UAB
Universitat Autònoma de Barcelona

Unraveling the Genetic History of Northeastern Iberians

Daniel Ruiz de la Cuesta Aguirre

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  https://creativecommons.org/licenses/?lang=ca

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  https://creativecommons.org/licenses/?lang=es

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  https://creativecommons.org/licenses/?lang=en



Unraveling the Genetic History of Northeastern Iberians

Daniel Ruiz de la Cuesta Aguirre

Doctoral thesis

2025

Supervisors:

Assumpció Malgosa Morera

Cristina Maria Pereira dos Santos

Universitat Autònoma de Barcelona

PhD in Biodiversity

Unraveling the Genetic History of Northeastern Iberians

PhD dissertation presented by Daniel Ruiz de la Cuesta Aguirre in fulfillment of the requirements for the Doctorate in Biodiversity of Departament de Biologia Animal, de Biologia Vegetal i d'Ecologia, Universitat Autònoma de Barcelona, supervised by:

Dra. Cristina Maria Pereira dos Santos, Associate Professor at Unitat d'Antropologia Biològica, Departament de Biologia Animal, de Biologia Vegetal i d'Ecologia, Universitat Autònoma de Barcelona.

Dra. Assumpció Malgosa, Full Professor at Unitat d'Antropologia Biològica, Departament de Biologia Animal, de Biologia Vegetal i d'Ecologia, Universitat Autònoma de Barcelona.

Dra. Cristina Maria Pereira dos Santos

Dra. Assumpció Malgosa Morera

Daniel Ruiz de la Cuesta Aguirre

A mi madre

This research was supported by the Ministerio de Ciencia e Innovación, Gobierno de España (PGC2018-096666-B-I00).

It was also co-funded by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR-FI), Departament de Recerca i Universitats, Generalitat de Catalunya, and the European Social Fund (Fons Social Europeu) under the grant 2020 FI_B1 00641.

Additionally, it was founded by the Fundación Palarq (PALARQ23_55).

ACKNOWLEDGES / AGRADECIMIENTOS

Esta tesis han 5 años, 4 artículos, 3 meses en Australia, una pandemia, un ataque informático y un apagón, entre otras cosas. No está mal, ¿no? Además, también han sido un montón de momentos maravillosos y de gente increíble. Seguramente me dejaré a alguien, así que, si crees que, directa o indirectamente, has aportado a esta tesis, este agradecimiento también es para ti.

En primer lugar, gracias a mi madre, por escuchar mil veces cómo le explicaba lo que hacía. A mi padre, por sentir siempre curiosidad y preguntarme todas las dudas hasta entenderlo. A mi hermano, por apoyarme a pesar de no haber seguido el camino fácil. En general, a toda mi familia (y familia política).

A mi gente de Logroño: la de siempre, aquella con la que me he reencontrado durante estos años, y a la nueva. A mis amigas y amigos de Barcelona: a las "Terrates", a "El grupito del vino" / "el montón bueno" / "bé sé ene", a "Las Tartitas", al "Dr. Dr.", a "La isla". A todas mis compañeras de piso durante estos años de tesis, por aguantar mi agotamiento y mis chapadas sobre lo bien que había ido el día o cómo habíamos detectado una contaminación en el lab.

I would also like to thank the ACAD team (and the former ACAD members) who made my stay in Adelaide a thousand times better. Thanks to the people who sheltered me during those three months and helped me overcome my fear of cats and dogs, many people thank you for that. And especially, thank you to Bastien, who was the one that made this dream come true.

Gracias a todas y todos los miembros del GREAB. Gracias también a todas las profesoras (y profesores) del departamento, tanto las que han estado desde el principio como a las que han llegado al final. Gracias también a los miembros del tribunal de seguimiento. Gracias a Jéssica y Nuria, por toda vuestra ayuda. Gracias a postdocs, predocs y alumnos/as de máster y grado, que directa o indirectamente me ayudasteis durante este tiempo. Gracias a Diana, por introducirme en el laboratorio. Gracias a los/as alumnos/as de máster que he co-supervisado, porque también me habéis ayudado a aprender a ser un mejor investigador. Gracias a todas las personas involucradas en los cuatro artículos que conforman esta tesis, en especial a Cristina, Xavi y Rosa.

Y ya acabando, gracias a la siguiente generación de predocs, por esas comidas con sobremesa tan largas y por animarme siempre. Seguiré entrando en pánico a la mínima y hablando solo, eso no lo dudéis.

Por último, pero no menos importante, gracias a Assumpció y a Cristina, por darme esta oportunidad, por enseñarme cómo ser investigador, por estar cada día ahí y por toda vuestra ayuda en general.

P.D. Gracias también a Spotify y a los cantantes (A, AM, AS, CR, TS, etc.) que me ha acompañado durante horas y horas en los últimos 5 años.

ABSTRACT

The study of ancient DNA (aDNA) has transformed the understanding of human evolutionary history, enabling the direct analysis of genetic material preserved in archaeological remains. However, aDNA studies face substantial challenges due to the scarcity, fragmentation, and chemical modifications of ancient molecules, as well as contamination from modern sources. In the Iberian culture (Iron Age), additional obstacles arise from the predominant funerary practice of cremation, limiting the availability of preserved biological material. Consequently, despite extensive archaeological research, genetic studies on Iberian communities have remained sparse, fragmentary, and geographically biased.

This thesis addresses these gaps by combining methodological innovations with a focused paleogenomic investigation of Iron Age individuals in northeastern Iberian settlements. By leveraging the recent advances in paleogenomics, this work provides novel insights into both technical aspects of ancient DNA analysis and the demographic history of Iberian populations from the Early Iron Age to the Roman Conquest period.

The first part of the thesis is dedicated to developing and validating new methodological tools to enhance ancient mitochondrial DNA (mtDNA) analysis. A comprehensive *post-mortem* molecular damage profile was generated by compiling 427 ancient mtDNA sequences obtained through shotgun and capture techniques and sequenced with Next Generation Sequencing (NGS) technologies. This study identified a total of 2,933 molecular damage hotspots, of which 2,781 had not been previously reported as mutational, heteroplasmic, or damage-prone sites. While the influence of sequencing errors cannot be entirely ruled out, the profile provides a critical reference framework for distinguishing damage from mutational hotspots or heteroplasmic positions, thereby facilitating more accurate downstream genetic interpretations.

Furthermore, a user-friendly bioinformatic pipeline was developed for the analysis of mtDNA sequences generated via amplicon-based NGS methods, which are particularly suited for highly degraded samples. The pipeline automates key analytical steps, including contamination estimation, haplogroup assignment, and haplotype determination. Critical parameters were optimized, notably the implementation of fastp v0.23.2 for duplicate removal and a PMDS threshold of 1 using PMDtools v0.23.2, validated as effective cutoffs for authenticating aDNA even when samples display as little as 9.17% damaged reads.

The second part of the thesis focuses on the paleogenomic analysis of Iberian newborns from six key archaeological sites: the fortress of Vilars, Ullastret (Illa d'en Reixac and Puig de Sant Andreu), Sant Miquel d'Olèrdola, Castell de Besora, el poblat Ibèric de Burriac, and El Camp de les Lloses. Genomic data was recovered for 22 newborns, and mitochondrial DNA was reported for a total of 50 Iberian newborns, representing different Iberian tribes such as the Ilergetae, Indiketes, Cessetani, Layetani, Bergistani, and Ausetani. These individuals, whose burial contexts align with broader Iberian domestic rituals, serve as a valuable proxy for reconstructing the genetic structure of Iron Age northeastern Iberia.

The nuclear genomic data revealed that Iberian communities largely evolved from local Bronze Age populations, undergoing gradual changes rather than abrupt population replacement during the Iron Age. However, signatures of gene flow from Central/Eastern Mediterranean and North African sources were detected, highlighting the impact of trade, migration, and cultural contacts. The genetic landscape remained predominantly local, but external influences played a significant role in shaping Iberian diversity. The study of individuals from El Camp de les Lloses, dating to the Late Iron Age and early Roman period, illustrated how Roman expansion introduced increased Mediterranean genetic influences into the Iberian Peninsula.

The mitochondrial DNA analysis of 73 individuals, including previously published ancient Iberian mitochondrial DNA profiles, revealed subtle differences in haplogroup frequencies among tribes, although overall genetic differentiation remained absent. The relatively high mtDNA diversity observed across all groups supported the hypothesis of patrilocal mating systems and small-scale female mobility, which would have prevented important differentiation between tribes. Most detected mtDNA haplogroups had ancestral roots in the Iberian Peninsula before the Iron Age, confirming local maternal continuity. Nevertheless, the presence of distinct haplogroups linked to Near Eastern and North African populations provides strong support for sporadic long-distance female migration into Iberian communities during this period.

In summary, this thesis contributes substantially to the field of paleogenomics by addressing critical methodological challenges in ancient DNA analysis and providing the first systematic genomic characterization of northeastern Iberian communities during the Iron Age. By combining nuclear and mitochondrial data, and integrating archaeological and historical contexts, it reconstructs the complex demographic processes that shaped ancient Iberian societies. It emphasizes the importance of local continuity, the nuanced impact of

external interactions, and the role of women in mediating cultural and genetic exchanges. The methodological tools and findings presented herein lay the groundwork for future studies on poorly preserved or underrepresented ancient populations worldwide.

