN endógeno de molares de mamuts (*Mammuthus columbi*) excavados en Santa Lucía, Estado de México.

3.2. Específicos

- Extraer ADN de las muestras disponibles y construir las librerías genómicas correspondientes para su secuenciación en la plataforma Illumina NextSeq550.
- Secuenciar y autenticar la naturaleza del ADN obtenido de cada muestra.
- Estimar y comparar con base en las características de las muestras (profundidad de muestreo, tipo de material, sexo, etc.) la cantidad de ADN endógeno obtenido en las distintas extracciones.
- Identificar las características de aquellas muestras que den un mejor rendimiento.

4. METODOLOGÍA

4.1. Materiales

4.1.1. Muestras osteológicas

Se utilizaron muestras provenientes de molares de mamut recuperados como producto de las excavaciones del proyecto de Salvamento Arqueológico y Paleontológico del AIFA (Aeropuerto Internacional Felipe Ángeles) en Santa Lucía, Estado de México. Para la recolección de las muestras se contó con los permisos correspondientes por parte del Consejo de Paleontología del INAH para realizar la extracción de ADN de los materiales disponibles.

4.1.2. Oligonucleótidos

Para este trabajo se utilizaron los oligonucleótidos señalados en el Cuadro 1, los cuales sirvieron como adaptadores para la plataforma de secuenciación Illumina NextSeq550. Estos se compraron en la casa comercial T4 oligos, con purificación en HPLC a escala de 50 nM.

Cuadro 1. Oligonucleótidos para construcción de librerías.

Nombre	Secuencia 5'-3'	Función
P5	AATGATACGGCGACCACCGA	Adaptador para la plataforma Illumina
P7	CAAGCAGAAGACGGCATACGAGAT	Adaptador para la plataforma Illumina

4.1.3. Reactivos

En el presente trabajo se utilizaron los reactivos a continuación enlistados. Para la extracción de ADN, se preparó una solución tampón de lisis y una de tampón de unión. Para 25 mL de tampón de lisis se mezclaron 22.5 mL de EDTA 0.5M (Sigma-Aldrich. E7889. EUA), 1.863 mL de H₂O y 12.5 µL de Tween 20 (Sigma-

Aldrich. P9416. EUA). Esta solución se expuso 5 minutos a luz UV. Previo a su uso, se añadieron 625 μL de proteinasa K a 10 mg/mL (New England Biolabs P8107S, EUA) y se almacenó a temperatura ambiente. Para el tampón de unión se mezclaron 500 mL de tampón PB (Qiagen 19065. Alemania) que consiste en clorhidrato de guanidina e isopropanol. También se añadieron 500 μL fenol rojo (Sigma-Aldrich P0290. EUA), 2.5 mL de NaCl 5M (Sigma-Aldrich S3014. EUA) y 9 mL de acetato de Sodio 3M (Sigma-Aldrich S7899. EUA). También se utilizó la solución tampón PE (Qiagen 19065. Alemania) para los lavados. De igual manera, se utilizó el tampón EB (Tris-Cl 10mM, pH 8.5) para la elución del ADN (Qiagen 19086. Alemania).

Para la construcción de librerías, se utilizó el tampón BST, preparado con 8 μL de dNTPs 25 mM y 500 μL de Tampón ThermoPol 10X (New England Biolabs Inc. M0275L. EUA). También se utilizó el tampón Quick Ligation 2X (10 μL), así como 5 μL de ADN ligasa T4 Quick (New England Biolabs Inc. E6056L. EUA) para las reacciones de ligación de adaptadores. Para el paso de rellenado de espacios, se utilizaron 3 μL de ADN polimerasa de fragmento largo *Bst* (New England Biolabs Inc. M0275L. EUA).

Para la cuantificación de los extractos de ADNa, se utilizó el kit Qubit dsDNA HS Assay Kit (Invitrogen Q32854. EUA), que permite cuantificar ADN con alta sensibilidad por medio de fluorescencia.

4.2. Métodos

4.2.1. Muestreo

Esta tesis forma parte del proyecto de investigación registrado ante el INAH “La prehistoria y paleoambiente del noroeste de la Cuenca de México”, por lo que las muestras se extrajeron de los ejemplares de molares de mamuts recuperados como producto de las excavaciones del proyecto de Salvamento Arqueológico y Paleontológico del AIFA (Aeropuerto Internacional Felipe Ángeles). Dichos restos

fósiles se encuentran almacenados en el edificio militar correspondiente al Museo del Mamut en la base militar de Santa Lucía, Estado de México (Museo Quinamétzin). Es importante mencionar que los restos fueron colectados oportunísticamente de cada yacimiento por el equipo de paleontólogos, es decir, no se originaron de ningún ensamblaje fósil en particular, por lo que se considera un muestreo aleatorio del registro fósil disponible de estos animales en la NOCMex.

El muestreo se realizó en la base militar de Santa Lucía, dentro de instalaciones dedicadas exclusivamente a esta tarea, dentro de un edificio donde no se habían procesado muestras de ADN moderno ni se habían llevado a cabo técnicas de PCR para evitar la contaminación de las muestras con ADN moderno. La manipulación de los fósiles se llevó a cabo en un laboratorio provisional dentro de una tienda de campaña, el cual contaba con mesas y campanas de acrílico en su interior. Para evitar la contaminación de las muestras, se utilizó equipo de protección estándar (cubrebocas, traje desechable, guantes y botas quirúrgicas).

Los molares se limpiaron para remover restos de tierra adherida. La superficie lateral del molar se limpió con cloro comercial (5%), seguido de un lavado con una solución de etanol (75%). El muestreo se realizó utilizando un dremel para generar un pequeño agujero entre las capas de esmalte, de manera que se recuperó polvo de dentina (50-100 mg). Para cada molar se establecieron dos puntos de muestreo, es decir, de perforación, y se obtuvieron dos muestras de cada uno. De cada punto se extrajo dentina a distintas profundidades, una más superficial que la otra. Para evitar su contaminación, cada muestra de polvo se guardó en un tubo Eppendorf LoBind de 1.5mL. Dichas muestras se almacenaron a 4 °C en las instalaciones del laboratorio de Paleogenómica (exclusivas para almacenamiento de muestras antiguas), ubicado en un edificio independiente del LIIGH.

Después se unió una columna MinElute (Qiagen, Alemania) junto con el reservorio Zymospin limpio, se aseguró con parafilm y se colocó en un tubo cónico de 50 mL. Los tubos colectores de las columnas MinElute se conservaron para su uso posterior.

Se transfirió la solución del tubo cónico de 15 mL al reservorio Zymospin que estaba dentro del tubo cónico de 50 mL. Se centrifugaron los tubos cónicos de 50 mL con el reservorio a 1000 rpm durante 15 minutos o hasta que hubiera pasado todo el volumen a través de la columna MinElute.

Se tomó la columna MinElute, se retiró el parafilm y se transfirió la columna al tubo colector de la columna MinElute reservado anteriormente. Después, se añadieron 720 μ L de tampón PE a la columna MinElute y se centrifugó a 13,000 rpm por un minuto. Luego, se desechó el tubo colector y se centrifugó a velocidad máxima por un minuto. Finalmente, el líquido de la columna MinElute se transfirió a un nuevo tubo Eppendorf LoBind de 1.5 mL.

4.2.6. Elución de ADN

Se agregaron 27 μ L de tampón EB en el centro de la columna MinElute, evitando tocar el filtro. Se incubó a 37 °C por 10 minutos y se centrifugó a máxima velocidad por 1 minuto. La solución eluida se transfirió a un tubo Eppendorf LoBind de 1.5 mL nuevo y etiquetado para su almacenamiento y posterior uso. Dicho procedimiento se repitió una vez más. El volumen total de ADN obtenido esperado era de ~54 μ L, y la concentración de ADN recuperado se midió por fluorescencia utilizando el equipo Qubit (Invitrogen, EUA).

4.2.7. Cuantificación de ADN

La cuantificación de los extractos de ADN obtenidos se realizó en el equipo Qubit, con los reactivos del kit Qubit dsDNA High sensitivity Assay Kit, de acuerdo al protocolo especificado por el fabricante. Este protocolo está basado en la

detección de fluorescencia de un agente intercalante. El procedimiento consistió en preparar un mix de tampón Qubit, donde se utilizaron 199 μL por muestra, además de 1 μL del colorante por cada muestra. Primero se prepararon dos estándares, mezclando en cada tubo Qubit (provisto por el proveedor) 190 μL del mix y 10 μL de cada solución estándar. Posteriormente se prepararon los tubos de las muestras, utilizando 199 μL del mix y 1 μL de muestra. Por último, se leyeron los estándares en el equipo para generar una curva estándar y realizar las lecturas de las muestras.

4.2.8. Construcción de librerías

La construcción de librerías para secuenciación en plataforma Illumina se llevó a cabo de acuerdo al protocolo propuesto por Meyer y Kircher (2010) , detallado a continuación.

4.2.8.1. Preparación de tampón BST

En un tubo de microcentrífuga de 1.5 mL se añadieron 8 μL de dNTP mix 25mM a 500 μL de Buffer ThermoPol 10X para obtener un stock de tampón Thermopol y mix de dNTPs. Con esta mezcla se obtuvo una concentración final de 0.4 mM de cada dNTP durante la etapa de llenado de espacios.

4.2.8.2. Reparación de extremos

Se prepararon los siguientes componentes en un master mix para añadirlo a las muestras en tubos de PCR. Se aseguró que las sales del buffer estuvieran en suspensión antes de pipetear. La tira de tubos de PCR se colocó en una gradilla congelada debido a la termosensibilidad de las enzimas. Cada reacción tenía un volumen total de 50 μL (Cuadro 2). La reacción se mezcló por pipeteo y se centrifugaron a baja velocidad para bajar el líquido de las paredes.

Cuadro 2. Componentes de reacción de reparación de extremos.

Reactivo	Volumen por reacción (µL)
H ₂ O estéril	17.5
Buffer de reparación NEBNext (10X)	5
Mix de enzimas de reparación de extremos NEBNext	2.5
ADN	25
Volumen total	50

Se incubó a 12°C por 20 minutos, seguido por 37°C durante 15 minutos. Después se realizó una limpieza, la cual consistió en transferir los 50 µL de la reacción a un tubo Eppendorf LoBind de 1.5mL y se añadieron 500 µL de buffer PB de Qiagen. Esta mezcla se mezcló por pipeteo.

Luego se transfirió la mezcla del paso anterior a una columna Qiagen QIAquick y se centrifugó a máxima velocidad por un minuto. Luego se desechó el líquido del tubo colector. Después se añadieron 750 µL de tampón PE, se centrifugó la columna y se desechó el líquido. Luego se repitió la centrifugación y se repitió la centrifugación para eliminar el buffer remanente. Se transfirió la columna a un nuevo tubo Eppendorf LoBind de 1.5 mL.

Después se añadieron 31 µL de buffer EB al centro de la columna evitando tocar el filtro. Esto se incubó a 37°C durante 15 minutos. Se centrifugó a máxima velocidad por un minuto para eluir.

4.2.8.3. Ligación de adaptadores.

Se mezclaron los reactivos señalados en el Cuadro 3 en tubos de PCR para obtener el master mix con el buffer de ligasa y la enzima. Se mezcló por pipeteo.

Cuadro 3. Componentes de reacción de ligación de adaptadores.

Componente	Volumen por reacción (µL)
Buffer Quick Ligation NEBNext (5X)	10 µL
Mix de adaptadores P5/P7 (2.5 µM)	5 µL
ADN Ligasa Quick T4 NEBNext	5 µL
ADN reparado de la reacción anterior	31 µL
Volumen total	51 µL

Para añadir los componentes de la reacción, el ADN se mezcló con los adaptadores por separado, y se mezcló por pipeteo. Se centrifugó a baja velocidad para coleccionar el líquido de las paredes. Ya con todos los componentes en el tubo de PCR, se incubó en el termociclador a 20°C por 20 minutos.

Después se realizó una limpieza por columna QIAQuick de QIAGEN. Los 50 µL de la reacción se transfirieron a un tubo Eppendorf LoBind de 1.5 mL y después se añadieron 250 µL de Buffer PB. Esto se mezcló por pipeteo.

La mezcla del paso anterior se transfirió a una columna QIAGEN QIAQuick y se centrifugó a máxima velocidad por un minuto. Luego se desechó el líquido del tubo colector. Después se añadieron 750 µL de tampón PE, se centrifugó la columna y se desechó el líquido. Luego se repitió la centrifugación y se repitió la centrifugación para eliminar el buffer remanente. Se transfirió la columna a un nuevo tubo Eppendorf LoBind de 1.5 mL.

4.2.2. Extracción de ADN

Previo a la recuperación de ADN, se prepararon las soluciones tampón de lisis y tampón de unión descritas en la sección 4.1.3. La extracción de ADN se llevó a cabo de acuerdo a los protocolos publicados por Dabney y col. (2013), y Rohland y col. (2018).

4.2.3. Limpieza de Reservoirio Zymo spin

Primero, se retiró el filtro del reservorio y se desechó. Luego, se enjuagó el reservorio con cloro comercial al 5% (v/v). Posteriormente se realizó un enjuague con una solución de etanol al 70% (v/v). El exceso de etanol de las muestras se eliminó colocándolas sobre una toalla de papel absorbente. Finalmente, los reservorios se expusieron a luz UV durante 3 min por ambos lados. Se utilizó un reservorio por cada muestra.

4.2.4. Reacción de lisis

A cada tubo Eppendorf LoBind de 1.5 mL conteniendo 0.1-0.7 mg de muestra en polvo, se añadió 1 mL de tampón de lisis. Se incubó en rotación a 37 °C durante ~16 h en la estufa de hibridación con plataforma de apilamiento Shake 'n' Stack (Thermo Scientific, EUA) a velocidad media. Transcurrido este tiempo, se centrifugaron los tubos a 2,500 g durante 5 min en la microcentrífuga ventilada modelo LEGEND MICRO 17 (Thermo Scientific, EUA).

4.2.5. Extracción de ADN

En un tubo cónico de 15 mL se añadieron 13 mL de tampón de unión y ~1 mL del sobrenadante de la reacción de lisis (se evitó tomar restos de la muestra, ya que podía tapan la columna del siguiente paso). Esto se mezcló por inversión o en vortex.

Después se añadieron 31 μL de buffer EB al centro de la columna evitando tocar el filtro. Esto se incubó a 37°C durante 15 minutos. Se centrifugó a máxima velocidad por un minuto para eluir.

4.2.8.4. Rellenado de espacios.

Se mezclaron los siguientes componentes en un master mix para después añadirlo a las muestras en tubos de PCR (Cuadro 4).

Cuadro 4. Componentes de reacción de relleno de espacios.

Reactivo	Volumen por reacción (μL)
Tampón BST	5.8 μL
ADN polimerasa Bst de fragmento largo	3 μL
ADN con adaptadores ligados de la reacción anterior	30 μL
H ₂ O estéril	11.2 μL
Volumen total	50 μL

La reacción se incubó por 20 minutos a 65°C y luego por 20 minutos a 80°C. Después de la incubación las librerías se pueden transferir a tubos de 1.5 mL y luego congelar a -20°C para su posterior uso.

4.2.9. Reacción de qPCR

Luego de la construcción de librerías, se realizó una reacción de qPCR para determinar el número de ciclos adecuado para añadir el índice a cada librería. Se

mezclaron los siguientes componentes por reacción en un tubo para qPCR (Cuadro 5).

Cuadro 5. Componentes de reacción de qPCR.

Reactivo	Volumen por reacción (µL)
ADN molde (Librería)	1
Oligo sentido (PE) 10 µM	0.4
Oligo antisentido (índice 10) 10 µM	0.4
KAPA SYBR FAST MIX	10
ROX 50X (high)	0.4
H ₂ O estéril	7.8
Volumen total	20

Los componentes del tubo se mezclaron por pipeteo y después se centrifugó ligeramente para colectar el líquido de las paredes del tubo. Posteriormente las reacciones se llevaron a cabo en el equipo de qPCR (laboratorio de ADN moderno) de acuerdo al Cuadro 6.

Una vez terminado el programa, se realizó el análisis de los resultados para obtener los gráficos de amplificación. El número de ciclos para el indexado de librerías se seleccionó tomando en cuenta el ciclo de umbral (C_T) de cada librería.

Cuadro 6. Programa de reacción de qPCR.

Paso	Temperatura	Duración	Ciclos
Activación de enzima	95°C	10 min	1
Desnaturalización	95°C	15 s	1
Alineamiento/extensión	60°C	20 s	30

4.2.10. Indexado de librerías.

Esta reacción se realizó utilizando la enzima Phusion U Polymerase, como se muestra en el Cuadro 7. Se incluyó un control negativo para la reacción utilizando agua. Cada uno de los índices utilizados era un oligonucleótido de 6 pares de bases.

Cuadro 7. Componentes de reacción de indexado de librerías.

Componente	µL por reacción
ADN (librería)	12
Oligo PE (sentido)	1.5
Oligo antisentido (cualquier índice disponible a 10 µM)	1.5
Master Mix Phusion U Hot 2X	15
Volumen total	30

Se añadieron 16.5 uL de Master mix al tubo y después el volumen correspondiente de cada librería a su tubo correspondiente. Al final se añadieron los índices seleccionados para cada librería.

Cuadro 8. Programa de PCR para indexado de librerías.

Temperatura	Duración	Ciclos
98°C	1 min	1
98°C	10 s	* De acuerdo a C_T obtenido en la reacción de qPCR
72°C	45 s	
72°C	5 min	1
4°C	Infinito	1

Una vez terminado el programa de amplificación (Cuadro 8), el DNA (librería indexada) fue purificado utilizando SPRI beads de acuerdo a las instrucciones del fabricante, con una proporción de 1.2x, para recuperar fragmentos mayores a 150 pb, tomando en cuenta la longitud del par de adaptadores ~120 pb.

Después de la purificación, se determinó la concentración de las librerías indexadas mediante Qubit. Asimismo, se analizó 1 µL de librería indexada con el equipo BioAnalyzer 2100 para determinar la longitud de los fragmentos.

4.2.11. Manejo de residuos

Los residuos generados en el laboratorio de ADN fueron separados de acuerdo a su naturaleza y colocados en los contenedores destinados para su adecuado desecho.

4.2.12. Secuenciación de nueva generación

Se llevó a cabo la secuenciación shotgun y multiplex de las librerías en la plataforma Illumina NextSeq550 en el Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO), en una corrida de alto rendimiento (high output), donde se generaron alrededor de ~320 Millones de reads en sentido y antisentido (PE).

4.2.13. Cálculo de contenido endógeno

Para llevar a cabo esta tarea, se utilizó el clúster de súpercomputo del Laboratorio de Visualización Científica Avanzada (LAVIS). Primero se removieron los adaptadores de las lecturas generadas utilizando AdapterRemoval (Lindgreen, 2012). Posteriormente las secuencias se mapearon con un mapeo competitivo contra el genoma de referencia de elefante (LoxAfr4) y el genoma humano (hg19) para quitar contaminación, utilizando el programa BWA (Burrows-Wheeler Aligner) (Li y Durbin, 2010), utilizando los parámetros `-l 16500 -n 0.01 -o 2 -t 16` (optimizados para alineamiento de ADN), siguiendo el protocolo de Pečnerová y col., 2017. Después, para autenticar la naturaleza de las secuencias se examinaron los cambios de bases con respecto al genoma de referencia en las formas $C \rightarrow T$ y $G \rightarrow A$ en los extremos 5' y 3', respectivamente. También se evaluó la complejidad de la librería, así como los patrones de fragmentación y daño post-mortem (referidos como "patrones de daño" utilizando el software mapDamage 2.0 (Jónsson y col., 2013).

