UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE ECOLOGÍA

BIOLOGÍA EXPERIMENTAL

Estudio paleogenómico de patógenos humanos en población prehispánica y colonial de México

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

MIRIAM JETZABEL BRAVO LÓPEZ

TUTORA PRINCIPAL DE TESIS: DRA. MARÍA DEL CARMEN ÁVILA ARCOS
LIIGH-UNAM

COMITÉ TUTOR: DR. DAVID ROMERO CAMARENA
CENTRO DE CIENCIAS GENÓMICAS, UNAM
DR. DANIEL PIÑERO DALMAU
INSTITUTO DE ECOLOGÍA, UNAM

CD. MX., ABRIL 2021

Universidad Nacional Autónoma de México
UNAM
Biblioteca Central
Dirección General de Bibliotecas de la UNAM

UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

INSTITUTO DE ECOLOGÍA

BIOLOGÍA EXPERIMENTAL

Estudio paleogenómico de patógenos humanos en población prehispánica y colonial de México

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

MIRIAM JETZABEL BRAVO LÓPEZ

TUTORA PRINCIPAL DE TESIS: DRA. MARÍA DEL CARMEN ÁVILA ARCOS
LIIGH-UNAM

COMITÉ TUTOR: DR. DAVID ROMERO CAMARENA
CENTRO DE CIENCIAS GENÓMICAS, UNAM
DR. DANIEL PIÑERO DALMAU
INSTITUTO DE ECOLOGÍA, UNAM

MÉXICO, CD. MX., ABRIL 2021

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS
ENTIDAD INSTITUTO DE ECOLOGÍA
OFICIO CPCB/260/2021
ASUNTO: Oficio de Jurado

M. en C. Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM
P r e s e n t e

Me permito informar a usted que en la reunión ordinaria del Subcomité de Subcomité de Biología Experimental y Biomedicina del Posgrado en Ciencias Biológicas, celebrada el día **15 de febrero de 2021** se aprobó el siguiente jurado para el examen de grado de **DOCTORA EN CIENCIAS** de la estudiante **BRAVO LÓPEZ MIRIAM JETZABEL** con número de cuenta **514012410** con la tesis titulada: **"ESTUDIO PALEOGENÓMICO DE PATÓGENOS HUMANOS EN POBLACIÓN PREHISPÁNICA Y COLONIAL DE MÉXICO",** realizada bajo la dirección de la **DRA. MARÍA DEL CARMEN ÁVILA ARCOS,** quedando integrado de la siguiente manera:

Presidente:	DRA.	ROSA IRENE FREGEL LORENZO
Vocal:	DR.	FEDERICO SÁNCHEZ QUINTO
Vocal:	DR.	LUIS DAVID ALCARAZ PERAZA
Vocal:	DRA.	MARGARITA MEZA MANZANILLA
Secretario:	DR.	DAVID ROMERO CAMARENA

Sin otro particular, me es grato enviarle un cordial saludo.

A T E N T A M E N T E
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Cd. Universitaria, Cd. Mx., a 26 de marzo de 2021

COORDINADOR DEL PROGRAMA

POSGRADO EN CIENCIAS BIOLOGICAS
COORDINACIÓN

DR. ADOLFO GERARDO NAVARRO SIGÜENZA

AGRADECIMIENTOS ACÁDEMICOS

Al Posgrado en Ciencias Biológicas UNAM, que me permitió formarme como investigadora.

A CONACYT, por la beca proporcionada que me permitió dedicarme tiempo completo al Doctorado.

A mi tutora principal, la Dra. María Ávila Arcos, por permitirme continuar con mis estudios de Doctorado en su laboratorio, por sus enseñanzas y paciencia, y por ser parte imprescindible en el proceso más enriquecedor de mi vida.

A las instituciones que financiaron este proyecto de investigación; Organización *Wellcome Trust*, con número de proyecto 208934/Z/17/Z; Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), con número de proyecto IA201219, y al programa *The International Human Frontier Science Program Organization (HFSP)*, con número de proyecto RGY0075/2019.

A mi Comité Tutor, al Dr. David Romero Camarena y al Dr. David Piñero Dalmau, por su apoyo en todo momento durante la realización de mi Doctorado, por proporcionarme valiosos comentarios y sugerencias que permitieron enriquecer este proyecto de investigación.

Al Consejo de Arqueología, a la Arqlga. Elizabeth Mejía, a el Arqlgo. Alberto Herrera, Arqlgo. Alejandro Meraz, Arqlga. María de la Luz Moreno Cabrera, Dr. Jorge Gómez Valdés, Antr. Joel Olvera, Arqlga. Julia Perez Perez, y a la Arqlga. Gabriela Zepeda, por no sólo permitir llevar a cabo el análisis de aDNA de parte de nuestro patriminonio cultural, si no por enriquecer el contenido de los capítulos de constituyen esta tesis.

Al Dr. Rafael Palacios, por crear un espacio (LIIGH) que me permitierá crecer tanto personal como acádemicamente.

A los honorables miembros del jurado; Dra. Rosa Fregel Lorenzo, Dr. David Romero Camarena, Dr. Federico Sánchez Quinto, Dr. Luis David Alcaraz Peraza, y a la Dra. Margarita Meza Manzanilla, por sus valiosos comentarios y observaciones en esta tesis.

Al Dr. Diego Ortega del Vecchyo, a la Dra. Daniela Robles Espinoza, a la Dra. Alejandra Medina Rivera, y a la Dra. Lucia Morales Reyes, por contribuir en mi formación acádemica y por su apoyo en mi estancia en el LIIGH.

A la Dra. María Nieves Colón, por sus valiosas enseñanzas durante mis estudios de Doctorado.

A la Dra. Åshild Vågene por compartir sus conocimientos en el área de genómica de patógenos antiguos.

A Alejandra Castillo y a Carina Uribe, por su disposición incondicional y ayuda durante la realización de este proyecto de investigación.

Al Laboratorio Nacional de Visualización Científica Avanzada (LAVIS), en particular a Luis Alberto Aguilar Bautista, y a Jair García Sotelo, por sus asesorías y disposición en todo momento durante la realización de esta tesis.

A Erika Rodríguez, por su apoyo en la parte administrativa de mis estudios de Doctorado en Ciencias Biológicas.

Al Dr. Simon Rasmussen y a su grupo de investigación, por recibirme de la manera más amable en su laboratorio y por las observaciones que enriquecieron este proyecto de investigación

AGRADECIMIENTOS A TÍTULO PERSONAL

A los pilares de mi vida y las personas que más admiro, a mis padres Luz López Manzo y a Gustavo Bravo Mendoza, y a mi hermana Eugenia Bravo López, esta vida no me alcanzara para agredecerles todo lo que han hecho por mi.

A mis grandes amigas y *roomies*, Ana B y Karen, por su invaluable amistad, por enseñarme tanto y convertirse en mi soporte durante la realización de esta tesis.

A Ernesto, por tu amistad, apoyo incondicional y amor indefinido, me has ayudado a crecer tanto profesional como personalmente. Tantos años enfrente de mí y sin poder notarlo.

A Dircé, por tu valiosa amistad y apoyo desinteresado, gracias por estar presente a pesar de la distancia.

A mis compañeras (os) del laboratorio, a Viri, Barbara, Alán, y Angie, y a los que alguna vez formaron parte del grupo, a Camila, Che, y Axel, por su amistad y apoyo en todo momento.

A mis amigos del LIIGH, a Isaac y María Trejo, por su amistad y ayuda bionformática.

A mis amigos del INMEGEN, a Ricardo, Elvira, y Omar, por su amistad y por hacer llevadera mi estancia en ese lugar. Adicionalmente, al Dr. Hugo y a Dafné, por mostrarme que nada estaba perdido. No sé que hubiera hecho sin todos ustedes en esa etapa de mi vida.

Y por último, a los seres más valiosos en mi vida, a Paco, Goliath, Fanilu, Chiquis y Panchita, por su cariño, el más sincero que podría tener, jamás los olvidaré.

ÍNDICE

ABREVIATURAS

aDNA	Ácido desoxirribonucleico antiguo
a.C.	Antes de Cristo
AP	Antes del presente
d.C.	Después de Cristo
DNA	Ácido desoxirribonucleico
NGS	Secuenciación de siguiente generación (*Next Generation Sequencing)*
PCR	Reacción en cadena de la polimerasa (*Polymerase Chain Reaction*)
pb	Pares de bases
qPCR	Reacción en cadena de la polimerasa cuantitativa (Quantitative *Polymerase Chain Reaction)*
RNA	Ácido ribonucleico
SNP	Polimorfismos de un solo nucleótido (*Single Nucleotide Polymorphism*)

RESUMEN

El análisis genómico de agentes causantes de enfermedades infecciosas en poblaciones humanas del pasado ha contribuido al entendimiento de su estado de salud. En México existe abundante evidencia arqueológica e histórica de la existencia de enfermedades infecciosas en los periodos prehispánico y colonial. La aplicación de tecnologías de siguiente generación ha permitido recuperar genomas completos de patógenos a partir de comunidades microbianas complejas que se encuentran en restos arqueológicos humanos. En esta tesis doctoral— que consiste en un artículo de revisión (capítulo 4), un artículo de investigación (capítulo 5) y dos reportes de proyectos de investigación (capítulos 6 y 7)— muestra el alcance de esta estrategia en la identificación de patógenos presentes en individuos del periodo prehispánico y colonial en México, así como en cazadores-recolectores de la Costa Norte de la Patagonia Argentina. En el artículo de revisión (capítulo 4) se sintetiza el panorama general de la genómica de patógenos antiguos, incluyendo las estrategias metodológicas utilizadas y los desafíos presentes en esta subrama de la genómica. En el artículo de investigación (capítulo 5) se presenta la reconstrucción de 12 genomas de *Tannerella forsythia* en muestras de cálculo dental y dientes de individuos del periodo prehispánico y colonial de México. *T. forsythia* es una bacteria asociada con el desarrollo de enfermedad periodontal que, de acuerdo con los resultados generados en esta tesis, acompañó a los primeros pobladores de América. Nuestros resultados también muestran que tras la llegada de Europeos y Africanos durante la conquista española, ocurrió un posible reemplazo de la cepa nativa por la cepa importada. Durante el periodo colonial, hubo un importante incremento de la presencia de enfermedades infecciosas en relación al periodo prehispánico; en particular en regiones de la Nueva España donde había hacinamiento. En capítulo 6, se describe la reconstrucción del genoma de *Salmonella enterica Paratyphi C* a partir del análisis metagenómico de dientes de individuos del periodo colonial del Templo de la Inmaculada Concepción, localizado en Ciudad de México. Al evaluar la relación filogenética del genoma reconstruido con otros genomas antiguos y modernos reportados en la bibliografía, se observó una cercanía filogenética con cepas antiguas provenientes del sur de México.

Las herramientas analitícas empleadas en esta tesis no sólo permitieron identificar aDNA de patógenos a partir de restos arqueológicos de individuos de México, sino también de otras regiones de América. Como parte de las colaboraciones establecidas durante el doctorado, en el capítulo 7, se describe el análisis del metagenoma de dientes y huesos de cazadores-recolectores de la Patagonia Argentina. En ellos se identificó aDNA de patógenos zoonóticos como *Erysipelothrix rhusiopathiae*, así como patógenos asociados a enfermedad periodontal, entre ellos *Rothia dentocariosa*, *Streptococcus sanguinis*, *Tannerella forsythia*. Por lo que este análisis aporta conocimiento sobre el estado de salud en una población antigua poco estudiada desde la paleogenómica.

ABSTRACT

The genomic analysis of the causative agents of infectious diseases that human populations have faced in the past, have contributed to the understanding of their health status through time. As regards Mexico, there is abundant archeological and historic evidence of the presence of infectious diseases during Pre-Hispanic and Colonial period. Next generation sequencing technologies have allowed pathogen genome reconstruction from complex microbial communities commonly found in archeological remains. This dissertation—consisting of a review paper (chapter 4), one research paper (chapter 5) and two project reports (chapters 6 and 7)— shows the scope of these strategies to pathogen aDNA identification from Pre-hispanic and Colonial individuals from Mexico, also out-of hunter-gatherers from Central Coastal Argentinian Patagonian. The review paper (chapter 4), synthesizes the whole picture of ancient pathogen genomics, including the description of the main methodological strategies and the challenges found in this subfield. The research paper (chapter 5) presents the reconstruction of 12 *Tannerella forsythia* genomes from dental calculus and teeth samples from Pre-Hispanic and Colonial individuals from Mexico. *T. forsythia* is associated with the development of periodontal disease, which according to the results obtained in this dissertation, this bacterium arrived with the first human migrations to the Americas. Our results also show that new strains were introduced with the arrival of European and African populations during the Spanish colonization, and a possible replacement occurred of the native strain by the imported one. During the Colonial period, the increase of infectious diseases was evident compared to the Pre-Hispanic period, in particular, in the Viceroyalty of New Spain overcrowded areas. As for the chapter 6, we reconstructed a *Salmonella enterica Paratyphi C* genome from a teeth sample, excavated from the Temple of the Immaculate Conception 'La Conchita', located in Mexico City. When evaluating its phylogenetic relationship with ancient genomes publicly reported, we observed a phylogenetic closeness between south Mexico strains.

The analytical strategies employed in this dissertation not only allowed the identification of pathogen aDNA from archaeological remains located in Mexico, but also from other America regions. As part of the collaborations established during my doctoral studies, we analyzed the metagenomes of hunter-gatherer's teeth and bones samples from Central Coastal Argentinian Patagonian. In the chapter 7, we described the identification of aDNA pathogenic taxa *Erysipelothrix rhusiopathiae,* and those associated to periodontal disease, like *Rothia dentocariosa*, *Streptococcus sanguinis*, *Tannerella forsythia,* among others. This analysis contributes to the understanding of the health status of hunter-gatherers, which remains little explored through paleogenomics.

1. INTRODUCCIÓN

1.1 Principales epidemias en la historia humana

Los humanos han coexistido y co-evolucionado con microorganismos patógenos. Éstos han sido una de las principales fuerzas selectivas sobre la evolución humana (Haldane, 1932; Fumagalli *et al.*, 2011; Siddle y Quintana-Murci, 2014). El impacto potencial que los patógenos pueden tener sobre nuestra especie queda perfectamente ejemplificado con la actual pandemia ocasionada por el virus SARS-CoV-2 (por sus en inglés, severe acute respiratory syndrome coronavirus 2), la cual tiene lugar al momento de la escritura de esta tesis. Sin embargo, numerosas epidemias nos han afectado a lo largo de nuestra historia poblacional, dejando evidencia en nuestros genomas (Siddle y Quintana-Murci, 2014).

Hace aproximadamente 60,000-100,000 años, los humanos anatómicamente modernos se dispersaron dentro de África y hacia otros continentes. Esto propició una exposición a diversas condiciones ambientales y por lo tanto a una variedad de patógenos (Wolfe, Dunavan y Diamond, 2007; Cagliani y Sironi, 2013). Sin embargo, hace ~12,000 años en América, Asia y Europa, la transición epidemiológica y demográfica provocó una transformación en el estilo de vida de caza-recolección a uno basado en la agricultura y ganadería (Stone, 2020). La presencia de enfermedades infecciosas se vió potencializada por las condiciones de vida insalubres, asociadas al sedentarismo y el contacto con animales domesticados, y a su vez, esto propició un incremento en la tasa de mortalidad (Armelagos, Goodman y Jacobs, 1991; Gignoux, Henn y Mountain, 2011). Se propone que los agentes causantes de estas enfermedades infecciosas en este periodo eran principalmente de origen zoonótico como sarampión y tosferina, transmitidos por ganado y cerdos, respectivamente (Suzuki y Nei, 2002; Matthijnssens *et al.*, 2008; Furuse, Suzuki y Oshitani, 2010; Stone, 2020). Hoy en día se estima que el 61% de los patógenos humanos conocidos hasta el momento son zoonóticos (Woolhouse, 2002).

La relación continua hospedero-patógeno, ha propiciado una competencia y contra-adaptación en ambos a lo largo del tiempo (Siddle y Quintana-Murci, 2014). El impacto, en términos evolutivos de los patógenos sobre el genoma del hospedero, depende de la duración y virulencia de las infecciones. Al ocurrir un evento demográfico drástico en una población humana, como es la presencia de un brote epidémico de alta letalidad, el tamaño poblacional puede reducirse considerablemente, lo que limita la variación genética y aumenta la fuerza de la deriva génica en una población (Figura 1) (Domínguez-Andrés y Netea, 2019).

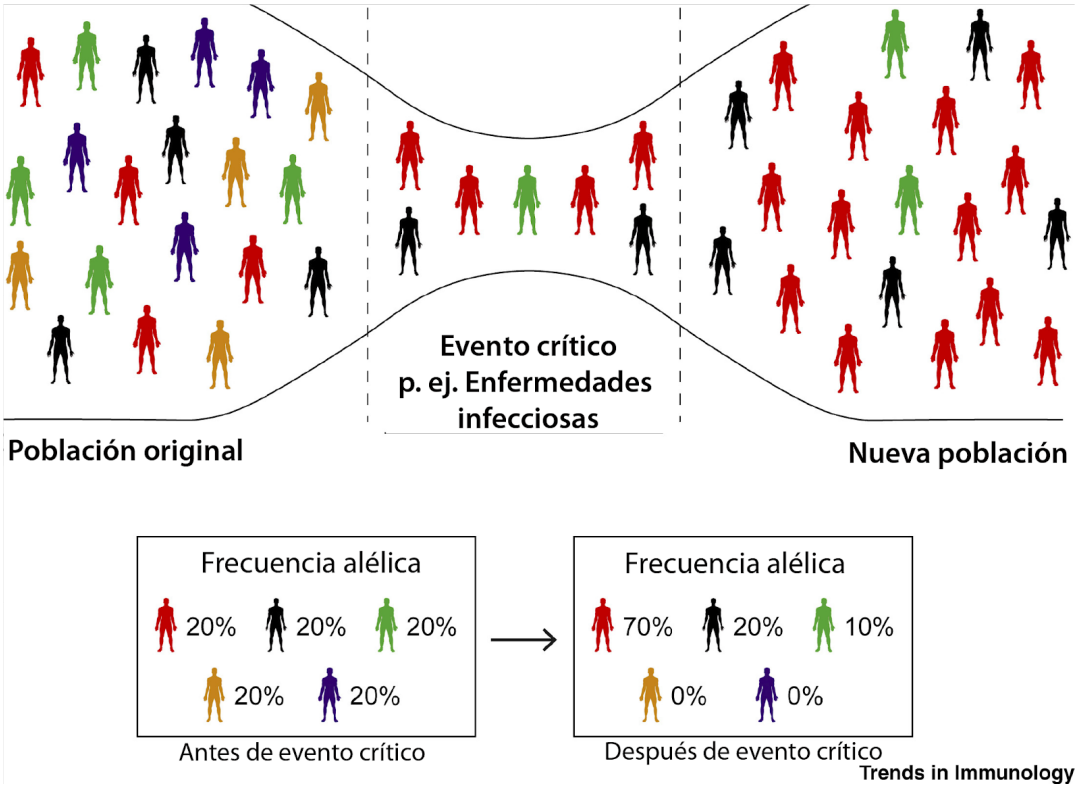


Figura 1. Cuello de botella en poblaciones humanas. Modificado de (Domínguez-Andrés y Netea, 2019). Al ocurrir un evento catastrófico sobre una población humana, como un brote epidémico, disminuye el número de individuos y por lo tanto su diversidad genética. El decremento de la diversidad genética es temporal, los efectos sobre la variación genética de la descendecia puede ser duraderos.

De acuerdo con la evidencia documental, numerosos eventos epidémicos de relevancia histórica se propagaron a través del comercio y rutas militares (Cossar, 1994). La peste bubónica es uno de ellos, ocasionada por la bacteria *Yersinia pestis,* la cual surgió en China hace más de 2600 años, y se extendió en múltiples radiaciones a Europa y al norte de África, a través de la Ruta de Seda (Gage y Kosoy, 2005; Sherman, 2020). También se propone que en 1495, la sífilis, ocasionada por el patógeno bacteriano *Treponema pallidum*, se extendió a lo largo de Europa debido a la invasión de Nápoles por parte de las tropas del rey Carlos VIII de Francia, en las que se hallaban mercenarios de Suiza, Italia y España (Meyer *et al.*, 2002). Por otra parte, en 1826 se registró una pandemia de cólera, ocasionada por *Vibrio cholerae*, que inició en la India y se propagó a través de las rutas comerciales y militares, desde Asia Central al Medio Oriente, después a Europa y posteriormente a Norteamérica (Chan, Tuite y Fisman, 2013). Asimismo, la

Primera Guerra Mundial propició una pandemia de influenza en 1918, ocasionada por una cepa de virus influenza H1N1 (Saunders-Hastings y Krewski, 2016).

Por otra parte, la colonización europea del continente americano, representó uno de los eventos con mayor impacto epidemiológico en la historia humana (Dobyns, 1993). Sin embargo, se desconoce la mayoría de los agentes causantes de las enfermedades infecciosas presentes durante este periodo (Koch *et al.*, 2019). Por lo que su estudio requiere la integración de distintos tipos de evidencias como la documental, arqueológica y osteopatológica. La aportación de estas evidencias en el entendimiento de las enfermedades infecciosas en poblaciones humanas antiguas se describe en la siguiente sección.

1.2. Estudio de enfermedades infecciosas en poblaciones humanas antiguas

La paleopatología se encarga del estudio de enfermedades en las poblaciones humanas antiguas por medio del análisis de restos orgánicos como son tejidos, cálculo dental, y coprolitos (Inhorn y Brown, 1990). La integración de la evidencia documental, arqueológica y osteológica han contribuido a este campo de estudio (Grauer, 2018).

La descripción documental de los signos de enfermedades infecciosas ha permitido inferir su existencia a través del tiempo. Sin embargo, esta evidencia se ve influenciada por el contexto social y cultural de los autores (Siek, 2013). Además, la presencia de errores de transcripción, traducción e interpretación, y ambigüedades en la descripción de la enfermedad y la temporalidad de los acontecimientos, pueden generar conclusiones erróneas (Hong, 1997).

Otro tipo de evidencia que ha permitido el estudio de enfermedades infecciosas en el pasado, es el contexto arqueológico de los restos funerarios, como el tipo de objetos asociados y la ubicación de los entierros. Un ejemplo de esto, es la identificación de cal en entierros humanos de los siglos XVI-XVIII en la Ciudad de México. La cal era utilizada para promover una degradación rápida de los restos en individuos que padecieron un proceso infeccioso durante su vida, y así atenuar su propagación (Salas Contreras, 2007; Bianucci *et al.*, 2008). Adicionalmente, los enterramientos colectivos o fosas comunes han sido asociados con la presencia de brotes epidémicos, ya que sugieren un número elevado de muertes simultáneas (Antoine, 2008).

Finalmente, la evidencia osteopatológica se basa en la descripción y clasificación de lesiones en restos humanos que indican la presencia de una o múltiples enfermedades infecciosas durante la vida del individuo (Pinhasi y Mays, 2008). Sin embargo, las lesiones óseas patognomónicas (determinantes en el diagnóstico de una enfermedad) son limitadas, como en el caso del *mal de Pott*, que es característico en la tuberculosis, y la *caries sicca* en las treponematosis (Klunk, 2018). Un tipo de lesión inespecífica comúnmente identificada en restos arqueológicos es la reacción perióstica, que se define como una inflamación de la membrana del tejido conectivo que rodea al hueso. No obstante, su etiología es diversa, ya que puede ser inducida por una infección bacteriana, una enfermedad metabólica o un trauma (Ortner D. J., 2003; Davies-Barrett, Antoine y Roberts, 2019).

Dentro del análisis osteológico de enfermedades infecciosas, han surgido planteamientos que cuestionan su interpretación, como la "paradoja osteológica", propuesta por Wood *et al.* (1992). Esta paradoja sugiere que las lesiones esqueléticas predominan en un individuo saludable y están ausentes en un individuo enfermo. Bajo este supuesto, la presencia de lesiones inactivas o cicatrizadas, indican que el individuo pasó por un proceso infeccioso y sanó, por lo que se puede inferir que poseía un sistema inmune competente. En contraste, se podría asumir que los individuos sin lesiones esqueléticas, no presentaron una resistencia ante la enfermedad que les permitiera sobrevivir lo suficiente para desarrollar lesiones óseas (Wood *et al.*, 1992). Sin embargo, varios autores han discutido la validez de la paradoja osteológica, argumentando que la interpretación de la presencia o ausencia de lesiones esqueléticas no puede ser limitado a un estado de salud binario—salud o enfermedad—del individuo, y que deben considerarse distintos factores y líneas de evidencia. Esta discusión permite reconocer que la inferencia de enfermedades en el pasado debe realizarse de forma multidisciplinaria e integrativa (Goodman, 1993; Civera Cerecedo, 2006; Pinhasi y Mays, 2008).

Si bien las tres fuentes de evidencia antes descritas: documental, arqueológica y osteológica, han sido el eje de la investigación paleopatológica desde sus inicios, varias de sus limitaciones han sido contrarrestadas gracias a la implementación de herramientas moleculares. El estudio de Spigelman y Lemma, 1993, es considerado como pionero al identificar el DNA de *Mycobacterium tuberculosis* a través de PCR en una vértebra del periodo bizantino (siglo IV-1453) proveniente de Turquía. Esto marcó un punto de inflexión en el desarrollo de estudios subsecuentes sobre la identificación de agentes causantes de enfermedades infecciosas por

medio del análisis de DNA antiguo (aDNA). La siguiente sección describe las características del aDNA, así como las herramientas que han permitido su estudio.

1.3. DNA antiguo

El aDNA es un término empleado para describir el material genético recuperado de restos orgánicos antiguos (Marciniak y Perry, 2017).

El primer estudio de aDNA involucró la recuperación de secuencias de DNA mitocondrial del quagga (una subespecie extinta de la cebra) a través de clonación bacteriana (Higuchi *et al.*, 1984). La implementación de la reacción en cadena de la polimerasa (PCR) en 1989, para la amplificación del aDNA a partir de restos antiguos, marcó el primer parteaguas en el desarrollo de esta disciplina (Mullis y Faloona, 1989; Paabo, 1989).

El aDNA se caracteriza por presentar un daño físico y químico acumulativo *postmortem,* como son la rupturas de cadena sencilla y cadena doble. En su conjunto, estas lesiones *postmortem* son producidas por la degradación hidrolítica y oxidativa del DNA, que propicia la fragmentación (menor a 100 bp), la pérdida de bases nitrogenadas y la presencia de nucleótidos atípicos (Lindahl, 1993; Willerslev y Cooper, 2005; Lan y Lindqvist, 2018). La desaminación hidrolítica de citosina a uracilo, causa una alta tasa de transiciones C-> T o G -> A durante la amplificación por PCR (Willerslev y Cooper, 2005).

Las condiciones ambientales influyen en la integridad del aDNA, así como la proporción del contenido endógeno con respecto al exógeno (Kistler *et al.*, 2017). Los restos arqueológicos que se preservan en un ambiente seco y frío como en el *permafrost*, se caracterizan por tener fragmentos de DNA más largos, un mayor contenido endógeno, y un menor grado de contaminación. Mientras que el DNA recuperado a partir de restos excavados en regiones tropicales, presenta una menor tasa de sobrevivencia y fragmentos de DNA más cortos (Hofreiter *et al.*, 2015).

Otra de las características del aDNA, es la susceptibilidad a la contaminación de fuentes externas, como el DNA moderno proveniente de la manipulación del resto arqueológico, a partir de la excavación y durante su procesamiento en el laboratorio. De acuerdo con especificaciones establecidas y consensuadas para evitar introducir alguna fuente de contaminación, se requiere

un laboratorio exclusivo para su procesamiento, aislado de cualquier otro laboratorio en donde se procesen muestras modernas y se generen productos de PCR (Cooper y Poinar, 2000) (Figura 2).

Después de la PCR, la innovación con mayor impacto en el campo de la paleogenómica ha sido la implementación de tecnologías de siguiente generación (NGS, por sus siglas en inglés) en el análisis de aDNA. Dicha innovación dio lugar al campo de la paleogenómica. Esta disciplina ha proporcionado una resolución y escala sin precedentes para estudiar de manera directa los cambios en la composición genética de poblaciones humanas a través del tiempo (Shapiro y Hofreiter, 2014).

El uso de NGS es ideal para el análisis de aDNA, ya que este tipo de metodología requiere fragmentos cortos de DNA, por lo que no es necesario llevar a cabo una fragmentación del DNA de la muestra antigua. Una ventaja adicional es que al secuenciar aDNA con estas tecnologías es posible estimar la frecuencia de desaminación de citosinas a través de análisis bioinformáticos y a partir de ello revelar si este patrón de daño es consistente con lo esperado para una muestra antigua. De hecho, observar dichos patrones de daño ha sido considerado como un estándar para autentificar la naturaleza antigua del DNA (Ginolhac *et al.*, 2011).

La implementación de distintas estrategias metodológicas y analitícas dentro del área de la paleogenómica, no solo ha permitido el análisis de genomas humanos, sino también de varios patógenos de importancia histórica (Spyrou *et al.*, 2019).

En cuanto al desarrollo de la paleogenómica en México, este ha sido paulatino. El análisis de aDNA de muestras arqueológicas mexicanas es realizado en su mayoría por grupos de investigación extranjeros, lo cual ha aminorado su crecimiento a nivel local. Sin embargo, existen antecedentes de la realización de proyectos de investigación enfocados en el análisis de aDNA en México—descritos en la siguiente sección.

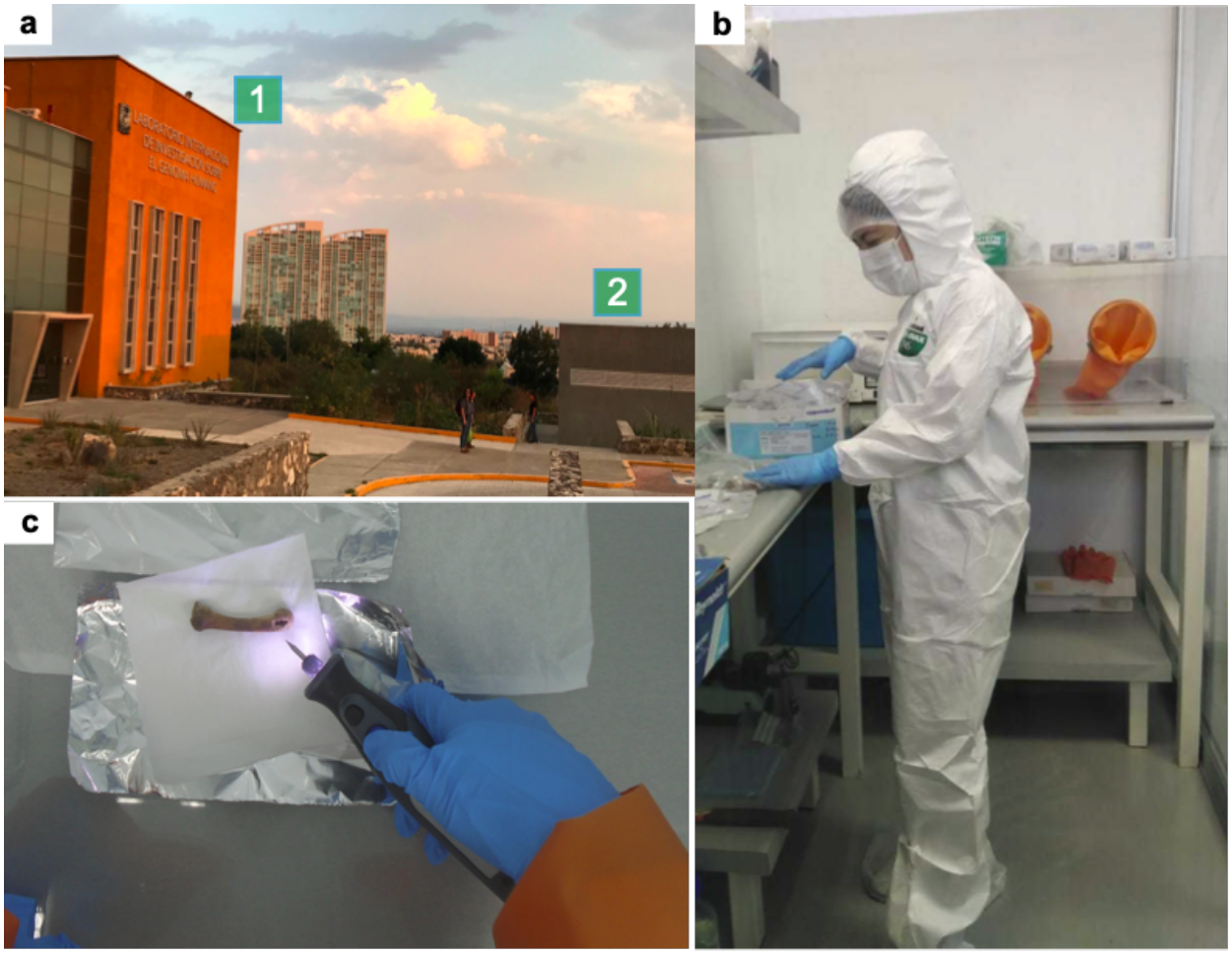


Figura 2. Laboratorio de Paleogenómica del LIIGH y aspectos a considerar en el análisis de aDNA. **a)** Localización del laboratorio exclusivo para el análisis de aDNA y físicamente aislado del laboratorio de DNA moderno. **1**.Laboratorio de DNA moderno, **2**.Laboratorio de Paleogenómica del LIIGH; **b)** Vestimenta adecuada para el procesamiento de muestras antiguas dentro del laboratorio, que consiste en traje de cuerpo completo, guantes, cofia, cubrebocas y cubrezapatos; **c)** Ejemplificación de la toma de muestra a partir de un hueso antiguo, usando una herramienta de mano giratoria llamada Dremel.

1.3.1. Investigación de aDNA en México

El estudio de (Vargas Sanders y Alcalá, 1990) fue pionero al lograr extraer DNA genómico de restos óseos de un individuo de ~700 años de antigüedad proveniente del sitio del Peñón del Marqués, en la Ciudad de México. Posteriormente, la investigación de aDNA en el país se ha centrado en el análisis de la diversidad genética de poblaciones prehispánicas, a través de la amplificación de marcadores de haplogrupos mitocondriales americanos, en restos de

Teotihuacan, Ciudad de México (Álvarez-Sandoval *et al.*, 2015; Samudio *et al.*, 2017); Xcaret, Quintana Roo (González-Oliver *et al.*, 2001); y en otras poblaciones mayas (Ochoa-Lugo *et al.*, 2016).

Por otra parte, el estudio De La Cruz et al. 2008 determinó el sexo, mediante amplificación de un fragmento del gen de la amelogenina, en individuos sacrificados al dios Ehécatl-Quetzalcóatl, del sitio Tlatelolco, Ciudad de México. El análisis reveló la predominancia de individuos del sexo masculino, lo que es consistente con la veneración a un dios masculino (Cruz *et al.*, 2008).

Además del análisis de aDNA en restos humanos, otro tipo de sustratos que han sido analizados de manera exitosa, es el caso del maíz. En Vallebueno-Estrada et al. 2016, se logró reconstruir el genoma de un maíz (*Zea mays*) con una antigüedad de 5,000 años AP por medio de NGS. La muestra (olote) se excavó en un sitio arqueológico en el Valle de Tehuacán, Puebla, muy cerca del supuesto centro de domesticación del maíz. El análisis del genoma antiguo reveló que hace 5,000 años el maíz estaba parcialmente domesticado y pertenecía a un grupo de plantas emparentadas, lo que sugiere el uso de prácticas de mejoramiento tradicional por medio de la autopolinización (Vallebueno-Estrada *et al.*, 2016). Por su parte, Pérez-Zamorano et al. 2017 lograron reconstruir el genoma plastídico y mitocondrial a partir de la misma muestra de maíz que Vallebueno-Estrada et al. 2016. Los análisis filogenéticos muestran que el genoma recuperado es basal con respecto a otros genomas de *Zea mays* (Pérez-Zamorano *et al.*, 2017).

A pesar del escaso número de estudios de aDNA realizados a nivel local, han sentado precedente en el desarrollo de la paleogenómica en México. Por lo que los resultados que se derivan de la presente tesis, muestran los avances tanto a nivel analítico como metodológico, de este campo a nivel local.

1.3.2. Identificación de patógenos a través del aDNA

El análisis de aDNA a partir de muestras arqueológicas, no sólo han permitido plantear inferencias sobre la historia evolutiva humana, si no también sobre los patógenos que han sido parte de ésta.

El primer estudio de aDNA de patógenos fue el de Spigelman y Lemma, 1993, en el que se analizó aDNA de *Mycobacterium tuberculosis,* a partir de una vértebra del periodo bizantino (395-

1453 d.C.) de Turquía (Spigelman y Lemma, 1993). Seguido de esto, Salo et al. 1994 identificaron aDNA del mismo patógeno, a partir de un pulmón momificado de 1000 años de antigüedad proveniente de Perú (Salo *et al.*, 1994). Posteriormente, el análisis de aDNA de patógenos se extendió a otro tipo de sustratos, cómo la pulpa dental (tejido vascularizado localizado dentro del diente). El primer estudio en usar este sustrato, identificó DNA de *Yersinia pestis* en seis individuos (1600-1800 d.C.) provenientes de Francia (Drancourt *et al.*, 1998). Éste y otros estudios demostraron que el diente es un sustrato idóneo para este tipo de análisis, ya que al estar en contacto directo con la circulación sanguínea es posible identificar el DNA del patógeno en individuos con infección sistémica (Nguyen-Hieu, Aboudharam y Drancourt, 2011).

Entre 1993 y 2008, diversos estudios reportaron la recuperación de aDNA de al menos 20 patógenos antiguos mediante PCR a partir de diversos sustratos (diente, hueso, tejido momificado, tejido embebido en parafina y coprolitos) (Raoult y Drancourt, 2008). Esto sentó las bases para el desarrollo de la Paleomicrobiología (Drancourt y Raoult, 2005). Sin embargo, fue evidente la falta del establecimiento de medidas preventivas experimentales y analíticas durante la realización de este tipo de estudios (Campana *et al.*, 2014).

Durante la última década, el refinamiento de estrategias experimentales y analíticas en NGS, ha permitido la identificación inequívoca de aDNA de patógenos antiguos y, en algunos casos, la reconstrucción de sus genomas completos (Spyrou *et al.*, 2019). Otro desarrollo metodológico relevante en el estudio genómico de patógenos es la captura-enriquecimiento por hibridación, que ha permitido analizar los genomas de patógenos asociados con brotes epidémicos de relevancia histórica (Bos *et al.,* 2026) y ha permitido realizar inferencias sobre su evolución y patogenicidad (Schuenemann *et al.*, 2018).

En el artículo de revisión desarrollado en esta tesis, se describen los avances y limitaciones de la genómica de patógenos antiguos, mientras que el artículo de investigación y el primer reporte del proyecto, describen ejemplos específicos de la aplicación de los métodos de captura-enriquecimiento para la reconstrucción de genomas de patógenos (*Tannerella forsythia* y *Salmonella enterica*, respectivamente).

Hasta el momento se han caracterizado aproximadamente 175 genomas antiguos de patógenos causantes de enfermedades infecciosas; en su mayoría bacterias y en menor proporción virales y eucariotas, principalmente de muestras de origen europeo (Spyrou *et al.*, 2019; Fellows Yates

et al., 2020). Las inferencias obtenidas a través de la genómica comparativa de patógenos antiguos y actuales, han contribuido al entendimiento de su historia evolutiva, orígenes, diversidad, macroecología y de su rango geográfico a través del tiempo.

La reconstrucción de genomas de patógenos antiguos puede delimitar la ventana de tiempo estimada para la emergencia de patógenos que afectan a poblaciones actuales, así como revelar linajes extintos (Ho y Duchêne, 2020). Un ejemplo ilustrativo es el análisis del genoma antiguo del virus *Measles morbillivirus*, causante de paperas, obtenido a partir de un espécimen de museo (tejido pulmonar) de un niño que murió en 1912. El análisis filogenético reveló que este patógeno divergió del virus de la peste bovina en el siglo VI a.C. (Düx *et al.*, 2020). Por su parte, en el estudio de Duggan *et al.*, (2016) recuperaron el genoma del virus de la viruela a partir de tejido momificado de origen europeo, con una temporalidad de entre 1643 y 1655. En este estudio se determinó que el tiempo evolutivo del virus de la viruela (1588-1645) es más reciente de lo que se pensaba (1145 a.C.). Otra investigación relevante en este contexto es la de Zhou et al., (2008) en que la reconstrucción de un genoma de *Salmonella enterica* paratyphi C, a partir de un diente de 800 años de antigüedad de origen europeo, permitió estimar que el tiempo de divergencia a partir del ancestro común ocurrió hace 4000 años, lo que coincide con el tiempo de domesticación de los cerdos en Europa y representa la especialización de un patógeno zoonótico de animales domesticados a humanos (Zhou *et al.*, 2018).

El estudio de enfermedades infecciosas en América a través del análisis de aDNA, ha sido limitado en comparación con Europa (Darling y Donoghue, 2014); en particular, no se ha hecho uso del análisis de aDNA para la identificación de agentes causantes de procesos infecciosos en poblaciones cazadoras-recolectoras. El segundo reporte del proyecto, contribuye en este sentido al reportar la identificación del aDNA de patógenos en la región de la costa norte de la Patagonia Argentina en grupos de cazadores-recolectores (6000-300 AP).

1.4. Evidencia documental, arqueológica y genética de enfermedades infecciosas humanas en México

El presente trabajo doctoral centra su interés en el estudio de patógenos en individuos antiguos de México. A través de la integración del cuerpo de conocimiento generado por otras disciplinas, principalmente por la arqueología, antropología y paleopatología, permitirá abordar los resultados obtenidos desde una perspectiva interdisciplinaria. A continuación se describe la evidencia

documental, osteológica y genética de las enfermedades infecciosas presentes en el periodo prehispánico y colonial de México.

1.4.1. Presencia de enfermedades infecciosas durante el periodo prehispánico

Gran parte del registro documental correspondiente al periodo prehispánico fue destruido durante la conquista española, por lo que el registro de enfermedades infecciosas durante este periodo es limitado (Eleodoro Balboa, 2014). La mayoría del registro documental disponible respecto a enfermedades infecciosas se deriva de escritos testimoniales de los años subsecuentes a la conquista española (Treviño, 2005).

De acuerdo con fuentes históricas del siglo XVI, como el "Libro Chilam Balam de Chumayel", las condiciones de salud en la zona maya durante el periodo prehispánico eran idóneas y las enfermedades infecciosas eran poco comunes (Roys, 1933; Carrera Stampa, 1969; Treviño, 2005). En contraste, Viesca (1992; 1998) y Mandujano, Camarillo y Mandujano (2003) describen la evidencia documental a partir de crónicas, la presencia de enfermedades infecciosas respiratorias y gastrointestinales, descritas como tos con esputo sanguinolento y diarreas. De acuerdo a estos reportes, dichos padecimientos ocurrieron durante 1450-1454, en los sitios de México-Tenochtitlán, Teotihuacán y Cuicuilco (Viesca, 1992). Se sugiere que las condiciones de vida, la falta de higiene, el consumo de agua contaminada, y la presencia de sequías y heladas extemporáneas propiciaron el desarrollo de estas enfermedades (Viesca, 1998; Mandujano, Camarillo y Mandujano, 2003).

Existen dos documentos que señalan la presencia de brotes epidémicos que afectaron a la población de manera importante. La *Tercera Relación*, por Chimalpahin, reporta la presencia de posible difteria en 1456, la cual ocasionó el despoblamiento de Chalco (Chimalpáhin Cuauhtlehuanitzin, 1998; Mandujano, Camarillo y Mandujano, 2003). Por otro lado, el códice Chimalpopoca reporta otro brote epidémico en 1496; se infiere que se trató de tifo exantemático, que afectó a la población de Xochitlán, Tehuantepec y Amixtlán, actualmente ubicados en el estado de Puebla (Mandujano, Camarillo y Mandujano, 2003; Velázquez y León Portilla, 2019).
La evidencia osteológica ha contribuido en gran medida sobre el estudio de enfermedades infecciosas del pasado. La tabla 1 describe las inferencias de enfermedades infecciosas basadas en evidencia osteológica. El estudio de Cisneros (2013), identificó la presencia de reacción perióstica en los restos arqueológicos de individuos de Ecatepec, Tlalpizáhuac y Tepozoco,

Estado de México, con una temporalidad de 300 a.C.-750 d.C. Esta evidencia hace suponer que existieron enfermedades infecciosas recurrentes. Asimismo, la reacción perióstica ha sido identificada en otros individuos del Estado de México, con una temporalidad de 300 a.C.-100 d.C. (Cisneros, 2010).

La tuberculosis, causada por *Mycobacterium tuberculosis,* es otra enfermedad que se ha identificado con frecuencia en restos del periodo prehispánico (Tabla 1). Otras lesiones reportadas con cierta frecuencia en restos arqueológicos del periodo prehispánico, son aquéllas sugerentes de sífilis, como son *caries sicca*, tibia de sable, dientes de Hutchinson, principalmente (Barker, 1977). Se han identificado individuos con este tipo de lesiones provenientes de Tehuacán, Puebla, con una temporalidad de 2500 a.C.-200 d.C. (Anderson 1967). También se han identificado en individuos encontrados en la cueva de Cecilia, Sonora, en 900 d.C. - 1521 d.C. (Dávalos, 1964), en la Cueva de la Candelaria, Coahuila (Romano, 1956), en Tlatelolco (Dávalos, 1964); en Xochicalco, Morelos (Stewart, 1956), en Yagul, Oaxaca (Estrada *et al.*, 1970) y en Mitla, Oaxaca (Robles García y Molina Villegas, 1998).

Otro tipo de enfermedades infecciosas de relevancia desde el punto de vista paleopatológico son las periodontales. Por medio de la medición de la unión esmalte-cemento al reborde óseo, Velázquez (2006) reporta la presencia de este tipo de padecimiento en individuos con una temporalidad de 1400-1500 d.C. de Tlatelolco, Ciudad de México, de Cholula, Puebla, de Jaina, Veracruz, y de Los Cabos, Baja California Sur (Velázquez C., 2006). Asimismo, Cisneros (2010) reporta la enfermedad periodontal en individuos excavados en el Cerro de Ehécatl, Ecatepec, que corresponden al 300 a.C - 750 d.C. (Cisneros, 2010). En el artículo de investigación de la presente tesis, se describe el primer análisis paleogenómico de patógenos asociados con el desarrollo de enfermedad periodontal en individuos del periodo prehispánico y colonial de México.

Con respecto a la evidencia molecular, Martínez et al. (2014) amplificaron por PCR un elemento de inserción, IS6110, característico del complejo *M. tuberculosis*, en individuos con lesiones de tuberculosis vertebral, provenientes de Pajones, Zacatecas y Tamtoc, San Luis Potosí. Sin embargo, el procesamiento de las muestras no se llevó a cabo en un laboratorio específico para el análisis de aDNA (Garfias, 2016), por lo que los resultados pudieron deberse a una posible contaminación de DNA moderno de *M. tuberculosis.* En un estudio independiente Steffani-Vallejo (2014), analizó estos mismos individuos, mediante PCR de los marcadores IS6110, IS1081, el

gen *oxyR* y la región 16S de *M. tuberculosis*. El resultado obtenido fue una secuencia nucleotídica a partir de un individuo de Tamtoc, con un porcentaje de identidad del 77% con IS1081. De igual forma, el resultado obtenido probablemente sea un falso positivo, ya que la secuencia nucleotídica obtenida tiene mayor similitud con bacterias de origen ambiental (Garfias, 2016). En un tercer estudio se analizaron estos mismos individuos usando los marcadores *tpp15* de *T. pallidum* e IS6110 de *M. tuberculosis*, y no se logró amplificar ninguno de ellos (Garfias, 2016).

En resumen, existe evidencia documental y osteológica de enfermedades infecciosas durante el periodo prehispánico, e incluso de posibles brotes epidémicos. Sin embargo, la cantidad de información es limitada comparada con la disponible para el periodo colonial, la cual se resume en la siguiente sección.

Tabla 1. Enfermedades infecciosas durante el periodo prehispánico en México.

Enfermedad	Temporalidad	Distribución geográfica	Tipo de Evidencia	Referencia
Sífilis	2500 a.C.-200 d.C	Tehuacán, Puebla.	Osteológica	Anderson, 1967
Enfermedad periodontal	300 a.C.-100 d.C	el Cerro de Ehécatl, Ecatepec	Esquelética	Cisneros, 2010
Inespecifíca	300 a.C.-750 d.C	Ecatepec, Tlalpizáhuac y Tepozoco,	Osteológica	Favila Cisneros, 2013
Gran pestilencia	901 d.C. – 1000 d.C.	Tula	Histórica	Mandujano et al. 2003
Sífilis	900 d.C. - 1521 d.C.	Tlatelolco	Osteológica	Hurtado, 1964
	900 d.C. - 1521 d.C.	Xochicalco, Morelos	Osteológica	Stewart 1956
	900 d.C. - 1521 d.C.	Yagul, Oaxaca	Osteológica	Estrada et al. 1970
	900 d.C. - 1521 d.C.	Cueva de Cecilia, Sonora	Osteológica	Hurtado, 1964
	900 d.C. - 1521 d.C.	Cueva de la Candelaria, Coahuila	Osteológica	Romano,1956
Tuberculosis	200-650 d.C	Pajones, Zacatecas	Osteológica	Martínez Mora, et.al., 2014
	900 d.C. - 1521 d.C.	Tlatelolco, Cd. de México	Osteológica	Hurtado, 1964
Tuberculosis/ Sífilis	1100 - 1521 d.C.	Tamtoc, San luis Potosí	Osteológica	Hernández, 2012
	1200 d.C.- 400 a.C.	Tlatilco, Estado de México	Osteológica	Faulhaber,1965
	1325 -1521 d.C.	Cd. de México	Osteológica	Salas ,1982
	1337-1521 d.C.	Tlatelolco, Cd. de México	Osteológica	Contreras, 2010
Sífilis Enfermedad periodontal	1400 d.C.	Llaadzié, Mitla Oaxaca	Osteológica	Robles García, 1892
Catarro pestilential	1440 d.C.	Valle de México	Histórica	Viesca, 1998
Enfermedad periodontal	1400-1500 d.C.	Tlatelolco, Ciudad de méxico; Cholula, Puebla, Jaina, Veracruz; Los Cabos, Baja California Sur	Esquelética	Velázquez, 2006
Esputo sanguinolento y enfermedades respiratorias	1450-1454 d.C.	México-Tenochtitlán, Teotihuacán y Cuicuilco	Histórica	Viesca 1992;1998; Mandujano et al. 2003
Difteria	1456 d.C.	Tenochtitlán	Histórica	Tercera Relación por Chimalpahín. Chimalpáhin Cuauhtlehuanitzin, 1998
Tifo exantemático	1496 d.C.	Xochtlán, Tehuantepec y Amaxtlan	Histórica	Mandujano et al. 2003

1.4.2. Presencia de enfermedades infecciosas durante el periodo colonial

El registro documental de las enfermedades infecciosas durante el periodo colonial comprende principalmente matrículas de tributos, relaciones geográficas, crónicas de conquistadores y misioneros, registros parroquiales y de hospitales (Bravo y Guerra, 1570; Beaumont, 1932; Telleriano-Remensis, 1995; Cook y Lovell, 2000; Inegi, 2005; Gimmel, 2008; Baracs, 2015; Velázquez y León Portilla, 2019). En conjunto, estos documentos han permitido reconstruir la propagación de ciertas enfermedades como la viruela y el sarampión (Magaña, 2013), así como revelar su impacto demográfico en la población (Cook y Lovell, 2000).

De acuerdo con la historiografía médica, durante el siglo XVI la presencia de brotes epidémicos ocurrieron frecuentemente en la Nueva España (Malvido y Viesca, 1985; de Montellano, 1992). Las condiciones de vivienda de las poblaciones en el centro de México eran insalubres por el hacinamiento. La desnutrición y condiciones de vida detrimentales, ocasionadas por las hambrunas durante crisis agrícolas, así como la esclavitud y la explotación económica, propiciaron el desarrollo de enfermedades infecciosas y, por lo tanto, una alta mortalidad en la población indígena (Cook y Lovell, 2000; Crespo, 2010; Meza Manzanilla, 2013).

De acuerdo con (Malvido, 1992), la historia epidemiológica en la Nueva España se divide en dos épocas: “Contacto”, que corresponde de 1520 a 1560, y “Colonial”, de 1561 a 1821 (Malvido, 1992). Las epidemias más severas reportadas durante el periodo de Contacto (1520-1560) son las siguientes (Tabla 2):

i) La primera epidemia registrada fue ocasionada por la viruela (*“hueyzahuatl”* en náhuatl o gran lepra), durante 1520-1521. Se presentó principalmente en el centro de México (Cook y Lovell, 2000). Se estima que ocasionó la muerte de entre 5 y 8 millones de individuos (Acuña-Soto y Stahle, 2002). Se le atribuye como una de las causas que propició la caída de México-Tenochtitlan frente a los españoles (Mandujano, Camarillo y Mandujano, 2003).

ii) En 1531 el sarampión (*“mathaltotonqui”* o “*tepitonzáhuatl*” que significa pequeña lepra) se propagó principalmente en el centro de México y en Tabasco (Motolinía y Toribio, 1971), y tuvo un alta tasa de mortalidad infantil (Nicholson, 1966).

iii) Durante 1545-1548 se describe una enfermedad caracterizada por hemorragias intensas y fiebre elevada, los indígenas la denominaron *"Cocoliztli"*, que significa pestilencia en náhuatl (García-Martínez, 2005; Mendieta, 1945). Se estima que entre 5 y 15 millones de individuos de la Nueva España murieron a causa de esta enfermedad. En Tlaxcala, Puebla y Ciudad de México se registró el mayor número de muertes (Cook y Lovell, 2000).

iv) En 1550 se registró un brote de paperas ("quechopotzahualiztli") que afectó a una gran cantidad de la población. El Códice Aubin, lo describe cómo un abultamiento doloroso en el área del cuello y fiebre (Lehmann, Kutscher y Vollmer, 1981; Telleriano-Remensis, 1995). Se desconoce el número aproximado de muertes ocasionadas por este brote.

Por otra parte, en el periodo Colonial (1561 a 1821) se tiene registro de los siguientes brotes epidémicos:

i) Entre 1576-1581 se registró un brote epidémico denominado "*hueycocoliztli*" (gran pestilencia), este afectó principalmente el centro de México (Alonso de Molina, 1977; Guevara Flores, 2017). Esta enfermedad pudo haber sido ocasionada por tifo (Márquez, 1993), producida por varias especies de bacteria del género *Rickettsia*. También se piensa que pudo tratarse de la peste (Malvido y Viesca, 1985). Se estima que este brote epidémico ocasionó la muerte de 2 a 2.5 millones de individuos (Acuña-Soto y Stahle, 2002). Se ha sugerido que ambos brotes epidémicos, el de 1545-1548 (*"Cocoliztli"*) y el de 1576-1581, pudieron ser epidemias compuestas, ocasionadas por la co-infección de diversos patógenos (Cook y Lovell, 2000).

ii) Durante 1736-1739 se presentó una enfermedad con signos similares a tifo exantemático, fiebre amarilla (ocasionada por un virus del género *Flavivirus*) y malaria (causada por diferentes especies de *Plasmodium*), fue nombrada por los indígenas como *"matlazahuatl"* que significa erupción cutánea de color azul violeta en forma de red (Cuenya, 1996). Se desconoce la tasa de mortalidad de este brote.

Por el contrario, el análisis osteológico ha permitido la identificación de otro tipo de enfermedades infecciosas, como es el caso de sífilis. La identificación de lesiones específicas de sífilis (*caries sicca*, engrosamiento del periostio en la parte anterior de los huesos largos, periostitis gomatosa

y esclerosis en las diáfisis), se han reportado en las colecciones esqueléticas del Hospital de San Juan de Dios y del Hospital San José de los Naturales, que corresponden a una temporalidad de entre 1531 y 1822 (Márquez Morfín, 2015; Márquez Morfín y Meza Manzanilla, 2015). La sífilis congénita, que se transmite de la madre infectada al feto, ocasionó una alta mortalidad en niños durante 1915. Uno de los casos reportados de esta enfermedad (basado en la identificación de osteocondritis), se encontró en el convento de Santa Isabel, en la Ciudad de México (1654 - 1669) (Márquez Morfín y Sosa Márquez, 2016).

Tabla 2. Principales brotes epidémicos durante el periodo Colonial en México.

Epidemia	Términología náhuatl*	Periodo	Distribución geográfica	Fuentes históricas	Referencia
Viruela	*huey zahuatl*	1520-1521	Costa del Golfo de México, Sierra Central	Anales de Tecamachalco 1903:7; Códice Florentino	Kubler et.al, 1952
Sarampión	*mathaltotonqui*	1531	Centro de México, Tabasco	Códice Aubin	Motolinía y Toribio, 1971
	Cocoliztli o matlazahuatl	1545-1548	Centro de México	Códice Telleriano-Remensis, Tira de Tepecpán	Telleriano-Remensi, 1995; Noguez, 1978; Mendieta, 1945
Paperas	*quechopotzahualiztli*	1550		Códice Aubin	Lehmann et al., 1981
Posible Tifo exantemático	*Cocoliztli*	1576-1580	Centro de México	Tira de Petechpán, Codex Mexicanus, lámina 86; Anales antiguos de México no.16)	Alonso de Molina, 1977; Guevara Flires, 2017

En cuanto a la evidencia molecular de los patógenos presentes durante el periodo colonial se han publicado tres estudios realizados por grupos de investigación establecidos en el extranjero. El primer estudio, Schuenemann et al. (2018), identificó un genoma de *T. pallidum ssp. pertenue* y dos genomas de *ssp. pallidum*, a partir de infantes con lesiones características de sífilis, recuperados del convento de Santa Isabel, en la Ciudad de México (1654 - 1669).

El segundo estudio (Vågene *et al.*, 2018), reconstruyó el genoma de *Salmonella enterica* Paratyphi C, bacteria responsable de fiebre paratifoidea, a partir de restos encontrados en un cementerio de la época colonial, en Teposcolula-Yucunda, Oaxaca. De acuerdo con la evidencia documental y arqueológica, en este sitio se sepultaron víctimas de "*Cocoliztli*" (Spores *et al.*, 2007). Esto llevó a los autores a sugerir que *Salmonella enterica* Paratyphi C podría ser el agente causante de *"Cocoliztli"*. Sin embargo, los signos de esta enfermedad, descritos en documentos históricos, no coinciden con los signos característicos de fiebre paratifoidea, ocasionado por Paratyphi C.

Por último, el tercer estudio es el de Barquera et al., (2020) en el que caracterizó el genoma de *T. pallidum pertenue* y del virus Hepatitis B, a partir de dos individuos de origen africano del Hospital San José de los Naturales, en la Ciudad de México. El análisis filogenético de ambos patógenos, apoyan que el origen de los individuos analizados fue de África occidental.

Hasta el momento son muy escasos los estudios paleogenómicos realizados en México. En el estudio de Guzmán-Solís et al., (2020) reconstruyeron dos genomas antiguos del virus parvovirus B19 humano y uno del virus Hepatitis B, a partir de individuos excavados en el mismo sitio que Barquera et al., (2020), y un genoma del parvovirus humano B19 en un individuo encontrado en el Templo de la Inmaculada Concepción "La Conchita", Ciudad de México. Los genomas virales reconstruidos en Guzmán-Solís et al., (2020) mostraron similitud con cepas de origen africano, lo cual es consistente con la ancestría africana inferida por los análisis morfológicos, isotópicos y genéticos de los individuos analizados. Adicionalmente, el artículo de investigación que se deriva de la presente tesis doctoral, representa la primera evidencia directa del aDNA de *Tannerella forsythia*, patógeno asociado con el desarrollo de enfermedad periodontal, en individuos del periodo prehispánico y colonial de México. A través de la reconstrucción parcial de 11 genomas de *T. forsythia*, se observó una distribución filogenética diferencial entre estos, lo que sugiere que las cepas presentes en el periodo prehispánico acompañaron a los primeros pobladores de América y nuevas cepas fueron introducidas durante la conquista española de México (Bravo-Lopez *et al.*, 2020). Adicionalmente, en esta tesis se describe, a manera de reporte, la identificación del aDNA de *Salmonella enterica* Paratyphi C, en el sitio del Templo de la Inmaculada Concepción "La Conchita", Ciudad de México.

En resumen, la conquista española impactó de manera determinante en el estado de salud de la población nativa en la Nueva España, en particular, a través de la introducción de enfermedades

infecciosas. La evidencia documental, osteopatológica y arqueológica, han permitido reconstruir la etiología de algunas de estas enfermedades. Recientemente, la paleogenómica ha permitido plantear inferencias a un nivel de resolución más fino sobre las enfermedades infecciosas presentes durante el periodo Colonial, así como entender la historia evolutiva de los agentes causantes de estas enfermedades infecciosas. Sin embargo, es de suma importancia considerar que el análisis de restos arqueológicos debe realizarse dentro de un marco ético y considerar estrategias que permitan su preservación.

1.5. Metodologías para la preservación de muestras arqueológicas

El análisis de aDNA en muestras arqueológicas es un proceso destructivo, por lo que es necesario el uso de metodologías que permitan respaldar y preservar este tipo de muestras que, por su naturaleza, son únicas e irremplazables. Algunas de estas estrategias, son el registro fotográfico y la elaboración de moldes.

Recientemente, la implementación de la técnica de imagenología 3D acoplada con microtomografía de rayos X, ha permitido registrar la morfología exterior e interior de las muestras a alta resolución (Machado *et al.*, 2017). Esta técnica involucra la recolección de proyecciones 2D desde diferentes direcciones de visualización y, a través de algoritmos matemáticos, reconstruye una imagen digital tridimensional, donde cada vóxel (unidad cúbica que compone un objeto tridimensional) representa la atenuación de los rayos-X en ese punto. La ventaja de este método es que la información cuantitativa como el volumen, tamaño, forma, distribución, y conectividad de los poros, puede ser obtenida a través del volumen 3D de las muestras, incluso a nivel de microescala y nanoescala (Lantes-Suárez y Prieto-Martínez, 2017). Así, este análisis de microtomografía permite generar características microestructurales internas de restos arqueológicos como son dientes y huesos; artefactos de metal, cerámica, piedra, textiles, madera, y papel (Tuniz y Zanini, 2018).

1.6. Consideraciones éticas en el análisis de DNA de restos arqueológicos en México

Como se ha mencionado en los párrafos anteriores, los avances metodológicos y analiticos dentro del campo de aDNA antiguo han permitido la recuperación de genomas completos de

agentes causantes de enfermedades infecciosas. Sin embargo, poco se ha reflexionado sobre las implicaciones de los avances de la Paleogenómica en México.

Los restos arqueológicos humanos forman parte de nuestro patrimonio cultural y representan una fuente única e invaluable para entender nuestro pasado. Por lo que su estudio debe desarrollarse bajo lineamientos éticos. En el caso de México, el Instituto Nacional de Antropología e Historia (INAH) es responsable de regular su implementación. Sin embargo, fue hasta el 2019 que se publicó un reglamento elaborado por el INAH llamado "Lineamientos generales para el manejo y resguardo de restos humanos" donde se señalan las directrices para regular la protección, el manejo, traslado, destino, depósito, seguridad y gestión de los restos arqueológicos humanos en el territorio nacional (*Lineamientos generales para el manejo y resguardo de restos humanos*, 18 de Octubre del 2019).

En México, contamos con una amplia riqueza cultural, la cuál ha sido atractiva no sólo para arqueólogos y antropólogos a nivel nacional e internacional, sino también para investigadores del área de la Paleogenómica. No obstante, el presupuesto en México destinado a su investigación, en particular el análisis genómico, es escaso. Esto ha motivado que el análisis de aDNA de restos arqueológicos sea financiado y realizado por grupos de investigación establecidos en el extranjero (Alemania y Estados Unidos, principalmente).