RESUMEN

El estudio del ADN antiguo ha transformado la comprensión de la historia evolutiva humana, permitiendo el análisis directo del material genético preservado en restos arqueológicos. Sin embargo, los estudios de ADN antiguo se enfrentan a importantes desafíos debido a la escasez, fragmentación y modificaciones químicas de las moléculas antiguas, así como a la contaminación por fuentes modernas. En la cultura ibera (Edad del Hierro), existen obstáculos adicionales debido a la práctica funeraria predominante de la cremación, que limita la disponibilidad de material biológico preservado. En consecuencia, a pesar de la extensa investigación arqueológica, los estudios genéticos sobre los iberos han sido escasos, fragmentarios y geográficamente sesgados.

Esta tesis aborda estas lagunas combinando innovaciones metodológicas con una investigación paleogenómica centrada en individuos perinatales procedentes de enterramientos realizados en poblados íberos del noreste peninsular. Aprovechando los recientes avances en paleogenómica, este trabajo aporta nuevas perspectivas tanto en aspectos técnicos del análisis de ADN antiguo como en la historia demográfica de las poblaciones iberas desde la Primera Edad del Hierro hasta el periodo de la conquista romana.

La primera parte de la tesis se dedica al desarrollo y validación de nuevas herramientas metodológicas para mejorar el análisis del ADN mitocondrial (mtDNA) antiguo. Se generó un perfil completo de daño molecular *post-mortem* mediante la compilación de 427 secuencias antiguas de mtDNA obtenidas mediante técnicas de captura y *shotgun* y secuenciadas con tecnologías de nueva generación (NGS). Este estudio identificó un total de 2.933 *hotspot* de daño molecular, de los cuales 2.781 no habían sido previamente reportados como posiciones altamente mutadas, heteroplásmicas o propensas al daño. Aunque no puede descartarse completamente la influencia de errores de secuenciación, el perfil proporciona un marco de referencia crucial para distinguir daño de mutaciones o heteroplasmia, facilitando interpretaciones genéticas más precisas.

Además, se desarrolló una *pipeline* bioinformática de uso sencillo para el análisis de secuencias de mtDNA generadas mediante métodos de NGS basados en amplicones, especialmente adecuados para muestras altamente degradadas. La *pipeline* automatiza pasos clave como la estimación de contaminación, determinación de haplotipos y asignación de haplogrupos. Se optimizaron parámetros críticos, como la implementación de fastp v0.23.2 para la eliminación de duplicados y un umbral de PMDS de 1 utilizando

PMDtools v0.23.2, validado como un punto de corte efectivo para autenticar ADN antiguo incluso en muestras con apenas un 9,17% de lecturas dañadas.

La segunda parte de la tesis se centra en el análisis paleogenómico de individuos perinatales iberos de seis yacimientos arqueológicos clave: la fortaleza de Els Vilars d'Arbeca, Ullastret (Illa d'en Reixac y Puig de Sant Andreu), Sant Miquel d'Olèrdola, el Castell de Besora, el poblado ibero de Burriac, y El Camp de les Lloses. Se recuperaron datos genómicos de 22 perinatales iberos, y se reportaron datos de mtDNA de un total de 50 perinatales iberos, representando diferentes tribus como los ilergetes, indiketes, cesetanos, layetanos, bergistanos y ausetanos. Estos individuos, cuyos contextos funerarios se alinean con rituales domésticos iberos, constituyen un valioso proxy para reconstruir la estructura genética del noreste ibero en la Edad del Hierro.

Los datos genómicos nucleares revelaron que las comunidades iberas evolucionaron en gran medida a partir de poblaciones locales de la Edad del Bronce, experimentando cambios graduales más que reemplazos poblacionales abruptos durante la Edad del Hierro. Sin embargo, se detectaron señales de flujo génico procedente del Mediterráneo Central/Oriental y de fuentes norteafricanas, lo que refleja el impacto del comercio, la migración y los contactos culturales. Aunque el paisaje genético se mantuvo mayoritariamente local, las influencias externas desempeñaron un papel significativo en la diversidad ibera. El estudio de individuos de El Camp de les Lloses, fechados en la fase final de la Edad del Hierro y principios del periodo romano, ilustró cómo la expansión romana introdujo un aumento de influencias genéticas mediterráneas en la península ibera.

El análisis de mtDNA de 73 iberos, incluyendo perfiles de mtDNA de individuos previamente publicados, reveló diferencias sutiles en la frecuencia de haplogrupos entre tribus, aunque en general no se observó diferenciación genética significativa. La alta diversidad de mtDNA observada en todas las tribus apoya la hipótesis de sistemas de apareamiento patrilocales y movilidad femenina a pequeña escala, que habrían impedido una diferenciación importante entre tribus. La mayoría de los haplogrupos detectados tienen raíces ancestrales en la península ibera antes de la Edad del Hierro, confirmando la continuidad materna local. No obstante, la presencia de haplogrupos vinculados a poblaciones del este del Mediterráneo y el norte de África respalda la existencia de migraciones femeninas esporádicas de larga distancia hacia las comunidades iberas.

En resumen, esta tesis contribuye de manera sustancial al campo de la paleogenómica al abordar desafíos metodológicos críticos en el análisis de ADN antiguo y

ofrecer la primera caracterización genómica sistemática de comunidades iberas del noreste durante la Edad del Hierro. Al combinar datos nucleares y mitocondriales, e integrar contextos arqueológicos e históricos, reconstruye los procesos demográficos complejos que moldearon las sociedades iberas antiguas. Destaca la importancia de la continuidad local, el impacto matizado de las interacciones externas y el papel de las mujeres en la mediación de los intercambios culturales y genéticos. Las herramientas metodológicas y hallazgos presentados aquí sientan las bases para futuros estudios sobre poblaciones antiguas pobremente preservadas o subrepresentadas a nivel mundial.

RESUM

L'estudi de l'ADN antic ha transformat la comprensió de la història evolutiva humana, permetent l'anàlisi directa del material genètic conservat en restes arqueològiques. No obstant això, els estudis d'ADN antic s'enfronten a reptes importants a causa de l'escassetat, la fragmentació i les modificacions químiques de les molècules antigues, així com de la contaminació procedent de fonts modernes. En la cultura ibèrica (Edat del Ferro), es donen obstacles addicionals degut a que la pràctica funerària predominant, la cremació, limita la disponibilitat de material biològic conservat. En conseqüència, malgrat la recerca arqueològica extensa, els estudis genètics sobre les comunitats ibèriques han estat escassos, fragmentaris i geogràficament esbiaixats.

Aquesta tesi aborda aquestes mancances combinant innovacions metodològiques amb una investigació paleogenòmica centrada en individus perinatals enterrats en els poblats ibèrics del nord-est de la península Ibèrica. Aprofitant els avenços recents en paleogenòmica, aquest treball aporta noves perspectives tant en aspectes tècnics de l'anàlisi d'ADN antic com en la història demogràfica de les poblacions ibèriques des de la Primera Edat del Ferro fins al període de la conquesta romana.

La primera part de la tesi es dedica al desenvolupament i validació de noves eines metodològiques per a millorar l'anàlisi de l'ADN mitocondrial (mtDNA) antic. Es va generar un perfil complet de dany molecular *post-mortem* mitjançant la recopilació de 427 seqüències antigues de mtDNA obtingudes mitjançant tècniques de captura i *shotgun* i seqüenciades amb tecnologies de nova generació (NGS). Aquest estudi va identificar un total de 2.933 *hotspots* de dany molecular, dels quals 2.781 no havien estat prèviament reportats com a posicions altament mutacionals, heteroplàsmics o propenses al dany. Tot i que no es pot descartar completament la influència d'errors de seqüenciació, el perfil proporciona un marc de referència crític per a distingir el dany de les mutacions o les heteroplàsmies, facilitant interpretacions genètiques més precises.

A més, es va desenvolupar una pipeline bioinformàtica fàcil d'utilitzar per a l'anàlisi de seqüències de mtDNA generades mitjançant mètodes NGS basats en amplicons, especialment adequats per a mostres altament degradades. La pipeline automatitza passos clau com l'estimació de contaminació, la determinació d'haplotips i l'assignació d'haplogrups. Es van optimitzar paràmetres crítics, com la implementació de fastp v0.23.2 per a l'eliminació de duplicats i un llindar de PMDS d'1 mitjançant PMDtools v0.23.2, validat

com a punt de tall efectiu per autenticar aDNA fins i tot en mostres amb només un 9,17% de lectures danyades.

La segona part de la tesi se centra en l'anàlisi paleogenòmica de perinatals ibèrics procedents de sis jaciments arqueològics clau: la fortalesa dels Vilars, Ullastret (Illa d'en Reixac i Puig de Sant Andreu), Sant Miquel d'Olèrdola, el Castell de Besora, el poblat ibèric de Burriac, i El Camp de les Lloses. Es van recuperar dades genòmiques de 22 perinatals ibèrics i es van reportar dades de mtDNA d'un total de 50 perinatals ibèrics, representant diverses tribus com els ilergets, indikets, cesetans, layetans, bergistans i ausetans. Aquests individus, els contextos funeraris dels quals s'alineen amb rituals domèstics ibèrics, constitueixen un valuós model per reconstruir l'estructura genètica del nord-est de la península durant l'Edat del Ferro.