Tomando en cuenta los resultados de este procesamiento de datos y utilizando la fracción de lecturas que mapean al genoma de elefante, se determinó la cantidad de ADN endógeno de cada librería, el cual se refiere a la proporción de secuencias que mapearon al genoma de referencia, y que se considerarán como del organismo del cual la muestra fue obtenida (Damgaard y col., 2015). Este cálculo se llevó a cabo con el software SAMtools (Li y col., 2009). Para el cálculo se

consideraron las bases que tuvieran un puntaje de calidad de 30 en la escala Phred. La ecuación que representa este cálculo es la siguiente:

$$\text{Contenido endógeno} = \frac{\# \text{ de Lecturas mapeadas a } M. \text{ columbi}}{\# \text{ de Lecturas antes de mapeo}} * 100$$

También se realizó el sexamiento de los individuos siguiendo la metodología propuesta por Pecnerova y col. (2017), la cual consiste en calcular la probabilidad de que un individuo sea XX o XY tomando en cuenta la proporción de lecturas que mapearon contra el cromosoma X en comparación con el cromosoma 8 debido a la similitud de su tamaño.

En resumen, las estadísticas del mapeo por estimar fueron: el número de lecturas que mapearon contra el genoma humano (para evaluar contaminación), el contenido de ADN endógeno (estimado al dividir la fracción de las lecturas que mapearon contra LoxAfr4 y no mapearon contra el genoma humano, entre el número total de secuencias), y la clonalidad de las secuencias (proporción de secuencias duplicadas), así como el sexo cromosómico de los individuos.

Para graficar los resultados del contenido endógeno de las distintas muestras, se utilizó el programa RStudio para generar los gráficos correspondientes (RStudio Team, 2012), utilizando principalmente el paquete ggplot2 (Wickham, 2016). También se utilizó R para realizar pruebas estadísticas (U de Mann-Whitney) para las variables cualitativas y los diagramas de caja, así como los gráficos de residuales para evaluar la relación lineal entre el contenido endógeno y las variables cualitativas (profundidad mínima y máxima, y resultados de cuantificación utilizando Qubit y BioAnalyzer). El código que se generó para realizar estas pruebas se encuentra en el Anexo 4.

5. RESULTADOS

5.1. Obtención de muestras

En las Figura 4 y 5 se muestran las fotografías disponibles de los restos fósiles que fueron muestreados y procesados para su secuenciación. Dichas fotografías fueron obtenidas con la ayuda de la Dra. Miriam J. Bravo, la Mtra. Viridiana Villa y el Dr. Federico Sánchez Quinto en la Base Militar Aérea de Santa Lucía. Los datos de excavación y notas arqueológicas fueron proporcionados por Alejandro López Jiménez, del equipo de trabajo del INAH. El nombre o ID asignado a cada una de las muestras corresponde a la nomenclatura de la Figura 3. Por ejemplo, se tiene la muestra McM1_2139; Mc corresponde a la especie *M. columbi*, la segunda M corresponde a molar, el 1 corresponde al número de tubo, es decir el 1.1, por último, el número 2139 corresponde al número de bolsa, que es un número que fue proporcionado por el equipo de arqueología para la identificación de las piezas.

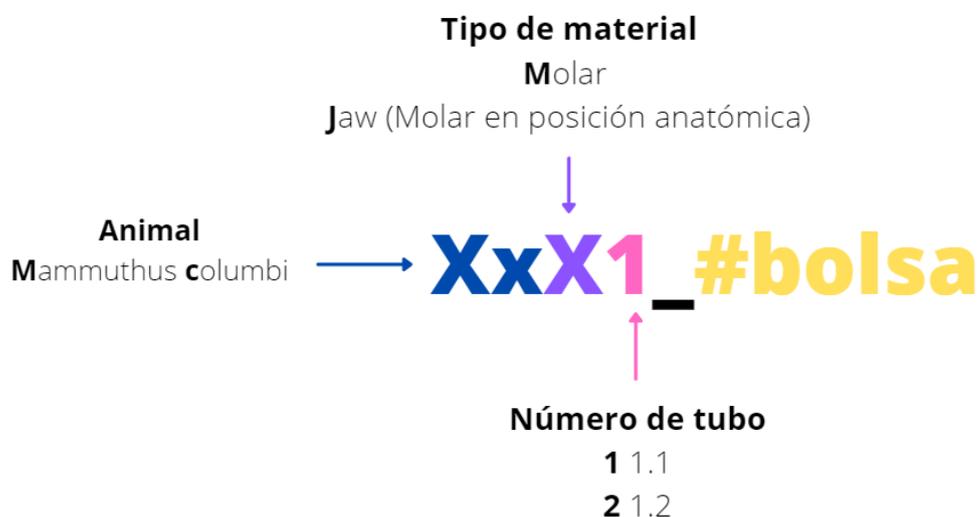


Figura 3. Nomenclatura para la identificación de muestras procesadas.



Figura 4. Fotografías de molares y molares en posición anatómica muestreados en la Base Militar Aérea de Santa Lucía.



Figura 5. Fotografías de molares y molares en posición anatómica muestreados en la Base Militar Aérea de Santa Lucía (continuación).

En las fotografías de las Figuras 4 y 5 se pueden apreciar distintas características físicas de los restos fósiles. Por ejemplo, el color: la muestra McM2_19147 muestra una coloración naranja, la muestra McM2_Molar_negro presenta manchas de color café oscuro y negro, y el resto de los molares tienen una coloración entre amarillenta y café. También se puede apreciar que algunas muestras tienen una apariencia más porosa que otras, por ejemplo, la muestra McM2_2371 y McJ2_20343 tienen un aspecto más dañado que el de la McM2_2140 o la McJ2_21602, que está prácticamente íntegra. Estas características recobrarán más importancia más adelante.

5.2. Extracción de ADN de molares y mandíbulas

Se obtuvo el ADN de 70 muestras de dentina en 14 grupos independientes, cada uno con el control negativo correspondiente (H₂O en lugar de dentina). El extracto

se cuantificó utilizando el Qubit dsDNA High sensitivity Assay Kit, el cual es un método fluorométrico. Todas las muestras tuvieron una concentración mayor al límite inferior de cuantificación (0.01 ng/μl), excepto los controles negativos, los cuales fueron indetectables, esto debido a la nula o muy baja concentración de ADN, ya fuera de la muestra, contaminación cruzada, o del medio ambiente. El registro de las extracciones se encuentra en el Anexo 1. Para fines prácticos, en el Cuadro 9 se muestran aquellas que fueron seleccionadas para secuenciación.

Cuadro 9. Concentración de ADN obtenido en Qubit de las muestras de mamut.

Muestra	Concentración (ng/μL)	Muestra	Concentración (ng/μL)
McM1_2139	3.64	McM2_Molar_negro	5.86
McM1_21920	3.9	McM2_19147	1.01
McJ1_23463	7.64	McJ2_1639	24.8
McJ2_5917	3.42	McJ2_9524	4.16
McJ2_21602	5.44	McJ2_16695	39.6
McJ2_3279	5.3	McJ2_20343	4.26
McJ2_2021	4.52	McM2_2140	1.25
McJ2_14852	7.96	McM2_23171	1.09
McM2_24983	3	McJ2_18728	4.9
McJ2_42114	39.4	McJ2_3158	8.84
McM2_22683	1.61	McCN1	No detectable

De las 70 extracciones realizadas, se seleccionaron 22 para la construcción de librerías a partir del extracto y posterior secuenciación. Se tomaron en cuenta aquellas muestras con la mayor concentración de ADN.

5.3. Construcción de librerías, indexado y secuenciación de nueva generación

Posteriormente se llevaron a cabo los experimentos descritos en la sección 4.2.8 a la 4.2.12. Para seleccionar las librerías candidatas a indexado y, por lo tanto, a secuenciación, se tomó en cuenta el CT de la librería, es decir, su umbral de amplificación en qPCR. Posterior al indexado las librerías seleccionadas fueron colocadas en un pool (una mezcla de librerías) para su secuenciación en la plataforma Illumina NextSeq550.

5.4. Procesamiento Bioinformático

Este procesamiento se llevó a cabo de acuerdo al script que se muestra en el Anexo 2. Para el procesamiento de las secuencias, primero se llevó a cabo el demultiplex, el cual consistió en separar los datos crudos en cada una de las librerías que se mandaron a secuenciar tomando en cuenta los índices que fueron añadidos previo a la preparación del pool. Posteriormente se llevó a cabo el corte de adaptadores, ya que dichas secuencias no forman parte del contenido genético del organismo analizado. Esto se llevó a cabo con el software AdapterRemoval (Lindgreen, 2012). Es importante mencionar que el tipo de secuenciación llevado a cabo fue *paired-end*, lo que quiere decir que las lecturas se realizan en ambas direcciones de los fragmentos, es por eso que en este paso tomaron en cuenta las lecturas que se sobrelaparon, generando secuencias consenso. Después de esto se realizó el alineamiento contra los genomas de referencia con ayuda del programa BWA (Li y Durbin, 2010). El alineamiento se realizó en el cluster del LAVIS utilizando parámetros relajados para ADN, como especifica la sección 4.2.13. Como referencia se utilizó una referencia combinada del genoma del elefante africano (*Loxodonta africana*) Loxv4 y el genoma humano (hg19) para estimar y remover cualquier tipo de contaminación de secuencias humanas en lo que se denomina un mapeo competitivo.

Este procedimiento generó archivos con la extensión .bam, los cuales después fueron filtrados con el software SAMtools (Danecek y col., 2021), con el cual se

lograron remover las secuencias paralogas y las secuencias duplicadas generadas por la amplificación por PCR. Las estadísticas de secuenciación para las muestras se muestran en el Cuadro 10 y 11, las cuales fueron obtenidas mediante SAMtools.

Para la autenticación de la naturaleza antigua del ADN se utilizó el programa mapDamage 2.0 (Jónsson y col., 2013), con el cual se generaron gráficos de patrones de daño. Los gráficos de las muestras y el respectivo control negativo se encuentran en el Anexo 3.

Cuadro 10. Estadísticas de secuenciación obtenidas a partir de mapeo a genoma de referencia. Marcadas con un asterisco (*) se encuentran las dos muestras que presentaron mejor rendimiento de contenido endógeno.

Muestra	Lecturas antes de mapeo	Lecturas mapeadas a <i>M. columbi</i>	Clonalidad	% de contenido endógeno	Longitud de lectura (pb)	Lecturas mapeadas a humano	% de contaminación humana
McM1_2139	8,173,328	361	22.863	0.0	43.25	75	0.001
McM1_21920	10,751,810	5,867	18.672	0.04	40.54	351	0.003
McJ1_23463	8,600,764	30,045	3.002	0.23	53.21	119	0.001
McJ2_5917	12,703,746	6,276	12.712	0.04	40.04	233	0.002
McJ2_21602	30,867,715	17,330	5.067	0.04	45.16	512	0.002
McJ2_3279	22,209,754	988	14.827	0.0	46.16	124	0.001
McJ2_2021	13,430,702	14,639	9.546	0.09	42.87	464	0.003
McJ2_14852	22,384,093	5,197	16.848	0.02	57.68	362	0.002
McM2_24983	8,374,783	6,679	10.481	0.06	41.06	181	0.002
McJ2_42114	5,729,178	32,285	2.350	0.21	49.71	108	0.002
McM2_22683	14,728,484	30,566	10.041	0.11	44.14	2727	0.019
McM2_Molar_negro	7,311,853	183	80.937	0.0	34.93	300	0.004
McM2_19147	5,181,809	163,253	3.146	2.22*	50.23	1680	0.032

Cuadro 11. Estadísticas de secuenciación obtenidas a partir de mapeo a genoma de referencia. Marcadas con un asterisco (*) se encuentran las dos muestras que presentaron mejor rendimiento de contenido endógeno (continuación).

Muestra	Lecturas antes de mapeo	Lecturas mapeadas a M. columbi	Clonalidad	% de contenido endógeno	Longitud de lectura (pb)	Lecturas mapeadas a humano	% de contaminación humana
McJ2_1639	14,695,907	22,722	3.277	0.14	47.85	166	0.001
McJ2_9524	11,957,531	258,768	1.963	1.85 *	42.62	473	0.004
McJ2_16695	10,118,372	683	42.020	0.01	47.23	190	0.002
McJ2_20343	8,679,765	5,011	21.995	0.05	39.04	494	0.006
McM2_2140	7,060,738	882	10	0.01	43.77	65	0.001
McM2_23171	4,845,489	28,063	2.260	0.37	42.81	150	0.003
McJ2_18728	5,719,475	115	67.329	0.13	42.60	179	0.003
McJ2_3158	5,976,076.0	3158	2.598	0.0	47.29	43	0.001
McCN1	15,204	59	11.940	0.06	48.89	8	0.053

Posterior a obtener los estadísticos de secuenciación se realizó el sexamiento de los individuos tomando en cuenta la cantidad de lecturas al cromosoma X entre las lecturas al cromosoma 8. El cromosoma 8 del elefante es de un tamaño muy similar al cromosoma X, por lo que se esperaría que la proporción de lecturas chrX/chr8 fuera cercana a 1 en hembras y cercana a 0.5 en machos de acuerdo a la metodología propuesta por (Pečnerová y col., 2017). Los resultados de dicho sexamiento se muestran en el Cuadro 12. De las muestras analizadas, se identificaron 8 machos, 8 hembras, y de tres muestras no se pudo identificar el sexo molecular.

Obtenidos estos datos, se procedió a procesarlos en R Studio utilizando el paquete ggplot2 (Wickham, 2016), además de ggpubr y ggsignif, que derivan de éste mismo.

Cuadro 12. Resultados del sexamiento de los individuos muestreados.

Muestra	Proporción chrX/chr8	Sexo molecular
McM1_2139	0.671	NA
McM1_21920	0.452	Macho
McJ1_23463	0.958	Hembra
McJ2_5917	0.425	Macho
McJ2_21602	0.457	Macho
McJ2_3279	0.771	NA
McJ2_2021	0.494	Macho
McJ2_14852	0.449	Macho
McM2_24983	0.974	Hembra
McJ2_42114	0.498	Macho
McM2_22683	0.883	Hembra
McM2_Molar_negro	0.403	Macho
McM2_19147	1.056	Hembra
McJ2_1639	0.452	Macho
McJ2_9524	0.876	Hembra
McJ2_16695	0.671	NA
McJ2_20343	1.173	Hembra
McM2_2140	0.784	Hembra
McM2_23171	0.903	Hembra
McJ2_18728	1.343	Hembra
McJ2_3158	0.899	Hembra
McCN1	0.358	NA

Esto se realizó utilizando el script del Anexo 4. Es importante mencionar que, como se puede ver en los Cuadros 10 y 11, las muestras McJ2_9524 y McM2_19147 fueron aquellas que obtuvieron un mejor rendimiento en cuanto al porcentaje del contenido endógeno con 2.22% y 1.85% respectivamente, mientras que el contenido endógeno del resto estuvo en un rango de 0 a 0.63%. Los datos de profundidad máxima y mínima fueron compartidos por el equipo de paleontología, y se muestran en el Cuadro 13.

Cuadro 13. Profundidades de excavación de muestras de mamuts.

Muestra	Profundidad mínima de excavación (m)	Profundidad máxima de excavación (m)	Muestra	Profundidad mínima de excavación (m)	Profundidad máxima de excavación (m)
McM1_2139	NA	NA	McM2_Molar_negro	NA	NA
McM1_21920	1.45	1.53	McM2_19147*	1.41	1.46
McJ1_23463	1.13	1.79	McJ2_1639	NA	3.3
McJ2_5917	1.25	1.65	McJ2_9524*	1.07	NA
McJ2_21602	3.22	3.53	McJ2_16695	1.5	1.68
McJ2_3279	1.83	1.85	McJ2_20343	2.14	2.62
McJ2_2021	2.04	2.1	McM2_2140	2.21	2.24
McJ2_14852	3.62	3.82	McM2_23171	0.76	0.84
McM2_24983	1.7	1.74	McJ2_18728	2.68	3.02
McJ2_42114	1.38	1.57	McJ2_3158	1.85	2.7
McM2_22683	NA	NA	McCN1	NA	NA

Para realizar los gráficos en R Studio, se definieron como variables cualitativas el sexo cromosómico de los individuos y el tipo de material, es decir, si la muestra se había obtenido de un molar en posición anatómica (aún adherido a la mandíbula) o de un molar “suelto”. Las tablas correspondientes a cada gráfico se

encuentran en el Anexo 1. Primero se realizó un gráfico de caja tomando en cuenta el tipo de material y el porcentaje de contenido endógeno (Figura 4). En dicho gráfico se observa que la distribución del contenido endógeno es muy similar para ambos tipos de material, y, de hecho, la diferencia entre ellos no es significativa, lo que está indicado por la leyenda NS (No significativo). Se puede afirmar que no es significativa ya que se realizó una prueba de U de Mann Whitney para muestras no paramétricas, la cual se obtuvo mediante la función *wilcox.test()* de R. El valor de p obtenido en esta prueba fue de 0.6965.

Posteriormente se realizó la misma comparación, pero con el sexo cromosómico de los individuos. La gráfica de caja se muestra en la Figura 6. Se puede observar en dicho gráfico que las medianas son muy similares, la diferencia entre la distribución del contenido endógeno entre machos y hembras no es significativa, como indica la leyenda NS de la parte superior del gráfico. Para corroborar esto se llevó a cabo una prueba de U de Mann-Whitney, de la misma forma que para el sexo molecular. El valor de p obtenido en esta prueba fue de 1, lo que indica que no hay diferencia entre la cantidad de contenido endógeno obtenido de las hembras y de los machos.

Además de considerar las anteriores variables cualitativas, también se consideraron las profundidades de excavación mínima y máxima como variables cuantitativas. La profundidad mínima de excavación se considera la distancia entre la superficie de la tierra y el primer contacto con resto fósil, y en la profundidad máxima se considera la distancia que existe entre la superficie del suelo y la parte más profunda que alcanzara el resto fósil.

Para estas variables se generaron gráficos de dispersión entre la profundidad de excavación y la cantidad de contenido endógeno (Figuras 8 y 10). Además de esto, en la Figura 8 (contenido endógeno vs. profundidad mínima de excavación), se puede observar que casi todas las muestras, con excepción de los dos *outliers*, tienen una cantidad similar de contenido endógeno independientemente de la profundidad de excavación (rango de 0% a 0.6%).

Para descartar que estos datos se ajustaran a un modelo de tipo lineal se realizaron gráficos de residuales en RStudio utilizando el paquete ggplot2 (Figuras 9 y 1). Ambos gráficos de residuales tienen un comportamiento similar: los residuales, es decir, el error entre los valores esperados de un comportamiento lineal y el real de las muestras no sigue un patrón, ni tiene un comportamiento normal. Por lo tanto, se puede descartar que el porcentaje de contenido endógeno tenga alguna relación con las profundidades mínima y máxima de excavación.