La investigación liderada por grupos de investigación extranjeros, conlleva una serie de problemáticas discutidas en otros estudios (Bardill *et al.*, 2018; Prendergast y Sawchuk, 2018; Wagner *et al.*, 2020) que es pertinente discutir en el contexto de la investigación que conforma esta tesis:

> i) Exclusión de arqueólogos y antropólogos especialistas en la investigación local. Esto puede generar una interpretación equivocada de los resultados obtenidos al no tomar en cuenta fuentes originales del contexto histórico—la mayoría no traducidas al inglés— de los individuos analizados.
>
> ii) Falta de capacitación de recursos humanos a nivel local. El desarrollo e implementación de nuevas tecnologías en el estudio de aDNA por grupos de investigación extranjeros, promueve que su aprendizaje sea enfocado a una minoría. Una iniciativa actual para difundir y enseñar los avances metodológicos del área de aDNA, involucra la organización

de cursos en conjunto con la ENAH y el Laboratorio de Paleogenómica (LIIGH-UNAM) dirigido a estudiantes con diferentes formaciones (Arqueología, Antropología, Biología y Ciencias Genómicas) desde el año 2018. Asimismo, otras iniciativas han logrado incluir a los descendientes de los actores principales de este tipo de estudios, como es SING (Summer internship for INdigenous peoples in Genomics) (Bardill *et al.*, 2018) en EUA, Canadá y Nueva Zelanda.

iii) Dinámica de poder entre laboratorios internacionales y nacionales. El principal interés por la publicación de artículos en revistas de alto impacto, ha promovido un ambiente competitivo con un trasfondo de conflicto de intereses.

iv) Acceso limitado de los resultados generados de la investigación. La difusión de los resultados de la investigación están enfocados a su publicación en una revista de divulgación científica, dejando a un lado la difusión de la información dirigida al público en general. La integración de diversos puntos de vista, es necesaria para enriquecer el alcance de la investigación.

En resumen, es necesario abrir el diálogo sobre las consideraciones éticas en el análisis de aDNA en México, tomando en cuenta la participación de arqueólogos, antropólogos y otros genetistas, con la finalidad de enriquecer y promover la investigación de la paleogenómica en México.

2. OBJETIVOS

2.1. Objetivo general

Analizar el genoma de patógenos humanos presentes en individuos del periodo prehispánico y colonial.

2.2. Objetivos particulares

Los objetivos particulares se desglosan a través de los resultados desarrollados en los artículo de investigación y los dos reportes del proyecto:

1. Caracterizar el metagenoma de restos arqueológicos del periodo prehispánico y colonial (capítulo 5, 6 y 7).

2. Identificar la presencia de DNA de patógenos humanos y diseñar estrategias de captura y enriquecimiento que permitan aumentar su contenido de DNA endógeno (capítulo 5 y 6).

3. Reconstruir el genoma de patógenos y analizar genes relevantes en procesos de patogenicidad, así como la relación filogenética entre cepas antiguas y actuales (capítulo 5 y 6).

3. SÍNTESIS DE ARTÍCULOS Y REPORTES

Esta tesis comprende el análisis paleogenómico de patógenos en individuos antiguos— como primera premisa partimos de una revisión de los avances y desafíos en esta rama de la Genómica (capítulo 4)—contribuimos al entendimiento de la evolución de patógenos asociados a enfermedades infecciosas de la cavidad oral (capítulo 5) y sistémicas (capítulo 6) en individuos del periodo prehispánico y colonial de México. Las herramientas analíticas empleadas en esta tesis, permitieron extender el área de estudio a otra región de América poco explorada, como es el caso de la Costa Norte de la Patagonia Argentina, habitada por cazadores-recolectores (capítulo 7). A continuación se desglosa su contenido.

3.1. Artículo de revisión. What have we learned about past human infectious diseases using paleogenomics?

En preparación
Miriam Bravo-Lopez y María C. Ávila-Arcos

Resumen

El DNA extraído de restos humanos arqueológicos contiene información más allá del componente genético del hospedero; también contiene el DNA de patógenos presentes durante la vida del individuo. La reconstrucción de genomas de patógenos antiguos ha permitido un mayor entendimiento de la etiología, naturaleza e historia evolutiva de diversas enfermedades infecciosas que han sido parte de la historia humana. La caracterización genómica de patógenos antiguos, junto con la evidencia arqueológica y documental, nos puede informar sobre su origen y su expansión geográfica. En esta revisión presentamos los principales hallazgos en el campo de genómica de patógenos antiguos, así como las estrategias metodológicas y las limitaciones para su identificación.

Contribuciones de autoría

Bravo-López realizó el borrador del manuscrito, María C. Ávila-Arcos realizó correcciones del manuscrito.

3.2. Artículo de investigación. Paleogenomic insights into the red complex bacteria *Tannerella forsythia* in Pre-Hispanic and Colonial individuals from Mexico

Publicado:
(Bravo-Lopez et al. 2020)

Miriam Bravo-Lopez, Viridiana Villa-Islas, Carolina Rocha Arriaga, Ana B. Villaseñor-Altamirano, Axel Guzmán-Solís, Marcela Sandoval-Velasco, Julie K. Wesp, Keitlyn Alcantara, Aurelio López- Corral, Jorge Gómez-Valdés, Elizabeth Mejía, Alberto Herrera, Alejandro Meraz-Moreno, Maria de la Luz Moreno-Cabrera, Andrés Moreno-Estrada, Maria A. Nieves-Colón, Joel Olvera, Julia Pérez-Pérez, Katrine Højholt Iversen, Simon Rasmussen, Karla Sandoval, Gabriela Zepeda and María C. Ávila-Arcos.

Resumen

El “complejo rojo” es un agregado de tres bacterias orales (*Tannerella forsythia, Porphyromonas gingivalis* y *Treponema denticola*) responsable de la manifestación clínica severa de enfermedad periodontal. En este estudio, reportamos la primera evidencia directa de DNA antiguo de *T. forsythia* en muestras arqueológicas de dentina y cálculo dental, correspondientes al periodo prehispánico y colonial de México. Recuperamos 11 genomas parciales antiguos de *T. forsythia* y observamos una distribución filogenética distinta de las muestras, lo que sugiere que las cepas presentes en los individuos del periodo prehispánico llegaron al continente Americano junto con las primeras migraciones humanas y nuevas cepas fueron introducidas con la llegada de poblaciones europeas y africanas durante el siglo XVI. Asimismo, identificamos la presencia diferencial de genes de *T. forsythia* entre periodos, en el que ciertos genes se encontraban presentes en individuos del periodo prehispánico y ausentes del periodo colonial, y viceversa. Este estudio destaca el potencial de estudiar el genoma antiguo de *T. forsythia* para revelar interacciones sociales a través del análisis de la transmisión de la enfermedad. Nuestros resultados ilustran una relación antigua entre este patógeno oral y su hospedero humano, así como evidencia clave en el entendimiento de su historia evolutiva en México prehispánico y colonial.

Contribuciones de autoría

M.B.L. diseñó y realizó experimentos, analizó datos y realizó el borrador del manuscrito; V.V.I. Realizó experimentos y analizó datos; C.R.A. analizó datos; A.B.V.A. analizó datos; A.G.S. realizó experimentos; M.S.-V. diseñó experimentos; K.A., J.G.V, A.H., E.M., A.M.M., J.O., J.P.-P., K.S., J.K.W., G.Z. y A.LC. proporcionaron las muestras arqueológicas e información del contexto arqueológico; M.A.N.-C. diseñó experimentos; K.H.I. analizó datos; S.R. diseñó análisis de datos; M.C.Á.-A. concibió la investigación, diseño experimentos, analizó datos y realizó el borrador del manuscrito.

3.3. Reporte de proyecto. Recovery of an ancient *Salmonella enterica* genome from a Colonial individual from Mexico City.

Miriam Bravo-Lopez, Åshild Joanne Vågene, Carolina Rocha-Arriaga,Viridiana Villa-Islas, Maria de la Luz Moreno Cabrera, Alejandro Meraz Moreno, Jorge Gomez-Valdes, María C. Ávila-Arcos.

Resumen

A partir de la colonización europea en México, ocurrieron diversos brotes epidémicos que ocasionaron la muerte de millones de individuos indígenas. Con la finalidad de contribuir a la caracterización de la diversidad de patógenos humanos presentes durante el periodo post-contacto, analizamos dientes de individuos provenientes del periodo Colonial (1700-1900 CE) de la Ciudad de México. Reconstruimos el genoma (10X) de *Salmonella enterica* Paratyphi C, la bacteria responsable de la fiebre paratifoidea. Realizamos una comparación filogenética con cepas antiguas previamente reportadas, provenientes del sur de México y de Europa. Identificamos una cercanía filogenética con cepas antiguas de México previamente reportadas.

Contribuciones de autoría

M.B.-L. diseñó y realizó experimentos, analizó datos y realizó el borrador del manuscrito; A.J.-V. diseño análisis de datos y analizó datos; R.A.-C. analizó datos; V.V.-I.. Realizó experimentos y analizó datos; M.S.-V. diseñó experimentos; M.C.M.-L., M.M.-A. G.V.-J; proporcionaron los restos arqueológicos e información del contexto arqueológico;M.C.Á.-A. concibió la investigación, diseño experimentos, analizó datos y realizó el borrador del manuscrito.

3.4. Reporte de proyecto. Paleogenomic analysis of pathogenic bacteria in hunter-gatherers from central coastal Argentinian Patagonian.

Miriam Bravo-Lopez, Carolina Rocha-Arriaga, Camila Tamburrini, Viridiana Villa-Islas, Rosa Fregel, Silvia Dahinten, Julieta Gómez-Otero, Gabriela Millán, Carlos Bustamante, María C. Ávila- Arcos, María Laura Parolin.

Resumen

La costa norte de Patagonia Argentina ha sido habitada por cazadores-recolectores desde el Holoceno medio, de acuerdo con datación directa de huesos humanos. Las condiciones biogeográficas y ambientales de esta región, propiciaron abundancia y diversificación de la actividad de subsistencia, influyendo en el estado de salud de los cazadores-recolectores. En este estudio, analizamos muestras de dientes y huesos de cazadores-recolectores provenientes de la Costa Norte Central de la Patagonia (6000-300 AP) en búsqueda de patógenos antiguos. Llevamos a cabo la extracción de aDNA, construcción de librerías y secuenciación de tipo shotgun a baja profundidad, y las secuencias genéticas resultantes fueron clasificadas taxonómicamente por medio de una base de referencia de genomas completos de bacterias, arqueas y virus de NCBI RefSeq. Identificamos aDNA de la bacteria *Erysipelothrix rhusiopathiae,* un patógeno zoonótico que ocasiona lesiones en la piel llamada erisipeloide. Adicionalmente, la identificación de patógenos periodontopaticas como *Rothia dentocariosa, Eikenella corrodens, Streptococcus sanguinis, Streptococcus gordonii, Capnocytophaga gingivalis, Capnocytophaga ochracea, Filifactor alocis, Fusobacterium nucleatum; Capnocytophaga sputigena, Veillonella parvula, Aggregatibacter aphrophilus, Filifactor alocis,* y *Tannerella forsythia*; y junto con la evidencia paleopatológica, sugieren la presencia de enfermedad periodontal en estos individuos. El resultado obtenido a través del análisis de isótopos estables (13C y 15N), muestran una dependencia dietaria de fuentes marítimas y terrestres, que pudieron haber promovido la presencia de estos patógenos orales. Adicionalmente, identificamos aDNA de *Clostridium perfringens*, agente causante de gangrena e intoxicación por alimentos, y de otra bacteria *Clostridium tetani*, que causa la enfermedad del tétanos. La realización de análisis genómicos a mayor profundidad, contribuirá al entendimiento de cómo las estrategias de sobrevivencia influyeron en la salud de los cazadores-recolectores, y como el genoma de estos patógenos ha evolucionado en los últimos 6,000 años.

Contribuciones de autoría

M.B.-L. analizó datos y realizó el borrador del manuscrito; R.A.-C. analizó datos; P.L.-M concibió la investigación y realizó experimentos; T.-C. analizó datos; V.V.-I. Realizó experimentos y analizó datos; F.-R., G.O.-J,M.-G., M.M.-A. G.V.-J; proporcionaron las muestras arqueológicas, así como el contexto arqueológico de los individuos; M.C.Á.-A. Diseño de análisis de datos y realizó el borrador del manuscrito.

4. Artículo de revisión. What have we learned about past human infectious diseases using paleogenomics?

En preparación
Miriam Bravo-Lopez y María C. Ávila-Arcos

Abstract

DNA extracted from ancient human remains contains information beyond the host's genetic component; it also harbors DNA from pathogens present at the time of an individual's death. The recovery of ancient DNA from pathogens offers new insights into the etiology, nature and evolutionary history of infectious diseases that have been a part of our human history. The genomic characterization of ancient pathogens, along with historical and archeological evidence, can inform us of the origins and geographical spread of ancient diseases. In this review we present the major findings in ancient human pathogen genomics and the analytical and methodological approaches used to identify ancient pathogen DNA, as well as some relevant limitations for its retrieval.

4.1 Introduction

Infectious diseases have profoundly marked our human history. Traditionally, ethnohistorical descriptions and paleopathological evidence have been considered the main sources for understanding past human population health (Raoult and Drancourt, 2008). However, these sources have certain limitations that can be counteracted with the use of molecular approaches, as it is the retrieval and analysis of ancient DNA (aDNA).

The term of aDNA is used to describe the genetic material from paleontological, archaeological, historical or biological specimens. Such genetic material is often fragmented, damaged (e.g., deamination of cytosines), and is recovered in small quantities (Marciniak and Perry, 2017). The information revealed by aDNA coupled with Next Generation Sequencing (NGS) technologies has brought new insights into the longstanding co-evolutionary history of humans and pathogens (Spyrou *et al.*, 2019). These have also allowed the identification of an individual's disease without the diagnostic anatomical pathology (Harkins and Stone, 2015). In addition, ancient pathogen DNA can document past mixed-strain infections (Kay *et al.*, 2015) and those that are no longer extant at a higher level of resolution compared to other paleomicrobiological strategies. Here, we highlight the major findings in ancient pathogen recovery through paleogenomics, as well as the current methodological and analytical approaches for its study. Lastly, we discuss the challenges

and perspectives of the field, as well as the ethical considerations of working with archaeological material for the study of ancient pathogen DNA.

4.2. From amplicons to whole genomes in aDNA pathogen research

Over the last 30 years, the analytical and methodological approaches used for the recovery of aDNA have matured beyond single-locus studies and have yielded promising insights into the evolutionary past of a wide range of organisms across space and time (Cappellini *et al.*, 2018).

In its early beginnings in the 1980s, aDNA studies were mainly based on PCR, often combined with traditional Sanger sequencing. However, the biggest limitation of the Sanger method is the reduced amount of DNA that can be processed in any period of time (e.g., one sample produces one sequence) (Linderholm, 2016). Therefore, the recovery of aDNA was restricted mainly to mitochondrial DNA (mtDNA) fragments, as its high copy number per cell facilitates PCR amplification and its maternal inheritance without recombination simplifies reconstruction of the phylogenetic relationships among extinct and extant species (Orlando and Cooper, 2014; MacHugh, Larson and Orlando, 2017). Although some of the results from early aDNA studies were irreproducible (Pääbo *et al.*, 2004), these suggested that endogenous DNA, which is generally limited to very low levels of short and damaged fragments, could be retrieved from ancient specimens (Lan and Lindqvist, 2018).

The first molecular detection of an ancient pathogen may go back to the 1990s, when *Mycobacterium tuberculosis* DNA was detected from a vertebra dated to the Byzantine period—ie. ~1000 years old (Spigelman and Lemma, 1993), and from a 1000-year-old South American mummy (Salo *et al.*, 1994). This study paved the way to paleomicrobiology, a subfield where microbiology, evolution, anthropology, and history converge (Raoult and Drancourt, 2008). Several reports followed using the same methodological approach, including the recovery of *Plasmodium falciparum* aDNA from Egyptian mummy tissues (Nerlich *et al.*, 2008) and the detection of *Treponema pallidum subsp. pallidum* aDNA, the causative agent of venereal syphilis, from a 200-year-old skeletal specimen from Easter Island (Kolman *et al.*, 1999). Although, to our knowledge none of these studies have been replicated using NGS, therefore contamination should not be ruled out from these studies.

The insights into the evolutionary history of ancient pathogens gained from PCR-based studies were limited, in addition modern DNA contamination was often identified and generated big skepticism about these studies (Cano and Toranzos, 2018). This changed dramatically with the

introduction of NGS to the aDNA field, as it allowed the characterization of a large number of ancient genomes, not only from geographic regions with favorable climatic conditions for DNA preservation, but also from sites all over the world (Achilli *et al.*, 2018).

Despite NGS technologies having transformed aDNA research, they are not free of problems. For example, sequencing error rates (~0.1-15%) are higher than traditional Sanger sequencing, and the damaged nucleotides in aDNA further inflate these error rates (Lan and Lindqvist, 2018). Therefore, in order to take full advantage of NGS technologies in aDNA research, comprehensive studies seek to improve the efficiency of the key experimental steps, such as DNA extraction (Damgaard *et al.*, 2015; Sirak *et al.*, 2017, 2020), library construction (Sandoval-Velasco *et al.*, 2017; Gansauge and Meyer, 2019), and hybridization enrichment (Cruz-Dávalos *et al.*, 2018; Furtwängler *et al.*, 2020). Further improvements in the bioinformatic tools used of its analysis have also contributed to the expansion of the field. So far, the number of available human genomes recovered from ancient remains is estimated to be at least 3,000 (Racimo *et al.*, 2020), and around 264 for ancient microbial genomes (Fellows Yates *et al.*, 2020).

In 2011, the first fully reconstructed ancient bacterial genome using NGS was reported; it belonged to *Yersinia pestis,* the causative agent of the Black Death (Bos *et al.*, 2011). Further refinements in experimental protocols together with strict criteria for determining the authenticity of data, have allowed the genome reconstruction of additional ancient bacterial pathogens such as *Brucella melitensis (Kay et al., 2014)*, *Tannerella forsythia* (Warinner *et al.*, 2014; Bravo-Lopez *et al.*, 2020; Philips *et al.*, 2020), *Helicobacter pylori (Maixner et al., 2016), Mycobacaterium leprae (Schuenemann, Avanzi, et al., 2018; Blevins et al., 2020; Fotakis et al., 2020)*, *Mycobacaterium tuberculosis (Bos et al., 2014)*, *Salmonella enterica (Vågene et al., 2018; Zhou et al., 2018; Key et al., 2020)* and *Treponema pallidum ssp. pallidum* and *pertenue (Schuenemann, Kumar Lankapalli, et al., 2018)*.

Moreover, the identification of aDNA from viral pathogens has also been reported; including Variola virus (VARV) (Duggan *et al.*, 2016; Pajer *et al.*, 2017), Hepatitis B virus (Mühlemann, Jones, *et al.*, 2018) and Human Parvovirus B19 (Mühlemann, Margaryan, *et al.*, 2018; Guzmán-Solís *et al.*, 2020). Of notice, the majority of ancient genomes reconstructed so far has a strong bias towards pathogens from European samples; only a handful of pathogens from the Americas.

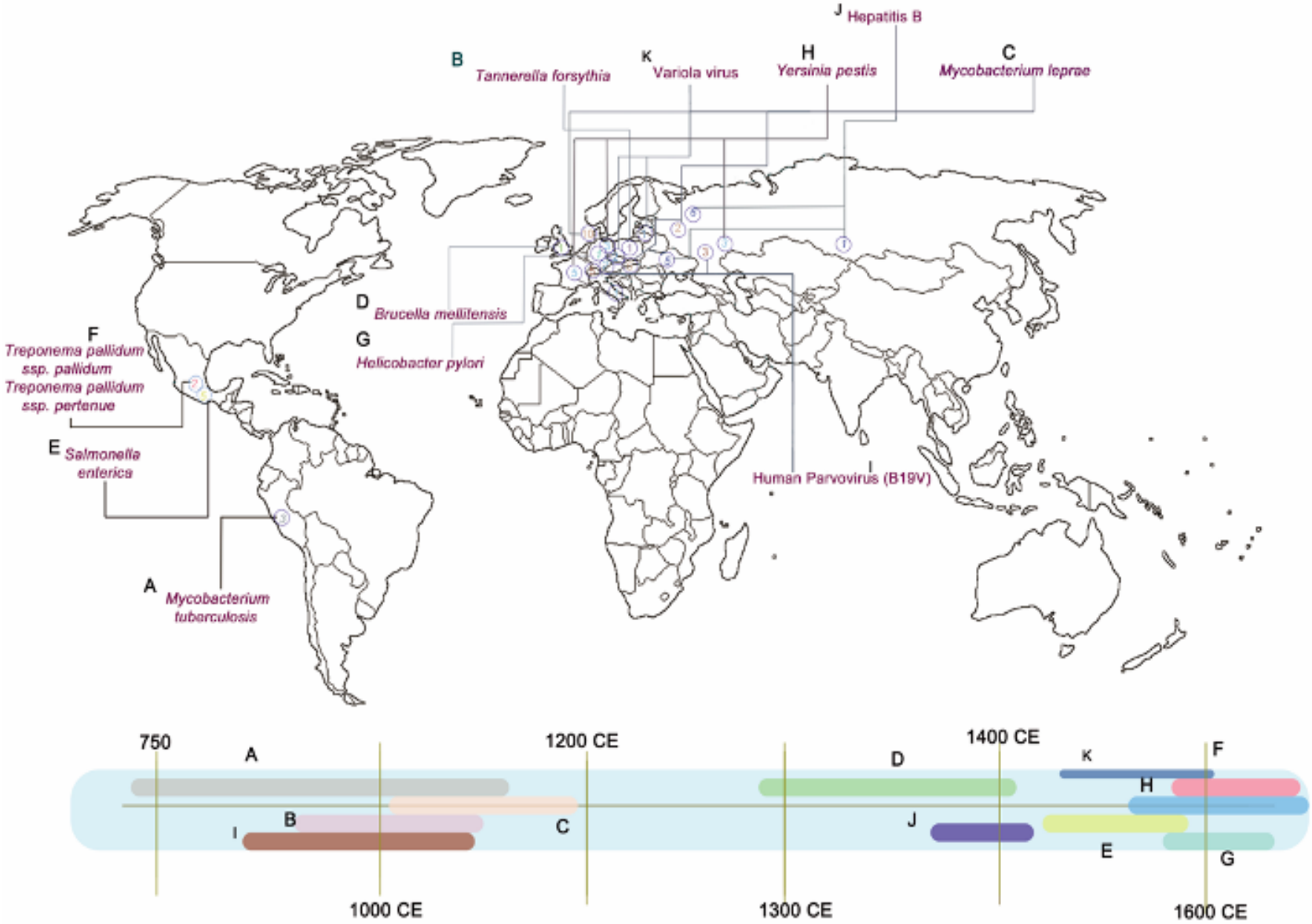


Figure 1. Geographic locations and temporal periods of pathogen genomes sequenced to date. The number of

has been studied through NGS (Schuenemann, Kumar Lankapalli, *et al.*, 2018; Vågene *et al.*, 2018; Barquera *et al.*, 2020; Bravo-Lopez *et al.*, 2020; Guzmán-Solís *et al.*, 2020) (Figure 1).

4.3. Overview of ancient pathogen paleogenomics studies

The knowledge on pathogen evolution gained from studying the genomes of ancient pathogens has complemented previous insights gained from historical and archeological sources (Andam *et al.*, 2016). As the number of ancient pathogen genomes rapidly increases (Arning and Wilson, 2020), so does the understanding of the molecular mechanisms underlie the causes of ancient epidemics of historical relevance.

Perhaps the most widely studied ancient pandemic is the Black Death, which took place in Europe and Asia during the mid-1300s and was caused by the bacterial pathogen *Yersinia pestis*. The genome of this highly virulent bacteria has been broadly studied from a paleogenomic approach, since it left a profound mark in human history (Bos *et al.*, 2011; Schuenemann *et al.*, 2011; Wagner *et al.*, 2014; Rasmussen *et al.*, 2015; Feldman *et al.*, 2016; Kirsten I Bos *et al.*, 2016; Andrades Valtueña *et al.*, 2017; Damgaard *et al.*, 2018; Namouchi *et al.*, 2018; Spyrou *et al.*, 2018, 2019; Keller *et al.*, 2019). Through genome sequencing of ancient *Yersinia pestis* strains from infected humans in the Bronze Age Eurasia (~5800 years ago) it was determined that the acquisition of the *Yersinia* murine toxin gene (*ymt*) played a crucial role in the epidemiological history of plague, as it allowed the survival of *Yersinia pestis* in the flea vector, giving rise to the large-scale pandemics observed in the past millennium (Spyrou *et al.*, 2018). The study of ancient *Y. pestis* has certainly been a driver for the development of the ancient pathogen genomics field and it has been thoroughly reviewed elsewhere (Namouchi *et al.*, 2018; Spyrou *et al.*, 2019; Morozova *et al.*, 2020; McCormick, 2021), however, less attention has been given to studies involving other ancient pathogens from different contexts. We describe below how some of these studies have revealed outstanding facts about past infectious diseases elsewhere.

The retrieval of ancient pathogen genomes has allowed researchers to address long- standing questions about the causative agents of past epidemics outside Europe. This is the case of the 16th century outbreak, locally known as *"Cocoliztli"*, which dramatically affected the Native population of New Spain (present-day Mexico). *"Cocoliztli"* was perhaps the deadliest outbreak experienced upon colonization of the Americas.The exact cause (or causes) of the outbreak has remained elusive for the past 500 years (Guevara Flores, 2017). A recent study, however, was able to reconstruct the genome of *Salmonella enterica Paratyphi C* from indigenous individuals

buried in a cemetery dated to the early contact era in Teposcolula-Yucundaa, Oaxaca, in southern Mexico. Based on historical and archaeological evidence, this cemetery was linked to the 1545–1550 CE "*Cocoliztli"* epidemic that affected large parts of Mexico. These observations led the authors to suggest *Salmonella enterica* to be one of the causative agents of this deathly outbreak (Vågene *et al.*, 2018). Nonetheless, more samples from more archeological sites linked to cocoliztli outbreaks would be needed to unequivocally assert this.

Interestingly, a recent study also recovered the genome of *Salmonella Paratyphi C* from a 800-year-old skeleton of a young woman excavated from a church cemetery in Trondheim, Norway (Zhou *et al.*, 2018). This study suggested that the young Norwegian died of an infection with *Salmonella Paratyphi C,* which has long caused enteric fever across northern Europe. Intriguingly, *Salmonella Paratyphi C* is less common in North America and Europe compared to South and East Asia or Africa (Achtman *et al.*, 2012; Zhou *et al.*, 2018). These observations clearly open up new questions regarding how and from where this pathogen entered the Americas.

Likewise, the genetic analysis of skeletons dated back to Colonial Mexico with characteristic skeletal manifestations for congenital treponematosis allowed the identification of aDNA from *T. pallidum ssp. Pertenue*, which causes yaws, in an infant individual estimated to be six months of age at time of death (Schuenemann, Kumar Lankapalli, *et al.*, 2018). Moreover, this study put forward the possible existence of congenital yaws, which contradicts previous assumptions of the inability of this strain to cross the placental barrier (Giacani and Lukehart, 2014). Consequently, this study opened a new hypothesis to test regarding the etiology of the disease.

Other revealing results include those obtained from the analysis of genomes of less studied pathogens, like *Helicobacter pylori,* which has been associated with the development of peptic ulcers and stomach cancer (Kabamba, Tuan and Yamaoka, 2018). An ancient genome of this pathogenic bacteria was reconstructed from the preserved stomach content of a 5,300-year-old European natural ice mummy. This finding proposed the existence of a nearly-pure representative of the bacterial population of Asian origin in Europe prior to hybridization with an African strain (Maixner *et al.*, 2016). Another example is *Brucella melitensis*, identified by Kay and colleagues in 2014. The recovery of the DNA of this medieval strain of *B. melitensis* revealed the continuous presence of this zoonotic disease in the same region over centuries (Kay *et al.*, 2014).

4.4. Evolutionary insights of paleoviruses genome reconstruction

In addition to characterizing ancient pathogenic bacterial genomes, the field has shown enormous potential in the study of ancient pathogenic viruses (Tsangaras and Greenwood, 2018). To date, only a handful of genomes from ancient viruses —all of them with single or double-stranded DNA—has been recovered. These include VARV (variola virus) sampled from a Lithuanian child mummy dating between 1643 and 1665 AD (Duggan *et al.*, 2016) and from human tissues kept in a museum for 160 and 60 years (Pajer *et al.*, 2017), HBV (Hepatitis B Virus) from 800 and 4500 years old remains (Kahila Bar-Gal *et al.*, 2012; Patterson Ross *et al.*, 2018); from a 16th century African individual (Barquera *et al.*, 2020), and human parvovirus B19 from 500 to 6,000-year-old samples (Mühlemann, Margaryan, *et al.*, 2018); from 16th century Africans individuals (Guzmán-Solís *et al.*, 2020).

Regarding VARV, the causing agent of smallpox, a number of recent studies have successfully reported the recovery of aDNA from this virus. Duggan *et al.* captured, sequenced, and reconstructed a draft genome of an ancient strain of VARV, sampled from a Lithuanian child mummy dating between 1643 and 1665 AD. Phylogenetic analysis of the ancient strain, placed it basal to all previously sequenced orthopoxvirus strains with the closest being the ancient 300-year-old Siberian VARV sequence previously reported by Biagini and colleagues (Biagini, Theves and Balaresque, 2012). This study provided a new epidemiological calibration point for smallpox and the variola virus lineages responsible for historical smallpox outbreaks, which have been in circulation for approximately 200 years (Duggan *et al.*, 2016). Additionally, Pajer and colleagues reconstructed two complete variola virus genomes from human tissues kept in a museum in Prague, dating back 60 and 160 years. The 60-year-old variola virus genome likely represents a new endemic European variant of variola virus circulating in the middle of the 19th century in Europe (Pajer *et al.*, 2017).

In regards to HBV, three studies to date have reconstructed successfully HBV virus genomes from ancient remains. The first complete genome sequence of an HBV was from the mummified remains of a child discovered in the 16th century from Naples, Italy (Patterson Ross *et al.*, 2018). The analysis of the genome showed that HBV evolution over the last ~450 years was characterized by a marked lack of temporal structure that hinders efforts to resolve the evolutionary time-scale of this human pathogen (Patterson Ross *et al.*, 2018). Additionally, Mühlemann and colleagues also reconstructed twelve ancient VARV genomes from Central to Eestern Eurasia ranging from 800 to 4,500 years old (Mühlemann *et al.*, 2020). The geographical

and temporal patterns observed in this study in ancient and modern HBV genotypes are compatible with well-documented human migrations during the Bronze and Iron Ages.

Another virus for which ancient genomes have been reconstructed is human parvovirus B19 (B19V). This virus causes a number of illnesses, including the childhood rash known as “fifth disease”. Ancient DNA from B19V was recovered from teeth dating back to 500- and 6,900-years old belonging to individuals from Eurasia, Southeast Asia and Greenland (Mühlemann et al., 2018). One of the most notable findings from this study was that this virus, once assumed to have a rapid rate of evolution comparable to that of RNA viruses, actually evolves at a rate approximately an order of magnitude lower than previously believed. In addition, B19V has been identified in three individuals from the 16th century found in Mexico City, interestingly, the origin of two of them were on the African continent (Guzmán-Solís *et al.*, 2020).

Virus paleogenomics is a young and promising field, though some important challenges prevail. Many viruses of historical relevance have RNA genomes (e.g., influenza and HIV) and are generally unstable outside of a living host. In consequence, these studies will likely remain limited to DNA viruses, and only under specific conditions of preservation. Advances in genomic enrichment techniques and sequencing technologies offer great potential to expand this challenging field (Tsangaras and Greenwood, 2018). As further research continues on the pathogens involved in major historical infectious disease outbreaks, additional biosafety considerations, ethical discussions, and potential risk and benefit analysis should be warranted since the reconstruction of ancient viruses may pose a biological threat to public health.

4.5. Substrates for the recovery of ancient pathogen DNA

The genetic material extracted from ancient remains is a mixture of DNA from microorganisms that colonize the sample during the burial, as well as other contaminants occurring during excavation and processing, often representing > 90% of the total DNA (Arning and Wilson, 2020). The pathogen DNA fraction often constitutes less than 1% of the total metagenomic background (Devault *et al.*, 2014; Duchêne *et al.*, 2020). In consequence, retrieving viable aDNA from skeletal tissues (e.g., bone or teeth) or from secondary substrates (dental calculus and coprolites) is challenged by temporal and environmental factors that compromise the molecular integrity of DNA (Mann *et al.*, 2018).

The archaeological material available for aDNA studies is valuable and often highly limited; hence, it is crucial to carefully select the appropriate source material for the retrieval of ancient pathogen DNA. The rationale for selecting a sample for this type of study includes, but it is not limited to; (i) a tissue (bone or mummified tissue) that displays physical changes consistent with particular infections; or (ii) tissues with a direct contact with the circulatory system; or (ii) samples where the historical or archeological context links them to a past epidemic or death by infection (Harkins and Stone, 2015).

In the majority of microbial aDNA studies, postcranial bones or teeth are the first-choice substrates for aDNA extraction, possibly because these have an abundant network of blood vessels/circulation and become directly exposed to blood-borne pathogens (Raoult and Drancourt, 2008). Additionally, most cells in teeth are found within the dental pulp—the inner part of a tooth made up of living connective tissue; pathogen DNA can easily reach this tissue by direct invasion into blood vessels, via lymphatic vessels draining a focus of infection (ie, abscess) (Nguyen-Hieu, Aboudharam and Drancourt, 2011).

Several investigations that have evaluated tissue-specific aDNA preservation across the human skeleton, consistently point to the inner ear of the petrous bone as being the richest and most reliable source of endogenous molecules (Pinhasi *et al.*, 2015). However, Margaryan et al., (2018) compared the potential for ancient pathogen DNA retrieval from teeth and petrous bone, and found limited potential for the analyses of blood-borne ancient pathogen DNA from the latter, due to having lower blood circulation levels compared to other parts of the human skeleton (Margaryan *et al.*, 2018).

Calcified dental plaque is another rich source of well-preserved endogenous aDNA from both the host and the oral microbiome (Warinner *et al.*, 2014). Studies reporting the genome reconstruction of ancient pathogen genomes from dental plaque, include those on *Tannerella forsythia* (Warinner *et al.*, 2014; Bravo-Lopez *et al.*, 2020), *Neisseria meningitidis* (Eerkens *et al.*, 2018), and *Mycobacterium leprae* (Fotakis *et al.*, 2020).

On the other hand, there are a few studies that have made use of nonhuman substrates for aDNA pathogen recovery. One of such studies involves the reconstruction of variola virus (VARV) DNA from bodily fluid material and medical instruments from a vaccination kit (Duggan *et al.*, 2020). Additionally, the recovery of aDNA from a 5,700-year-old chewing gum made from birch pitch

found in Denmark, enabled the characterization of the oral, as well as the reconstruction of the genome of bacterial pathogen *Streptococcus pneumoniae* (Jensen *et al.*, 2019).

4.6. Ancient pathogen genomes reconstruction through hybridization-enrichment methods

The ancient pathogen DNA fraction approximately represents 0.5% (or less) of the total metagenome in archeological specimens (Carpenter *et al.*, 2013; Spyrou *et al.*, 2019), although there are some exceptions like in the case of calcified nodules, where the endogenous bacterial DNA can be up to 66% (Devault *et al.*, 2017). Hence, a groundbreaking approach in ancient pathogen genomics research has been the implementation and the improvement of hybridization capture strategies (Ávila-Arcos *et al.*, 2015; Bossert and Danforth, 2018)). This technique seeks to increase the target DNA compared to contaminant DNA and thus can lower the cost of sequencing whole genomes from degraded samples.

The capture-enrichment of ancient pathogen DNA usually follows one of three main approaches: i) DNA-microarrays, ii) in-solution capture with DNA baits, and iii) in-solution capture with RNA baits. These methods are based on the hybridization of homologous fragments from the target species genomes to custom DNA or RNA single-stranded biotinylated oligonucleotides (referred as 'baits'). The baits are usually 80-300 nucleotides long and its design and synthesis have been facilitated by different commercial companies such as Agilent, Arbor Biosciences and NimbleGen (Gaudin and Desnues, 2018; Furtwängler *et al.*, 2020). These baits can be immobilized (in microarrays) or in solution. The hybrids are retrieved in the solid phase of microarrays or through streptavidin-coated magnetic beads for the in-solution approach, while unhybridized molecules (e.g., off-target DNA) are washed away.

The use of DNA array-based methods in aDNA has allowed the recovery of a *Mycobacterium leprae* genome from five medieval individuals from the United Kingdom, Sweden and Denmark (Schuenemann *et al.*, 2013). Furthermore, Bos *et al.,* (2014) designed a DNA array-based method to screen in parallel 92 different pathogens in ancient samples, recovering *Mycobacterium leprae* from a medieval sample from the United Kingdom, previously analyzed in Schuenemann *et al.,* (2013). This approach has also proven useful for the DNA recovery of other pathogens like *Yersinia pestis* (Bos *et al.*, 2011), *Mycobacterium tuberculosis* (Bos *et al.*, 2014), S*almonella enterica* (Vågene *et al.*, 2018), and *Vibrio cholerae* (Devault *et al.*, 2014).

The in solution-capture enrichment approach, either with DNA or RNA probes has allowed the characterization of genome-wide data of several ancient pathogens such as *Helicobacter pylori (Maixner et al., 2016)*, *Mycobacterium leprae* (Krause-Kyora, Nutsua, *et al.*, 2018), among many others (Bos *et al.*, 2014; Duggan *et al.*, 2016; Mühlemann, Jones, *et al.*, 2018; Mühlemann, Margaryan, *et al.*, 2018; Schuenemann, Kumar Lankapalli, *et al.*, 2018; Vågene *et al.*, 2018; Barquera *et al.*, 2020; Bravo-Lopez *et al.*, 2020; Guzmán-Solís *et al.*, 2020).

Furtwängler *et al.,* 2020, compared array-based hybridization capture and in-solution capture using either RNA or DNA baits in ancient and modern *Mycobacterium leprae* and *Treponema pallidum samples*. The RNA bait in-solution strategy showed a higher efficiency, reproducibility and specificity in ancient and modern samples. Additionally, this study evaluated the impact of one and two rounds of hybridization on the capture efficiency. A second round of hybridization coupled with capture-enrichment, resulted in an increase of the enrichment factor, at the cost of loss of library complexity. The implementation of a second round of hybridization should be avoided if the starting endogenous DNA content is high, a single round round of capture coupled with deeper sequencing would be more efficient (Furtwängler *et al.*, 2020).

The capture-enrichment of aDNA pathogen has contributed to the growing pace of the ancient pathogen genomics field. Nevertheless, the impact of the starting libraries characteristics—the endogenous content, libraries complexity, and read lengths—over the efficiency of the capture-enrichment of ancient pathogens remains to be explored further.

4.7. Perspectives in ancient pathogen genomics

Only five years ago it was difficult to anticipate the potential of paleogenomics to better understand the origins, evolution, and molecular basis of pathogenicity of bacterial and viral agents responsible for major historical pandemics in human history (Spyrou *et al.*, 2019).

The field of ancient pathogen genomics has faced the challenge of the recovery of pathogenic microorganisms that are characterized by an unstable nature of its genome, e.g., viruses, in particular single-stranded DNA (ssDNA) and RNA viruses (Enard and Petrov, 2020). In the case of single-stranded DNA, Mühlemann *et al.,* (2018) recovered Parvovirus B19 from Eurasia and Greenland, between ~0.5 and 6.9 thousand years old. In addition, Guzmán-Solís et al., (2020) reconstructed three Parvovirus B19 genomes from tooth samples excavated from Mexico City

dated to 1700-1900 CE. The reason why these could be recovered is that during replication, the B19V genome goes through a double-stranded DNA (dsDNA) phase, which are likely the fragments recovered with the library building method (which takes double stranded DNA as the template). Other viruses with ssDNA genomes having intermediate dsDNA during replication, e.g., *Anelloviridae* viruses represent good candidates for the recovery of their aDNA.

As for the RNA viruses, Düx *et al.,* (2018), reconstructed the viral genome of *Measles morbillivirus,* which is the causative agent of measles, from a lung museum specimen dated to 1912. Another example is the recovery of Human immunodeficiency virus 1 genome from serum samples dating back to 1978–1979 (Worobey *et al.*, 2016). In contrast, an increased number of DNA viral genomes have been recovered from archaeological remains (Biagini P, Theves C. and Balaresque P., 2012; Duggan *et al.*, 2016; Krause-Kyora, Susat, *et al.*, 2018; Patterson Ross *et al.*, 2018; Barquera *et al.*, 2020; Guzmán-Solís *et al.*, 2020). The recovery of ancient RNA from a number of sources including a 14,300-year-old Pleistocene canid preserved in permafrost (Smith *et al.*, 2019) opens a promising window of opportunity for the eventual retrieval of ancient viral RNA. Further methodological and analytical strategies are needed to increase the resolution power on the recovery of ancient viruses (Ferrari *et al.*, 2020).

An additional analytical strategy that can be improved in the field of ancient pathogen genomics, is capture-enrichment. Furtwängler *et al.,* 2020 compared the efficiency, specificity and reproducibility of in-solution (using DNA and RNA baits) and DNA-array based capture methods for *M. leprae* and *T. pallidum.* The in-solution approach, using the RNA bait capture, showed the best performance compared to the other methods. Nevertheless, the efficiency of these methods remains to be evaluated in different sources and for other taxa (e.g., viruses).

Lastly, the 95% of the ancient pathogen genomes reconstructed so far come from European individuals, and this could be mainly explained by the funding facilities of the research groups from this geographic area. However, the main consequences of this bias, is the limitation of the local capacity building and a valuable knowledge exchange. The establishment of collaborative projects with local scientists, may stimulate further discussions towards this newly established sub-field of paleogenomics.

In conclusion, methodological improvements for the isolation, sequencing, and authentication of ancient pathogen genomes will continue to further the field. While such advances have also

facilitated the potential for sequencing ancient genomes, paleogenomics should not be considered as a panacea**.** Archeological and historical approaches as well as modern microbiology are central for contextualizing disease in past human populations.

4.8. References

Achilli, A., Olivieri, A., Semino, O., & Torroni, A. (2018). Ancient human genomes—keys to understanding our past. *Science*, *360*(6392), 964–965.

Achtman, M., Wain, J., Weill, F.-X., Nair, S., Zhou, Z., Sangal, V., Krauland, M. G., Hale, J. L., Harbottle, H., Uesbeck, A., Dougan, G., Harrison, L. H., Brisse, S., & S. Enterica MLST Study Group. (2012). Multilocus sequence typing as a replacement for serotyping in Salmonella enterica. *PLoS Pathogens*, *8*(6), e1002776.

Andam, C. P., Worby, C. J., Chang, Q., & Campana, M. G. (2016). Microbial Genomics of Ancient Plagues and Outbreaks. *Trends in microbiology*, *24*(12), 978–990.

Andrades Valtueña, A., Mittnik, A., Key, F. M., Haak, W., Allmäe, R., Belinskij, A., Daubaras, M., Feldman, M., Jankauskas, R., Janković, I., Massy, K., Novak, M., Pfrengle, S., Reinhold, S., Šlaus, M., Spyrou, M. A., Szécsényi-Nagy, A., Tõrv, M., Hansen, S., … Krause, J. (2017). The Stone Age Plague and Its Persistence in Eurasia. *Current biology: CB*, *27*(23), 3683–3691.e8.

Arning, N., & Wilson, D. J. (2020). The past, present and future of ancient bacterial DNA. *Microbial Genomics*, *6*(7). https://doi.org/10.1099/mgen.0.000384

Ávila-Arcos, M. C., Sandoval-Velasco, M., Schroeder, H., Carpenter, M. L., Malaspinas, A.-S., Wales, N., Peñaloza, F., Bustamante, C. D., & Gilbert, M. T. P. (2015). Comparative performance of two whole-genome capture methodologies on ancient DNA Illumina libraries. *Methods in ecology and evolution / British Ecological Society*, *6*(6), 725–734.

Barquera, R., Lamnidis, T. C., Lankapalli, A. K., Kocher, A., Hernández-Zaragoza, D. I., Nelson, E. A., Zamora-Herrera, A. C., Ramallo, P., Bernal-Felipe, N., Immel, A., Bos, K., Acuña-Alonzo, V., Barbieri, C., Roberts, P., Herbig, A., Kühnert, D., Márquez-Morfín, L., & Krause, J. (2020). Origin and Health Status of First-Generation Africans from Early Colonial Mexico. *Current Biology: CB*. https://doi.org/10.1016/j.cub.2020.04.002

Biagini P, Theves C., & Balaresque P. (2012). Variola virus in a 300-year-old Siberian mummy. *The New England journal of medicine*, *367*, 2057–2059.

Blevins, K. E., Crane, A. E., Lum, C., Furuta, K., Fox, K., & Stone, A. C. (2020). Evolutionary history of Mycobacterium leprae in the Pacific Islands. En *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 375, Número 1812, p. 20190582). https://doi.org/10.1098/rstb.2019.0582

Bos, K. I., Harkins, K. M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S. A., Bryant, J. M., Harris, S. R., Schuenemann, V. J., Campbell, T. J., Majander, K., Wilbur, A. K., Guichon, R. A., Wolfe Steadman, D. L., Cook, D. C., Niemann, S., Behr, M. A., Zumarraga, M., … Krause, J. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature*, *514*(7523), 494–497.

Bos, K. I., Schuenemann, V. J., Golding, G. B., Burbano, H. A., Waglechner, N., Coombes, B. K., McPhee, J. B., DeWitte, S. N., Meyer, M., Schmedes, S., Wood, J., Earn, D. J. D., Herring, D. A., Bauer, P., Poinar, H. N., & Krause, J. (2011). A draft genome of Yersinia pestis from victims of the Black Death. *Nature*, *478*(7370), 506–510.

Bossert, S., & Danforth, B. N. (2018). On the universality of target-enrichment baits for phylogenomic research. *Methods in ecology and evolution / British Ecological Society*, *0*(0). https://doi.org/10.1111/2041-210X.12988

Bravo-Lopez, M., Villa-Islas, V., Arriaga, C. R., Villaseñor-Altamirano, A. B., Guzmán-Solís, A., Sandoval-Velasco, M., Wesp, J. K., Alcantara, K., López-Corral, A., Gómez-Valdés, J., Mejía, E., Herrera, A., Meraz-Moreno, A., de la Luz Moreno-Cabrera, M., Moreno-Estrada, A., Nieves-Colón, M. A., Olvera, J., Pérez-Pérez, J., Iversen, K. H., … Ávila-Arcos, M. C. (2020). Paleogenomic insights into the red complex bacteria Tannerella forsythia in Pre-Hispanic and Colonial individuals from Mexico. En *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 375, Número 1812, p. 20190580). https://doi.org/10.1098/rstb.2019.0580

Cano, R. J., & Toranzos, G. A. (Eds.). (2018). Paleomicrobiology: a Snapshot of Ancient Microbes and Approaches to Forensic Microbiology. En *Environmental Microbial Forensics* (pp. 63–90). American Society of Microbiology.

Cappellini, E., Prohaska, A., Racimo, F., Welker, F., Pedersen, M. W., Allentoft, M. E., Damgaard, P. de B., Gutenbrunner, P., Dunne, J., Hammann, S., Roffet-Salque, M., Ilardo, M., Moreno-Mayar, J. V., Wang, Y., Sikora, M., Vinner, L., Cox, J., Evershed, R. P., & Willerslev, E. (2018). Ancient Biomolecules and Evolutionary Inference. *Annual review of biochemistry*, *87*(1), null.

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., Leshtakov, K., Dimitrova, D., Theodossiev, N., Pettener, D., Luiselli, D., Sandoval, K., Moreno-Estrada, A., Li, Y., Wang, J., … Bustamante, C. D. (2013). Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *American journal of human genetics*, *93*(5), 852–864.

Cruz-Dávalos, D. I., Nieves-Colón, M. A., Sockell, A., Poznik, G. D., Schroeder, H., Stone, A. C., Bustamante, C. D., Malaspinas, A.-S., & Ávila-Arcos, M. C. (2018). In-solution Y-chromosome capture-enrichment on ancient DNA libraries. *BMC genomics*, *19*(1), 608.

Damgaard, P. B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., & Allentoft, M. E. (2015). Improving access to endogenous DNA in ancient bones and teeth. *Scientific reports*, *5*(1). https://doi.org/10.1038/srep11184

Damgaard, P. de B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J. V., Pedersen, M. W., Goldberg, A., Usmanova, E., Baimukhanov, N., Loman, V., Hedeager, L., Pedersen, A. G., Nielsen, K., Afanasiev, G., Akmatov, K., Aldashev, A., Alpaslan, A., … Willerslev, E. (2018). 137 ancient human genomes from across the Eurasian steppes. *Nature*, *557*(7705), 369–374.

Devault, A. M., McLoughlin, K., Jaing, C., Gardner, S., Porter, T. M., Enk, J. M., Thissen, J., Allen, J., Borucki, M., DeWitte, S. N., Dhody, A. N., & Poinar, H. N. (2014). Ancient pathogen DNA in archaeological samples detected with a Microbial Detection Array. *Scientific Reports*, *4*, 4245.

Devault, A. M., Mortimer, T. D., Kitchen, A., Kiesewetter, H., Enk, J. M., Golding, G. B., Southon, J., Kuch, M., Duggan, A. T., Aylward, W., Gardner, S. N., Allen, J. E., King, A. M., Wright, G., Kuroda, M., Kato, K., Briggs, D. E., Fornaciari, G., Holmes, E. C., … Pepperell, C. S. (2017). A molecular portrait of maternal sepsis from Byzantine Troy. *eLife*, *6*. https://doi.org/10.7554/eLife.20983

Duchêne, S., Ho, S. Y. W., Carmichael, A. G., Holmes, E. C., & Poinar, H. (2020). The Recovery, Interpretation and Use of Ancient Pathogen Genomes. En *Current Biology* (Vol. 30, Número 19, pp. R1215–R1231). https://doi.org/10.1016/j.cub.2020.08.081

Duggan, A. T., Klunk, J., Porter, A. F., Dhody, A. N., Hicks, R., Smith, G. L., Humphreys, M., McCollum, A. M., Davidson, W. B., Wilkins, K., Li, Y., Burke, A., Polasky, H., Flanders, L., Poinar, D., Raphenya, A. R., Lau, T. T. Y., Alcock, B., McArthur, A. G., … Poinar, H. N. (2020). The origins and genomic diversity of American Civil War Era smallpox vaccine strains. *Genome Biology*, *21*(1), 175.

Duggan, A. T., Perdomo, M. F., Piombino-Mascali, D., Marciniak, S., Poinar, D., Emery, M. V., Buchmann, J. P., Duchêne, S., Jankauskas, R., Humphreys, M., Golding, G. B., Southon, J., Devault, A., Rouillard, J.-M., Sahl, J. W., Dutour, O., Hedman, K., Sajantila, A., Smith, G. L., …

Poinar, H. N. (2016). 17 th Century Variola Virus Reveals the Recent History of Smallpox. *Current biology: CB*, *26*(24), 3407–3412.
Eerkens, J. W., Nichols, R. V., Murray, G. G. R., Perez, K., Murga, E., Kaijankoski, P., Rosenthal, J. S., Engbring, L., & Shapiro, B. (2018). A probable prehistoric case of meningococcal disease from San Francisco Bay: Next generation sequencing of Neisseria meningitidis from dental calculus and osteological evidence. *International Journal of Paleopathology*. https://doi.org/10.1016/j.ijpp.2018.05.001
Enard, D., & Petrov, D. A. (2020). Ancient RNA virus epidemics through the lens of recent adaptation in human genomes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1812), 20190575.
Feldman, M., Harbeck, M., Keller, M., Spyrou, M. A., Rott, A., Trautmann, B., Scholz, H. C., Päffgen, B., Peters, J., McCormick, M., Bos, K., Herbig, A., & Krause, J. (2016). A High-Coverage Yersinia pestis Genome from a Sixth-Century Justinianic Plague Victim. *Molecular Biology and Evolution*, *33*(11), 2911–2923.
Fellows Yates, J. A., Valtueña, A. A., Vågene, A. J., Cribdon, B., Velsko, I. M., Borry, M., Bravo-López, M. J., Fernandez-Guerra, A., Green, E. J., Ramachandran, S. L., Heintzman, P. D., Spyrou, M. A., Hübner, A., Gancz, A. S., Hider, J., Allshouse, A. F., & Warinner, C. (2020). *Community-curated and standardised metadata of published ancient metagenomic samples with AncientMetagenomeDir* (p. 2020.09.02.279570). https://doi.org/10.1101/2020.09.02.279570
Ferrari, G., Neukamm, J., Baalsrud, H. T., Breidenstein, A. M., Ravinet, M., Phillips, C., Rühli, F., Bouwman, A., & Schuenemann, V. J. (2020). Variola virus genome sequenced from an eighteenth-century museum specimen supports the recent origin of smallpox. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1812), 20190572.
Fotakis, A. K., Denham, S. D., Mackie, M., Orbegozo, M. I., Mylopotamitaki, D., Gopalakrishnan, S., Sicheritz-Pontén, T., Olsen, J. V., Cappellini, E., Zhang, G., Christophersen, A., Gilbert, M. T. P., & Vågene, Å. J. (2020). Multi-omic detection of Mycobacterium leprae in archaeological human dental calculus. En *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 375, Número 1812, p. 20190584). https://doi.org/10.1098/rstb.2019.0584
Furtwängler, A., Neukamm, J., Böhme, L., Reiter, E., Vollstedt, M., Arora, N., Singh, P., Cole, S. T., Knauf, S., Calvignac-Spencer, S., Krause-Kyora, B., Krause, J., Schuenemann, V. J., & Herbig, A. (2020). *Comparison of target enrichment strategies for ancient pathogen DNA* (p. 2020.07.09.195065). https://doi.org/10.1101/2020.07.09.195065
Gansauge, M.-T., & Meyer, M. (2019). A Method for Single-Stranded Ancient DNA Library Preparation. *Methods in Molecular Biology* , *1963*, 75–83.
Gaudin, M., & Desnues, C. (2018). Hybrid Capture-Based Next Generation Sequencing and Its Application to Human Infectious Diseases. En *Frontiers in Microbiology* (Vol. 9). https://doi.org/10.3389/fmicb.2018.02924
Giacani, L., & Lukehart, S. A. (2014). The Endemic Treponematoses. *Clinical microbiology reviews*, *27*(1), 89–115.
Guevara Flores, S. E. (2017). *La construcción sociocultural del cocoliztli en la epidemia de 1545 a 1548 en la Nueva España*.
Guzmán-Solís, A. A., Blanco-Melo, D., Villa-Islas, V., Bravo-López, M. J., Sandoval-Velasco, M., Wesp, J. K., Gómez-Valdés, J. A., de la Luz Moreno-Cabrera, M., Meraz-Moreno, A., Solís-Pichardo, G., Schaaf, P., tenOever, B. R., & Ávila-Arcos, M. C. (2020). *Ancient viral genomes reveal introduction of HBV and B19V to Mexico during the transatlantic slave trade* (p. 2020.06.05.137083). https://doi.org/10.1101/2020.06.05.137083
Harkins, K. M., & Stone, A. C. (2015). Ancient pathogen genomics: insights into timing and adaptation. *Journal of human evolution*, *79*, 137–149.

Jensen, T. Z. T., Niemann, J., Iversen, K. H., Fotakis, A. K., Gopalakrishnan, S., Vågene, Å. J., Pedersen, M. W., Sinding, M.-H. S., Ellegaard, M. R., Allentoft, M. E., Lanigan, L. T., Taurozzi, A. J., Nielsen, S. H., Dee, M. W., Mortensen, M. N., Christensen, M. C., Sørensen, S. A., Collins, M. J., Gilbert, M. T. P., … Schroeder, H. (2019). A 5700 year-old human genome and oral microbiome from chewed birch pitch. *Nature Communications*, *10*(1), 5520.

Kabamba, E. T., Tuan, V. P., & Yamaoka, Y. (2018). Genetic populations and virulence factors of Helicobacter pylori. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, *60*, 109–116.

Kahila Bar-Gal, G., Kim, M. J., Klein, A., Shin, D. H., Oh, C. S., Kim, J. W., Kim, T.-H., Kim, S. B., Grant, P. R., Pappo, O., Spigelman, M., & Shouval, D. (2012). Tracing hepatitis B virus to the 16th century in a Korean mummy. *Hepatology* , *56*(5), 1671–1680.

Kay, G. L., Sergeant, M. J., Giuffra, V., Bandiera, P., Milanese, M., Bramanti, B., Bianucci, R., & Pallen, M. J. (2014). Recovery of a Medieval Brucella melitensis Genome Using Shotgun Metagenomics. *mBio*, *5*(4), e01337–14 – e01337–14.

Kay, G. L., Sergeant, M. J., Zhou, Z., Chan, J. Z.-M., Millard, A., Quick, J., Szikossy, I., Pap, I., Spigelman, M., Loman, N. J., Achtman, M., Donoghue, H. D., & Pallen, M. J. (2015). Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nature communications*, *6*(1). https://doi.org/10.1038/ncomms7717

Keller, M., Spyrou, M. A., Scheib, C. L., Neumann, G. U., Kröpelin, A., Haas-Gebhard, B., Päffgen, B., Haberstroh, J., Lacomba, A. R. i., Raynaud, C., Cessford, C., Durand, R., Stadler, P., Nägele, K., Bates, J. S., Trautmann, B., Inskip, S. A., Peters, J., Robb, J. E., … Krause, J. (2019). Ancient Yersinia pestis genomes from across Western Europe reveal early diversification during the First Pandemic (541–750). *Proceedings of the National Academy of Sciences of the United States of America*, *116*(25), 12363–12372.

Key, F. M., Posth, C., Esquivel-Gomez, L. R., Hübler, R., Spyrou, M. A., Neumann, G. U., Furtwängler, A., Sabin, S., Burri, M., Wissgott, A., Lankapalli, A. K., Vågene, Å. J., Meyer, M., Nagel, S., Tukhbatova, R., Khokhlov, A., Chizhevsky, A., Hansen, S., Belinsky, A. B., … Krause, J. (2020). Emergence of human-adapted Salmonella enterica is linked to the Neolithization process. *Nature Ecology & Evolution*, *4*(3), 324–333.

Kirsten I Bos, Alexander Herbig, Jason Sahl, Nicholas Waglechner, Mathieu Fourment, Stephen A Forrest, Jennifer Klunk, Verena J Schuenemann, Debi Poinar, Melanie Kuch, G Brian Golding, Olivier Dutour, & Hendrik N Poinar. (2016). Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. *eLife*, *5*, 1–11.

Kolman, C. J., Centurion-Lara, A., Lukehart, S. A., Owsley, D. W., & Tuross, N. (1999). Identification of Treponema pallidum subspecies pallidum in a 200-year-old skeletal specimen. *The Journal of Infectious Diseases*, *180*(6), 2060–2063.

Krause-Kyora, B., Nutsua, M., Boehme, L., Pierini, F., Pedersen, D. D., Kornell, S.-C., Drichel, D., Bonazzi, M., Möbus, L., Tarp, P., Susat, J., Bosse, E., Willburger, B., Schmidt, A. H., Sauter, J., Franke, A., Wittig, M., Caliebe, A., Nothnagel, M., … Nebel, A. (2018). Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. *Nature communications*, *9*(1). https://doi.org/10.1038/s41467-018-03857-x

Krause-Kyora, B., Susat, J., Key, F. M., Kühnert, D., Bosse, E., Immel, A., Rinne, C., Kornell, S.-C., Yepes, D., Franzenburg, S., Heyne, H. O., Meier, T., Lösch, S., Meller, H., Friederich, S., Nicklisch, N., Alt, K. W., Schreiber, S., Tholey, A., … Krause, J. (2018). Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *eLife*, *7*. https://doi.org/10.7554/eLife.36666

Lan, T., & Lindqvist, C. (2018). *Paleogenomics: Genome-Scale Analysis of Ancient DNA and Population and Evolutionary Genomic Inferences*. Springer International Publishing.

Linderholm, A. (2016). Ancient DNA: the next generation - chapter and verse. En *Biological Journal of the Linnean Society* (Vol. 117, Número 1, pp. 150–160). https://doi.org/10.1111/bij.12616

MacHugh, D. E., Larson, G., & Orlando, L. (2017). Taming the Past: Ancient DNA and the Study of Animal Domestication. *Annual Review of Animal Biosciences*, *5*(1), 329–351.
Maixner, F., Krause-Kyora, B., Turaev, D., Herbig, A., Hoopmann, M. R., Hallows, J. L., Kusebauch, U., Vigl, E. E., Malfertheiner, P., Megraud, F., OSullivan, N., Cipollini, G., Coia, V., Samadelli, M., Engstrand, L., Linz, B., Moritz, R. L., Grimm, R., Krause, J., … Zink, A. (2016). The 5300-year-old Helicobacter pylori genome of the Iceman. *Science*, *351*(6269), 162–165.
Mann, A. E., Sabin, S., Ziesemer, K., Vågene, Å. J., Schroeder, H., Ozga, A. T., Sankaranarayanan, K., Hofman, C. A., Fellows Yates, J. A., Salazar-García, D. C., Frohlich, B., Aldenderfer, M., Hoogland, M., Read, C., Milner, G. R., Stone, A. C., Lewis, C. M., Krause, J., Hofman, C., … Warinner, C. (2018). Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. *Scientific reports*, *8*(1). https://doi.org/10.1038/s41598-018-28091-9
Marciniak, S., & Perry, G. H. (2017). Harnessing ancient genomes to study the history of human adaptation. *Nature reviews. Genetics*, *18*(11), 659–674.
Margaryan, A., Hansen, H. B., Rasmussen, S., Sikora, M., Moiseyev, V., Khoklov, A., Epimakhov, A., Yepiskoposyan, L., Kriiska, A., Varul, L., Saag, L., Lynnerup, N., Willerslev, E., & Allentoft, M. E. (2018). Ancient pathogen DNA in human teeth and petrous bones. *Ecology and evolution*, *8*(6), 3534–3542.
McCormick, M. (2021). Gregory of Tours on Sixth-Century Plague and Other Epidemics. *Speculum*, *96*(1), 38–96.
Morozova, I., Kasianov, A., Bruskin, S., Neukamm, J., Molak, M., Batieva, E., Pudło, A., Rühli, F. J., & Schuenemann, V. J. (2020). New ancient Eastern European Yersinia pestis genomes illuminate the dispersal of plague in Europe. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1812), 20190569.
Mühlemann, B., Jones, T. C., Damgaard, P. de B., Allentoft, M. E., Shevnina, I., Logvin, A., Usmanova, E., Panyushkina, I. P., Boldgiv, B., Bazartseren, T., Tashbaeva, K., Merz, V., Lau, N., Smrčka, V., Voyakin, D., Kitov, E., Epimakhov, A., Pokutta, D., Vicze, M., … Willerslev, E. (2018). Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature*, *557*(7705), 418–423.
Mühlemann, B., Margaryan, A., Damgaard, P. de B., Allentoft, M. E., Vinner, L., Hansen, A. J., Weber, A., Bazaliiskii, V. I., Molak, M., Arneborg, J., Bogdanowicz, W., Falys, C., Sablin, M., Smrčka, V., Sten, S., Tashbaeva, K., Lynnerup, N., Sikora, M., Smith, D. J., … Jones, T. C. (2018). Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(29), 7557–7562.
Mühlemann, B., Vinner, L., Margaryan, A., Wilhelmson, H., de la Fuente Castro, C., Allentoft, M. E., de Barros Damgaard, P., Hansen, A. J., Holtsmark Nielsen, S., Strand, L. M., Bill, J., Buzhilova, A., Pushkina, T., Falys, C., Khartanovich, V., Moiseyev, V., Jørkov, M. L. S., Østergaard Sørensen, P., Magnusson, Y., … Sikora, M. (2020). Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. *Science*, *369*(6502). https://doi.org/10.1126/science.aaw8977
Namouchi, A., Guellil, M., Kersten, O., Hänsch, S., Ottoni, C., Schmid, B. V., Pacciani, E., Quaglia, L., Vermunt, M., Bauer, E. L., Derrick, M., Jensen, A. Ø., Kacki, S., Cohn, S. K., Jr, Stenseth, N. C., & Bramanti, B. (2018). Integrative approach using Yersinia pestis genomes to revisit the historical landscape of plague during the Medieval Period. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(50), E11790–E11797.
Nerlich, A. G., Schraut, B., Dittrich, S., Jelinek, T., & Zink, A. R. (2008). *Plasmodium falciparum* in Ancient Egypt. *Emerging infectious diseases*, *14*(8), 1317–1319.
Nguyen-Hieu, T., Aboudharam, G., & Drancourt, M. (2011). *Mini Review: Dental Pulp as a Source for Paleomicrobiology*. *5*(1), 7.
Orlando, L., & Cooper, A. (2014). Using Ancient DNA to Understand Evolutionary and Ecological Processes. *Annual review of ecology, evolution, and systematics*, *45*(1), 573–598.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., & Hofreiter, M. (2004). Genetic Analyses from Ancient DNA. *Annual review of genetics*, *38*(1), 645–679.