Les dades genòmiques nuclears van revelar que les comunitats ibèriques van evolucionar en gran mesura a partir de poblacions locals de l'Edat del Bronze, experimentant canvis graduals en lloc de reemplaçaments poblacionals sobtats durant l'Edat del Ferro. Tanmateix, es van detectar senyals de flux genètic procedent del Mediterrani Central/Oriental i de fonts nord-africanes, reflectint l'impacte del comerç, la migració i els contactes culturals. Tot i que el paisatge genètic es va mantenir predominantment local, les influències externes van jugar un paper significatiu en la diversitat ibèrica. L'estudi d'individus d'El Camp de les Lloses, datats a la fase final de l'Edat del Ferro i primers temps del període romà, va il·lustrar com l'expansió romana va introduir un augment de les influències genètiques mediterrànies a la península.

L'anàlisi del mtDNA de 73 ibers, incloent perfils de mtDNA d'individus prèviament publicats, va revelar diferències subtils en les freqüències d'haplogrups entre tribus, tot i que en general no es va observar diferenciació genètica significativa. L'elevada diversitat de mtDNA observada en tots els grups dona suport a la hipòtesi de sistemes de matrimoni patrilocals i mobilitat femenina a petita escala, que haurien evitat una diferenciació significativa entre tribus. La majoria dels haplogrups detectats tenen arrels ancestrals a la península abans de l'Edat del Ferro, confirmant la continuïtat materna local. Tot i així, la presència d'haplogrups vinculats a l'est del Mediterrani i al nord d'Àfrica reforça la idea de migracions femenines esporàdiques de llarga distància cap a les comunitats ibèriques.

En resum, aquesta tesi contribueix de manera substancial al camp de la paleogenòmica abordant reptes metodològics crítics en l'anàlisi d'ADN antic i oferint la primera caracterització genòmica sistemàtica de comunitats ibèriques del nord-est durant

l'Edat del Ferro. Combinant dades nuclears i mitocondrials i integrant contextos arqueològics i històrics, reconstrueix els processos demogràfics complexos que van modelar les societats ibèriques antigues. També, subratlla la importància de la continuïtat local, l'impacte matisat de les interaccions externes, i el paper de les dones en la mediació dels intercanvis culturals i genètics. Les eines metodològiques i els resultats presentats en aquest treball estableixen les bases per a futurs estudis sobre poblacions antigues poc conservades o poc representades a escala mundial.

INDEX

1. **INTRODUCTION** **1**
   - 1.1. The study of ancient DNA: from paleogenetics to paleogenomics 3
   - 1.2. Bias in the ancient human genomes published 7
   - 1.3. From bones to genotypes: how to study ancient genomes 10
   - 1.4. Advantages of mitochondrial DNA (mtDNA) when working with aDNA 14
   - 1.5. Overview of paleogenomic studies in the Iberian Peninsula 17
   - 1.6. History of the Western Mediterranean and the Iberian territories during the Iron Age 19
   - 1.7. Iberian Archaeological sites analyzed in this thesis 25

2. **THE RATIONALE AND OBJECTIVES OF THE THESIS** **31**
   - 2.1. The rationale of the thesis 33
   - 2.2. Objectives of the thesis 34

3. **MATERIAL AND METHODS** **35**

4. **RESULTS** **39**
   - 4.1. Article 1: A *Post-Mortem* Molecular Damage Profile in the Ancient Human Mitochondrial DNA 43
   - 4.2. Article 2: An easy-to-use pipeline to analyze amplicon-based Next Generation Sequencing results of human mitochondrial DNA from degraded samples 59
   - 4.3. Article 3: The Genetic Landscape of Northeastern Iberian Communities from the Early to Late Iron Age 81
   - 4.4. Article 4: Mitochondrial DNA diversity in Northeast Iberians during the Iron Age ... 151

5. **GENERAL DISCUSSION** **199**
   - 5.1. Development of New Tools for Ancient Mitochondrial DNA Analysis 202
   - 5.2. Genetic Analysis of Northeast Iberian Communities 208
   - 5.3. Limitations 218

6. **CONCLUSIONS** **221**

7. **BIBLIOGRAPHY** **225**

8. **SUPPLEMENTARY MATERIAL** **271**



1.

INTRODUCTION

1. INTRODUCTION

1.1. The study of ancient DNA: from paleogenetics to paleogenomics

The study of ancient DNA (aDNA) — DNA extracted from preserved or fossilized biological materials, such as bones, teeth, mummified tissues, and coprolites (Brown & Brown, 1994) — faces significant challenges inherent to its nature. The aDNA is characterized by its scarcity, fragmentation, and chemical modifications that accumulate over time. These characteristics pose substantial obstacles to extraction, sequencing, and analysis, requiring specialized techniques and careful interpretation of results (Dabney et al., 2013a; Orlando et al., 2021). In fact, DNA extracts from ancient samples are often complex mixtures containing bacterial, fungal, and modern human contaminants, leaving the endogenous DNA — the genuine genetic material from the ancient specimen — at levels typically ranging from 1% to 10%. Higher proportions are occasionally observed in exceptionally well-preserved samples, such as those found in permafrost or dry caves (Ovchinnikov et al., 2000; Schwarz et al., 2009; van der Valk et al., 2021).

Post-mortem damage is a primary concern in aDNA research. DNA molecules have limited chemical stability and are susceptible to damage from hydrolysis, oxidation, and nonenzymatic methylation. In living cells, enzymatic repair mechanisms continuously maintain genomic integrity by correcting damage (Lindahl, 1993a). However, when an organism dies, these repair processes cease, leaving the genome vulnerable to degradation due to intracellular nucleases, microbial activity, and environmental factors. The *post-mortem* damage manifests in various forms, including extensive fragmentation, resulting in DNA molecules typically ranging from 40 to 500 base pairs in length (Pääbo, 1989). Particularly in ancient skeletal remains, the average size of aDNA molecules can be shorter

than 60 bp (Glocke & Meyer, 2017; Noonan et al., 2006). Abasic sites, where nucleobases are lost while the sugar-phosphate backbone remains intact — most commonly affecting purines: Adenine (A) and Guanine (G) — and single-strand breaks further contribute to DNA degradation (Lindahl, 1993a). Different chemical modifications can occur. However, the most prevalent is the one generated by the hydrolytic deamination of cytosine (C) to uracil (U) (Brotherton et al., 2007; Noonan et al., 2006; Pääbo, 1989; Stiller et al., 2006). Deamination of C to U causes DNA polymerases to incorporate an A as a complementary base to the U, which turns into thymine (T) during replication, causing C-to-T substitution (Briggs et al., 2007; Dabney et al., 2013a; Meyer et al., 2012). This type of damage tends to accumulate at the 5' ends of molecules, while a reverse modification — G-to-A changes — commonly occurs at the 3' ends (Briggs et al., 2007). These modifications decrease exponentially from the ends along the DNA molecule (Briggs et al., 2007). These damage patterns can introduce biases in sequence alignment, complicate variant calling, and potentially lead to misinterpretations of genetic diversity and phylogenetic relationships (Briggs et al., 2007; Dabney et al., 2013b; Dalén et al., 2023).

The study of aDNA emerged in the 1980s with the advent of DNA sequencing technologies, marking the beginning of the paleogenetics field. The first breakthrough occurred in 1984 when Higuchi and colleagues extracted and sequenced 229 bp of mitochondrial DNA (mtDNA) from a dried muscle of a 140-year-old specimen of the extinct quagga preserved in salt at the South African Museum in Cape Town, demonstrating that genetic material could survive in ancient remains (Higuchi et al.,, 1984). Shortly after, Svante Pääbo's pioneering work on mummified human specimens from around 3,000-2,400 BCE, confirmed that DNA could be retrieved from samples thousands of years old (Pääbo, 1985a, 1985b, 1986). The development of the Polymerase Chain Reaction (PCR) in 1987 revolutionized the field, allowing the amplification of small amounts of aDNA (Mullis & Faloona, 1987; Raxworthy & Smith, 2021; Saiki et al., 1988). This technique quickly substituted cloning amplification previously used, since PCR was faster, more sensitive, and avoided the error-prone DNA repair mechanisms inherent in cloning, making it a more reliable method for analyzing ancient genetic material (Pääbo & Wilson, 1988).

The use of PCR in aDNA studies revealed the challenge of working with ancient human samples due to modern human contamination, as noted by Svante Pääbo (Gitschier, 2008). Despite this, the technique facilitated the retrieval of aDNA from various ancient species, including maize (Rollo et al., 1988), the Tasmanian wolf (Thomas et al., 1989), the New Zealand moa (Cooper et al., 1992), and the woolly mammoth (Hagelberg et al., 1994). In the

meantime, in 1990, a silica-based DNA extraction method was developed for modern DNA. This method used guanidine hydrochloride salt to facilitate DNA binding to silica particles, improving the yield of extracted DNA (Boom et al., 1990). This approach was later adapted for aDNA, enhancing DNA recovery from ancient samples (Höss & Pääbo, 1993).