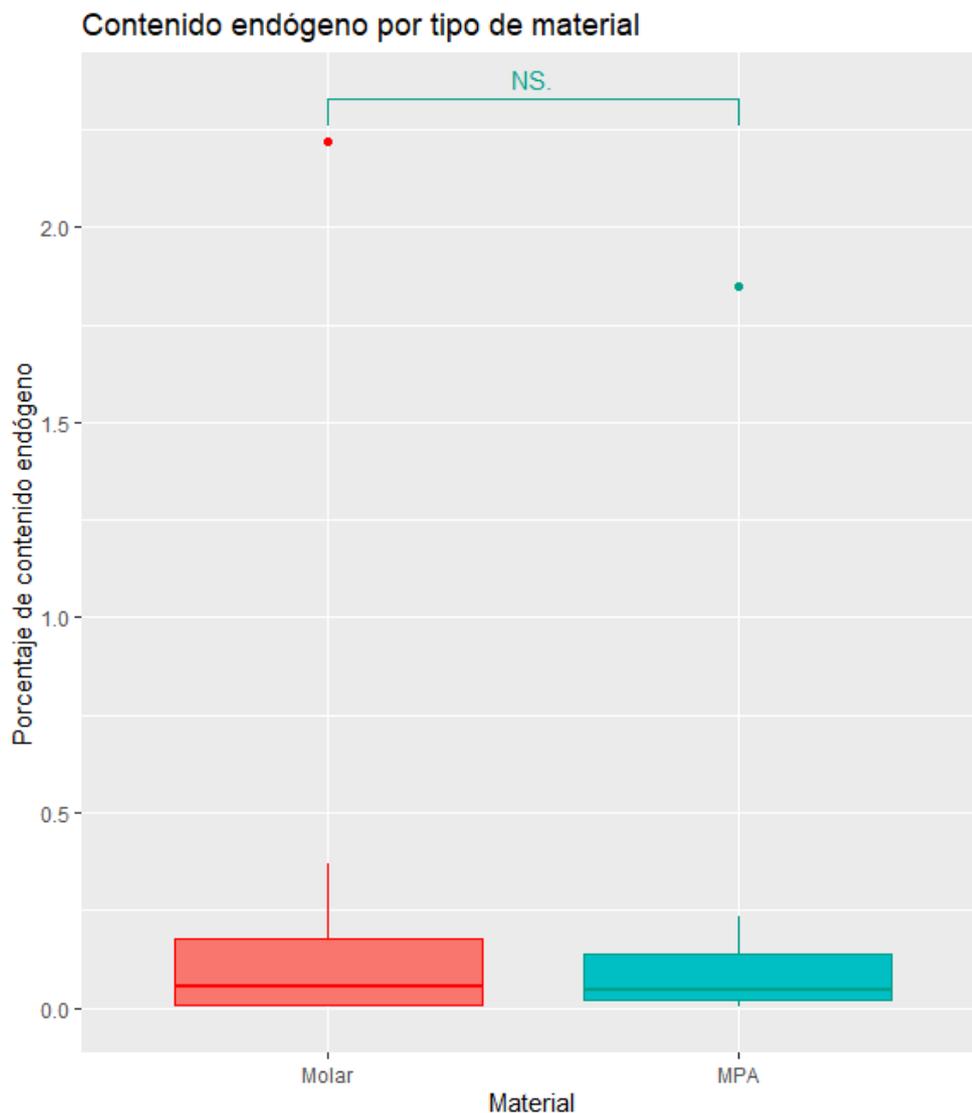


Figura 6. Gráfico de caja del contenido endógeno de acuerdo al tipo de material del que se obtuvo la muestra, molar suelto (Molar, azul) o molar en posición anatómica (MPA, rojo).

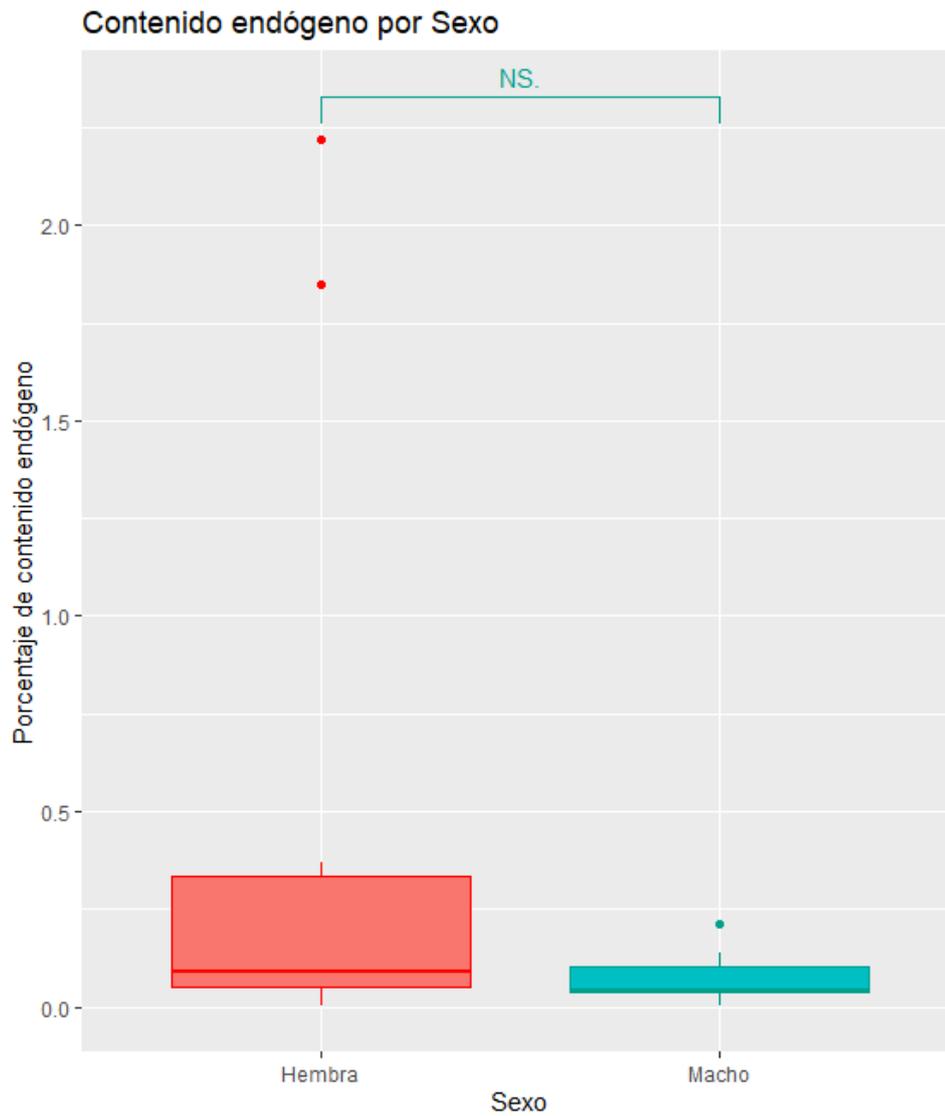


Figura 7. Gráfico de caja del contenido endógeno en relación al sexo cromosómico del individuo al que pertenece la muestra (hembra en rojo y macho en azul).

Por otro lado, también se realizó el mismo análisis, pero esta vez entre el porcentaje de contenido endógeno y la cuantificación de ADN mediante el método fluorométrico Qubit. Se realizó un diagrama de dispersión (Figura 12) y un gráfico de residuales (Figura 13) para descartar que los datos del diagrama de dispersión se ajustaran a un modelo lineal.

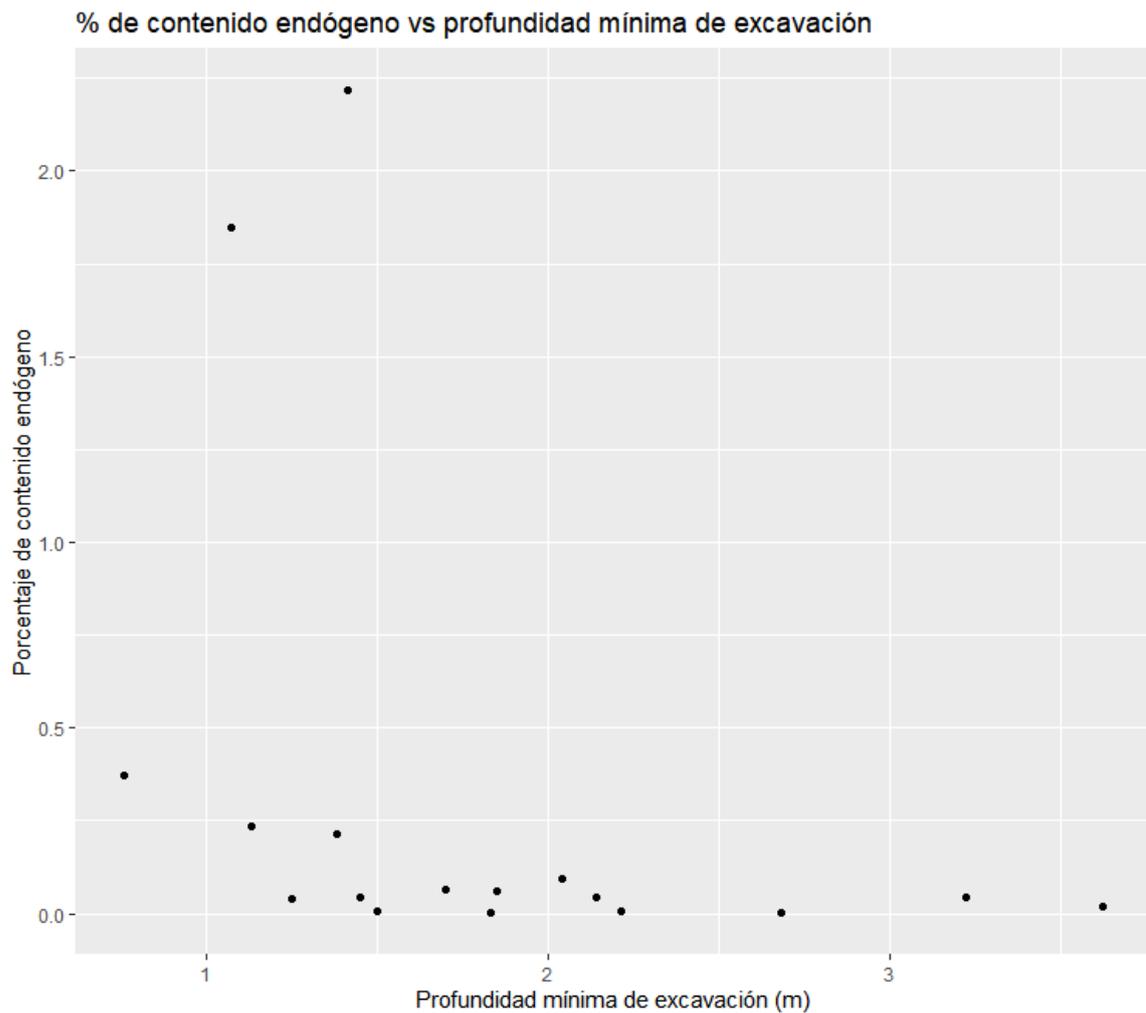


Figura 8. Diagrama de dispersión de la relación entre el porcentaje de contenido endógeno y la endógeno y la profundidad mínima de excavación.

Se observa en la Figura 13 que las muestras no tienen una distribución normal, además de que hay dos *outliers*, correspondientes a las dos muestras con mejor rendimiento, por lo que no existe una relación entre la cantidad de ADN cuantificado en Qubit y la cantidad de contenido endógeno final encontrada tras la secuenciación. Esta comparación resulta de interés ya que se podría pensar que entre mayor sea la concentración de ADN, hay una mayor probabilidad de encontrar ADN perteneciente al individuo, sin embargo, los resultados nos indican que esto no sucede así. Es importante recordar que el ADN extraído de las muestras es ADN total, pues no hay un método de extracción selectivo que

permita obtener solamente ADN endógeno. Los extractos pueden contener contaminación, además de que los procedimientos por los que pasa en el laboratorio pueden modificar el rendimiento final (por ejemplo, el PCR del paso de indexado).

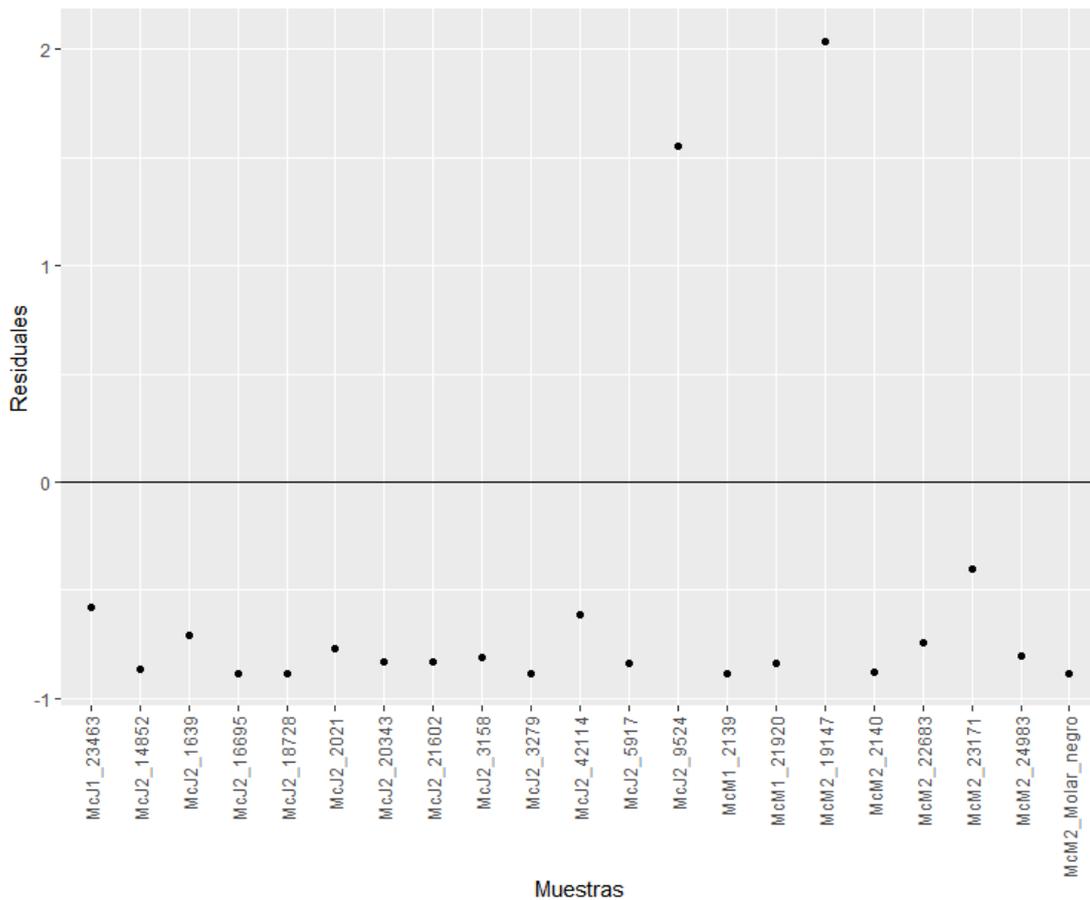


Figura 9. Gráfico de residuales para la relación entre porcentaje de contenido endógeno y profundidad mínima de excavación.

De igual forma se realizó un análisis similar considerando el porcentaje de contenido endógeno y la cuantificación de ADN mediante BioAnalyzer de las librerías, la cual se realiza previa a secuenciar las muestras. En la Figura 13 se observa el diagrama de dispersión, y en la Figura 14, el gráfico de residuales. En el gráfico de residuales es evidente que los datos no tienen una distribución normal, además de que hay dos *outliers*, correspondientes a las dos muestras con un mejor rendimiento. Esto nos indica que el comportamiento del contenido

endógeno en relación al ADN cuantificado en BioAnalyzer tras el indexado no presenta una relación de tipo lineal.

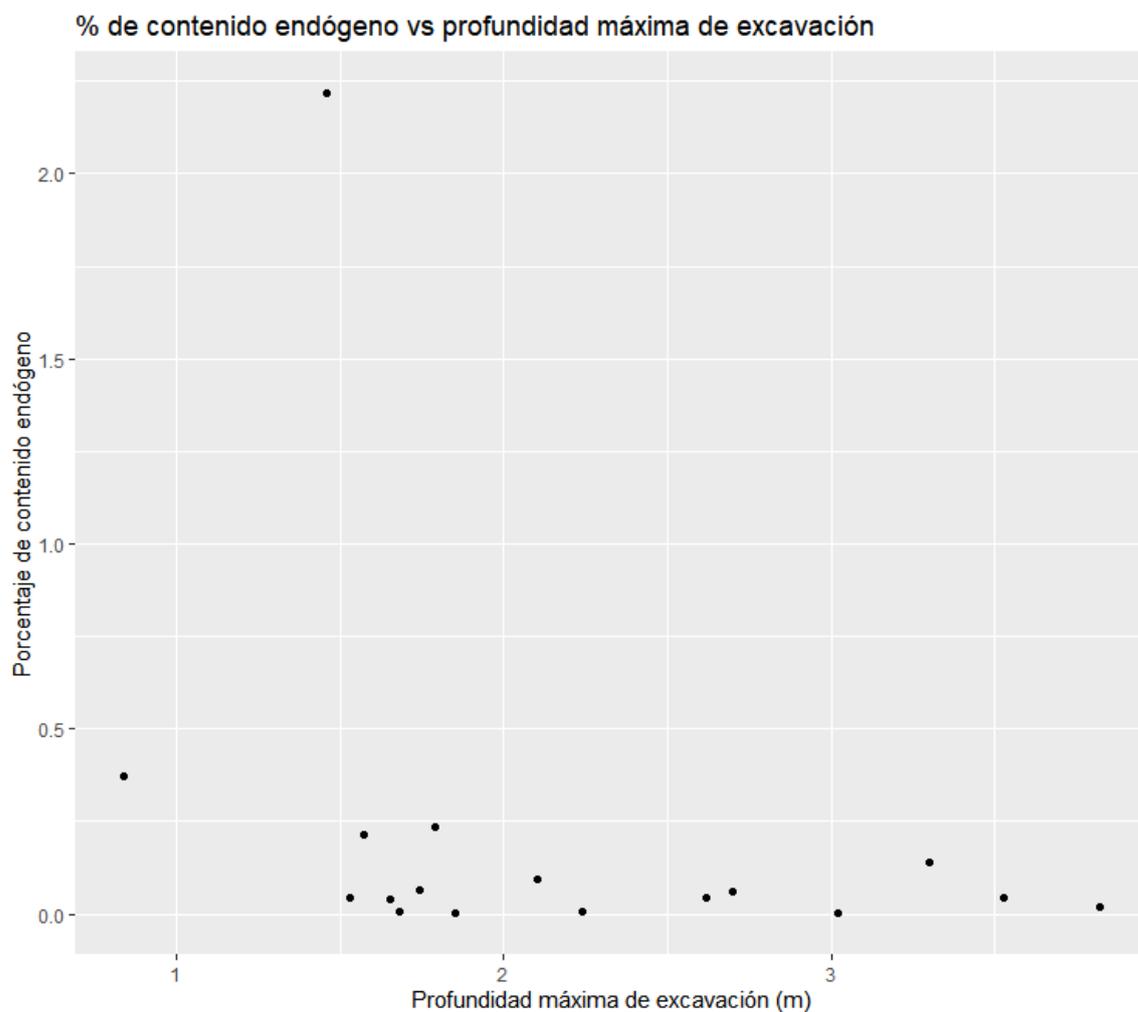


Figura 10. Diagrama de dispersión de la relación entre el porcentaje de contenido endógeno y la profundidad máxima de excavación.