Pajer, P., Dresler, J., Kabíckova, H., Písa, L., Aganov, P., Fucik, K., Elleder, D., Hron, T., Kuzelka, V., Velemínsky, P., Klimentova, J., Fucikova, A., Pejchal, J., Hrabakova, R., Benes, V., Rausch, T., Dundr, P., Pilin, A., Cabala, R., … Meyer, H. (2017). Characterization of Two Historic Smallpox Specimens from a Czech Museum. *Viruses*, *9*(8), 200.

Patterson Ross, Z., Klunk, J., Fornaciari, G., Giuffra, V., Duchêne, S., Duggan, A. T., Poinar, D., Douglas, M. W., Eden, J.-S., Holmes, E. C., & Poinar, H. N. (2018). The paradox of HBV evolution as revealed from a 16th century mummy. *PLoS pathogens*, *14*(1), e1006750.

Philips, A., Stolarek, I., Handschuh, L., Nowis, K., Juras, A., Trzciński, D., Nowaczewska, W., Wrzesińska, A., Potempa, J., & Figlerowicz, M. (2020). Analysis of oral microbiome from fossil human remains revealed the significant differences in virulence factors of modern and ancient Tannerella forsythia. *BMC Genomics*, *21*(1), 402.

Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczky, P., Anders, A., Pietrusewsky, M., Rollefson, G., Jovanovic, M., Trinhhoang, H., Bar-Oz, G., Oxenham, M., Matsumura, H., & Hofreiter, M. (2015). Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PloS One*, *10*(6), e0129102.

Racimo, F., Sikora, M., Vander Linden, M., Schroeder, H., & Lalueza-Fox, C. (2020). Beyond broad strokes: sociocultural insights from the study of ancient genomes. *Nature Reviews. Genetics*, *21*(6), 355–366.

Raoult, D., & Drancourt, M. (Eds.). (2008). *Paleomicrobiology: past human infections*. Springer.

Rasmussen, S., Allentoft, M. E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.-G., Pedersen, A. G., Schubert, M., Van Dam, A., Kapel, C. M. O., Nielsen, H. B., Brunak, S., Avetisyan, P., Epimakhov, A., Khalyapin, M. V., Gnuni, A., Kriiska, A., Lasak, I., Metspalu, M., … Willerslev, E. (2015). Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. *Cell*, *163*(3), 571–582.

Salo, W. L., Aufderheide, A. C., Buikstra, J., & Holcomb, T. A. (1994). Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy. *Proceedings of the National Academy of Sciences*, *91*(6), 2091–2094.

Sandoval-Velasco, M., Lundstrøm, I. K. C., Wales, N., Ávila-Arcos, M. C., Schroeder, H., & Gilbert, M. T. P. (2017). Relative performance of two DNA extraction and library preparation methods on archaeological human teeth samples. *STAR: Science & Technology of Archaeological Research*, *3*(1), 80–88.

Schuenemann, V. J., Avanzi, C., Krause-Kyora, B., Seitz, A., Herbig, A., Inskip, S., Bonazzi, M., Reiter, E., Urban, C., Dangvard Pedersen, D., Taylor, G. M., Singh, P., Stewart, G. R., Velemínský, P., Likovsky, J., Marcsik, A., Molnár, E., Pálfi, G., Mariotti, V., … Krause, J. (2018). Ancient genomes reveal a high diversity of Mycobacterium leprae in medieval Europe. *PLoS pathogens*, *14*(5), e1006997.

Schuenemann, V. J., Bos, K., DeWitte, S., Schmedes, S., Jamieson, J., Mittnik, A., Forrest, S., Coombes, B. K., Wood, J. W., Earn, D. J. D., White, W., Krause, J., & Poinar, H. N. (2011). Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. *Proceedings of the National Academy of Sciences*, *108*(38), E746–E752.

Schuenemann, V. J., Kumar Lankapalli, A., Barquera, R., Nelson, E. A., Iraíz Hernández, D., Acuña Alonzo, V., Bos, K. I., Márquez Morfín, L., Herbig, A., & Krause, J. (2018). Historic Treponema pallidum genomes from Colonial Mexico retrieved from archaeological remains. *PLoS neglected tropical diseases*, *12*(6), e0006447.

Schuenemann, V. J., Singh, P., Mendum, T. A., Krause-Kyora, B., Jager, G., Bos, K. I., Herbig, A., Economou, C., Benjak, A., Busso, P., Nebel, A., Boldsen, J. L., Kjellstrom, A., Wu, H., Stewart,

G. R., Taylor, G. M., Bauer, P., Lee, O. Y.-. C., Wu, H. H. T., … Krause, J. (2013). Genome-Wide Comparison of Medieval and Modern Mycobacterium leprae. *Science*, *341*(6142), 179–183.

Sirak, K. A., Fernandes, D. M., Cheronet, O., Novak, M., Gamarra, B., Balassa, T., Bernert, Z., Cséki, A., Dani, J., Gallina, J. Z., Kocsis-Buruzs, G., Kővári, I., László, O., Pap, I., Patay, R., Petkes, Z., Szenthe, G., Szeniczey, T., Hajdu, T., & Pinhasi, R. (2017). A minimally-invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis. *BioTechniques*, *62*(6), 283–289.

Sirak, K., Fernandes, D., Cheronet, O., Harney, E., Mah, M., Mallick, S., Rohland, N., Adamski, N., Broomandkhoshbacht, N., Callan, K., Candilio, F., Lawson, A. M., Mandl, K., Oppenheimer, J., Stewardson, K., Zalzala, F., Anders, A., Bartík, J., Coppa, A., … Pinhasi, R. (2020). Human auditory ossicles as an alternative optimal source of ancient DNA. *Genome Research*, *30*(3), 427–436.

Smith, O., Dunshea, G., Sinding, M.-H. S., Fedorov, S., Germonpre, M., Bocherens, H., & Gilbert, M. T. P. (2019). Ancient RNA from Late Pleistocene permafrost and historical canids shows tissue-specific transcriptome survival. *PLoS Biology*, *17*(7), e3000166.

Spigelman, M., & Lemma, E. (1993). The use of the polymerase chain reaction (PCR) to detect Mycobacterium tuberculosis in ancient skeletons. *International Journal of Osteoarchaeology*, *3*(2), 137–143.

Spyrou, M. A., Bos, K. I., Herbig, A., & Krause, J. (2019). Ancient pathogen genomics as an emerging tool for infectious disease research. *Nature Reviews. Genetics*, *20*(6), 323–340.

Spyrou, M. A., Tukhbatova, R. I., Wang, C.-C., Valtueña, A. A., Lankapalli, A. K., Kondrashin, V. V., Tsybin, V. A., Khokhlov, A., Kühnert, D., Herbig, A., Bos, K. I., & Krause, J. (2018). Analysis of 3800-year-old Yersinia pestis genomes suggests Bronze Age origin for bubonic plague. *Nature communications*, *9*(1). https://doi.org/10.1038/s41467-018-04550-9

Tsangaras, K., & Greenwood, A. D. (2018). *Paleovirology: Viral Sequences from Historical and Ancient DNA*. Springer International Publishing.

Vågene, Å. J., Herbig, A., Campana, M. G., Robles García, N. M., Warinner, C., Sabin, S., Spyrou, M. A., Andrades Valtueña, A., Huson, D., Tuross, N., Bos, K. I., & Krause, J. (2018). Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology & Evolution*, *2*(3), 520–528.

Wagner, D. M., Klunk, J., Harbeck, M., Devault, A., Waglechner, N., Sahl, J. W., Enk, J., Birdsell, D. N., Kuch, M., Lumibao, C., Poinar, D., Pearson, T., Fourment, M., Golding, B., Riehm, J. M., Earn, D. J. D., DeWitte, S., Rouillard, J.-M., Grupe, G., … Poinar, H. (2014). Yersinia pestis and the Plague of Justinian 541–543 AD: a genomic analysis. *The Lancet infectious diseases*, *14*(4), 319–326.

Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R. Y., Fiddyment, S., Speller, C., Hendy, J., Charlton, S., Luder, H. U., Salazar-García, D. C., Eppler, E., Seiler, R., Hansen, L. H., Castruita, J. A. S., … Cappellini, E. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics*, *46*(4), 336–344.

Worobey, M., Watts, T. D., McKay, R. A., Suchard, M. A., Granade, T., Teuwen, D. E., Koblin, B. A., Heneine, W., Lemey, P., & Jaffe, H. W. (2016). 1970s and “Patient 0”HIV-1 genomes illuminate early HIV/AIDS history in North America. *Nature*, *539*(7627), 98–101.

Zhou, Z., Lundstrøm, I., Tran-Dien, A., Duchêne, S., Alikhan, N.-F., Sergeant, M. J., Langridge, G., Fotakis, A. K., Nair, S., Stenøien, H. K., Hamre, S. S., Casjens, S., Christophersen, A., Quince, C., Thomson, N. R., Weill, F.-X., Ho, S. Y. W., Gilbert, M. T. P., & Achtman, M. (2018). Pan-genome Analysis of Ancient and Modern Salmonella enterica Demonstrates Genomic Stability of the Invasive Para C Lineage for Millennia. *Current biology: CB*, *28*(15), 2420–2428.e10.

5. Capítulo 2. Artículo de investigación. Paleogenomic insights into the red complex bacteria *Tannerella forsythia* in Pre-Hispanic and Colonial individuals from Mexico.

Publicado:
(Bravo-Lopez et al. 2020)

Miriam Bravo-Lopez, Viridiana Villa-Islas, Carolina Rocha Arriaga, Ana B. Villaseñor-Altamirano, Axel Guzmán-Solís, Marcela Sandoval-Velasco, Julie K. Wesp, Keitlyn Alcantara, Aurelio López-Corral, Jorge Gómez-Valdés, Elizabeth Mejía, Alberto Herrera, Alejandro Meraz-Moreno, Maria de la Luz Moreno-Cabrera, Andrés Moreno-Estrada, Maria A. Nieves-Colón, Joel Olvera, Julia Pérez-Pérez, Katrine Højholt Iversen, Simon Rasmussen, Karla Sandoval, Gabriela Zepeda and María C. Ávila-Arcos.

Abstract

Background

The 'red complex' is an aggregate of three oral bacteria (T*annerella forsythia, Porphyromonas gingivalis and Treponema denticola*) responsible for severe clinical manifestation of periodontal disease. The unique morphological characteristics of *T. forsythia,* like a distinctive glycosylated S-layer, promotes its adherence to the host's gingival cell surfaces and attenuation of the immune response, positioned as the main contributor to the development of periodontitis. Although the recovery of *T. forsythia* DNA from ancient humans and archaic hominins suggests a long term (thousands of years) relationship between them, no study to our knowledge has leveraged genetic information from ancient teeth or dental calculus to explicitly explore the molecular properties of this infectious condition from a spatio-temporal framework. Therefore, we aim to characterize the genome of *T. forsythia* spanning the Pre-Hispanic and Colonial periods in present-day Central Mexico and combined these data with other ancient and modern strains, can shed light on its evolutionary history and the coevolutionary dynamics between this oral pathogen and humans.

Results

To retrieve ancient genomic data for *T. forsythia*, we used a hybridization capture-enrichment design targeting 234 genes to reconstruct partial genomes (approx. 259 Kb) for seven Pre-Hispanic and five Colonial individuals where the pathogen was identified. We recovered twelve partial ancient *T. forsythia* genomes and observed a distinct phylogenetic placement of samples, suggesting that the strains present in Pre-Hispanic individuals likely arrived with the first human migrations to the Americas and that new strains were introduced with the arrival of European and African populations in the sixteenth century.

Conclusions

This study provides the first account of evolutionary patterns in the 'red complex' bacteria *T. forsythia* by leveraging ancient and present-day genomic data. The identification of this pathogen in Pre-Hispanic individuals from Mexico is indicative of its long-standing relationship between the human host. Moreover, we were able to recover informative amounts of aDNA from this pathogen by means of a capture-enrichment strategy that considerably increased the depth of coverage for 234 informative genes. Though limited by small sample size and the targeting of only a fraction of the genes in the *T. forsythia's* genome, our results unveil many future directions for the study of this pathogen. Further characterization of whole ancient *T. forsythia* genomes from different spatio-temporal contexts and diverse populations can provide valuable data to validate the usability of this pathogen as a proxy for inferring past social interactions and to characterize in depth host–pathogen coevolutionary dynamics through extended periods of time.

PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Bravo-Lopez M *et al*. 2020 Paleogenomic insights into the red complex bacteria *Tannerella forsythia* in Pre-Hispanic and Colonial individuals from Mexico. *Phil. Trans. R. Soc. B* **375**: 20190580.
http://dx.doi.org/10.1098/rstb.2019.0580

Accepted: 8 June 2020

One contribution of 14 to a theme issue 'Insights into health and disease from ancient biomolecules'.

Subject Areas:
genomics, health and disease and epidemiology, evolution

Keywords:
paleogenomics, *Tannerella forsythia*, ancient pathogens, periodontal disease, red complex, capture-enrichment

Author for correspondence:
María C. Ávila-Arcos
e-mail: mavila@liigh.unam.mx

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5096265.

THE ROYAL SOCIETY PUBLISHING

Paleogenomic insights into the red complex bacteria *Tannerella forsythia* in Pre-Hispanic and Colonial individuals from Mexico

Miriam Bravo-Lopez[1], Viridiana Villa-Islas[1], Carolina Rocha Arriaga[1], Ana B. Villaseñor-Altamirano[1], Axel Guzmán-Solís[1], Marcela Sandoval-Velasco[2], Julie K. Wesp[3], Keitlyn Alcantara[4], Aurelio López-Corral[5], Jorge Gómez-Valdés[6], Elizabeth Mejía[7], Alberto Herrera[7], Alejandro Meraz-Moreno[8], Maria de la Luz Moreno-Cabrera[8], Andrés Moreno-Estrada[9], Maria A. Nieves-Colón[9,10], Joel Olvera[11], Julia Pérez-Pérez[6], Katrine Højholt Iversen[12], Simon Rasmussen[12], Karla Sandoval[9], Gabriela Zepeda[13] and María C. Ávila-Arcos[1]

[1]International Laboratory for Human Genome Research, National Autonomous University of México (UNAM), Querétaro, 76230, Mexico
[2]Section for Evolutionary Genomics, GLOBE Institute, University of Copenhagen, Copenhagen, 1350, Denmark
[3]Department of Sociology and Anthropology, North Carolina State University, Raleigh, NC 27695, USA
[4]Department of Anthropology, Indiana University-Bloomington, Bloomington, IN 47405, USA
[5]Department of Archeology, National Institute of Anthropology and History, Tlaxcala, 90000, Mexico
[6]National School of Anthropology and History, Mexico City, 14030, Mexico
[7]National Institute of Anthropology and History, Querétaro, 76000, Mexico
[8]National Institute of Anthropology and History, Mexico City, 06600, Mexico
[9]National Laboratory of Genomics for Biodiversity, Unit of Advanced Genomics (LANGEBIO), CINVESTAV, Irapuato, Guanajuato 36824, Mexico
[10]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85281, USA
[11]Graduate Program of Physical Anthropology, National School of Anthropology and History, Mexico City, 14030, Mexico
[12]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 2200, Denmark
[13]National Institute of Anthropology and History, Guanajuato, 36250, Mexico

JKW, 0000-0002-7744-6401; JG-V, 0000-0001-6996-2732

The 'red complex' is an aggregate of three oral bacteria (*Tannerella forsythia*, *Porphyromonas gingivalis* and *Treponema denticola*) responsible for severe clinical manifestation of periodontal disease. Here, we report the first direct evidence of ancient *T. forsythia* DNA in dentin and dental calculus samples from archaeological skeletal remains that span from the Pre-Hispanic to the Colonial period in Mexico. We recovered twelve partial ancient *T. forsythia* genomes and observed a distinct phylogenetic placement of samples, suggesting that the strains present in Pre-Hispanic individuals likely arrived with the first human migrations to the Americas and that new strains were introduced with the arrival of European and African populations in the sixteenth century. We also identified instances of the differential presence of genes between periods in the *T. forsythia* ancient genomes, with certain genes present in Pre-Hispanic individuals and absent in Colonial individuals, and *vice versa*. This study highlights the potential for studying ancient *T. forsythia* genomes to unveil past social interactions through analysis of disease transmission. Our results illustrate the long-standing relationship between this oral pathogen and its human host, while also unveiling key evidence to understand its evolutionary history in Pre-Hispanic and Colonial Mexico.

This article is part of the theme issue 'Insights into health and disease from ancient biomolecules'.

© 2020 The Author(s) Published by the Royal Society. All rights reserved.

1. Introduction

Paleopathological evidence and ethnohistorical descriptions have traditionally served as the main sources for understanding population health in the past [1]. They provide rich information about the human context of diseases and the spatio-temporal variation that affect disease dynamics [2]; however, these approaches provide limited information regarding the specific pathogens associated with past diseases. The use of biomolecular tools, such as the retrieval and analysis of ancient DNA (aDNA), now provides new sources of evidence for understanding disease environments in the past.

In particular, the recovery of high-throughput sequencing and analysis of pathogen DNA from archeological remains have revealed a number of complex evolutionary patterns of the causative agents for various infectious diseases that were not apparent when solely considering historical and archeological information [3–8]. Most research in this area has focused on ancient pathogens from European samples, with a particular emphasis on the bubonic plague bacterium *Yersinia pestis* [5,8–10]. Only a handful of studies have explored the genetic makeup of ancient pathogens from the Americas [4,6,7], a region of interest considering the dramatic impact that newly introduced pathogens had on Indigenous populations after European contact. This research builds on these previous studies with a diachronic study of ancient pathogens from across different regions in Central Mexico.

Teeth are often the first-choice substrate for aDNA extraction as they are known to be a good source of endogenous aDNA [11] as well as ancient pathogen DNA [12], particularly for oral pathogens [3,13,14]. Furthermore, dental calculus (i.e. mineralized biofilm) has recently been recognized as a valuable reservoir for microbial, dietary and host DNA [13,15]. This is important considering that oral infections represent some of the most common human diseases worldwide [16], have affected human populations throughout their evolutionary history [17] and increased in frequency with the advent of agriculture [18,19]. In addition, it is also known that oral infections can cause severe harm at the systemic level when left untreated [20].

Periodontitis (also known as periodontal disease) is a severe type of oral infection that can cause major damage to the alveolar bone [21,22]. The manifestation of this condition is the result of different bacterial complexes in the oral environment. The most pathogenic is the 'red complex', which includes the Gram-negative bacteria *Treponema denticola*, *Porphyromonas gingivalis* and *Tannerella forsythia* [23]. These bacteria act in synergy causing the destruction of alveolar bone and soft tissues, which can ultimately lead to tooth loss if untreated [24]. Furthermore, the bacteria produce endotoxins that can cause systemic health conditions such as vascular disorders [25,26] and respiratory tract infections, among others [20]. In particular, *T. forsythia* is identified as the main contributor to the development of periodontitis, since it is the only member of the 'red complex' with a distinctive glycosylated S-layer that promotes adherence to the host's gingival cell surfaces [27] and attenuation of the immune response [28,29]. The relevance of understanding the molecular principles of this disease is well acknowledged in the clinical context [30]; however, it has received less attention from an evolutionary standpoint. Although the recovery of *T. forsythia* DNA from ancient humans [3,31] and archaic hominins [32] suggests that this pathogen has infected humans for thousands of years, no study to our knowledge has leveraged genetic information from ancient teeth or dental calculus to explicitly explore the molecular properties of this infectious condition from a spatio-temporal framework. Therefore, a more in-depth characterization of the genome of *T. forsythia* at different points in time and in different geographical regions can shed light on its evolutionary history and the coevolutionary dynamics between this oral pathogen and humans.

Since the European colonization of the Americas imposed one of the most dramatic social, cultural and demographic transitions in human history, exploring the genetic makeup of oral pathogens spanning this shift is of deep interest to unveiling the evolutionary mechanisms underlying pathogenicity and adaptation to new environments. As a first approach to address this issue, we characterized partial *T. forsythia* genomes from twelve individuals spanning the Pre-Hispanic and Colonial periods in present-day Central Mexico and combined these data with other ancient [3,31] and modern [33] strains. While the first contact between Europeans and different Native groups in the Mexican territory varied geographically, we use the term 'Colonial' to refer to the period between 1519 and 1810 CE, which is the standard for Mesoamerican archaeology and the sites included in this study [34].

To retrieve ancient genomic data for this oral pathogen, we used a hybridization capture-enrichment design targeting 234 genes to reconstruct partial genomes (approx. 259 Kb) of *T. forsythia* for seven Pre-Hispanic and five Colonial individuals where the pathogen was identified. This allowed us to assess the phylogenetic relationship between Pre-Hispanic and Colonial strains, as well as variations between these and the other ancient and modern strains.

2. Identification of *Tannerella forsythia* in Pre-Hispanic and colonial individuals from Mexico

We screened 53 aDNA Illumina libraries (49 from teeth and 4 from dental calculus; electronic supplementary material table S1) by sequencing at low depth in the NextSeq550 platform, producing between approximately 2 million and approximately 20 million paired-end reads per library (electronic supplementary material table S6). Twenty samples belong to Pre-Hispanic (540 BCE–1519 CE) individuals from the modern states of Guanajuato, Queretaro, Tlaxcala and Veracruz and 33 samples to Colonial individuals from Mexico City and Queretaro (figure 1). Detailed information for each sample is found in electronic supplementary material table S1. Sequenced reads that did not map to the human reference genome were taxonomically classified using Kraken2 [35] through comparison to a database composed of whole bacterial, archaeal and viral genomes from the NCBI's Reference Sequence Database (see §7). As a spatial and temporal reference, we also included in our analysis data from four ancient (950 CE–1850 CE) European individuals [3,31] previously reported as positive for 'red complex' bacteria.

The per cent of reads assigned to a taxon by Kraken2 ranged from 1% to 34%, with a predominant component of environmental and human microbiome bacteria in varying proportions (figure 2 and electronic supplementary material, table S2). We identified the presence of 'red complex' bacteria; *T. forsythia* (1065 to 307 840 reads); *P. gingivalis* (14 to 7422 reads) and *T. denticola* (140 to 6807 reads) in

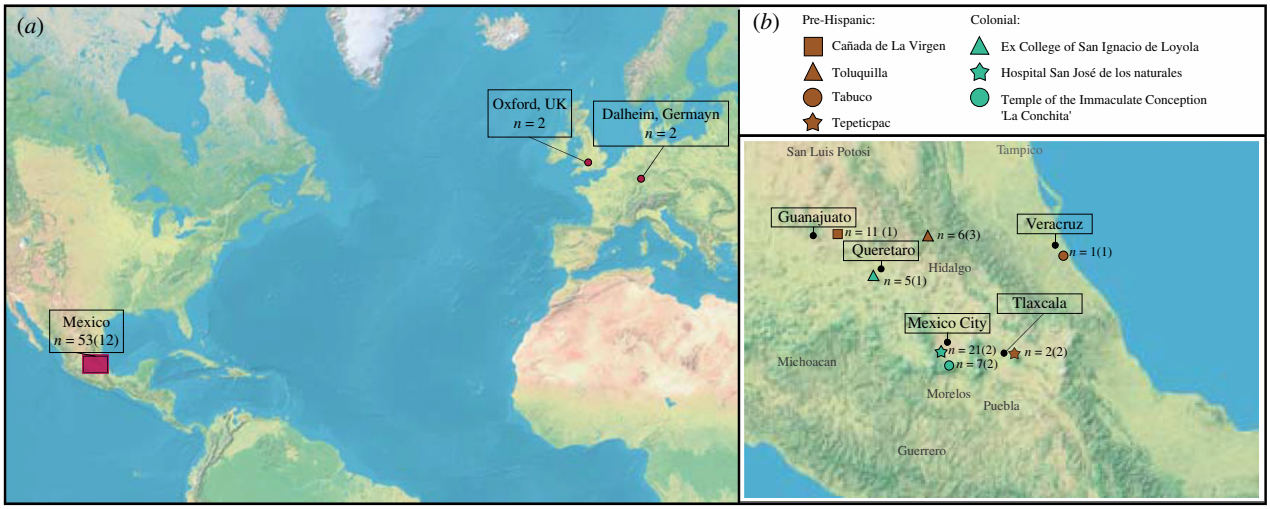


Figure 1. Geographic locations and temporal periods of ancient individuals included in this study. (*a*) Study sites with ancient individuals positive for *T. forsythia*. (*b*) Map of Mexico indicating the position of the archeological sites included in this study. The time periods of the samples analysed are indicated by colour. Yellow: Pre-Hispanic, Turquoise: Colonial. The number of samples that are positive for *T. forsythia* are indicated in brackets for each archeological site location.

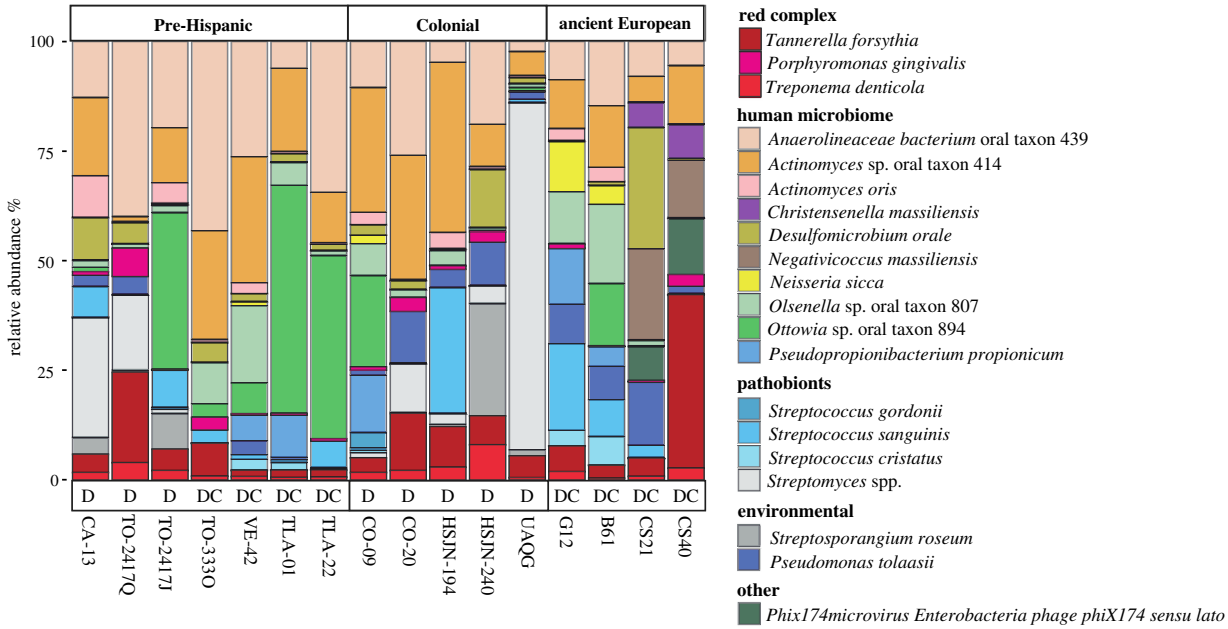


Figure 2. Red complex bacteria in Pre-Hispanic and Colonial individuals. Relative abundance as determined by Kraken2 [35] of the main taxa observed in each ancient individual in this study. 'Red complex' bacteria are coloured in shades of red, reads assigned to *T. forsythia* are distinguished by the most intense red colour. D = dentin sample and DC = dental calculus sample. 'Ancient European' individuals correspond to previously published data; G12 and B61 [3], CS21 and CS40 [31]. The main taxa categorization was based on references found in table S3.

twelve individuals: seven Pre-Hispanic and five Colonial (figure 1, tables 1 and 2, electronic supplementary material, table S2). Notably, *T. forsythia* was the dominant bacterium in this analysis in comparison to the other two members of the 'red complex'. By contrast, the taxonomic profiling of reads from twelve DNA extraction and library preparation blanks (negative controls) assigned a low number of reads (2 and 6) to *T. forsythia 92A2* in two blanks (electronic supplementary material table S4; see also electronic supplementary material, Discussion), giving us confidence that our results were unlikely to be false positives. Moreover, when only considering teeth, only 8 out of the 49 libraries had hits to *T. forsythia*, while all four dental calculus samples did. When contrasting the frequency of *T. forsythia*-positive samples per period, we observe 7 out of 20 samples for the Pre-Hispanic period and 5 out of 33 samples for the Colonial period.

Interestingly, the taxonomic profiles of the 'red complex'-positive libraries displayed by non-metric multidimensional scaling (NMDS) reveal a clustering based on the substrate analysed (electronic supplementary material figure S1). The exceptions to this clustering pattern are samples TO-2417J and CO-09, which are dentine samples that cluster with calculus samples (electronic supplementary material figure S1, electronic supplementary material table S5). The taxa that seem to be driving this are *Actinomyces* sp. *oral taxon 894*, *Olsenella* and sp. *oral taxon 807*, *Anaerolineaceae bacterium oral taxon 439*, *Streptococcus gordonii* and *T. forsythia*.

We mapped the low-depth sequencing data that did not map to the human reference genome of the twelve individuals with reads assigned *T. forsythia* by Kraken2 to the available *T. forsythia* reference genome (NC_016610.1) to confirm the presence of authentic aDNA from these bacteria and to rule

out potential modern contamination (see §7). This yielded between 1065 and 769 591 uniquely mapped reads (electronic supplementary material, table S6). By contrast, only three out of twelve experimental negative controls yielded between 1 and 30 reads mapping to *T. forsythia* (electronic supplementary material, table S6). We observed the characteristic damage and fragmentation patterns expected for aDNA libraries [36], namely the elevated C to T transitions at the molecule ends (electronic supplementary material, figure S2) and short-read lengths between 69 and 94 base pairs (bp). Additionally, the mapped reads were distributed across the entire genome and not concentrated in conserved or specific regions. Together, these observations support that the reads represent authentic aDNA from *T. forsythia*.

3. Capture design and targeted enrichment of aDNA libraries

To increase the genomic coverage of *T. forsythia*, we designed a custom set of in-solution capture-enrichment baits targeting 234 genes (258 696 bp) found in the *T. forsythia* RefSeq genome assembly (NC_016 610.1). The 234 genes were selected based on functional annotations in the Pathosystems Resource Integration Center (PATRIC) database [37]. These include 207 genes annotated as 'essential', 24 genes with an annotation related to antibiotic resistance, two genes annotated as transporters and one gene annotated as a virulence factor (electronic supplementary material, figure S3). A description of the regions used for the probe design is available in electronic supplementary material, table S7.

T. forsythia DNA was enriched in 11 out of the 12 libraries and sequenced between approximately 461 thousand and approximately 58 million reads for each with two sequencing rounds on the Illumina NextSeq550. The library for individual TO-2417J was not included in the capture since it was sequenced at high depth, together with the pre-capture library for individual TO-2417Q on a NovaSeq instrument in order to reach whole-genome coverage for both samples (see §7). Upon merging the sequence data available for each sample, between 12 326 and 179 824 unique reads mapping to the reference genome were retrieved, which resulted in an average on-target depth between approximately 1.50 X and approximately 42 X. The captured libraries show between 16- and 213-fold enrichment of on-target sequences compared to the pre-capture libraries (figure 3 and electronic supplementary material, figure S4 and table S6 and S8). For individuals TO-2417Q and TO-2417J, for which we sequenced pre-capture libraries at a higher depth, we obtained whole genomes with depths of approximately 28.6 X and approximately 9.8 X, respectively. The depth for sample TO-2417Q was calculated from the merged data for all pre-capture and post-capture reads.

We did not find a correlation between the type of substrate (dentin or calculus) or the pre-capture endogenous content and the performance of the capture strategy; however, the pre-capture library with the shortest fragments (HSJN-194, an average of 69 bp) had the highest fold enrichment (approx. 213 fold) post-capture. Overall, there was a general trend in which pre-capture libraries with shorter average fragment lengths yielded higher fold enrichment values (electronic supplementary material table S8). However, we did not observe any significant correlation between fragment length and fold enrichment (Pearson's $r = 0.10$, p-value = 0.89) when considering only the four dental calculus samples, either when considering only the dentin samples (Pearson's $r = -0.67$, p-value = 0.09) or all samples together (Pearson's $r = -0.56$, p-value = 0.06).

4. Phylogenetic analysis of ancient *Tannerella forsythia* genomes

We selected the samples with depths above 2.5 X, which excluded Pre-Hispanic sample CA-13, to investigate their phylogenetic relationship to available *T. forsythia* genome-wide data. To this end, the data for the remaining 11 Pre-Hispanic ($n = 6$) and Colonial ($n = 5$) samples were merged with published ancient [3,31] and modern [38] sequence data for this oral pathogen. The comparative data included two ancient individuals (950–1200 CE) from Dalheim, Germany [3] and two ancient individuals (1770–1855 CE) from the Radcliffe Infirmary Burial Ground collection in Oxford, UK [31] (figure 1, electronic supplementary material, table S1). We also included data for two present-day *T. forsythia* genomes, one retrieved from a periodontitis patient [33] and *T. forsythia*'s reference genome sequence (NC_016610.1), which belongs to strain 92A2; both modern genomes were obtained from US patients. In sum, the phylogenetic analysis considered a total of seventeen *T. forsythia* partial genomes, 15 ancient and 2 modern.

The raw read data from the previous studies were mapped to the *T. forsythia* genome with the same parameters used for our data. From the mapped data, we generated consensus sequences for 170 genes (electronic supplementary material, table S7) that were in the capture design and were also orthologous between the *T. forsythia* and the *P. gingivalis* reference genome (NZ_CP025932.1), which was used as an outgroup in the phylogenetic analysis (see §7).

The maximum-likelihood tree integrates the consensus sequences from all eighteen samples: six Pre-Hispanic, five Colonial, four ancient European, two modern *T. forsythia* [33,39] and *P. gingivalis*. Out of the 15 bootstrap support values for the bifurcations in the tree, 11 were above 75, and the lowest value was 55, consistent with an overall well-supported topology. There is a phylogenetic positioning that corresponds with the dates and geographic origin of the samples, with one clade grouping Colonial, European and modern individuals, and a second clade that includes all of the Pre-Hispanic *T. forsythia* samples (figure 4). In the *T. forsythia* Pre-Hispanic cluster, there is no clear phylogenetic proximity according to geographical location, except two samples from Toluquilla, Queretaro (TO-2417Q and TO-333O). Interestingly, TO-2417J, which is from the same site, is basal to the remaining Pre-Hispanic samples. The Colonial samples cluster together with the ancient European individuals and the two modern *T. forsythia* samples from the USA forming a monophyletic group. Noticeably, the reference *T. forsythia* genome is closest to one of the historical samples from Germany (CS40), while the second sample from this site (CS21) is basal to a group consisting of the modern and Colonial samples. The placement of the remaining samples does not seem to follow a clear temporal or geographical trend; however, individual HSJN-194 is basal to the Modern/European/Colonial clade and is closely related to individual C0–09. This is noteworthy since morphological analyses of HSJN-194 [40] and mitochondrial ancestry of both hosts

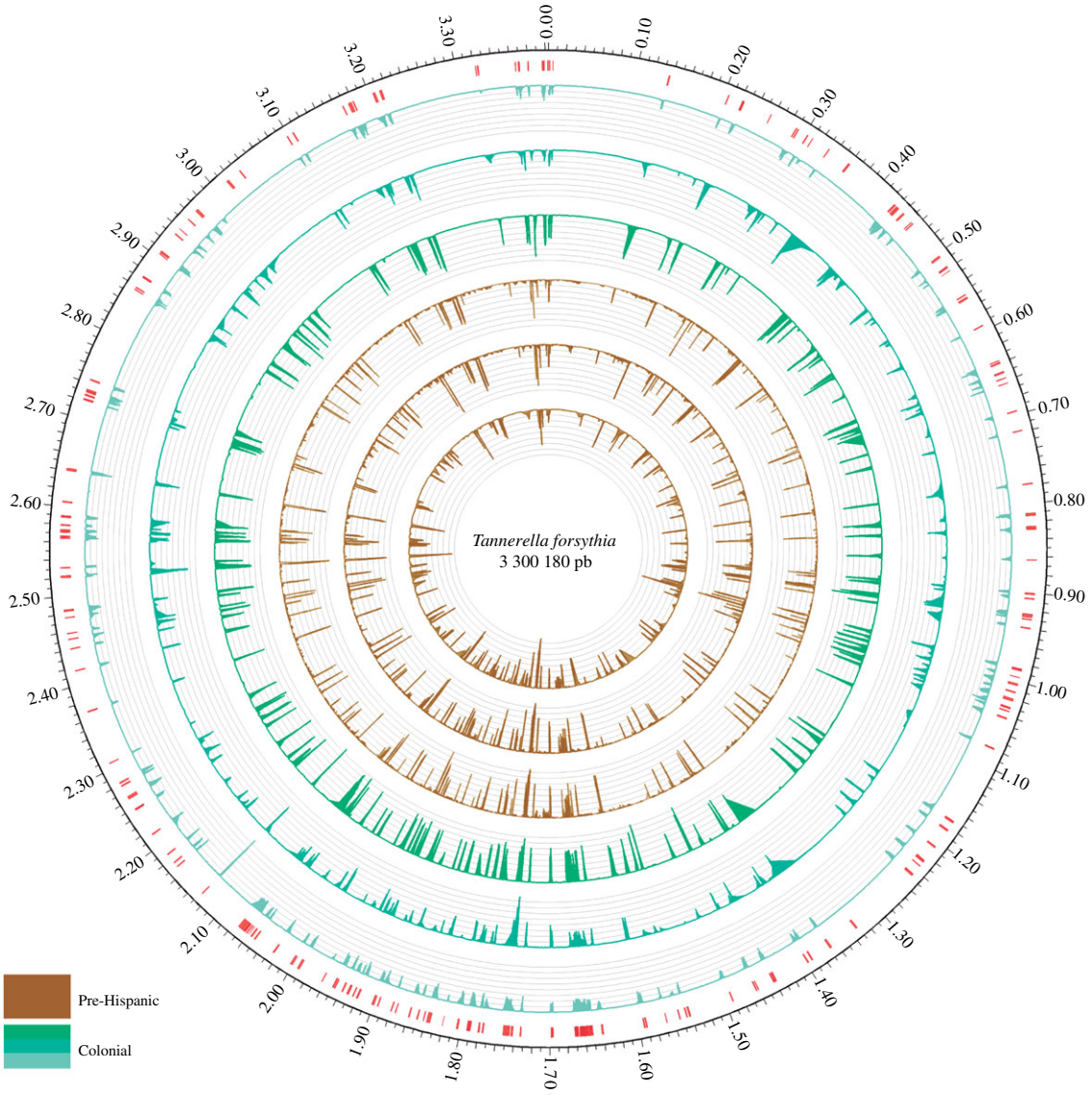


Figure 3. Depth of coverage of the *T. forsythia* genome in three Pre-Hispanic and three Colonial samples. The depth is shown in 100 bp windows and it is considerably higher at the 234 genes that were included in the targeted capture design (red lines in outer ring). Each grey line represents approx. 2 x. From inner ring: Pre-Hispanic individuals TO-2417J, TO-333O, TLA-01; Colonial individuals: CO-20, HSJN-240, UAQG. The outmost ring represents the location of the probes. All the plots for Pre-Hispanic and Colonial individuals are shown in figure S4.

(electronic supplementary material, table S1) suggests these individuals are of African ancestry. The next most basal sample from this clade corresponds to individual B61, which is one of the two most ancient (900–1200 CE) samples in the dataset and may explain its basal placement in the tree.

5. Differential presence of genes

To explore the genomic characteristics of the Pre-Hispanic and Colonial *T. forsythia*, we evaluated the presence or absence of the 234 genes included in our custom bait design across samples. The depth distribution across genes for each sample was contrasted (electronic supplementary material figures S5–S8), though samples CA-13 and CS21 were excluded from this analysis given the low on-target depth of coverage (1.53 X and 2.5 X, respectively). For the remaining samples, the depth values were normalized (electronic supplementary material figure S6 and electronic supplementary material table S9) and the genes with normalized values of 0 were considered as absent (See §7).

Out of the 207 genes annotated as 'essential' in the PATRIC database, there is a differential presence of only two genes between Pre-Hispanic and Colonial individuals (figure 5, electronic supplementary material figure S6). These include the absence of *pncB* (BFO-2125), a nicotinate phosphoribosyltransferase in all the Pre-Hispanic individuals, except TO-333O (figure 5), while it is present in all Colonial, ancient European and modern sequences. The second gene, *rplI* (BFO-3362), is absent in all of the Colonial individuals except CO-09, while it is present in all Pre-Hispanic, ancient Europeans and modern sequences.

In addition to these two genes, all of the ancient and modern *T. forsythias* analysed in this study lack *tetQ* (BFO-1235), a gene associated with antibiotic resistance by protecting the bacterial ribosome from binding tetracycline (figure 5). This pattern had been previously reported for the *T. forsythia* ancient genomes reconstructed from the two ancient individuals from Dalheim, Germany (G12 and B61) [3]. This

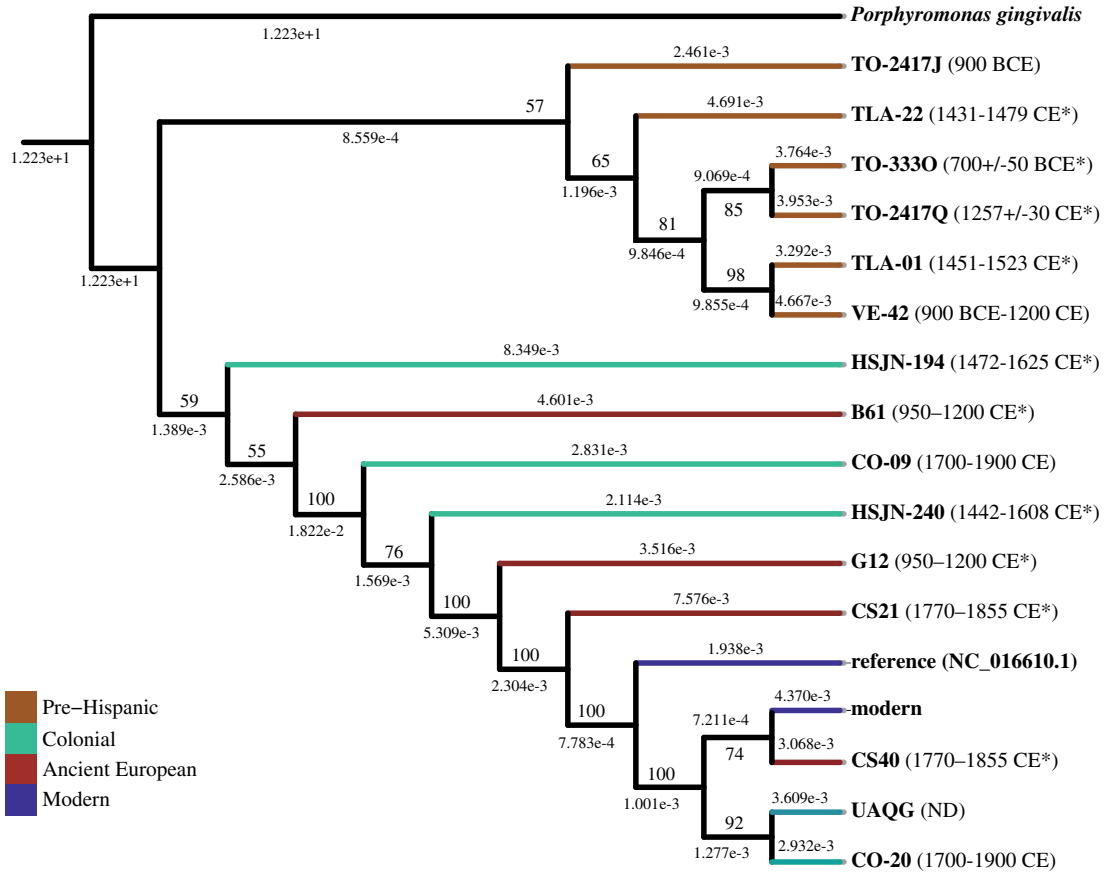


Figure 4. Phylogenetic tree of 15 ancient and two modern *T. forsythia* genomes. A maximum-likelihood three built with 1000 bootstraps is shown. Bootstrap values are shown for each bifurcation and branch lengths are shown on top of each branch. Colours in the branch reflect the period according to the legend. Estimated or C14 dates are shown in parentheses next to the sample's names.

gene is, however, annotated in the *T. forsythia* reference genome (NC_016610.1).

6. Discussion

We identified 'red complex' bacteria (*Tannerella forsythia*, *Porphyromonas gingivalis* and *Treponema denticola*) in seven Pre-Hispanic and five Colonial individuals from Mexico. Although previous studies have reported the presence of these pathogenic bacteria in ancient European individuals [3,31,32,41], to our knowledge this study is the first direct account of *T. forsythia* in Pre-Hispanic and Colonial individuals from the Americas.

Notably, *T. forsythia* was the dominant bacteria in comparison to the other two members of the 'red complex'. This bacterium is present in all four dental calculus samples tested in this study, as opposed to only 8 out of 49 ancient teeth (dentin) samples. This is consistent with previous reports of dental plaque being better at preserving aDNA from oral microorganisms than dentin [35]. Interestingly, we observed a clustering based on the substrates of 'red complex'-positive samples in the NMDS (electronic supplementary material figure S1). Dental calculus samples showed a more condensed cluster compared to dentin samples; this pattern has been identified previously in ancient remains [41], suggesting a low influence of exogenous microbes and a more uniform microbial composition. Interestingly, we identified that two dentin samples (TO-2417J and CO-09) clustered in dental calculus samples, mainly driven by the presence of *Actinomyces* sp. *oral taxon 894*, *Olsenella* and sp. *oral taxon 807*, *Anaerolineaceae bacterium oral taxon 439*, *Streptococcus gordonii* and *T. forsythia* (electronic supplementary material table S5). The latter two pathogens act as 'late colonizers', commonly found in dental calculus [31].

This observation is congruent with the ecological plaque hypothesis [42], which proposes that while certain oral pathogens, like *T. forsythia*, could be found in dental calculus as an innocuous member of the microbiome, their presence in dentin could be reflective of an infectious process involving periodontal disease. Further testing involving larger sample sizes with associated paleopathological indicators of periodontal disease is needed to unequivocally assess this hypothesis.

When only considering dentin, 3 out of 16 (18.7%) Pre-Hispanic individuals are positive for *T. forsythia*; a slightly lower proportion is observed for Colonial individuals (5 out of 33, 14.7%). Two out of the three Pre-Hispanic teeth that are positive for *T. forsythia* (TO-2417Q and CA-13) belong to individuals who were likely of high-ranking status according to their archaeological context. The burial of individual TO-2417Q at the site of the Toluquilla (in Queretaro State) displayed an hematite mirror and rich burial offerings (Elizabeth Mejía 2017, personal communication), while individual CA-13 was found with snail shell beads, a conch shell pectoral and pottery in the burial [43]. Both burial contexts suggest an upper social rank for these individuals (electronic supplementary material). Thus, high-ranking status could be linked to differential access to certain types of food that increased the presence of *T. forsythia* in these individuals [44]. It has been suggested that the excessive intake of fermentable carbohydrates results in dental biofilms experiencing extended periods of low pH [42,45], which can favour the proliferation of obligatory anaerobic species, like *T. forsythia*. Unfortunately,

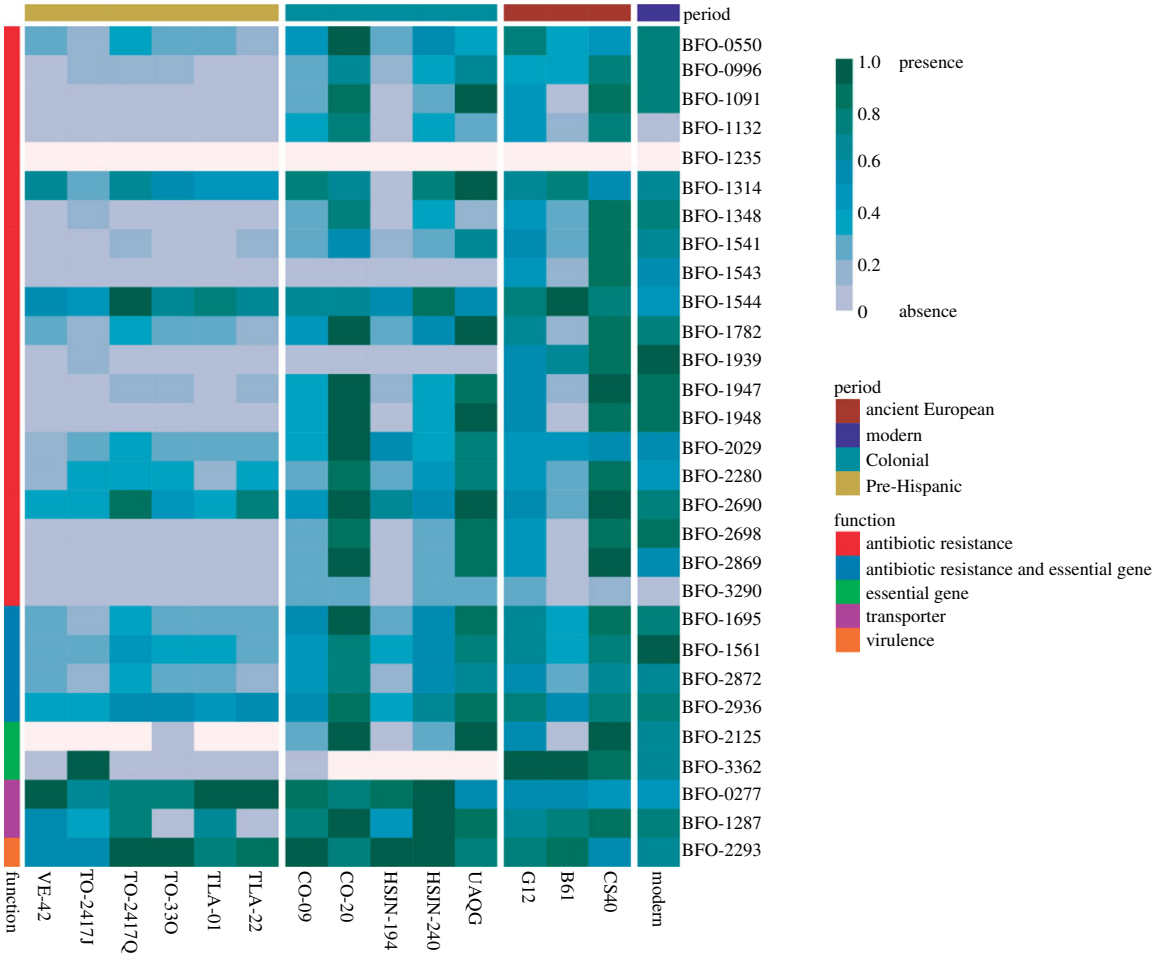


Figure 5. Presence/absence of selected essential, virulence and antibiotic resistance genes in ancient and modern *T. forsythia* genomes. Normalized average depth over the 234 captured genes is shown in a heatmap for a subset of 20 antibiotic resistance genes, 1 virulence factor, 4 essential and antibiotic resistance genes, 2 essential and 2 transporter genes (rows) in 6 Pre-Hispanic, 5 Colonial, 4 ancient European and one Modern *T. forsythia* sequences (columns). Normalized depth of the 234 genes of all individuals analysed is shown in figure S6 and table S9.

dental calculus, from which we could in principle characterize diet, was not available for these two individuals, therefore we cannot directly test this scenario. Nonetheless, these observations could inform future studies to directly test the potential associations of social status and diet with the presence of this pathogen.

To explore in depth the evolutionary relations between the different *T. forsythia* genomes, the depth of coverage of the ancient bacterial genomes was increased using a custom set of baits for targeted capture enrichment. This strategy, which targeted 234 genes, increased the on-target depth of coverage by one to two orders of magnitude and was effective regardless of the substrate. There is a trend of pre-capture libraries with shorter fragments (69–75 bp) that generally resulted in better fold enrichments. This is noteworthy given that the synthesized baits were 60 bp in length. Previous reports show that the length of baits influences the efficiency of enrichment by biasing the distribution of the read fragments obtained post-capture [46]. Although it remains to be tested how different factors, including bait length, can be adjusted to improve the efficiency of targeted capture-enrichment experiments, our approach proved quite useful for enriching libraries from different sources and with varying parameters of pre-capture endogenous content and average fragment lengths. Therefore, this study is a valuable precedent for future research examining oral health and periodontal disease in ancient human populations.

Characterizing a subset of genes at higher coverage allowed us to carry out phylogenetic analyses that revealed the existence of two clusters that distinguish the Pre-Hispanic and Colonial *T. forsythia* sequences. The Colonial samples cluster together with ancient European individuals and the two modern *T. forsythia* sequences (including the reference genome) from the USA, while all Pre-Hispanic individuals formed a separate monophyletic clade. This phylogenetic distribution of the sequences by period leads to interesting questions regarding the transmission dynamics of the newly introduced strain and whether a replacement of *T. forsythia* 'Pre-Hispanic strains' by 'Colonial strains' may have taken place. While no individual from the Colonial period carries a 'Pre-Hispanic strain', our sample size is too small to unambiguously assert this and future studies with larger sample sizes are needed to explicitly test this hypothesis. It is unclear how *T. forsythia* is transmitted; however, other periodontal pathogens (e.g. *Porphyromonas gingivalis* and *Actinobacillus actinomycetemcomitans*) follow a cohabitation-based transmission pattern [47]. These observations allow us to formulate hypotheses about the use of this pathogen as a proxy to infer past human interactions and should be addressed by future research. Additional studies characterizing ancient and present-day *T. forsythia* strains from several geographical regions and time periods could draw a more complete picture of these transmission dynamics and elucidate how integrating

knowledge of host ancestry and the putative geographical origins of their pathogenic strains could be used to infer human interaction networks.

When evaluating the presence/absence of the 234 captured genes, we observed a differential presence of two genes (*pncB* and *rpll*) annotated as having essential functions between Pre-Hispanic, Colonial and ancient European individuals.

The gene *pncB* is absent in all Pre-Hispanic genomes, except one (TO-333O). This gene is involved in NAD biosynthesis and in most bacterial species follows two pathways, one by genes *nadB-nadA-nadC* and a second one regulated by genes *pncA-pncB* [48]. While most bacteria have both pathways, some are strictly dependent on one or the other [48]. Our capture design included the three genes *nadB*, *nadA* and *nadC*, and we did not see absence of any of these. Therefore, we can speculate that *nadB-nadC* genes could be compensating for the lack of gene *pncB*. An additional plausible explanation is that in spite of *pncB* gene being annotated as essential in *T. forsythia* 92A2 reference genome, its enzymatic function could be compensated by other enzymes, a phenomenon that has been observed frequently among bacteria [49]. Lastly, gene *rpll* (ribosomal protein L9) is absent in all Colonial individuals except one (UAQG). Despite being a ribosomal protein and annotated as essential, this gene has also been found to be absent in other bacterial genomes [50].

The *tetQ* gene (BFO-1235) involved in resistance to antibiotic tetracycline is absent in all of the ancient and one of the modern genomes in this analysis. This finding replicates what was observed in the study of the two ancient samples from Dalheim, Germany that we included in our analyses [3]. This gene is present in *T. forsythia*'s reference genome and has been found at varying frequencies (8–80%) among periodontitis patients in Europe [51,52] and the USA [53], while in Latin American populations, it is present at high frequencies [54]. Our observations suggest that this gene might have arisen under recent selective pressures; however, this is an open question that remains to be evaluated once a larger dataset becomes available.

To evaluate the possibility that our capture design may have been unsuccessful at targeting and enriching some genes and mistakenly marking these as absent, we leveraged the whole-genome data for *T. forsythia* (13 X average depth from the higher depth sequenced pre-capture library only) for individual TO-2417Q. When applying the same presence/absence criteria on the whole-genome data and contrasting the results with those from the post-capture library, the same genes are identified as absent (electronic supplementary material, figure S6). Therefore, we can rule out any considerable bias introduced by the capture efficiency that would mistakenly render specific genes as missing. In addition, any systematic error in the probes would be revealed in having the same gene as missing in all the samples from this study, which is not observed.

Overall, while limited to only the genes considered in the capture design, the patterns identified with the differential presence of genes point to some of the evolutionary changes experienced by *T. forsythia* during a wide temporal range. Since this pathogen has accompanied humans for at least tens of thousands of years, it is not surprising that the genome of *T. forsythia* co-adapted to the changing environments (diet, lifestyle, oral hygiene, etc.) experienced by their human hosts as they migrated, interacted with other populations and modified their environments. It is also notable that in this study the antibiotic resistance gene *tetQ* is absent in all *T. forsythia* genomes, except in the reference one, which could reflect an early acquisition of this gene. This also reveals how the selective pressures posed by antibiotics can differ geographically and temporarily, and that the presence or absence of antibiotic resistance genes in samples with spatio-temporal diversity can reflect the mode and tempo of such pressures [3].

In conclusion, this study provides the first account of evolutionary patterns in the 'red complex' bacterium *T. forsythia* by leveraging ancient and present-day genomic data. The identification of this pathogen in Pre-Hispanic individuals from Mexico is indicative of the long-standing relationship between *T. forsythia* and its human host. Moreover, we were able to recover informative amounts of aDNA from this pathogen by means of a capture-enrichment strategy that considerably increased the depth of coverage for 234 informative genes. In turn, these data allowed us to dissect relevant evolutionary information and uncover the phylogenetic relationships between ancient sequences spanning the transition from Pre-Hispanic to Colonial times. This study, therefore, illustrates the potential use of ancient oral pathogen genomes as a proxy to infer patterns of social networks in the past.

Though limited by small sample size and the targeting of only a fraction of the genes in the *T. forsythia*'s genome, our results unveil many future directions for the study of this pathogen. Further characterization of whole ancient *T. forsythia* genomes from different spatio-temporal contexts and diverse populations can provide valuable data to validate the usability of this pathogen as a proxy for inferring past social interactions and to characterize in depth host–pathogen coevolutionary dynamics through extended periods of time.

7. Material and methods

(a) Sample provenance

Teeth and dental calculus samples ($N = 53$) were obtained from the archaeological collections of four Pre-Hispanic sites and three Colonial sites. The Pre-Hispanic individuals are from Cañada de la Vírgen, Guanajuato ($n = 11$ teeth); Toluquilla and Ranas, Querérato ($n = 1$ dental calculus, $n = 5$ teeth); Tepeticpac, Tlaxcala ($n = 2$ dental calculus); and Veracruz ($n = 1$, dental calculus). The Colonial individuals are from the Temple of the Immaculate Conception 'La Conchita', Mexico City, Mexico ($n = 7$ teeth), the Hospital San Jose de los Naturales, Mexico City ($n = 21$ teeth) and Ex College of San Ignacio de Loyola, Queretaro ($n = 5$ teeth) (figure 1 and table 1). Sample HSJN-194 was radiocarbon (C14) dated at the Accelerator Mass Spectrometry Laboratory (LEMA), at the Institute of Physics, National Autonomous University of Mexico (IF-UNAM) in Mexico City. Permits to carry out DNA analyses with these samples were granted by the Archaeology Council of the *Instituto de Antropología e Historia* (INAH) of Mexico. Additional information on the context of the samples is available in the electronic supplementary material.

(b) DNA extraction and library preparation

Samples were processed in a dedicated aDNA laboratory at the International Laboratory for Human Genome research (LIIGH-UNAM, Queretaro, Mexico), following strict procedures to avoid modern DNA contamination [55,56]. Dental plaque samples were collected using a dental scaler. Teeth and dental calculus samples were UV irradiated (256 nm) for 3 min using a UVP CL-1000 cross-linker. Teeth were cleaned with a 1% sodium hypochlorite solution. A Dremel tool was used to remove the outer surfaces of teeth and to slice transversely at

Table 1. Characteristics of the ancient individuals analysed in this study.

period	sample ID	substrate	date	archeological site	reference
Pre-Hispanic	CA-13	dentin	770-400 BCE*	Cañada de La virgen, Guanajuato, Mexico	this study
Pre-Hispanic	TO-2417Q	dentin	1257 ± 30 CE*	Toluquilla, Queretaro, Mexico	this study
Pre-Hispanic	TO-2417 J	dentin	900 BCE	Toluquilla, Queretaro, Mexico	this study
Pre-Hispanic	TO-3330	dental calculus	700 ± 50 BCE*	Toluquilla, Queretaro, Mexico	this study
Pre-Hispanic	VE-42	dental calculus	900 BCE–1200 CE	Tabuco, Veracruz, Mexico	this study
Pre-Hispanic	TLA-01	dental calculus	1451–1523 CE* †	Tepeticpac, Tlaxcala, Mexico	this study
Pre-Hispanic	TLA-22	dental calculus	1431–1479 CE* †	Tepeticpac, Tlaxcala, Mexico	this study
Colonial	CO-09	dentin	1700–1900 CE	Temple of the Immaculate Conception 'La Conchita', Mexico City, Mexico	this study
Colonial	CO-20	dentin	1700–1900 CE	Temple of the Immaculate Conception 'La Conchita', Mexico City, Mexico	this study
Colonial	HSJN-194	dentin	1472–1625 CE*	Hospital San Jose de los Naturales, Mexico City, Mexico	this study
Colonial	HSJN-240	dentin	1442–1608 CE*	Hospital San Jose de los Naturales, Mexico City, Mexico	this study
Colonial	UAQG	dentin	ND	Ex College of San Ignacio de Loyola, Queretaro, Mexico	this study
Ancient European	G12	dental calculus	950–1200 CE*	Dalheim, Germany	Warinner *et al.* [3]
Ancient European	B61	dental calculus	950–1200 CE*	Dalheim, Germany	Warinner et al. [3]
Ancient European	CS21	dental calculus	1770–1855 CE*	Radcliffe Infirmary Burial Ground collection, UK	Velsko *et al.* [31]
Ancient European	CS40	dental calculus	1770–1855 CE*	Radcliffe Infirmary Burial Ground collection, UK	Velsko *et al.* [31]

†López Aurelio and Santacruz Cano Ramón, Anexo 2: Fechamientos, Proyecto Arqueológico Tepeticpac, 2020. Informe de Actividades de Campo y Análisis de Materiales 2012–2018, Documento en Archivo Técnico del INAH, Ciudad de México.

C14 dating is indicated by asterisks (*).

ND, Not determined.

the cemento-enamel junction [57]. The roots were then covered in aluminium foil and pulverized using a hammer, as in [58]. The ancient DNA extraction for both teeth and dental calculus samples was carried out using a silica-based method [59,60], as described in [61], using approximately 200 mg of pulverized tooth or approximately 20 mg of calculus [3,41].

In order to detect contaminants in reagents or by human manipulation, extraction, library and PCR indexing blanks were processed in parallel. DNA extracts and extraction blanks were quantified with the Qubit 2.0 High Sensitivity assay [62]. For all the samples, we created barcoded (6 bp), double-stranded Illumina libraries, as previously reported [63]. Pre-capture and captured sample libraries, and 12 negative controls, carried along with DNA extraction, library preparation and PCR indexing were sequenced on an Illumina NextSeq550 (2 × 75 cycles) at the LANGEBIO's genomics core facility (National Laboratory of Genomics for Biodiversity, Irapuato, Guanajuato). Additionally, two pre-capture libraries (TO-2417Q and TO-2417J) were sequenced at higher depth to obtain whole genomes for *T. forsythia* on the NovaSeq instrument (S1, 2 × 100 cycles) available at the LANGEBIO's genomics core facility.

(c) Sequence data quality control and taxonomic classification

AdapterRemoval v. 2 [64] was used to process the fastq files to clip adapter sequences and merge sequence pairs (with at least 11 bp overlap). Only reads 30 bp and longer and with a quality above 33 were retained for downstream analyses. For pre-capture libraries, reads were first mapped to the human reference genome (version GRCh37, hg19) using 'bwa aln' option with (flags –l500 –t 16 –q 25) and the unmapped reads were then retrieved for downstream analyses. Unmapped reads from dentin, dental calculus and negative control libraries were then taxonomically binned with Kraken 2 [35], using only-classified-output and NCBI RefSeq bacterial, archaea and viral genomes as the reference database. To extract the counts assigned at the species level we used Kraken-biom (https://github.com/smdabdoub/kraken-biom). The species that were present at less than 0.02% relative abundance were removed from downstream analyses, as in [29]. Metagenomic profiles were analysed in-depth with Pavian [65] to identify known pathogens. The visualization of the relative abundance of taxonomic groups and non-metric multidimensional scaling (NMDS) was performed with the phyloseq library [66,67]. The taxonomic assignments were evaluated manually to identify known pathogens.