Significant achievements came in the 1990s when Handt et al. (1994) successfully extracted DNA from the intestines of Ötzi the Tyrolean Iceman, a 5,300-year-old mummy found in the Alps, identifying a unique mitochondrial subhaplogroup of K absent in modern populations (Handt et al., 1994), and few years later, Krings et al. (1997) reconstructed the first complete mtDNA sequence from *Homo neanderthalensis*, providing the first genetic insight into extinct hominins (Krings et al., 1997). Also in 1998, Poinar et al. successfully amplified DNA from a coprolite of an extinct ground sloth, revealing information about its diet (Poinar et al., 1998). During the same years several ambitious studies claimed to retrieve DNA from million-year-old specimens, including fossil plants, such as a 17-20 million-year-old Magnolia (Golenberg et al., 1990), insects in amber, including 25-40 million-year-old bees (Cano et al., 1992), a 120-135-million-year-old weevil (Cano et al., 1993), and 25-30 million-year-old termites (DeSalle et al., 1992, 1993), and even an 80-million-years old dinosaur (Woodward et al., 1994). However, these claims faced increasing skepticism due to the inability of other research groups to replicate the results (Austin et al., 1997; Walden & Robertson, 1997). Doubts intensified when a phylogenetic tree revealed that the "dinosaur" DNA was more closely related to humans than to birds or reptiles, introducing further doubt on the authenticity of these findings (Hedges & Schweitzer, 1995). Moreover, chemical analyses of DNA degradation processes suggested that genetic material could not survive for more than a few thousand years, let alone millions (Lindahl, 1993a, 1993b; Pääbo & Wilson, 1991). These challenges highlighted the importance of establishing rigorous standards. In 2000, Cooper and Poinar formally stressed the importance of strict laboratory protocols to ensure the authenticity of aDNA sequences, establishing criteria to avoid contamination and false positives (Cooper & Poinar, 2000). This call for accurate protocols marked a turning point in aDNA research, leading to more reliable and reproducible studies. These included a physically isolated work area, environmental and extraction controls (avoiding modern DNA as positive control), independent replication, etc. (Cooper & Poinar, 2000). In 2003, the first analysis of ancient environmental DNA was performed using 400,000 to 10,000-year-old permafrost and temperate sediments, revealing extinct plants and megafauna, and demonstrating the potential of environmental aDNA in reconstructing past ecosystems (Willerslev et al., 2003).

The transition from paleogenetics to paleogenomics was marked by the introduction of Next-Generation Sequencing (NGS) technologies. This shift made sequencing of entire genomes from ancient specimens faster and cheaper. The first paleogenomic study that applied NGS technologies involved sequencing 28 million bp, from which 13 million bp were identified as endogenous, from a woolly mammoth's mandible (Poinar et al., 2006). An in-silica column-based aDNA extraction method that combined high DNA yields with convenient handling was developed in 2007 (Rohland & Hofreiter, 2007), making it easier to recover aDNA from archaeological samples. 2010 was a pivotal year for paleogenomics. The first ancient human genome was sequenced from a 4,000-year-old Palaeo-Eskimo individual, successfully recovering 79% (2.4 billion bp) of the diploid genome with an average depth of coverage of approximately 20x (Rasmussen et al., 2010). The first draft of the complete Neanderthal genome, a composite of three individuals, with an average of 1.3x depth coverage of more than 4 billion bp was published (Green et al., 2010). Moreover, researchers identified a genetically distinct hominin group related to Neanderthals, known as the Denisovans, with an average coverage of around 1.9x (Reich et al., 2010). This year also saw the practical application of in-solution enrichment in aDNA, a method involving the usage of artificially synthesized oligonucleotides as baits to hybridize with complementary sequences in a DNA library, allowing for the selective enrichment of targeted regions. This technique was applied to capture 14,000 bp of Neanderthal DNA from a poorly preserved sample with low endogenous DNA (Burbano et al., 2010). The first ancient pathogen genome was presented in 2011, and was the genome of *Yersinia pestis*, obtained from London victims of the Black Death, dating between 1348 and 1350 (Bos et al., 2011). In 2012, the first high-quality ancient genome with 30x coverage was achieved using a single-stranded DNA library preparation technique on a Denisovan sample (Meyer et al., 2012). One year later, a method that allowed for the recovery of ultrashort DNA fragments (50 bp) using silica columns was developed (Dabney et al., 2013b). In 2015, Haak et al. (2015) used a panel targeting 390,577 single nucleotide polymorphisms (SNPs) across the human genome designed to capture informative sites relevant to studying human population dynamics and genetic variation. Most of the SNPs were from the Affymetrix Human Origins array, which includes around 629,000 SNPs selected to provide a clean and well-documented set of genetic markers for population inference, helping minimize biases that can occur with SNPs selected for genome-wide association studies. Some months later, Fu et al. (2015) captured samples with different panels, one composed of 842,630 SNPs, covering all SNPs from the Illumina 610-Quad array, all SNPs from the Affymetrix 50k array, and some other SNPs. By the end of that

year, the two arrays had been merged to generate the '1240k panel', which included a total of 1,237,207 SNPs: nearly all SNPs from the Affymetrix Human Origins and Illumina 610-Quad arrays, as well as SNPs on chromosomes X and Y, and around 47,000 functional SNPs (Mathieson et al., 2015). However, few laboratories had the knowledge, ability, and economic power to apply this technique. In 2019, a DNA extraction protocol was developed for recovering highly fragmented molecules (as short as 35 base pairs long) preserved within bones and teeth (Dabney & Meyer, 2019). This method employed an EDTA/Proteinase K lysis buffer and spin columns with silica membranes for DNA purification (Dabney & Meyer, 2019). In the same year, the genome of a 5,700-year-old female and her oral microbiome was recovered from a chewed birch pitch (Jensen et al., 2019).

An enormous jump in the field was made in 2021 when researchers sequenced DNA from mammoth molars in Siberia, dating back over 1 million years (van der Valk et al., 2021). The same year, the true revolution in human paleogenomics arrived when two biotechnology companies (Daicel Arbor Biosciences and Twist Bioscience) released a commercial in-solution enrichment kit targeting the same set of 1240k SNPs as Mathieson used in 2015 (Rohland et al., 2022). This made the capture technique accessible to more research groups. Nonetheless, allelic bias when performing capture with Arbor's kit was discovered a couple of years later (Davidson et al., 2023, 2024). In 2022, Svante Pääbo received the Nobel Prize in Physiology or Medicine for his discoveries concerning the genomes of extinct hominins and human evolution. The same year, two-million-year-old genetic material was recovered, allowing the recreation of the ancient Greenland ecosystem (Kjær et al., 2022). In 2023, a group was able to successfully recover ancient human DNA from a 20,000-year-old deer tooth pendant using a non-destructive method that extracts DNA from the pores of bone and tooth artifacts, which can absorb DNA from skin cells, sweat, and other bodily fluids over time (Essel et al., 2023). These and other achievements have opened new avenues for studying long-term evolutionary processes and extinct species.

1.2. Bias in the ancient human genomes published

The rapid growth of ancient population genomics, with thousands of new genomes analyzed yearly, has revealed an evident geographic bias in the distribution of published ancient human genomes (Figure 1). This bias can be attributed to environmental, ethical, institutional, and financial factors.