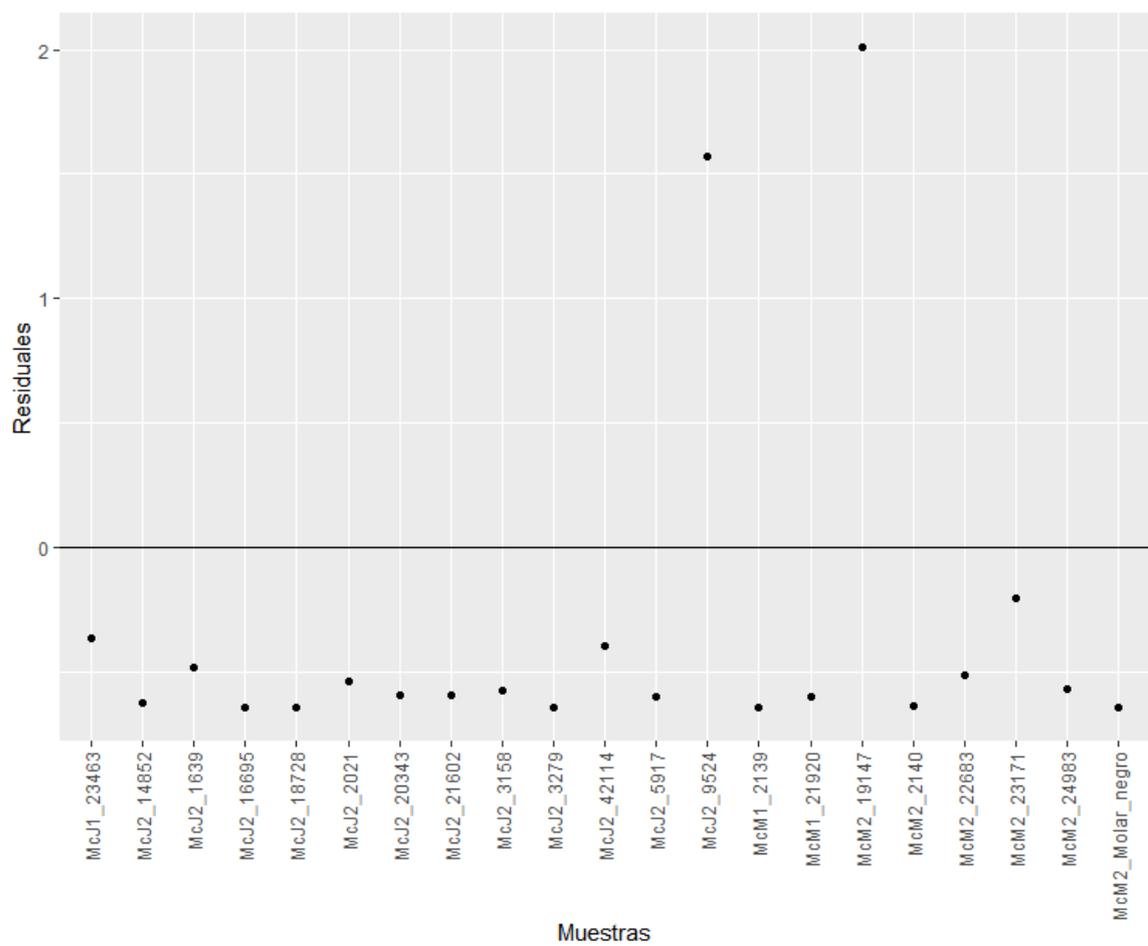


Figura 11. Gráfico de residuales para la relación entre porcentaje de contenido endógeno y profundidad máxima de excavación.

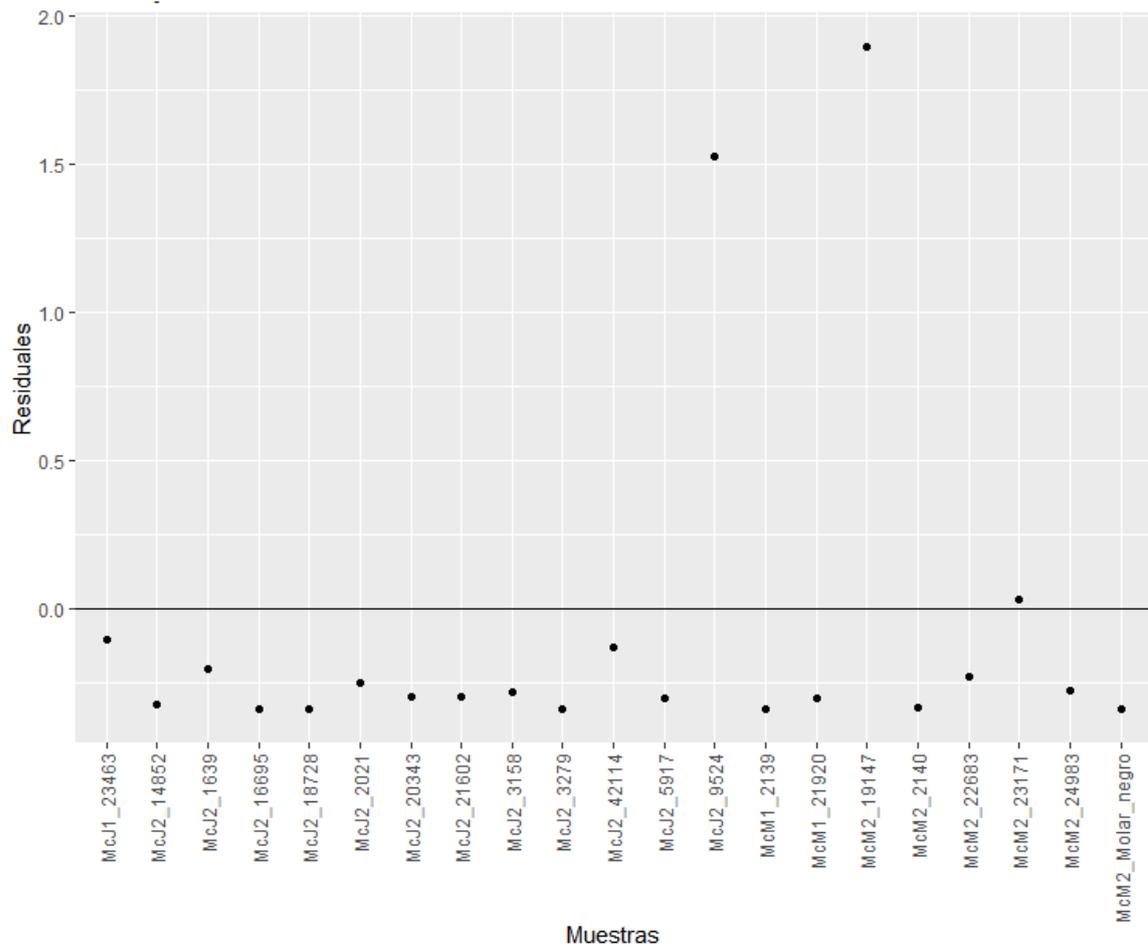


Figura 13. Gráfico de residuales de la relación entre contenido endógeno y el ADN cuantificado por medio de Qubit.

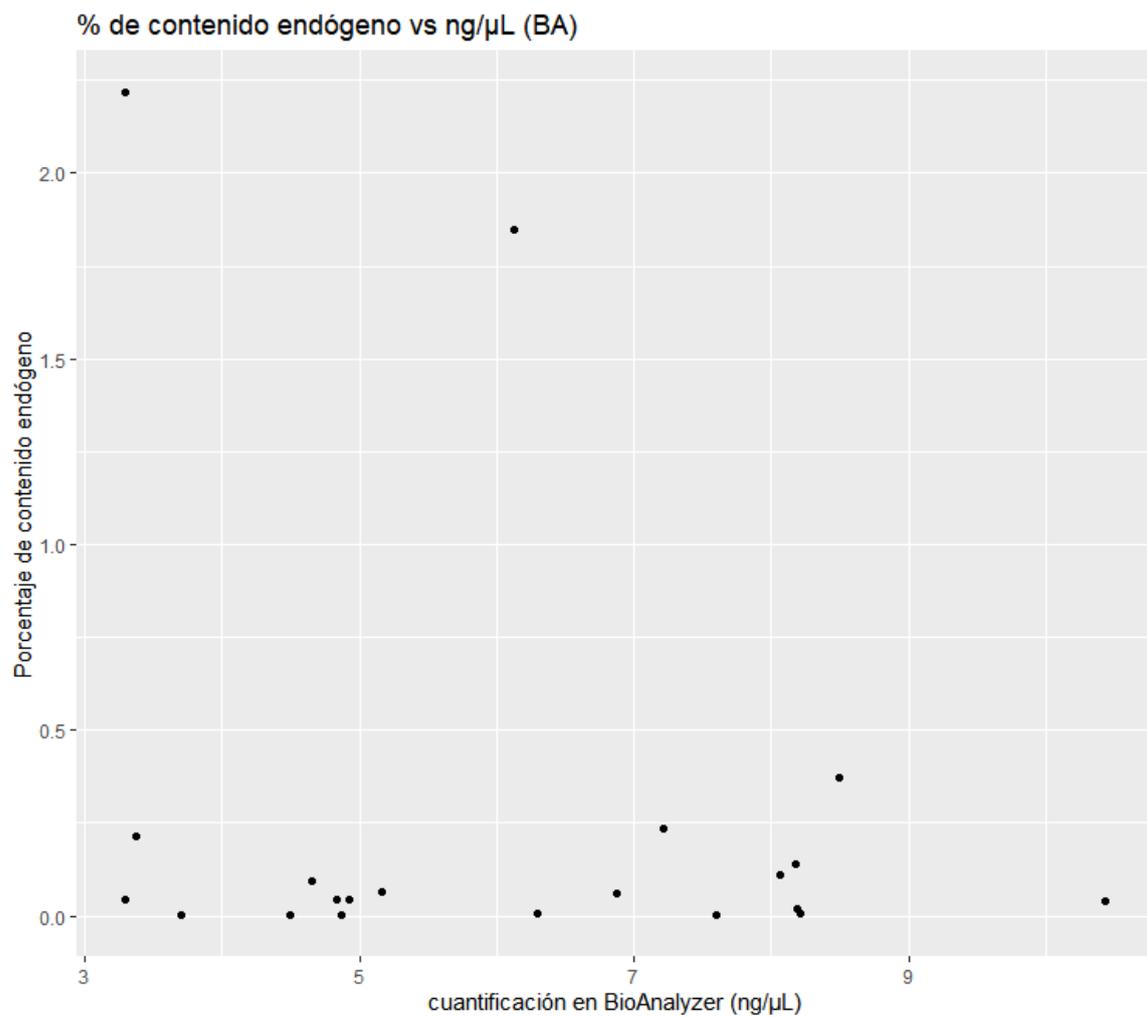


Figura 14. Porcentaje de contenido endógeno con relación a la cuantificación en BioAnalyzer realizada después de indexado de librerías para secuenciación.

Gráfico de residuales de relación de % de contenido endógeno y cuantificación en BioAnalyzer

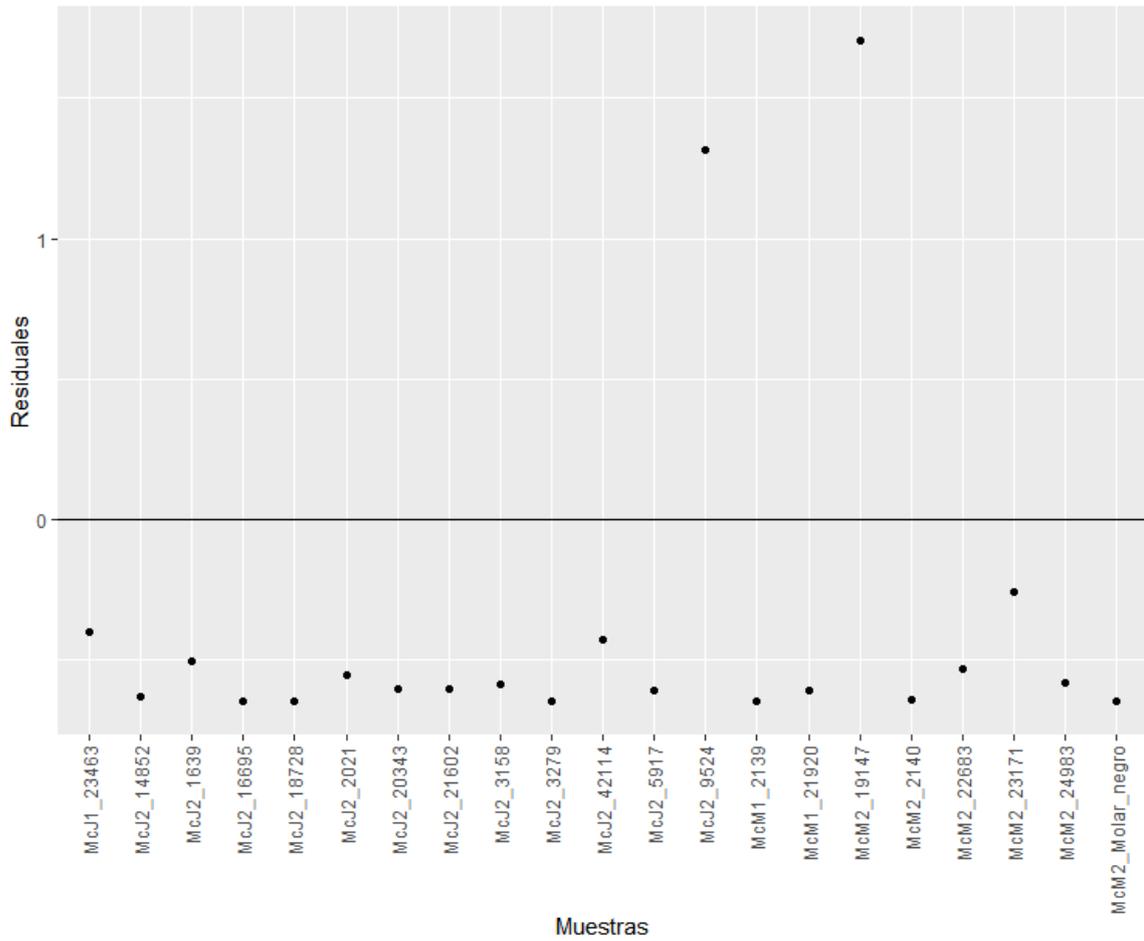


Figura 15. Gráfico de residuales de la relación entre porcentaje de contenido endógeno y cuantificación en BioAnalyzer.

6. DISCUSIÓN

6.1. Retos para obtener contenido endógeno de muestras antiguas.

Como se ha mencionado con anterioridad en este trabajo, la especie *M. columbi*, considerada endémica del continente americano habitó el territorio nacional casi en su totalidad, excepto por la península de Yucatán y el estado de Tabasco, de acuerdo al registro fósil (Arroyo-Cabrales y col., 2007). De todo el territorio, es la cuenca de México donde mayor abundancia de restos existe. Considerando la gran disponibilidad de material en la zona de Santa Lucía, Estado de México además del elevado costo que representa para un laboratorio el procesar muestras para su análisis paleogenómico, y la naturaleza destructiva de los métodos actuales, resulta una cuestión importante el tener una manera de identificar aquellas muestras que tengan una mayor probabilidad de arrojar datos de secuenciación de alta calidad, que es donde yace el propósito de este trabajo.

En general, uno de los mayores retos del campo paleogenómico es la degradación del material genético con el paso del tiempo. Tras la muerte de los organismos, una serie de reacciones químicas tienen lugar, además de que los mecanismos moleculares de reparación del material genético interrumpen su funcionamiento (Shapiro y col., 2019). Como consecuencia de esto, el ADN acumula una gran cantidad de daño, siendo el más importante la fragmentación, pues la longitud de los fragmentos de ADN a suelen ser menores a 100 pb (Orlando y col., 2021), la cual puede ser consecuencia de la presencia de nucleasas, microorganismos, o condiciones climáticas como la temperatura o la humedad (Shapiro y col., 2019).

La hidrólisis también es una de las mayores fuentes de daño al material genético, pues no sólo se escinde directamente el esqueleto de azúcar-fosfato del ADN, generando sitios abásicos, sino que provoca uno de los daños más característicos del ADN, que son las sustituciones $C \rightarrow T$ y $G \rightarrow A$. Esta sustitución ocurre por la desaminación de las bases nitrogenadas, provocando que las citosinas pasen

de cromosoma Y para aumentar el contenido endógeno de dichas librerías. Para comparar de una manera más precisa, se realizó un submuestreo de las lecturas, es decir, se realizaron proyecciones del total de lecturas contra la cantidad de lecturas *on-target*, donde se evidenció que aquellas librerías con mayor contenido endógeno y menor clonalidad tenían mejores rendimientos. En el caso de este estudio, la librería con menos lecturas fue McM2_23171, por lo que el número de lecturas del resto se deberían submuestrear al número que tiene esta muestra. Si eso se realizara, se podría haber hecho un análisis más certero de la calidad y complejidad de las librerías.

6.2. Extracción de ADN

Actualmente no existe una manera certera de determinar qué muestras tendrán un mejor rendimiento en cuanto a contenido endógeno (Baleka, 2019; Damgaard y col., 2015). No obstante, en el estudio de (Wadsworth y col., 2017) se relacionó la cantidad de proteínas con la presencia de ADN, en los que se ha relacionado la presencia de proteínas no colagénicas con la presencia de ADN, a pesar de no encontrar una correlación directa. Por ejemplo, se podría pensar que aquellos extractos que dieron un mayor rendimiento en la cuantificación por medio de Qubit darían un mayor porcentaje de contenido endógeno. Por ejemplo, las muestras McJ2_42114 y McJ2_16695, que tuvieron una concentración de 39.4 y 39.6 ng/ μ L respectivamente arrojaron un contenido endógeno de 0.21 y 0.01% respectivamente. Por otro lado, las dos muestras que dieron el mayor rendimiento, es decir, la McM2_19147 y la McJ2_9524 tuvieron una concentración de 1.01 y 4.16 ng/ μ L respectivamente. Esto también se puede observar en las Figuras 12 y 14, donde se puede ver que no hay una tendencia que correlacione el contenido endógeno y la concentración de ADN del extracto, tanto obtenido por Qubit como en BioAnalyzer.

Como tal, la recuperación de contenido endógeno es complicada debido a que representa sólo una pequeña fracción del ADN total extraído. Aunque existen excepciones (Fu y col., 2014; Seguin-Orlando y col., 2014), en la mayoría de

extractos de ADN el contenido endógeno puede llegar a representar menos del 1% (Damgaard y col., 2015), tal como ocurrió en las muestras que se procesaron, donde sólo 2 del total de muestras procesadas obtuvieron un porcentaje mayor a 1% de contenido endógeno. Esto no significa que el resto de las muestras no se puedan utilizar para posteriores análisis poblaciones, pues en un estudio de Baleka (2019), realizado con restos de un elefante siciliano antiguo, logró obtener solamente 0.04% de contenido endógeno después de realizar un protocolo de captura y enriquecimiento (originalmente había obtenido 0.02%), lo que le permitió realizar la reconstrucción del genoma mitocondrial del animal, sin embargo se tuvieron que desechar el resto de las muestras debido al nulo contenido endógeno que presentaron.

Al menos en cuanto a la extracción, esto sucede porque aunque a la fecha existen muchos métodos novedosos para la extracción de ADN, aún no existe un método que extraiga de manera selectiva el ADN endógeno, sino que se extrae el ADN total de la muestra, el cual puede incluir ADN de microorganismos, ambiental, y ADN contaminante (Damgaard y col., 2015; Orlando y col., 2021) proveniente de la manipulación del laboratorio o proveniente del momento en el que el reto fósil es excavado y extraído para su almacenamiento. No obstante, tomando en cuenta el porcentaje de contaminación humana que se encontró en las muestras tras el mapeo (Cuadro 10), es evidente que el procesamiento en el laboratorio de ADN fue realizado de la manera correcta, por lo que podemos inferir que el ADN no endógeno en los extractos proviene de las fuentes contaminantes anteriormente mencionadas.

6.3. Construcción de librerías genómicas.

Para obtener secuencias de ADN, fue necesario construir librerías genómicas a partir de los extractos. Posteriormente se llevó a cabo la secuenciación en la plataforma Illumina NextSeq 550, que obtiene las lecturas mediante síntesis. Para este trabajo, como se especifica en la sección 4.2.8, se construyeron librerías genómicas de doble cadena tomando en cuenta el protocolo de (Meyer y Kircher,

2010) Este protocolo tiene tres pasos básicos: reparación de extremos, ligación de adaptadores, y rellenado de espacios. El primer paso consiste en hacer que la ADN polimerasa T4 repare los extremos dañados para obtener extremos romos. Aprovechando dichos extremos, se puede realizar la ligación de los adaptadores P5 y P7 a los extremos de las moléculas reparadas con la ayuda de la ADN ligasa T4. Por último, se realiza el rellenado de espacios con ayuda de la Bst polimerasa.

A pesar de que este método ha demostrado ser eficiente y ha sido utilizado en mucho estudios de proboscídeos antiguos (Chang y col., 2017; Dehasque y col., 2021; Pečnerová y col., 2017; van der Valk y col., 2021), una de sus mayores desventajas es que la ligación de adaptadores no es direccional, lo que quiere decir que es posible que una molécula tenga dos adaptadores iguales en ambos extremos, lo que promueve la pérdida de moléculas. Adicionalmente, el protocolo requiere de varios pasos de limpieza, lo que también ocasiona la pérdida de moléculas que pueden contener secuencias únicas del organismo a secuenciar. Esta es otra de las razones que explican el bajo rendimiento de la mayoría de las muestras procesadas.