(d) Mapping and authenticity of ancient pathogen DNA

The reads from libraries that were *T. forsythia* positive were mapped to the *Tannerella forsythia* genome 92A2 (assembly NC_016610.1), *Porphyromonas gingivalis* ATCC 33277 (assembly NC_010729.1) and *Treponema denticola* ATCC 35405 (assembly NC_002967.9) using bwa aln option [68] (with the flags –l 1024 –n 0.03, –q 37). Clonal duplicates were removed using samtools rmdup [69]. Damage and fragmentation patterns were assessed using mapDamage2 [70] with default parameters. The coverage distribution of reads across the reference genome was visualized using Artemis [71].

Table 2. Identification of "red complex" bacteria in ancient individuals from Mexico.

			Tannerella forsythia		*Porphyromonas gingivalis*		*Treponema denticola*	
period	sample	total reads*	Kraken (count reads)	uniquely mapped reads	Kraken (count reads)	uniquely mapped reads	Kraken (count reads)	uniquely mapped reads
Pre-Hispanic	CA-13	4 983 380	2 303	1 733	422	236	890	464
Pre-Hispanic	TO-2417Q	7 070 465	14 559	18 867	4 490	7 422	2 760	2 767
Pre-Hispanic	TO-2417J	598 299 054	2 043 487	307 840	64 218	4 786	870 528	6 807
Pre-Hispanic	TO-3330	2 378 046	32 040	11 031	12 056	4 470	3 309	1 181
Pre-Hispanic	VE-42	2 377 859	5 295	1 206	1 096	532	2 623	743
Pre-Hispanic	TLA-01	2 383 724	10 998	3 268	2 445	737	3 235	1 069
Pre-Hispanic	TLA-22	2 383 724	7 464	1 992	2 157	263	2 597	629
Colonial	CO-09	19 446 371	40 547	6 796	8 852	1 779	19 777	3 920
Colonial	CO-20	15 681 711	5 493	4 344	1 323	920	876	767
Colonial	HSJN-194	5 321 515	8 650	1 065	736	4 600	2 641	1 803
Colonial	HSJN-240	4 773 622	1 770	4 600	616	1 065	2 136	1 305
Colonial	UAQG	7 329 193	3 275	2 438	31	14	305	140
Ancient European	G12	47 043 719	407 359	268 432	65 957	43 065	127 287	6 7028
Ancient European	B61	46 633 826	234 385	178 848	11 696	6 997	28 427	18 957
Ancient European	CS21	5 852 657	296 749	280 244	26 280	14 530	53 558	14 364
Ancient European	CS40	4 826 624	596 602	342 820	39 297	851	40 027	2 493

*Total unmapped reads to human.

(e) Capture-enrichment of *Tannerella forsythia* genes

A custom set of in-solution capture-enrichment baits was designed targeting 234 genes of *T. forsythia* based on genome assembly NC_016610.1. The 234 genes (258 696 total bp) were selected based on annotations in the Pathosystems Resource Integration Center (PATRIC) database [37] (electronic supplementary material table S5). The probes were designed by Arbor Biosciences (myBaits Custom kit, Ann Arbor, MI, USA) using their proprietary pipeline, which yielded *ca*. 20 000 probes, 60 bp in length, and with 6 x tiling density for each genomic region. Each bait candidate was aligned using BLAST [72] against the reference genomes of *Treponema denticola* (NC_002967.9), *Porphyromonas gingivalis* (NC_010729.1) and *Tannerella forsythia* (NC_016610.1). The hybridization melting temperature (Tm) was estimated for each hit assuming standard myBaits Custom kit buffers and conditions.

DNA capture was performed on the indexed libraries following the manufacturer's protocol (myBaits Kit protocol v4). Libraries were amplified with Phusion U Hot Start DNA Polymerase (Thermo Fisher Scientific), purified with SPRISelect Magnetic Beads (Beckman Coulter) and quantified with a Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA). Captured libraries were amplified for 12 cycles. Amplified libraries were pooled in equimolarity and sequenced on the NextSeq550 (2 × 75 cycles). Resulting fastq files were processed and mapped as described above for the pre-capture libraries, except that for post-capture libraries we mapped directly to *T. forsythia*'s reference genome and skipped the human genome mapping step.

We used BEDTools intersect [73] to determine the number of reads on target. The input to BEDTools intersect were the bam file and a file in bed format with the coordinates of the targeted genes; the output was a bam file with the intersecting reads. The output bam files were then processed with SAMtools 'depth' [69] function to calculate the genome depth and coverage of the alignments to the *T. forsythia* reference genome. To evaluate the success of the capture-enrichment strategy, the total numbers of unique reads mapping to the target regions before and after capture were compared (electronic supplementary material table S6). Since different numbers of reads were obtained in the pre- and post-capture sequencing runs, an equal number of reads (the lowest between the pre- and post-capture runs) for each individual were subsampled using seqtk (https://github.com/lh3/seqtk). The genome coverage was calculated in 100 bp windows and was plotted using Circos [74].

(f) Phylogenetic analysis

Consensus sequences for six Pre-Hispanic Mexico (CA-13, TO-2417Q, TO-2417J, TO-333O, VE-42, TLA-01, TLA-22); five Colonial Mexico (CO-09, CO-20, HSJN-194, HSJN-240, UAQG); four ancient European (G12, B61, CS21, CS40) and two publicly available modern *T. forsythia* genomes (modern and reference) [33,39] were generated using ANGSD [75] (flags -doCounts 1 -doFasta 2), for 170 orthologous genes between *T. forsythia* and *P. gingivalis ATCC 33277* (NC_010729.1). This set of 170 genes was defined by selecting the orthologous genes between both species in Orthologs Database (OrthologueDB) [76] that were also present in the custom capture-enrichment design. Gene consensus sequences were concatenated and aligned using MAFFT [77], with 1000 as the maximum number of iterations. The poorly aligned sequences were trimmed using trimAl [78], with a strict set of parameters, based on an automatically selected similarity threshold. trimAl uses the residue similarity scores distribution from the MSA (multiple sequence alignments) and selects the values at percentiles 20 and 80 of the alignment length. The lower and upper boundaries allow the retention of the 20% most conserved columns in the alignment, whereas the 20% most dissimilar columns are discarded [78]. RaXML [79] was used to construct a maximum-likelihood phylogeny based on the multiple sequence alignment, with the GTRGAMMA substitution model [79] and 1000 bootstrap replicates.

(g) Gene presence/absence

Reads were re-mapped to the *T. forsythia* genome (assembly NC_016610.1) using bwa aln [68] with the flags –l 1024 –n 0.03, –q 0 as in [6]. The rationale is when BWA identifies a read that maps to more than one location it randomly selects one and assigns it a 0 mapping quality, so when filtering by quality one it would disregard these reads, which in turn could increase the chances of a gene with paralogous sequences in the genome being considered as absent. Coverage at the 234 genes selected for enrichment was calculated using the BEDTools [73]. We used a min–max normalization to reduce the scale values to 0 and 1 and make data comparable between individuals. The following formula was used:

$$Yi = \frac{Xi - \min(x)}{\max(x) - \min(x)},$$

where min and max are the minimum and maximum values in X given its range. So Xi converts to Yi. We considered genes to be absent only if they had a normalized depth value of 0. These criteria are equivalent to defining a gene as absent if its depth is zero; however, the normalization step allows us to contrast samples with very different average depth ranges across genes (figure 5, figures S5–S8). The normalized depth of each gene in the ancient and modern genomes in our dataset was plotted using the pheatmap library [80] in R [67].

Ethics. The handling of the archeological remains analysed in this study was based on the Protocols for the Conservation and Protection of Cultural Heritage of Mexico. The permissions for the analysis of samples were acquired in accordance with the regulations determined by the Archaeological Council of the National Institute of Anthropology and History through the Ministry of Culture in Mexico. The official notice number of the permissions granted to analyse the individuals from the Hospital San Jose de los Naturales, Temple of Immaculate Conception 'La Conchita' and Tepeticpac, Tlaxcala collections are 401.1S.3-2018/1373, 401.1S.3.2020/1310 and 401.1S.3/2017/1995, respectively. The permit to analyze the individual from Tabuco, Veracruz was included in the project IN302219 PAPIIT-DGAPA-UNAM, headed by Carlos Serrano Sanchez.

Data accessibility. Raw data (fastq files) are available upon request. Mapped reads (bam files) that support this study have been deposited to the NCBI Short Read Archive (SRA) under the project accession PRJNA660267.

Authors' contributions. M.B.-L designed experiments, performed experiments, analysed data and drafted the manuscript; V.V.-I. performed experiments and analysed data; C.R.A. analysed data; A.B.V.A. analysed data; A.G.S. performed experiments; M.S.-V. designed experiments; K.A., J.G.V, A.H., E.M., A.M.M., J.O., J.P.-P., K.S., J.K.W., G.Z. and A.LC. provided archaeological samples and provided archeological information. M.A.N.-C. designed experiments; K.H.I. analysed data; S.R. designed data analyses; M.C.Á.-A. conceived of the investigation, designed experiments, analysed data and drafted the manuscript. All authors contributed to and approved the final version of the manuscript

Competing interests. We declare we have no competing interests.

Funding. The project was funded by the Wellcome Trust Seed Award in Science 208934/Z/17/Z, and by project IA201219 PAPIIT-DGAPA-UNAM. We thank the PhD program of Biological Sciences at UNAM and CONACyT for providing funding to M.B.-L.

Acknowledgements. We are grateful to Diego Ortega Del Vecchyo and Barbara Moguel for comments on the manuscript, as well as to Luis Alberto Aguilar Bautista, Alejandro de León Cuevas, Carlos Sair Flores Bautista and Jair Garcia Sotelo of the Laboratorio Nacional de Visualización Científica Avanzada for IT support. We thank Alejandra Castillo Carbajal and Carina Uribe Díaz for technical support throughout the project. We also thank to Ernesto Garfias Morales for his support on figure editing and to María Guadalupe Trejo Arellano for advising on data analyses. Simon Rasmussen was supported by the Novo Nordisk Foundation (NNF14CC0001).

References

1. Grauer AL. 2018 A century of paleopathology. *Am. J. Phys. Anthropol.* **165**, 904–914. (doi:10.1002/ajpa.23366)
2. Pinhasi R, Mays S (eds). 2008 Advances in human palaeopathology. Chichester, UK: John Wiley & Sons.
3. Warinner C *et al.* 2014 Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* **46**, 336–344. (doi:10.1038/ng.2906)
4. Bos KI *et al.* 2014 Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497. (doi:10.1038/nature13591)
5. Rasmussen S *et al.* 2015 Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. *Cell* **163**, 571–582. (doi:10.1016/j.cell.2015.10.009)
6. Vågene ÅJ *et al.* 2018 *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2**, 520–528. (doi:10.1038/s41559-017-0446-6)
7. Schuenemann VJ *et al.* 2018 Historic *Treponema pallidum* genomes from Colonial Mexico retrieved from archaeological remains. *PLoS Negl. Trop. Dis.* **12**, e0006447. (doi:10.1371/journal.pntd.0006447)
8. Rascovan N, Sjögren K-G, Kristiansen K, Nielsen R, Willerslev E, Desnues C, Rasmussen S. 2019 Emergence and spread of basal lineages of *Yersinia pestis* during the Neolithic decline. *Cell* **176**, 295–305. (doi:10.1016/j.cell.2018.11.005)
9. Bos KI *et al.* 2016 Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *eLife* **5**, 1–11. (doi:10.7554/elife.12994.010)
10. Spyrou MA *et al.* 2018 Analysis of 3800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague. *Nat. Commun.* **9**, 2234. (doi:10.1038/s41467-018-04550-9)
11. Adler CJ, Haak W, Donlon D, Cooper A. 2011 Survival and recovery of DNA from ancient teeth and bones. *J. Archaeol. Sci.* **38**, 956–964. (doi:10.1016/j.jas.2010.11.010)
12. Margaryan A *et al.* 2018 Ancient pathogen DNA in human teeth and petrous bones. *Ecol. Evol.* **8**, 3534–3542. (doi:10.1002/ece3.3924)
13. Adler CJ *et al.* 2013 Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat. Genet.* **45**, 450–455. (doi:10.1038/ng.2536)
14. Damgaard PB, Margaryan A, Schroeder H, Orlando L, Willerslev E, Allentoft ME. 2015 Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.* **5**, 11184. (doi:10.1038/srep11184)
15. Brealey JC, Leitão HG, van der Valk T, Xu W, Bougiouri K, Dalén L, Guschanski K. 2019 A roadmap to mammalian oral microbiome evolution with dental calculus. *bioRxiv*. (doi:10.1101/59679)
16. Haque M, Sartelli M, Haque SZ. 2019 Dental infection and resistance—global health consequences. *Dent. J.* **7**, 22. (doi:10.3390/dj7010022)
17. Lukacs JR. 2012 Oral health in past populations: context, concepts and controversies. In *A companion to paleopathology* (ed. AL Grauer), pp. 553–581. Chichester, UK: Wiley-Blackwell. (doi:10.1002/9781444345940.ch30)
18. Larsen CS. 1995 Biological changes in human populations with agriculture. *Annu. Rev. Anthropol.* **24**, 185–213. (doi:10.1146/annurev.an.24.100195.001153)
19. Lukacs JR. 1992 Dental paleopathology and agricultural intensification in south Asia: new evidence from Bronze Age Harappa. *Am. J. Phys. Anthropol.* **87**, 133–150. (doi:10.1002/ajpa.1330870202)
20. Bui FQ, Almeida-da-Silva CLC, Huynh B, Trinh A, Liu J, Woodward J, Asadi H, Ojcius DM. 2019 Association between periodontal pathogens and systemic disease. *Biomed. J.* **42**, 27–35. (doi:10.1016/j.bj.2018.12.001)
21. Könönen E, Gursoy M, Gursoy UK. 2019 Periodontitis: a multifaceted disease of tooth-supporting tissues. *J. Clin. Med.* **8**, 1135. (doi:10.3390/jcm8081135)
22. Helmi MF, Huang H, Goodson JM, Hasturk H, Tavares M, Natto ZS. 2019 Prevalence of periodontitis and alveolar bone loss in a patient population at Harvard School of Dental Medicine. *BMC Oral Health* **19**, 254. (doi:10.1186/s12903-019-0925-z)
23. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent Jr RL. 1998 Microbial complexes in subgingival plaque. *J. Clin. Periodontol.* **25**, 134–144. (doi:10.1111/j.1600-051X.1998.tb02419.x)
24. Khan SA, Kong EF, Meiller TF, Jabra-Rizk MA. 2015 Periodontal diseases: bug induced, host promoted. *PLoS Pathog.* **11**, e1004952. (doi:10.1371/journal.ppat.1004952)
25. Cullinan M, Ford P, Seymour G. 2009 Periodontal disease and systemic health: current status. *Aust. Dent. J.* **54**, S62–S69. (doi:10.1111/j.1834-7819.2009.01144.x)
26. Sudhakara P, Gupta A, Bhardwaj A, Wilson A. 2018 Oral dysbiotic communities and their implications in systemic diseases. *Dent. J.* **6**, 10. (doi:10.3390/dj6020010)
27. Sakakibara J, Nagano K, Murakami Y, Higuchi N, Nakamura H, Shimozato K, Yoshimura F. 2007 Loss of adherence ability to human gingival epithelial cells in S-layer protein-deficient mutants of *Tannerella forsythensis*. *Microbiology* **153**, 866–876. (doi:10.1099/mic.0.29275-0)
28. Sekot G, Posch G, Messner P, Matejka M, Rausch-Fan X, Andrukhov O, Schäffer C. 2011 Potential of the *Tannerella forsythia* S-layer to delay the immune response. *J. Dent. Res.* **90**, 109–114. (doi:10.1177/0022034510384622)
29. Sára M, Sleytr UB. 2000 S-Layer proteins. *J. Bacteriol.* **182**, 859–868. (doi:10.1128/JB.182.4.859-868.2000)
30. Divaris K. 2019 Searching deep and wide: advances in the molecular understanding of dental caries and periodontal disease. *Adv. Dent. Res.* **30**, 40–44. (doi:10.1177/0022034519877387)
31. Velsko IM *et al.* 2019 Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. *Microbiome* **7**, 102. (doi:10.1186/s40168-019-0717-3)
32. Weyrich LS *et al.* 2017 Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* **544**, 357–361. (doi:10.1038/nature21674)
33. Stafford GP, Chaudhuri RR, Haraszthy V, Friedrich V, Schäffer C, Ruscitto A, Honma K, Sharma A. 2016 Draft genome sequences of three clinical isolates of *Tannerella forsythia* isolated from subgingival plaque from periodontitis patients in the United States. *Genome Announc.* **4**(6), e01286-16. (doi:10.1128/genomeA.01286-16)
34. Palka JW. 2009 Historical archaeology of indigenous culture change in mesoamerica. *J. Archaeol. Res.* **17**, 297. (doi:10.1007/s10814-009-9031-0)
35. Wood DE, Lu J, Langmead B. 2019 Improved metagenomic analysis with Kraken 2. *bioRxiv*. 762302. (doi:10.1101/762302)
36. Gansauge M-T, Meyer M. 2014 Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res.* **24**, 1543–1549. (doi:10.1101/gr.174201.114)
37. Wattam AR *et al.* 2014 PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* **42**, D581–D591. (doi:10.1093/nar/gkt1099)
38. Friedrich V, Pabinger S, Chen T, Messner P, Dewhirst FE, Schäffer C. 2015 Draft genome sequence of *Tannerella forsythia* type strain ATCC 43037. *Genome Announc.* **3**(3), e00660-15. (doi:10.1128/genomeA.00660-15)
39. Hanson-Drury S, To TT, Liu Q, Vo AT, Kim M, Watling M, Bumgarner RS, McLean JS. 2017 Draft genome sequence of *Tannerella forsythia* clinical isolate 9610. *Genome Announc.* **5**, e00024-17. (doi:10.1128/genomeA.00024-17)
40. Wesp JK. 2017 Caring for bodies or simply saving souls: the emergence of institutional care in Spanish Colonial America. In *New developments in the bioarchaeology of care: further case studies and expanded theory* (eds L Tilley, AA Schrenk), pp. 253–276. Cham, Switzerland: Springer International Publishing.
41. Mann AE *et al.* 2018 Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. *Sci. Rep.* **8**, 9822. (doi:10.1038/s41598-018-28091-9)
42. Naginyte M, Do T, Meade J, Devine DA, Marsh PD. 2019 Enrichment of periodontal pathogens from the biofilms of healthy adults. *Sci. Rep.* **9**, 5491. (doi:10.1038/s41598-019-41882-y)
43. Zepeda García Moreno G. 2005 Cañada de la Virgen, Allende, Guanajuato 'La casa de los trece cielos'. *Arqueología Mexicana* **73**, 54–57.

44. Lourdes MM. 2012 Vida urbana y salud en la Mesoamérica prehispánica. *Boletín de Antropológia. Universidad de Antiquia* **26**, 215–238.
45. Nyvad B, Takahashi N. 2020 Integrated hypothesis of dental caries and periodontal diseases. *J. Oral Microbiol.* **12**, 1710953. (doi:10.1080/20002297.2019.1710953)
46. Ávila-Arcos MC, Sandoval-Velasco M, Schroeder H, Carpenter ML, Malaspinas A-S, Wales N, Peñaloza F, Bustamante CD, Gilbert MTP. 2015 Comparative performance of two whole-genome capture methodologies on ancient DNA Illumina libraries. *Methods Ecol. Evol* **6**, 725–734. (doi:10.1111/2041-210X.12353)
47. Van Winkelhoff AJ, Boutaga K. 2005 Transmission of periodontal bacteria and models of infection. *J. Clin. Periodontol* **32**, 16–27. (doi:10.1111/j.1600-051X.2005.00805.x)
48. Sorci L *et al.* 2009 Nicotinamide mononucleotide synthetase is the key enzyme for an alternative route of NAD biosynthesis in *Francisella tularensis*. *Proc. Natl Acad. Sci. USA* **106**, 3083–3088. (doi:10.1073/pnas.0811718106)
49. Ghosh S, O'Connor TJ. 2017 Beyond paralogs: the multiple layers of redundancy in bacterial pathogenesis. *Front. Cell. Infect. Microbiol.* **7**, 467. (doi:10.3389/fcimb.2017.00467)
50. Yutin N, Puigbò P, Koonin EV, Wolf YI. 2012 Phylogenomics of prokaryotic ribosomal proteins. *PLoS ONE* **7**, e36972. (doi:10.1371/journal.pone.0036972)
51. Ioannidis I, Sakellari D, Spala A, Arsenakis M, Konstantinidis A. 2009 Prevalence of *tetM*, *tetQ*, *nim* and bla_{TEM} genes in the oral cavities of Greek subjects: a pilot study. *J. Clin. Periodontol.* **36**, 569–574. (doi:10.1111/j.1600-051X.2009.01425.x)
52. Freitas d, Gonçalves L, Coelho MJ, Ferraz MP, Magalhães R, Pina C, Cardoso IL. 2019 Prevalence of antibiotic (β-lactams, tetracycline, metronidazole, erythromycin) resistance genes in periodontic infections. *Clin Res Trials* **5**, 1–4. (doi:10.15761/CRT.1000280)
53. Kim S-M, Kim HC, Lee S-WS. 2011 Characterization of antibiotic resistance determinants in oral biofilms. *J. Microbiol.* **49**, 595–602. (doi:10.1007/s12275-011-0519-1)
54. Collins JR, Arredondo A, Roa A, Valdez Y, León R, Blanc V. 2016 Periodontal pathogens and tetracycline resistance genes in subgingival biofilm of periodontally healthy and diseased Dominican adults. *Clin. Oral Investig.* **20**, 349–356. (doi:10.1007/s00784-015-1516-2)
55. Cooper A, Poinar HN. 2000 Ancient DNA: do it right or not at all. *Science* **289**, 1139. (doi:10.1126/science.289.5482.1139b)
56. Willerslev E, Cooper A. 2005 Ancient DNA. *Proc. R. Soc. B* **272**, 3–16. (doi:10.1098/rspb.2004.2813)
57. Cruz-Dávalos DI, Nieves-Colón MA, Sockell A, Poznik GD, Schroeder H, Stone AC, Bustamante CD, Malaspinas A-S, Ávila-Arcos MC. 2018 In-solution Y-chromosome capture-enrichment on ancient DNA libraries. *BMC Genomics* **19**, 608. (doi:10.1186/s12864-018-4945-x)
58. Schuenemann VJ *et al.* 2011 Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proc. Natl Acad. Sci. USA* **108**, E746–E752. (doi:10.1073/pnas.1105107108)
59. Rohland N, Hofreiter M. 2007 Comparison and optimization of ancient DNA extraction. *Biotechniques* **42**, 343–352. (doi:10.2144/000112383)
60. Dabney J *et al.* 2013 Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15 758–15 763. (doi:10.1073/pnas.1314445110)
61. Sandoval-Velasco M, Lundstrøm IKC, Wales N, Ávila-Arcos MC, Schroeder H, Gilbert MTP. 2017 Relative performance of two DNA extraction and library preparation methods on archaeological human teeth samples. *STAR: Sci. Technol. Archaeol. Res.* **3**, 80–88. (doi:10.1080/20548923.2017.1388551)
62. Simbolo M, Gottardi M, Corbo V, Fassan M, Mafficini A, Malpeli G, Lawlor RT, Scarpa A. 2013 DNA qualification workflow for next generation sequencing of histopathological samples. *PLoS ONE* **8**, e62692. (doi:10.1371/journal.pone.0062692)
63. Meyer M, Kircher M. 2010 Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448. (doi:10.1101/pdb.prot5448)
64. Schubert M, Lindgreen S, Orlando L. 2016 AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88. (doi:10.1186/s13104-016-1900-2)
65. Breitwieser FP, Salzberg SL. 2020 Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics* **36**, 1303–1304. (doi:10.1093/bioinformatics/btz715)
66. McMurdie PJ, Holmes S. 2013 phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217. (doi:10.1371/journal.pone.0061217)
67. R Core Team. 2018 R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. (https://www.Rproject.org/)
68. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760. (doi:10.1093/bioinformatics/btp324)
69. Li H *et al.* 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
70. Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. 2011 mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* **27**, 2153–2155. (doi:10.1093/bioinformatics/btr347)
71. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012 Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**, 464–469. (doi:10.1093/bioinformatics/btr703)
72. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. (doi:10.1016/S0022-2836(05)80360-2)
73. Quinlan AR, Hall IM. 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842. (doi:10.1093/bioinformatics/btq033)
74. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009 Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645. (doi:10.1101/gr.092759.109)
75. Korneliussen TS, Albrechtsen A, Nielsen R. 2014 ANGSD: Analysis of next generation sequencing data. *BMC Bioinf.* **15**, 356. (doi:10.1186/s12859-014-0356-4)
76. Whiteside MD, Winsor GL, Laird MR, Brinkman FSL. 2013 OrthologueDB: a bacterial and archaeal orthology resource for improved comparative genomic analysis. *Nucleic Acids Res.* **41**, D366–D376. (doi:10.1093/nar/gks1241)
77. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
78. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009 trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973. (doi:10.1093/bioinformatics/btp348)
79. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (doi:10.1093/bioinformatics/btu033)
80. Kolde R. 2012 Pheatmap: pretty heatmaps. *R package version*.

Supplementary Information

Paleogenomic insights into the red complex bacteria *Tannerella forsythia* in Pre-Hispanic and Colonial individuals from Mexico

Miriam Bravo-Lopez[1], Viridiana Villa-Islas[1], Carolina Rocha Arriaga[1], Ana B. Villaseñor Altamirano[1], Axel Guzmán-Solís[1], Marcela Sandoval-Velasco[2], Julie K. Wesp[3], Keitlyn Alcantara[4], Aurelio Lopez-Corral[5], Jorge Gomez-Valdes[6], Elizabeth Mejia[7], Alberto Herrera[7], Alejandro Meraz-Moreno[8], María de la Luz Moreno-Cabrera[8], Andrés Moreno-Estrada[9], Maria A. Nieves-Colón[9,10], Joel Olvera[11], Julia Perez-Perez[6], Katrine Højholt Iversen[12], Simon Rasmussen[12], Karla Sandoval[9], Gabriela Zepeda[13] and Maria C. Ávila-Arcos[1].

[1]International Laboratory for Human Genome Research, National Autonomous University of Mexico (UNAM), Queretaro, Mexico
[2]Section for Evolutionary Genomics, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
[3]Department of Sociology and Anthropology, North Carolina State University, USA
[4]Department of Anthropology, Indiana University-Bloomington, Bloomington, IN, USA
[5]Department of Archeology, National Institute of Anthropology and History, Tlaxcala, Mexico.
[6]National School of Anthropology and History, Mexico City, Mexico
[7]National Institute of Anthropology and History-Queretaro, Mexico
[8]National Institute of Anthropology and History-Mexico City, Mexico
[9]National Laboratory of Genomics for Biodiversity, Unit of Advanced Genomics (LANGEBIO), CINVESTAV, Irapuato, Guanajuato 36824, Mexico.
[10]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA.
[11]Graduate Program of Physical Anthropology, National School of Anthropology and History, Mexico City
[12]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
[13]National Institute of Anthropology and History-Guanajuato, México

CONTENT

1. Samples and archeological sites description

1.1. Pre-Hispanic archeological sites

1.1.1. Cañada de la Virgen, Guanajuato

The Cañada de la Virgen Pre-Hispanic settlement is situated in the Mexican state of Guanajuato, Mexico's central highlands. Cañada de la Virgen was an important ceremonial site in the region, as revealed by the architectural features of monuments built according to astronomical criteria, religious beliefs and agricultural cycles.(Zepeda and Barrales 2008) Nineteen burials were excavated from Cañada de la Virgen, of which eleven were recovered from Complex A, a space considered to be designated for celestial observation. At the top of this complex is the Red Temple, a large pyramidal base topped with a structure, from which individual CA-13 was excavated (Moreno 2008).

Individual CA-13 was located at the top of the Complex A, and according to the osteological analyses, he was a ~50-year-old male who could have undergone a preservation treatment to be praised as a relic in a funerary bulge and revered by generations. Paleopathological examination of this individual suggests the presence of cribra orbitalia and porotic hyperostosis. Additionally, the skeletal remains showed a bone injury or trauma in the facial region, as well an antemortem nasal fracture with deviation of the vomer bone and nasal spine (Zepeda and Barrales 2008).

1.1.2. Toluquilla, Queretaro

Toluquilla is a Pre-Hispanic settlement located in the Mountain range in the state of Queretaro, in Central Mexico. The occupational range of this archeological site is larger compared to other Pre-Columbian societies (300 BCE to 1,500 CE). Individuals TO-2417Q and TO-2417J were excavated from this site.

Individual TO-2417Q was identified as a 40-50-year-old male, with a severe degree of dental caries. Interestingly, the objects found in its burial were two isolated bones, a vertebrate and coxal bones, and a pyrite mosaic that would have formed a mirror (Elizabeth Mejía, personal communication, December 3, 2019). Pyrite mirrors have been associated to high-ranking individuals, mostly found in the center of Mexico (Rojo and Murrieta, n.d.).
Individual TO-2417J was identified to have lumbar osteophytes and bone tumors in both parietal lobes.

1.1.3. Tepeticpac, Tlaxcala

The archaeological site of Tepeticpac, from which samples TLA-1 and TLA-22 were excavated, is located in the present-day state of Tlaxcala in Central Mexico. The constructions of the site correspond to the late Post-Classic period, from the XIII century to the arrival of the Spaniards. It is characterized by the presence of 150 terraces, 15 residential buildings and five government centers. The site was founded by a group of migrants called Chichimecas around the XIII century, who invaded the center of Tlaxcala and managed to expand their territory and consolidate the political and military hegemony in the region. Individuals TLA-1 and TLA-22 were located in the area known as “La plaza de la Terraza” (Operation 13; N22E82 and N24E82; -.27m/-.57 m) (Lopez Corral et al. 2019; Cano and Corral 2015).

Individual TLA-1 corresponds to a male individual, with an estimated age between 30-40 years at the time of death. According to the paleopathological analysis, both parietal bones showed evidence of severe porotic hyperostosis. Regarding oral pathologies, it was identified severe wear of the occlusal surface of teeth, as well as a generalized brown coloring (moderate fluorosis). No injuries or skeletal trauma were observed in the skull (Russel Alcántara, personal communication, May 3, 2020).

Individual TLA-22 was identified as a 40-50-year-old male. Additional objects found in this burial included fragments of a brazier with pigments, a burned adult human bone and some bones of a subadult individual. The paleopathological analysis of TLA-22 revealed that the individual presented porotic hyperostosis in the frontal bone, in the area of the coronal suture and also in the occipital bone, close to the lambdoid suture. Regarding his oral health, the first upper right molar and the third upper right maxillary molar showed a severe carious affection, and there was antemortem loss of the second upper right molar. In the upper section of the left side, the second premolar had severe occlusal caries that even destroyed the crown. Additionally, a possible perimortem trauma of the right parietal bone was identified.

1.1.4. Tabuco, Veracruz

The Tabuco archaeological site, where individual VE-42 was excavated, is located in the surroundings of the Tuxpan river in the state of Veracruz; approximately 5 km from the coastal zone. It is considered as an important seaport of the Mesoamerican Post-Classic period (Lira López and Sánchez 2004). The VE-42 burial belongs to a 40-45-year-old male.

1.2. Colonial archeological sites

1.2.1. Temple of the Immaculate Conception “La Conchita”, Mexico City

The Temple of the Immaculate Conception is located in the south of Mexico City. It is a baroque style building constructed in the mid-eighteenth century, though there are records suggesting the construction of a Chapel at the same site as early as the sixteenth century. The skeletal remains analyzed in this study (CO-09 and CO-20) were recovered in the archaeological rescue project of the Temple of the Immaculate Conception between 2012 and 2013. The remains were found in secondary deposits that, archeologically, date to between the eighteenth and late nineteenth centuries.

1.2.2. Hospital Real de San Jose de los Naturales, Mexico City

The Hospital Real San José de los Naturales (HSJN) was located in the center of Mexico City. Individuals HSJN-194 and HSJN-240 were excavated from this site. The HSJN was founded by order of King Philip II in 1531 and closed in 1822. The HSJN was dedicated to serving only the Indigenous population.

Individual HSJ-194 is a middle adult (between 35–49 years old) male. The osteological analyses show evidence of moderate biomechanical stress, including osteoarthritic bone growth on the upper and lower limbs. Regarding the analyses of oral pathologies, he has moderate amounts of dental wear and dental calculus (calcified plaque), as well as a small degree of periodontal disease with at least two maxillary teeth that were lost *antemortem*.

Individual HSJN-240 is also a middle male adult, with moderate amounts of biomechanical stress on his upper limb and early signs of osteoarthritic changes on his lower lumbar vertebrae. He has light amounts of dental wear and dental calculus; however, three maxillary teeth were lost *antemortem* and another molar has a severe cavity that would have likely resulted in another tooth loss.

1.2.3. Ex College of San Ignacio de Loyola, Queretaro

The Ex College is considered a historical building of the Jesuit order, located in the Mexican city of Santiago de Querétaro in the State of Querétaro. The excavation of this site was part of a rescue project, in which 29 individuals were recovered. The UAQG individual, who was excavated from this site, had a moderate degree of dental abscess and a deviation of the lower extremities. Additionally, UAQG had a shovel shape of the upper incisors, a feature common in Native Americans (Mejía Pérez, 2019).

2. Taxonomic assignment of blanks and contamination assessment

Along with the processing of the ancient samples analyzed in this study, from DNA extraction, library preparation, and PCR indexing, we processed negative controls (blanks); we carried out NGS for 12 libraries for which we report results. We analysed these data to get taxonomic assignments (Table S4) and mapping statistics to *T. forsythia 92A2* (Table S6).

Three out of the twelve blanks had some reads mapping to *T. forsythia* 92A2 (1, 11 and 30 in blanks CO-LIB, TO-iPCR and HSJN-DNA, respectively). For two of these, CO-LIB and TO-iPCR, the percent of reads mapping to *T. forsythia*'s genome was 0.08% and 0.21% respectively. Of note, these values are higher than the values observed for some of the ancient libraries (Table S6). For library HSJN-DNA, however, the value was 0.01%, which is lower than any observed for other libraries.

To gain insight into the extent at which contamination could be affecting our observations, we explored the fraction of reads assigned to *T. forsythia 92A2* by Kraken2 (Wood *et al.,* 2019) in all the blanks (Table S4). We followed this strategy since the percentage of endogenous DNA in libraries with different amounts of starting reads can be misleading (because of clonality). Also, because the mapping itself can cause some bias by only including one reference; whereas the taxonomic classification by Kraken2 provides a reference database (with thousands of sequences and also different *T. forsythia* strains), hence a less biased account of the number of likely *T. forsythia 92A2* reads. All blanks had 0 reads assigned to *T. forsythia 92A2* by Kraken2, except TO-iPCR which had 6 reads (out of 5209, 0.11%) and HSJN-DNA with 1 out of 202,267 (0.0009%). We considered that the one read for library HSJN-DNA could be disregarded, but we decided to look in more detail at the libraries that were processed together with TO-iPCR, i.e. all the "Toluquilla" libraries. Apart from library TO-2417Q (in which we detected *T. forsythia* DNA and processed further) only one library, TO-11R, had just one read (0.00001%) assigned to *T. forsythia 92A2.* It is important to mention that sample TO-2417J, which is also from Toluquilla, was processed two years later and blanks were not sequenced. Because of the previous observations, we think only the blank TO-iPCR could be a reason of concern, and since it belongs to an indexing PCR blank, we could assume some cross-contamination likely arose at the PCR setup step when processing the "Toluquilla" libraries. We could also assume this most likely stemmed from library TO-2417Q, but since no other library in the batch showed a considerable fraction of *T. forsythia 92A2* reads (just one read in one library), we have reasons to think that this had no effect on our analysis and results.

References

Cabrera, María de la Luz Moreno, and Alejandro Meraz Moreno. n.d. “Las Investigaciones Arqueológicas Y Nuevos Descubrimientos En Coyoacán, DF.” http://www.academia.edu/download/34372412/Moreno_y_Meraz_-_Las_investigaciones_arqueologicas_y_nuevos_descubrimientos_en_Coyoacan__D.F.pdf.

Cano, Ramón Santacruz, and Aurelio López Corral. 2015. “Tepeticpac, Conservación Integral de Un Sitio Arqueológico En Tlaxcala, México.” In *México. Restauración Y Protección Del Patrimonio Cultural*, 199–223. Aracne Editrice.

Lira López, Yamile, and Carlos Serrano Sánchez. 2004. “Prácticas Funerarias En La Costa Del Golfo de México.” México, Universidad Veracruzana/UNAM/AMAB.

Lopez Corral, Aurelio, Ivonne Velasco Almanza, Thania E. Ibarra Narvaez, and Ramon Santacruz Cano. 2019. “Iconography and Collective Government during the Late Postclassic Period in Tepeticpac and Tlaxcallan, Mexico.” *LATIN AMERICAN ANTIQUITY* 30 (2): 333–53.

Moreno, Gabriela Zepeda García. 2008. “Cañada de La Virgen, Guanajuato.” *Arqueología Mexicana* 16 (92): 48–51.

Rojo, Danna Levin, and Cynthia Radding Murrieta. n.d. *The Oxford Handbook of Borderlands of the Iberian World*. Oxford University Press.

Wood DE, Lu J, Langmead B. 2019 Improved metagenomic analysis with Kraken 2. bioRxiv. , 762302. (doi:10.1101/762302)

Zepeda, G., and D. Barrales. 2008. “Arqueología E Identidad En Cañada de La Virgen, Guanajuato.” *Tiempo Y Región, Estudios Históricos Y Sociales, Ana*.

6. Capítulo 3. Reporte de proyecto. Recovery of an ancient *Salmonella enterica* genome from a Colonial individual from Mexico City.

Miriam Bravo-Lopez, Åshild Joanne Vågene, Carolina Rocha-Arriaga,Viridiana Villa-Islas, Maria de la Luz Moreno Cabrera, Alejandro Meraz Moreno, Julie Wesp, Marcela Sandoval-Velasco, Jorge Gomez-Valdes, María C. Ávila-Arcos.

Abstract

Background

Upon European colonization of Mexico successive outbreaks of infectious diseases occurred, which are estimated to have killed millions of Indigenous individuals. One of the deadliest outbreaks was *"Cocoliztli"*, which was responsible for 15 million deaths shortly after colonization (1545-1550 CE) in New Spain (present-day Mexico). According to paleogenomic evidence, *Salmonella enterica* Paratyphi C has been suggested to be one of the causative agents of the *"Cocoliztli"* epidemic (Vågene *et al.*, 2018). Additionally, *Salmonella enterica* Paratyphi C has been detected in 6,500 years old skeletal remains from Russia, Turkey and Switzerland (Key et al. 2020) and from 800-year-old remains from Norway (Zhou et al. 2018). Here we aim to gain insights into the geographic distribution and spread of *Salmonella enterica* Paratyphi C in the early Colonial period in Mexico.

Results

To gain insights into causative agents of infectious diseases during the post-contact period, we screened and generated paleogenomic data for seven teeth from Colonial (320-120 BP) individuals from the Temple of the Immaculate Conception 'La Conchita', Mexico City, for which records of having served as a cemetery for *"Cocoliztli"* victims exist. Through the taxonomic profiling of the paleogenomic data, we identified numerous reads assigned to *S.* Paratyphi C, which causes paratyphoid fever in humans. We designed an in-house capture-enrichment strategy of *S. Paratyphi* C DNA, which enabled the reconstruction of a 10-fold coverage genome. We determined its phylogenetic relationship with other *S.* Paratyphi C genomes from southern Mexico and Europe.

Conclusions

We reconstructed *a S.* Paratyphi C genome (~10X) from human remains from Mexico City. Through phylogenetic analyses we determined a close relationship with two strains recovered from south Mexico linked to the *"Cocoliztli"* epidemic outbreak during the Colonial period. In-depth

gene presence/absence alyses are needed to understand its virulence mechanisms and to gain evolutionary insights of *S.* Paratyphi C in Mexico.

6.1. Introduction

Salmonella enterica serovar Paratyphi C is a human-restricted pathogen that causes enteric fever, which is associated with a broad spectrum of clinical symptoms, from asymptomatic to severe, such as fever, malaise, headache, and can cause perforation and profuse hemorrhage of ileal ulcers (Baker *et al.*, 2020).

During the pre-antibiotic era, enteric fever caused a mortality rate between 10 and 30% (Saha *et al.*, 2020). The availability of first-line antimicrobials (chloramphenicol, ampicillin, and trimethoprim-sulfamethoxazole) over the past nearly 70 years has reduced the overall mortality rate to less than 1% (Gibani, Britto and Pollard, 2018). In addition, the vaccine against *S. Typhi* (an injectable Vi-capsular polysaccharide) is also effective for Paratyphi C infection, since it carries the genes for the Vi capsular polysaccharide. Hence people vaccinated for Typhi will also have protection against Paratyphi C to some degree (Roggelin *et al.*, 2015). Therefore, the presence and impact of Paratyphi C on ancient populations could have been greater than previously expected (Nair *et al.*, 2020).

The Spanish conquest of Mexico in 1519, was a complex process of change, which not only impacted religion and political organization, but also the lifestyle and health status of the Indegenous societies (Hassig, 2006). During 1519, the population size was approximately 25.2 million people in central Mexico (which extends from the desert north of Mexico City to the Isthmus of Tehuantepec)(Cook and Borah, 1971). At the end of the sixteenth century, the population had dropped by 95 percent (Acuña-Soto Rodolfo and David W. Stahle, 2002). Part of this dramatic depopulation was caused by warfare, but mostly by the infectious diseases that came with the arrival of Europeans and Africans.

The geographical conditions throughout Mesoamerica, combined with the high density of the Indigenous population, permitted rapid and effective transmission of infectious diseases (Holloway, 2011). According to the historical evidence, the first epidemic in New Spain, caused by smallpox, occurred between 1520 and 1521, in which 5-8 million of individuals died (Acuña-Soto Rodolfo and David W. Stahle, 2002). This was followed by mumps in 1531. In addition, during 1545-1548, an epidemic called *“Cocoliztli”* swept through Mexico City. *“Cocoliztli”* was

reported as a deadly disease of unknown cause, which was characterized by an intense bleeding and high fever; it was estimated that between 5 to 15 million individuals died because of the disease (Guevara Flores, 2017). Additionally, other diseases swept through Mexico, including measles and influenza also killing in the order of millions (Cook and Lovell, 2000).

The identification of aDNA of *Salmonella enterica subsp. enterica* serovar Paratyphi C genome from brurials linked to the *"Cocoliztli"* epidemic outbreak during 1545–1550 CE, located in Teposcolula Yucundaa, Oaxaca, Mexico, suggested this pathogen could be one of the possible causes of *"Cocoliztli"* at Teposcolula Yucundaa (Vågene *et al.*, 2018).

On the other hand, further aDNA studies have identified *Salmonella enterica* Paratyphi C in 6,500-year old skeletal remains from Russia, Turkey and Switzerland (Key *et al.*, 2020) and from 800-year old remains from Norway (Zhou *et al.*, 2018). These findings reflect the long-standing relationship of this pathogen with ancient Eurasian populations, and show that it was circulating in Europe 300 years before it was first known to appear in Mexico.

In order to gain a broader understanding of *S.* Paratyphi C's evolution and its impact on ancient populations in Mexico, we screened the metagenomic composition of seven individuals excavated at the Temple of the Immaculate Conception 'La Conchita', Coyoacan, located in the southwest of Mexico City, dated to (320-120 BP) to screen for the presence of ancient pathogen DNA (Supplementary table 1). We extracted DNA, built sequencing libraries, and generated low-depth shotgun sequencing data from seven teeth samples. One metagenome (from individual CO-05) yielded approximately 42,000 reads mapping to *S.enterica* Paratyphi C. Based on this finding, we applied an in-house targeted capture-enrichment protocol to recover the whole *S.* Paratyphi C genome. This allowed us to assess its phylogenetic relationship with available genomic sequences of ancient and modern strains of *S.* Paratyphi C. The presence of *S.* Paratyphi C in Mexico City extends the known geographic distribution of this pathogen, allowing us to gain insight into its spread and distribution in the early Colonial period in Mexico.

Table 1. Overview of *S. enterica* positive samples.

Sample	Date (yr BP)	Archeological site	Uniquely mapped reads	Coverage	Study
CO-05	320-120	Temple of Immaculate Conception, Mexico	765,737	10.85	This study
Tepos 14	504-317	Teposcolula-Yocunda, Oaxaca, Mexico	4,104,504	14.41	Vagene *et al.,* 2018
Tepos 35	504-317	Teposcolula-Yocunda, Oaxaca, Mexico	5,888,701	25.5	Vagene *et al.,* 2018
Ragna	1200±50	Trondheim, Norway	48,481	2.0	Zhou *et al.*, 2018
SUA004	4,300–4,010	Su Asedazzu, Italy	222,506	3.04	Key *et al.,* 2020
ETR001	1,810–1,620	Chiusi, Italy	410,446	4.64	Key *et al.,* 2020
MUR009	6,500–6,350	Murzikhinsky II, Russia	553,044	7.04	Key *et al.,* 2020
MUR019	6,490–6,320	Murzikhinsky II, Russia	107,052	2.19	Key *et al.,* 2020
IV3002	5,580–5,080	Ipatovo 3, Russia	112,992	1.97	Key *et al.,* 2020
MK3001	2,990-2,870	Marinskaja 3, Russia	15,446	1.32	Key *et al.,* 2020
IKI003	5,290–5,050	İkiztepe, Turkey	17,695	1.18	Key *et al.,* 2020

Alignment statistics are based on Paratyphi C RKS4594 reference genome (NC_012125.1)

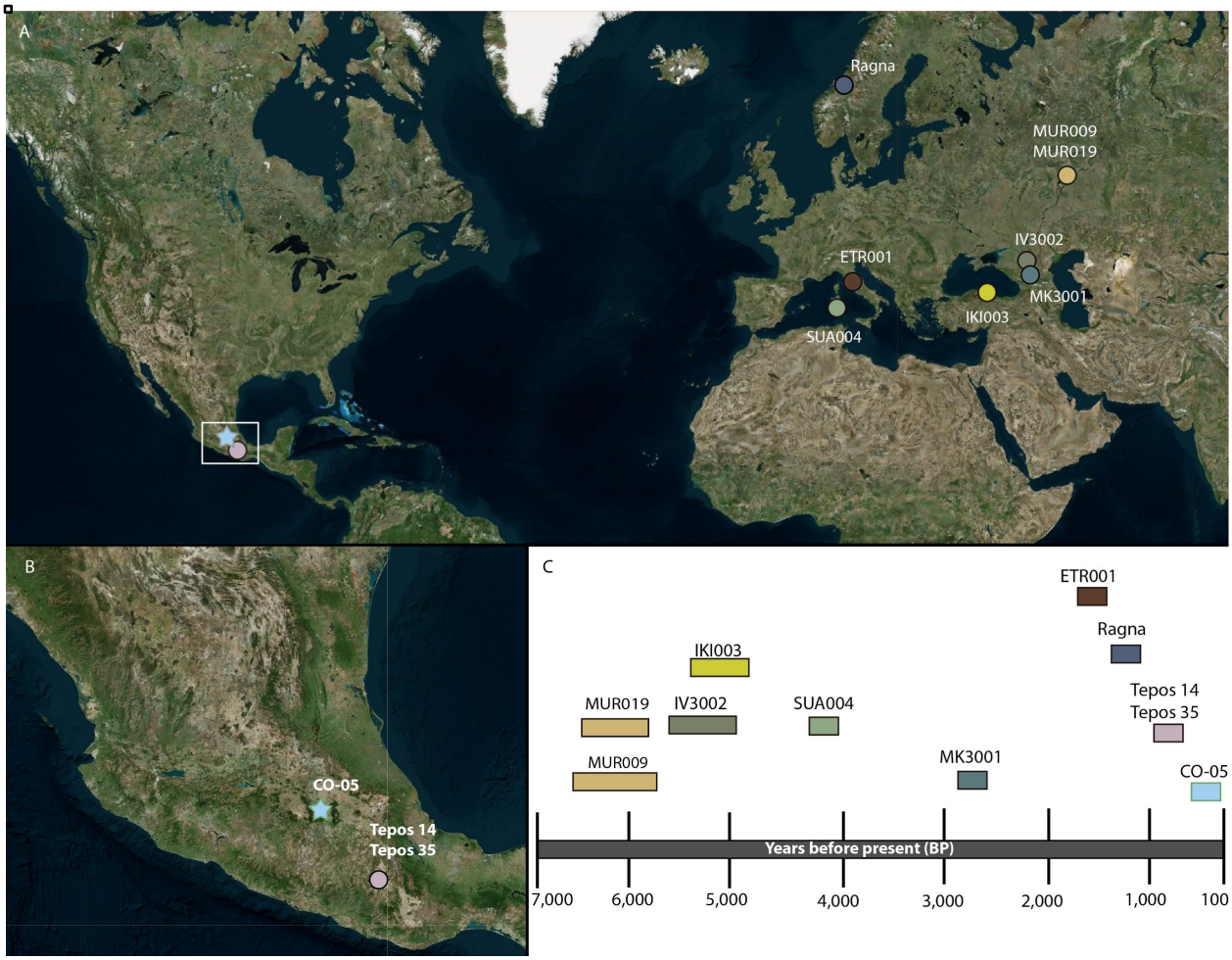


Figure 1. Geographic location and radiocarbon age of ancient human individuals infected with *S. enterica*. Previously published ancient genomes from Vagene *et al.,* 2018 (Tepos 14 and Tepos 35); Zhou, *et al.,* 2018 (Ragna), and Key, *et al.,* 2020 (MUR009, MUR019, IKI003, IV3002, SUA004, MK3001, ETR001).

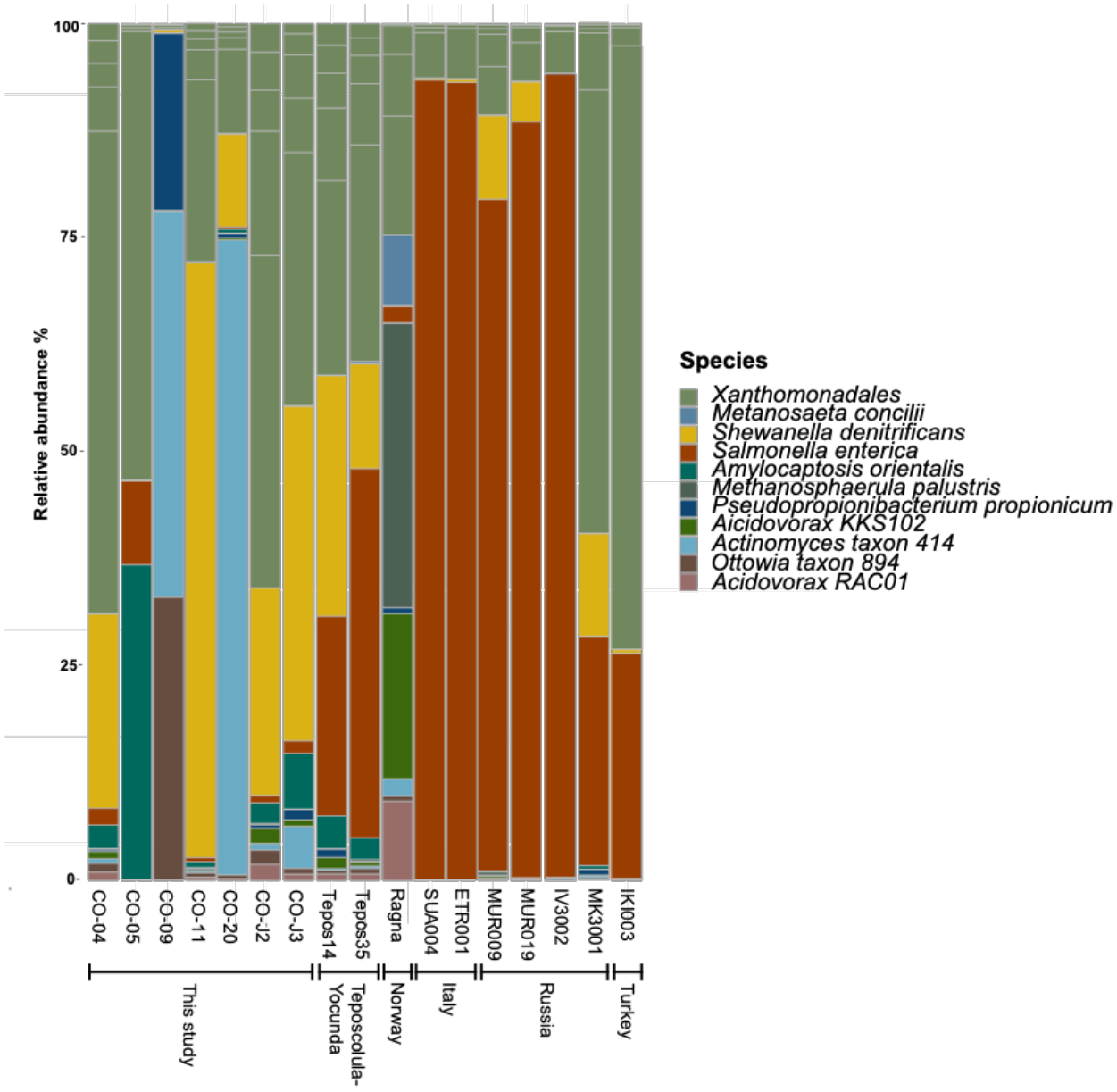


Figure 2. *S. enterica* DNA in Colonial individuals. Relative abundance as determined by Kraken2 of the main taxa observed in each ancient individual. *S. enterica* is shown in red color. The CO-05 individual had the highest number of assigned reads to *S. enterica* compared to all individuals analyzed in this study (Supplementary table 2). Previously published data were included; Teposcolula-Yocundaa (Vågene *et al.*, 2018); Norway (Zhou *et al.*, 2018); Italy, Russia and Turkey (Key *et al.*, 2020).

6.2. Results

6.2.1. Identification of *Salmonella enterica* Paratyphi C in Colonial individuals from Mexico City

We analyzed teeth samples from seven individuals from the Temple of the Immaculate Conception 'La Conchita', Mexico City, dated to 320-120 BP (Figure 1). DNA extracts were generated from teeth samples, converted into Illumina double-stranded libraries and shotgun sequenced to a minimum depth of two million paired-end reads per library on a NextSeq550 instrument. We determined the taxonomic profile of these individuals with program Kraken2 using a database composed of whole bacterial, archaeal and viral genomes available in NCBI's Reference Sequence Database (2017). The majority of reads belonged to bacterial taxa from soil and human oral microbiota at varying proportions (Figure 2 and Supplementary Table 1). Interestingly, for one individual (CO-05) Kraken2 assigned 67,650 reads to *S. enterica.* One DNA extraction and two library construction blanks did not show any hits to *S. enterica,* suggesting an authentic result (Supplementary table 3). We identified a higher number of reads mapping to the *S*. Paratyphi C RKS4594 genome (NC_012125.1) (42,506 reads), compared to *S. enterica* Typhi and Typhimurium (Supplementary Table 4). We analyzed the damage pattern characteristic of aDNA, the distribution of the evenness of reads across the genome, and obtained edit distance calculation to authenticate the ancient nature of the S. Paratyphi C DNA in CO-05 (Supplementary Figure 1).

6.2.2. Reconstruction of the *S. enterica* genome through hybridization capture.

To increase the DNA fraction belonging to *S.* Paratyphi C we performed an in-solution hybridization capture using in-house biotinylated RNA baits transcribed from *Salmonella enterica* serovar *Typhimurium strain SO2*, isolated from a Mexican patient (Silva *et al.*, 2016). The genetic similarity (homologous genes having over 97% DNA sequence identity) among Salmonella serovars and sharp pathogenic differences (self-limited local infections vs potentially fatal systemic infections) suggest a very close relation between Salmonella serovars. The *S. enterica* enriched library was sequenced to 2 million reads on the Illumina NextSeq550. The captured libraries show a 5.3-fold enrichment compared to the pre-capture libraries (Supplementary Table 5 and 6). However, we acknowledge that using S. enterica *Typhimurium* baits, could hinder the capture of regions specific to *S.* Paratyphi C, therefore, we sequenced at higher depth both pre-

capture and capture approximately to 245 million reads. This strategy allowed the reconstruction of a 10-fold Paratyphi C genome (Table 1).

6.2.3. Phylogenetic analysis of *S. enterica* genomes.

For the phylogenetic analysis, we built a multi-sequence alignment of 18 *S. enterica* modern genomes (Supplementary Table 7) and the six ancient *S. enterica* genomes (Table 1). We selected sequences with depths above 3X, which excluded IKI003, IV3002, MUR019 and Ragna, to ensure reliable SNP calling. The phylogenetic analysis was performed with a set of 70,620 SNPs that could be called from at least 90% of the genomes.

All *S. enterica* genomes from ancient individuals are phylogenetically close to the previously designated “Para C lineage” (Alikhan *et al.*, 2018; Key *et al.*, 2020). As is shown in the phylogenetic tree, this lineage includes Paratyphi C and a serovar invasive for pigs (Choleraesuis).

The ancient genome recovered in this study (CO-05) together with the previously reported ancient *S.* Paratyphi genomes from south Mexico (Tepos 14 and Tepos 35), clustered with modern Paratyphi C. As for the ancient *S. enterica* genomes from western Eurasia, the ETR001 genome, dated to 1,700 BP from Italy, which its human subsistence practice is associated with agro-pastoralist, defines a novel sub-branch of Choleraesuis. The early Bronze Age genome SUA004 (Italy; 4,200 BP), also from a agro-pastoralist subsistence activity, defines a new branch that is basal to the entire Para C Lineage, while MUR009 (Russia; 6500 BP) *S. enterica genome* is basal to Para C Lineage and Choleraesuis (human- and pig-adapted serovars).

6.3. Discussion

Here we confidently report a *S. Paratyphi* C genome (~10X) recovered from a tooth archeologically dated to 320-120 BP from Mexico City, combining an in-house bait capture-enrichment protocol using *S. typhimurium* modern DNA and deep shotgun sequencing.

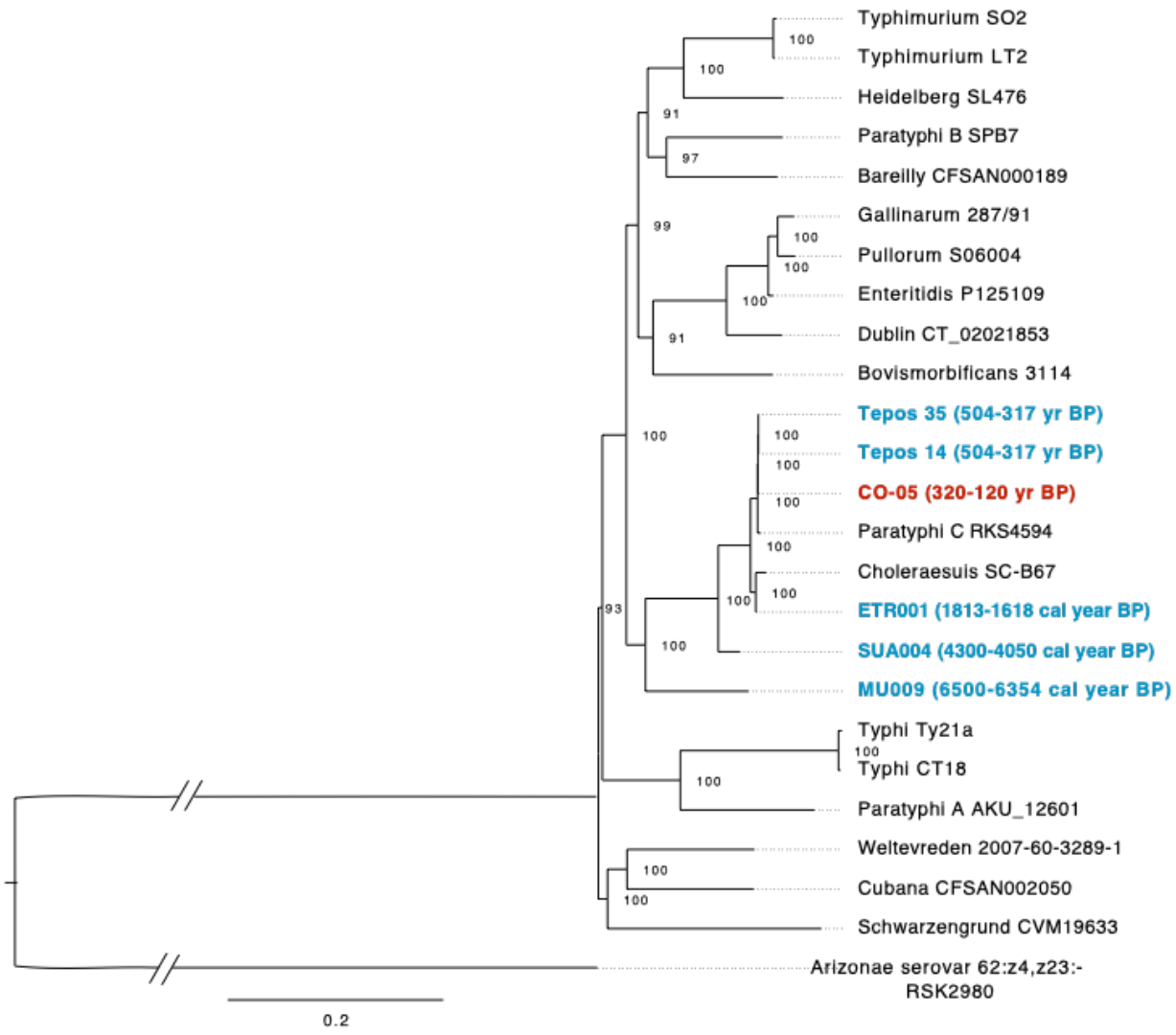


Figure 3. Phylogenetic relationships of reconstructed ancient and modern *S. enterica* genomes. Maximum likelihood tree, including the ancient genomes (>3X coverage), considering 70,620 SNP positions. The sample analyzed in this study is in red color, and the ancient samples previously reported are represented in blue colors. Bootstrap values are shown in black at each node (1,000 bootstraps). Arizonae serovar was used as the outgroup.

The individuals analyzed in this study were excavated from Coyoacan, located in the southwest of Mexico City, where the first New Spain government was established. Between 1550 and 650 CE, the characteristic landscape condition of Coyoacan—fertile soil and abundant water—allowed a large-scale agricultural production and livestock raising (Bautista Martínez, 1986; Horn, 1997). This could have promoted an increased occurrence, persistence and spread of Salmonella among the population (Hasan *et al.*, 2019)

Additionally, some of the burials excavated at the Temple of the Immaculate Conception 'La Conchita', were covered in a thin layer of lime (Moreno et.al., In Preparation), which has been reported to be used to accelerate the rate of body decomposition, thus preventing the spread of infectious diseases during the Colonial period (Salas Contreras, 2007). Nonetheless, this site has no archeological or historical evidence linking it to a specific epidemic outbreak.

The identification of *S. enterica* in ancient transitional foragers and agro-pastoralist individuals from western Eurasia suggest that the Neolithization process influenced the emergence of human-specific Paratyphi C (Key *et al.*, 2020).

Our identification of *S. Paratyphi* C in Mexico City extends the known geographic and temporal distribution of this pathogen in Mexico in the past. Its presence in the Temple of the Immaculate Conception 'La Conchita' during the 320-120 BP period provides direct evidence that *S.* Paratyphi C affected people elsewhere besides only the Teposcolula-Yucundaa site in southern Mexico during the 1545-1450 *"Cocoliztli"* epidemic.

6.4. Conclusion

The in-house bait protocol followed in this study and together with deeper sequencing, proved to be a successful strategy in capturing a *S.* Paratyphi C genome (~10X) from an individual (320-120 BP) from Mexico City. Through phylogenetic analyses we determined a close relationship with two strains recovered from south Mexico linked to the *"Cocoliztli"* epidemic outbreak during the Colonial period.

Further analyses, like the presence and absence of S. Paratyphi C genes when compared to other ancient and modern strains, could provide clues about the virulence mechanisms and to gain evolutionary insights of *S.* Paratyphi C in Colonial Mexico.

6.5. Materials and Methods

6.5.1. Samples and DNA extraction.

Tooth samples were excavated from the Temple of the Immaculate Conception 'La Conchita', Mexico City, Mexico (n = 7 teeth, ca. 320-120 BP) and were obtained from the Laboratory of Osteology, Post Graduate Studies Division, National School of Anthropology and History (ENAH), Mexico (Supplementary Table S1). Permits to carry out DNA analyses with these samples were

granted by the Archaeology Council of the Instituto de Antropología e Historia (INAH) of Mexico (401.1S.3-2020/1310). The samples were processed in the Human Paleogenomics Laboratory at the International Laboratory for Human Genome Research, UNAM. Teeth were cut at the enamel-dentin junction, and ~200 mg of tooth root were drilled. DNA was extracted using a standard protocol specifically for aDNA (Dabney *et al.*, 2013).