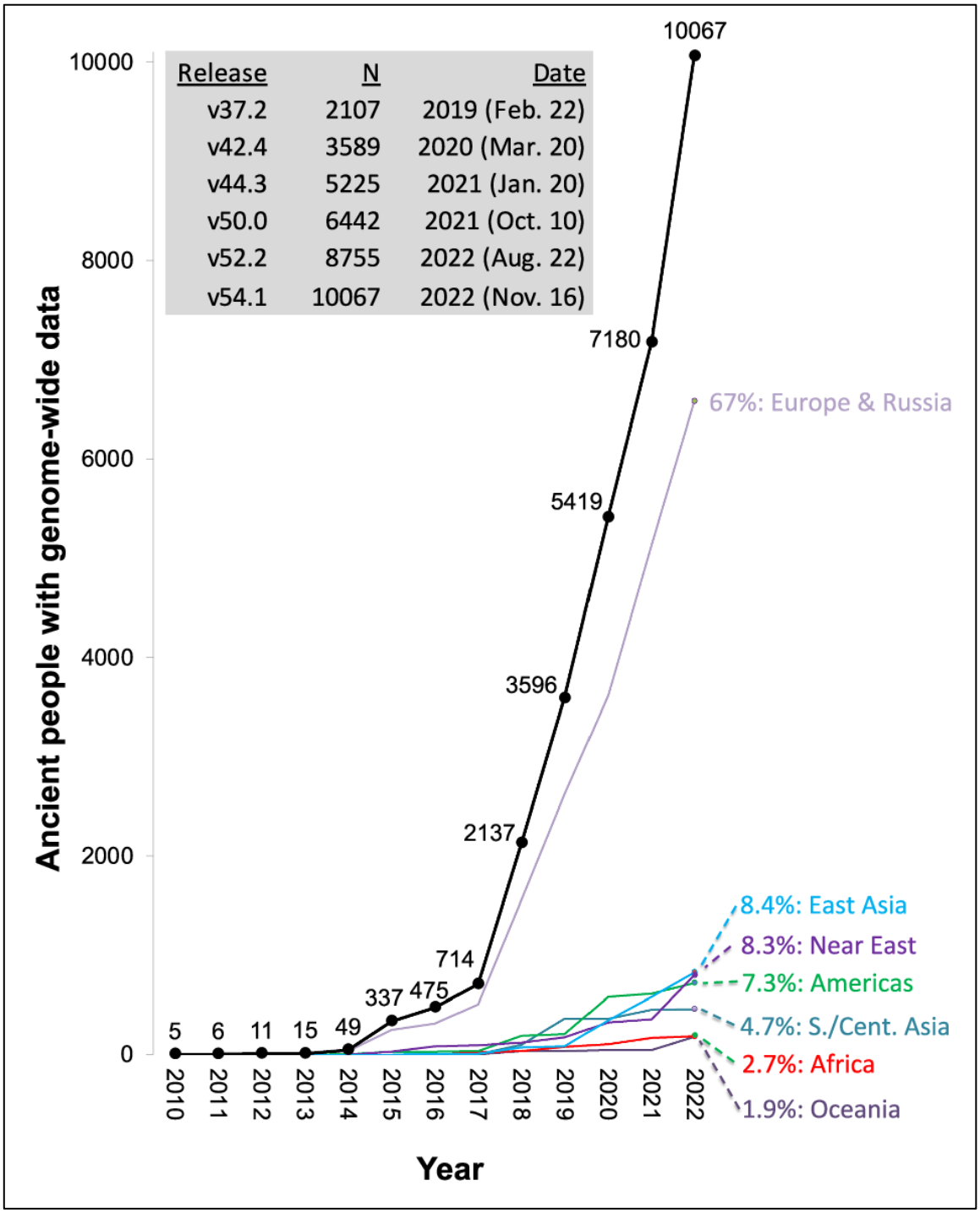


Figure 1. *Number of published ancient human genome-wide data by year of publication (broken down by geography). The grey square indicates the different versions of the "Allen Ancient DNA Resource" database (Release) with the number of ancient genomes (N) and the date of release (Date). Modified from* (Mallick et al., 2024).

Firstly, environmental conditions significantly impact DNA preservation. Factors such as hydrolysis, oxidation, microbial activity, pH, salinity, humidity, and temperature degrade DNA over time, making sample recovery from warm and humid regions particularly challenging (Dabney et al., 2013b; Gutiérrez-García et al., 2014; Molak & Ho, 2011; Olivieri et al., 2010; Poinar et al., 1996; Smith et al., 2001). In contrast, rapid desiccation or freezing can help preserve DNA, but such conditions are rare outside permafrost or arid environments (Dabney et al., 2013b). Consequently, regions such as Europe and Siberia, where cold or dry conditions favor preservation, are overrepresented in the genetic record against the tropical areas.

Secondly, ethical frameworks and cultural considerations are rightly prioritized when studying ancient human remains. Many countries have established regulations to protect the dignity and rights of Indigenous communities, ensuring respectful and responsible research. For instance, the *Native American Graves Protection and Repatriation Act* (NAGPRA) in the United States (Claw et al., 2018; Fleskes et al., 2019) and the *Aboriginal and Torres Strait Islander Heritage Protection Act 1984* in Australia (Department of Agriculture, Water and the Environment, 2021) mandate consultation with affiliated tribes before studying their ancestral remains. Similar frameworks exist in Canada, New Zealand, and the Philippines, where guidelines such as New Zealand's *Te Mata Ira* (Hudson et al., 2016) emphasize the importance of cultural values and Indigenous authority. These practices foster ethical research and community engagement, even if they occasionally limit the availability of samples for aDNA studies. In fact, in the 2024 EMBO | EMBL Symposium "Reconstructing the human past: using ancient and modern genomics" one of the sessions was dedicated to ethical considerations and research practices in paleogenomics. There, Jennifer Raff, the associate professor of anthropology at the University of Kansas, explained how her laboratory engaged the paleogenomic research with Indigenous communities in North America, and how much some tribes are interested in knowing about their ancestors (Raff, 2024).

Thirdly, local research autonomy is increasingly emphasized, particularly in Global South countries such as Chile, India, Mexico, and Puerto Rico (Ávila-Arcos et al., 2022). These countries have focused on developing aDNA research programs, both independently and in collaboration with established international initiatives. Efforts include strengthening cultural and heritage management regulations, securing local funding and infrastructure, building research and training capacity, and prioritizing consultation with Indigenous and descendant communities (Ávila-Arcos et al., 2022). This approach ensures that research priorities align with local interests and cultural heritage protection. For instance, Mexico's National Institute

of Anthropology and History (INAH) strictly regulates access to archaeological samples and enforces a general ban on the export of archaeological, artistic, historical, and fossil remains (Ávila-Arcos et al., 2022). This ensures that ancient Mexican samples are analyzed within the country, preserving national heritage. However, the emphasis on developing local expertise and infrastructure, while crucial, has meant that significant time and resources were invested in establishing and refining aDNA laboratories and training local researchers. This can partly explain the lower number of published genomes from these regions compared to European areas. However, in the past two years, two pioneering studies on ancient Mexican human DNA were performed and published by Mexican researchers (Barquera et al., 2024; Villa-Islas et al., 2023). These achievements reflect the growing strength of local research programs and highlight the importance of encouraging autonomous aDNA research in regions with rich archaeological and cultural heritage.

Lastly, financial and logistical challenges remain key factors. The specialized techniques required for aDNA research — extraction, library preparation, sequencing, and capture — are costly and often require advanced infrastructure and trained personnel. These limitations can restrict the number of samples analyzed, particularly in large-scale studies, limiting the development of state-of-the-art aDNA research in small local labs.

All these factors contribute to the geographic bias observed in ancient human genomic research. Understanding and respecting these considerations is essential for conducting responsible and meaningful studies, while ongoing efforts to address technical and logistical challenges can help broaden the scope of aDNA research worldwide.

1.3. From bones to genotypes: how to study ancient genomes

The analysis of aDNA involves several critical steps, each designed to address the unique challenges posed by degraded and often contaminated genetic material. While each laboratory may have its own protocols and preferred methodologies, certain key steps are widely used across the field (Zhu et al., 2024).

The process begins with the careful selection of ancient samples, typically from bones, teeth, or other biological remains, chosen based on their preservation quality and archaeological context. Petrous bone, in particular, is known to yield the highest quality and most abundant aDNA due to its dense, protected nature and its high concentrations of osteocytes (Haarkötter et al., 2023; Ibrahim et al., 2022; Pinhasi et al., 2015; Vinueza-

Espinosa et al., 2020; Zupanič Pajnič et al., 2025). These samples are then prepared for DNA extraction in specialized clean-room laboratories to minimize contamination risks (Cooper & Poinar, 2000). Commonly used DNA extraction methods include silica-based protocols, which are particularly effective at recovering short, degraded DNA fragments from ancient samples (Boom et al., 1990; Dabney et al., 2013b; Höss & Pääbo, 1993; Rohland & Hofreiter, 2007; Vinueza-Espinosa et al., 2020). Once DNA is extracted, state-of-the-art technology consists of converting DNA into sequencing libraries. These libraries can be based on double-stranded or single-stranded molecules, both of which follow the same principle: adapters (known sequences) are added to the DNA fragments to facilitate amplification and sequencing. Each sample receives a unique pair of indexes (a specific identifier), allowing multiple samples to be pooled and sequenced together in a single run. Shotgun sequencing is typically employed, where the entire DNA library is sequenced, often using Illumina platforms. This approach provides the deep coverage necessary for analyzing low-quality aDNA. The sequencers identify the indexes and separate the samples, storing the resulting sequence reads in FASTQ files, which contain the name of each molecule, its sequence, and its quality score (Cock et al., 2010).

Two optional steps in the workflow can help enhance the quality of the data: Uracil-DNA-Glycosylase (UDG) treatment and in-solution capture. UDG treatment involves the removal of Uracils from DNA sequences after extraction, improving the mappability of ancient molecules by reducing the changes caused by deamination (Briggs et al., 2007). However, this treatment can compromise the authenticity of DNA as ancient, which is based on C-to-T misincorporation patterns, as mentioned before (Briggs et al., 2007; Rohland et al., 2015). To address this, three approaches have been developed: Total UDG treatment, which removes all Uracils for the highest mappability; Partial UDG treatment, which removes only some Uracils, maintaining a balance between mappability and authenticity; and Non-UDG treatment, where the Uracil removal step is skipped in favor of preserving the ability to authenticate the DNA as ancient, at the cost of reduced mappability. In-solution capture is the technique that allows for the enrichment of specific regions of interest in the genome, even when the amount of endogenous DNA is low. Commercial kits have made this technique more accessible to many laboratories (Rohland et al., 2022), but it requires specialized infrastructure, trained personnel, and significant time and financial resources. Many laboratories have been able to achieve excellent results with shotgun sequencing alone, without relying on in-solution capture (Allentoft et al., 2024; Antonio et al., 2019; Margaryan et al., 2020; Moots et al., 2023; Simões et al., 2023).