Existen otros protocolos de librerías de cadena doble, como BEST, que tiene la ventaja de que se realiza en el mismo tubo, utilizando desnaturalización por temperatura en lugar de pasos de limpieza entre reacciones, lo que se ha demostrado que aumenta la complejidad de las librerías, además de reducir el tiempo y costo de procesamiento (Carøe y col., 2018; Kapp y col., 2021)

Otra solución para esta cuestión sería optar por la construcción de librerías de cadena sencilla (Sección 1.6). Esta última metodología ha demostrado tener mayor sensibilidad para muestras con poca biomasa, además de que permite una mayor selección de fragmentos con daño post-mortem presentes en el extracto, lo que da como resultado un mayor rendimiento de contenido endógeno en las librerías (Gansauge y Meyer, 2013; Orlando y col., 2021). Al menos para mamuts, este método se ha utilizado con éxito en el estudio de (van der Valk y col., 2021) para procesar algunas de las muestras que analizaron. Sin embargo, en dicho

estudio también se utilizó el protocolo de Meyer y Kircher para procesar las muestras más antiguas (más de 1 Ma), las cuales se recuperaron del Permafrost. Por otro lado, en el estudio de Baleka (2019), donde se utilizaron muestras que provenían de entornos templados y subtropicales, y la mayoría de las muestras se procesaron siguiendo un protocolo de construcción de librerías de cadena sencilla. Considerando el tipo de clima predominante en México sería conveniente pensar en una estrategia de construcción de librerías de cadena sencilla para aumentar el rendimiento de ADN endógeno de estas muestras. Desafortunadamente, una de las limitaciones de este tipo de metodologías es su alto costo y el tiempo que pueden llegar a tomar (hasta dos días) (Gansauge y col., 2017; Gansauge y Meyer, 2013; Orlando y col., 2021).

Para solucionar este problema, Kapp y col., (2021) han desarrollado una nueva metodología que utiliza una ligación direccional y entablillada, lo que permite que ambos adaptadores se ligan de manera simultánea a ambos lados del fragmento de ADN en la orientación correcta, denominada reacción de Santa Cruz. Esta metodología ha probado ser eficiente, rápida y no tan elevada en costos, lo que permitiría aumentar el rendimiento de las muestras que se utilizaron en este estudio, considerando que se ha utilizado en megafauna proveniente del Pleistoceno, como bisontes, perezosos antiguos (*Mylodon darwini*), e inclusive un oso polar de más de 100,000 años de antigüedad (Shapiro y col., 2004; van Geel y col., 2022; Wang y col., 2022).

6.4. Preservación del ADN en dentina

Tomando en cuenta los objetivos del presente estudio, se realizaron las comparaciones entre la cantidad de contenido endógeno y las características de las piezas de las que se obtuvieron las muestras. Las primeras comparaciones que se hicieron fueron entre el tipo de material del que se obtuvieron, es decir, si provenían de la dentina de un molar en posición anatómica o de la dentina de un molar “suelto”. Al estar en posición anatómica, se creyó posible que la preservación fuera mejor al estar en un ambiente menos expuesto en

comparación con los otros molares. De acuerdo a los resultados obtenidos en la gráfica de caja de la Figura 6 y la prueba estadística descrita en la Sección 5.5, se puede observar que la distribución de las cantidades de contenido endógeno es similar en ambos materiales. En cuanto a los Cuadros 10 y 11, se muestra que, de las dos muestras con mayor contenido endógeno, es decir, la McM2_19147, la que obtuvo una mayor cantidad de contenido endógeno fue un molar “suelto”. No obstante, en la Sección 5.5 se obtuvo que el valor de p de la diferencia entre los dos tipos de materiales no es significativo, por lo que podemos decir que no es determinante en cuanto a la cantidad de contenido endógeno que se puede obtener de ellas, al menos considerando molares en su posición anatómica o separados de la mandíbula.

Esto sucede porque ambas muestras fueron muestras de dentina. El diente de mamut, como se muestra en la Figura 1, está formado por tres tipos de tejido: dentina, cemento y esmalte. A pesar de que la dentina suele ser el material preferido para análisis paleogenómicos, es importante comprender la composición del hueso, así como las diferencias entre estos tres tipos de tejido para tratar de entender las diferencias que puede haber entre la preservación de ADN en cada uno de ellos.

Se debe mencionar que los dientes también están formados por tejido óseo, por lo que es importante conocer la estructura del hueso para entender la preservación de ADN en la dentina. La mayoría de la estructura del hueso está compuesta de dos fracciones: la orgánica y la inorgánica. La fracción inorgánica está compuesta de hidroxiapatita (HAP) carbonatada (donde se postula que el ADN se adsorbe), mientras que la fracción orgánica está compuesta mayoritariamente por colágeno tipo 1 (90% aprox.), así como otras proteínas no colagénicas y glicoproteínas (Campos y col., 2012; Kendall y col., 2018). La bioapatita y el colágeno se relacionan de la siguiente manera: el colágeno forma fibrillas, las cuales se estabilizan unas a otras mediante los enlaces que se forman al agregarse entre sí. Al entrelazarse, las moléculas de colágeno dejan pequeños espacios entre ellas (40 nm aprox), donde la apatita tiene oportunidad de formar

su estructura cristalina, y mineralizar la estructura de colágeno por dentro y sobre la superficie, formando la estructura de los huesos, como se muestra en la Figura 11 (Campos y col., 2012). Es importante también mencionar que este crecimiento es a expensas del agua que pueda contener la estructura fibrilar del colágeno; es decir, a medida que la apatita aumenta su estructura de tamaño, la hidratación del hueso disminuye. El tejido maduro de hueso tiene, en promedio, 46% de colágeno, 46% de HAP y 8% de agua (Campos y col., 2012). No obstante, harían falta estudios del papel que juegan otro tipo de proteínas en la preservación de ADN, pues en el estudio conducido por Cappellini y col. (2012) se encontró que el proteoma del mamut, al menos lo obtenido de una muestra de fémur, estaba compuesto en su mayoría por proteínas de la matriz extracelular no colagénicas y sólo un 9% de colágeno.

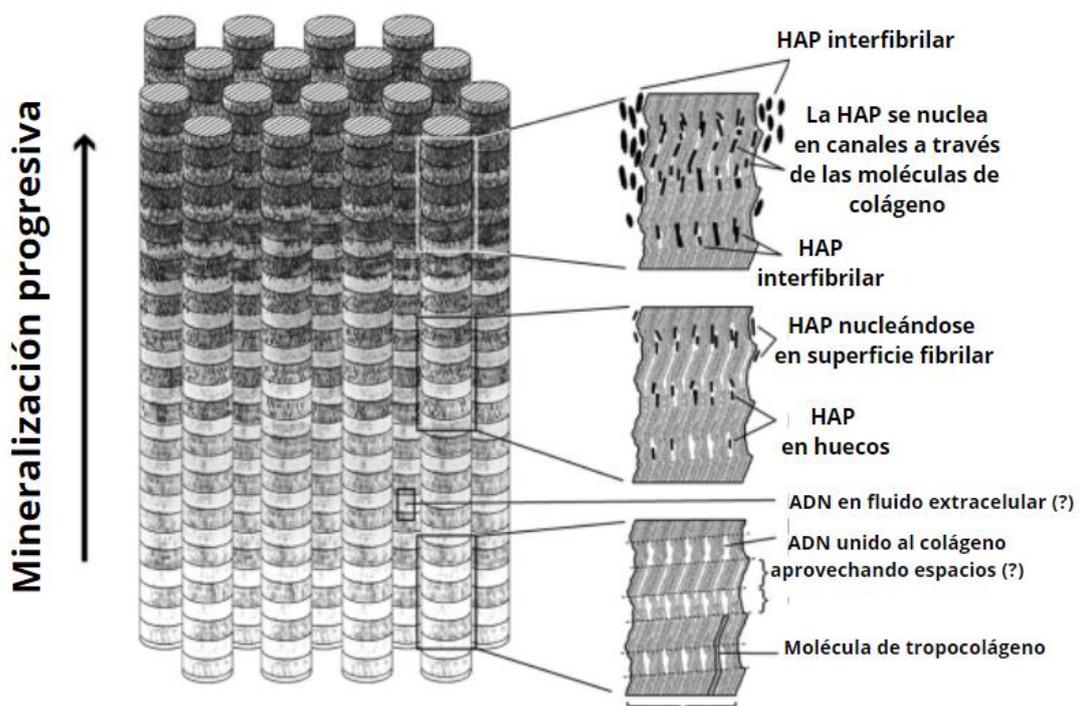


Figura 17. Proceso de mineralización del tejido óseo mediante la unión del colágeno y la HAP. Imagen modificada de Campos y col. (2012).

Algo importante a considerar en cuanto a los restos fósiles es la porosidad del hueso que se analiza. El hueso es un tejido vascularizado, por lo que en cuanto el organismo comienza los procesos de degradación de tejidos blandos, como las venas y arterias, se generan poros que pueden dejar al hueso vulnerable a filtraciones del ambiente, o microorganismos, lo que quiere decir que un tejido mineralizado más compacto tendrá una mejor conservación del material genético (Campos y col., 2012; Kendall y col., 2018).

Observando las fotografías disponibles de las piezas dentales (Cuadro 9), es difícil hacer alguna inferencia a simple vista de la porosidad de las piezas; por ejemplo, la pieza de la muestra McJ2_16695 aparentemente se ve porosa, grisácea, y parece una piedra. Sin embargo, obtuvo un mejor rendimiento que piezas que se observan mejor preservadas, como la McM2_Molar_negro o la McM1_2139. También obtuvo un rendimiento similar a la McM2_2140, que se ve aparentemente mejor preservada. Esto se debe a que el tamaño de los poros es muy pequeño, de la orden de nanómetros (Kendall y col., 2018), por lo que para analizar cada muestra se debería hacer un análisis microscópico, aunque su viabilidad debería considerarse debido a que podría ser costoso para los laboratorios. Esto coincide con estudios previos, donde se ha encontrado que el hueso petroso, el más denso del cuerpo de los mamíferos, suele presentar mejor rendimiento de contenido endógeno que otro tipo de huesos (Gamba y col., 2014; Jørkov y col., 2009). De hecho, en el estudio de Baleka (2019), la muestra que dio mejor rendimiento de ADN endógeno y la única que no fue desechada fue la que provenía del hueso petroso del elefante siciliano. Sería recomendable para obtener un mejor rendimiento de ADN endógeno de los individuos de este estudio comparar el porcentaje de ADN endógeno del hueso petroso de los mamuts.

A la fecha, no se conocen del todo los procesos moleculares que intervienen en la preservación del ADN durante su degradación, sin embargo se sabe que el ADN se preserva al asociarse a la estructura cristalina de la HAP por medio de adsorción, así como al asociarse con la estructura fibrilar del colágeno (a esta

última no se tiene certeza de cómo suceda esta asociación) (Kendall y col., 2018). Comprender mejor los procesos de degradación del tejido dental y de las proteínas dentro de él es de suma importancia para tener más certeza en cuanto a la selección de muestras para su análisis paleogenómico.

En cuanto a los distintos tejidos que hay en los dientes, que también son huesos, podemos encontrar que los que más se han analizado en estudios de ADN y su rendimiento son la dentina y el cemento. En un estudio de 2015, (Damgaard y col., 2015) realizaron la comparación del ADN endógeno que se obtenía del cemento y de la dentina de dientes humanos antiguos. En dicho estudio, encontraron que el mayor rendimiento de ADN endógeno se encontraba en el cemento de las raíces en comparación con la dentina. La explicación que propusieron es que la dentina es la zona del diente que mayor contacto tiene con la cavidad pulpar, que es vulnerable a ataques por parte de microorganismos, lo que puede provocar la mayor degradación del material genético, al contrario del cemento, que se encuentra en la zona más exterior de los dientes (Damgaard y col., 2015).

Aún así, para muestras de mamuts, es complicado llegar a una zona cercana a la cavidad pulpar debido al tamaño de los molares; de hecho, en el Cuadro 9 se puede observar que no todos los molares estaban completos, o no en todos se muestran las raíces (donde estaban las cavidades pulpares). Aún así, se puede observar que la muestra McM2_19147 es un fragmento de raíz de molar, lo que podría explicar su alto rendimiento de contenido endógeno en comparación con el resto de las muestras. De la muestra McJ2_9524 no hay fotografía disponible, por lo que no se puede comparar con respecto al resto de las muestras.

Por otro lado, el esmalte, aunque también forma parte del tejido de los dientes, es casi completamente mineral, por lo que suele preferirse para análisis de isótopos y elementos traza (Kendall y col., 2018). Una mejor comparación entre tipos de materiales sería entre el cemento y la dentina de los molares que se pudieran obtener.

6.5. Sexo de los individuos

De acuerdo a lo observado en los resultados de la Figura 7, se podría decir que el ADN endógeno de las hembras alcanzó un rango mayor que aquel alcanzado por las muestras pertenecientes a individuos machos. En esta gráfica se removieron las dos muestras que mejor rendimiento tuvieron, ambas pertenecientes a hembras, para una mejor visualización de los resultados. En especies con dimorfismo sexual los procesos tafonómicos como la descomposición o erosión pueden llevar a estados diferenciales de preservación de fósiles de machos y hembras (Pečnerová y col., 2017). De cualquier forma, a pesar de que los restos fósiles más grandes suelen tener una degradación más lenta (Behrensmeyer, 1978) (en este caso, los de los machos, que suelen ser de mayor tamaño), no se han encontrado diferencias de velocidad de degradación entre los sexos de megafauna, como los mamuts (Ioannidou, 2003), cuya proporción macho hembra, además, suele ser 1:1 (Pečnerová y col., 2017). Esto está de acuerdo con los resultados obtenidos, donde, a pesar del mayor rango alcanzado por las hembras, la diferencia con los machos no es significativa, lo que quiere decir que la preservación de ADN en los fósiles de mamut no es dependiente del sexo, o al menos no se pudo observar dicha dependencia en estas muestras.

6.6. Condiciones climáticas y del suelo.

Los métodos experimentales y el proceso de degradación del hueso no son los únicos factores involucrados en la conservación del ADN en restos fósiles. Los factores del clima y el suelo juegan un papel sumamente importante en esta cuestión. Ya que el daño al ADN se acumula a lo largo del tiempo, su preservación es dependiente del tiempo, por lo que se espera que a mayor edad haya un menor contenido de material genético (Shapiro y col., 2019). No obstante, esto depende también de la temperatura y la humedad. Por ejemplo, el ADN ha demostrado conservarse muy bien en ambientes como el permafrost (Campos y col., 2010; Shapiro y col., 2004; van der Valk y col., 2021) o en cuevas (Bollongino y col., 2008), donde la temperatura es baja, constante y la humedad mínima. Hace

algunos años no se pensaba que en regiones con climas cálidos y húmedos se pudiera preservar el ADN, sin embargo, estudios demuestran que sí es posible encontrar ADN en estas regiones (Bravo-Lopez y col., 2020; Guzmán-Solís y col., 2021; Taketani y col., 2022; Villa-Islas y col., 2022).

Además de la temperatura y la humedad, es importante tomar en cuenta el suelo donde los restos antiguos fueron depositados. De acuerdo a los resultados obtenidos, se encontró que al menos para nuestras muestras, no existe correlación ni relación lineal entre la profundidad a la que las muestras fueron excavadas y el porcentaje de contenido endógeno presente en ellas. Aún así, las dos muestras con mayor porcentaje de contenido endógeno están a profundidades menores que aquellas que no presentaron o presentaron muy bajo contenido endógeno (Cuadros 10 y 11), por lo que teniendo una cantidad mayor de muestras se podrían establecer cortes de profundidad en los que se podría esperar un mejor rendimiento.

Estos resultados están de acuerdo con el estudio de sedaDNA (ADN antiguo sedimentario) realizado por Moguel y col. (2021), en el que se realizaron análisis metagenómicos de la diversidad existente en el lago de Chalco durante el Holoceno. En dicho estudio, se observó que a medida que se alejaban de la superficie, se encontraba menor biodiversidad, probablemente debido a que poco ADN podía sobrevivir a grandes profundidades. Por otro lado, Kendall y col., (2018), establece que los restos que son enterrados a grandes profundidades suelen sobrevivir por más tiempo, principalmente debido al poco acceso al oxígeno, lo que las hace inaccesibles a bacterias aeróbicas que pueden llegar a introducirse en los poros del hueso, y por tanto, a degradar el complejo de colágeno y HAP del hueso, lo que afectaría directamente la cantidad de ADN que se puede obtener de los restos. También es necesario recordar que las placas de la tierra están en constante movimiento, lo que puede ocasionar que ciertos restos queden expuestos u ocultos dependiendo del movimiento que se produzca (Johnson y Harrison, 1990). Los restos de los mamuts no fueron enterrados intencionalmente, de hecho, de acuerdo a los informes del equipo de

paleontología, se debe determinar si los mamuts cayeron en trampas naturales y fueron después sepultados a lo largo del tiempo, por lo que se podría esperar, al menos en este caso, que aquellos restos que se encuentren a mayor profundidad sean más antiguos, y por lo tanto, tengan un menor rendimiento de ADN.

Además de eso, se deben considerar las características del suelo donde los restos fueron encontrados. De acuerdo al informe anual del proyecto “La prehistoria y paleoambientes del noreste de la Cuenca de México” (no publicado), del cual este proyecto forma parte, los estudios tafonómicos para los restos de mamut encontrados en Santa Lucía no han sido realizados. Un estudio tafonómico se refiere al análisis de la fosilización de algún resto, es decir, de los procesos que influyen la preservación de un cuerpo tras su muerte. Esto puede incluir analizar los organismos que intervienen en la descomposición, las condiciones del entierro, los elementos, del suelo, etc.

A pesar de que no existen dichos estudios sobre los individuos de este estudio, el antecedente más cercano es el del mamut de Santa Ana Tlacotenco. Gracias a este estudio se logró definir que el mamut cayó en una trampa, donde murió y posteriormente sus restos fueron sepultados por una serie de lahares, que son flujos de sedimentos volcánicos que se movilizan desde las laderas de los volcanes (Guilbaud y col., 2015).

Los estudios tafonómicos son de suma importancia para elucidar las condiciones en las que un cuerpo se descompuso a través del tiempo. En el estudio de Guilbaud se realizaron análisis del pH y de otros elementos que podrían haber afectado la composición del suelo. Al menos para este estudio, sería importante conocer el pH del suelo donde se enterraron los muestras, pues se ha demostrado que cuando los cuerpos se descomponen, modifican el pH del suelo, volviéndolo más ácido al producir los ácidos húmico y tánico (Gill-King, 1997.; Janaway, 2008). Los suelos ácidos son más agresivos que los suelos alcalinos con los restos fósiles, mermando su conservación, esto a pesar de que el colágeno es más susceptible al pH alcalino que al ácido (Kendall y col., 2018).

Otro aspecto a considerar sobre estas muestras es que, al igual que el mamut de Santa Ana Tlacotenco, se encontraron en una zona volcánica, y de hecho, de acuerdo al informe anual del proyecto, en la zona de excavación se encontró una capa de ceniza volcánica. Por ejemplo, en el estudio de (Guilbaud y col., 2015) se encontró que la ceniza volcánica no tuvo efecto en modificar el pH del suelo respecto a otras zonas donde no había ceniza volcánica. Sin embargo, pudo haber tenido un efecto “protector” al aislar al cadáver del exterior y haber sido enterrado debajo de esta. Esto es algo que se ha visto en los cadáveres encontrados en Pompeya, de donde se ha extraído ADN con éxito, y donde se ha sugerido que las cenizas volcánicas han tenido un papel importante en el aislamiento de los restos para su protección contra microorganismos (Cipollaro y col., 2005; Scorrano y col., 2022).