6.5.2. Sequencing library preparation and sequence data quality control.

The aDNA extract (20 ul) was transformed into double-stranded Illumina libraries, and shotgun sequenced on an Illumina NextSeq550 (2 × 75 cycles) at the LANGEBIO's genomics core facility (National Laboratory of Genomics for Biodiversity, Irapuato, Guanajuato). Additionally, three negative controls, carried along with DNA extraction, library preparation and PCR indexing, respectively, were sequenced. Processing of fastq files (clipping adapter and merging sequence pairs) was done using AdapterRemoval v. 2 (Schubert, Lindgreen and Orlando, 2016) to merge pairs (11 nucleotides minimum overlap), remove adapter sequences and low-quality reads. Only reads 30 bp and longer and with an average quality above 33 were retained for downstream analyses. We mapped to the human reference genome (version GRCh37, hg19), using the 'bwa aln' option with (flags −l500 −t 16 −q 25). We selected the unmapped reads for taxonomic classification with Kraken2 (Wood, Lu and Langmead, 2019) and we used the NCBI RefSeq database including bacterial, archaea and viral genomes. Metagenomic profiles were analyzed with Pavian (Breitwieser and Salzberg, 2020), focusing on pathogenic taxa. The visualization of relative abundance of taxonomic groups was performed with the R's phyloseq library (McMurdie and Holmes, 2013).

6.5.3. Identification and reconstruction of *S. enterica* genome.

S. enterica positive samples were mapped to the *S. Paratyphi C* reference (NC_012125.1) using bwa aln option (with the flags −l 32 −n 0.04, −q 37). Clonal duplicates were removed using samtools rmdup (Li and Durbin, 2009). Damage and fragmentation patterns were assessed using mapDamage2 (Ginolhac *et al.*, 2011).

We enriched the one candidate library from CO-05 individual, for *S. enterica* DNA using an in-house bait protocol based on (Carpenter *et al.*, 2013) protocol. We generated biotinylated RNAs probes transcribed from *Salmonella enterica* serovar *Typhimurium strain SO2*, isolated from a

Mexican patient (Silva *et al.*, 2016), and sequenced at higher depth both pre-capture and capture libraries on the Illumina NextSeq550 (2 × 75 cycles).

6.5.4. *S. enterica* phylogenetic analysis

For the phylogenetic placement of the modern genomes, we download the sequencing data of previous studies (Vågene *et al.*, 2018; Zhou *et al.*, 2018; Key *et al.*, 2020) (Supplementary Table 3). We generated artificial reads for the modern genomes, splitting them into 100 bp reads with 1bp tiling density, using an in-house script and mapped against the Paratyphi C RKS4594 reference following the same criteria used for the ancient genomes (Vågene *et al.*, 2018). In addition, we include two previously published 16th century genomes from South Mexico (Tepos 14 and Tepos 35), seven individuals from Eurasia and one medieval Norwegian (Ragna) genome using the publicly available raw reads (Vågene *et al.*, 2018; Zhou *et al.*, 2018; Key *et al.*, 2020). We used GATK UnifiedGenotyper v3.5.0 (McKenna *et al.*, 2010) to perform SNP discovery for all dataset genomes using the 'EMIT_ALL_SITES' output mode, in order to obtain allele calls at both variant and invariant sites in the vcf file output.

A multi-sequence alignment of all variable sites was built for all modern and ancient genomes using the tool multivcfanalyzer (v. 0.87-alpha)(Bos *et al.*, 2014). The repetitive and highly conserved regions of the Paratyphi C RKS4594 reference genome were excluded for SNP calling to avoid spurious read mapping as in Vågene *et al.*, (2018). In order to avoid spurious SNP calls, we removed ancient genomes with a genome-wide coverage below 3-fold (MK3001, IKI003, IV3002, MUR019 and Ragna) and a genotype support of at least 90%, leading to an alignment with overall 70,620 SNPs. A maximum-likelihood tree was built with RAxML using the GTRCAT model and 1000 bootstraps (Figure 3).

6.6. References

Acuña-Soto Rodolfo, & David W. Stahle. (2002). Megadrought and Megadeath in 16th Century Mexico. *Emerging infectious diseases*, *8*(4).

Alikhan, N.-F., Zhou, Z., Sergeant, M. J., & Achtman, M. (2018). A genomic overview of the population structure of Salmonella. *PLoS Genetics*, *14*(4), e1007261.

Baker, S., Blohmke, C. J., Maes, M., Johnston, P. I., & Darton, T. C. (2020). The Current Status of Enteric Fever Diagnostics and Implications for Disease Control. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *71*(Supplement_2), S64–S70.

Bautista Martínez, J. (1986). *Los antiguos pobladores de Coyoacan, DF: estudio osteológico y cultural*. Tesis de Licenciatura en Antropología Física. México, DF: Escuela Nacional ….

Bos, K. I., Harkins, K. M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S. A., Bryant, J. M., Harris, S. R., Schuenemann, V. J., Campbell, T. J., Majander, K., Wilbur, A. K., Guichon, R. A., Wolfe Steadman, D. L., Cook, D. C., Niemann, S., Behr, M. A., Zumarraga, M., … Krause, J. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature*, *514*(7523), 494–497.

Breitwieser, F. P., & Salzberg, S. L. (2020). Pavian: Interactive analysis of metagenomics data for microbiomics and pathogen identification. *Bioinformatics*, 4.

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., Leshtakov, K., Dimitrova, D., Theodossiev, N., Pettener, D., Luiselli, D., Sandoval, K., Moreno-Estrada, A., Li, Y., Wang, J., … Bustamante, C. D. (2013). Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *American journal of human genetics*, *93*(5), 852–864.

Cook, N. D., & Lovell, W. G. (2000). *Juicios secretos de Dios: epidemias y despoblación indígena en Hispanoamérica colonial*. Editorial Abya Yala.

Cook, S. F., & Borah, W. W. (1971). *Essays in Population History: Mexico and the Caribbean*. University of California Press.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.-L., & Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, *110*(39), 15758–15763.

Gibani, M. M., Britto, C., & Pollard, A. J. (2018). Typhoid and paratyphoid fever: a call to action. *Current opinion in infectious diseases*, *31*(5), 440.

Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E., & Orlando, L. (2011). mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* , *27*(15), 2153–2155.

Guevara Flores, S. E. (2017). *La construcción sociocultural del cocoliztli en la epidemia de 1545 a 1548 en la Nueva España*.

Hasan, R., The Aga Khan University, Nordin, A., Shakoor, S., Keenum, I., Vinneras, B., London School of Hygiene and Tropical Medicine, Swedish University of Agricultural Sciences, The Aga Khan University, Tech, V., & Swedish University of Agricultural Sciences. (2019). Salmonella, Enteric Fevers, and Salmonellosis. En *Global Water Pathogen Project*. https://doi.org/10.14321/waterpathogens.27

Hassig, R. (2006). *Mexico and the Spanish Conquest*. University of Oklahoma Press.

Holloway, T. H. (2011). *A Companion to Latin American History*. John Wiley & Sons.

Horn, R. (1997). *Postconquest Coyoacan: Nahua-Spanish Relations in Central Mexico, 1519-1650*. Stanford University Press.

Key, F. M., Posth, C., Esquivel-Gomez, L. R., Hübler, R., Spyrou, M. A., Neumann, G. U., Furtwängler, A., Sabin, S., Burri, M., Wissgott, A., Lankapalli, A. K., Vågene, Å. J., Meyer, M., Nagel, S., Tukhbatova, R., Khokhlov, A., Chizhevsky, A., Hansen, S., Belinsky, A. B., … Krause, J. (2020). Emergence of human-adapted Salmonella enterica is linked to the Neolithization process. *Nature Ecology & Evolution*, *4*(3), 324–333.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PloS one*, *8*(4), e61217.

Nair, S., Fookes, M., Corton, C., Thomson, N. R., Wain, J., & Langridge, G. C. (2020). Genetic Markers in S. Paratyphi C Reveal Primary Adaptation to Pigs. *Microorganisms*, *8*(5). https://doi.org/10.3390/microorganisms8050657

Roggelin, L., Vinnemeier, C. D., Fischer-Herr, J., Johnson-Weaver, B. T., Rolling, T., Burchard, G. D., Staats, H. F., & Cramer, J. P. (2015). Serological response following re-vaccination with Salmonella typhi Vi-capsular polysaccharide vaccines in healthy adult travellers. *Vaccine*, *33*(33), 4141–4145.

Saha, S., Sajib, M. S. I., Garrett, D., & Qamar, F. N. (2020). Antimicrobial Resistance in Typhoidal Salmonella: Around the World in 3 Days. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *71*(Supplement_2), S91–S95.

Salas Contreras, C. (2007). Prácticas funerarias en la ex-iglesia de la Encarnación, "Antigua Biblioteca Iberoamericana". *Arqueología mexicana*, *36*, 116–134.

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC research notes*, *9*(1), 88.

Silva, C., Calva, E., Puente, J. L., Zaidi, M. B., & Vinuesa, P. (2016). Complete Genome Sequence of Salmonella enterica Serovar Typhimurium Strain SO2 (Sequence Type 302) Isolated from an Asymptomatic Child in Mexico. *Genome Announcements*, *4*(2). https://doi.org/10.1128/genomeA.00253-16

Vågene, Å. J., Herbig, A., Campana, M. G., Robles García, N. M., Warinner, C., Sabin, S., Spyrou, M. A., Andrades Valtueña, A., Huson, D., Tuross, N., Bos, K. I., & Krause, J. (2018). Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology & Evolution*, *2*(3), 520–528.

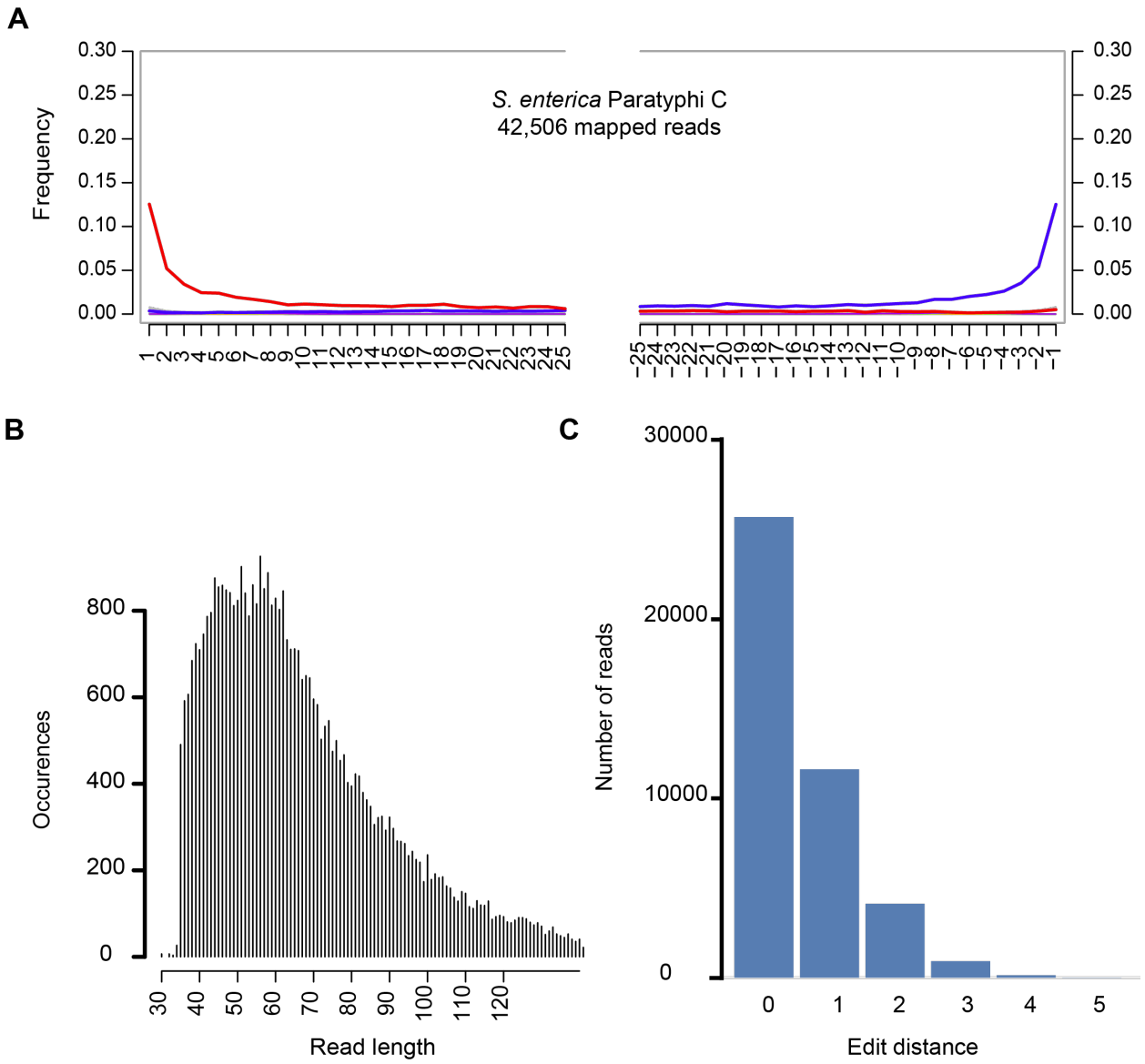
Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. En *bioRxiv* (p. 762302). https://doi.org/10.1101/762302

Zhou, Z., Lundstrøm, I., Tran-Dien, A., Duchêne, S., Alikhan, N.-F., Sergeant, M. J., Langridge, G., Fotakis, A. K., Nair, S., Stenøien, H. K., Hamre, S. S., Casjens, S., Christophersen, A., Quince, C., Thomson, N. R., Weill, F.-X., Ho, S. Y. W., Gilbert, M. T. P., & Achtman, M. (2018). Pan-genome Analysis of Ancient and Modern Salmonella enterica Demonstrates Genomic Stability of the Invasive Para C Lineage for Millennia. *Current biology: CB*, *28*(15), 2420–2428.e10.

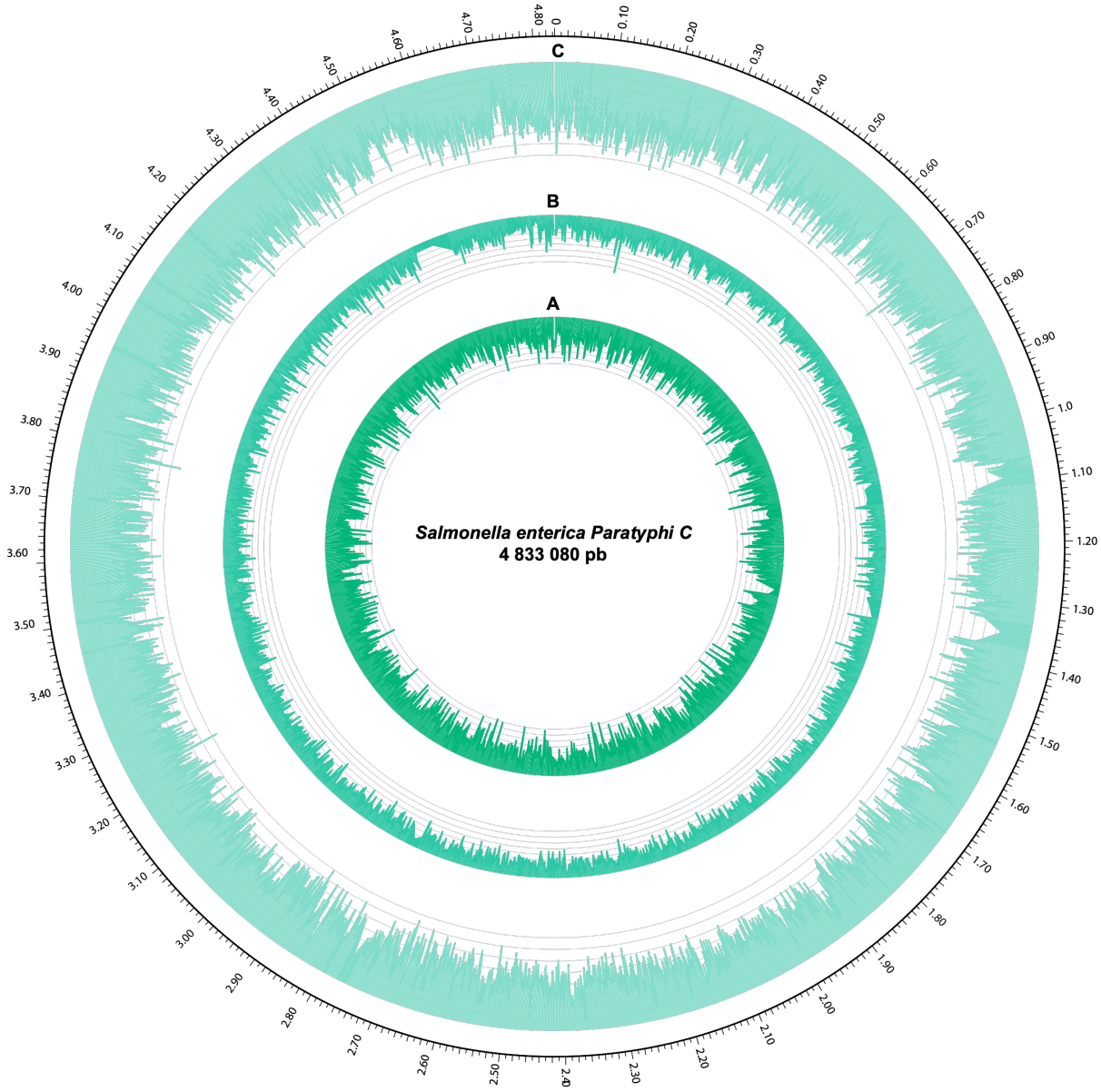
Supplementary Information

Reporte del proyecto. Recovery of an ancient *Salmonella enterica* genome from a Colonial individual from Mexico City

Supplementary Figures



Supplementary Figure 1. Authentication of ancient *S. enterica* Paratyphi C DNA of CO-05. **A**. Deamination profile obtained from mapping *S. enterica* Paratyphi C reference genome (NC_010067.1) against CO-05 shotgun sequencing data. The X axis shows the position (nt) on the 5' and 3' end of the read, and the Y axis the frequency of C > T changes at 5' end of fragments (red line), whereas G > A substitutions at 3' end of fragments (blue line). **B.** Deamination profiles were produced using mapDamage 2.0. B. DNA fragment length distribution of the mapped reads. **C.** Distribution of edit distance of CO-05 reads mapped to *S. enterica* Paratyphi C reference genome (NC_010067.1).



Supplementary Figure 2. Coverage plot S. enterica Paratyphi C genome in CO-05. The depth is shown in 100 bp windows. Each grey line represents approx. 2x. **A.** Shotgun sequencing data; **B**. First round of capture-enrichment; **C.** Merge data of second round of capture-enrichment and deep sequencing.

Supplementary Tables

Supplementary Table 1. Archeological description of individuals from the Temple of Immaculate Conception "La Conchita" analyzed in this study.

Sample ID	Substrate	Teeth used for DNA extraction	Date	Age at death	Sex estimate skeleton	Sex estimate DNA	Host mtDNA haplogroup
CO-04	Tooth	1st premolar mandibular Right	320-120 BP	NA	NA	NA	B2
CO-05	Tooth	3rd Molar Maxillary Right	320-120 BP	NA	NA	NA	NA
CO-09	Tooth	3rd Molar Maxillary Left	320-120 BP	NA	NA	NA	L3b1a11
CO-11	Tooth	1st premolar mandibular Right	320-120 BP	NA	NA	NA	NA
CO-20	Tooth	2nd Premolar Mandibular Right	320-120 BP	30-35 years	NA	NA	NA
CO-J2	Tooth	2nd Premolar Maxillary Left	320-120 BP	NA	NA	NA	NA
CO-J3	Tooth	1st Premolar Maxillary Left	320-120 BP	NA	NA	NA	NA

Supplementary Table 2. Taxonomic classification based on Kraken2 of shotgun screening data

#OTU ID	CO-11	CO-20	C0-04	CO-05	CO-09	CO-J2	CO-J3	Kingdom	Phylum	Class	Order	Family	Genre	Specie
1813	184	96	481	362871	173	152	295	k__Bacteria	p__Actinobacteria	c__Actinobacteria	o__Pseudonocardiales	f__Pseudonocardiaceae	g__Amycolatopsis	s__
31958	143	82	479	254458	158	112	306	k__Bacteria	p__Actinobacteria	c__Actinobacteria	o__Pseudonocardiales	f__Pseudonocardiaceae	g__Amycolatopsis	s__orientalis
208439	67	29	156	127335	64	63	103	k__Bacteria	p__Actinobacteria	c__Actinobacteria	o__Pseudonocardiales	f__Pseudonocardiaceae	g__Amycolatopsis	s__japonica
129921	56	31	158	71397	37	62	89	k__Bacteria	p__Actinobacteria	c__Actinobacteria	o__Pseudonocardiales	f__Pseudonocardiaceae	g__Amycolatopsis	s__keratiniphila
28901	97	40	335	67650	96	38	67	k__Bacteria	p__Proteobacteria	c__Gammaproteobacteria	o__Enterobacterales	f__Enterobacteriaceae	g__Salmonella	s__enterica
1485	1484	54	28	22175	501	77	60	k__Bacteria	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	s__
1491	1570	75	35	19277	210	102	97	k__Bacteria	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	s__botulinum
1513	2239	72	38	16909	64	103	30	k__Bacteria	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	s__tetani
1542	104	31	15	15049	72	21	32	k__Bacteria	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	s__novyi
1492	71	5	1	13385	22	4	13	k__Bacteria	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	s__butyricum
75697	652	230	587	11637	202	458	118	k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Alcaligenaceae	g__Castellaniella	s__defragrans
632569	172	57	320	10585	159	147	148	k__Bacteria	p__Actinobacteria	c__Actinobacteria	o__Pseudonocardiales	f__Pseudonocardiaceae	g__Saccharomonospora	s__marina
1007105	70	144	171	9682	67	301	28	k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Alcaligenaceae	g__Pusillimonas	s__sp. T7-7
2070	345	138	726	8314	238	202	293	k__Bacteria	p__Actinobacteria	c__Actinobacteria	o__Pseudonocardiales	f__Pseudonocardiaceae	g__	s__
33910	290	138	646	8247	237	226	249	k__Bacteria	p__Actinobacteria	c__Actinobacteria	o__Pseudonocardiales	f__Pseudonocardiaceae	g__Amycolatopsis	s__mediterranei
29442	9384	4970	10203	6910	13827	2004	8107	k__Bacteria	p__Proteobacteria	c__Gammaproteobacteria	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	s__tolaasii
1883	20154	4624	24079	4476	14164	1356	19095	k__Bacteria	p__Actinobacteria	c__Actinobacteria	o__Streptomycetales	f__Streptomycetaceae	g__Streptomyces	s__

46867	83	6	2	4051	70	6	79	k__Bacteria	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	s__chauvoei
1496	537	14	7	3485	1504	67	12	k__Bacteria	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Peptostreptococcaceae	g__Clostridioides	s__difficile
590	0	0	1	3442	2	0	4	k__Bacteria	p__Proteobacteria	c__Gammaproteobacteria	o__Enterobacterales	f__Enterobacteriaceae	g__Salmonella	s__
29341	244	4	6	3401	50	15	10	k__Bacteria	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	s__argentinense
80840	3427	969	4012	3033	2739	2732	769	k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__	g__	s__
2014542	1233	352	741	3017	165	1669	123	k__Bacteria	p__Proteobacteria	c__Gammaproteobacteria	o__Oceanospirillales	f__Alcanivoracaceae	g__Alcanivorax	s__sp. N3-7A
543	167	111	453	2923	220	204	117	k__Bacteria	p__Proteobacteria	c__Gammaproteobacteria	o__Enterobacterales	f__Enterobacteriaceae	g__	s__
1814	348	102	517	2805	196	191	208	k__Bacteria	p__Actinobacteria	c__Actinobacteria	o__Pseudonocardiales	f__Pseudonocardiaceae	g__Amycolatopsis	s__methanolica
666685	1729	618	9048	2304	940	579	661	k__Bacteria	p__Proteobacteria	c__Gammaproteobacteria	o__Xanthomonadales	f__Rhodanobacteraceae	g__Rhodanobacter	s__denitrificans
286	4233	1585	9759	2263	2282	1785	1386	k__Bacteria	p__Proteobacteria	c__Gammaproteobacteria	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	s__
1494	1546	47	6	2187	24	79	19	k__Bacteria	p__Firmicutes	c__Clostridia	o__Clostridiales	f__Clostridiaceae	g__Clostridium	s__cochlearium

Supplementary Table 3. Taxonomic classification by Kraken2 of the DNA extraction and library construction blanks.

Kingdom	Phylum	Class	Order	Family	Genre	Specie	CO-DNA	CO-LIB	CO-iPCR
k__Bacteria	p__Proteobacteria	c__Gammaproteobacteria	o__Enterobacterales	f__Enterobacteriaceae	g__Salmonella	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Gammaproteobacteria	o__Enterobacterales	f__Enterobacteriaceae	g__Salmonella	s__enterica	0	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__	g__	s__	1	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Burkholderiaceae	g__	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Burkholderiaceae	g__Burkholderia	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Burkholderiaceae	g__Paraburkholderia	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Burkholderiaceae	g__Ralstonia	s__	0	0	0

k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Burkholderiaceae	g__Cupriavidus	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Burkholderiaceae	g__Pandoraea	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Comamonadaceae	g__	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Comamonadaceae	g__Acidovorax	s__	1	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Comamonadaceae	g__Variovorax	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Comamonadaceae	g__Delftia	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Comamonadaceae	g__Comamonas	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Oxalobacteraceae	g__Janthinobacterium	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Oxalobacteraceae	g__Herbaspirillum	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Betaproteobacteria	o__Burkholderiales	f__Alcaligenaceae	g__Bordetella	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhizobiales	f__	g__	s__	1	0	0
k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhizobiales	f__Bradyrhizobiaceae	g__	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhizobiales	f__Bradyrhizobiaceae	g__Bradyrhizobium	s__	0	1	0
k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhizobiales	f__Methylobacteriaceae	g__Methylobacterium	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhizobiales	f__Rhizobiaceae	g__Rhizobium	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhizobiales	f__Brucellaceae	g__	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Rhizobiales	f__Brucellaceae	g__Ochrobactrum	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Sphingomonadales	f__	g__	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Sphingomonadales	f__Sphingomonadaceae	g__	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Sphingomonadales	f__Sphingomonadaceae	g__Sphingomonas	s__	0	0	0
k__Bacteria	p__Proteobacteria	c__Alphaproteobacteria	o__Sphingomonadales	f__Sphingomonadaceae	g__Sphingopyxis	s__	0	0	0

CO-DNA, DNA extraction blank; CO-LIB, library construction blank; CO-iPCR, indexing blank.

Supplementary Table 4. Mapping statistics of CO-05 against *Salmonella enterica* serovars human-adapted.

Serovar	Total sequences	Uniquely mapped reads	Average depth genomewide	Positions covered at least once	Mapped reads average length	% damage of first 5-prime base	% damage of second 5-prime base	% damage of first 3-prime base	% damage of second 3-prime base
Paratyphi C RKS4594 (NC_012125.1)	9,772,063	42,506	1.3424	2,087,866	66.2963	0.1254	0.0520	0.1252	0.0541
Typhimurium str. LT2 (NC_003197.2)	9,772,063	35,185	1.3141	1,751,514	65.8125	0.1123	0.0475	0.1102	0.0516
Typhi str. Ty2 (NC_004631.1)	9,772,063	30,966	1.2912	1,564,729	65.6489	0.1123	0.0471	0.1052	0.0527

Supplementary Table 5. Mapping statistics of ancient individuals included in this study against *Salmonella enterica* Paratyphi C.

Sample_ID	Total sequences	Uniquely mapped reads	Average depth genomewide	Mapped reads average length	% damage of first 5-prime base	% damage of second 5-prime base	% damage of first 3-prime base	% damage of second 3-prime base	Reference
CO-05†	9,772,063	42,506	1.3424	66.2963	0.1254	0.0520	0.1252	0.0541	This study
CO-05*	2,409,110	57,516	1.6152	68.5408	0.1279	0.0565	0.1345	0.0573	This study
CO-05**	245,379,619	765,737	10.7213	66.0463	0.1328	0.0590	0.1378	0.0638	This study
Tepos 14	160,571,850	4,104,504	46.3885	54.625	0.2068	0.2083	0.2333	0.1200	Vagene *et al.*, 2018
Tepos 35	197,940,170	5,888,701	86.0062	70.5921	0.2127	0.1084	0.2045	0.1052	Vagene *et al.*, 2018
Ragna pulp	86,873,591	48,481	1.27317	46.0476	0.086	0.0387	0.0713	0.0299	Zhou *et al.*, 2018
MUR019	1,737,507	107,052	2.19237	59.4002	0.0250	0.0092	0.0240	0.0095	Key *et al.*, 2020
MUR009	48,177,904	553,044	7.04167	54.047	0.0244	0.0097	0.0259	0.0097	Key *et al.*, 2020
IV3002	1,798,465	112,992	1.9748	48.6732	0.0274	0.0078	0.0255	0.0066	Key *et al.*, 2020
IKI003	502,118	17,695	1.18187	45.393	0.1590	0.0905	0.0091	0.0058	Key *et al.*, 2020
SUA004	1,802,652	222,506	3.04235	48.2857	0.0152	0.0048	0.0152	0.0048	Key *et al.*, 2020
ETR001	2,096,278	410,446	4.6447	45.3816	0.0207	0.0051	0.0228	0.0051	Key *et al.*, 2020
MK3001	7,195,996	15,446	1.32774	45.326	0.1353	0.0123	0.142	0.0126	Key *et al.*, 2020

†Shotgun sequencing
*First round of hybridization capture
**Deep sequencing and second round of hybridization capture
Individuals in bold text were included in the phylogenetic analysis

Supplementary Table 6. Overview of mapping statistics of down-sampled pre and post-capture libraries.

Library	Total sequences	Before collapse	After collapse	Uniquely mapped reads	Average depth	Positions covered at least once	Mapped reads average length	Endogenous DNA (%)	Fold coverage	Clonality	Reads on target	Fold enrichment
Pre-capture	2,409,110	10,917	10,836	10,836	1.0755	661,758	66.053	0.45%	0.15	0.7%	10,836	5.30786268
Capture	2,409,110	563,173	57,516	57,516	1.61521	2,427,346	68.5408	2.39%	0.81	89.8%	57,516	

Supplementary Table 7. Reference genomes included in the phylogenetic tree construction.

Identifier(s)	Name
NC_010067.1	Salmonella enterica subsp. arizonae serovar 62:z4,z23:- str. RSK2980 chromosome, complete genome
NC_021820.1	Salmonella enterica subsp. enterica serovar Typhimurium str. 08-1736, complete genome
NC_011149.1	Salmonella enterica subsp. enterica serovar Agona str. SL483 chromosome, complete genome
NC_021844.1	Salmonella enterica subsp. enterica serovar Bareilly str. CFSAN000189, complete genome
NC_022241.1	Salmonella enterica subsp. enterica serovar Bovismorbificans str. 3114 complete genome
NC_006905.1	Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67 chromosome, complete genome
NC_021818.1	Salmonella enterica subsp. enterica Serovar Cubana str. CFSAN002050, complete genome
NC_011205.1	Salmonella enterica subsp. enterica serovar Dublin str. CT_02021853 chromosome, complete genome
NC_011294.1	Salmonella enterica subsp. enterica serovar Enteritidis str. P125109 chromosome, complete genome
NC_011274.1	Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91 chromosome, complete genome
NC_011083.1	Salmonella enterica subsp. enterica serovar Heidelberg str. SL476 chromosome, complete genome
NC_011147.1	Salmonella enterica subsp. enterica serovar Paratyphi A str. AKU_12601 chromosome, complete genome
NC_006511.1	Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC 9150 chromosome, complete genome
NC_010102.1	Salmonella enterica subsp. enterica serovar Paratyphi B str. SPB7 chromosome, complete genome
NC_012125.1	Salmonella enterica subsp. enterica serovar Paratyphi C strain RKS4594 chromosome, complete genome
NC_021984.1	Salmonella enterica subsp. enterica serovar Pullorum str. S06004, complete genome
NC_011094.1	Salmonella enterica subsp. enterica serovar Schwarzengrund str. CVM19633 chromosome, complete genome
NC_003198.1	Salmonella enterica subsp. enterica serovar Typhi str. CT18, complete genome

NC_003197.1 Salmonella enterica subsp. enterica serovar Typhimurium str. LT2 chromosome, complete genome

CP014356.1 * Salmonella enterica subsp. enterica serovar Typhimurium strain SO2, complete genome

NC_016832.1 Salmonella enterica subsp. enterica serovar Typhi str. P-stx-12, complete genome

NC_021176.1 Salmonella enterica subsp. enterica serovar Typhi str. Ty21a, complete genome

NC_004631.1 Salmonella enterica subsp. enterica serovar Typhi str. Ty2 chromosome, complete genome

NT_187069.1-
NT_187134.1 Salmonella enterica subsp. enterica serovar Weltevreden str. 2007-60-3289-1

* Used for baits construction

7. Reporte de proyecto. Paleogenomic analysis of pathogenic bacteria in hunter-gatherers from Central Coastal Argentinian Patagonian.

Miriam Bravo-Lopez, Carolina Rocha-Arriaga, Camila Tamburrini, Viridiana Villa-Islas, Rosa Fregel, Silvia Dahinten, Julieta Gómez-Otero, Gabriela Millán, Carlos Bustamente, María C. Ávila-Arcos, María Laura Parolin

Abstract

Background

The Central-Northern Coast of Argentinian Patagonia has been occupied by hunter-gatherers since at least the middle Holocene based on archeological records and direct dating of human bones. The biogeographical and environmental conditions, with a relatively late European contact would have provided abundance and diversification of the subsistence activity, influencing the hunter-gatherer's health status.

Results

We characterized the metagenomes from 19 individuals from the Central-Northern Coast of Patagonia (6000-300-year BP), using complete bacterial, archaeal and viral genomes in the NCBI's RefSeq database. We identified numerous reads (~1,400 to ~65,000) assigned to *Erysipelothrix rhusiopathiae*, which is a zoonotic pathogenic bacterium that causes erythematous skin lesions known as erysipeloid. In addition, we identified periodontopathic bacteria like *Rothia dentocariosa, Eikenella corrodens, Streptococcus sanguinis, Streptococcus gordonii, Capnocytophaga gingivalis, Capnocytophaga ochracea, Filifactor alocis, Fusobacterium nucleatum, Capnocytophaga sputigena, Veillonella parvula, Aggregatibacter aphrophilus, Filifactor alocis,* and *Tannerella forsythia*. These findings, together with the oral paleopathological evidence, indicate the presence of periodontal disease in these individuals. The results obtained from stable isotope (13C and 15N) analyses show a dietary dependence on terrestrial and marine resources, and an increase of carbohydrate consumption after 1000 years BP, when pottery technology was adopted which could have promoted the presence of these oral pathogens. Additionally, we identified in one individual aDNA from *Clostridium perfringens* and *Clostridium chauvoei*, a causative agent of gangrene, food poisoning and non-foodborne diarrhea, and *Clostridium tetani*, which causes tetanus disease.

Conclusions

Overall, our study demonstrates the presence of ancient pathogen DNA associated with zoonotic pathogenic bacteria and oral cavity infectious diseases in hunter-gatherer individuals from the Central-Northern Coast of Patagonia, which is in concordance with their lifestyle (mobility and diet), particularly before European contact. Further genomic analyses will allow the reconstruction of these pathogen genomes to provide additional insights into their evolution in the past ca. 6000 years, as well as their influence on the hunter-gatherer's lifestyle.

7.1. Introduction

The Argentine Patagonia covers the southernmost end of South America and is divided into three main regions: Andean, central Plateau and Atlantic Patagonia. According to archaeological evidence of human activity in the Arroyo Verde site and by direct dating of human bones, hunter-gatherers' groups occupied the Central-Northern Coast of Patagonia since at least the middle Holocene (7,000 to 5,000 yr BP) (Gómez Otero, 2006). Hunter-gatherer communities have been named as "Tehuelches" by ethnologists of the 19th century. However, this term has been questioned by Nacuzzi et al., (2005), and many others authors since it was a term imposed by other native groups and European conquerors.

The Central-Northern Coast of Patagonia, encompasses three types of environments; the coast, the Plateau and the Chubut river lower valley, providing diverse food resources including plant species, marine and terrestrial animals. All together had a great impact on hunter gatherers' subsistence activities, mobility patterns, demography and human-organism co-residence, influencing their health status (Otero and Novellino, 2011; Flensborg, 2015).

In this regard the isotopic analysis together with artifact, and archeofauna evidence found in the Atlantic coast of Central Patagonia support that, from the Middle Holocene until the European contact in16th century, these hunter-gatherers groups had an intensive utilization of the coastal area and variable exploitation of marine foods (Gomez Otero and Moreno, 2015). Meanwhile, after the horse adoption between the 18th and 19th centuries, a less intensive use of the coastal space is registered, increasing intake of terrestrial proteins and carbohydrates from the inland plateau (Gómez Otero, 2006). These new mobility patterns and lifestyle changes were probably related to the horse ecology and requirements (e.g., water), and the reduction of prey acquisition costs (Gomez Otero and Moreno, 2015). Moreover, the late Holocene was characterized by an

economic intensification, reduction of residential mobility, incorporation of new technologies, increase of social complexity and expansion of inter-ethnic contacts inside and outside Patagonia (Gómez Otero, 2006).

The paleopathological and bioarchaeological evidence of hunter-gatherer has provided insights into the presence of infectious diseases, through the description of skeletal indicators as cribra orbitalia, porotic hyperostosis, and lineal hypoplasia of the dental enamel, suggesting a low frequency of infectious diseases among hunter-gatherers (Constantinescu, 1997; Aspillaga *et al.*, 1999; Otero and Novellino, 2011). However, the vast majority of these studies have been analyzed individuals mainly from the Southern Patagonia (Suby, 2020).

Therefore, a refined analysis of the identification of ancient pathogens in hunter-gatherers from the Central-Northern Coast of Patagonia is needed. Recent developments in paleogenomics have allowed the identification, recovery and reconstruction of ancient pathogen genomes from a variety of archeological remains (Spyrou *et al.*, 2019; Arning and Wilson, 2020). However, most of the pathogen genomes reconstructed are from European samples. To our knowledge, this is the first study reporting the characterization of the metagenomes in search of pathogenic taxa in Argentinian Patagonia samples. We extracted and analyzed aDNA from teeth and bones of 19 individuals from the Central coast of Patagonia (6000-300 BP). We used Kraken2 (Wood, Lu and Langmead, 2019) and a reference database composed of bacterial, fungi, viral genomes (Refseq, November 2017).

7.2. Results and discussion

7.2.1. Taxonomic characterization of hunter-gatherer's metagenomes.

In order to identify hits to pathogenic taxa, we characterized the metagenomic profiles of shotgun data obtained from DNA extracts of 19 individuals from the Central-Northern Coast of Patagonia (n=5 bones and n=14 tooth samples) (Figure 1, Table 1) using the software Kraken2. According to their microbial composition, 85% belongs to Bacterial domain, and to a lesser extent Viral (10%) and Archeal (5%) domains (Supplementary Figure 1).

At the phylum level, 13 out of 19 individuals were enriched for *Actinobacteria* (53%), except AEH148, AEH151, AEH154, AEH157, AEH158 and AEH159. The majority of *Actinobacteria* are organisms present in both terrestrial and aquatic environments (Barka *et al.*, 2016). In AEH148,

Proteobacteria were present in high proportion (40%) compared to the other individuals. This phylum includes organisms with a great impact on medical, veterinary, industrial and agricultural areas (Kersters *et al.*, 2006). In addition, AEH151, AEH154, AEH157, AEH158 and AEH159 individuals were enriched for *Firmicutes* (49%) (Supplementary figure 1). Interestingly, *Firmicutes* has been reported as a main component of the intestinal microflora (Vesth *et al.*, 2013). At the genus level, the taxa composition was diverse among the individuals analyzed (Supplementary figure 2). Most individuals showed an enrichment of *Streptomyces*, except AEH144, which was enriched for *Mycobacterium*, represented by 30%. Additionally, AEH146 showed a high abundance of *Brevibacterium*, mostly represented by the *Brevibacterium sandarakinum* (Gram-stain-positive environmental bacteria) (Kämpfer *et al.*, 2010), as for the AEH148 showed a high abundance of *Pseudomonas*, represented by *Pseudomonas stutzeri* (Gram-negative bacteria and is widely distributed in the environment) (Lalucat *et al.*, 2006).

Individual AEH152 showed high abundance of *Jiangell*, represented by *Jiangella alkaliphila*. The AEH153 individual showed a high abundance of *Nocardiopsis*, represented by the species *Nocardiopsis dassonvillei*. AEH157 showed a high abundance of *Planococcus*, represented by the *Planococcus rifietoensis*. Individuals AEH149 and AEH150 showed a varying composition of *Rothia, Ottowia* and *Neisseria*, and *Micromonospora*, *Mycobacterium*, *Jiangella* and *Halomonas.*

The taxonomic profiles of bones (n=5) and teeth (n=14) from the ancient individuals, did not show any clustering based on the taxa found in each sample (Supplementary Figure 3). This is explained by the varying proportion of the taxa present in the samples analyzed.

Table 1. Individuals analyzed from the Central-Northern Coast of Argentinian Patagonia.

Sample ID	Region	Date ^{14}C yr BP	Diet (^{13}C^{15}N)	References
AEH141	Lower Valley of Chubut river	6070±80	mixed*	Gómez Otero 2007
AEH142	Golfo Nuevo coast	2640±50	mixed	Gómez Otero 2007, Millán et al., 2013

AEH143	Golfo Nuevo coast	2410±50	marine	Gómez Otero 2007; Millán et al., 2013
AEH144	Península Valdés	2010±50	marine	Gómez Otero 2007
AEH145	Lower Valley of Chubut river	1990±60	terrestrial	Gómez Otero 2007; Millán et al., 2013
AEH146	Lower Valley of Chubut river	1480±60	terrestrial	Gómez Otero 2007; Millán et al., 2013
AEH147	Lower Valley of Chubut river	1410±70	n.d.	Gómez Otero 2006; Millán et al., 2013
AEH148	Nuevo Gulf coast	1400±60	n.d.	Gómez Otero 2006; Millán et al., 2013
AEH149	Lower Valley of Chubut river	1390±60	terrestrial	Gómez Otero 2006; Millán et al., 2013
AEH150	Lower Valley of Chubut river	1090±70	n.d.	Gómez Otero 2006; Millán et al., 2013
AEH151	Punta León coast	1050±50	marine	Gómez Otero 2007; Millán et al., 2013
AEH152	Península Valdés	880±50	marine	Gómez Otero 2007; Millán et al., 2013
AEH153	Península Valdés	880±50	marine	Gómez Otero 2007; Millán et al., 2013
AEH154	Nuevo Gulf coast	770±50	mixed	Gómez Otero 2007; Millán et al., 2013
AEH155	North San Matías coast	770±70	mixed	Gómez Otero 2007; Millán et al., 2013
AEH156	Nuevo Gulf coast	550±60	n.d.	Gómez Otero 2007; Millán et al., 2013

AEH157	Nuevo Gulf coast	370±50	mixed	Gómez Otero 2007; Millán et al., 2013
AEH158	Gastre, North Central plateaus	350±50	mixed	Gómez Otero 2007
AEH159	Lower Valley of Chubut river	310±70	mixed	Gómez Otero 2007; Millán et al., 2013

N.d. No data
*marine and terrestrial diet

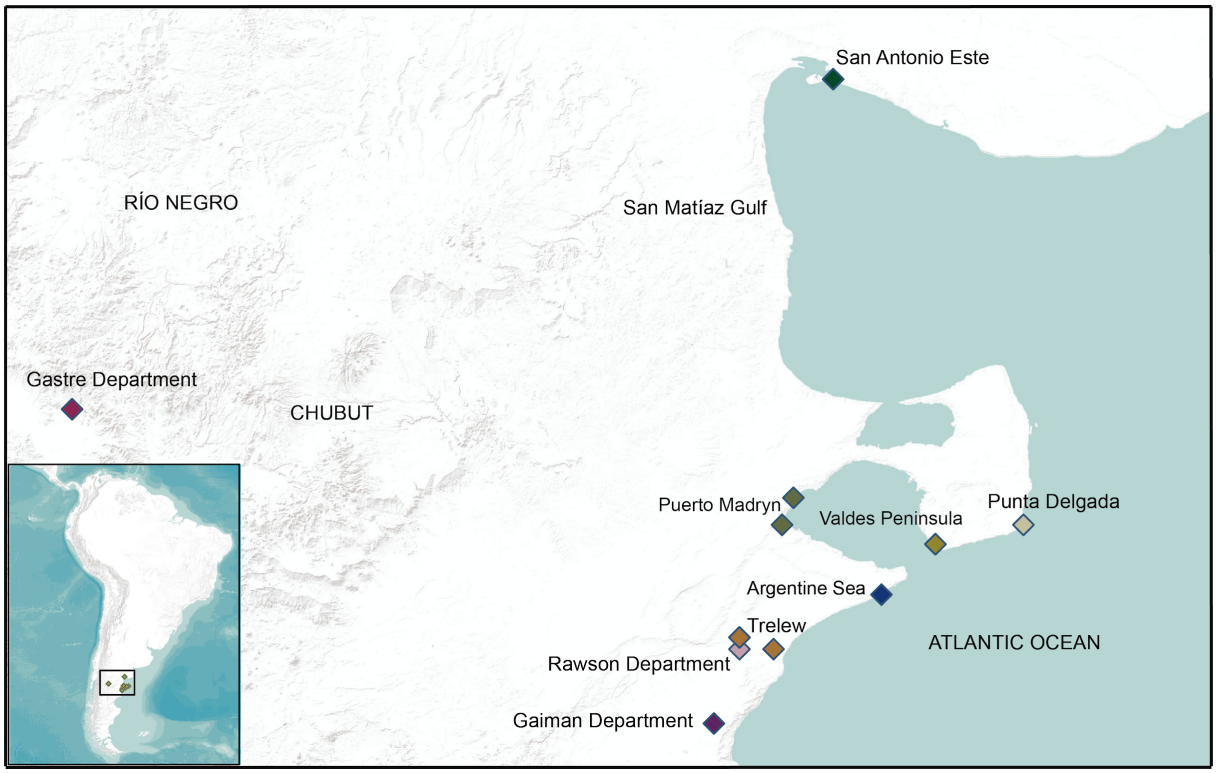


Figure 1. Location of the archeological sites analyzed from the Central-Northern Coast of Patagonia.

7.2.2. Identification of aDNA pathogens associated with systemic infection

The metagenomic composition of the 19 individuals was analyzed in depth in search of pathogenic taxa (Figure 2). For all the putative ancient pathogen DNA identified, we identified the characteristic aDNA damage pattern and a wide distribution across the reference genome (Supplementary figure 4, supplementary table 2).

Of note, we were able to detect aDNA from pathogens associated with periodontal disease, as well as zoonotic pathogens associated with consumption of raw food. These suggest the presence of different infectious diseases circulating in this population. Further studies will allow the reconstruction of these pathogen genomes to contribute to understanding of their evolution as well as their potential influence on the lifestyle hunter-gatherers from Patagonia.

Individuals AEH146 (1480±60 yBP), AEH154 (770±50), and AEH159 (310±70), had hits to bacteria *Erysipelothrix rhusiopathiae*, yielding ~1,400 to ~65,000 mapped reads, showing the characteristic damage pattern of aDNA, namely GC → AT misincorporations (Supplementary figure 3). *E. rhusiopathiae* is a Gram positive opportunistic pathogen in both humans and animals (mainly swine, sheep, cattle, fish, mollusks, and crustaceans), which causes erythematous skin lesions known as erysipeloid, and in some cases the infection becomes generalized, causing septicemia and endocarditis (Forde *et al.*, 2016).

Human erysipeloid has been associated with the handling of raw meat and seafood products, which could have been a common practice among hunter-gatherers from the plateau and eastern Patagonia, since its diet was based on terrestrial species (e.g. land plants, guanaco and choique) and marine species like pinniped, penguins, mollusks and fish (Flensborg, 2015; Forde *et al.*, 2016). Thus, is supported by the C/N isotopic studies on AEH146 individual, who was characterized by a terrestrial diet, and for AEH154, AEH155 and AEH159, by a mixed diet (marine-terrestrial) (Otero, 2007; Otero and Novellino, 2011).

Additionally, according to the paleopathological analysis of the individual AEH159, a periosteal reaction on the left tibia was detected, suggesting the presence of an infectious process during her lifetime.

Clostridium chauvoei was identified in AEH151 (1050±50 yBP) (~ 2,200 mapped reads). Among the Clostridium species, *C. chauvoei* is considered as one of the most pathogenic causing fulminant gas gangrene in humans (Nagano *et al.*, 2008). The infection by *C. chauvoei* is responsible for myonecrosis, and fever that is rapidly followed by lameness and death (Rychener *et al.*, 2017). It is known to mainly infect cattle, but it has been suggested to be a zoonotic pathogen common among modern pastoralists, based on Indigenous knowledge from Fula people in Cameroon and Maasai people in Tanzania (Gaddy, 2020). In addition, we identified numerous reads (~2,500 mapped reads) assigned to *Clostridium butyricum*, bacteria associated with botulism—a neuroparalytic illness caused by botulinum toxin intoxication— in infants (Cassir, Benamar and La Scola, 2016). However, the presence of *C. butyricum* in AEH151 could not necessarily have caused an infectious disease during her lifetime, since it has been identified as part of the commensal bacteria in animals and humans (Cassir, Benamar and La Scola, 2016).

AEH158 (350±50 yBP) had hits to *Clostridium perfringens* (~56,00 reads) and *Clostridium tetani* (~12,000 mapped reads).These pathogenic bacteria produce one or more unique exotoxins that sicken the host and can eventually result in death (Cohen *et al.*, 2017). *C. perfringens* has been isolated from diverse environments such as soils, food, sewage, and from the gastrointestinal tract in both diseased, and non-diseased humans and animals. This bacterium causes gas gangrene, food poisoning and non-foodborne diarrhea, and enterocolitis (Kiu and Hall, 2018). The virulence of this pathogen is associated with the production of >20 identified toxins or enzymes (Revitt-Mills, Rood and Adams, 2015). On the other hand, *Clostridium tetani*'s natural habitat is soil, dust, and intestinal tracts of various animals, and is the causative agent of the tetanus disease (Brüggemann, Bäumer and Fricke, 2003). Interestingly, it has been detected in the gastrointestinal tract from a >5000-year-old neolithic mummified tissue, found in an Alpine glacier in 1991 (Lugli *et al.*, 2017).

Proteus mirabilis was also identified in AEH159 (~20,000 mapped reads). This pathogen causes a variety of human infections, including those of wounds, the eye, the gastrointestinal tract, and the urinary tract. Additionally, according to the paleopathological analysis of the individual AEH159, a periosteal reaction on the left tibia was detected, suggesting the presence of an infectious process during lifetime.

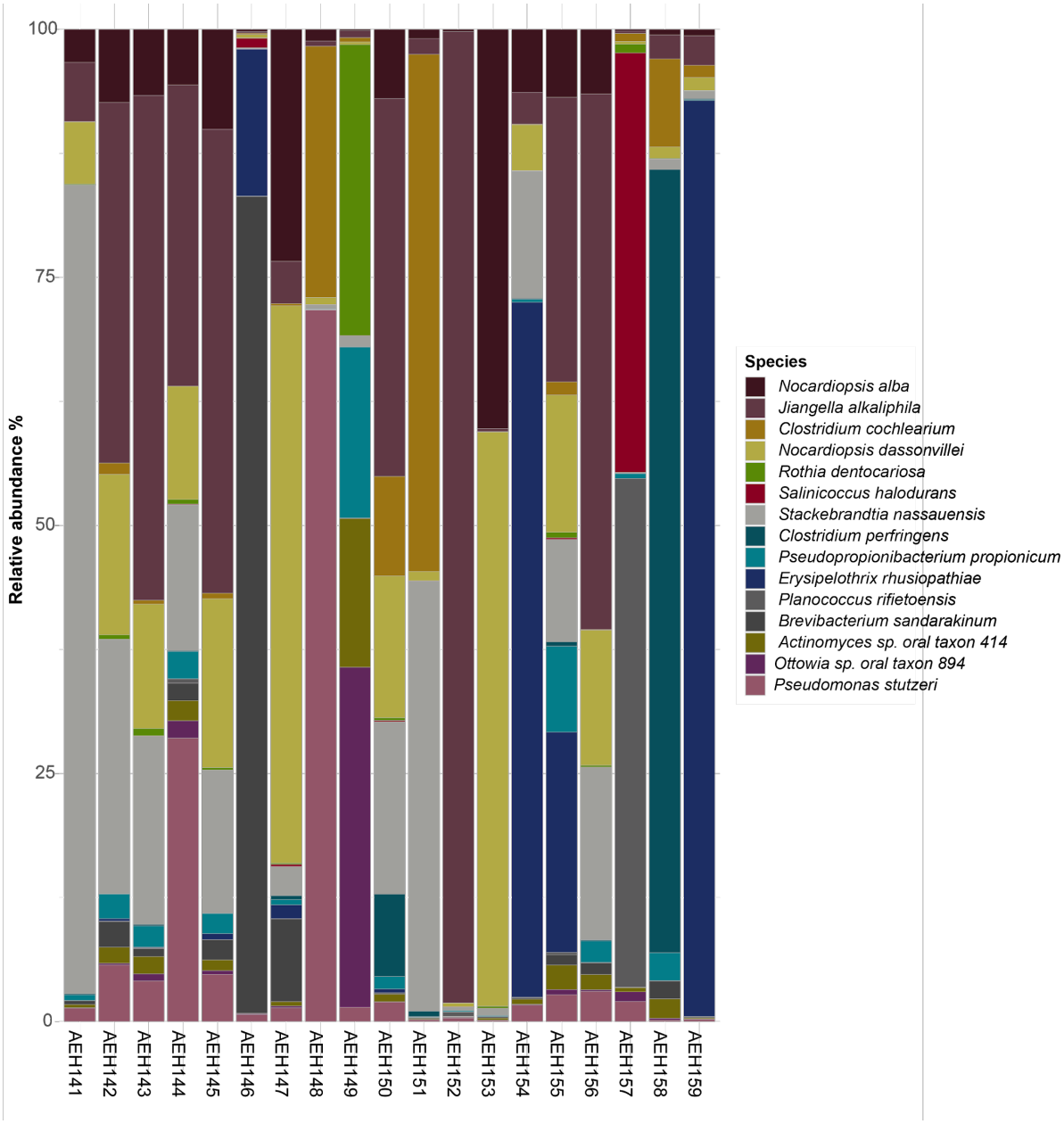


Figure 2. Major species present in ancient individuals from Central Patagonia analyzed in the present study. Taxonomic assignments were obtained using Kraken2.

7.2.3. Identification of aDNA from pathogens associated with oral cavity infections

We identified periodontopathic bacteria in AEH149 (1390±60 yBP), AEH157 (370±50) and AEH158 (310±70). Individual AEH149 had a coinfection of *Rothia dentocariosa* (~72,000 mapped reads), *Eikenella corrodens* (34,000 mapped reads); *Streptococcus sanguinis* (~33,000 mapped reads), *Capnocytophaga gingivalis* (~16,000 mapped reads); *Capnocytophaga ochracea*

(~14,000 reads); *Actinomyces oris* (~14,000 mapped reads); *Fusobacterium nucleatum* (~13,000 mapped reads); *Capnocytophaga sputigena, Veillonella parvula* (~11,000 mapped reads); *Aggregatibacter aphrophilus* (~5,000 mapped reads); and *Filifactor alocis* (~1,000 mapped reads).

The coinfection of periodontopathogenic pathogens causes periodontal disease, which damages the periodontal ligament and alveolar bone (Chen, Feng and Slots, 2020; Curtis, Diaz and Van Dyke, 2020). The co-aggregation of *C. gingivalis* and *C. ochracea* with other oral bacteria, like *S. sanguinis*, is regulated by their adhesins, causing adhesion and colonization on bone (Piau *et al.*, 2013). *Eikenella corrodens* and *Aggregatibacter aphrophilus* share an enhanced capacity to produce endocarditis as well as arthritis, osteomyelitis or periodontitis. Translocation of these *Capnocytophaga spp.* can promote bone infections, persistent inflammation and tissue damage. The paleopathological analyses of AEH149, revealed severe tooth wear, together with a periodontal lesion and an abscess in the right mandibular permanent first molar, which is consistent with the molecular evidence of the oral pathogens detected in this individual. Additionally, a cyst was present in the upper right and lower left teeth.

The etiology of periodontal disease is multifactorial, although a key mechanism for its progressive tissue damage is the production of organic acids from dietary carbohydrates by bacterial fermentation (Dlamini, Morris and Sealy, 2016). According to C/N isotopic composition the AEH149 individual diet was based on terrestrial products, therefore this type of diet could be an important source of carbohydrates. According to Constantinescu et al., (2019), hunter-gatherers from Northern Tierra del Fuego, had a high frequency of temporomandibular joint disorders (TMJ), caused by an elevated mechanical masticatory force characterized to be present in terrestrial dietary patterns (Constantinescu, 1999; Suby, 2020). In modern populations, a correlation between TMJ and periodontal disease has been observed (Kumar, Yashoda and Puranik, 2017).

Individual AEH157 also had hits to *Rothia dentocariosa* (~2,000 mapped reads); *Streptococcus sanguinis* (~1,000 mapped reads), which are pathogenic bacteria associated with periodontal disease development. As for the AEH158 individual, we identified hits to *Tannerella forsythia* (~1,000 mapped reads), this bacterium is identified as one of the main contributors to the progression of periodontitis. *T. forsythia* acts in synergy with two other bacteria (*Treponema denticola, Porphyromonas gingivalis*), resulting in the so-called ‘red complex’ (Amano *et al.*, 2014). Previous aDNA studies have proved a longstanding relationship with humans for

thousands years (Warinner *et al.*, 2014; Weyrich *et al.*, 2017; Velsko *et al.*, 2019; Bravo-Lopez *et al.*, 2020; Philips *et al.*, 2020).

The interaction of certain aspects related to hunter-gatherer's lifestyle, like the type of food consumed and oral hygiene habits, could have influenced the development of infectious disease in the oral cavity (Otero and Novellino, 2011). In addition, carbohydrate consumption increased progressively after 1000 BP, when pottery technology was adopted and consequently the processed food consumption could have promoted the development of oral diseases in hunter gatherers during pre-contact times (Otero and Novellino, 2011; Suby, 2020).

7.3. Conclusions

This is the first study demonstrating the presence of ancient pathogen DNA in hunter-gatherers from the Central-Northern Coast of Argentinian Patagonia. The molecular evidence of pathogens associated with zoonotic infectious diseases (*E. rhusiopathiae),* represents an opportunity to screen pathogen aDNA from sea and land fauna samples in this area.

According to the characterization of 19 hunter-gatherer's metagenomes, three and six individuals out of 19 individuals analyzed, had bacterial aDNA associated with oral cavity and systemic infectious, respectively. These findings provide a finer analysis into the understanding of the infectious diseases that circulated in the hunter-gatherer's societies from the north of Patagonia. The integration of paleopathological, bioarchaeological and paleogenomics evidence from the individuals analyzed in this study, allowed to gain insights into the hunter-gatherer's health status and in some cases a likely cause of death. Of note, we propose that AEH158 and AEH159 individuals could have died because of the gravity of the infectious diseases caused by the ancient pathogen detected, *C. perfringens and C. tetani*, and *Clostridium perfringens* (~56,00 reads) and *Proteus mirabilis*, respectively. This is consistent with the estimated young age death for the AEH158 and AEH159 individuals was 14-16 and 20-22 years old, respectively.

Further genomic analyses will provide additional insights into how subsistence strategies influenced the hunter-gatherer's health, and how the genomes of these pathogens have evolved in the past ca. 6000 years.

7.4. Materials and Methods

7.4.1. Ethics Statement

The sample processing for aDNA analysis was approved by the Ethics Committees of Puerto Madryn Zonal Hospital (Resol.004/2017) and by the Cultural Secretariat of Chubut Province, Argentina, following the Law 25.517 and its regulatory decree 701/1of 0. The transportation archeological samples from Argentina to Mexico, were approved by the National Institute of Anthropology (INAPL, Argentina), and the National Registry of Archaeological Sites, Collections and Objects (RENYCOA, Argentina).

7.4.2. Archaeological samples provenance

The remains of the individuals analyzed in this study are deposited at the Repositorio de Antropología Biológica del Instituto de Diversidad y Evolución Austral del IDEAus-CONICET-CCT, which is the official repository of the Culture Secretary of Chubut province, based on the Law XI N° 11 “Régimen de las Ruinas y Yacimientos Arqueológicos, Antropológicos y Paleontológicos”.

7.4.2.1. Central Coastal Argentinian Patagonian landscape

The study area encompasses two regions; the coast between 42º S and 43º 15’ S–64º 58’ W, and the lower valley of Chubut river (43º 22’ S–65º 51’ W). Tertiary plateaus are located in the coastal landscape, interrupted by the valley of the Chubut river. As the sea level decreased the seashore is exposed, facilitating the recollection of marine invertebrates (Otero and Novellino, 2011). The predominant climate is arid temperate, and the permanent fres-water sources are scarce, which are the Chubut River and the center of the Valdés peninsula. The environmental conditions allow the availability of diverse food sources, like plants species rich in carbohydrate consumed by the hunter-gathers from this area, are “macachín” (*Arjona tuberosa*), “piquillín” (*Condalia microphylla*), “alpataco” (*Prosopis alpataco*) and “algarrobillo” (*Prosopis denudans*). In addition, terrestrial animals were valuable food resources, in particular, the guanaco (*Lama guanicoe*) and the lesser Patagonian rhea (*Pterocnemia pennata*). Their consumption was dependent on their seasonal energetic supply (Casamiquela, 1983; Otero and Novellino, 2011). Another nutrimental source of food was the marine fauna, including southern sea lions (*Otaria flavescens*), sea elephants *(Mirounga leonina*), southern right whale (*Eubalaena australis*), and orcas (*Orcinus orca*). Along the lower valley of the Chubut river, other species were accessible,

like perches and catfish, waterfowl (ducks, *Chloephaga geese*, swans) and a large rodent called coypu *(Myocastor coypus*) (Otero and Novellino, 2011).

7.4.3. DNA extraction and library preparation

To avoid contamination, both DNA extraction and library preparation steps were performed in dedicated clean lab facilities at the Paleogenomics Laboratory, University of California Santa Cruz. Well-preserved bones and teeth were selected. Bone samples were decontaminated by removing the superficial layer using a Dremel tool and a metallic bit, and UV irradiated for 30 min in all surfaces. After decontamination, a piece of bone was pulverized using a Retsch™ MM 400 Mixer Mill. Regarding teeth samples, these were decontaminated by wiping the surface with 10% bleach solution, rinsed with 70% ethanol and UV irradiated for 15 min on each side. To obtain the root cementum, teeth were cut in the middle line and drilled using a Dremel tool and a metallic bit. Finally, a piece of the root cementum was cut with a Dremel and pulverized with a mixer mill. All metallic material was thoroughly cleaned with a bleach solution after use, rinsed with 70% ethanol and UV-irradiated for one hour to avoid cross-contamination between samples. DNA was extracted following the protocol in Dabney *et al.* 2013. Up to 120 mg of bone/teeth powder was incubated overnight at 37ºC, using a proteinase K/EDTA solution, and DNA extracted using a silica-based and guanidine method. Finally, DNA was eluted in 50 µl of the elution buffer. The DNA extracts were then built into double-stranded DNA libraries, with 7-bp single-index barcoding to allow for multiplexing sequencing, according to the protocol used in Meyer and Kircher 2010. Libraries were sequenced on an Illumina NextSeq 500 at Stanford University (Palo Alto, CA, USA) in paired-end mode to generate approximately 1 million reads per library (2 x 75nt).

7.4.2. Sequencing data quality control and taxonomic classification

Processing of fastq files (clipping adapter and merging sequence pairs) was done using AdapterRemoval v. 2 (Schubert, Lindgreen and Orlando, 2016) . Only reads 30 bp and longer and with an average quality above 33 were retained for downstream analyses. We mapped to the human reference genome (version GRCh37, hg19), using the ‘bwa aln’ option with (flags −l500 −t 16 −q 25). We selected the unmapped reads for taxonomic classification with Kraken2 (Wood, Lu and Langmead, 2019) usingNCBI’s RefSeq database, including bacterial, archaea and viral genomes. Metagenomic profiles were analyzed with Pavian (Breitwieser and Salzberg, 2020), focusing on pathogenic taxa. The visualization of relative abundance of taxonomic groups was performed with the phyloseq library (McMurdie and Holmes, 2013).