On what concerns data analysis, currently, most aDNA studies use the nf-core/eager pipeline (Ewels et al., 2020; Fellows Yates et al., 2021; Peltzer et al., 2016), which standardizes the bioinformatic analysis of aDNA, ensuring consistency across studies and reducing errors. However, to accurately interpret the results, it remains essential to understand each step of the process and its purpose:

1) Adapter trimming: adapters added during library preparation must be removed since they are artificial additions to the ends of the sequences (Illumina, 2024). Failure to remove adapters can lead to misalignments and sequencing artifacts, particularly in aDNA studies where reads are often shorter than adapter sequences (Lien et al., 2023). Reads can be sequenced as single-end (read once) or paired-end (read twice). Paired-end sequencing is especially valuable for aDNA, as overlapping regions between forward and reverse reads allow error correction and reconstruction of longer sequences (Lien et al., 2023). The same tools used to trim adapters can merge overlapping reads into a single, longer read, reducing sequencing errors in the overlapped regions (Lien et al., 2023). A quality filter is often applied to discard sequences below a certain length or quality threshold. For example, reads with excessive terminal damage (e.g., cytosine deamination) or too short reads are removed to improve downstream analysis accuracy.

2) Alignment against a reference genome: in human paleogenomics, the most used reference genomes are hg19 and GRCh37. These two versions are functionally identical, differing only in chromosome naming conventions (e.g., hg19 uses “chr1” while GRCh37 uses “1”). Although newer versions of the human genome are available, researchers continue to use these older references to maintain compatibility with previous studies. The most widely used alignment tool is BWA-backtrack, designed specifically for Illumina reads up to 100 bp (Li & Durbin, 2009). This tool has been optimized for aDNA, and recommended parameters are well-documented to ensure optimal alignment (Oliva et al., 2021a; Oliva et al., 2021b). For circular molecules, such as mitochondrial DNA, chloroplast DNA, or plasmids, CircularMapper can improve read mapping by accounting for the circular nature of the sequences (Peltzer et al., 2016).

3) Deduplication: consists of the elimination of duplicate molecules that may arise during PCR amplification or sequencing. Unlike modern DNA studies (Ebbert et al., 2016), duplicates in aDNA analyses can artificially inflate allele frequencies, leading to erroneous conclusions (Hofreiter et al., 2021). This is because high-quality molecules, including modern contaminants, are preferentially amplified. Several tools are used for this purpose, including Dedup (Peltzer et al., 2016), MarkDuplicates from Picard, markdup from Samtools (Danecek et al., 2021), and the dedup parameter from Fastp (Chen et al., 2018). While most of these tools identify duplicates based on read coordinates, others, like Fastp, use sequence similarity.

4) Authentication of aDNA: is a crucial step which relies on *post-mortem* damage patterns, such as cytosine deamination. The characteristic C-to-T (and G-to-A) substitutions at fragment ends help differentiate genuine aDNA from modern contaminants. Popular tools for this purpose include MapDamage2 (Jónsson et al., 2013), Schmutzi (Renaud et al., 2015), DamageProfiler (Neukamm et al., 2021), and PMDtools (Skoglund et al., 2014a). For untreated aDNA (without UDG), true ancient sequences should display a high frequency of damage at molecule ends. Additional authentication techniques focus on assessing contamination levels using mitochondrial DNA (Renaud et al., 2015) and the X chromosome (Korneliussen et al., 2014). Once authenticated, misincorporation patterns are removed to prevent damage-derived substitutions from being misinterpreted as genuine variants (Jun et al., 2015).

5) Genotype calling: while working with ancient human populations, genotypes are often called using a pseudohaploid approach. This method assumes that each position in the genome has a single allele, randomly choosing one from the observed reads. It is a common strategy for aDNA due to the typically low-depth coverage — the number of times a base in a genome is sequenced — of ancient genomes. This approach minimizes the impact of errors from damage patterns while still providing valuable insights into genetic variation.

1.4. Advantages of mitochondrial DNA (mtDNA) when working with aDNA

Given the aDNA challenges, mtDNA was the most studied region in paleogenetic studies, and in the genomic era, it is still offering several advantages. Its high copy number per cell (hundreds to thousands of copies) increases the likelihood of recovery from degraded samples (Allentoft et al., 2012). Moreover, the circular structure of mtDNA provides better protection against degradation than linear nuclear DNA (Allentoft et al., 2012). Additionally, the smaller size and higher mutation rate of the mitochondrial genome make it particularly useful for population genetics and phylogenetic studies (Brown et al., 1979; Giles et al., 1980). Another advantage is that human mtDNA phylogeography (Figure 2) is well established (Hernández, 2023). The Most Recent Common Ancestor of all human mtDNAs (mtMRCA) is macro-haplogroup L, dating back to around 200,000 years ago, with its root located in Africa (Fu et al., 2013; Kaneva et al., 2022). This macro-haplogroup was differentiated into eight major haplogroups (from L0 to L7) found predominantly in Africa, supporting the "Out of Africa" theory of human origins (Cerezo et al., 2012; van Oven & Kayser, 2009). The L3 haplogroup originated around 60,000-80,000 years ago and is considered the one that went outside Africa (Soares et al., 2012) and the most recent common ancestor of all non-African mtDNA lineages. The direct descendants of L3 are M and N (Soares et al., 2012). Haplogroup R differentiated shortly after the haplogroup N split from L3, around 50,000-60,000 years ago, during the early stages of the "Out of Africa" migration in Southeast Asia (Fregel et al., 2015; Larruga et al., 2017). It gave rise to several haplogroups, such as Asiatic lineages (B), and the most common lineages in Europe: R0 (includes HV, H, and V), JT (includes J and T), and U (includes K). This well-known phylogeography is a valuable tool to address some questions about the origin of ancient humans and the role of women during human migrations and population origins (Hernández, 2023; Jorde et al., 1998; Maca-Meyer et al., 2001; Merheb et al., 2019; Torres, 2016; Witas & Zawicki, 2004).

MtDNA is certainly the most studied region of the human genome. It is so well studied that the probability of mutation has been estimated for both non-coding and coding regions (Santos et al., 2005, 2008) and also for each position, and some mutational hotspots — positions more likely to mutate than expected — have been identified (Soares et al., 2009). The study of mutational hotspots has also been applied to mtDNA positions detected in heteroplasmy, a condition when a cell, tissue, or individual presents different mtDNA haplotypes (Li et al., 2015; Ramos et al., 2013). Nevertheless, *post-mortem* molecular DNA damage poses a challenge in mitochondrial aDNA data analysis, as it can be difficult to distinguish from mtDNA heteroplasmy, contamination, and errors arising from DNA

amplification and sequencing. Evaluating a mixed profile in ancient mtDNA is highly complex due to the challenge of distinguishing genuine endogenous variants from artifacts. This underscores the importance of identifying *post-mortem* damage-derived hotspots for accurately interpreting base mixtures in ancient samples with low-depth coverage. While numerous studies have explored the extent and nature of *post-mortem* damage (Briggs et al., 2007; Brotherton et al., 2007; Olivieri et al., 2010; Sawyeret al., 2012), others have focused on identifying *post-mortem* damage hotspots in mtDNA from plants (Staats et al., 2011) and animals, including bison (Gilbert et al., 2005), Neanderthals (Vives et al., 2008), and humans (Gilbert et al., 2003; Lamers et al., 2009). The studies on ancient humans (Gilbert et al., 2003; Lamers et al., 2009) employed cloning technology to generate multiple sequences from the same region, often focusing on the mtDNA Hypervariable Region I or small regions of the mtDNA coding region. Using samples from fewer than 50 individuals per study, these investigations demonstrated that *post-mortem* damage is not randomly distributed, revealing damage hotspots that were not observed as hypermutable positions in modern datasets (Gilbert, 2006; Gilbert et al., 2003). More recent studies have concentrated on analyzing the non-coding region of mtDNA, amplified by PCR and sequenced using NGS from artificially degraded samples (Holland et al., 2021; Rathbun et al., 2017). Rathbun et al. (2017) showed that DNA damage can be induced both passively, through storage conditions, and actively, through accelerated hydrolysis. They identified 27 recurring position changes, suggesting these sites may be particularly prone to damage. Holland et al. (2021) evaluated the extent of damage in free-diluted aDNA, emphasizing the importance of distinguishing between damage and heteroplasmy. In their study, 35 potential damage hotspots were identified. Notably, both studies considered all base changes rather than focusing solely on C-to-T or G-to-A misincorporations. However, no study has yet investigated molecular damage hotspots across the entire human mitogenome.

The analysis of mitochondrial DNA (mtDNA), whether obtained through PCR, shotgun sequencing, or targeted capture, has proven effective for samples where nuclear DNA recovery was limited or impossible, offering valuable insights into ancient population dynamics, migration patterns, and evolutionary relationships (Diroma et al., 2021). However, as mtDNA only traces maternal lineages, it provides a partial view of the genetic makeup of past populations. Interpretations should therefore be approached with caution and, whenever possible, complemented with nuclear DNA data. Nevertheless, mtDNA remains a powerful tool, particularly for identifying patterns of female mobility and assessing affinities

between ancient and modern populations, thus contributing to our understanding of long-term demographic trends (Achilli et al., 2004).