Otro aspecto que se puede comparar es el color de los restos, que puede aportar información sobre las características del suelo de donde los restos fueron obtenidos. Las capas del suelo pueden exhibir distintos regímenes químicos, por ejemplo, en las capas superiores hay más presencia de oxígeno, y a medida que aumenta la profundidad pueden tener lugar reacciones de reducción de Mn^{4+} , Fe^{3+} , y en los niveles más profundos, reducción de sulfatos, y presencia de metano. Los huesos que están en las capas anóxicas suelen presentar colores cafés o negros debido a la adsorción de los óxidos de hierro y manganeso a la superficie, sumado a la interacción con los ácidos húmico y fúlvico, que también pueden dar al hueso tal coloración (Fernandez-Jalvo y Andrews, 2016; Kendall y col., 2018). También, por ejemplo, el color negro puede estar relacionado al ataque por parte de hongos, lo que puede provocar la degradación del ADN (Fernandez-Jalvo y Andrews, 2016). Esto es evidente en la muestra McM2_Molar_negro, pero también en la McJ2_2021, la McM1_21920 y la McM1_2139. Por otro lado, la muestra con mejor rendimiento de ADN, es decir, la aMcM2_19147, muestra un color naranja brillante (Cuadro 9), lo que indica que la tierra era rica en óxido de hierro (Fernandez-Jalvo y Andrews, 2016). Aún así, se debe reconocer que el número de muestras es reducido, por lo que no se puede establecer de manera

certera esta correlación, sería bueno tener más muestras con características similares, y evaluarlas.

6.7. Preservación de ADN

Al día de hoy no se conoce a ciencia cierta cuáles son los mecanismos de degradación o conservación del ADN dentro del tejido del hueso. (Allentoft y col., 2012) realizaron un estudio que intentó dilucidar el decaimiento exponencial del ADN en el que compararon 158 muestras antiguas datadas con C¹⁴. En dicho estudio, si bien pudieron establecer que el ADN mitocondrial se preserva mejor que el nuclear, no se estableció una relación concluyente entre la edad de las muestras y la degradación de ADN. Los autores concluyeron que la edad no era necesariamente determinante en la conservación de ADN, pero que era una mezcla de las condiciones tafonómicas (profundidad del entierro, pH, flujo de agua, fluctuaciones del clima, incorporación del cadáver al sedimento), y las condiciones del tejido óseo.

También en el 2017, (Kistler y col., 2017) realizaron un meta análisis tomando en consideración datos genómicos existentes para analizar el decaimiento del ADN a lo largo del tiempo. Ellos determinaron que no existe una correlación entre la cantidad de fragmentación de ADN y la edad, aún cuando las condiciones ambientales son controladas, por lo que resaltaron la importancia de conocer cómo las estructuras del tejido óseo y el ambiente crean un ambiente “cerrado” que permita una mejor preservación del material genético.

Esto no quiere decir que la edad no se pueda tomar en consideración. Los resultados de datación de las muestras que se analizaron en este estudio no están disponibles, sin embargo, no se descarta que se pueda encontrar alguna correlación entre la edad y el porcentaje de contenido endógeno que se pueda encontrar de ellas, considerando además que los estudios de Allentoft y col. (2012) y Kistler y col. (2017) consideraron en su mayoría muestras obtenidas en lugares que presentan climas fríos. Otra consideración en cuanto a edad, podría

ser la edad de los individuos, es decir, si se trataba de un organismo adulto, infantil, o juvenil, considerando que la dentadura de los proboscídeos pasa por un proceso de renovación a medida que estos animales se desarrollan. En este estudio sólo tenemos la certeza de que los individuos correspondientes a las muestras McM1_2139, McM2_2317, McJ2_952, McM2_2140 y McJ2_18728 eran considerados individuos infantiles. Son muy pocas muestras para realizar una aseveración sobre su rendimiento, a pesar de que una de ellas tuvo uno de los dos mejores porcentajes (McJ2_9524, 1.85%), sin embargo, sería de interés realizar una comparación entre estas muestras y el resto si se tuviera la edad del resto de los individuos.

6.8. Importancia y perspectivas de este estudio

La paleogenómica es un área del conocimiento que ha tenido un crecimiento exponencial en las últimas décadas. Es una disciplina que abre una ventana hacia el pasado para poder responder preguntas que emergen en el presente. Los mamuts son un género de proboscídeos que, como se dijo anteriormente, poblaron casi todos los continentes y cuya extinción sigue siendo tema de debate entre científicos. A pesar de que las especies *M. trogontherii* y *M. primigenius* han sido ampliamente estudiadas debido a la amplia disponibilidad de restos, y su mejor conservación gracias a los climas fríos donde se encuentran, esto ha generado un sesgo respecto al conocimiento que existe con respecto a *M. columbi* y su historia evolutiva. A pesar de esto, creemos que es posible que este sesgo se reduzca, pues gracias a las nuevas tecnologías y al desarrollo de nuevos métodos experimentales, como los presentados aquí, será posible echar un vistazo al panorama del Pleistoceno en México a través de los datos generados a partir de estas muestras. Por ejemplo, se podrán realizar estudios de reconstrucción de paleoambientes, afinidades genéticas, estructuras poblacionales, y se podrán generar todavía más datos si las librerías construidas para este estudio son sometidas a experimentos de captura y enriquecimiento, con lo que se podrán obtener datos de mayor resolución.

Es de suma importancia reconocer que el análisis de datos no es posible sin la generación de estos, y para eso es necesario que los métodos experimentales sean realizados con precisión, desde el muestreo hasta la construcción de librerías. A través de este trabajo se puede comprender el papel que juega cada uno de estos pasos en la generación de datos paleogenómicos de calidad. También se hace evidente que es necesario realizar más estudios sobre los mecanismos moleculares y bioquímicos que rigen la conservación del ADN dentro del tejido óseo, pues si se tiene mayor entendimiento de estos mecanismos, se podría realizar una mejor selección de aquellas muestras que se deben priorizar para su análisis. Esto es necesario especialmente en lugares cercanos a los trópicos como México, considerando además la naturaleza destructiva de los muestreos y las cuestiones éticas que implica la destrucción de las muestras, pues son patrimonio invaluable.

Por último, hasta nuestro conocimiento, este es el primer estudio paleogenómico de *M. columbi* en México que se ha realizado. A pesar de que las condiciones climáticas son retadoras, este estudio demuestra que es posible realizar paleogenómica sobre *M. columbi* en lugares donde hace algunos años no se pensaba que se pudiera recuperar este material genético. Es emocionante pensar en todo el conocimiento que esta disciplina tiene el potencial de generar para la mejor comprensión del género *Mammuthus* y el orden de los proboscídeos.

7. CONCLUSIONES

Se lograron obtener muestras de molares y mandíbulas pertenecientes a la especie *M. columbi*, de las cuales se extrajo el ADN, y se construyeron librerías genómicas para su secuenciación en la plataforma Illumina NextSeq550.

Se secuenciaron las muestras de ADN, se mapearon contra la referencia Loxv4 perteneciente a *Loxodonta africana* y se autenticó la naturaleza antigua de dichas secuencias, lo que representa la primera evidencia paleogenómica de *M. columbi* en México, algo que debido a las condiciones climáticas de este país y edad de las muestras se llegó a pensar que no podría realizarse.

Se obtuvieron estadísticos de secuenciación, poniendo especial atención al porcentaje de contenido endógeno, ya que representa la fracción perteneciente a los individuos. De acuerdo a esos datos, se realizaron comparaciones del rendimiento considerando el tipo de material (molar suelto y molar en posición anatómica), sexo y profundidad de excavación.

Se observó que para estas muestras no existe una relación de tipo lineal entre el rendimiento de contenido endógeno y las variables evaluadas. No obstante, es necesario considerar más datos que aún no están disponibles, como condiciones tafonómicas, edad de las muestras y características físicas del tejido óseo de dichas muestras, pues es la combinación de todos esos factores lo que realmente dicta la preservación del ADN a lo largo del tiempo.

8. BIBLIOGRAFÍA

Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, Campos PF, Samaniego JA, Gilbert MTP, Willerslev E, Zhang G, Scofield RP, Holdaway, RN, y Bunce M. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings. Biological Sciences / The Royal Society*. **2012**; 279: 4724–4733.

Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, Malaspinas AS, Margaryan A, Higham T, Chivall D, Lynnerup N, Harvig L, Baron J, Della Casa P, Dąbrowski P, Willerslev E. Population genomics of Bronze Age Eurasia. *Nature*. **2015**; 522:167–172.

Arroyo-Cabrales J, Johnson E, Morett L. Mammoth bone technology at Tocuila in the Basin of Mexico. *The World of Elephants. Proceedings of the 1st International Congress*. **2001**; 419–423.

Arroyo-Cabrales J, Polaco OJ, Laurito C, Johnson E, Teresa Alberdi M, Valerio-Zamora AL. The proboscideans (Mammalia) from Mesoamerica. *Quaternary international: the journal of the International Union for Quaternary Research*. **2007**; 169-170:17–23.

Ávila-Arcos MC, Sandoval-Velaco M, Schroeder H, Carpenter ML, Malaspinas A, Waless N, Peñaloza F, Bustamante CD, Gilbert MT. Comparative performance of two whole-genome capture methodologies on ancient DNA Illumina libraries. *Methods in Ecology and Evolution/British Ecological Society*. **2015**; 6-6: 725-734.

Baleka SI. Palaeogenetic analyses of extinct Elephantidae from temperate and subtropical climates [monografía en internet]. Universität Potsdam, 2019. [Consultado 2022 julio 25]. Disponible en <https://publishup.uni-potsdam.de/frontdoor/index/index/docId/43056>

Behrensmeyer AK. Taphonomic and ecologic information from bone weathering. *Paleobiology*. **1978**; 4:150–162.

Boessenkool S, Hanghøj K, Nistelberger, HM, Der Sarkissian C, Gondek, AT, Orlando L, Barrett JH, Star. Combining bleach and mild predigestion improves ancient DNA recovery from bones. *Molecular Ecology Resources*. **2017**; 17:742–751.

Bollongino R, Tresset A, Vigne JD. Environment and excavation: Pre-lab impacts on ancient DNA analyses. *Comptes Rendus. Palevol*. **2008**; 7: 91–98.

Brace S, Palkopoulou E, Dalén L, Lister AM, Miller R, Otte M, Germonpré M, Blockley, SPE, Stewart JR, Barnes I. Serial population extinctions in a small mammal indicate Late Pleistocene ecosystem instability. *Proceedings of the National Academy of Sciences of the United States of America*, **2012**; 109: 20532–20536.

Bravo-Cuevas VM, Ortiz-Caballero E, Jiménez-Hidalgo E, Barrón-Ortiz CI, Theodor JM. Taxonomía y hábito alimentario de ejemplares de *Mammuthus columbi* (Proboscidea: Elephantidae) del centro y sur de México. *Boletín de la Sociedad Geológica Mexicana*. **2020**; 1:72.

Bravo-López M, Villa-Islas V, Rocha Arriaga C, Villaseñor-Altamirano AB, Guzmán-Solís A, Sandoval-Velasco M, Wesp JK, Alcantara K, López-Corral A, Gómez-Valdés J, Mejía E, Herrera A, Meraz-Moreno A, Moreno-Cabrera, M de la L, Moreno-Estrada A, Nieves-Colón MA, Olvera J, Pérez-Pérez J, Iversen KH, Ávila-Arcos MC. Paleogenomic insights into the red complex bacteria *Tannerella forsythia* in Pre-Hispanic and Colonial individuals from Mexico. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. **2020**; 375: 20190580.

Campos PF, Craig OE, Turner-Walker G, Peacock E, Willerslev E, Gilbert MTP. DNA in ancient bone - where is it located and how should we extract it? *Annals of Anatomy = Anatomischer Anzeiger: Official Organ of the Anatomische Gesellschaft*. **2012**; 194: 7–16.

Campos PF, Willerslev E, Sher A, Orlando L, Axelsson E, Tikhonov A, Aaris-Sørensen K, Greenwood AD, Kahlke RD, Kosintsev P, Krakhmalnaya T,

Kuznetsova T, Lemey P, MacPhee R, Norris CA, Shepherd K, Suchard MA, Zazula GD, Shapiro B, Gilbert MTP. Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **2010**; 107: 5675–5680.

Cappellini E, Jensen LJ, Szklarczyk D, Ginolhac A, da Fonseca RAR, Stafford TW, Holen SR, Collins MJ, Orlando L, Willerslev E, Gilbert MTP, Olsen JV Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. *Journal of Proteome Research*. **2012**; 11: 917–926.

Carøe C, Gopalakrishnan S, Vinner L, Mak SST, Sinding MHS, Samaniego JA, Wales N, Sicheritz-Pontén T, Gilbert MTP. Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution / British Ecological Society*. **2018**; 9: 410–419.

Chang D, Knapp M, Enk J, Lippold S, Kircher M, Lister A, MacPhee RDE, Widga C, Czechowski P, Sommer R, Hodges E, Stümpel N, Barnes I, Dalén L, Derevianko A, Germonpré M, Hillebrand-Voiculescu A, Constantin S, Kuznetsova T, Shapiro B. The evolutionary and phylogeographic history of woolly mammoths: a comprehensive mitogenomic analysis. *Scientific Reports*. **2017**; 7: 44585.

Cipollaro M, Galderisi U, Di Bernardo G. Ancient DNA as a multidisciplinary experience. *Journal of Cellular Physiology*. **2005**; 202: 315–322.

Cruz-Dávalos DI, Nieves-Colón MA, Sockell A, Poznik GD, Schroeder H, Stone AC, Bustamante CD, Malaspina AS, Ávila-Arcos MC. In-solution Y-chromosome capture-enrichment on ancient DNA libraries. *BMC genomics*. **2018**; 19: 608.

Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga JL, Meyer M. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America*. **2013**; 110:15758–15763.

Daley T y Smith AD. Predicting the molecular complexity of sequencing libraries. *Nature Methods*. **2013**; 10: 325-327.

Damgaard PB, Margaryan A., Schroeder H, Orlando L, Willerslev E, Allentoft M E. Improving access to endogenous DNA in ancient bones and teeth. *Scientific Reports*. **2015**; 5:11184.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham, A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. *GigaScience*. **2021**; 10: giab008.

Dehasque M, Pečnerová P, Muller H, Tikhonov A, Nikolskiy P, Tsigankova VI, Danilov GK, Díez-del-Molino D, Vartanyan S, Dalén L, Lister AM. Combining Bayesian age models and genetics to investigate population dynamics and extinction of the last mammoths in northern Siberia. *Quaternary Science Reviews*. **2021**; 59: 265-269.

Enk J, Devault A, Widga C, Saunders J, Szpak P, Southon J, Rouillard JM, Shapiro B, Golding GB, Zazula G, Froese D, Fisher DC, MacPhee RDE, Poinar H. Mammuthus Population Dynamics in Late Pleistocene North America: Divergence, Phylogeography, and Introgression. *Frontiers in Ecology and Evolution*. **2016**; 4:42.

Fernandez-Jalvo Y, Andrews P. *Atlas of Taphonomic Identifications: 1001+ Images of Fossil and Recent Mammal Bone Modification*. Nueva York: Springer. **2016**.

Fisher DC. Paleobiology of Pleistocene Proboscideans. *Annual Review of Earth and Planetary Sciences*. **2018**; 46:229-260.

Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PLF, Aximu-Petri A, Prüfer K, de Filippo C, Meyer M, Zwyns N, Salazar-García D C, Kuzmin YV, Keates SG, Kosintsev PA, Razhev DI, Richards MP, Peristov NV, Pääbo S. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*. **2014**; 514: 445–449.

Gamba C, Hanghøj K, Gaunitz C, Alfarhan AH, Alquraishi SA, Al-Rasheid KAS,

Bradley DG, Orlando L. Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Molecular Ecology Resources*. **2016**; 16:459–469.

Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, Domboróczki L, Kővári I, Pap I, Anders A. Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications* **2014**; 5: 5257.

Gansauge MT, Gerber T, Glocke I, Korlevic P, Lippik L, Nagel S, Riehl LM, Schmidt A, Meyer M. Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Research*. **2017**; 45: 79.

Gansauge MT, Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols*. 2013; 8: 737–748.

Gill-King. Chemical and ultrastructural aspects of decomposition. *Forensic Taphonomy: The Postmortem Fate of Human*. Reino Unido: CRC Press. **1997**.

Glocke I, Meyer M. Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome Research*. **2017**; 27: 1230–1237.

Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews. Genetics*. 2016; 17: 333–351.

Graham F. Daily briefing: Dozens of mammoths found under the new Mexico City airport. [monografía en internet]. *Nature*, 2020. [Consultado 2021 julio 26]. Disponible en: <https://doi.org/10.1038/d41586-020-01663-4>

Green MR, Sambrook J. Isolation and Quantification of DNA. *Cold Spring Harbor Protocols*. **2018**; 6:403-414

Guilbaud MN, Arana-Salinas L, Siebe C, Barba-Pingarrón LA, Ortiz A. Volcanic stratigraphy of a high-altitude *Mammuthus columbi* (Tlacotenco, Sierra Chichinautzin), Central México. *Bulletin of Volcanology*, **2015**; 77: 1-16.

Guzmán-Solís AA, Villa-Islas V, Bravo-López MJ, Sandoval-Velasco M, Wesp J K, Gómez-Valdés JA, Moreno-Cabrera M de la L, Meraz A, Solís-Pichardo G, Schaaf P, TenOever BR, Blanco-Melo D, Ávila Arcos MC. Ancient viral genomes reveal introduction of human pathogenic viruses into Mexico during the transatlantic slave trade. *eLife*. **2021**; 10: 1-16.

Harney É, Cheronet O, Fernandes DM, Sirak K, Mah M, Bernardos R, Adamski, N, Broomandkhoshbacht N, Callan K, Lawson AM, Oppenheimer J, Stewardson K, Zalzal F, Anders A, Candilio F, Constantinescu M, Coppa A, Ciobanu I, Dani J, Pinhasi R. A minimally destructive protocol for DNA extraction from ancient teeth. *Genome Research*. **2021**; 31:472–483.

Hofreiter M. Nondestructive DNA extraction from museum specimens. *Methods in Molecular Biology*. **2012**; 840:93–100.

Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*. **2001**; 29: 4793–4799.

Höss M, Pääbo S. DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Research*. **1993**. 21:3913–3914.

Ioannidou E. Taphonomy of Animal Bones: Species, Sex, Age and Breed Variability of Sheep, Cattle and Pig Bone Density. *Journal of Archaeological Science*. **2003**; 30: 355–365.

Janaway RC. The decomposition of materials associated with buried cadavers. *Soil Analysis in Forensic Taphonomy*. Florida: CRC Press. 2008.

Johnson CA, Harrison CGA. Neotectonics in central Mexico. *Physics of the Earth and Planetary Interiors*. **1990**; 64: 187–210.

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.

Jørkov MLS, Heinemeier J, Lynnerup N. The petrous bone--a new sampling site for identifying early dietary patterns in stable isotopic studies. *American Journal of Physical Anthropology*. **2009**; 138: 199–209.

Kapp JD, Green RE, Shapiro B. A Fast and Efficient Single-stranded Genomic Library Preparation Method Optimized for Ancient DNA. *The Journal of Heredity*. **2021**; 112: 241–249.

Kendall C, Eriksen AMH, Kontopoulos I, Collins MJ, Turner-Walker G. Diagenesis of archaeological bone and tooth. *Palaeogeography, Palaeoclimatology, Palaeoecology*. **2018**; 491:21–37.

Kistler L, Ware R, Smith O, Collins M, Allaby RG. A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Research*. **2017**; 45: 6310–6320.

Kobun R. Extraction and analytical methods for the identification of parasites in food. En *Advanced Food Analysis Tools*. Oxford: Academic Press. **2021**: 101-125.

Lee PC, Moss CJ. The social context for learning and behavioural development among wild African elephants. En *Mammalian social learning: Comparative and ecological perspectives*. Cambridge: Cambridge University Press. **1999**: 102–125.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics*. **2009**; 25: 2078-2079.

Lindgreen S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Research Notes*, **2012**; 5: 337.

Lister AM. Evolution and taxonomy of Eurasian mammoths. *The Proboscidea: evolution and palaeoecology of elephants and their relatives*. Oxford: Oxford University Press. **1996**: 203–213.