7.5. References

Amano, A., Chen, C., Honma, K., Li, C., Settem, R. P., & Sharma, A. (2014). Genetic characteristics and pathogenic mechanisms of periodontal pathogens. *Advances in Dental Research*, *26*(1), 15–22.

Arning, N., & Wilson, D. J. (2020). The past, present and future of ancient bacterial DNA. *Microbial Genomics*, *6*(7). https://doi.org/10.1099/mgen.0.000384.

Aspillaga, E., Ocampo, C., Rivas, P., & Others. (1999). Restos óseos humanos de contextos arqueológicos del área de Isla de Navarino: indicadores de estilos de vida en indígenas canoeros. *Anales del Instituto de la Patagonia*. http://bibliotecadigital.umag.cl/handle/20.500.11893/1345

Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Meier-Kolthoff, J. P., Klenk, H.-P., Clément, C., Ouhdouch, Y., & van Wezel, G. P. (2016). Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiology and Molecular Biology Reviews: MMBR*, *80*(1), 1–43.

Bravo-Lopez, M., Villa-Islas, V., Arriaga, C. R., Villaseñor-Altamirano, A. B., Guzmán-Solís, A., Sandoval-Velasco, M., Wesp, J. K., Alcantara, K., López-Corral, A., Gómez-Valdés, J., Mejía, E., Herrera, A., Meraz-Moreno, A., de la Luz Moreno-Cabrera, M., Moreno-Estrada, A., Nieves-Colón, M. A., Olvera, J., Pérez-Pérez, J., Iversen, K. H., … Ávila-Arcos, M. C. (2020). Paleogenomic insights into the red complex bacteria Tannerella forsythia in Pre-Hispanic and Colonial individuals from Mexico. En *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 375, Número 1812, p. 20190580). https://doi.org/10.1098/rstb.2019.0580

Breitwieser, F. P., & Salzberg, S. L. (2020). Pavian: Interactive analysis of metagenomics data for microbiomics and pathogen identification. *Bioinformatics*, 4.

Brüggemann, H., Bäumer, S., & Fricke, W. F. (2003). The genome sequence of Clostridium tetani, the causative agent of tetanus disease. *Proceedings of the*. https://www.pnas.org/content/100/3/1316/

Casamiquela, R. M. (1983). La significación del guanaco (Lama guanicoe) en el ámbito pampeano-patagónico: aspectos corológico, ecológicos, etológicos y etnográficos. *Mundo Ameghiniano*, *4*, 21–46.

Cassir, N., Benamar, S., & La Scola, B. (2016). Clostridium butyricum: from beneficial to a new emerging pathogen. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *22*(1), 37–45.

Chen, C., Feng, P., & Slots, J. (2020). Herpesvirus-bacteria synergistic interaction in periodontitis. *Periodontology 2000*, *82*(1), 42–64.

Cohen, J. E., Wang, R., Shen, R.-F., Wu, W. W., & Keller, J. E. (2017). Comparative pathogenomics of Clostridium tetani. *PloS One*, *12*(8), e0182909.

Constantinescu, F. (1997). Hombres y mujeres de cerro Los Onas: presentes, ausentes... los relatos de sus huesos. *Anales del Instituto de la Patagonia*, *59*, 74.

Constantinescu, F. (1999). Evidencias bioantropológicas para modos de vida cazador recolector terrestre y marítimo en los restos óseos humanos de Tierra del Fuego. *Anales del Instituto de la Patagonia*. http://www.bibliotecadigital.umag.cl/handle/20.500.11893/1335

Curtis, M. A., Diaz, P. I., & Van Dyke, T. E. (2020). The role of the microbiota in periodontal disease. *Periodontology 2000*, *83*(1), 14–25.

Dlamini, N., Morris, A. G., & Sealy, J. (2016). Carbon Isotopes and Dental Caries as Evidence for Regional Variation in the Diets of Early Farming Communities from Katanga, Democratic Republic of the Congo. *Journal of African Archaeology*, *14*(2), 135–153.

Flensborg, G. (2015). Health and disease of hunter-gatherer groups from the eastern Pampa–Patagonia transition (Argentina) during the Late Holocene. *Anthropological science: journal of*

the Anthropological Society of Nippon = Jinruigaku zasshi, *advpub*. https://doi.org/10.1537/ase.151001

Forde, T., Biek, R., Zadoks, R., Workentine, M. L., De Buck, J., Kutz, S., Opriessnig, T., Trewby, H., van der Meer, F., & Orsel, K. (2016). Genomic analysis of the multi-host pathogen Erysipelothrix rhusiopathiae reveals extensive recombination as well as the existence of three generalist clades with wide geographic distribution. En *BMC Genomics* (Vol. 17, Número 1). https://doi.org/10.1186/s12864-016-2643-0

Gaddy, H. G. (2020). Using local knowledge in emerging infectious disease research. *Social Science & Medicine*, *258*, 113107.

Gómez Otero, J. (2006). *Dieta, uso del espacio y evolución en poblaciones cazadoras-recolectoras de la costa centro-septentrional de Patagonia durante el Holoceno medio y tardío*. http://repositorio.filo.uba.ar/handle/filodigital/1274

Gomez Otero, J., & Moreno, J. E. (2015). *Archaeological evidence for hunter-gatherers mobility and diet changes during eighteen and nineteen centuries in the central Patagonian Atlantic coast*. https://ri.conicet.gov.ar/handle/11336/5613

Kämpfer, P., Schäfer, J., Lodders, N., & Busse, H.-J. (2010). Brevibacterium sandarakinum sp. nov., isolated from a wall of an indoor environment. *International Journal of Systematic and Evolutionary Microbiology*, *60*(Pt 4), 909–913.

Kersters, K., De Vos, P., Gillis, M., Swings, J., Vandamme, P., & Stackebrandt, E. (2006). Introduction to the Proteobacteria. En *The Prokaryotes : a handbook on the biology of bacteria: Vol. 5 : Proteobacteria : alpha and beta subclasses* (pp. 3–37). Springer.

Kiu, R., & Hall, L. J. (2018). An update on the human and animal enteric pathogen Clostridium perfringens. *Emerging Microbes & Infections*, *7*(1), 141.

Kumar, V., Yashoda, R., & Puranik, M. (2017). Association between periodontal disease temporomandibular disorders and rheumatoid arthritis among patients visiting rheumatology centers in Bengaluru City: A cross-sectional study. En *Journal of Indian Association of Public Health Dentistry* (Vol. 15, Número 4, p. 312). https://doi.org/10.4103/jiaphd.jiaphd_125_17

Lalucat, J., Bennasar, A., Bosch, R., García-Valdés, E., & Palleroni, N. J. (2006). Biology of Pseudomonas stutzeri. *Microbiology and Molecular Biology Reviews: MMBR*, *70*(2), 510–547.

Lugli, G. A., Milani, C., Mancabelli, L., Turroni, F., Ferrario, C., Duranti, S., van Sinderen, D., & Ventura, M. (2017). Erratum to: Ancient bacteria of the Ötzi's microbiome: a genomic tale from the Copper Age. *Microbiome*, *5*(1), 23.

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PloS one*, *8*(4), e61217.

Nagano, N., Isomine, S., Kato, H., Sasaki, Y., Takahashi, M., Sakaida, K., Nagano, Y., & Arakawa, Y. (2008). Human fulminant gas gangrene caused by Clostridium chauvoei. *Journal of Clinical Microbiology*, *46*(4), 1545–1547.

Otero, J. G. (2007). Isótopos estables, dieta y uso del espacio en la costa atlántica centro-septentrional y el valle inferior del río Chubut (Patagonia, Argentina). En *Arqueología de Fuego-Patagonia. Levantando Piedras, desenterrando huesos… y develando arcanos* (pp. 151–161). CEQUA Punta Arenas.

Otero, J. G., & Novellino, P. (2011). Diet, nutritional status and oral health in hunter-gatherers from the central-northern coast of Patagonia and the Chubut river lower valley, Argentina. *International journal of osteoarchaeology*, *21*(6), 643–659.

Philips, A., Stolarek, I., Handschuh, L., Nowis, K., Juras, A., Trzciński, D., Nowaczewska, W., Wrzesińska, A., Potempa, J., & Figlerowicz, M. (2020). Analysis of oral microbiome from fossil human remains revealed the significant differences in virulence factors of modern and ancient Tannerella forsythia. *BMC Genomics*, *21*(1), 402.

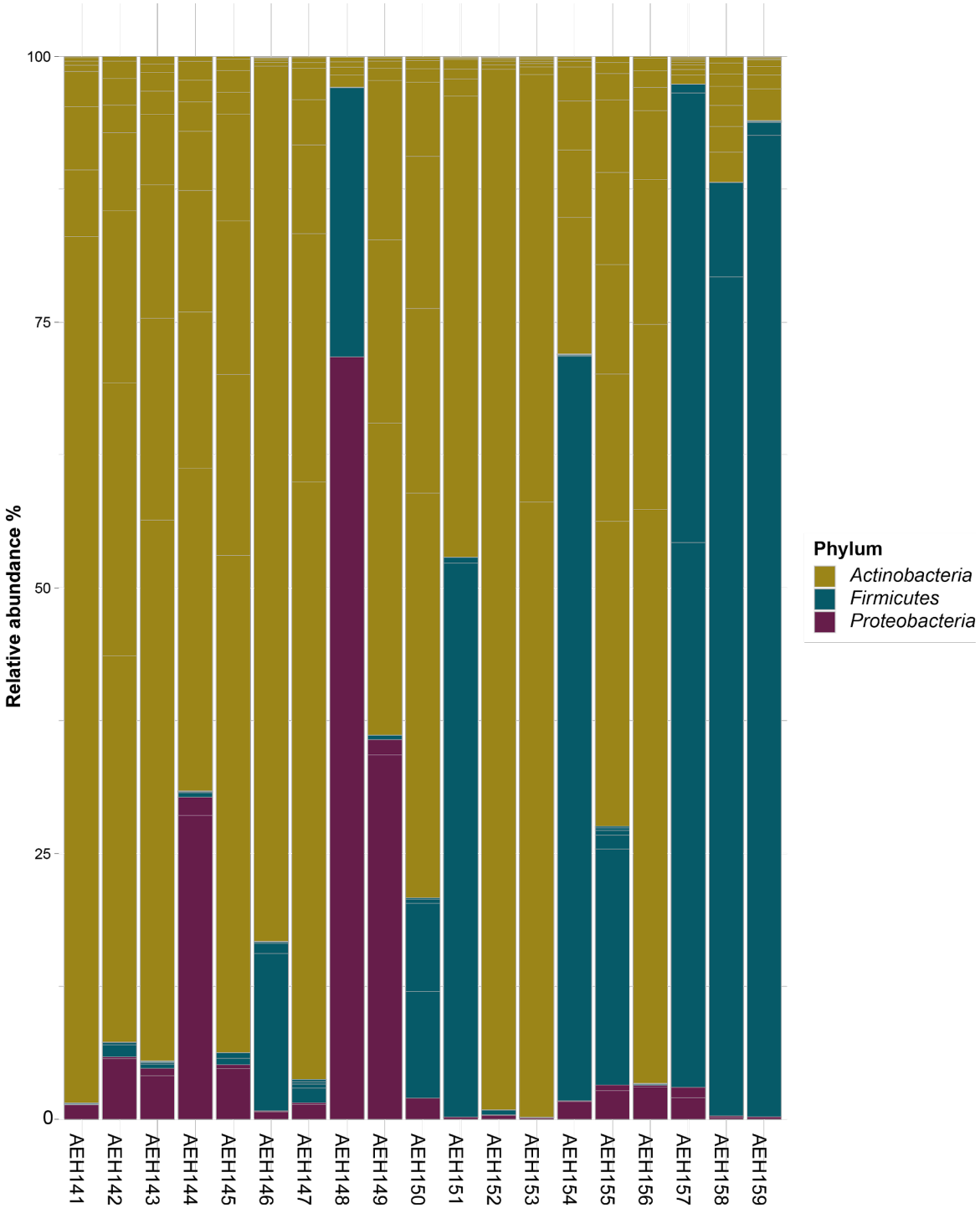
Piau, C., Arvieux, C., Bonnaure-Mallet, M., & Jolivet-Gougeon, A. (2013). Capnocytophaga spp. involvement in bone infections: a review. *International Journal of Antimicrobial Agents*, *41*(6), 509–515.

Revitt-Mills, S. A., Rood, J. I., & Adams, V. (2015). Clostridium perfringens extracellular toxins and enzymes: 20 and counting. *Microbiology Australia*, *36*(3), 114–117.
Rychener, L., InAlbon, S., Djordjevic, S. P., Chowdhury, P. R., Ziech, R. E., de Vargas, A. C., Frey, J., & Falquet, L. (2017). Clostridium chauvoei, an Evolutionary Dead-End Pathogen. *Frontiers in Microbiology*, *8*, 1054.
Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC research notes*, *9*(1), 88.
Spyrou, M. A., Bos, K. I., Herbig, A., & Krause, J. (2019). Ancient pathogen genomics as an emerging tool for infectious disease research. *Nature Reviews. Genetics*, *20*(6), 323–340.
Suby, J. A. (2020). Paleopathological Research in Southern Patagonia: An Approach to Understanding Stress and Disease in Hunter-Gatherer Populations. *Latin American Antiquity*, *31*(2), 392–408.
Velsko, I. M., Fellows Yates, J. A., Aron, F., Hagan, R. W., Frantz, L. A. F., Loe, L., Martinez, J. B. R., Chaves, E., Gosden, C., Larson, G., & Warinner, C. (2019). Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. *Microbiome*, *7*(1), 102.
Vesth, T., Ozen, A., Andersen, S. C., Kaas, R. S., Lukjancenko, O., Bohlin, J., Nookaew, I., Wassenaar, T. M., & Ussery, D. W. (2013). Veillonella, Firmicutes: Microbes disguised as Gram negatives. *Standards in Genomic Sciences*, *9*(2), 431–448.
Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R. Y., Fiddyment, S., Speller, C., Hendy, J., Charlton, S., Luder, H. U., Salazar-García, D. C., Eppler, E., Seiler, R., Hansen, L. H., Castruita, J. A. S., … Cappellini, E. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics*, *46*(4), 336–344.
Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A. G., Alt, K. W., Caramelli, D., Dresely, V., Farrell, M., Farrer, A. G., Francken, M., Gully, N., Haak, W., Hardy, K., Harvati, K., Held, P., Holmes, E. C., … Cooper, A. (2017). Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature*, *544*(7650), 357–361.
Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. En *bioRxiv* (p. 762302). https://doi.org/10.1101/762302.

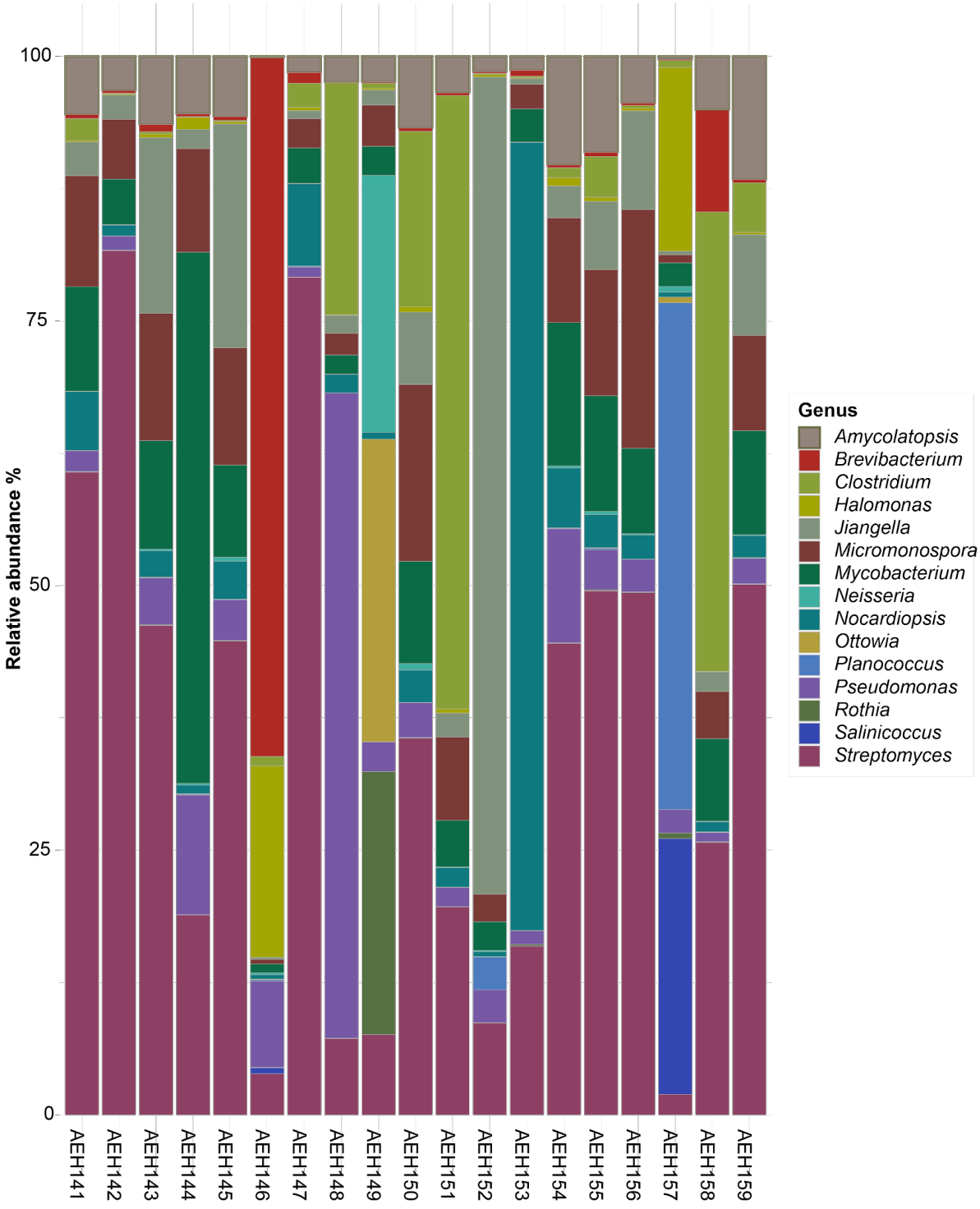
Supplementary Information

Reporte de proyecto. Paleogenomic analysis of pathogenic bacteria in hunter-gatherers from Central Coastal Argentinian Patagonian.

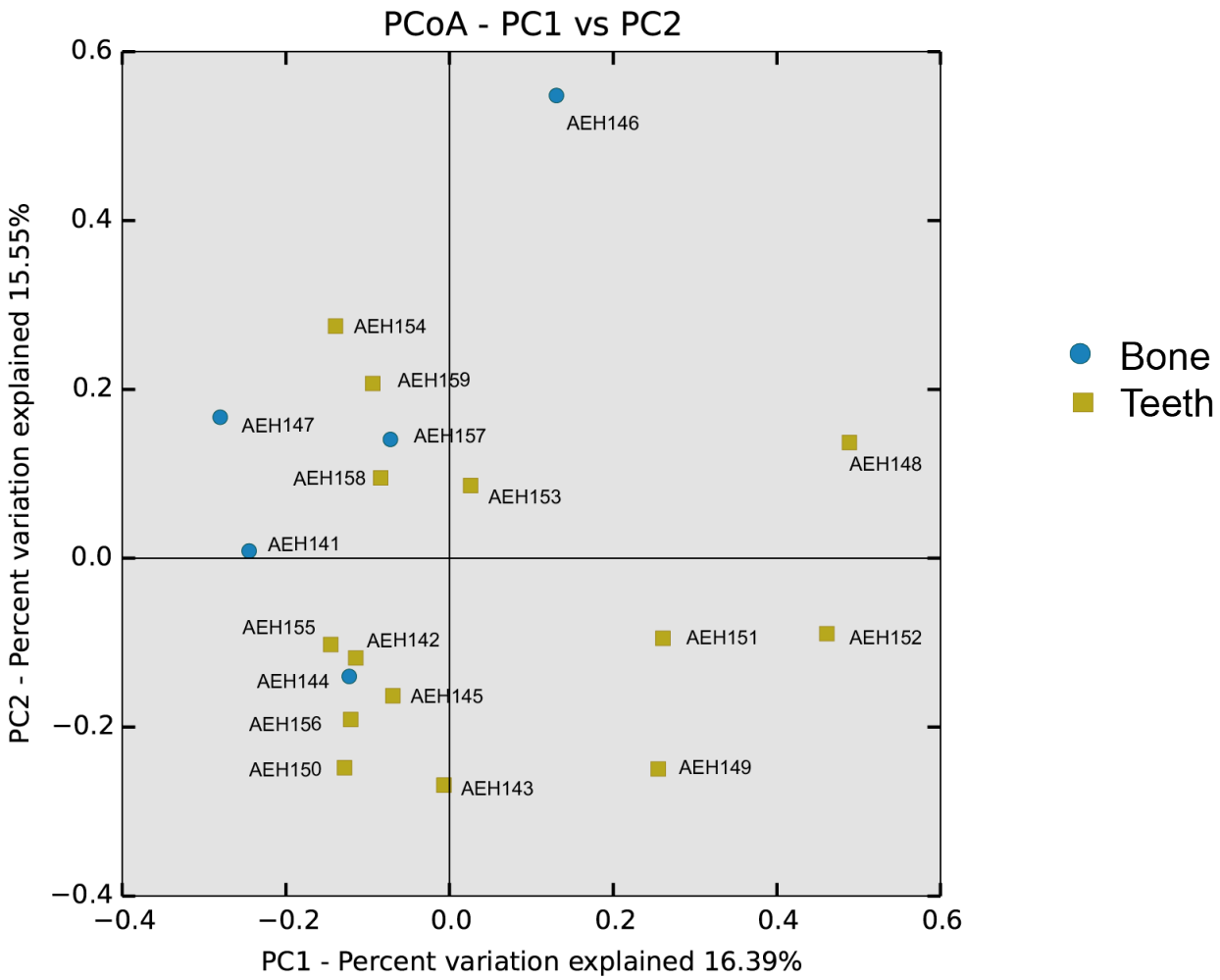
Supplementary Figures



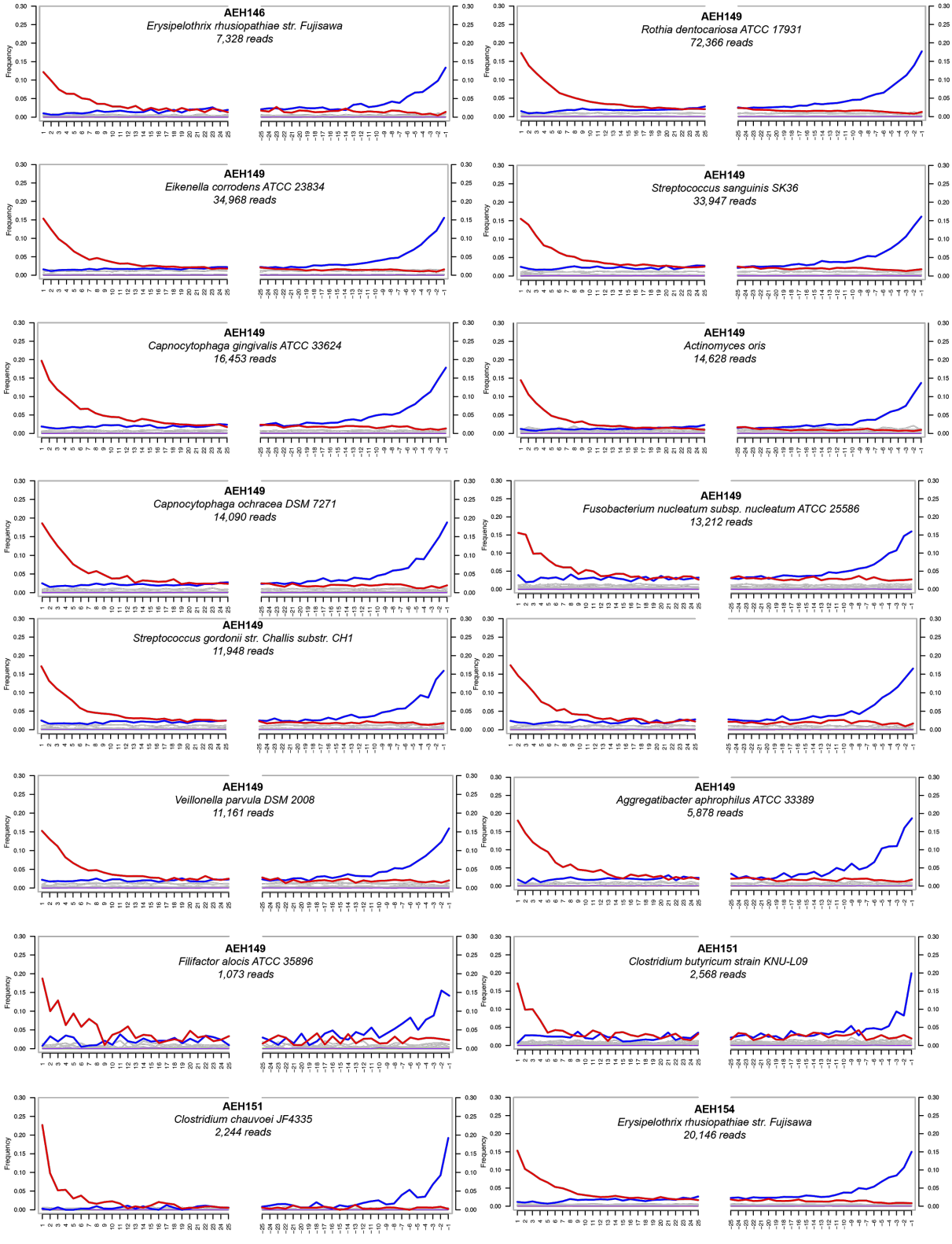
Supplementary figure 1. Relative abundance of the top 3 bacterial phyla from Central Patagonia analyzed in this study. Taxonomic assignments were obtained using Kraken2.

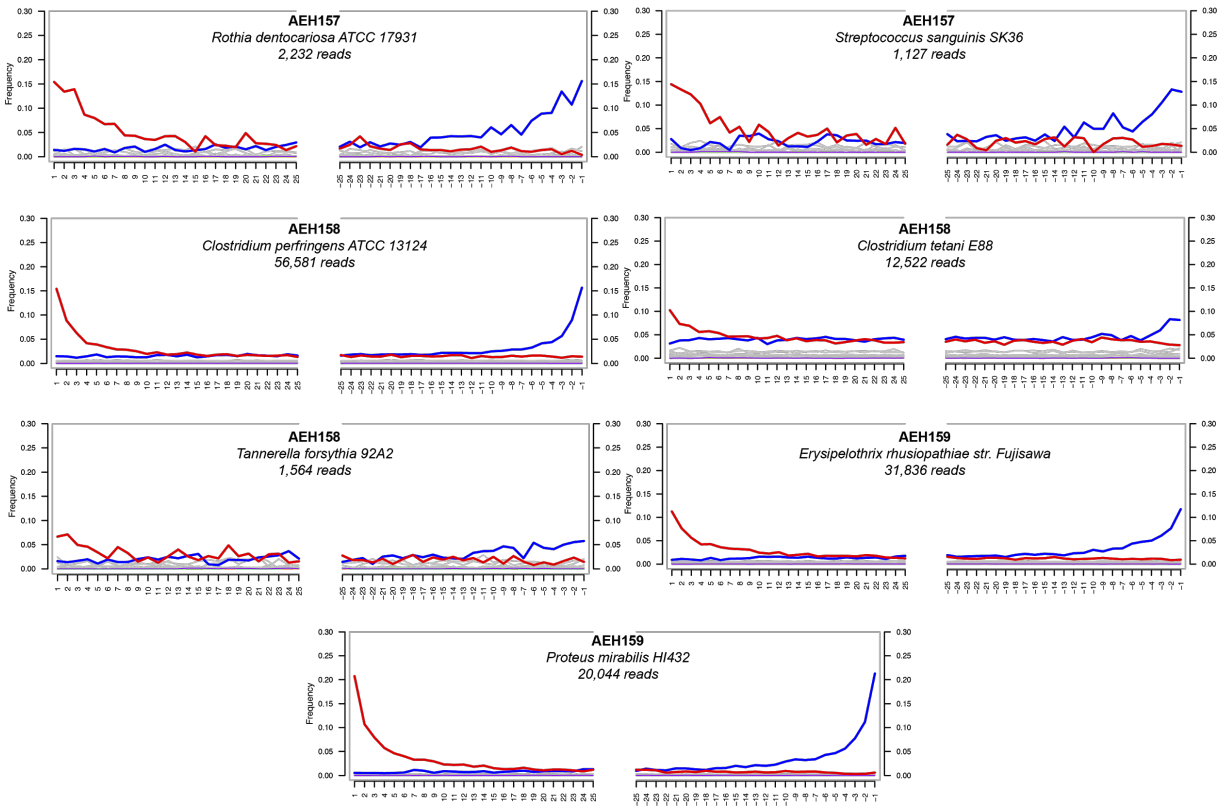


Supplementary figure 2. Relative abundance of the top-ranked 15 bacterial genera detected in the samples analyzed in this study. Taxonomic assignments were obtained using Kraken2.



Supplementary figure 3. Non-metric multidimensional scaling based on Kraken2 relative abundance of taxa differentiated by substrate (bone and teeth).


AEH146
Erysipelothrix rhusiopathiae str. Fujisawa
7,328 reads
AEH149
Rothia dentocariosa ATCC 17931
72,366 reads
AEH149
Eikenella corrodens ATCC 23834
34,968 reads
AEH149
Streptococcus sanguinis SK36
33,947 reads
AEH149
Capnocytophaga gingivalis ATCC 33624
16,453 reads
AEH149
Actinomyces oris
14,628 reads
AEH149
Capnocytophaga ochracea DSM 7271
14,090 reads
AEH149
Fusobacterium nucleatum subsp. nucleatum ATCC 25586
13,212 reads
AEH149
Streptococcus gordonii str. Challis substr. CH1
11,948 reads
AEH149
Veillonella parvula DSM 2008
11,161 reads
AEH149
Aggregatibacter aphrophilus ATCC 33389
5,878 reads
AEH149
Filifactor alocis ATCC 35896
1,073 reads
AEH151
Clostridium butyricum strain KNU-L09
2,568 reads
AEH151
Clostridium chauvoei JF4335
2,244 reads
AEH154
Erysipelothrix rhusiopathiae str. Fujisawa
20,146 reads
Frequency



Supplementary figure 4. Damage plots of ancient pathogen DNA identified in the samples analyzed in this study. In each plot the pathogen identified and the reads obtained after mapping to the corresponding reference genome are indicated. The X axis shows the position (nt) on the 5' and 3' end of the read, and the Y axis the frequency of C > T changes at 5' end of fragments (red line), whereas G > A substitutions at 3' end of fragments (blue line).

Supplementary Tables

Supplementary Table 1. Bioarcheological information and pathogens identified in the samples analyzed in this study.

ID	Site/Individual	Region/environment	long/lat	Province	Date ^{14}C yr BP	Estimated age (years)	Morphological and molecular sex estimation	Diet (^{13}C/^{15}N)	Material analyzed	Paleopathology lesions	Pathogens associated	
											Systemic infection	Oral cavity infection
AEH141	Chacra 375 Ind. 1	Lower Valley of Chubut river	65°18'W 43°18'S	Chubut	6070±80	30-45	Male	Mixed*	Bone (metatarsal)	Periapical lesions		
AEH142	Punta Cuevas 2	Golfo Nuevo coast	65°03'W 42°46'S	Chubut	2640±50	42-49	Male	Mixed	Teeth (incisor)	N.D.		
AEH143	Tehuelches St.	Golfo Nuevo coast	65°03'W 42°46'S	Chubut	2410±50	30-45	Female	Marine	Teeth (molar)	N.D.		
AEH144	Punta Delgada	Península Valdés	63°38'W 42°46'S	Chubut	2010±50	>50	Male	Marine	Bone (ulna)	Periosteal reaction in the right tibia		
AEH145	El Elsa	Lower Valley of Chubut river	65°06'W 43°18'S	Chubut	1990±60	35-52	Female	Terrestrial	Teeth (molar)	Periapical lesions and periodontitis		
AEH146	Loma Grande-Chacra 151 Ind. 1	Lower Valley of Chubut river	65°18'W 43°15'S	Chubut	1480±60	>45	Female	Terrestrial	Bone (tarsal)	N.D.	*Erysipelothrix rhusiopathiae str. Fujisawa*	
AEH147	Chacra 375 Ind. 2	Lower Valley of Chubut river	65°18'W 43°18'S	Chubut	1410±70	adult	Female	N.D.	Bone (phalange)	Periodontitis		
AEH148	BonLee neighborhood	Nuevo Gulf coast	65°03'W 42°46'S	Chubut	1400±60	30-40	Female	N.D.	Teeth (molar)	Periapical lesions		
AEH149	Loma grande-Chacra 151 Ind. 2	Lower Valley of Chubut river	65°18'W 43°15'S	Chubut	1390±60	>50	Male	Terrestrial	Teeth (molar)	N.D.		*Rothia dentocariosa ATCC 17931* *Actinomyces oris* *Veillonella parvula DSM 2008* *Aggregatibacter aphrophilus ATCC 33389* *Capnocytophaga gingivalis ATCC 33624* *Capnocytophaga ochracea DSM 7271* *Capnocytophaga sputigena* *Eikenella corrodens ATCC 23834* *Filifactor alocis ATCC 35896* *Fusobacterium nucleatum subsp. nucleatum ATCC 25586* *Streptococcus gordonii str. Challis substr. CH1* *Streptococcus sanguinis SK36*
AEH150	Los Lobos	Lower Valley of Chubut river	65°27'W 43°37'S	Chubut	1090±70	adult	Male	N.D.	Teeth (molar)	Periodontitis		
AEH151	Punta Leon	Punta León coast	64°28'W 43°04'S	Chubut	1050±50	20-22	Female	Marine	Teeth (molar)	N.D.	*Clostridium chauvoei JF-4335* *Clostridium butyricum strain KNU-L09*	
AEH152	La Azucena Ind. 1	Península Valdés	64°09'W 42°51'S	Chubut	880±50	30-39	Female	Marine	Teeth (incisor)	N.D.		
AEH153	La Azucena Ind. 2	Península Valdés	64°09'W 42°51'S	Chubut	880±50	25-35	Female	Marine	Teeth (molar)	N.D.		
AEH154	El Golfito Ind. 2	Nuevo Gulf coast	65°03'W 42°46'S	Chubut	770±50	35-45	Male	Mixed	Teeth (molar)	N.D.	*Erysipelothrix rhusiopathiae str. Fujisawa*	
AEH155	Abril	North San Matías coast	64°45'W 40°49'S	Río Negro	770±70	adult	Female	Mixed	Teeth (incisor)	N.D.		
AEH156	Villarino St.	Nuevo Gulf coast	65°03'W 42°46'S	Chubut	550±60	30-45	Female	N.D.	Teeth (molar)	Lesiones periapicales y enfermedad periodontal		
AEH157	El Doradillo Ind. 1	Nuevo Gulf coast	64°59'W 42°39'S	Chubut	370±50	30-40	Male	Mixed	Bone (metacarpal)	N.D.		*Capnocytophaga gingivalis ATCC 33624* *Rothia dentocariosa ATCC 17931* *Streptococcus sanguinis SK36*
AEH158	Gastre Ind. 1	Gastre, North Central plateaus	69°13'W 42°16'S	Chubut	350±50	14-16	Male	Mixed	Teeth (incisor)	N.D.	*Clostridium perfringens ATCC 13124* *Clostridium tetani E88*	*Tannerella forsythia 92A2*
AEH159	Barranca Norte Ind. 1	Lower Valley of Chubut river	65°05'W 43°18'S	Chubut	310±70	20-22	Female	Mixed	Teeth (canine)	Periosteal reaction in the left tibia	*Erysipelothrix rhusiopathiae str. Fujisawa* *Proteus mirabilis HI432*	

Supplementary table 2. Mapping statistics of the aDNA pathogen identified in the samples analyzed in this study.

ID	Pathogen	Causing disease	Genome size	Total sequences (after adapter removal)	Uniquely mapped reads	Average depth genomewide	Positions covered at least once	Mapped reads average length	% damage of first 5-prime base	% damage of second 5-prime base	% damage of first 3-prime base	% damage of second 3-prime base
AEH146	*Erysipelothrix rhusiopathiae str. Fujisawa*	Erysipeloid	1.77806	9,737,311	7,318	0.202748	314,598	49.5384	0.1217	0.0989	0.1336	0.0984
AEH149	*Rothia dentocariosa ATCC 17931*	Periodontitis	2.49282	21,284,302	72,366	1.31417	1,599,802	45.5149	0.1727	0.1381	0.1771	0.1393
	Eikenella corrodens ATCC 23834	Periodontitis, head and neck infectios	2.24314	21,284,302	34,968	0.73257	943,090	45.4943	0.1534	0.1245	0.1555	0.1206
	Streptococcus sanguinis SK36	Periodontitis	2.36222	21,284,302	33,947	0.64388	956,662	45.3074	0.1549	0.1391	0.1609	0.1349
	Capnocytophaga gingivalis ATCC 33624	Periodontitis	2.75289	21,284,302	16,453	0.289571	618,807	46.4196	0.197	0.145	0.1782	0.1463
	Actinomyces oris	Periodontitis	3.1092	21,284,302	14,628	0.205664	455,759	42.7884	0.1447	0.1059	0.137	0.1067
	Capnocytophaga ochracea DSM 7271	Periodontitis	2.65865	21,284,302	14,090	0.249084	543,006	46.1962	0.1859	0.1535	0.1882	0.1501
	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	Periodontitis	2.4247	21,284,302	13,212	0.284373	450,662	46.8122	0.1558	0.1509	0.1598	0.147
	Streptococcus gordonii str. Challis substr. CH1	Periodontitis	2.19256	21,284,302	11,948	0.245689	408,548	45.1764	0.1711	0.1319	0.159	0.1353
	Capnocytophaga sputigena	Periodontitis	2.99784	21,284,302	11,266	0.180839	440,678	45.8362	0.174	0.1455	0.1652	0.1403
	Veillonella parvula DSM 2008	Periodontitis, meningitis, endocarditis, osteomyelitis,	2.14419	21,284,302	11,161	0.239691	386,855	45.7954	0.1527	0.1303	0.1591	0.1231
	Aggregatibacter aphrophilus ATCC 33389	Periodontitis	2.33287	21,284,302	5,878	0.115941	245,569	46.1022	0.1805	0.1458	0.1868	0.1604
	Filifactor alocis ATCC 35896	Periodontitis	1.93101	21,284,302	1,073	0.025642	34,226	46.1566	0.1872	0.1	0.1415	0.1552
AEH151	*Clostridium butyricum strain KNU-L09*	Botulism	4.5972	5,793,170	2,568	1.04195	105,861	43.8516	0.1711	0.0986	0.199	0.0824
	Clostridium chauvoei JF4335	Fulminant human gas gangrene	2.8837	5,793,170	2,244	0.034072	95,404	43.8458	0.2269	0.0979	0.192	0.0924
AEH154	*Erysipelothrix rhusiopathiae str. Fujisawa*	Erysipeloid	1.77806	16,949,811	20,146	0.535976	670,520	47.5697	0.1533	0.102	0.1499	0.1069
AEH157	*Rothia dentocariosa ATCC 17931*	Periodontitis	2.49282	16,846,943	2,232	0.0422	100,504	47.3907	0.154	0.1341	0.156	0.1075
	Streptococcus sanguinis SK36	Periodontitis	2.36222	16,846,943	1,127	0.021798	50,111	46.2067	0.1441	0.133	0.1282	0.133
AEH158	*Clostridium perfringens ATCC 13124*	Gas gangrene, food poisoning and non-foodborne diarrhoea, and enterocolitis	3.41292	23,446,224	56,581	0.733108	1,543,019	42.2004	0.1541	0.0886	0.1566	0.0895
	Clostridium tetani E88	Tetanus disease	2.83879	23,446,224	12,522	0.190267	306,260	43.6739	0.1025	0.0732	0.0816	0.0833
	Tannerella forsythia 92A2	Periodontitis	3.30018	23,446,224	1,564	0.021555	70,684	46.945	0.0666	0.0712	0.0576	0.0552
AEH159	*Erysipelothrix rhusiopathiae str. Fujisawa*	Erysipeloid	1.77806	8,578,176	31,836	0.868969	891,064	48.8061	0.1124	0.0765	0.1173	0.0762
	Proteus mirabilis HI432	Infections, including those of wounds, the eye, the gastrointestinal tract, and the urinary tract.	4.04145	8,578,176	20,044	0.210112	749,038	42.5985	0.2074	0.107	0.2129	0.1117

8. DISCUSIÓN

La realización de esta tesis permitió generar avances pioneros en el área de genómica de patógenos. A través de la caracterización metagenómica de 53 individuos de México (20 del periodo prehispánico y 33 del colonial) y 19 de la Costa Norte de la Patagonia Argentina; logramos identificar patógenos asociados a enfermedades sistémicas y a infecciones de la cavidad oral. Con la finalidad de aumentar el contenido endógeno de estos patógenos, diseñamos estrategias de captura y enriquecimiento en el genoma de *Tannerella forsythia* en siete individuos del periodo prehispánico y cinco del periodo colonial. También aplicamos dicha estrategia para caracterizar el genoma antiguo del agente causante de fiebre paratifoidea (*Salmonella enterica* Paratyphi C) presente en un individuo del periodo colonial de la Ciudad de México. Por otra parte, se realizó el análisis bioinformático de datos obtenidos a partir de individuos cazadores-recolectores provenientes de la costa norte de la Patagonia Argentina, logrando identificar el aDNA de patógenos asociados a enfermedades infecciosas sistémicas y de la cavidad oral, principalmente (Figura 3). A continuación se discuten los principales hallazgos obtenidos en el artículo de investigación, y los dos reportes de proyecto, desarrollados en esta tesis.

8.1. Identificación de agentes causantes de enfermedades infecciosas en poblaciones humanas a través de la paleogenómica

Los avances en la Paleogenómica han permitido el desarrollo de una subrama especializada en la genómica de patógenos antiguos, cuyo objetivo principal es la identificación y el análisis del genoma de agentes causantes de enfermedades infecciosas en poblaciones antiguas.

Las estrategias metodológicas y analitícas dentro de esta área han permitido la reconstrucción de 264 genomas de patógenos humanos, recuperados a partir de restos arqueológicos. De los cuales 202 son bacterianos, cuatro de eucariontes y 60 de virus (Fellows Yates *et al.*, 2020). La mayoría de los restos arqueológicos analizados provienen de Europa. Sin embargo, hasta el momento se han reconstruido 12 genomas de patógenos a partir de restos arqueológicos de México (Schuenemann *et al.*, 2018; Vågene *et al.*, 2018; Barquera *et al.*, 2020).

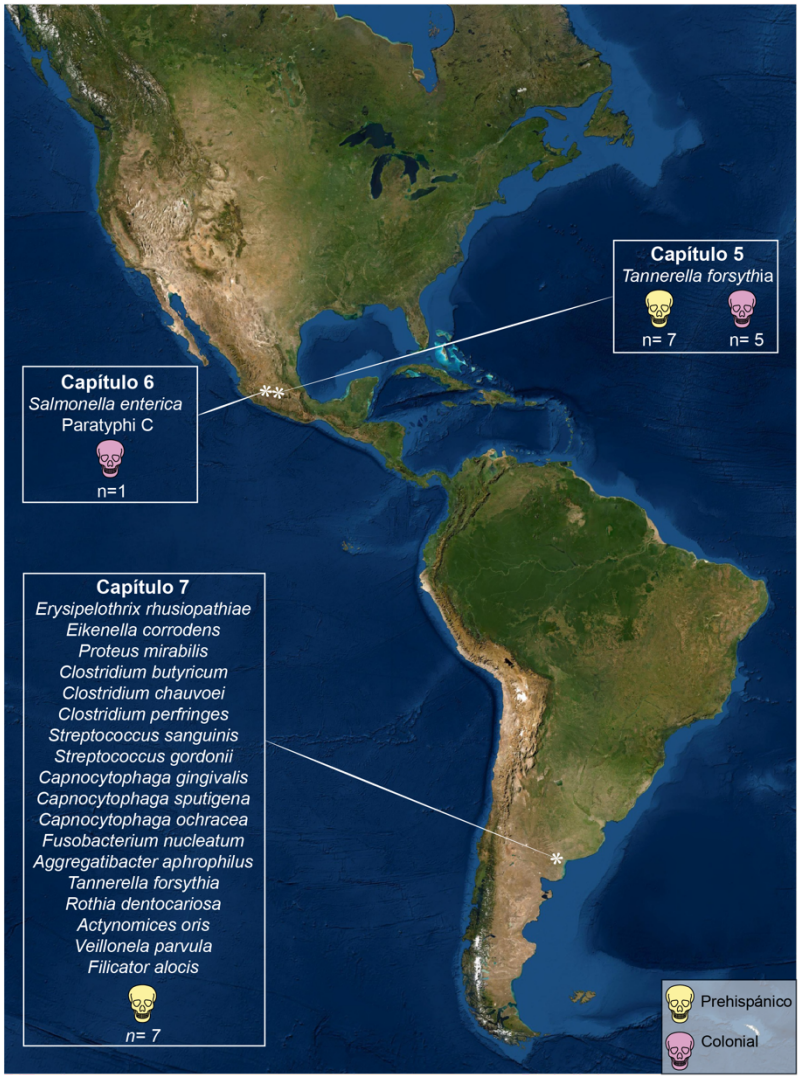


Figura 3. Ubicación geográfica de los individuos considerados en este estudio. Se muestran los patógenos identificados en cada región, así como el número de individuos y el periodo en el que fueron identificado. Se indica el capítulo en el que se describen.

Dentro del trabajo desarrollado en esta tesis doctoral, se llevó a cabo la extracción de aDNA, construcción de librerías, secuenciación de tipo *shotgun* y análisis metagenómico a partir de muestras de 53 individuos de México y 19 individuos de la Patagonia Argentina. De las 53 muestras de México, 20 pertenecen al periodo prehispánico (540 BCE-1519 CE), provenientes de los estados de Guanajuato, Querétaro, Tlaxcala y Veracruz. Asimismo, se analizaron 33 individuos del periodo colonial (1472-1900 CE) de la Ciudad de México y Querétaro (Figura 1 en capítulo 5). En cuanto a las muestras de Argentina, se analizó el metagenoma de individuos

cazadores-recolectores de la Costa Norte de la Patagonia en un rango temporal de 6000 a 300 AP.

El análisis metagenómico de estos 72 individuos antiguos nos permitió identificar la presencia de aDNA de patógenos asociados a enfermedades infecciosas sistémicas y de la cavidad oral. Con respecto a patógenos asociados a enfermedades infecciosas sistémicas, identificamos *Salmonella enterica Paratyphi C* en un individuo del periodo colonial (320-120 BP) de la Ciudad de México. Además, identificamos *Erysipelothrix rhusiopathiae, Clostridium tetani, Clostridium chauvoei, Clostridium perfringens* y *Proteus mirabilis* en tres individuos cazadores-recolectores de la costa norte de la Patagonia.

Por otra parte, identificamos patógenos asociados con procesos infecciosos en la cavidad oral, como es el complejo rojo (*Tannerella forsythia, Treponema denticola* y *Porphyromonas gingivalis*) en 11 individuos del periodo Prehispánico y Colonial de México. En los cazadores-recolectores identificamos *Rothia dentocariosa, Eikenella corrodens, Streptococcus sanguinis, Capnocytophaga gingivalis, Capnocytophaga ochracea, Fusobacterium nucleatum, Capnocytophaga sputigena, Veillonella parvula, Aggregatibacter aphrophilus, Filifactor alocis* y *Tannerella forsythia* en dos individuos.

El aDNA de los patógenos antes mencionados se identificó en muestras de dientes y cálculo dental. El diente es un sustrato idóneo para este tipo de análisis, ya que al estar en contacto directo con la circulación sanguínea durante la vida del individuo, permite identificar DNA del patógeno causante de una infección a nivel sistémico. Además,el diente proporciona las condiciones idóneas para la preservación del DNA, ya que las capas externas (el esmalte y el cemento) lo protegen de fuentes exógenas de contaminación (Adler *et al.*, 2011). Por su parte, el cálculo dental—placa fosilizada que almacena el DNA del microbioma oral y los restos alimenticios—ha sido ampliamente utilizado en análisis de DNA de agentes causantes de enfermedades infecciosas sistémicas y de la cavidad oral (Warinner *et al.*, 2014; Fotakis *et al.*, 2020).

8.2. Estrategias metodológicas y analiticas en la identificación de aDNA de patógenos humanos

8.2.1. Análisis metagenómico de restos arqueológicos

La caracterización del metagenoma de restos arqueológicos no sólo ha permitido identificar el microbioma presente al momento de la muerte de un individuo, sino también realizar inferencias evolutivas sobre los agentes causantes de enfermedades infecciosas, como por ejemplo de *Helicobacter pylori* (Maixner *et al.*, 2016), *Brucella melitensis* (Kay *et al.*, 2014) y *Mycobacterium leprae* (Neukamm *et al.*, 2020).

El análisis metagenómico de individuos del periodo prehispánico y colonial de México, permitió la identificación del "complejo rojo" bacteriano (*Tannerella forsythia, Porphyromonas gingivalis* y *Treponema denticola*), asociados con enfermedad periodontal (ver capítulo 2); y *Salmonella enterica* Paratyphi C, agente causante de fiebre paratifoidea (ver capítulo 3). Adicionalmente, el análisis metagenómico se extendió a cazadores-recolectores del Norte de la Patagonia Argentina, en el cual se identificaron patógenos asociados a enfermedades sistémicas y enfermedad periodontal.

Algunas de las herramientas analíticas que han permitido la caracterización metagenómica en restos antiguos son *Basic Local Alignment Search Tool* (BLAST) (Altschul *et al.*, 1990), *MEGAN Alignment Tool* (MALT) (Vågene *et al.*, 2018) y Kraken2 (Wood y Salzberg, 2014). Estos algoritmos se basan en un agrupamiento taxonómico de acuerdo a una base de datos de genomas actuales, como las Secuencias de Referencia de NCBI (RefSeq) (O'Leary *et al.*, 2016). La elección del tipo de algoritmo y la base de datos en la clasificación taxonómica de comunidades microbianas de restos antiguos, depende del objetivo del estudio ya que cada uno presenta ventajas y desventajas en su uso (Velsko *et al.*, 2018). El análisis metagenómico realizado en esta tesis se obtuvo a través de Kraken2, el cual se basa en la similitud de los k-meros entre el DNA de interés y la base de referencia de RefSeq, que en este caso incluyó 9,415 genomas bacterianos, 258 genomas de arqueas y 7,540 de virus.

Un programa comúnmente empleado en el análisis de metagenomas antiguos es MALT (p. Ej. en (Vågene *et al.*, 2018; Jensen *et al.*, 2019; Barquera *et al.*, 2020; Guzmán-Solís *et al.*, 2020), dada su sensibilidad y exactitud. Sin embargo, la capacidad computacional para su

funcionamiento, requiere ~500 GB de memoria, lo que rebasaba nuestros recursos computacionales.

Recientemente, se desarrolló una tubería de análisis llamada *HOPS* (Heuristic Operations for Pathogen Screening) (Huebler *et al.*, 2019), en la que, a través del análisis de metagenómico obtenido por *MALT*, se evalúa la autenticidad de aDNA a través de la determinación del número de discordancias entre el DNA de interés con un microorganismos *(edit distance*) y patrones característicos de aDNA. Seguramente, el diseño de distintas estrategias metodológicas para la clasificación taxonómica del metagenoma en restos antiguos continuará su desarrollo e innovación.

Como se puede apreciar, el área de genómica de patógenos ha crecido de manera acelerada en los últimos años, en los que distintas estrategias metodológicas y analíticas han sido implementadas. Una estrategia clave para la reconstrucción de genomas completos de patógenos antiguos es la de captura y enriquecimiento, la cual se discute en la siguiente sección.

8.1.2. Reconstrucción de genomas de patógenos a través de captura y enriquecimiento

La estrategia de captura y enriquecimiento se basa en aumentar el contenido de DNA endógeno a través de oligonucleótidos (referidos como *"baits"*) de DNA o RNA biotinilados, que pueden encontrarse inmovilizados en una fase sólida o en solución. La unión complementaria de los baits y el DNA de interés, es capturada por perlas magnéticas de estreptavidina, que a través de lavados se descarta el DNA inespecífico (ver capítulo 1).

La reconstrucción parcial o completa del genoma de patógenos antiguos ha permitido plantear inferencias evolutivas y sobre los mecanismos involucrados en su patogenicidad. Esto ha permitido reconstruir el genoma de un amplio número de patógenos, como *Yersinia pestis*, (Schuenemann *et al.*, 2011) *Salmonella enterica (Vågene et al., 2018)*, *Mycobacterium leprae (Sabin et al., 2020)* e incluso virus, como Parvovirus Humano B19, Hepatitis B (Guzmán-Solís *et al.*, 2020), entre otros.

En los capítulos 2 y 3, empleamos captura y enriquecimiento en solución, utilizando baits de RNA. De acuerdo al estudio de Furtwängler et al. (2020), este tipo de estrategia es más eficiente

en la captura de aDNA de patógenos comparado con el uso de baits de DNA y captura por medio de microarreglos. Con respecto al artículo de investigación (capítulo 5), empleamos baits de RNA diseñados para enriquecer genes relevantes en mecanismos de patogenicidad y procesos esenciales de *T. forsythia*. Esto nos permitió capturar de manera parcial once genomas del periodo prehispánico y colonial de México. Con respecto al capítulo 6, logramos reconstruir un genoma de 10X de *Salmonella enterica* Paratyphi C, a través de un protocolo de captura casero en el que sintetizamos los baits a partir del DNA moderno de *Salmonella enterica* Typhimurium.

8.3. Composición microbiana diferencial entre diente y cálculo dental

La caracterización microbiana del diente y cálculo dental de individuos del periodo prehispánico y colonial de México, permitió evidenciar un agrupamiento de acuerdo con el sustrato. A través del análisis NMDS (Escalamiento multidimensional no métrico), se observó una diferenciación de las muestras de diente y las de cálculo dental. Estudios previos han sugerido que la composición microbiana en el cálculo dental es uniforme (Mann *et al.*, 2018), y en la dentina es variable. No obstante, dos muestras de diente (TO-2417J y CO-09) se agruparon con las muestras de cálculo dental (Figura suplementaria 1 en capítulo 2). Se observó que *Actinomyces sp. oral taxon 894, Olsenella and sp. oral taxon 807, Anaerolineaceae bacterium oral taxon 439, Streptococcus gordonii* y *T. forsythia* eran las especies responsables de desviar el agrupamiento hacia el clúster de las muestras de cálculo dental. Esto puede explicarse con base a lo que que postula la hipótesis de la “placa ecológica”, la cual sugiere que determinados patógenos orales, como *T. forsythia*, comúnmente identificados en cálculo dental, se pueden co-localizar en otro microambiente de la cavidad oral, como el diente, dependiendo de su rol durante un proceso infeccioso (enfermedad periodontal). La evaluación inequívoca de esta hipótesis requiere de un número mayor de muestras de cálculo dental y diente.

Por otra parte, en los individuos provenientes de la costa norte de la Patagonia Argentina, analizamos muestras de hueso y diente. Sin embargo, de acuerdo con el perfil taxonómico obtenido (Figura 2 en el capítulo 4), no fue posible observar un agrupamiento diferencial de su composición microbiana acorde con el tipo de muestra analizada.

8.4. Asociación entre estilo de vida en poblaciones humanas y la presencia de patógenos antiguos

Dos dientes de individuos del periodo prehispánico, TO-2417Q y CA-13, provenientes de Querétaro y Guanajuato, respectivamente, presentaron una alta abundancia de la bacteria *T. forsythia* (capítulo 2). De acuerdo con el contexto arqueológico, ambos individuos probablemente pertenecían a un alto rango jerárquico. En el entierro del individuo TO-2417Q, proveniente de Toluquilla, Querétaro, se encontró un espejo de hematita y ofrendas, lo que es un indicador de que este individuo tenía un alto rango social en comparación con los demás individuos del mismo sitio. Por otra parte, el entierro del individuo CA-13 proveniente de Cañada de la Virgen, Guanajuato, se encontraron cuentas de caracol, un pectoral de concha y cerámica.

Se ha sugerido que una posición jerárquica alta dentro de la sociedad prehispánica, puede estar asociada con el acceso a cierto tipo de alimentación (Morfín, 2012), lo cual pudo favorecer la presencia de *T. forsythia.* El consumo excesivo de carbohidratos fermentables, resulta en la formación de *biofilms*, que puede promover la proliferación de dicha bacteria en la cavidad oral (Naginyte *et al.*, 2019). El análisis de cálculo dental de estos individuos permitiría identificar algunos de los alimentos que consumieron durante su vida y así poder confirmar esta asociación. Asimismo, el análisis de aDNA de diente y cálculo dental de individuos con una descripción de su rango jerárquico por medio del contexto arqueológico, contribuirá a esclarecer la posible correlación entre el rango jerárquico y la presencia de *T. forsythia.*

Otra observación de interés es la identificación de *Erysipelothrix rhusiopathiae,* en cazadores-recolectores de la costa norte Patagonia Argentina (6000-300 BP) (capítulo 4). Esta bacteria causa lesiones localizadas en la piel llamadas erisipeloide, que pueden progresar a una septicemia y endocarditis. El desarrollo de esta enfermedad se presenta con mayor frecuencia en individuos con una manipulación de carne cruda y productos de origen marítimo (Forde *et al.*, 2016). De acuerdo con el análisis de isótopos estables (13C y 15N), los cazadores-recolectores tenían una dieta terrestre-marina, incluyendo carne de guanaco, plantas, moluscos y pinipedos (Otero y Novellino, 2011). Por lo tanto la presencia de *E. rhusiopathiae* podría relacionarse con este estilo de vida.

9. PERSPECTIVAS

La paleogenómica ha contribuido al entendimiento de la diversidad y estructura genética, y de los procesos selectivos en humanos, plantas, animales y patógenos antiguos. Por lo que consideramos que el análisis realizado en esta tesis, representa un primer acercamiento en el entendimiento de los procesos selectivos en individuos antiguos de México.

El uso conjunto de tecnologías de siguiente generación y de métodos de captura-enriquecimiento en el análisis de aDNA acoplado a esta, han permitido reconstruir genomas completos de patógenos antiguos en poblaciones humanas del pasado. En el artículo de investigación (capítulo 5), reconstruimos el genoma de *T. forsythia* de forma parcial, por lo que un siguiente paso será el diseño de una estrategia de captura y enriquecimiento que permita la reconstrucción de su genoma completo de *Tannerella forsythia* en individuos del periodo prehispánico y colonial de México. Ello permitirá realizar un análisis más fino sobre la presencia y ausencia de genes y, por lo tanto, inferir los posibles mecanismos de adaptación al hospedero antes y después de la conquista española en México. Asimismo, la evaluación de genes asociados a mecanismos importantes de virulencia del genoma de *S. enterica* Paratyphi C en un individuo Colonial, permitirá contrastar su evolución con cepas identificadas en México y Europa. Con respecto a los cazadores-recolectores de la Patagonia Argentina, una perspectiva incluye el diseño de una estrategia de captura y enriquecimiento para los patógenos presentes en mayor abundancia (*Erysipelothrix rhusiopathiae* y *Rothia dentocariosa)*, con el objetivo de analizar la evolución de estos patógenos durante 6000-300 BP y contribuir al entendimiento de la presencia de enfermedades infecciosa, integrando la evidencia arqueológica e histórica de los cazadores-recolectores de la costa Norte de la Patagonia Argentina.

Por otra parte, el análisis de la interacción hospedero-patógeno en los individuos analizados en esta tesis, a través del análisis de genes relacionados con el sistema inmune, permitirá evaluar el impacto de los patógenos sobre la composición genética de su hospedero.

Sin duda, el área de genómica de patógenos en México se encuentra en una fase de crecimiento. Los resultados generados en esta tesis muestran su potencial en el entendimiento del estado de salud de poblaciones antiguas de México.

10. CONCLUSIONES

El estudio del estado de salud en poblaciones antiguas requiere un enfoque multidisciplinario, considerando principalmente la integración de la evidencia histórica, documental, arqueológica y recientemente, el análisis de aDNA. Los avances de la paleogenómica han permitido generar inferencias sobre la presencia de enfermedades infecciosas en poblaciones humanas antiguas, a un nivel de resolución fino, dando lugar al área de genómica de patógenos. El crecimiento de esta área en México ha ocurrido a un ritmo pausado, sin embargo, los resultados mostrados en esta tesis han contribuido al entendimiento de la evolución de los agentes causantes de enfermedades infecciosas en poblaciones humanas. Esta disciplina ha presenciado un crecimiento acelerado en los últimos años, así como diversos retos descritos en el artículo de revisión. El desarrollo de nuevas tecnologías, tanto metodológicas como analitícas han permitido la identificación y reconstrucción del genoma diversos patógenos, como los estudiados en la presente tesis. Los resultados obtenidos como parte de esta tesis permiten el planteamiento de futuros estudios sobre inferencias evolutivas y de patogenicidad durante el periodo prehispánico y colonial de México.

11. LITERATURA CITADA

Adler, C. J., Haak, W., Donlon, D., & Cooper, A. (2011). Survival and recovery of DNA from ancient teeth and bones. *Journal of archaeological science*, *38*(5), 956–964.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Álvarez-Sandoval, B. A., Manzanilla, L. R., González-Ruiz, M., Malgosa, A., & Montiel, R. (2015). Genetic Evidence Supports the Multiethnic Character of Teopancazco, a Neighborhood Center of Teotihuacan, Mexico (AD 200-600). *PloS One*, *10*(7), e0132371.

Antoine, D. (2008). 5 The Archaeology of “Plague”. *Medical history*, *52*(S27), 101–114.

Armelagos, G. J., Goodman, A. H., & Jacobs, K. H. (1991). The origins of agriculture: Population growth during a period of declining health. *Population and environment*, *13*(1), 9–22.

Baracs, R. M. (2015). *El largo descubrimiento del Opera medicinalia de Francisco Bravo*. Fondo de Cultura Economica.

Bardill, J., Bader, A. C., Garrison, N. A., Bolnick, D. A., Raff, J. A., Walker, A., Malhi, R. S., & the Summer internship for INdigenous peoples in Genomics (SING) Consortium. (2018). Advancing the ethics of paleogenomics. *Science*, *360*(6387), 384–385.

Barker, B. C. W. (1977). *Paleopathological Diagnosis and Interpretation: Bone Diseases in Ancient Human Populations*. JSTOR.

Barquera, R., Lamnidis, T. C., Lankapalli, A. K., Kocher, A., Hernández-Zaragoza, D. I., Nelson, E. A., Zamora-Herrera, A. C., Ramallo, P., Bernal-Felipe, N., Immel, A., Bos, K., Acuña-Alonzo, V., Barbieri, C., Roberts, P., Herbig, A., Kühnert, D., Márquez-Morfín, L., & Krause, J. (2020). Origin and Health Status of First-Generation Africans from Early Colonial Mexico. *Current Biology: CB*. https://doi.org/10.1016/j.cub.2020.04.002

Beaumont, F. (1932). Pablo: Crónica de Michoacán. *Publicaciones del Archivo General de la Nación, México*.

Bianucci, R., Rahalison, L., Massa, E. R., Peluso, A., Ferroglio, E., & Signoli, M. (2008). Technical note: a rapid diagnostic test detects plague in ancient human remains: an example of the interaction between archeological and biological approaches (southeastern France, 16th-18th centuries). *American Journal of Physical Anthropology*, *136*(3), 361–367.

Bravo, F., & Guerra, F. (1570). *The Opera Medicinalia, 1570. Limited facsimile edition. With An Introduction to Opera Medicinalia*.

Bravo-Lopez, M., Villa-Islas, V., Arriaga, C. R., Villaseñor-Altamirano, A. B., Guzmán-Solís, A., Sandoval-Velasco, M., Wesp, J. K., Alcantara, K., López-Corral, A., Gómez-Valdés, J., Mejía, E., Herrera, A., Meraz-Moreno, A., de la Luz Moreno-Cabrera, M., Moreno-Estrada, A., Nieves-Colón, M. A., Olvera, J., Pérez-Pérez, J., Iversen, K. H., … Ávila-Arcos, M. C. (2020). Paleogenomic insights into the red complex bacteria Tannerella forsythia in Pre-Hispanic and Colonial individuals from Mexico. En *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 375, Número 1812, p. 20190580). https://doi.org/10.1098/rstb.2019.0580

Cagliani, R., & Sironi, M. (2013). Pathogen-driven selection in the human genome. *International Journal of Evolutionary Biology*, *2013*, 204240.