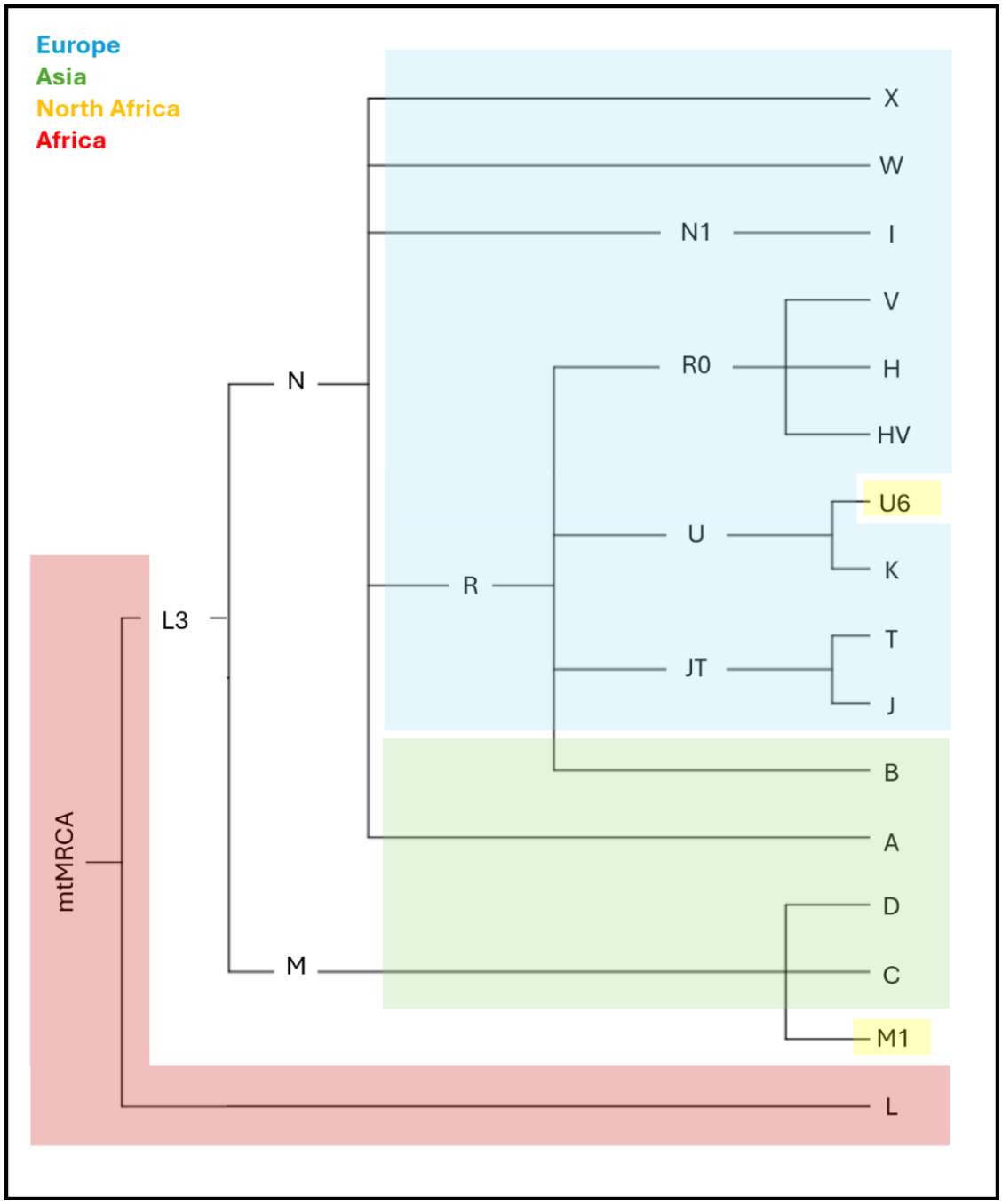


Figure 2. *Simplified phylogeography of human mitochondrial DNA. The mtMRCA refers to the Most Recent Common Ancestor of all human mtDNAs. Modified from Hernández (2023).*

The study of female mobility has revealed diverse kinship and residence systems across prehistoric Europe, reflecting the evolving social structures of ancient communities. In many Neolithic and Bronze Age contexts, societies often followed patrilocal models characterized by female exogamy. For instance, in Corded Ware groups from Bohemia (2,900–

2,400 BCE), genetic and isotopic evidence showed the integration of non-local women into communities with stable Y-chromosome lineages but diverse mtDNA, consistent with patrilocal residence and female mobility (Papac et al., 2021). A patrilocal mating system during the Bronze Age, where women moved to the husband's community after marriage, was also suggested in the Iberian Peninsula (Simón et al., 2011; Villalba-Mouco et al., 2022). In contrast, matrilocal or female-centered systems have been identified in other contexts. In Neolithic Greece, a marked reduction in mtDNA diversity by the 4th millennium BCE suggested increasing female sedentism, likely associated with matrilocality (Silva et al., 2022). Similarly, in Iron Age southern England (100 BCE – 100 CE), mitochondrial data from a Celtic community showed that 64% of individuals, across both sexes, shared a single founding maternal lineage, supporting a matrilocal model. The presence of non-local grave goods in male burials suggested that men moved into established female kin groups through marriage (Cassidy et al., 2025).

Beyond residence systems, female mobility also played a key role in cultural transmission and demographic integration. In Phoenician settlements across the Mediterranean (700–400 BCE), mitochondrial genomes from southern Sardinia and Lebanon revealed the presence of both indigenous and non-local haplogroups, as Near Eastern N1b1a5 and European W5 in Sardinia, and European lineages in Lebanon. These patterns indicate bidirectional female migration between North Africa, the Levant, and Europe, highlighting the active role of women in trade and settlement networks and challenging the assumption of exclusively male-driven migrations (Matisoo-Smith et al., 2018).

These examples illustrate how institutionalized sex-biased mobility shaped past European societies' genetic and social landscapes.

1.5. Overview of paleogenomic studies in the Iberian Peninsula

Recent advancements in paleogenomic techniques have provided new opportunities to explore the genetic history of ancient populations in various archaeological contexts and how the local ancestry mixed with ancestries from different sources (Figure 3). In this context, the Iberian Peninsula's genetic history has been well analyzed during the last years.

Paleolithic samples from the Iberian Peninsula, including those from the south (Cueva del Malalmuerzo, Granada, 23 kya BP) (Villalba-Mouco et al., 2023) and the north (El Mirón, Cantabria, 19 kya BP) (Fu et al., 2016), have helped identify the Iberian Peninsula as a

refuge during the Last Glacial Maximum (LGM, 26–19 kya BP). These findings suggest genetic continuity from pre-LGM to post-LGM populations (Posth et al., 2023; Villalba-Mouco et al., 2023). During the Mesolithic, hunter-gatherers from the North of the Iberian Peninsula exhibited a higher genetic influence from Western Hunter-Gatherers (WHG) than those in the south (Olalde et al., 2019). Anatolian farmers expanded into Europe during the Early Neolithic (Figure 3), and their coexistence with local hunter-gatherers led to an increase in Neolithic ancestry in early European populations. However, this ancestry was gradually diluted in favor of WHG ancestry by the end of the Neolithic and the beginning of the Chalcolithic. This event has been documented in populations from all over Europe and even in the Iberian Peninsula (Fu et al., 2016; Gamba et al., 2014; Haak et al., 2015; Lipson et al., 2017; Martiniano et al., 2017; Mathieson et al., 2015; Olalde et al., 2019; Posth et al., 2023; Szécsényi-Nagy et al., 2017; Valdiosera et al., 2018). The Chalcolithic in the Iberian Peninsula was a period of continuity from the Neolithic (Olalde et al., 2018, 2019; Roca-Rada et al., 2024; Szécsényi-Nagy et al., 2017).