Lister AM, Sher AV. Evolution and dispersal of mammoths across the Northern Hemisphere. *Science*. **2015**; 350:805–809.

Lister AM, Sher AV, van Essen H, Wei G. The pattern and process of mammoth evolution in Eurasia. *Quaternary international: the journal of the International Union for Quaternary Research*. **2005**; 126-128: 49–64.

McDaniel G E, Jefferson GT. Dental variation in the molars of *Mammuthus columbi* var. *M. imperator* (Proboscidea, Elephantidae) from a Mathis gravel quarry, southern Texas. *Quaternary International: The Journal of the International Union for Quaternary Research*. **2006**; 142: 166–177.

Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*. 2010; 6: 5448.

Miller W, Drautz DI, Ratan A, Pusey B, Qi J, Lesk AM, Tomsho LP, Packard M D, Zhao F, Sher A, Tikhonov A, Raney B, Patterson N, Lindblad-Toh K, Lander ES, Knight JR, Irzyk GP, Fredrikson KM, Harkins TT, Schuster SC. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature*. **2008**; 456:387–390.

Moguel B, Pérez L, Alcaraz LD, Blaz J, Caballero M, Muñoz-Velasco I, Becerra A, Lacleste JP, Ortega-Guerrero B, Romero-Oliva CS, Herrera-Estrella L, Lozano-García S. Holocene life and microbiome profiling in ancient tropical Lake Chalco, Mexico. *Scientific Reports*. **2021**; 11:13848.

Morrone JJ. *Encyclopedia of Ecology. Endemism*. Oxford: Elsevier. **2008**.

Orlando L, Allaby R, Skoglund P, Der Sarkissian C, Stockhammer PW, Ávila-Arcos MC, Fu Q, Krause J, Willerslev E, Stone AC, Warinner C. Ancient DNA analysis. *Nature Reviews Methods Primers*. **2021**; 1:1–26.

Orlando L, Gilbert MTP, Willerslev E. Reconstructing ancient genomes and epigenomes. *Nature Reviews. Genetics*. **2015**; 16: 395–408.

Palkopoulou E, Lipson M, Mallick S, Nielsen S, Rohland N, Baleka S, Karpinski, E, Ivancevic AM, To TH, Kortschak RD, Raison JM, Qu Z, Chin TJ, Alt KW, Claesson S, Dalén L, MacPhee RDE, Meller H, Roca AL, Reich D. A comprehensive genomic history of extinct and living elephants. *Proceedings of the National Academy of Sciences of the United States of America*. **2018**; 115: E2566–E2574.

Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, Li H, Omrak A, Vartanyan S, Poinar H, Götherström A, Reich D, Dalén L. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current Biology: CB*. **2015**; 25:1395–1400.

Pečnerová P, Díez-Del-Molino D, Dussex N, Feuerborn T, von Seth J, van der Plicht J, Nikolskiy P, Tikhonov A, Vartanyan S, Dalén L. Genome-Based Sexing Provides Clues about Behavior and Social Structure in the Woolly Mammoth. *Current Biology: CB*. **2017**; 27:3505–3510.

Pérez-Crespo VA, Arroyo-Cabrales J, Benammi M, Johnson E, Polaco O J, Santos-Moreno A, Morales-Puente P, Cienfuegos-Alvarado E. Geographic variation of diet and habitat of the Mexican populations of Columbian Mammoth (*Mammuthus columbi*). *Quaternary international: the journal of the International Union for Quaternary Research*. **2012**; 276: 8–16.

Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RDE, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, Rampp M, Miller W, Schuster SC. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science*. **2006**; 311: 392–394.

Quiles JL, Romero-Márquez JM, Navarro-Hortal MD, Battino M, Varela-López A. *Aging*. Nueva York: Academic Press. **2020**.

RStudio Team, RStudio: Integrated Development for R. [monografía en internet], RStudio PBC, **2020**. [Consultado 2021 octubre 10]. Disponible en <http://www.rstudio.com/>.

Rohland N, Glocke I, Aximu-Petri A, Meyer M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nature Protocols*. **2018**; 13: 2447–2461.

Rohland N, Hofreiter M. Ancient DNA extraction from bones and teeth. *Nature Protocols*. **2007**; 2: 1756–1762.

Rohland N, Siedel H, Hofreiter M. Nondestructive DNA extraction method for

mitochondrial DNA analyses of museum specimens. *BioTechniques*. **2004**; 36: 814–816, 818–821.

Schwarz C, Debruyne R, Kuch M, McNally E, Schwarcz H, Aubrey AD, Bada J, Poinar H. New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Research*. **2009**; 37: 3215–3229.

Scorrano G, Viva S, Pinotti T, Fabbri PF, Rickards O, Macciardi F. Bioarchaeological and palaeogenomic portrait of two Pompeians that died during the eruption of Vesuvius in 79 AD. *Scientific Reports*. **2022**; 12: 6468.

Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspinas AS, Manica A, Moltke I, Albrechtsen A, Ko A, Margaryan A, Moiseyev V, Goebel T, Westaway M, Lambert D, Khartanovich V, Wall JD, Nigst PR, Foley RA, Lahr MM, Nielsen R, Willerslev E. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science*. **2014**; 346: 1113–1118.

Shapiro B, Barlow A, Heintzman PD, Hofreiter M, Paijmans JLA, Soares AER. *Ancient DNA: Methods and Protocols*. New York: Springer New York. **2019**.

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MTP, Barnes I, Binladen J, Willerslev E, Hansen AJ, Baryshnikov GF, Burns JA, Davydov S, Driver JC, Froese DG, Harington C R, Keddie G, Cooper A. Rise and fall of the Beringian steppe bison. *Science*. **2004**; 306: 1561–1565.

Shapiro B, Hofreiter M. *Ancient DNA: Methods and Protocols*. New York: Humana Press. **2012**.

Taketani RG, Beier S, Dini-Andreote F, Fernandez C. Advancements in the Understanding of Anthropogenic Impacts on the Microbial Ecology and Function of Aquatic Environments. *Frontiers in Microbiology*. **2022**; 12: 1-14.

van der Valk T, Pečnerová P, Díez-Del-Molino D, Bergström A, Oppenheimer J, Hartmann S, Xenikoudakis G, Thomas JA, Dehasque M, Sağlıcan E, Fidan FR, Barnes I, Liu S, Somel M, Heintzman PD, Nikolskiy P, Shapiro B, Skoglund P,

Hofreiter M, Dalén L. Million-year-old DNA sheds light on the genomic history of mammoths. *Nature*. **2021**; 591: 265–269.

van Geel B, van Leeuwen JFN, Nooren K, Mol D, den Ouden N, van der Knaap PWO, Seersholm FV, Rey-Iglesia A, Lorenzen ED. Diet and environment of *Mylodon darwini* based on pollen of a Late-Glacial coprolite from the Mylodon Cave in southern Chile. *Review of Palaeobotany and Palynology*, **2022**; 296: 104549.

Vidya TNC, Sukumar R. Social and reproductive behaviour in elephants. *Current science*. **2005**; 89: 1200–1207.

Villa-Islas V, Izarraras-Gomez A, Larena M, Campos EMP, Sandoval-Velasco M, Rodríguez-Rodríguez JE, Bravo-Lopez M, Moguel B, Fregel R, Tretmanis JM, Velázquez-Ramírez DA, Herrera-Muñoz A, Sandoval K, Nieves-Colón MA, Zepeda G, Villanea FA, Medina EFV, Aguayo-Haro R, Valdiosera C, Ávila-Arcos MC. Demographic history and genetic structure in pre-Hispanic Central Mexico. *BiorXiv*. **2022**; 10: 496730.

Wadsworth C, Procopio N, Anderung C, Carretero JM, Irirarte E, Valdiosera C, Elburg R, Penkman K, Buckley M. Comparing ancient DNA survival and proteome content in 69 archaeological cattle tooth and bone samples from multiple European sites. *Journal of Proteomics*. **2017**; 158: 1-8.

Wales N, Andersen K, Cappellini E, Avila-Arcos MC, Gilbert MTP. Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *PloS One*. **2014**; 9: e86827.

Wales N, Kistler L. Extraction of Ancient DNA from Plant Remains. *Methods in Molecular Biology*. **2019**; 1963: 45–55.

Wang MS, Murray GGR, Mann D, Groves P, Vershinina AO, Supple MA, Kapp JD, Corbett-Detig R, Crump SE, Stirling I, Laidre KL, Kunz M, Dalén L, Green RE, Shapiro B. A polar bear paleogenome reveals extensive ancient gene flow from polar bears into brown bears. *Nature Ecology y Evolution*, **2022**; 6: 936–944.

Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Nueva York: Springer

International Publishing. **2016**.

Yang DY, Eng B, Waye JS, Dudar JC, Saunders SR. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *American Journal of Physical Anthropology*. **1998**; 105: 539–543.

ANEXOS

Anexo 1. Tablas utilizadas para construir las gráficas de *M. columbi*

Los resultados experimentales de este trabajo, es decir, extracciones, construcción de librerías e indexado de las muestras se encuentra en la siguiente liga. Debido al tamaño del documento, dicha base de datos se encuentra en el siguiente enlace para su mejor visualización.

<https://drive.google.com/drive/folders/1jA9Xc1baEramLkeGZXe-oKtiDTqqqKxx?usp=sharing>

A pesar de que la tabla maestra fue la que se utilizó para construir las gráficas de caja y diagramas de dispersión, continuación se presentan una por una para su mejor visualización.

Cuadro A1.1. Relación entre contenido endógeno y tipo de material (Figura 4).

MPA: molar en posición anatómica.

ID Muestra	Tipo de Material	Porcentaje de contenido endógeno (%)
McM1_2139	Molar	0.0
McM1_21920	Molar	0.1
McJ1_23463	MPA	0.4
McJ2_5917	MPA	0.1
McJ2_21602	MPA	0.1
McJ2_3279	MPA	0.0
McJ2_2021	MPA	0.1
McJ2_14852	MPA	0.0
McM2_24983	Molar	0.1
McJ2_42114	MPA	0.6
McM2_22683	Molar	0.2
McM2_Molar_negro	Molar	0.0
McM2_23171	Molar	0.6
McM2_19147	Molar	3.5
McJ2_1639	MPA	0.2
McJ2_9524	MPA	2.2
McJ2_16695	MPA	0.0
McJ2_20343	MPA	0.1
McM2_2140	Molar	0.0
McJ2_18728	MPA	0.0
McJ2_3158	MPA	0.1

Cuadro A1.2. Relación entre contenido endógeno y sexo molecular (Figura 5).

ID Muestra	Sexo	Porcentaje de contenido endógeno (%)
McM1_2139	NA	0.0
McM1_21920	Macho	0.1
McJ1_23463	Hembra	0.4
McJ2_5917	Macho	0.1
McJ2_21602	Macho	0.1
McJ2_3279	NA	0.0
McJ2_2021	Macho	0.1
McJ2_14852	Macho	0.0
McM2_24983	Hembra	0.1
McJ2_42114	Macho	0.6
McM2_22683	Hembra	0.2
McM2_Molar_negro	Macho	0.0
McM2_23171	Hembra	0.6
McM2_19147	Hembra	3.5
McJ2_1639	Macho	0.2
McJ2_9524	Hembra	2.2
McJ2_16695	NA	0.0
McJ2_20343	Hembra	0.1
McM2_2140	Hembra	0.0
McJ2_18728	Hembra	0.0
McJ2_3158	Hembra	0.1

Cuadro A1.3. Relación entre contenido endógeno y profundidad mínima de excavación (Figura 6).

ID Muestra	Profundidad mínima de excavación (m)	Porcentaje de contenido endógeno (%)
McM1_2139	NA	0.0
McM1_21920	1.45	0.1
McJ1_23463	1.13	0.4
McJ2_5917	1.25	0.1
McJ2_21602	3.22	0.1
McJ2_3279	1.83	0.0
McJ2_2021	2.04	0.1
McJ2_14852	3.62	0.0
McM2_24983	1.7	0.1
McJ2_42114	1.38	0.6
McM2_22683	NA	0.2
McM2_Molar_negro	NA	0.0
McM2_23171	0.76	0.6
McM2_19147	1.41	3.5
McJ2_1639	NA	0.2
McJ2_9524	1.07	2.2
McJ2_16695	1.5	0.0
McJ2_20343	2.14	0.1
McM2_2140	2.21	0.0
McJ2_18728	2.68	0.0
McJ2_3158		0.1

Cuadro A1.4. Relación entre contenido endógeno y profundidad máxima de excavación (Figura 7).

ID Muestra	Profundidad máxima de excavación (m)	Porcentaje de contenido endógeno (%)
McM1_2139	NA	0.0
McM1_21920	1.53	0.1
McJ1_23463	1.79	0.4
McJ2_5917	1.65	0.1
McJ2_21602	3.53	0.1
McJ2_3279	1.85	0.0
McJ2_2021	2.1	0.1
McJ2_14852	3.82	0.0
McM2_24983	1.74	0.1
McJ2_42114	1.57	0.6
McM2_22683	NA	0.2
McM2_Molar_negro	NA	0.0
McM2_23171	0.84	0.6
McM2_19147	1.46	3.5
McJ2_1639	3.3	0.2
McJ2_9524	NA	2.2
McJ2_16695	1.68	0.0
McJ2_20343	2.62	0.1
McM2_2140	2.24	0.0
McJ2_18728	3.02	0.0
McJ2_3158	2.7	0.1

Cuadro A1.5. Relación entre contenido endógeno y cuantificación de ADN realizada en Qubit (Figura 8).

ID Muestra	Concentración de ADN (ng/μL)	Porcentaje de contenido endógeno (%)
McM1_2139	3.64	0.0
McM1_21920	3.9	0.1
McJ1_23463	7.64	0.4
McJ2_5917	3.42	0.1
McJ2_21602	5.44	0.1
McJ2_3279	5.3	0.0
McJ2_2021	4.52	0.1
McJ2_14852	7.96	0.0
McM2_24983	3	0.1
McJ2_42114	39.4	0.6
McM2_22683	1.61	0.2
McM2_Molar_negro	5.86	0.0
McM2_23171	1.09	0.6
McM2_19147	1.01	3.5
McJ2_1639	24.8	0.2
McJ2_9524	4.16	2.2
McJ2_16695	39.6	0.0
McJ2_20343	4.26	0.1
McM2_2140	1.25	0.0
McJ2_18728	4.9	0.0
McJ2_3158	8.84	0.1

Cuadro A1.6. Relación entre contenido endógeno y cuantificación de ADN realizada en BioAnalyzer.

ID Muestra	Concentración de ADN (ng/μL)	Porcentaje de contenido endógeno (%)
McM1_2139	4.50	0.0
McM1_21920	4.93	0.1
McJ1_23463	7.22	0.4
McJ2_5917	10.43	0.1
McJ2_21602	3.30	0.1
McJ2_3279	7.60	0.0
McJ2_2021	4.65	0.1
McJ2_14852	8.19	0.0
McM2_24983	5.16	0.1
McJ2_42114	3.37	0.6
McM2_22683	8.07	0.2
McM2_Molar_negro	3.70	0.0
McM2_23171	8.50	0.6
McM2_19147	3.30	3.5
McJ2_1639	8.18	0.2
McJ2_9524	6.13	2.2
McJ2_16695	8.21	0.0
McJ2_20343	4.83	0.1
McM2_2140	6.30	0.0
McJ2_18728	4.87	0.0
McJ2_3158	6.88	0.1

Anexo 2. Scripts de procesamiento bioinformático

Script para mapeo contra genoma de referencia.

```
### Written by Maria (maricugh@gmail.com)
```

```
###
```

```
### New modification August 2013 nit now uses mapDamage2 and rescales
```

```
### May 2014
```

```
### This script is based on Clean-map-call.V1.1.sh modified by Maria for
```

```
### the analysis of Cony's samples at Stanford
```

```
### May 2015 further modified to include trimming/merging of pairs with adapter removal at the beginning
```

```
###Adapted to LIIGH cluster
```

```
#Your job name
```

```
#$-N Trim_map_script_v
```

```
#Use current working directory
```

```
#$-cwd
```

```
#Join stdout and stderr
```

```
#$-j y
```

```
# Run job through bash shell
```

```
#$-S /bin/bash
```

```
#Send an email after the job has finished
```

```
#$-m e
```

```
#$-M angeles.tvgz@gmail.com
```

```
#If modules are needed, source modules environment (Do not delete the next line):
```

```
./etc/profile.d/modules.sh
```

```
### Arguments are
```

```
### $1=sample
```

```
### $2=path to reference
```

```
#A report is sent to the email below
```

```
# change email to receive results there
```

```

EMAIL='angeles.tvgz@gmail.com'

####Preparing data: Cleaning and reporting stats

## $1 is the file that is taken as the first parameter of the program (e.g.in: ./program.sh file.txt ; file.txt is
saved in the $1 variable)

for file in *bz2

do

base=`basename $file .bz2`

## Run adapterRemoval

#module load adapterremoval/1.5.4

#AdapterRemoval --file1 <(bzcata $file1) --trimns --trimqualities --qualitybase 30 --minlength 30 --discarded
$base1.trim.discarded --settings $base1.trim.settings | bzip2 > $base1.txtcleaned.fq.bz2

#mail -s "$1 Trimming done" $EMAIL < $base1.trim.settings

### Mapping

#Change base

module load htlib/1.2.1

module load gcc/5.1.0

module load bwa/0.7.13

module load samtools/0.1.19

bwa aln -l 16 -n 0.01 -t 16 $1 <(bzcata $base.bz2) | bwa samse $1 -r '@RG\tID:$base\tSM:$base' - <(bzcata
$base.bz2) |samtools view -bSh - > $base.bwa.all.bam #to keep track of REAL endogenous content

echo "$base" >> $base\_REPORT.txt

#echo "Total sequences" >> $base\_REPORT.txt

samtools view -c $base.bwa.all.bam >> $base\_REPORT.txt

samtools view -bh -F4 -q 30 $base.bwa.all.bam > $base.bwa.bam

#Sort with samtools

samtools sort $base.bwa.bam $base.bwa.sort

#echo "Before collapse" >> $base\_REPORT.txt

samtools view -c $base.bwa.sort.bam >> $base\_REPORT.txt

#Rmdup

samtools rmdup -s $base.bwa.sort.bam $base.bwa.sort.rmdup.bam

```

```

#echo "After collapse" >> $base\_REPORT.txt

samtools view -c $base.bwa.sort.rmdup.bam >> $base\_REPORT.txt

#keep only uniquely mapped

samtools view -h $base.bwa.sort.rmdup.bam |grep -v 'XT:A:R'|grep -v 'XA:Z' |awk
'{if($0~/X1:i:0/||$0~/^@/)print $0}' | samtools view -bS - > $base.bwa.sort.rmdup.uniq.bam

#echo "After removing paralogs" >> $base\_REPORT.txt

samtools view -c $base.bwa.sort.rmdup.uniq.bam >> $base\_REPORT.txt

#Fastqc of mapped reads

module load fastqc/0.11.3

module load r/3.2.1

fastqc -f bam $base.bwa.sort.rmdup.uniq.bam y

samtools mpileup -f $1 -C 50 $base.bwa.sort.rmdup.uniq.bam | tee $base.bwa.sort.rmdup.uniq.pileup | awk
'{if($3!=""){sum+=$4}}END{print sum/NR "\n" NR}' >> $base\_REPORT.txt

#make index

#echo "Mapped reads average length" >> $base\_REPORT.txt

samtools view $base.bwa.sort.rmdup.uniq.bam |awk '{SUM+=length($10);DIV++}END{print SUM/DIV}' >>
$base\_REPORT.txt

mail -s "$base mapping finished" $EMAIL < $base\_REPORT.txt

module load mapdamage/2.0

##

mapDamage -i $base.bwa.sort.rmdup.uniq.bam -r $1

done

exit;