Campana, M. G., Robles García, N., Rühli, F. J., & Tuross, N. (2014). False positives complicate ancient pathogen identifications using high-throughput shotgun sequencing. *BMC research notes*, *7*(1), 111.

Carrera Stampa, M. (1969). Relaciones geográficas de Nueva España siglos XVI y XVIII. *Estudios de historia novohispana*, *2*(002). http://revistas.unam.mx/index.php/ehn/article/view/3212

Chan, C. H., Tuite, A. R., & Fisman, D. N. (2013). Historical epidemiology of the second cholera pandemic: relevance to present day disease dynamics. *PloS One*, *8*(8), e72498.

Chimalpáhin Cuauhtlehuanitzin, D. F. de S. A. M. (1998). Tercera relación. *Las ocho relaciones y el Memorial de Colhuacán*, *1*.

Cisneros, H. F. (2010). La salud en la antigüedad prehispánica: un acercamiento bioantropológico de las condiciones de vida del Ecatepec antiguo. *CIENCIA ergo-sum*, *17*(1), 37–45.

Civera Cerecedo, M. (2006). El enfoque paleoepidemiológico en la antropología física. *Anales de antropología*, *40*. http://revistas.unam.mx/index.php/antropologia/article/view/669

Cook, N. D., & Lovell, W. G. (2000). *Juicios secretos de Dios: epidemias y despoblación indígena en Hispanoamérica colonial*. Editorial Abya Yala.

Cooper, A., & Poinar, H. N. (2000). Ancient DNA: do it right or not at all [Review of *Ancient DNA: do it right or not at all*]. *Science*, *289*(5482), 1139.

Cossar, J. H. (1994). Influence of Travel and Disease: An Historical Perspective. *Journal of Travel Medicine*, *1*(1), 36–39.

Crespo, H. G. (2010). *Historia de Morelos: tierra, gente, tiempos del sur*.

Cruz, I. D. L., González-Oliver, A., Kemp, B. M., Román, J. A., Smith, D. G., & Torre-Blanco, A. (2008). Sex identification of children sacrificed to the ancient Aztec rain gods in Tlatelolco. *Current anthropology*, *49*(3), 519–526.

Cuenya, M. Á. (1996). Peste en una ciudad novohispana. El matlazahuatl de 1737 en la Puebla de los Ángeles. *Anuario de estudios americanos*, *53*(2), 51–70.

Darling, M. I., & Donoghue, H. D. (2014). Insights from paleomicrobiology into the indigenous peoples of pre-colonial America - A Review. *Memorias do Instituto Oswaldo Cruz*, *109*(2), 131–139.

Dávalos, E. (1964). La patología ósea prehispánica. *Actas y Memorias del V Congreso Internacional de Americanistas. México City, México: Instituto Nacional de Antropología e Historia*, *3*, 79–86.

Davies-Barrett, A. M., Antoine, D., & Roberts, C. A. (2019). Inflammatory periosteal reaction on ribs associated with lower respiratory tract disease: A method for recording prevalence from sites with differing preservation. *American journal of physical anthropology*, *168*(3), 530–542.

de Montellano, B. O. (1992). *Historia general de la medicina en México. Tomo I. México Antiguo*. JSTOR. https://www.jstor.org/stable/649314

Dobyns, H. F. (1993). Disease Transfer at Contact. *Annual review of anthropology*, *22*(1), 273–291.
Domínguez-Andrés, J., & Netea, M. G. (2019). Impact of Historic Migrations and Evolutionary Processes on Human Immunity. *Trends in Immunology*, *40*(12), 1105–1119.
Drancourt, M., Aboudharam, G., Signoli, M., Dutour, O., & Raoult, D. (1998). Detection of 400-year-old Yersinia pestis DNA in human dental pulp: An approach to the diagnosis of ancient septicemia. *Proceedings of the National Academy of Sciences*, *95*(21), 12637–12640.
Drancourt, M., & Raoult, D. (2005). Palaeomicrobiology: current issues and perspectives. *Nature Reviews. Microbiology*, *3*(1), 23–35.
Duggan, A. T., Perdomo, M. F., Piombino-Mascali, D., Marciniak, S., Poinar, D., Emery, M. V., Buchmann, J. P., Duchêne, S., Jankauskas, R., Humphreys, M., Golding, G. B., Southon, J., Devault, A., Rouillard, J.-M., Sahl, J. W., Dutour, O., Hedman, K., Sajantila, A., Smith, G. L., … Poinar, H. N. (2016). 17 th Century Variola Virus Reveals the Recent History of Smallpox. *Current biology: CB*, *26*(24), 3407–3412.
Düx, A., Lequime, S., Patrono, L. V., Vrancken, B., Boral, S., Gogarten, J. F., Hilbig, A., Horst, D., Merkel, K., Prepoint, B., Santibanez, S., Schlotterbeck, J., Suchard, M. A., Ulrich, M., Widulin, N., Mankertz, A., Leendertz, F. H., Harper, K., Schnalke, T., … Calvignac-Spencer, S. (2020). Measles virus and rinderpest virus divergence dated to the sixth century BCE. *Science*, *368*(6497), 1367–1370.
Eleodoro Balboa, A. (2014). El Archivo Histórico del Archivo General de la Nación: cambiando de rostro. En *Revista del Archivo General de la Nación* (Vol. 29, Número 1, pp. 471–513). https://doi.org/10.37840/ragn.v29i1.69
Estrada, A., McAuliffe, D., Eritta, E., & Brumbaugh, R. (1970). PHYSICAL ANTHROPOLOGY OF SKELETAL REMAINS FROM TOMBS 25, 26, AND 27 FROM YAGUL, OAXACA, MEXICO. *AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY*, *33*, 129 – +.
Fellows Yates, J. A., Valtueña, A. A., Vågene, A. J., Cribdon, B., Velsko, I. M., Borry, M., Bravo-López, M. J., Fernandez-Guerra, A., Green, E. J., Ramachandran, S. L., Heintzman, P. D., Spyrou, M. A., Hübner, A., Gancz, A. S., Hider, J., Allshouse, A. F., & Warinner, C. (2020). *Community-curated and standardised metadata of published ancient metagenomic samples with AncientMetagenomeDir* (p. 2020.09.02.279570). https://doi.org/10.1101/2020.09.02.279570
Forde, T., Biek, R., Zadoks, R., Workentine, M. L., De Buck, J., Kutz, S., Opriessnig, T., Trewby, H., van der Meer, F., & Orsel, K. (2016). Genomic analysis of the multi-host pathogen Erysipelothrix rhusiopathiae reveals extensive recombination as well as the existence of three generalist clades with wide geographic distribution. En *BMC Genomics* (Vol. 17, Número 1). https://doi.org/10.1186/s12864-016-2643-0
Fotakis, A. K., Denham, S. D., Mackie, M., Orbegozo, M. I., Mylopotamitaki, D., Gopalakrishnan, S., Sicheritz-Pontén, T., Olsen, J. V., Cappellini, E., Zhang, G., Christophersen, A., Gilbert, M. T. P., & Vågene, Å. J. (2020). Multi-omic detection of Mycobacterium leprae in archaeological human dental calculus. En *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 375, Número 1812, p. 20190584). https://doi.org/10.1098/rstb.2019.0584
Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admetlla, A., Pattini, L., & Nielsen, R. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genetics*, *7*(11), e1002355.
Furuse, Y., Suzuki, A., & Oshitani, H. (2010). Origin of measles virus: divergence from rinderpest virus between the 11th and 12th centuries. *Virology journal*, *7*(1), 52.
Garfias, M. E. (2016). *Búsqueda e identificación de DNA de Treponema pallidum en restos óseos humanos prehispánicos de México, que muestran lesiones sugerentes de sífilis* [Maestría]. Universidad Nacional Autónoma de México.
Gignoux, C. R., Henn, B. M., & Mountain, J. L. (2011). Rapid, global demographic expansions after the origins of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6044–6049.

Gimmel, M. (2008). Hacia una reconsideración del Códice de la Cruz Badiano: nuevas propuestas para el estudio de la medicina indígena en el período colonial. *Colonial Latin American review*, *17*(2), 273–283.

Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E., & Orlando, L. (2011). mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* , *27*(15), 2153–2155.

González-Oliver, A., Márquez-Morfín, L., Jiménez, J. C., & Torre-Blanco, A. (2001). Founding amerindian mitochondrial DNA lineages in ancient Maya from Xcaret, Quintana Roo. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *116*(3), 230–235.

Goodman, A. H. (1993). On the Interpretation of Health From Skeletal Remains. *Current anthropology*, *34*(3), 281–288.

Grauer, A. L. (2018). A century of paleopathology. *American journal of physical anthropology*, *165*(4), 904–914.

Guzmán-Solís, A. A., Blanco-Melo, D., Villa-Islas, V., Bravo-López, M. J., Sandoval-Velasco, M., Wesp, J. K., Gómez-Valdés, J. A., de la Luz Moreno-Cabrera, M., Meraz-Moreno, A., Solís-Pichardo, G., Schaaf, P., tenOever, B. R., & Ávila-Arcos, M. C. (2020). *Ancient viral genomes reveal introduction of HBV and B19V to Mexico during the transatlantic slave trade* (p. 2020.06.05.137083). https://doi.org/10.1101/2020.06.05.137083

Haldane, J. B. S. (1932). *The causes of evolution. New York and London*. Harper.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A., & Wilson, A. C. (1984). DNA sequences from the quagga, an extinct member of the horse family. *Nature*, *312*(5991), 282–284.

Hofreiter, M., Paijmans, J. L. A., Goodchild, H., Speller, C. F., Barlow, A., Fortes, G. G., Thomas, J. A., Ludwig, A., & Collins, M. J. (2015). The future of ancient DNA: Technical advances and conceptual shifts. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *37*(3), 284–293.

Hong, J. (1997). *Origin and spread of syphilis*. https://scholarworks.umt.edu/cgi/viewcontent.cgi?article=4166&context=etd

Ho, S. Y. W., & Duchêne, S. (2020). Dating the emergence of human pathogens [Review of *Dating the emergence of human pathogens*]. *Science*, *368*(6497), 1310–1311.

Huebler, R., Key, F. M., Warinner, C., Bos, K. I., Krause, J., & Herbig, A. (2019). HOPS: Automated detection and authentication of pathogen DNA in archaeological remains. *bioRxiv*. https://doi.org/10.1101/534198

Inegi. (2005). *Catálogo de Documentos históricos de las estadísticas en México (siglos XVI-XIX)*. Instituto Nacional de Estadística, Geografía e Informática México.

Inhorn, M. C., & Brown, P. J. (1990). The Anthropology of Infectious Disease. En *Annual Review of Anthropology* (Vol. 19, Número 1, pp. 89–117). https://doi.org/10.1146/annurev.an.19.100190.000513

Jensen, T. Z. T., Niemann, J., Iversen, K. H., Fotakis, A. K., Gopalakrishnan, S., Vågene, Å. J., Pedersen, M. W., Sinding, M.-H. S., Ellegaard, M. R., Allentoft, M. E., Lanigan, L. T., Taurozzi, A. J., Nielsen, S. H., Dee, M. W., Mortensen, M. N., Christensen, M. C., Sørensen, S. A., Collins, M. J., Gilbert, M. T. P., … Schroeder, H. (2019). A 5700 year-old human genome and oral microbiome from chewed birch pitch. *Nature Communications*, *10*(1), 5520.

Kay, G. L., Sergeant, M. J., Giuffra, V., Bandiera, P., Milanese, M., Bramanti, B., Bianucci, R., & Pallen, M. J. (2014). Recovery of a Medieval Brucella melitensis Genome Using Shotgun Metagenomics. *mBio*, *5*(4), e01337–14 – e01337–14.

Kistler, L., Ware, R., Smith, O., Collins, M., & Allaby, R. G. (2017). A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Research*, *45*(11), 6310–6320.

Klunk, J. (2018). *Genetic causes and consequences of ancient plague pandemics: pathogen evolution and human demographic reverberations*. https://macsphere.mcmaster.ca/handle/11375/23768

Koch, A., Brierley, C., Maslin, M. M., & Lewis, S. L. (2019). Earth system impacts of the European arrival and Great Dying in the Americas after 1492. *Quaternary science reviews*, *207*, 13–36.

Lantes-Suárez, O., & Prieto-Martínez, M. P. (2017). *Técnicas de imagen de rayos X y arqueología. Estado de la cuestión y potencialidad de la técnica*. Antrope.

Lan, T., & Lindqvist, C. (2018). *Paleogenomics: Genome-Scale Analysis of Ancient DNA and Population and Evolutionary Genomic Inferences*. Springer International Publishing.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, *362*(6422), 709–715.

Lineamientos generales para el manejo y resguardo de restos humanos. (18 de Octubre del 2019). Secretaría de Cultura. Instituto Nacional de Antropología e Historia. https://www.normateca.inah.gob.mx/pdf/01574791311.PDF

Machado, A. S., Oliveira, D. F., Gama Filho, H. S., Latini, R., Bellido, A. V. B., Assis, J. T., Anjos, M. J., & Lopes, R. T. (2017). Archeological ceramic artifacts characterization through computed microtomography and X-ray fluorescence. *X-Ray Spectrometry*, *46*(5), 427–434.

Magaña, M. A. (2013). *Epidemias y rutas de propagación en la Nueva España y México (siglo XVIII-XIX)*. Universidad Autónoma de Baja California, Instituto de Investigaciones ….

Maixner, F., Krause-Kyora, B., Turaev, D., Herbig, A., Hoopmann, M. R., Hallows, J. L., Kusebauch, U., Vigl, E. E., Malfertheiner, P., Megraud, F., OSullivan, N., Cipollini, G., Coia, V., Samadelli, M., Engstrand, L., Linz, B., Moritz, R. L., Grimm, R., Krause, J., … Zink, A. (2016). The 5300-year-old Helicobacter pylori genome of the Iceman. *Science*, *351*(6269), 162–165.

Malvido, E. (1992). ?` El arca de Noe o la caja de Pandora? Suma y recopilación de pandemias, epidemias y endemias en Nueva España, 1519-1810. *Nueva orientacion de la pesca. Serie de informes sobre temas del desarrollo*, 45–87.

Malvido, E., & Viesca, C. (1985). La epidemia de cocoliztli de 1576. *Historias México, DF*, *11*, 27–33.

Mandujano, A., Camarillo, L., & Mandujano, M. (2003). Historia de las epidemias en el México antiguo. Algunos aspectos biológicos y sociales. *Casa del Tiempo*, *5*, 9–21.

Mann, A. E., Sabin, S., Ziesemer, K., Vågene, Å. J., Schroeder, H., Ozga, A. T., Sankaranarayanan, K., Hofman, C. A., Fellows Yates, J. A., Salazar-García, D. C., Frohlich, B., Aldenderfer, M., Hoogland, M., Read, C., Milner, G. R., Stone, A. C., Lewis, C. M., Krause, J., Hofman, C., … Warinner, C. (2018). Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. *Scientific reports*, *8*(1). https://doi.org/10.1038/s41598-018-28091-9

Marciniak, S., & Perry, G. H. (2017). Harnessing ancient genomes to study the history of human adaptation. *Nature reviews. Genetics*, *18*(11), 659–674.

Márquez Morfín, L. (2015). La sífilis y su carácter endémico en la Ciudad de México. *Historia mexicana*, *64*(3), 1099–1161.

Márquez Morfín, L., & Meza Manzanilla, M. (2015). Sífilis en la Ciudad de México: análisis osteopatológico. *Cuicuilco*, *22*(63), 89–126.

Márquez Morfín, L., & Sosa Márquez, M. V. (2016). Mortalidad de niños y sífilis congénita en la Ciudad de México en 1915. *Estudios demograficos y urbanos*, *31*(1), 177–206.

Matthijnssens, J., Ciarlet, M., Heiman, E., & Arijs, I. (2008). Full genome-based classification of rotaviruses reveals a common origin between human Wa-Like and porcine rotavirus strains and human DS-1-like and bovine …. *Journal of*. https://jvi.asm.org/content/82/7/3204.short

Meyer, C., Jung, C., Kohl, T., Poenicke, A., Poppe, A., & Alt, K. W. (2002). Syphilis 2001--a palaeopathological reappraisal. *Homo: Internationale Zeitschrift Fur Die Vergleichende Forschung Am Menschen*, *53*(1), 39–58.

Meza Manzanilla, M. (2013). El declive de las condiciones de vida, salud y nutrición de los pobladores de Tetetzontilco en el siglo XVI. *Estudios de Antropología Biológica*, *11*(2). http://revistas.unam.mx/index.php/eab/article/view/43218

Morfín, L. M. (2012). Vida urbana y salud en la Mesoamérica prehispánica. *Revista de la Facultad de Odontologia Universidad de Antioquia*, *26*(43), 215–238.

Motolinía, B., & Toribio, F. (1971). Memoriales o Libro de las cosas de la Nueva España y de los naturales de ella, éd. *Edmundo O'Gorman, UNAM, IIH, México*.

Mullis, K. B., & Faloona, F. A. (1989). 9 - Specific Synthesis of DNA in Vitro via a Polymerase-Catalyzed Chain Reaction. En R. Wu, L. Grossman, & K. Moldave (Eds.), *Recombinant DNA Methodology* (pp. 189–204). Academic Press.

Naginyte, M., Do, T., Meade, J., Devine, D. A., & Marsh, P. D. (2019). Enrichment of periodontal pathogens from the biofilms of healthy adults. *Scientific Reports*, *9*(1), 5491.

Neukamm, J., Pfrengle, S., Molak, M., Seitz, A., Francken, M., Eppenberger, P., Avanzi, C., Reiter, E., Urban, C., Welte, B., Stockhammer, P. W., Teßmann, B., Herbig, A., Harvati, K., Nieselt, K., Krause, J., & Schuenemann, V. J. (2020). 2000-year-old pathogen genomes reconstructed from metagenomic analysis of Egyptian mummified individuals. *BMC Biology*, *18*(1), 108.

Nguyen-Hieu, T., Aboudharam, G., & Drancourt, M. (2011). *Mini Review: Dental Pulp as a Source for Paleomicrobiology*. *5*(1), 7.

Nicholson, H. B. (1966). Die Relationen Chimalpahin's zur Geschichte Mexico's; Teil 2: Das Jahrhundert nach der Conquista (1522-1615). Aztekischer Text herausgegeben von Günter zimmermann. Universtat Hamburg, Abhandlungen aus dem Gebiet der Auslandskunde, Band 69-Reihe B (Völkerkunde, Kulturgeschichte und Sprachen), Band 39, Hamburg, 1965. 207 pp. 48 DM. *American antiquity*, *31*(6), 886–887.

Ochoa-Lugo, M. I., Muñoz, M. de L., Pérez-Ramírez, G., Beaty, K. G., López-Armenta, M., Cervini-Silva, J., Moreno-Galeana, M., Meza, A. M., Ramos, E., Crawford, M. H., & Romano-Pacheco, A. (2016). Genetic Affiliation of Pre-Hispanic and Contemporary Mayas Through Maternal Linage. *Human Biology*, *88*(2), 136–167.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733–D745.

Ortner D. J. (2003). *Identification of pathological conditions in human skeletal remains* (2a ed.). Academic Press.

Otero, J. G., & Novellino, P. (2011). Diet, nutritional status and oral health in hunter-gatherers from the central-northern coast of Patagonia and the Chubut river lower valley, Argentina. *International journal of osteoarchaeology*, *21*(6), 643–659.

Paabo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences*, *86*(6), 1939–1943.

Pérez-Zamorano, B., Vallebueno-Estrada, M., González, J. M., Cook, A. G., Montiel, R., Vielle-Calzada, J.-P., & Delaye, L. (2017). Organellar Genomes from a ~5,000-Year-Old Archaeological Maize Sample Are Closely Related to NB Genotype. En *Genome Biology and Evolution* (Vol. 9, Número 4, pp. 904–915). https://doi.org/10.1093/gbe/evx048

Pinhasi, R., & Mays, S. (2008). *Advances in Human Palaeopathology*. John Wiley & Sons.

Prendergast, M. E., & Sawchuk, E. (2018). Boots on the ground in Africa's ancient DNA "revolution": archaeological perspectives on ethics and best practices. *Antiquity*, *92*(363), 803–815.

Raoult, D., & Drancourt, M. (2008). Molecular Detection of Past Pathogens. En D. Raoult & M. Drancourt (Eds.), *Paleomicrobiology: Past Human Infections* (pp. 55–68). Springer Berlin Heidelberg.

Robles García, N. M., & Molina Villegas, G. (1998). Exploración de una tumba prehispánica en el sitio Llaadzie en la comunidad de Mitla, Oaxaca. *Cuadernos del Sur*, *12*, 21–52.

Romano, A. (1956). Los restos óseos humanos de la cueva de La Candelaria, Coahuila. *MAThesis in Physical Anthropology. Mexico City: Escuela Nacional de Antropología e Historia/Instituto Nacional de Antropología e Historia*.

Roys, R. L. (1933). *THE BOOK OF CHILAM BALAM OF CHUMAYEL* (Carnegie Institution of Washington (Ed.); p. 142).

Sabin, S., Herbig, A., Vågene, Å. J., Ahlström, T., Bozovic, G., Arcini, C., Kühnert, D., & Bos, K. I. (2020). A seventeenth-century Mycobacterium tuberculosis genome supports a Neolithic emergence of the Mycobacterium tuberculosis complex. *Genome Biology*, *21*(1), 201.

Salas Contreras, C. (2007). Prácticas funerarias en la ex-iglesia de la Encarnación, "Antigua Biblioteca Iberoamericana". *Arqueología mexicana*, *36*, 116–134.

Salo, W. L., Aufderheide, A. C., Buikstra, J., & Holcomb, T. A. (1994). Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy. *Proceedings of the National Academy of Sciences*, *91*(6), 2091–2094.

Samudio, A. J. A., Sobrino, B. Z. G., Sandoval, B. A. A., Montiel, R., Sánchez, C. S., & Peñaloza, A. M. (2017). Genetic history of the classic period of Teotihuacan's burials in Central Mexico. *Revista argentina de antropología biológica*, *19*(1), 14–14.

Saunders-Hastings, P. R., & Krewski, D. (2016). Reviewing the History of Pandemic Influenza: Understanding Patterns of Emergence and Transmission. *Pathogens*, *5*(4). https://doi.org/10.3390/pathogens5040066

Schuenemann, V. J., Bos, K., DeWitte, S., Schmedes, S., Jamieson, J., Mittnik, A., Forrest, S., Coombes, B. K., Wood, J. W., Earn, D. J. D., White, W., Krause, J., & Poinar, H. N. (2011). Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. *Proceedings of the National Academy of Sciences*, *108*(38), E746–E752.

Schuenemann, V. J., Kumar Lankapalli, A., Barquera, R., Nelson, E. A., Iraíz Hernández, D., Acuña Alonzo, V., Bos, K. I., Márquez Morfín, L., Herbig, A., & Krause, J. (2018). Historic Treponema pallidum genomes from Colonial Mexico retrieved from archaeological remains. *PLoS neglected tropical diseases*, *12*(6), e0006447.

Shapiro, B., & Hofreiter, M. (2014). A Paleogenomic Perspective on Evolution and Gene Function: New Insights from Ancient DNA. *Science*, *343*(6169), 1236573–1236573.

Siddle, K. J., & Quintana-Murci, L. (2014). The Red Queen's long race: human adaptation to pathogen pressure. *Current Opinion in Genetics & Development*, *29*, 31–38.

Siek, T. (2013). The Osteological Paradox and Issues of Interpretation in Paleopathology. *Vis-à-vis: Explorations in Anthropology*, *12*(1). https://vav.library.utoronto.ca/index.php/vav/article/view/19464

Spigelman, M., & Lemma, E. (1993). The use of the polymerase chain reaction (PCR) to detect Mycobacterium tuberculosis in ancient skeletons. *International Journal of Osteoarchaeology*, *3*(2), 137–143.

Spores, R. A., Robles García, N., Diego Luna, L., Roldán López, L. L., & Ruiz Ríos, I. G. (2007). Avances de investigación de los entierros humanos del sitio Pueblo Viejo de Teposcolula y su contexto arqueológico. *Estudios de antropología biológica*, *13*(1). http://revistas.unam.mx/index.php/eab/article/view/21608

Spyrou, M. A., Bos, K. I., Herbig, A., & Krause, J. (2019). Ancient pathogen genomics as an emerging tool for infectious disease research. *Nature Reviews. Genetics*, *20*(6), 323–340.

Stewart, T. D. (1956). Skeletal remains from Xochicalco, Morelos. *Estudios antropológicos publicados en homenaje a Manuel Gamio*, 131–156.

Stone, A. C. (2020). Getting sick in the Neolithic [Review of *Getting sick in the Neolithic*]. *Nature Ecology & Evolution*, *4*(3), 286–287.

Suzuki, Y., & Nei, M. (2002). Origin and evolution of influenza virus hemagglutinin genes. *Molecular Biology and Evolution*, *19*(4), 501–509.

Telleriano-Remensis, C. (1995). Codex telleriano-remensis. *Ritual, Divination and*.

Treviño, C. V. (2005). Las enfermedades en Mesoamérica. *Arqueología mexicana*, *13*(74), 38–41.

Tuniz, C., & Zanini, F. (2018). Microcomputerized tomography (MicroCT) in archaeology. *Encyclopedia of Global Archaeology. Springer International Publishing, Cham*, 1–7.

Vågene, Å. J., Herbig, A., Campana, M. G., Robles García, N. M., Warinner, C., Sabin, S., Spyrou, M. A., Andrades Valtueña, A., Huson, D., Tuross, N., Bos, K. I., & Krause, J. (2018). Salmonella

enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology & Evolution*, *2*(3), 520–528.
Vallebueno-Estrada, M., Rodríguez-Arévalo, I., Rougon-Cardoso, A., Martínez González, J., García Cook, A., Montiel, R., & Vielle-Calzada, J.-P. (2016). The earliest maize from San Marcos Tehuacán is a partial domesticate with genomic evidence of inbreeding. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(49), 14151–14156.
Vargas Sanders, R., & Alcalá, R. S. (1990). Material genético de restos óseos humanos. *en: Estudios de Antropologia Biologica*, 219–242.
Velázquez C. (2006). *La enfermedad periodontal en diversas poblaciones del México prehispánico*. 225–230.
Velázquez, P. F., & León Portilla, M. (2019). *Códice Chimalpopoca. Anales de Cuauhtitlán y leyenda de los Soles*. Universidad Nacional Autónoma de México.
Velsko, I. M., Frantz, L. A. F., Herbig, A., Larson, G., & Warinner, C. (2018). *Selection of appropriate metagenome taxonomic classifiers for ancient microbiome research*. https://doi.org/10.1101/260042
Viesca, C. (1992). Hambruna y epidemia en Anáhuac (1450-1454) en la época de Moctezuma Ilhuicamina. *Florescano, E. y Malvido, E. (Comp.), Ensayos sobre la historia de las epidemias en México*, *1*.
Viesca, C. (1998). *Medicina prehispánica de México*. Panorama.
Wagner, J. K., Colwell, C., Claw, K. G., Stone, A. C., Bolnick, D. A., Hawks, J., Brothers, K. B., & Garrison, N. A. (2020). Fostering Responsible Research on Ancient DNA. *American Journal of Human Genetics*, *107*(2), 183–195.
Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R. Y., Fiddyment, S., Speller, C., Hendy, J., Charlton, S., Luder, H. U., Salazar-García, D. C., Eppler, E., Seiler, R., Hansen, L. H., Castruita, J. A. S., … Cappellini, E. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics*, *46*(4), 336–344.
Willerslev, E., & Cooper, A. (2005). Ancient DNA. *Proceedings. Biological Sciences / The Royal Society*, *272*(1558), 3–16.
Wolfe, N. D., Dunavan, C. P., & Diamond, J. (2007). Origins of major human infectious diseases. *Nature*, *447*(7142), 279–283.
Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biology*, *15*(3), R46.
Wood, J. W., Milner, G. R., Harpending, H. C., Weiss, K. M., Cohen, M. N., Eisenberg, L. E., Hutchinson, D. L., Jankauskas, R., Cesnys, G., Česnys, G., Katzenberg, M. A., Lukacs, J. R., McGrath, J. W., Roth, E. A., Ubelaker, D. H., & Wilkinson, R. G. (1992). The Osteological Paradox: Problems of Inferring Prehistoric Health from Skeletal Samples [and Comments and Reply]. *Current anthropology*, *33*(4), 343–370.
Woolhouse, M. E. J. (2002). Population biology of emerging and re-emerging pathogens. *Trends in Microbiology*, *10*(10 Suppl), S3–S7.
Zhou, Z., Lundstrøm, I., Tran-Dien, A., Duchêne, S., Alikhan, N.-F., Sergeant, M. J., Langridge, G., Fotakis, A. K., Nair, S., Stenøien, H. K., Hamre, S. S., Casjens, S., Christophersen, A., Quince, C., Thomson, N. R., Weill, F.-X., Ho, S. Y. W., Gilbert, M. T. P., & Achtman, M. (2018). Pan-genome Analysis of Ancient and Modern Salmonella enterica Demonstrates Genomic Stability of the Invasive Para C Lineage for Millennia. *Current biology: CB*, *28*(15), 2420–2428.e10.

12. ANEXOS

Durante la realización de mis estudios de Doctorado participé en dos artículos de investigación:

SCIENTIFIC DATA

Check for updates

OPEN

DATA DESCRIPTOR

Community-curated and standardised metadata of published ancient metagenomic samples with AncientMetagenomeDir

James A. Fellows Yates[1,2]✉, Aida Andrades Valtueña[1], Åshild J. Vågene[3], Becky Cribdon[4], Irina M. Velsko[1], Maxime Borry[1], Miriam J. Bravo-Lopez[5], Antonio Fernandez-Guerra[6,7], Eleanor J. Green[8,9], Shreya L. Ramachandran[10], Peter D. Heintzman[11], Maria A. Spyrou[1], Alexander Hübner[1,12], Abigail S. Gancz[13], Jessica Hider[14,15], Aurora F. Allshouse[16,17], Valentina Zaro[18] & Christina Warinner[1,16]✉

Ancient DNA and RNA are valuable data sources for a wide range of disciplines. Within the field of ancient metagenomics, the number of published genetic datasets has risen dramatically in recent years, and tracking this data for reuse is particularly important for large-scale ecological and evolutionary studies of individual taxa and communities of both microbes and eukaryotes. AncientMetagenomeDir (archived at https://doi.org/10.5281/zenodo.3980833) is a collection of annotated metagenomic sample lists derived from published studies that provide basic, standardised metadata and accession numbers to allow rapid data retrieval from online repositories. These tables are community-curated and span multiple sub-disciplines to ensure adequate breadth and consensus in metadata definitions, as well as longevity of the database. Internal guidelines and automated checks facilitate compatibility with established sequence-read archives and term-ontologies, and ensure consistency and interoperability for future meta-analyses. This collection will also assist in standardising metadata reporting for future ancient metagenomic studies.

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, 07745, Jena, Germany. [2]Institut für Vor- und Frühgeschichtliche Archäologie und Provinzialrömische Archäologie, Ludwig-Maximilians-Universität München, München, 80539, Germany. [3]Section for Evolutionary Genomics, GLOBE Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 1350, Denmark. [4]School of Life Sciences, University of Warwick, Coventry, CV4 7AL, United Kingdom. [5]International Laboratory for Human Genome Research, National Autonomous University of Mexico, Queretaro, 76230, Mexico. [6]Section for GeoGenetics, GLOBE Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, 1350, Denmark. [7]Microbial Genomics and Bioinformatics Research Group, Max Planck Institute for Marine Microbiology, Bremen, 28359, Germany. [8]BioArCh, Department of Archaeology, University of York, York, YO10 5DD, United Kingdom. [9]Department of Earth Sciences, Natural History Museum, London, SW7 5BD, United Kingdom. [10]Human Genetics, University of Chicago, Chicago, IL, 60637, USA. [11]The Arctic University Museum of Norway, UiT The Arctic University of Norway, Tromsø, 9037, Norway. [12]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, 04103, Germany. [13]Department of Anthropology, Pennsylvania State University, Pennsylvania, PA, 16802, USA. [14]Department of Anthropology, McMaster University, Hamilton, L8S4L9, Canada. [15]McMaster Ancient DNA Centre, McMaster University, Hamilton, L8S4L10, Canada. [16]Department of Anthropology, Harvard University, Cambridge, MA, 02138, USA. [17]Max Planck-Harvard Research Center for the Archaeoscience of the Ancient Mediterranean, Cambridge, MA, 02138, USA. [18]Department of Biology, Università degli Studi di Firenze, Florence, 50122, Italy. ✉e-mail: fellows@shh.mpg.de; warinner@shh.mpg.de

Background & Summary

A crucial, but sometimes overlooked, component of scientific reproducibility is the efficient retrieval of sample metadata. While the field of ancient DNA (aDNA) has been celebrated for its commitment to making sequencing data available through public archives[1], this data is not necessarily 'findable' (as defined in the FAIR principles[2]) - making the retrieval of relevant metadata time-consuming and complex. Metagenomic studies typically require large sample sizes, which are integrated with previously published datasets for comparative analyses. However, the current absence of standards in basic metadata reporting within ancient metagenomics can make data retrieval tedious and laborious, leading to analysis bottlenecks.

Ancient metagenomics can be broadly defined as the study of the *total* genetic content of samples that have degraded over time[3]. Areas of study that fall under ancient metagenomics include studies of host-associated microbial communities (e.g., ancient microbiomes[4]), genome reconstruction and analysis of specific microbial taxa (e.g., ancient pathogens[5]), and environmental reconstructions using sedimentary aDNA (sedaDNA)0[6]. Endogenous genetic material obtained from ancient samples has undergone a variety of degradation processes that can cause the original genetic signal to be overwhelmed by modern contamination. Therefore, to detect, quantify, and authenticate the remaining 'true' aDNA large DNA sequencing efforts are required[7,8]. These studies have only become feasible since the development of massively parallel 'next-generation sequencing', which enables the generation of large amounts of genetic data that are mostly uploaded to and stored on large generalised archives such as the European Bioinformatic Institute's (EBI) European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena/) or the US National Center for Biotechnology Information (NCBI)'s Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra). However, as these are generalised databases used for many kinds of genetic studies, searching for and identifying ancient metagenomic samples can be difficult and time consuming, partly because of the absence of standardised metadata reporting for ancient metagenomic data. Consequently, researchers must resort to repeated extensive literature searches of heterogeneously reported and inconsistently formatted publications to locate ancient metagenomic datasets. Overcoming the difficulty of finding previously published samples is particularly pertinent to studies of aDNA, as palaeontological and archaeological samples are by their nature limited, and avoiding repeated or redundant sampling is of high priority[9–11].

To address these issues, we established AncientMetagenomeDir, a CC-BY 4.0 licensed community-curated collection of annotated sample lists that aims to guide researchers to all published ancient metagenomics-related samples with publicly available sequence data. AncientMetagenomeDir was conceived by members of a recently established international and open community of researchers working in ancient metagenomics (Standards, Precautions and Advances in Ancient Metagenomics, or 'SPAAM' - https://spaam-community.github.io), whose aim is to foster research collaboration and define standards in analysis and reporting within the field. The collection aims to be comprehensive but lightweight, consisting of tab-separated value (TSV) tables for different major sub-disciplines of ancient metagenomics. These tables contain essential, sample-specific information for aDNA studies, including: geographic coordinates, temporal data, sub-discipline specific critical information, and public archive accession codes that guide researchers to associated sequence data (see Methods). This simple format, together with comprehensive guides and documentation, encourages continuous contributions from the community and facilitates usage of the resource by researchers coming from non-computational backgrounds, something common in interdisciplinary fields such as archaeo- and palaeogenetics.

AncientMetagenomeDir is designed to track the development of ancient metagenomics through regular releases. As of release v20.09, this includes 87 studies published since 2011, representing 443 ancient host-associated metagenome samples, 269 ancient microbial genome level sequences, and 312 sediment samples (Fig. 1) spanning 49 countries (Fig. 2). We expect AncientMetagenomeDir to deliver three key benefits. First, it will contribute to the longevity of important cultural heritage by guiding future sampling strategies, thereby reducing the risk of repeated or over-sampling of the same samples or regions. Second, it can serve as a starting point for the development of software to allow rapid aggregation of actual data files and field-specific data processing. Third, it will assist in expanding meta-analyses (such as[12,13]) to a wider range of sample types and DNA sources in order to tackle broader palaeogenetic, ecological, and evolutionary questions. Finally, as a community-curated resource designed specifically for widespread participation, AncientMetagenomeDir will help the field to define common standards of metadata reporting (such as with MIxS checklists[14]), facilitating the creation of future databases that are consistent, and richer, in useful metadata.

Methods

Repository Structure. AncientMetagenomeDir[15] is a community-curated set of tables maintained on GitHub containing metadata from published ancient metagenomic studies (https://github.com/SPAAM-community/AncientMetagenomeDir). While most submissions are made by SPAAM members, anyone with a GitHub account is welcome to propose (termed here 'proposer') and/or add publications for inclusion (termed 'contributor'). Proposers and contributors can be (but do not have to be) authors of the original publication(s) proposed for inclusion. Submitted studies must be published in a peer-reviewed journal because the purpose of AncientMetagenomeDir is not to act as a quality filter and we do not currently make assessments based on data quality. The tables are formatted as tab-separated value (TSV) files in order to maximize accessibility for all researchers and to allow portability between different data analysis software.

Valid samples for inclusion currently fall under three sub-fields: (1) host-associated metagenomes (i.e., host-associated or skeletal material microbiomes), (2) host-associated single genomes (i.e., pathogen or commensal microbial genomes), and (3) environmental metagenomes (e.g., sedaDNA). In addition, a fourth category is currently planned: (4) anthropogenic metagenomes (e.g., dietary and microbial DNA within pottery crusts, or microbial DNA and handling debris on parchment). The definitions under which a sample is considered 'ancient' is adapted on a per sub-field basis. Generally, samples are required to have had reported evidence of hydrolytic damage at molecule termini, short fragment lengths, and contain fraction of non-endogenous content

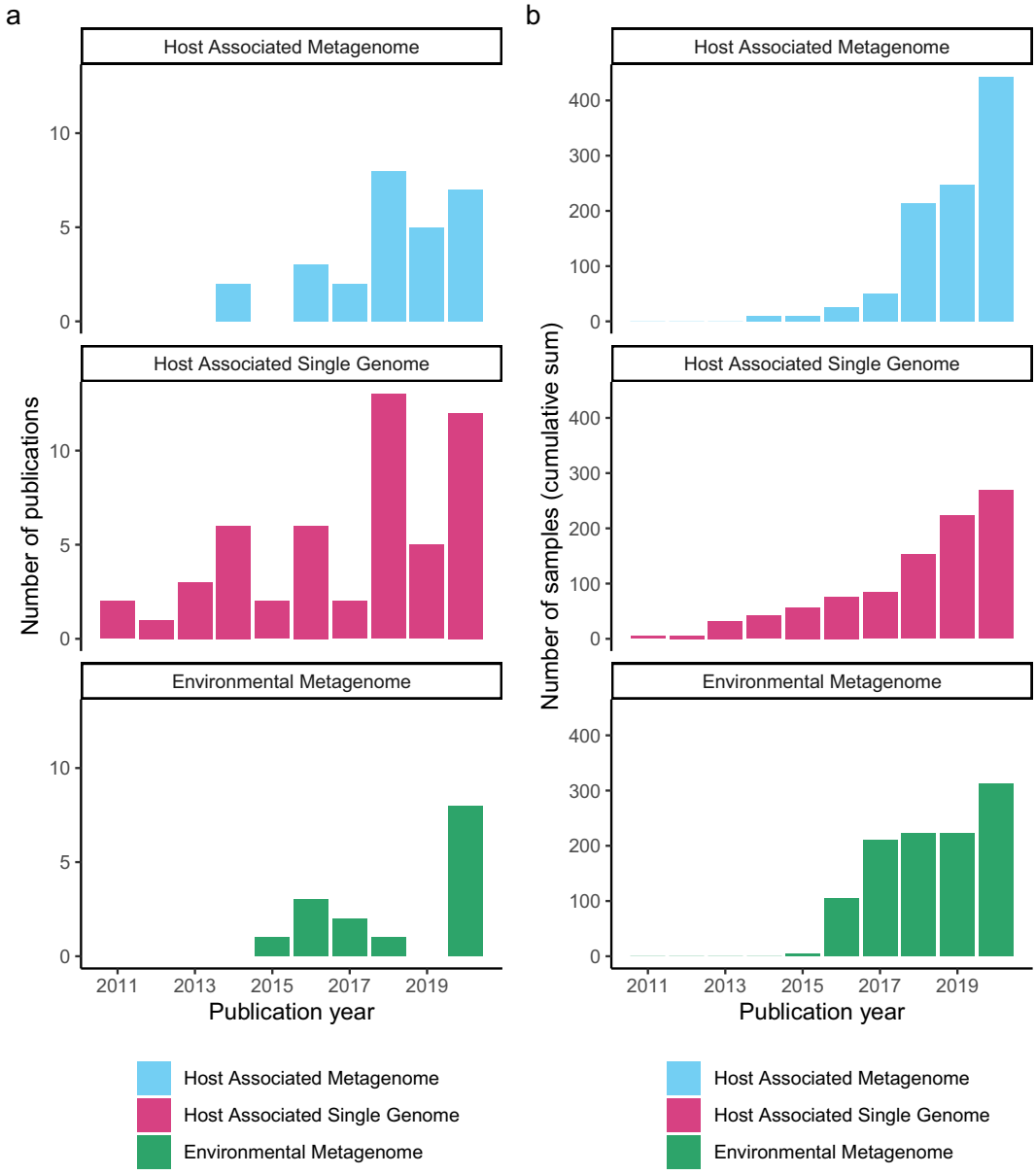


Fig. 1 Timelines depicting the development of the sub-disciplines of ancient metagenomics as recorded in AncientMetagenomeDir as per release v20.09. (**a**) Number of ancient metagenomic publications per year. (**b**) Cumulative sum of published samples with genetic sequencing data or sequences in publicly accessible archives.

(e.g. as summarised in[3]). However, for example, due to regular use in ancient pathogenomics studies, samples preserved in long-term medical collections from the last century that have limited degradation may also be included. In the first release of AncientMetagenomeDir, we have specified a minimum age of older than 1950 CE. Samples must have been sequenced using a shotgun metagenomic approach, or alternatively a whole organelle- or chromosome-level enrichment approach, and sequence data must be publicly available on an established or stable archive. INDSC-associated repositories such as the EBI's ENA or NCBI's SRA and Genbank databases are preferred, as they are the most accepted and commonly used archives for raw sequencing data. However, DOI-issuing long-term archives (such as Zenodo or Figshare), institutional repositories (such as institutional data services), or field-specific established repositories (e.g., TreeBASE) can also be accepted. Data on personal or lab websites are not accepted due to uncertain storage longevity. We currently do not include laboratory negative controls, as we consider these to be 'artefacts' of lab procedures and better addressed with experiment-level metadata. If required by a researcher, controls can be identified via sample-associated project accession codes.

Publications included in the current release of AncientMetagenomeDir were selected for inclusion based on direct contributions by authors of publications and also from literature reviews of each sub-field made by the SPAAM community. In this process, a proposer initially suggests a publication to be included via a GitHub 'Issue'. Publications may belong to multiple categories, and the corresponding issue is tagged with relevant category 'labels' to assist with faster evaluation and task distribution.

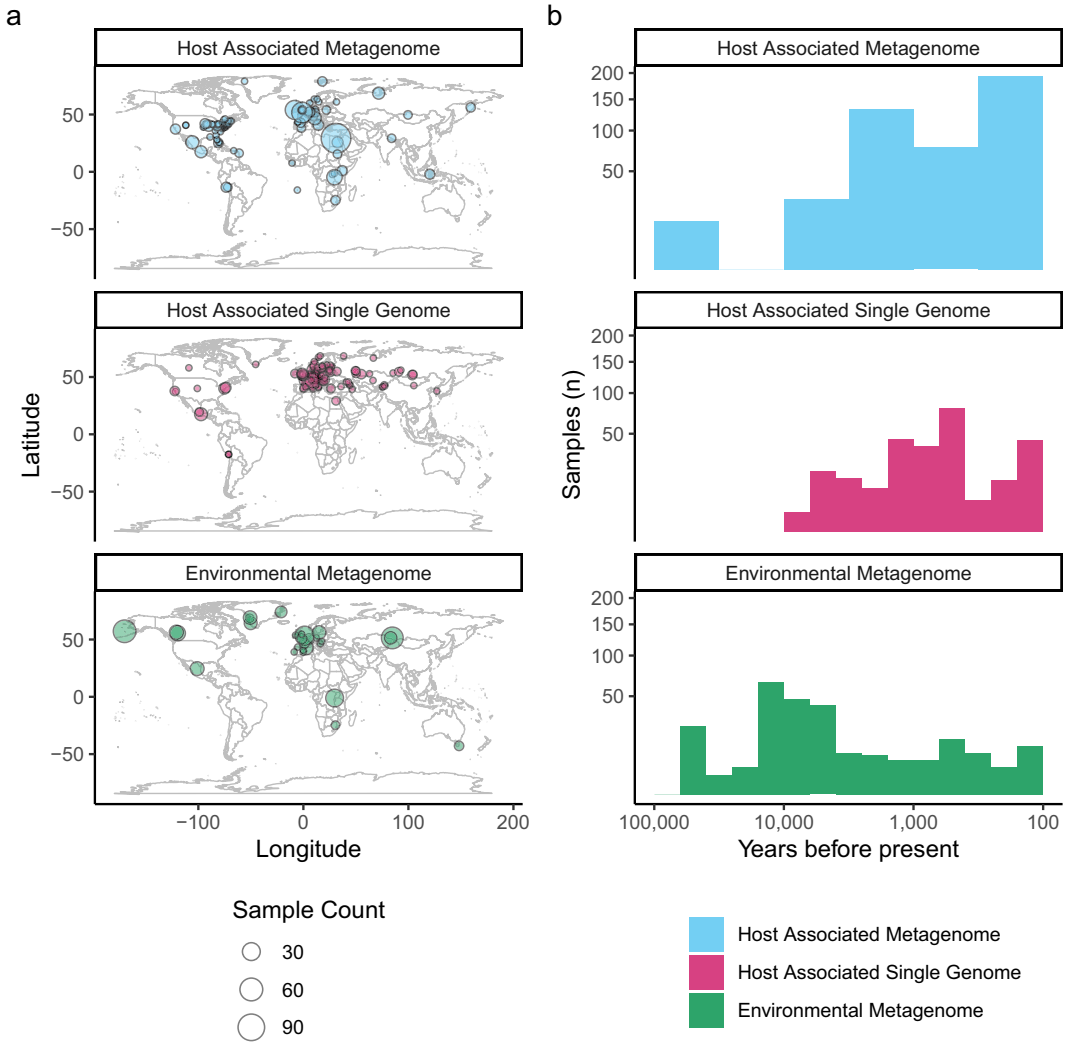


Fig. 2 Summary of temporal and spatial information of ancient metagenomic samples as recorded in AncientMetagenomeDir v20.09. (**a**) Maps depicting the geographic distribution of samples for each sub-discipline. (**b**) Histogram of sample ages for each sub-discipline. For visualisation purposes, plot axes are log-scaled, bins calculated using the 'Freedman–Diaconis' rule, and only samples dated to younger than 50,000 years are displayed.

Data acquisition. Members of the SPAAM community (termed 'curators') evaluate proposed publications for applicability under the criteria described above. Once approved, any member of the open SPAAM community can assign themselves to the corresponding Issue and will henceforth act as the contributor. A proposer from outside SPAAM who wishes to also be a contributor can be added to the SPAAM community by contacting a current member if desired. The contributor then creates a git branch from the main repository, manually extracts the relevant metadata from the given publication, and adds it to the assigned table (e.g., host associated metagenome, or environmental metagenome). Extensive documentation on submissions, including instructions on using GitHub, are available via tutorial documents and the associated repository wiki. Both are accessible via the main repository README under the 'Contributing' section. Furthermore, detailed documentation is also available to assist contributors and ensure correct entry of metadata, with one README file per table that contains column definitions and guidelines on how to interpret and record metadata.

The metadata in each table covers four main categories: publication metadata (project name, year, and publication DOI), geographic metadata (site name, coordinates, and country), sample metadata (sample name, sample age, material type, and (meta)genome type) and sequencing archive information (archive, sample archive accession ID). Due to inconsistency in the ways metadata are reported in publications and archives, and to maintain concise records, we have specified (standardised) approximations for the reporting of sample ages, geographic locations, and archive accessions, following MIxS[14] categories where possible. This approach allows researchers using the dataset to access sufficiently approximate information during search queries to identify samples of interest (e.g. all samples from Italy dating from between 4500-2500 Before Present (BP), i.e., from 1950), which they can subsequently manually check in the original publication to obtain the exact dating information (e.g., Late Bronze Age, 3725+/−15 BP). Due to inconsistency in dating and reporting methods, dates are reported

Field	Description	Field Type	Field Format
project_name	Unique AncientMetagenomeDir key for study	String	FirstAuthorYYYY
publication_year	Publication year of study	Integer	YYYY
publication_doi	Publication DOI (or library permalink)	String	Regex
site_name	Specific locality name where sample taken from	String	Free text
latitude	Latitude in decimal coordinate (WGS84 projection)	Number	Max. 3 decimals
longitude	Longitude in decimal coordinate (WGS84 projection)	Number	Max. 3 decimals
geo_loc_name	Present-day country name (INSDC) that locality resides in	String	Restricted enum
sample_name	Name of sample as reported in publication or archive	String	Free text
sample_age	Approximate date (before 1950, rounded to last 100 years)	Integer	YYYY
sample_age_doi	DOI of source of date. Can be more recent publication.	String	Regex
collection_date	Date of sampling of material for genetic analysis	Integer	YYYY
archive	Name of established data repository	String	Restricted enum
archive_accession	Sample-level accession code in data repository	String	Free text

Table 1. Core fields that are required for all AncientMetagenomeDir sub-discipline tables, including field type and standardised formatting description. Field formats are defined in a JSON schema, against which each new study submission is cross-checked by automated continuous integration (CI) checks and community peer-review. Further sub-discipline specific fields are included in the corresponding table, as required by the community.

(where relevant) as uncalibrated years BP, and rounded to the nearest 100 years, due to the range of calculation and reporting methods (radiocarbon dating vs. historical records, calibrated vs. uncalibrated radiocarbon dates, etc.). We hope that future extensions of AncientMetagenomeDir will include more exact dating information, such as raw dates and radiocarbon lab codes, to allow for consistent calibration of whole datasets for more precise dating information. Geographic coordinates are restricted to a maximum of three decimals, with fewer decimals indicating location uncertainty (e.g., if a publication only reports a region rather than a specific site). For sequence accession codes, we opted for using *sample* accession codes rather than direct sequencing data IDs. This is due to the myriad ways in which data are generated and uploaded to repositories (e.g., one sample accession per sample vs. one sample accession per library; or uploading raw sequencing reads vs. only consensus sequences). We found that in most cases sample accession codes are the most straightforward starting points for data retrieval. However, we did observe errors in some data accessions uploaded to public repositories, such as multiple sample codes assigned to different libraries of the same sample, and insufficient metadata to link accessions to specific samples reported in a study. Overall, we found that heterogeneity in sample (meta)data uploading was a common problem, which highlights the need for improvements in both training and community-agreed standards for data sharing and metadata reporting in public repositories (such as an ancient metagenomic MIxS extension). In addition to metadata recorded across all sample types, we have added table-specific metadata fields to individual categories as required (e.g., species for single genomes and community type for microbiomes). Such fields can be further extended or modified with the agreement of the community.

Data validation. After all metadata has been added, a contributor makes a Pull Request (PR) into the master branch. Every PR undergoes an automated 'continuous-integration' validation check via the open-source companion tool AncientMetagenomeDirCheck[16] (https://github.com/SPAAM-community/AncientMetagenomeDirCheck, License: GNU GPLv3). This tool automatically checks each submission for conformity against a specification schema of minimum required information and formatting consistency (see Technical Validation). Usage of controlled vocabularies, alongside stable linking (via DOIs), within the specifications ensures reliable querying of the dataset, and allows future expansion to include richer metadata by linking to other databases. Descriptions for the minimum required fields for an AncientMetagenomeDir table are provided in Table 1.

Once automated checks are cleared, a contributor then requests a minimum of one peer-review performed by another member of the SPAAM community (termed 'reviewer'). This reviewer checks the entered data for consistency against the table's README file and also for accuracy against the original publication. Once the automated and peer-review checks are both satisfied, the publication's metadata are then added to the master branch and the corresponding Issue is closed. For each added publication, a CHANGELOG is maintained to track the papers included in each release and to record any corrections that may have been made (e.g., if new radiocarbon dates are published for previously entered samples). The CHANGELOG or Issues pages on GitHub can be consulted to check whether a given publication has already been added (or excluded) from a table. Proposals and submissions can be made at any time, and contributed data is available on the main GitHub repository immediately after integration into the master branch. However, citable versions of the database are only made on each new (non-modifiable) release (see section Data Records). New submissions or corrections received after a release are included in subsequent versions.

Data Records

AncientMetagenomeDir (https://github.com/SPAAM-community/AncientMetagenomeDir) and AncientMetagenomeDirCheck (https://github.com/SPAAM-community/AncientMetagenomeDirCheck) are both maintained on GitHub. AncientMetagenomeDir has regular quarterly releases, each of which has a release-specific DOI assigned via the Zenodo long-term data repository. Both the collection and tools are

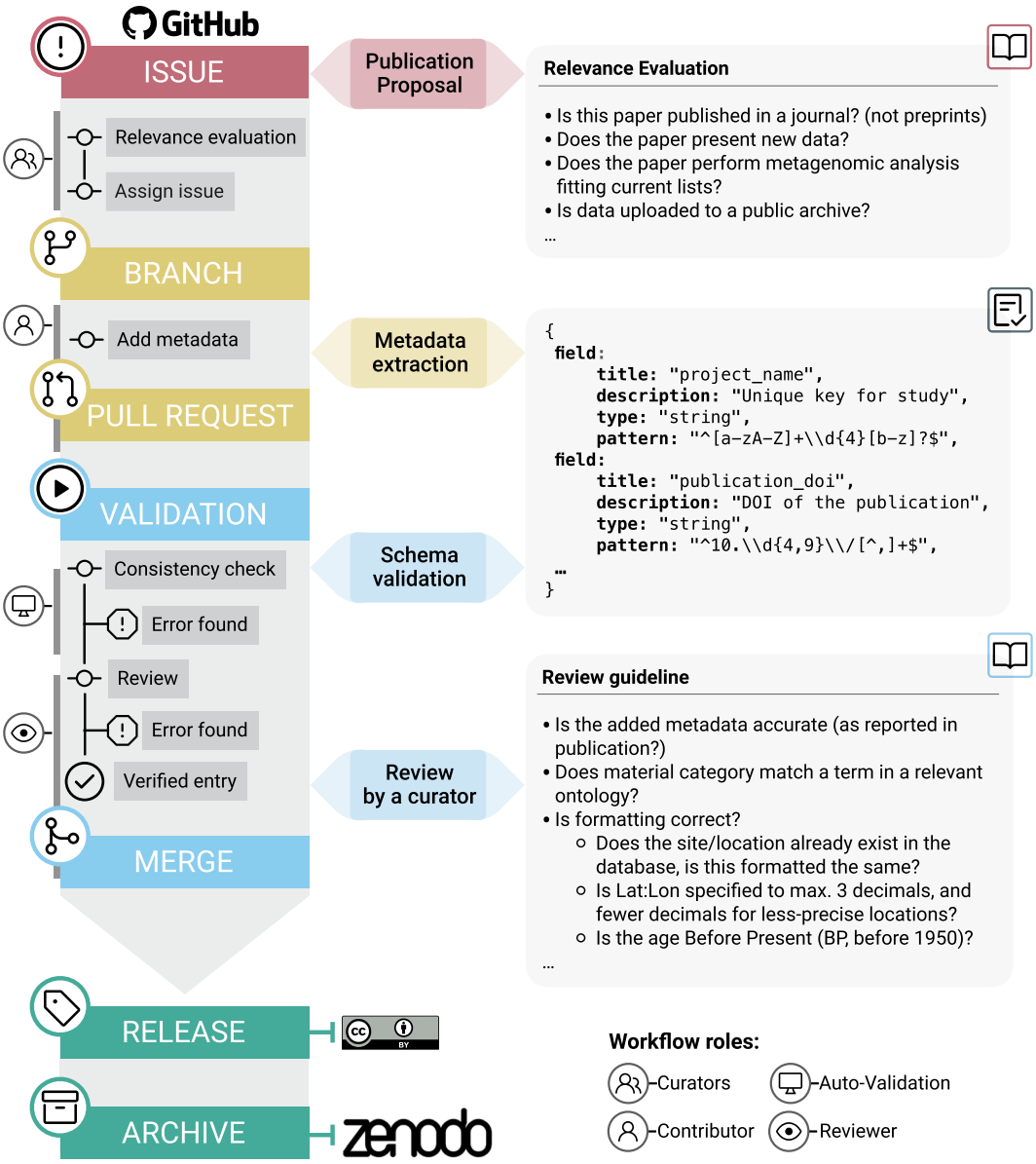


Fig. 3 AncientMetagenomeDir submission and update workflow. The submission workflow is carried out on GitHub, and final releases are archived at Zenodo. Submissions go through both automated computational validation and also peer-review for consistency and accuracy.

archived in the Zenodo repository with generalised DOIs[15] and[16], respectively. The full workflow can be seen in Fig. 3. Releases are made under a CC-BY 4.0 license (https://creativecommons.org/licenses/by/4.0/).

Technical Validation

All data entries to AncientMetagenomeDir undergo automated continuous-integration validation prior to submission into the protected main branch. These tests must pass before being additionally peer-reviewed by other member(s) of the community (see section Data Validation). Automated continuous-integration (CI) validation tests consist of regex patterns to control formatting of specified fields (e.g. DOIs, project IDs, date formats), and cross-checking of entries against controlled vocabularies defined in centralised JSON schema, often derived from established term-ontologies. For example, valid country codes are guided by the International Nucleotide Sequence Database Collaboration (INSDC) controlled vocabulary (http://www.insdc.org/country.html), host and microbial species names are defined by the NCBI's Taxonomy database (https://www.ncbi.nlm.nih.gov/taxonomy), and material types are defined by the ontologies listed on the EBI's Ontology Look Up service (https://www.ebi.ac.uk/ols/index) - particularly the Uberon[17] and Envo ontologies[18,19]. Entries must also have valid sample accession IDs corresponding to shotgun metagenomic, genome-enriched sequence data, or - when only available - consensus sequences, uploaded to established and stable public archives.

Usage Notes

Usage of the resource typically consists of loading the TSV file of interest in software such as Microsoft Excel, LibreOffice Calc, or R. The data table can be subsequently sorted or queried to identify datasets of interest. It should be noted that certain metadata fields (e.g., sample_age, latitude, and longitude) are approximate and do not provide *exact* values; rather, if exact values for these fields are required, they must be retrieved from the original publication or requested from the publications' authors. All selected data retrieved using AncientMetagenomeDir and used in subsequent studies should be cited using the original publication citation as well as AncientMetagenomeDir.

Retrieval of sequencing data using sample accession codes can be achieved manually via a given archive's website, or via archive-supplied tools (e.g., Entrez Programming Utilities for NCBI's SRA (https://www.ncbi.nlm.nih.gov/books/NBK179288/), or enaBrowserTools for EBI's ENA (https://github.com/enasequence/enaBrowserTools).

Contributions to the tables are also facilitated by extensive step-by-step documentation on how to use GitHub and AncientMetagenomeDir, the locations of which are listed on the main README of the repository.

Code availability

An R notebook used for generating images with package versions can be found in the AncientMetagenomeDir repository at https://github.com/SPAAM-community/AncientMetagenomeDir/tree/master/assets/analysis (commit 4308bb7). Code for validation of the dataset (with version 1 used for the first release of AncientMetagenomeDir) can be found at https://github.com/SPAAM-community/AncientMetagenomeDirCheck and https://doi.org/10.5281/zenodo.4003826.

Received: 18 September 2020; Accepted: 13 December 2020;
Published online: 26 January 2021

References

1. Anagnostou, P. *et al*. When data sharing gets close to 100%: what human paleogenetics can teach the open science movement. *PloS one* **10**, e0121409, https://doi.org/10.1371/journal.pone.0121409 (2015).
2. Wilkinson, M. D. *et al*. The FAIR guiding principles for scientific data management and stewardship. *Scientific data* **3**, 160018, https://doi.org/10.1038/sdata.2016.18 (2016).
3. Warinner, C. *et al*. A robust framework for microbial archaeology. *Annual review of genomics and human genetics* **18**, 321–356, https://doi.org/10.1146/annurev-genom-091416-035526 (2017).
4. Warinner, C., Speller, C., Collins, M. J. & Lewis, C. M. Jr. Ancient human microbiomes. *Journal of human evolution* **79**, 125–136, https://doi.org/10.1016/j.jhevol.2014.10.016 (2015).
5. Spyrou, M. A., Bos, K. I., Herbig, A. & Krause, J. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nature reviews. Genetics* **20**, 323–340, https://doi.org/10.1038/s41576-019-0119-1 (2019).
6. Edwards, M. E. The maturing relationship between quaternary paleoecology and ancient sedimentary DNA. *Quaternary Research* **96**, 39–47, https://doi.org/10.1017/qua.2020.52 (2020).
7. Dabney, J., Meyer, M. & Pääbo, S. Ancient DNA damage. *Cold Spring Harbor perspectives in biology* **5**, https://doi.org/10.1101/cshperspect.a012567 (2013).
8. Peyrégne, S. & Prüfer, K. Present-Day DNA contamination in ancient DNA datasets. *BioEssays: news and reviews in molecular, cellular and developmental biology* e2000081, https://doi.org/10.1002/bies.202000081 (2020).
9. Prendergast, M. E. & Sawchuk, E. Boots on the ground in africa's ancient DNA 'revolution': archaeological perspectives on ethics and best practices. *Antiquity* **92**, 803–815, https://doi.org/10.15184/aqy.2018.70 (2018).
10. Pálsdóttir, A. H., Bläuer, A., Rannamäe, E., Boessenkool, S. & Hallsson, J. H. Not a limitless resource: ethics and guidelines for destructive sampling of archaeofaunal remains. *Royal Society open science* **6**, 191059, https://doi.org/10.1098/rsos.191059 (2019).
11. Wagner, J. K. *et al*. Fostering responsible research on ancient DNA. *American journal of human genetics* **107**, 183–195, https://doi.org/10.1016/j.ajhg.2020.06.017 (2020).
12. Allentoft, M. E. *et al*. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society B: Biological Sciences* **279**, 4724–4733, https://doi.org/10.1098/rspb.2012.1745 (2012).
13. Kistler, L., Ware, R., Smith, O., Collins, M. & Allaby, R. G. A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic acids research* **45**, 6310–6320, https://doi.org/10.1093/nar/gkx361 (2017).
14. Yilmaz, P. *et al*. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature biotechnology* **29**, 415–420, https://doi.org/10.1038/nbt.1823 (2011).
15. Fellows Yates, J. A. *et al*. Spaam-community/ancientmetagenomedir: v20.09.1: Ancient ksour of ouadane. *Zenodo* https://doi.org/10.5281/zenodo.4011751 (2020).
16. Borry, M. & Fellows Yates, J. A. Spaam-commmunity/ancientmetagenomedircheck: Ancientmetagenomedircheck v1.0. *Zenodo* https://doi.org/10.5281/zenodo.4003826 (2020).
17. Mungall, C. J., Torniai, C., Gkoutos, G. V., Lewis, S. E. & Haendel, M. A. Uberon, an integrative multi-species anatomy ontology. *Genome biology* **13**, R5, https://doi.org/10.1186/gb-2012-13-1-r5 (2012).
18. Buttigieg, P. L. *et al*. The environment ontology: contextualising biological and biomedical entities. *Journal of biomedical semantics* **4**, 43, https://doi.org/10.1186/2041-1480-4-43 (2013).
19. Buttigieg, P. L. *et al*. The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. *Journal of biomedical semantics* **7**, 57, https://doi.org/10.1186/s13326-016-0097-6 (2016).