```

Script para sexamiento de los organismos.

```
module load samtools

#Bam files need to be indexed !!

#names list already exists in a .names file

name=$1;

ext=$2;

# Count reads in X and 8 chr

chrX=`samtools view -c -q 30 ${name}.${ext} Lx4chrX`

chr8=`samtools view -c -q 30 ${name}.${ext} Lx4chr8`

# Calculate proportion of reads to chr size --> Chr 8 is used because it's almost the size of chr X

chrXNorm=`echo ${chrX}/116625373 | bc -l | awk '{print "0"$0}'`

chr8Norm=`echo ${chr8}/125256028 | bc -l | awk '{print "0"$0}'`

# Obtain molecular sex. If the number calculated is near to 1, it's a female

#If number is near 0.5, it's a male --> they are haploid for X chr, and females diploid

#echo -e "$chrXNorm / $chr8Norm\n"

sex=`echo $chrXNorm/$chr8Norm | bc -l | awk '{print "0"$0}' | awk '{printf "%.3f", $0}'`

echo -e "${name}\t${sex}";
```

Script para estimar contaminación humana.

```
#!/bin/bash

module load samtools

name=$1;

ext=$2;

fqv1=`bzcat
/mnt/Cromosoma/mavila/atavares/Mammoths_SL/Mammoth_Analysis/Collapsed_sequences/${name}_L001
-4_R1-2_001.fastq.collapsed.bz2 | wc -l`;

FQV1=`echo ${fqv1}/4 | bc -l | awk '{printf "%0.0f", $0}'`;

cHum=`samtools view -F 4 -q 30
/mnt/Cromosoma/mavila/atavares/Mammoths_SL/Mammoth_Analysis/Loxv4_v2/${name}${ext} | awk '{if
(($0 !~ /^@/) yy ($3 ~ /HG/)) {print $0} }' | wc -l`;

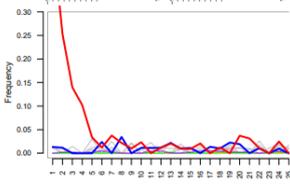
contH=`echo ${cHum}/${FQV1} | bc -l`;

contHF=`echo "${contH}*100" | bc -l | sed 's/^\./0./g'`;

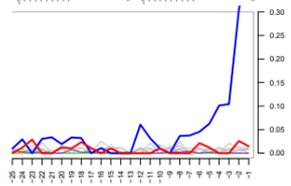
#echo -e "n: ${name}\tfq1: ${fqv1}\tfq2: ${FQV1}\tcont: ${cHum}\tcomt% ${contHF}";

echo -e "${name}\t${FQV1}\t${cHum}\t${contHF}";
```

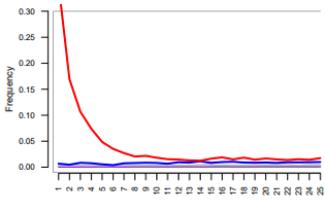
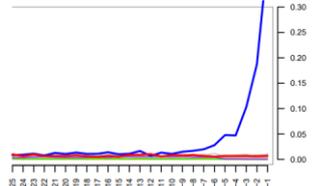
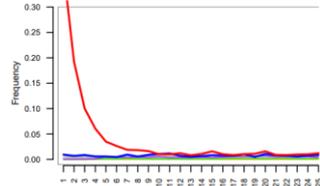
Anexo 3. Gráficos de patrones de daño de las muestras procesadas en este trabajo.



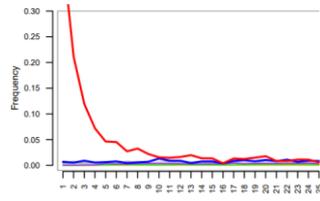
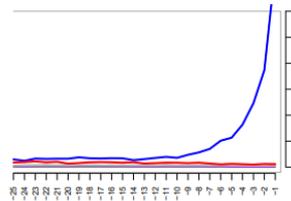
McM1_2139



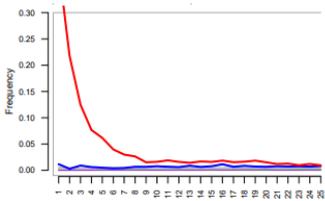
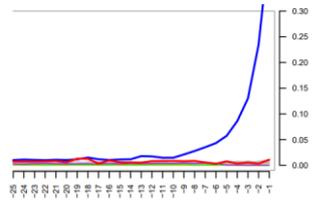
McM1_21920



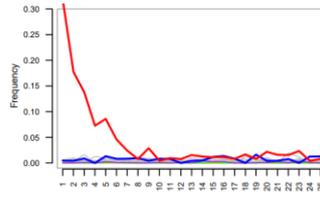
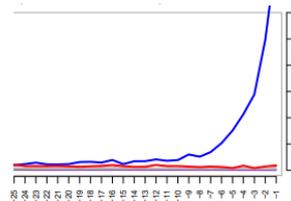
McJ1_23463



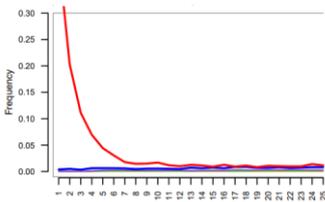
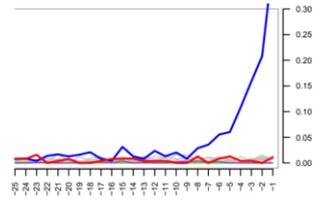
McJ2_5917



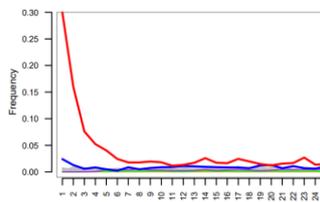
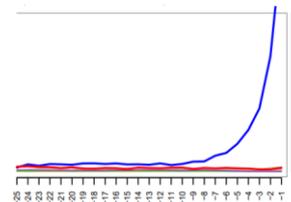
McJ2_21602



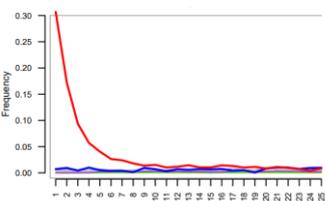
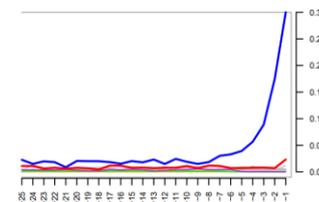
McJ2_3279



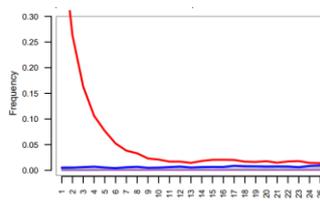
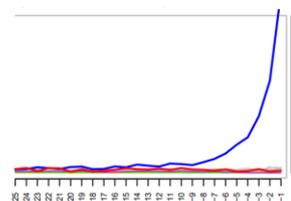
McJ2_2021



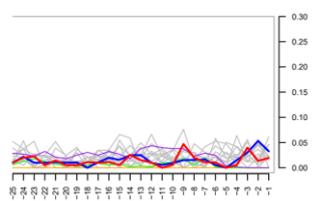
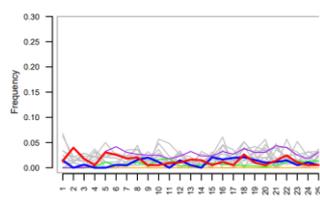
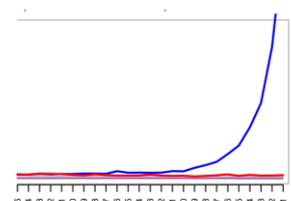
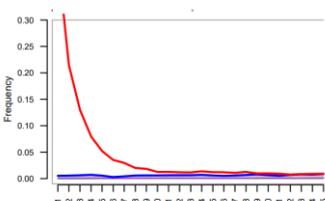
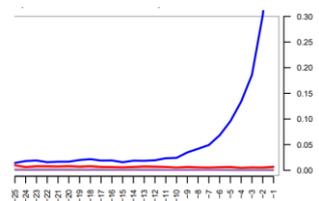
McJ2_14852



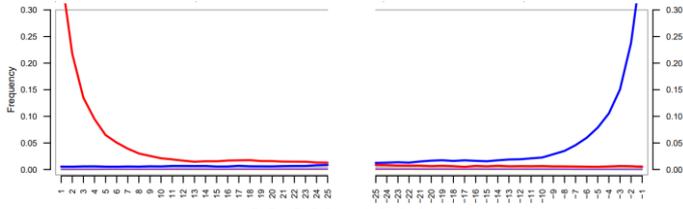
McM2_24983



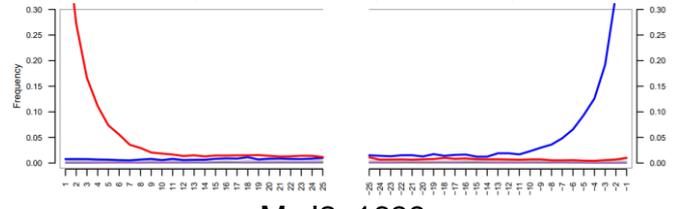
McJ2_42114



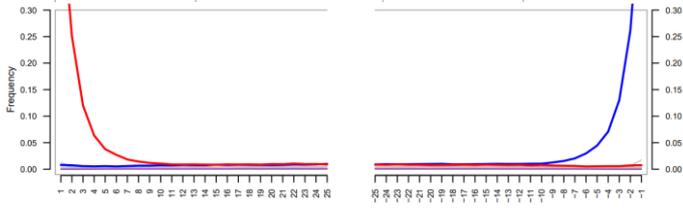
McM2_22683



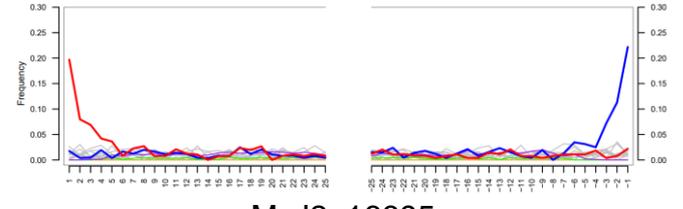
McM2_Molar_negro



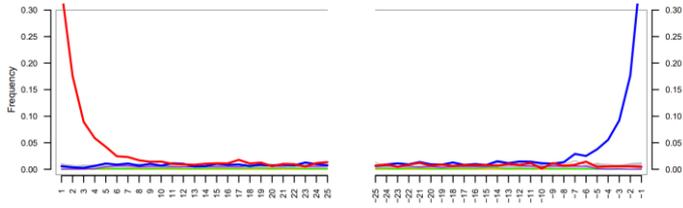
McM2_19147



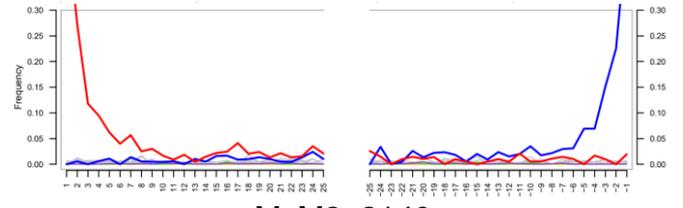
McJ2_1639



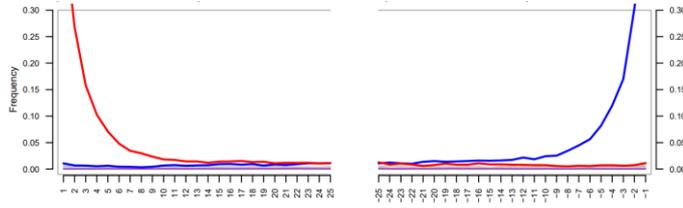
McJ2_9524



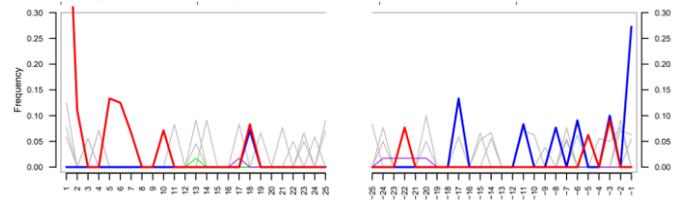
McJ2_16695



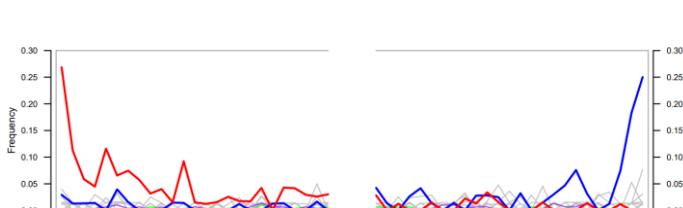
McJ2_20343



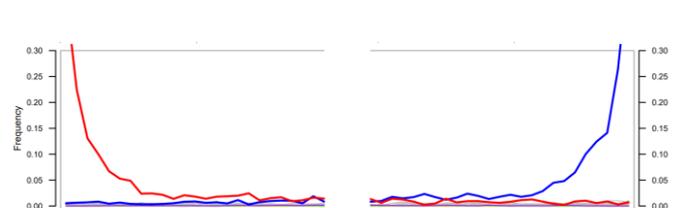
McM2_2140



McM2_23171



McCN1 (control negativo)



18728



3158



Anexo 4. Script para la obtención de gráficos en RStudio.

```
#Creado por Ángeles Tavares como parte del proyecto de tesis

#Análisis comparativo de la cantidad de ADN endógeno de molares de mamuts (Mammuthus columbi)

#excavados en Santa Lucía, Estado de México

#angeles.tvgz@gmail.com

#Crear dataframe GENERAL

Extracciones.df <- read.csv("Registro_Tesis_R.csv", sep = ",", header = TRUE)

#Crear dataframe GENERAL sin outliers

Extracciones.df.no <- read.csv("Registro_Tesis_R_NO.csv", sep = ",", header = TRUE)

#Cargar paquetes (deben estar previamente instalados utilizando el comando install.packages())

library(ggplot2)

library(wesanderson)

library(ggsignif)

library(ggpubr)

#Crear boxplots

#Boxplot de Tipo de material y porcentaje de contenido endógeno

ggplot(data = Extracciones.df, aes(x = Material, y = Endo_cont1, color = Material, na.rm = TRUE, fill =
Material)) +

  geom_boxplot(na.rm = TRUE) +

  scale_color_manual(values = wes_palette("Darjeeling1")) +

  geom_signif(comparisons = list(c("MPA", "Molar")), map_signif_level=TRUE)+

  labs(y = "Porcentaje de contenido endógeno") +

  ggtitle("Contenido endógeno por tipo de material")
```

```

#Boxplot de Sexo y porcentaje de contenido endógeno

ggplot(data = Extracciones.df, aes(x = Sexo, y = Endo_cont1, color = Sexo, na.rm = TRUE, fill = Sexo)) +

  geom_boxplot(na.rm = TRUE) +

  scale_color_manual(values = wes_palette("Darjeeling1")) +

  geom_signif(comparisons = list(c("Macho", "Hembra")), map_signif_level=TRUE)+

  labs(y = "Porcentaje de contenido endógeno") +

  ggtitle("Contenido endógeno por Sexo")

#SCATTERPLOTS

#scatterplot del contenido endógeno de ADN con relación a la profundidad mínima de excavación

ggplot(data = Extracciones.df, aes(x = Min_Depth, y = Endo_cont1, na.rm = TRUE)) +

  geom_point(na.rm = TRUE) +

  scale_color_manual(values = wes_palette("Darjeeling1")) +

  labs(y = "Porcentaje de contenido endógeno", x = "Profundidad mínima de excavación (m)") +

  ggtitle("% de contenido endógeno vs profundidad mínima de excavación")

#scatterplot del contenido endógeno de ADN con relación a la profundidad máxima de excavación

ggplot(data = Extracciones.df, aes(x = Max_Depth, y = Endo_cont1, na.rm = TRUE)) +

  geom_point(na.rm = TRUE) +

  scale_color_manual(values = wes_palette("Darjeeling1")) +

  labs(y = "Porcentaje de contenido endógeno", x = "Profundidad máxima de excavación (m)") +

  ggtitle("% de contenido endógeno vs profundidad máxima de excavación")

#scatterplot del contenido endógeno de ADN con relación a cuantificación en BioAnalyzer

ggplot(data = Extracciones.df, aes(x = Conc_BioAnalyzer_nguL, y = Endo_cont1, na.rm = TRUE)) +

  geom_point(na.rm = TRUE) +

  scale_color_manual(values = wes_palette("Darjeeling1")) +

```

```

labs(y = "Porcentaje de contenido endógeno", x = "cuantificación en BioAnalyzer (ng/μL)") +

ggtitle("% de contenido endógeno vs ng/μL (BA)")

#scatterplot del contenido endógeno de ADN con relación a cuantificación en Qubit tras extracción

ggplot(data = Extracciones.df, aes(x = Conc_1.2_nguL, y = Endo_cont1, na.rm = TRUE)) +

geom_point(na.rm = TRUE) +

scale_color_manual(values = wes_palette("Darjeeling1")) +

labs(y = "Porcentaje de contenido endógeno", x = "cuantificación en Qubit (ng/μL)") +

ggtitle("% de contenido endógeno vs ng/μL (Qubit)")

#Test de u de Mann-Whitney para observar si hay significancia en los boxplots

#de contenido endógeno por sexo y por tipo de material

#Crear vectores para tipo de Material

Molar = c(0.0031, 0.0421, 0.0635, 0.1123, 0.0019, 0.3705, 2.2187, 0.0079)

MPA = c(0.2335, 0.0412, 0.0429, 0.0037, 0.0930, 0.0209, 0.2125, 0.1380, 1.8499, 0.0050, 0.04596, 0.0013,
0.0620)

#Crear vectores para Sexo

Macho = c(0.4213, 0.4124, 0.4298, 0.9303, 0.209, 0.2125, 0.0019, 0.1380)

Hembra = c(0.2333, 0.0635, 0.1123, 0.3705, 2.2187, 1.8499, 0.4596, 0.0079, 0.0013, 0.0620)

#Realizar prueba para tipo de Material

wilcox.test(Molar, MPA)

#Realizar prueba para Sexo

wilcox.test(Macho, Hembra)

#ANÁLISIS DE RESIDUALES

# Dataframe de residuales

```

```

Residuales.df <- read.csv("Registro_Tesis_R_Residuales.csv", sep = ",", header = TRUE)

#Gráfico de residuales de profundidad mínima

ggplot(data = Residuales.df, aes(x = ID.Sample, y = residual_mind, na.rm = TRUE)) +

  geom_point(na.rm = TRUE) +

  labs(y = "Residuales", x = "Muestras") +

  ggtitle("Gráfico de residuales de relación de % de contenido endógeno

    y profundidad mínima") +

  guides(x = guide_axis(angle = 90)) +

  geom_hline(yintercept = 0)

#Gráfico de residuales de profundidad máxima

ggplot(data = Residuales.df, aes(x = ID.Sample, y = residual_maxd, na.rm = TRUE)) +

  geom_point(na.rm = TRUE) +

  labs(y = "Residuales", x = "Muestras") +

  ggtitle("Gráfico de residuales de relación de % de contenido endógeno

    y profundidad máxima") +

  guides(x = guide_axis(angle = 90))+

  geom_hline(yintercept = 0)

#Gráfico de residuales de cuantificación en Qubit

ggplot(data = Residuales.df, aes(x = ID.Sample, y = residual_Qubit, na.rm = TRUE)) +

  geom_point(na.rm = TRUE) +

  labs(y = "Residuales", x = "Muestras") +

  ggtitle("Gráfico de residuales de relación de % de contenido endógeno

    y cuantificación en Qubit") +

  guides(x = guide_axis(angle = 90)) +

```

```
geom_hline(yintercept = 0)

#Gráfico de residuales de cuantificación en BioAnalyzer

ggplot(data = Residuales.df, aes(x = ID.Sample, y = residual_BA, na.rm = TRUE)) +

geom_point(na.rm = TRUE) +

labs(y = "Residuales", x = "Muestras") +

ggtitle("Gráfico de residuales de relación de % de contenido endógeno

        y cuantificación en BioAnalyzer") +

guides(x = guide_axis(angle = 90)) +

geom_hline(yintercept = 0)
```