Acknowledgements

We would like to thank the wider SPAAM community (https://spaamcommunity.github.io) for their input in developing the project. J.A.F.Y., A.A.V., I.V., M.B., M.A.S, A.H. and C.W. acknowledge the Max Planck Society for financial support. J.A.F.Y was partly funded by the European Research Council (ERC) under the European Union's Horizon 2020 research innovation programme (ERC-2015-StG 678901-FoodTransforms to Philipp W. Stockhammer, Ludwig Maximilian University Munich, Germany). B.C. is supported by grant ERC-2014-ADG 670518 (to V. Gaffney, University of Bradford, United Kingdom). Å.J.V. is supported by Carlsbergfondet Semper Ardens grant CF18-1109 (to M. Thomas P. Gilbert, University of Copenhagen, Denmark). A.H. is partly supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's

Excellence Strategy–EXC 2051–Project-ID 390713860 (to C. Warinner, Friedrich Schiller University, Germany). E.J.G is supported by Arts & Humanities Research Council (grant number AH/N005015/1) and Natural History Museum (London, United Kingdom). M.J.B.-L. is supported by grant Wellcome Trust Seed Award in Science 208934/Z/17/Z, and by project IA201219 PAPIIT-DGAPA- UNAM (to María C. Ávila Arcos, LIIGH, Mexico). M.A.S. is supported by grant ERC-CoG 771234 PALEoRIDER (to Wolfgang Haak, Max-Planck-Institute for the Science of Human History, Germany). A.S.G is supported by NSF GRFP Grant No. DGE1255832 (any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation). S.L.R. is supported by NIH Genetics and Regulation Training Grant 5T32GM007197-46. I.V., M.B., and C.W. are supported by Werner Siemens Stiftung (Paleobiochemistry) (to C. Warinner, Leibnitz Institute for Natural Product Research and Infection Biology, Germany). Open Access funding enabled and organized by Projekt DEAL.

Author contributions

J.A.F.Y. and C.W. conceptualised the project. J.A.F.Y. designed the project and infrastructure with input from all co-authors. M.B. developed software. J.A.F.Y., A.A.V., Å.J.V., B.C., I.M.V., M.J.B.-L., A.F.-G., E.J.G., S.L.R., P.D.H., M.A.S., A.H., A.S.G., J.H., A.F.A., V.Z. and C.W. acquired data. J.A.F.Y. drafted the manuscript with input from all co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.A.F.Y. or C.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

The Creative Commons Public Domain Dedication waiver http://creativecommons.org/publicdomain/zero/1.0/ applies to the metadata files associated with this article.

© The Author(s) 2021

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Ancient viral genomes reveal introduction of HBV and B19V to Mexico during the transatlantic slave trade.

Axel A. Guzmán-Solís[1]*, Daniel Blanco-Melo[2*‡], Viridiana Villa-Islas[1], Miriam J. Bravo-López[1], Marcela Sandoval-Velasco[3], Julie K. Wesp[4], Jorge A. Gómez-Valdés[5], María de la Luz Moreno-Cabrera[6], Alejandro Meraz-Moreno[6], Gabriela Solís-Pichardo[7], Peter Schaaf[8], Benjamin R. tenOever[2,9,10] & María C. Ávila-Arcos[1‡].

1.International Laboratory for Human Genome Research, Universidad Nacional Autónoma de México (México); 2.Department of Microbiology, Icahn School of Medicine at Mount Sinai (NY, USA); 3.Section for Evolutionary Genomics, The Globe Institute, Faculty of Health, University of Copenhagen (Denmark); 4.Department of Sociology and Anthropology, North Carolina State University (USA); 5.Escuela Nacional de Antropología e Historia (México); 6.Instituto Nacional de Antropología e Historia (México); 7.Laboratorio Universitario de Geoquímica Isotópica (LUGIS), Instituto de Geología, Universidad Nacional Autónoma de México (México); 8.LUGIS, Instituto de Geofísica, Universidad Nacional Autónoma de México (México); 9.Virus Engineering Center for Therapeutics and Research (VECToR), Icahn School of Medicine at Mount Sinai (NY, USA); 10.Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai (NY, USA). * These authors contributed equally to this work. ‡ Correspondence: daniel.blancomelo@mssm.edu (D.B-M.), mavila@liigh.unam.mx (M.C.A-A.).

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ABSTRACT

After the European colonization of the Americas there was a dramatic population collapse of the Indigenous inhabitants caused in part by the introduction of new pathogens. Although there is much speculation on the etiology of the Colonial epidemics, direct evidence for the presence of specific viruses during the Colonial era is lacking. To uncover the diversity of viral pathogens during this period, we designed an enrichment assay targeting ancient DNA (aDNA) from viruses of clinical importance and applied it on DNA extracts from individuals found in a Colonial (16^{th} c. – 18^{th} c.) hospital and a Colonial chapel where records suggest victims of epidemics were buried during important outbreaks in Mexico City. This allowed us to reconstruct three ancient human parvovirus B19 genomes, and one ancient human hepatitis B virus genome from distinct individuals. The viral genomes are similar to African strains, consistent with the inferred morphological and genetic African ancestry of the hosts as well as with the isotopic analysis of the human remains, suggesting an origin on the African continent. This study provides direct molecular evidence of ancient viruses being transported to the Americas during the transatlantic slave trade and their subsequent introduction to New Spain. Altogether, our observations enrich the discussion about the etiology of infectious diseases during the Colonial period in Mexico.

INTRODUCTION

European colonization in the Americas resulted in a notable genetic exchange mainly between Native American populations, Europeans, and Africans[1–3]. Along

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

with human migrations, numerous new species were introduced to the Americas including bacterial and viral pathogens, which played a major role in the dramatic population collapse that afflicted the immunologically-naïve Indigenous inhabitants[4,5]. Among these pathogens, viral diseases, such as smallpox, measles and mumps have been proposed to be responsible for many of the devastating epidemics during the Colonial period[4]. Remarkably, the pathogen(s) responsible for the deadliest epidemics reported in New Spain (the Spanish viceroyalty that corresponds to Mexico, Central America, and the current US southwest states) remains unknown and is thought to have caused millions of deaths during the 16th century[4]. Indigenous populations were drastically affected by these mysterious epidemics, generically referred to as *Cocoliztli* ("pest" in Nahuatl)[6], followed by Africans and to a lesser extent European people[4,6,7]. Symptoms of the 1576 *Cocoliztli* epidemic were described in autopsy reports of victims treated at the "Hospital Real de San José de los Naturales" (HSJN)[6,8], the first hospital in Mexico dedicated specifically to treat the Indigenous population[6,8] (Figure 1a-b). The symptoms described included high fever, severe headache, neurological disorders, internal and external bleeding, hepatitis and intense jaundice[4,6,7]. This has led some scholars to postulate that the etiological agent of the *Cocoliztli* epidemic was a hemorrhagic fever virus[4,9], although others have suggested that the symptoms could be explained by bacterial infections[6,10].

The study of ancient viral genomes has revealed important insights into the evolution of specific viral families[11–19], as well as their interaction with human populations[20]. To explore the presence of viral pathogens in circulation during

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

epidemic periods in New Spain, we leveraged the vast historical and archeological information available for the Colonial HSJN. These include the skeletal remains of over 600 individuals recovered from mass burials associated with the hospital's architectural remnants (Figure 1b). Many of these remains were retrieved from burial contexts suggestive of an urgent and simultaneous disposal of the bodies, as in the case of an epidemic[8,21]. Prior bioarcheological research has shown that the remains of a number of individuals in the HSJN collection displayed dental modifications and/or morphological indicators typical of African ancestry[21], consistent with historical and archeological research that documents the presence of a large number of both free and enslaved Africans and their descendants in Colonial Mexico[3]. Indeed a recent paleogenomics study reported a Sub-Saharan African origin of three individuals from this collection[22].

Here we describe the screening for viral pathogens that circulated in New Spain during Colonial times, using ancient DNA (aDNA) techniques (Supplementary Figure 1). For this work, we sampled skeletal human remains recovered from the HSJN, and from a second site in Mexico City ("La Concepcion" chapel) belonging to one of the first catholic conversion centers in New Spain[23], where records suggest victims of epidemics were also buried (Figure 1a). We report the reconstruction of ancient hepatitis B virus (HBV) and human parvovirus B19 (B19V) genomes recovered from these remains. Our findings provide a direct molecular evidence of human viral pathogens of African origin being introduced to New Spain during the transatlantic slave trade.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

RESULTS

We sampled the skeletal remains from two archeological sites, a Colonial Hospital and a Colonial chapel in Mexico City (Figure 1a). For the HSJN (as a part of a broader ongoing study of African ancestry), 21 dental samples (premolar and molar teeth) were selected based on previous morphometric analyses and dental modifications that suggested an African ancestry[21,24–26]. The African presence in the Indigenous Hospital might reflect an urgent response to an epidemic outbreak, since hospitals treated patients regardless of the origin of the affected individuals during serious public health crises[21]. Dental samples of five additional individuals were selected (based on their conservation state) from “La Concepción” chapel (COY), which is located 10 km south of the HSJN in Coyoacán, a Pre-Hispanic Indigenous neighborhood that later became the first Spanish settlement in Mexico City after the fall of Tenochtitlan[23]. Following strict aDNA protocols, we processed these dental samples to isolate aDNA for next-generation sequencing (NGS) (Methods). Teeth roots (which are vascularized) can be assumed to be a good source of pathogen DNA[27], especially in the case of viruses that are widespread in the bloodstream during a systemic infection. Accordingly, a number of previous studies have successfully recovered ancient viral DNA from teeth roots[14–16,22].

Metagenomic analysis with MALT[10] on the NGS data using the Viral NCBI RefSeq database as a reference[28], revealed that sixteen samples contained traces of viral DNA, particularly of sequences related to *Hepadnaviridae*, *Herpesviridae*, *Parvoviridae* and *Poxviridae* (Figure 1c, Supplementary Figure 2, Methods). These viral hits revealed the potential to recover ancient viral genomes from these

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

samples. We selected thirteen samples for further screening based on the availability of the NGS library and the quality of the detected viral hit (best hit when using megaBLAST[29] to NCBI nr database) to a clinically important virus (HBV, B19V, Papillomavirus, Smallpox). To isolate and enrich the viral DNA fraction in the sequencing libraries, biotinylated single-stranded (ss) RNA probes designed to capture sequences from diverse human viral pathogens were synthesized (Supplementary Table 1). The selection of the viruses included in the capture design considered the following criteria: 1) DNA viruses previously retrieved from archeological human remains (i.e. Hepatitis B virus, Human Parvovirus B19, Variola Virus), 2) representative viruses from families capable of integrating into the human genome (i.e. *Herpesviridae*, *Papillomaviridae*, *Polyomaviridae*, *Circoviridae*) or 3) RNA viruses with a DNA intermediate (i.e. *Retroviridae*). Additionally, a virus-negative aDNA library, which showed no hits to any viral family included in the capture assay (except for a frequent *Poxviridae*-like region identified as an Alu repeat[30]), was captured and sequenced as a negative control (HSJN177) to estimate the efficiency of our capture assay. Four post-capture libraries had a ~100-300-fold increase of HBV-like hits or a ~50-100-fold increase of B19V-like hits (Figure 1c, Supplementary Table 2) compared to their corresponding pre-capture libraries (Methods). In contrast, the captured negative control (HSJN177) presented a negligible enrichment of these viral hits (Figure 1c, Supplementary Table 2).

We verified the authenticity of the viral sequences by querying the mapped reads against the non-redundant (nr) NCBI database using megaBLAST[29]. We only

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

retained reads for which the top hit was to either B19V or HBV, respectively (Supplementary Table 3). For each sample, the retained reads were mapped simultaneously to different available genotype sequences for their respective virus, and the genotype with the best sequence coverage was selected as the reference for final mapping (Supplementary Methods). To confirm the ancient origin of these viral reads, we evaluated the misincorporation damage patterns using the program mapDamage 2.0[31], which revealed an accumulation of C to T mutations towards their 5' terminal site with an almost symmetrical G to A pattern on the 3' end (Figure 2a, Supplementary Figure 3), as expected for aDNA[32]. Three ancient B19V genomes were reconstructed (Figure 2b, Supplementary Table 3) with sequence coverages between 92.37% and 99.1%, and average depths of 2.98-15.36X along their single stranded DNA (ssDNA) coding region, which excludes the double stranded DNA (dsDNA) hairpin regions at each end of the genome[33]. These inverse terminal repeats (ITRs) displayed considerably higher depth values compared to the coding region (<218X) consistent with the better *post-mortem* preservation of dsDNA compared to ssDNA[34] (Figure 2b). In addition, we reconstructed one ancient HBV genome (Figure 2c, Supplementary Table 3) at 30.8X average depth and with a sequence coverage of 89.9%, including its ssDNA region at a reduced depth (<10X). This genome presents a 6 nucleotide (nt) insertion in the core gene, that is characteristic of the genotype A[35]. Further phylogenetic analyses (Methods) revealed that the Colonial HBV genome clustered with modern sequences corresponding to sub-genotype A4 (previously named A6)[36] (Figure 3a, Supplementary Figure 4). The Genotype A (HBV/GtA) has a broad diversity in Africa reflecting its long history in this continent[35,37], while the

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

sub-genotype A4 has been recovered uniquely from African individuals in Belgium[38] and has never been found in the Americas. Regarding the three Colonial B19V genomes from individuals HSJN240, COYC4 and HSJNC81, these were phylogenetically closer to modern B19V sequences belonging to genotype 3 (Figure 3b, Supplementary Figure 5). This B19V genotype is divided into two sub-genotypes: 3a that is mostly found in Africa, and 3b, which is proposed to have spread outside Africa in the last decades[39]. The viral sequences from the individuals HSJN240 and COYC4 are similar to sub-genotype 3b genomes sampled from immigrants (Morocco, Egypt and Turkey) in Germany[40] (Figure 3b, Supplementary Figure 5); while the sequence of the individual HSJNC81 is more similar to a divergent sub-genotype 3a strain (Figure 3b, Supplementary Figure 5) retrieved from a child with severe anemia from France[41].

We recalibrated the molecular clock of HBV (Supplementary Figure 4b) and B19V (Supplementary Figure 5b) with a dated coalescent phylogeny analysis that included our ancient Colonial viral genomes, as well as other published ancient and modern HBV and B19V genomes from different genotypes in diverse geographical regions. The median evolutionary rate for HBV was estimated to be between 1.1×10^{-5} and 9.5×10^{-6} (mean of 7.1×10^{-6}) nucleotide substitutions per site per year (s/s/y) with a divergence time from last common ancestor between 8.3 and 13.61 (mean of 10.68) thousand years ago (ka) under a relaxed molecular clock model and a coalescent exponential population prior (Supplementary Figure 4b). For B19V, the estimated substitution rate is 1.3×10^{-5} and 9.9×10^{-6} (mean of 1.17×10^{-6}) s/s/y with a divergence time from last common ancestor between 9.45

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and 11.83 (mean 10.6) ka under a strict molecular clock model and a coalescent Bayesian skyline population prior (Supplementary Figure 5b). The recalibrated estimations are similar to previous reports[15,16] that included only Eurasian viral genomes considerably older (~0.2–24.0 ka) than our Colonial viral genomes.

The similarity of our ancient genomes to African strains is consistent with the previously reported morphological indicators of African ancestry[21,24–26] of the corresponding HSJN individuals, as well as to an ancient DNA analysis based on three individuals from the HSJN collection[22]. We thus used the *de novo* generated sequence data to determine the mitochondrial haplogroup of the hosts, as well as their autosomal genetic ancestry using the 1000 Genomes Project[42] as a reference panel (Figure 3c, Supplementary Table 4). The nuclear genetic ancestry analysis showed that all three HSJN individuals, from which the reconstructed viral genomes were isolated, fall within African genetic variation in a Principal Component Analysis plot (Figure 3c), while their mitochondrial DNA belong to the L haplogroup, which has high frequency in African populations (Supplementary Table 4). Additionally, we performed $^{87}Sr/^{86}Sr$ isotopic analysis on two of the HSJN individuals using teeth enamel as well as phalange (HSJN240) or parietal bone (HSJNC81) to provide insights on the places of birth (adult enamel) and where the last years of life were spent (phalange/parietal). The $^{87}Sr/^{86}Sr$ ratios measured on the enamel of the individual HSJNC81 (0.71098) and HSJN240 (0.71109) are similar to average $^{87}Sr/^{86}Sr$ ratios found in soils and rocks from West Africa (average of 0.71044, Supplementary Figure 6, Supplementary Tables 5 and 6). In

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

contrast, the $^{87}Sr/^{86}Sr$ ratios on the parietal and phalange bones from the HSJNC81 (0.70672) and HSJN240 (0.70755), show lower values similar to those observed in the Trans Mexican Volcanic Belt where the Mexico City Valley is located (0.70420 - 0.70550, Supplementary Figure 6, Supplementary Tables 5 and 6). Moreover, radiocarbon dating of HSJN240 (1442-1608 CE, years calibrated for 1σ) and HSJN194 (1472-1625 CE, years calibrated for 1σ) (Supplementary Table 4, Supplementary Figure 7) indicates that these individuals arrived during the first decades of the Colonial period, when the number of enslaved individuals arriving from Africa was particularly high[3]. Strikingly, Colonial individual COYC4, who was also infected with an African B19V strain, clusters with present-day Latin-American populations from the 1000 Genomes Project (Figure 3c), suggesting that following introduction from Africa, the virus (B19V) spread and infected people of different ancestries during Colonial times.

DISCUSSION

In this study we reconstructed one HBV and three B19V ancient genomes from four different individuals using NGS metagenomics and in-solution targeted enrichment methods (Figure 2b, c, Supplementary Figure 1). Several lines of evidence support that these ancient viral genomes are authentic and not an environmental contamination nor a capture artifact. First, our negative control did not show a considerable enrichment for B19V or HBV hits (Figure 1c). For those samples that showed an enrichment in viral sequences after capture, the reads covered the reference genomes almost in their entirety and displayed deamination patterns at the terminal ends of the reads as expected for aDNA (Figure 2a).

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Moreover, it is important to notice that B19V and HBV are blood-borne human pathogens that are not present in soil or the environment, and that DNA from these viruses had never been extracted before in the aDNA facilities used in this study.

We also described an unusual coverage pattern on the B19V genome, where the dsDNA hairpins at its terminal sites are highly covered reflecting a better stability of these regions over time (Figure 2b). Similarly, the partially circular dsDNA genome from HBV was poorly covered at the ssDNA region (Figure 2c), as in at least three previous ancient HBV genomes[14] (Supplementary Discussion 1). The variable coverage in both viruses argues against an integration event of these viruses, that would result as an uniform dsDNA coverage; further analysis are needed to elucidate if the aDNA retrieved in this and other studies is coming from circulating virions or from cell-free DNA intermediates[43] produced after viral replication in bone marrow or liver, for B19V and HBV, respectively[44,45].

The ancient B19V genomes were assigned to genotype 3. This genotype is considered endemic and the most prevalent in West Africa (Ghana: 100%, 11/11; Burkina Faso: 100%, 5/5)[39,46,47] and a potential African origin has been suggested[46]. This genotype has been also sporadically found outside Africa[46,47] in countries historically tied to this continent, like Brazil (50%, 6/12)[48,49], India (15.4%, 2/13)[50], France (11.4%, 9/79)[41,51], and USA (0.85%, 1/117)[47] as well as in immigrants from Morocco, Egypt, and Turkey in Germany (6.7%, 4/59)[40]. Two other genotypes, 1 and 2 exist for this virus. Genotype 1 is the most common and is found worldwide, while the almost extinct genotype 2 is mainly found in elderly people from Northern Europe[52]. Ancient genomes from genotypes 1 and 2 have

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

been recovered from Eurasian samples, including a genotype 2 B19V genome from a 10th century Viking burial in Greenland[16]. $^{87}Sr/^{86}Sr$ isotopes on individuals from the same burial revealed they were immigrants from Iceland[16], suggesting an introduction of the genotype 2 to contiguous North America during Viking explorations of Greenland. While serological evidence indicates that B19V currently circulates in Mexico, only the presence of genotype 1 has been formally described using molecular analyses[53]. Taken together, our results are consistent with an introduction of the genotype 3 to New Spain as a consequence of the transatlantic slave trade imposed by the European colonization. This hypothesis is supported by the $^{87}Sr/^{86}Sr$ isotopic analysis, which indicates that the individuals from the HSJN with B19V (HSJN240, HSJNC81) were born in West Africa and spent their last years of life in New Spain (Supplementary Figure 6). Furthermore, the radiocarbon ages of individuals HSJN240, HSJN194 (Supplementary Figure 7) support this notion as they correspond to the Early Colonial period, during which the number of enslaved Africans arriving was higher compared to later periods[3]. Remarkably, a similar B19V belonging to the genotype 3 was recovered from an Indigenous individual (COYC4) in an independent archeological site 10 Km south of the HSJN (Figure 1a), supporting that viral transmissions between African individuals and people of different ancestries occurred during the Colonial period in Mexico City.

The genotype A from HBV is highly diverse in Africa, reflecting its long evolutionary history, originated somewhere between Africa, Middle East and Central Asia[37]. The

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

287 introduction of the genotype A from Africa to America has been proposed based on
288 phylogenetic analysis of modern strains for Brazil[37,49] and Mexico[54], and more
289 precisely to the sub-genotype A1 for sequences from Martinique[55], Venezuela[56],
290 Haiti[57] and Colombia[58]. Recently, a similar introduction pattern was proposed for
291 the quasi genotype A3 based on an ancient HBV genome recovered from African
292 individuals sampled in Mexico[22]. The Colonial ancient HBV genome reconstructed
293 in our work is assigned to the genotype A4 (Figure 3a, Supplementary Figure 4),
294 which has never been reported in the Americas. A similar introduction pattern has
295 been proposed for other human-infecting viruses such as smallpox[7,59], based on
296 historical records; or Yellow fever virus[60], HTLMV-1[61], Hepatitis C virus (genotype
297 2)[62] and human herpes simplex virus[63] based on phylogenetic analysis of modern
298 strains from Afro-descendant or admixed human populations.
299
300 Although we cannot assert where exactly the African-born individuals in this study
301 contracted B19V or HBV (Africa, America, or the Middle Passage) nor if the cause
302 of their deaths can be attributed to such infections, the identification of ancient
303 B19V and HBV in contexts associated with Colonial epidemics in Mexico City is still
304 relevant in light of their paleopathological marks and the clinical information
305 available for the closest sequences in the phylogenetic analyses. The
306 reconstructed ancient B19V genome from individual HSJNC81 is closest to the V9
307 strain, which was isolated from an individual with severe anemia[41] (AJ249437)
308 (Figure 3b, Supplementary Figure 5). In fact, individual HSJNC81 displayed cribra

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

orbitalia in the eye sockets and porotic hyperostosis on the cranial vault (Supplementary Figure 8); these morphological changes to the skeleton are typically associated with anemias of varying different causes[64]. It is acknowledged that B19V can result in a severe or even fatal anemia due to the low level of hemoglobin when present simultaneously with other blood disorders, as thalassemia, sickle-cell anemia, malaria and iron deficiency[45,65]. Therefore, since B19V infects precursors of the erythroid lineage[45], it is possible that the morphological changes found in HSJNC81 might be the result of a severe anemia caused or enhanced by a B19V infection (Supplementary Discussion 2). Moreover, the identification of ancient B19V in a Colonial context is noteworthy considering there are historical records that describe the treatment of an outbreak of measles at the HSJN in 1531[21], that numerous cases of measles were reported during this period[4,8,59], and that several recent reports reveal that measles-like cases were actually attributable to B19V[66,67] (Supplementary Discussion 2). Nevertheless, this hypothesis requires future and more comprehensive studies aimed to characterize the presence of measles and rubella viruses from ancient remains. A task that imposes difficult technical challenges given that RNA is known to degrade rapidly, in fact the most ancient viral RNA genomes (1912 CE) have been recovered only from formalin-fixed tissue[17,18]. Additionally, historical records of the autopsies of victims of the 1576 *Cocoliztli* epidemic treated at the HSJN, describe the observation of enlarged hard liver and jaundice[4,6,7,9,68], which could be explained by severe liver damage or epidemic hepatitis[4,6]. This is noteworthy given both viruses HBV and B19V proliferate in the liver, have been isolated from liver tissue and are associated with hepatitis and jaundice[44,45]. However, it is important to acknowledge

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

that both viruses have also been previously identified in aDNA datasets not necessarily associated with disease or epidemic contexts[11,12,14,15], thus establishing a direct link would require additional samples and a more comprehensive pathogen screening to rule out the involvement of other pathogens. Finally, although our data is not a strong evidence that the reported manifestations of liver damage in *Cocolitzi* autopsies were directly caused by HBV or B19V, the identification of these viruses in likely victims of epidemic outbreaks in the Colonial period opens up new opportunities for investigating the presence of these viruses in similar contexts. This type of research is particularly relevant when considering previous hypotheses favoring the synergistic action of different pathogens in these devastating Colonial epidemics[7] (Supplementary Discussion 3).

It is important to emphasize that our findings should be interpreted with careful consideration of the historical and social context of the transatlantic slave trade. This cruel episode in history involved the forced transportation of millions of individuals to the Americas (ca. 250,000 to New Spain[3]) under inhumane, unsanitary and overcrowded conditions that, with no doubt, favored the spread of infectious diseases[59]. Therefore, the introduction of these and other pathogens from Africa to the Americas should be attributed to the brutal and harsh conditions of the Middle Passage that enslaved Africans were subjected to by traders and colonizers, and not to the African peoples themselves. Moreover, the adverse life conditions for enslaved Africans and Native Americans, especially during the first decades after colonization, surely favored the spread of diseases and emergence

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

of epidemics[59]. Integrative and multidisciplinary approaches are thus needed to understand this phenomenon at its full spectrum.

In summary, our study provides direct aDNA evidence of HBV and B19V introduced to the Americas from Africa during the transatlantic slave trade. The isolation and characterization of these ancient HBV and B19V genomes represent an important contribution to the recently reported ancient viral genome reported in the Americas[22] (only one before the present study). Our results expand our knowledge on the viral agents that were in circulation during Colonial epidemics like *Cocoliztli*, some of which resulted in the catastrophic collapse of the immunologically-naïve Indigenous population. Although we cannot assign a direct causality link between HBV and B19V and *Cocoliztli*, our findings confirm that these potentially harmful viruses were indeed circulating in individuals found in archeological contexts associated with this epidemic outbreak. Further analyses from different sites and samples will help understand the possible role of these and other pathogens in Colonial epidemics, as well as the full spectrum of pathogens that were introduced to the Americas during European colonization.

METHODS

Sample selection and DNA extraction

Dental samples (premolars and molars) were obtained from twenty-one individuals from the skeletal collection of the HSJN, selected based on their African-related skeletal indicators[21,24–26]. Five additional samples were taken from “La Concepción” chapel, based on their conservation state. Permits to carry out this sampling and

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

aDNA analyses were obtained by the Archeology Council of the National Institute of Anthropology and History (INAH).

DNA extraction and NGS library construction

Bone samples were transported to a dedicated ancient DNA clean-room laboratory at the International Laboratory for Human Genome Research (LIIGH-UNAM, Querétaro, Mexico), were DNA extraction and NGS-libraries construction was performed under the guidelines on contamination control for aDNA studies[69]. Previously reported aDNA extraction protocols were used for the HSJN[70] and COY[71] samples. Double-stranded DNA (dsDNA) indexed sequencing libraries were constructed from the DNA extract, as previously reported[72]. In order to detect contaminants in reagents or by human manipulation, extraction and library constructions protocols included negative controls (NGS blanks) that were analyzed in parallel with the same methodology. The resulting NGS dsDNA indexed libraries were quantified with a Bioanalyzer 2100 (Agilent) and pooled into equimolar concentrations.

NGS sequencing

Pooled libraries were paired-end sequenced on an Illumina NextSeq550 at the "Laboratorio Nacional de Genómica para la Biodiversidad" (LANGEBIO, Irapuato, Mexico), with a Mid-output 2x75 format. The reads obtained (R1 and R2) were merged (>11bp overlap) and trimmed with AdapterRemoval 1.5.4[73]. Overlapping reads (>30 bp in length) were kept and mapped to the human genome (hg19) using BWA 0.7.13[74], mapped reads were used for further human analysis (genetic

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ancestry, and mitochondrial haplogroup determination), whereas unmapped reads were used for metagenomic analysis and viral genome reconstruction.

Metagenomic analyses

The NCBI Viral RefSeq database was downloaded on February 2018; this included 7530 viral genomes. MALT 0.4.0[10] software was used to taxonomically classify the reads using the viral genomes database. The viral database was formatted automatically with malt-build once, and not human (unmapped) reads were aligned with malt-run (85 minimal percent identity). The produced RMA files with the viral abundances were normalized (default parameters) and compared to all the samples from the same archeological site with MEGAN 6.8.0[75].

Capture-enrichment Assay

Twenty-nine viruses were included in the in-solution enrichment design, the complete list of NCBI IDs is provided in Supplementary methods 5 and Supplementary Table 1. It contained viral genomes previously recovered from archeological remains like B19V, B19V-V9, and HBV (consensus genomes), selected VARV genes, as well as clinically important viral families that are able to integrate into the human genome, have dsDNA genomes, or dsDNA intermediates. The resulting design comprised 19,147 ssRNA 80 nt probes targeting with a 20 nt interspaced distance the whole or partial informative regions of eight viral families of clinical relevance (*Poxviridae, Hepadnaviridae, Parvoviridae, Herpesviridae, Retroviridae, Papillomaviridae, Polyomaviridae, Circoviridae*). To avoid a simultaneous false-positive DNA enrichment, low complexity regions and human-

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

like (hg38) sequences were removed (*in sillico*). The customized kit was produced by Arbor Biosciences (Ann Arbor, MI, USA). Capture-enrichment was performed on the indexed libraries based on the manufacturer's protocol (version 4) to pull-down aDNA with minimal modifications. Libraries were amplified with 18-20 cycles (Phusion U Hot Start DNA Polymerase by Thermo Fischer Scientific), purified with SPRISelect Magnetic Beads (Beckman Coulter) and quantified with a Bioanalyzer 2100 (Agilent). Amplified libraries were then pooled in different concentrations and deep sequenced yielding >1x10^6 non-human reads (Supplementary Table 4) in order to saturate the target viral genome. Reads generated from each enriched library were analyzed exactly as the shotgun (not-enriched) libraries. Normalized abundances between shotgun and captured libraries were compared in MEGAN 6.8.0[75] to evaluate the efficiency and specificity of the enrichment assay.

Viral datasets

The full list of accession numbers of the following datasets is given in Supplementary Methods 8.

HBV-Dataset-1 (HBV/DS1): comprises 35 HBV genomes from A-J human genotypes, 2 well-covered ancient HBV genomes (LT992443, LT992459) and a wholly monkey genome.

HBV-Dataset-2 (HBV/DS2): comprises 110 genomes based on a previous phylogenetic analysis[15], that included genomes from A-J genotypes as well as not-human primates HBV genomes (gibbon, gorilla, and chimpanzee), 16 ancient HBVgenomes[11,12,14,15] and one ancient HBV genome from this study (HSJN194).

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

B19V-Dataset-1 (B19V/DS1): comprises 13 B19V genomes from human genotypes 1-3 as well as a bovine parvovirus.

B19V-Dataset-2 (B19V/DS2): comprises 112 genomes from 1 to 3 B19V genotypes based on previous phylogenetic analysis[16], that included the 10 best-covered ancient genomes from genotype 1 and 2[16] as well as 3 ancient B19V from this study. Since many of the reported genomes are not complete, only the whole coding region (CDS) was used.

Genome Reconstruction and authenticity

HBV: Non-human reads were simultaneously mapped to HBV/DS1 with BWA (aln algorithm) with seedling disabled[76]. The reference sequence with the most hits was used to map uniquely to this reference and generate a BAM alignment without duplicates, from which damage patterns were determined and damaged sites rescaled using mapDamage 2.0[31], the rescaled alignment was used to produce a consensus genome. All the HBV mapped reads were analyzed through megaBLAST[29] using the whole NCBI nr database, in order to verify they were assigned uniquely to HBV (carried out with Krona 2.7[77]).

B19V: The reconstruction of the B19V ancient genome was done as previously reported from archeological skeletal remains[16], but increasing the stringency of some parameters as described next. Non-human reads were mapped against B19V/DS1 with BWA (aln algorithm) with seedling disabled[76], if more than 50% of the genome was covered, the sample was considered positive to B19V. Reads from the B19V-positive libraries were aligned with blastn (-evalue 0.001) to B19V/DS1 in order to recover all the parvovirus-like reads. To avoid local

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

alignments, only hits covering >85% of the read were kept and joined to the B19V mapped reads (from BWA), duplicates were removed. The resulting reads were analyzed with megaBLAST[29] using the whole NCBI nr database to verify the top hit was to B19V (carried out with Krona 2.7[77]). This pipeline was applied for two independent enrichments assays per sample and the filtered reads from the two capture rounds were joined. The merged datasets per sample were mapped using as a reference file the three known B19V genotypes with GeneiousPrime 2019.0.4[78] using median/fast sensibility and iterate up to 5 times. The genotype with the longest covered sequence was selected as the reference for further analysis. Deamination patterns for HBV and B19V were determined with mapDamage 2.0[31] and damaged sites were rescaled in the same program to produce a consensus whole genome using SAMtools 1.9[79].

Phylogenetic analyses

HBV/DS2 and B19V/DS2 were aligned independently in Aliview[80] (Muscle algorithm), and evolutionary models were tested under an AICc and BIC in jModelTest[81]. A neighbor joining tree with 1000 bootstraps was generated in GeneiousPrime 2019.0.4[78] using a Hasegawa-Kishino-Yano (HKY) model for both alignments. A maximum likelihood tree with 1000 bootstraps was produced in RAxML 8.2.10[82] using as a model a Generalized Time-Reversible (GTR)+G and GTR+G+I for B19V and HBV, respectively.

Since a temporal signal has been described for the coding region of B19V[16] and the whole genome of HBV[14,15], a Bayesian tree was generated to estimate the impact of the Colonial viral genomes on the divergence time from the most recent

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

499 common ancestor (MRCA). We used BEAST 1.8.4[83], a Bayesian skyline
500 population prior and a relaxed lognormal or strict molecular clock for B19V and a
501 Coalescent Exponential Population prior with relaxed lognormal molecular clock for
502 HBV as previously tested[15,16]. All parameters were mixed and converged into an
503 estimated sample size (ESS) >150 analyzed in Tracer 1.7[84]. The first 25% of trees
504 where discarded (burn in) and a Maximum Clade Credibility Tree was created with
505 TreeAnnotator[83]. The generated trees were visualized and edited in FigTree 1.4.3
506 with a midpoint root.
507
508 *Human population genetic analyses*
509 Human-mapped reads (BWA aln) obtained from the pre-capture sequence data of
510 viral-positive samples were used to infer the genetic ancestry of the hosts. A
511 Principal Components Analysis (PCA) was carried out using 16 populations from
512 the 1000 Genomes Project[42] reference panel including genotype data of 1,562,771
513 single nucleotide variants (SNVs) from 2,504 individuals (phase 3). Genomic
514 alignments of each ancient individual (HSJNC81, HSJN240, HSJN194 and
515 COYC4) were intersected with the positions of the SNVs present in the reference
516 panel genotype data. Pseudo haploid genotypes were called by randomly selecting
517 one allele at each intersected site and filtering by a base quality >30. Pseudo
518 haploid genotypes were also called for the complete reference panel. PCA was
519 performed on the merged ancient and modern dataset with smartpca (EIGENSOFT
520 package)[85,86] using the option *lsqproject* to project the ancient individuals into the
521 PC space defined by the modern individuals.
522

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Mitochondrial haplogroup and sex determination

NGS reads were mapped to the human mitochondrial genome reference (rCRS) with BWA (aln algorithm, -l default), the alignment file was then used to generate a consensus mitochondrial genome with program Schmutzi[87] The assignment of the mitochondrial haplogroup was carried out with Haplogrep[88,89] using the consensus sequence as the input. Assignment of biological sex was inferred based on the fraction of reads mapped to the Y-chromosome (Ry) compared to those mapping to the Y and X-chromosome[90]. Ry<0.016 and Ry>0.075 were considered XX or XY genotype, respectively. The resulting sex was coherent with the inferred morphologically (Supplementary Methods).

ACKNOWLEDGMENTS

This work was funded by the Wellcome Trust Sanger grant number 208934/Z/17/Z, and by project IA201219 PAPIIT-DGAPA-UNAM. D.B-M is an Open Philanthropy Fellow of the Life Sciences Research Foundation (LSRF). We thank the INAH Archeology Council for the sample permissions for this study. We are grateful with Teodoro Hernández Treviño, Gerardo Arrieta García from the “Laboratorio Universitario de Geoquímica Isotópica” (LUGIS-UNAM) for their technical support in performing the $^{87}Sr/^{86}Sr$ analyses and to Luis Alberto Aguilar Bautista, Alejandro de León Cuevas, Carlos Sair Flores Bautista and Jair Garcia Sotelo from the “Laboratorio Nacional de Visualización Científica Avanzada” (LAVIS/UNAM) who stored our data and provided the computational resources to perform this study.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We thank Alejandra Castillo Carbajal and Carina Uribe Díaz for technical support throughout the project.

DATA AVAILABILITY

Reconstructed genomes from this study are available in Genbank under accession number MT108214, MT108215, MT108216, MT108217. Accession number of sequences used in phylogenetic analysis are indicated in supplementary information. NGS data is available upon reasonable request.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

555 **REFERENCES:**

556

557 1. Rotimi, C. N., Tekola-Ayele, F., Baker, J. L. & Shriner, D. The African
558 Diaspora: History, Adaptation and Health. *Curr. Opin. Genet. Dev.* **41**, 77–84
559 (2016).
560 2. Salas, A. *et al.* The African Diaspora: Mitochondrial DNA and the Atlantic
561 Slave Trade. *Am. J. Hum. Genet.* **74**, 454–465 (2004).
562 3. Aguirre-Beltrán, G. La presencia del negro en México. *Rev. CESLA* **7**, 351–
563 367 (2005).
564 4. Acuna-Soto, R., Stahle, D. W., Therrell, M. D., Griffin, R. D. & Cleaveland,
565 M. K. When half of the population died: the epidemic of hemorrhagic fevers of 1576
566 in Mexico. *FEMS Microbiol. Lett.* **240**, 1–5 (2004).
567 5. Lindo, J. *et al.* A time transect of exomes from a Native American population
568 before and after European contact. *Nat. Commun.* **7**, 13175 (2016).
569 6. Malvido, E. & Viesca, C. La epidemia de cocoliztli de 1576. in *Ensayos*
570 *sobre la historia de las epidemias en México.* (eds. Florescano, E. & Malvido, E.)
571 27–32 (Instituto Mexicano del Seguro Social, 1982).
572 7. Somolinos d'Árdois, G. Las epidemias en México durante el siglo XVI. in
573 *Ensayos sobre la historia de las epidemias en México.* (eds. Florescano, E. &
574 Malvido, E.) 138–143 (Instituto Mexicano del Seguro Social, 1982).
575 8. Wesp, J. K. Caring for Bodies or Simply Saving Souls: The Emergence of
576 Institutional Care in Spanish Colonial America. in *New Developments in the*
577 *Bioarchaeology of Care: Further Case Studies and Expanded Theory* (eds. Tilley,
578 L. & Schrenk, A. A.) 253–276 (Springer International Publishing, 2017).
579 9. Marr, J. S. & Kiracofe, J. B. Was the huey cocoliztli a haemorrhagic fever?
580 *Med. Hist.* **44**, 341–362 (2000).
581 10. Vågene, Å. J. *et al.* Salmonella enterica genomes from victims of a major
582 sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2**, 520 (2018).
583 11. Kahila Bar-Gal, G. *et al.* Tracing hepatitis B virus to the 16th century in a
584 Korean mummy. *Hepatol. Baltim. Md* **56**, 1671–1680 (2012).
585 12. Patterson Ross, Z. *et al.* The paradox of HBV evolution as revealed from a

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

16th century mummy. *PLoS Pathog.* **14**, e1006750 (2018).

13. Duggan, A. T. *et al.* 17th Century Variola Virus Reveals the Recent History of Smallpox. *Curr. Biol. CB* **26**, 3407–3412 (2016).

14. Krause-Kyora, B. *et al.* Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. *eLife* **7**, (2018).

15. Mühlemann, B. *et al.* Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423 (2018).

16. Mühlemann, B. *et al.* Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 7557–7562 (2018).

17. Xiao, Y.-L. *et al.* High-throughput RNA sequencing of a formalin-fixed, paraffin-embedded autopsy lung tissue sample from the 1918 influenza pandemic. *J. Pathol.* **229**, 535–545 (2013).

18. Düx, A. *et al. The history of measles: from a 1912 genome to an antique origin*. http://biorxiv.org/lookup/doi/10.1101/2019.12.29.889667 (2019) doi:10.1101/2019.12.29.889667.

19. Pajer, P. *et al.* Characterization of Two Historic Smallpox Specimens from a Czech Museum. *Viruses* **9**, (2017).

20. Spyrou, M. A., Bos, K. I., Herbig, A. & Krause, J. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat. Rev. Genet.* **20**, 323–340 (2019).

21. Meza, A. Presencia africana en el cementerio del Hospital Real de San Jóse de los Naturales. *Arqueol. Mex.* **119**, 40–44 (2013).

22. Barquera, R. *et al.* Origin and Health Status of First-Generation Africans from Early Colonial Mexico. *Curr. Biol.* S0960982220304826 (2020) doi:10.1016/j.cub.2020.04.002.

23. Moreno-Cabrera, M. de la L., Meraz-Moreno, A. & Cervantes-Rosado, J. Relleno aligerado con vasijas cerámicas en el templo de la Inmaculada Concepción, en Coyoacán. in *Boletín de monumentos históricos 35* vol. 35 121–134 (Instituto Nacional de Antropología e Historia, 2015).

24. Ruíz-Albarrán, P. *Estudio de variabilidad biológica en la colección*

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

esquelética Hospital Real de San José de los Naturales. Un acercamiento a tráves de la técnica de morfometría geométrica. (Escuela Nacional de Antropología e Historia., 2012).

25. Hernández-Lopez, P. E. & Negrete, S. *¿Realmente Eran Indios? Afinidad biológica entre las personas atendidas en el Hospital Real San Jose de los Naturales, siglos XVI - XVIII.* (Escuela Nacional de Antropología e Historia., 2012).

26. Karam-Tapia, C. E. *Estimación del Mestizaje Mediante la Morfología Dental en la Ciudad de México (Siglo XVI al XIX).* (Escuela Nacional de Antropología e Historia., 2012).

27. Key, F. M., Posth, C., Krause, J., Herbig, A. & Bos, K. I. Mining Metagenomic Data Sets for Ancient DNA: Recommended Protocols for Authentication. *Trends Genet. TIG* **33**, 508–520 (2017).

28. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61-65 (2007).

29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

30. Tithi, S. S., Aylward, F. O., Jensen, R. V. & Zhang, L. FastViromeExplorer: a pipeline for virus and phage identification and abundance profiling in metagenomics data. *PeerJ* **6**, e4227 (2018).

31. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinforma. Oxf. Engl.* **29**, 1682–1684 (2013).

32. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14616–14621 (2007).

33. Luo, Y. & Qiu, J. Human parvovirus B19: a mechanistic overview of infection and DNA replication. *Future Virol.* **10**, 155–167 (2015).

34. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).

35. Kramvis, A. Genotypes and Genetic Variability of Hepatitis B Virus. *Intervirology* **57**, 141–150 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

36. Pourkarim, M. R., Amini-Bavil-Olyaee, S., Kurbanov, F., Van Ranst, M. & Tacke, F. Molecular identification of hepatitis B virus genotypes/subgenotypes: revised classification hurdles and updated resolutions. *World J. Gastroenterol.* **20**, 7152–7168 (2014).

37. Kostaki, E.-G. *et al.* Unravelling the history of hepatitis B virus genotypes A and D infection using a full-genome phylogenetic and phylogeographic approach. *eLife* **7**, (2018).

38. Pourkarim, M. R., Lemey, P., Amini-Bavil-Olyaee, S., Maes, P. & Van Ranst, M. Novel hepatitis B virus subgenotype A6 in African-Belgian patients. *J. Clin. Virol. Off. Publ. Pan Am. Soc. Clin. Virol.* **47**, 93–96 (2010).

39. Hübschen, J. M. *et al.* Phylogenetic analysis of human parvovirus b19 sequences from eleven different countries confirms the predominance of genotype 1 and suggests the spread of genotype 3b. *J. Clin. Microbiol.* **47**, 3735–3738 (2009).

40. Schneider, B. *et al.* Simultaneous persistence of multiple genome variants of human parvovirus B19. *J. Gen. Virol.* **89**, 164–176 (2008).

41. Nguyen, Q. T. *et al.* Novel Human Erythrovirus Associated with Transient Aplastic Anemia. *J. Clin. Microbiol.* **37**, 2483–2487 (1999).

42. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

43. Cheng, A. P. *et al.* A cell-free DNA metagenomic sequencing assay that integrates the host injury response to infection. *Proc. Natl. Acad. Sci.* **116**, 18738–18744 (2019).

44. Yuen, M.-F. *et al.* Hepatitis B virus infection. *Nat. Rev. Dis. Primer* **4**, 18035 (2018).

45. Broliden, K., Tolfvenstam, T. & Norbeck, O. Clinical aspects of parvovirus B19 infection. *J. Intern. Med.* **260**, 285–304 (2006).

46. Candotti, D., Etiz, N., Parsyan, A. & Allain, J.-P. Identification and Characterization of Persistent Human Erythrovirus Infection in Blood Donor Samples. *J. Virol.* **78**, 12169–12178 (2004).

47. Rinckel, L. A., Buno, B. R., Gierman, T. M. & Lee, D. C. Discovery and

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

679 analysis of a novel parvovirus B19 Genotype 3 isolate in the United States.
680 *Transfusion (Paris)* **49**, 1488–1492 (2009).
681 48. Sanabani, S., Neto, W. K., Pereira, J. & Sabino, E. C. Sequence Variability
682 of Human Erythroviruses Present in Bone Marrow of Brazilian Patients with
683 Various Parvovirus B19-Related Hematological Symptoms. *J. Clin. Microbiol.* **44**,
684 604–606 (2006).
685 49. Freitas, R. B. *et al.* Molecular characterization of human erythrovirus B19
686 strains obtained from patients with several clinical presentations in the Amazon
687 region of Brazil. *J. Clin. Virol. Off. Publ. Pan Am. Soc. Clin. Virol.* **43**, 60–65 (2008).
688 50. Jain, P. *et al.* Prevalence and genotypic characterization of human
689 parvovirus B19 in children with hemato-oncological disorders in North India. *J.*
690 *Med. Virol.* **87**, 303–309 (2015).
691 51. Servant, A. *et al.* Genetic Diversity within Human Erythroviruses:
692 Identification of Three Genotypes. *J. Virol.* **76**, 9124–9134 (2002).
693 52. Pyöriä, L. *et al.* Extinct type of human parvovirus B19 persists in tonsillar B
694 cells. *Nat. Commun.* **8**, 14930 (2017).
695 53. Valencia Pacheco, G. *et al.* Serological and molecular analysis of parvovirus
696 B19 infection in Mayan women with systemic lupus erythematosus in Mexico.
697 *Colomb. Médica* **v48**, 105–112 (2017).
698 54. Roman, S. *et al.* Occult hepatitis B in the genotype H-infected Nahuas and
699 Huichol native Mexican population. *J. Med. Virol.* **82**, 1527–1536 (2010).
700 55. Brichler, S. *et al.* African, Amerindian and European hepatitis B virus strains
701 circulate on the Caribbean Island of Martinique. *J. Gen. Virol.* **94**, 2318–2329
702 (2013).
703 56. Quintero, A. *et al.* Molecular epidemiology of hepatitis B virus in Afro-
704 Venezuelan populations. *Arch. Virol.* **147**, 1829–1836 (2002).
705 57. Andernach, I. E., Nolte, C., Pape, J. W. & Muller, C. P. Slave trade and
706 hepatitis B virus genotypes and subgenotypes in Haiti and Africa. *Emerg. Infect.*
707 *Dis.* **15**, 1222–1228 (2009).
708 58. Alvarado-Mora, M. V. *et al.* Phylogenetic analysis of complete genome
709 sequences of hepatitis B virus from an Afro-Colombian community: presence of

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

710 HBV F3/A1 recombinant strain. *Virol. J.* **9**, 244 (2012).
711 59. Mandujano-Sánchez, A., Solache, L. C. & Mandujano, M. A. Historia de las
712 Epidemias en el México Antiguo: Algunos Aspectos Biológicos y Sociales. in
713 *Ensayos sobre la historia de las epidemias en México.* (eds. Florescano, E. &
714 Malvido, E.) 9–21 (Instituto Mexicano del Seguro Social, 1982).
715 60. Bryant, J. E., Holmes, E. C. & Barrett, A. D. T. Out of Africa: A Molecular
716 Perspective on the Introduction of Yellow Fever Virus into the Americas. *PLoS*
717 *Pathog.* **3**, (2007).
718 61. Gadelha, S. R. *et al.* The origin of HTLV-1 in the South Bahia by
719 phylogenetic, mitochondrial DNA and β-globin analysis. *Retrovirology* **11**, P49
720 (2014).
721 62. Markov, P. V. *et al.* Phylogeography and molecular epidemiology of hepatitis
722 C virus genotype 2 in Africa. *J. Gen. Virol.* **90**, 2086–2096 (2009).
723 63. Forni, D. *et al.* Recent out-of-Africa migration of human herpes simplex
724 viruses. *Mol. Biol. Evol.* msaa001 (2020) doi:10.1093/molbev/msaa001.
725 64. Angel, J. L. Porotic hyperostosis, anemias, malarias, and marshes in the
726 prehistoric Eastern Mediterranean. *Science* **153**, 760–763 (1966).
727 65. Heegaard, E. D. & Brown, K. E. Human Parvovirus B19. *Clin. Microbiol.*
728 *Rev.* **15**, 485–505 (2002).
729 66. Rezaei, F. *et al.* Prevalence and genotypic characterization of Human
730 Parvovirus B19 in children with measles- and rubella-like illness in Iran. *J. Med.*
731 *Virol.* **88**, 947–953 (2016).
732 67. De Los Ángeles Ribas, M. *et al.* Identification of human parvovirus B19
733 among measles and rubella suspected patients from Cuba. *J. Med. Virol.* **91**,
734 1351–1354 (2019).
735 68. Acuna-Soto, R., Stahle, D. W., Cleaveland, M. K. & Therrell, M. D.
736 Megadrought and Megadeath in 16th Century Mexico. *Emerg. Infect. Dis.* **8**, 360–
737 362 (2002).
738 69. Warinner, C. *et al.* A Robust Framework for Microbial Archaeology. *Annu.*
739 *Rev. Genomics Hum. Genet.* **18**, 321–356 (2017).
740 70. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

741 Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl.*
742 *Acad. Sci. U. S. A.* **110**, 15758–15763 (2013).
743 71. Rohland, N. & Hofreiter, M. Ancient DNA extraction from bones and teeth.
744 *Nat. Protoc.* **2**, 1756–1762 (2007).
745 72. Meyer, M. & Kircher, M. Illumina Sequencing Library Preparation for Highly
746 Multiplexed Target Capture and Sequencing. *Cold Spring Harb. Protoc.* **2010**, 1–10
747 (2010).
748 73. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid
749 adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, (2016).
750 74. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–
751 Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
752 75. Huson, D. H. *et al.* MEGAN Community Edition - Interactive Exploration and
753 Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* **12**,
754 e1004957 (2016).
755 76. Schubert, M. *et al.* Improving ancient DNA read mapping against modern
756 reference genomes. *BMC Genomics* **13**, 178 (2012).
757 77. Ondov, B. D., Bergman, N. H. & Phillippy, A. M. Interactive metagenomic
758 visualization in a Web browser. *BMC Bioinformatics* **12**, 385 (2011).
759 78. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop
760 software platform for the organization and analysis of sequence data.
761 *Bioinformatics* **28**, 1647–1649 (2012).
762 79. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools.
763 *Bioinformatics* **25**, 2078–2079 (2009).
764 80. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for
765 large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
766 81. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more
767 models, new heuristics and high-performance computing. *Nat. Methods* **9**, 772
768 (2012).
769 82. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-
770 analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
771 83. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

772 Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973
773 (2012).
774 84. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A.
775 Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.*
776 **67**, 901–904 (2018).
777 85. Price, A. L. *et al.* Principal components analysis corrects for stratification in
778 genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
779 86. Patterson, N., Price, A. L. & Reich, D. Population Structure and
780 Eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
781 87. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of
782 contamination and endogenous mitochondrial consensus calling for ancient DNA.
783 *Genome Biol.* **16**, 224 (2015).
784 88. Kloss-Brandstätter, A. *et al.* HaploGrep: a fast and reliable algorithm for
785 automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25–32
786 (2011).
787 89. Weissensteiner, H. *et al.* HaploGrep 2: mitochondrial haplogroup
788 classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**,
789 W58–W63 (2016).
790 90. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex
791 identification of ancient human remains using DNA shotgun sequencing. *J.*
792 *Archaeol. Sci.* **40**, 4477–4482 (2013).
793

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

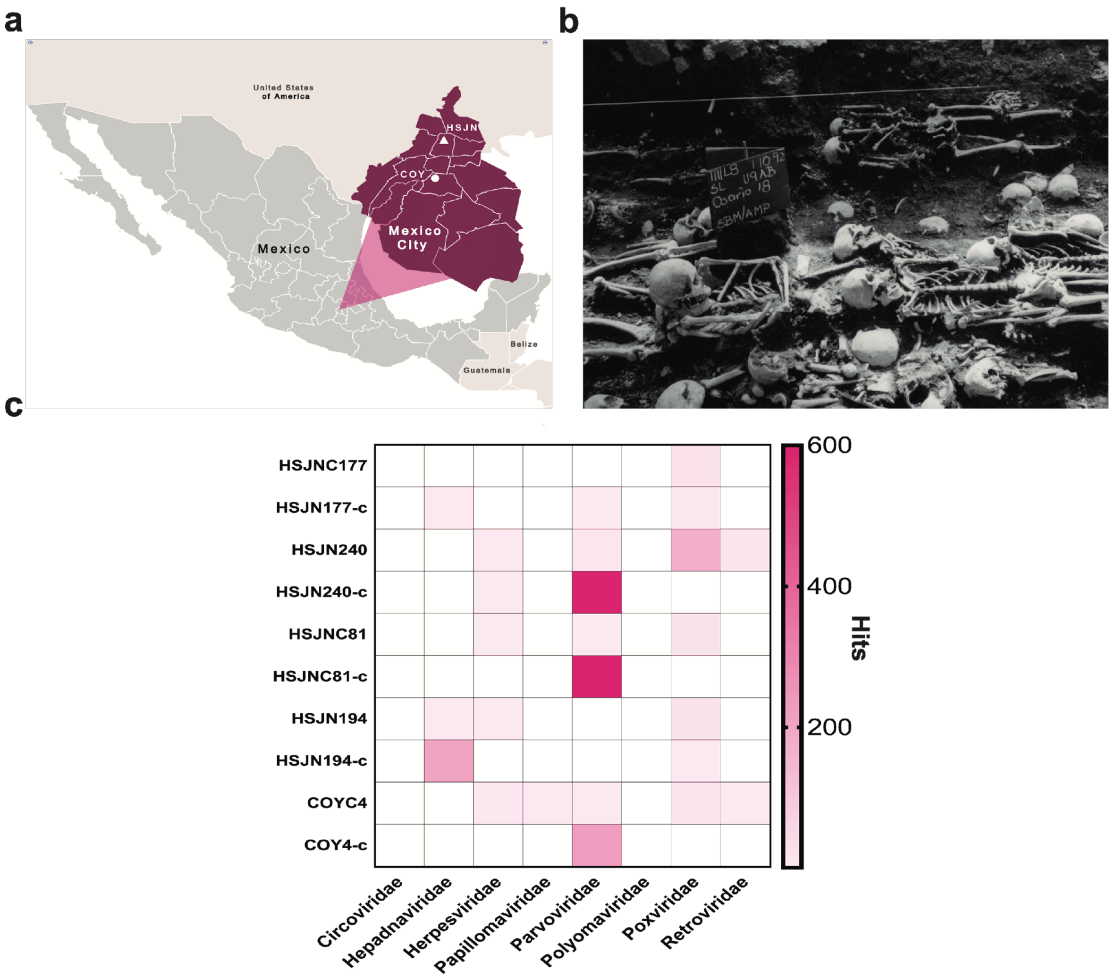


Figure 1. Metagenomic analysis of Colonial individuals reveal HBV-like and B19V-like hits. a, Location of the archeological sites used in this study, HSJN (19.431704, -99.141740) is shown as a triangle and COY (19.347079, -99.159017) as a circle, lines in pink map show current division of Mexico City. **b**, Several individuals discovered in massive burials archaeologically associated with the HSJN and colonial epidemics (photo courtesy of "Secretaria de Cultura INAH, SINAFO, Fototeca DSA"). **c**, Metagenomic analysis performed with MALT 0.4.0 based on the Viral NCBI RefSeq. Viral abundancies were compared and normalized automatically in MEGAN between shotgun (samplename) and capture (samplename-c) NGS data. Only targeted viral families are shown for HBV or B19V positive samples (all samples analysis shown in Supplementary Figure 3), a capture negative control (HSJN177) is shown as well that presented one B19V-like hit and three HBV-like hits after capture.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

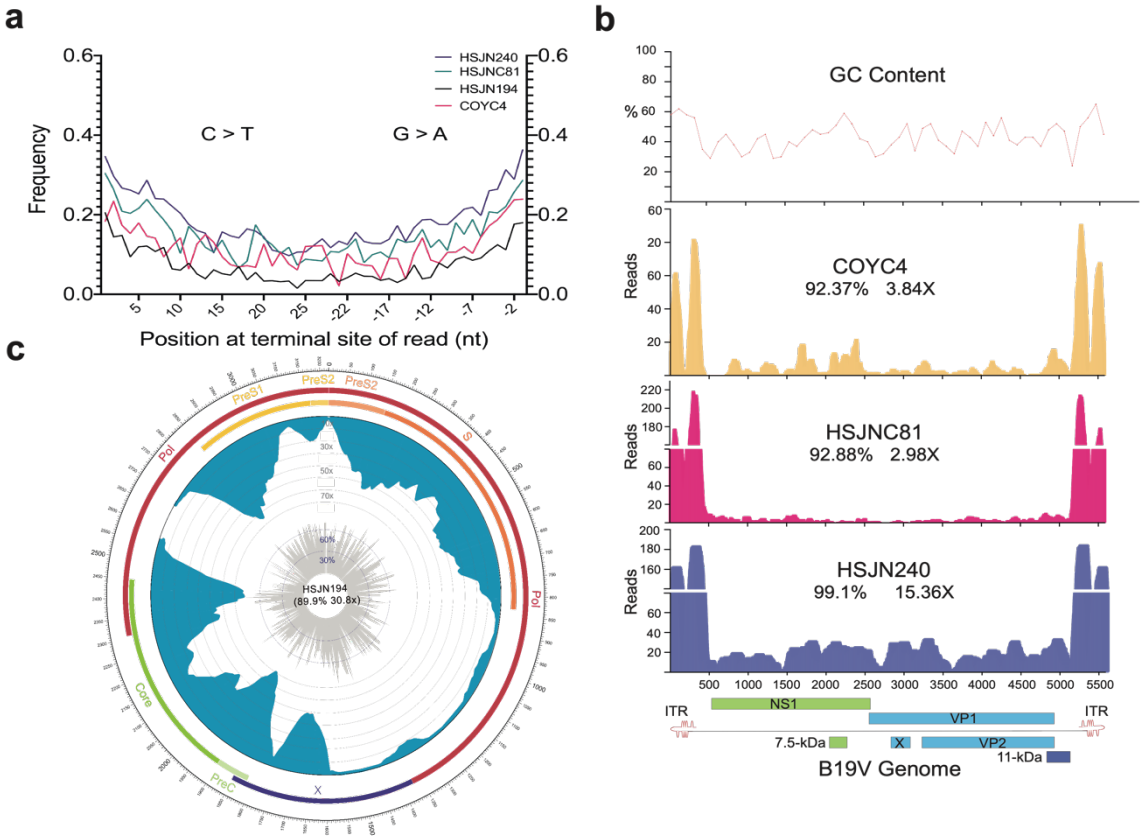


Figure 2. Ancient B19V and HBV ancient genomes. a, Superimposed damage patterns of ancient HBV (HSJN194) and B19V (HSJNC81, HSJN240, COYC4), X axis shows the position (nt) on the 5' (left) and 3' (right) end of the read, Y axis shows the damage frequency (raw damage patterns are shown on Supplementary Figure 4). **b**, B19V ssDNA linear genome, X axis shows position (nt) based on the reference genome (AB550331), and Y axis shows depth (as number of reads), GC content is shown as a percentage of each 100 bp windows, CDS coverage and average depth are shown under each individual ID. Schematic of the B19V genome is shown at the bottom. Highly covered regions correspond to dsDNA ITRs shown as a red hairpin. **c**, HBV circular genome, outer numbers show position (nt) based on reference genome (GQ331046), outer bars show genes with names, blue bars represent coverage and gray bars shows GC content each 10 bp windows. Coverage and average depth are shown in the center. Low covered region between S and X overlaps with ssDNA region.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.05.137083; this version posted June 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

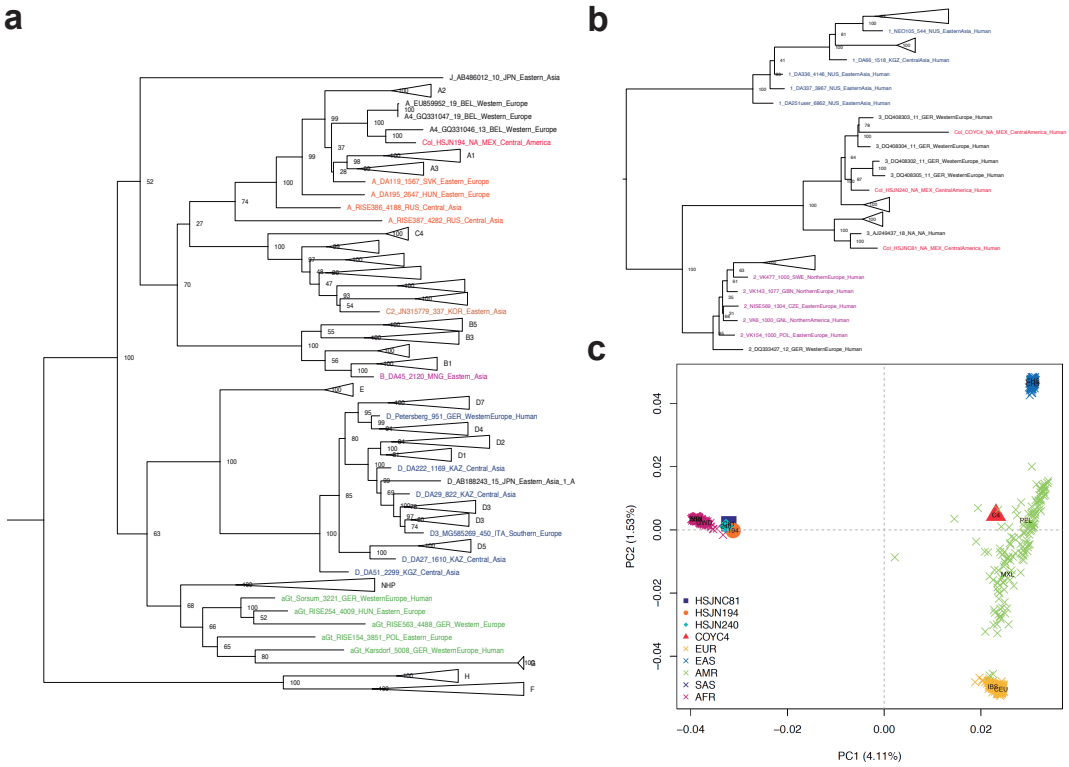


Figure 3. Viral Colonial genomes and their human hosts cluster with modern African genetic diversity. Maximum likelihood tree performed on RAxML 8.2.10 (1000 bootstraps) with a midpoint root based on the HBV whole genome (**a**) and B19V CDS (**b**), previously reported ancient genomes are colored, Colonial samples are shown in red. Sequences are named as follows: genotype_ID_sampling.year_country.of.origin_area.of.origin_host. HBV genotype nomenclature is based on letters, while for B19V is on numbers. **c**, PCA showing genetic affinities of ancient human hosts compared to 1000 Genomes Project reference panel, “x” show individuals from the reference panel while other shapes show human hosts from which ancient HBV (HSJN194) and B19V (HSJNC81, HSJN240, COYC4) were recovered.