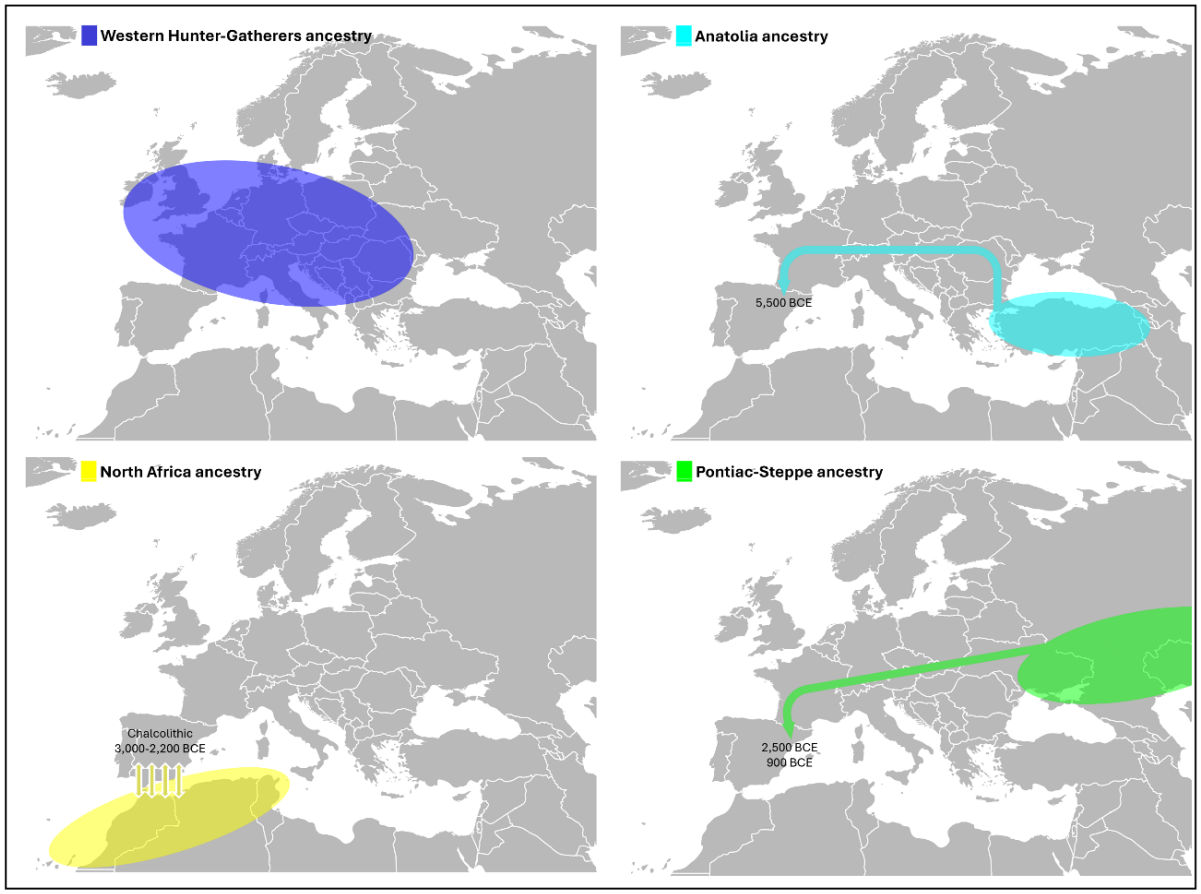


Figure 3. *Main genetic ancestries that influenced and shaped the Iberian Peninsula genetic pool. The moment when those components were supposed to arrive is indicated.*

During the Bronze Age, different groups carrying Pontic-Caspian Steppe ancestry spread westward throughout Europe. In the Iberian Peninsula, the Early Bronze Age (2200–1500 BCE) was marked by the introduction of Steppe-related ancestry. However, in the Iberian Peninsula, its influence varied regionally, being more prominent in the northern and northeastern regions than in the south (Olalde et al., 2019; Roca-Rada et al., 2024; Villalba-Mouco et al., 2021). This suggests that the Bronze Age genetic transition in the Iberian Peninsula was gradual and uneven. It should be noted that by the end of the Chalcolithic, individuals carrying "steppe-related ancestry", which are often but not exclusively linked to Bell Beaker-associated artifacts (Olalde et al., 2018, 2019; Villalba-Mouco et al., 2021), and/or African and Iran Neolithic-like ancestries (Fregel et al., 2018; Olalde et al., 2019; Villalba-Mouco et al., 2021) had already been reported (Figure 3), a clear indication of overseas relationships (Villalba-Mouco et al., 2021). Another significant genetic characteristic of this period was the shift in Y-chromosome lineage frequencies. The Steppe-related lineage, R1b, almost replaced earlier lineages such as I2, G2, and H, which persisted at low frequencies since then (Olalde et al., 2019; Roca-Rada et al., 2024; Villalba-Mouco et al., 2021).

The Iron Age in the Iberian Peninsula (800–100 BCE) was characterized by an increase in Steppe-related ancestry in northern populations compared to Bronze Age populations from the same region. This pattern was observed in both Indo-European (Olalde et al., 2019; Papac et al., 2023) and non-Indo-European-speaking populations (Olalde et al., 2019). The arrival of the Romans introduced additional genetic ancestries into the local genetic pool, incorporating gene flow from North Africa and the Eastern Mediterranean (Carrión et al., 2024; Olalde et al., 2019).

1.6. History of the Western Mediterranean and the Iberian territories during the Iron Age

Iberian culture emerged during the Iron Age around the 6th century BCE on the eastern and southern coasts of the Iberian Peninsula, along the Mediterranean Sea. It reached its peak between 450 and 200 BCE. Their funeral practices were based mainly on a rite of cremation. Nevertheless, from the new findings and studies, it is known that cremation necropolises do not represent all types of Iberian burial (Oliver Foix, 2004). For example, some newborns were buried under the pavement of houses and other remains had been found in silos and habitat sites (Afonso et al., 2019; Crespo et al., 2011; Oliver Foix, 2004;

Sandoval-Ávila et al., 2025; Subirà & Molist, 2007). In other contexts, it is hypothesized that the social differences, the difference in age, and the type of death probably marked the funeral rite (Eng & Smith, 1976; Gowland et al., 2014; Hassan et al., 2014). However, recent evidence suggests that the age profile of newborns aligns well with natural mortality causes (Martirosyan et al., 2024; Sandoval-Ávila et al., 2025). This civilization was highly developed, with its own alphabet and coin currency. Their economy was based on agriculture, livestock, and fishing in coastal areas. The Iberians excelled in iron metallurgy, producing a variety of tools. Additionally, they participated in Mediterranean trade networks (Almagro et al., 2001; Jordá Cerdá et al., 1986). The term "Iberians," coined by the Romans, refers to a heterogeneous reality rather than a uniform culture. Each group (Figure 4) — Ausetani, Cessetani, Ilergetae, Indiketes, Lacetani, Layetani, and others — underwent distinct sociopolitical developments, as reflected in the archaeological record, rather than following a single, linear trajectory (Aranegui, 2012; González Reyero, 2012).

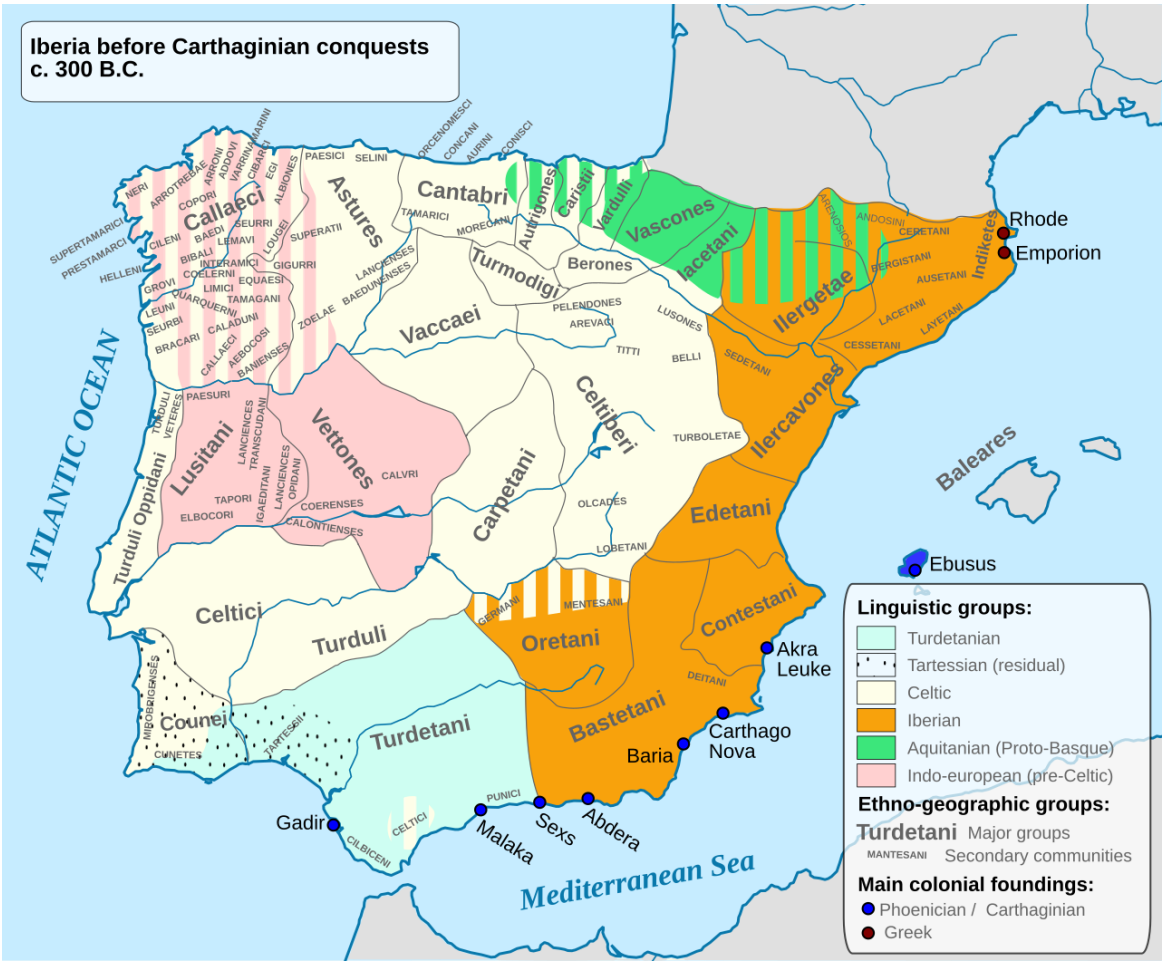


Figure 4. *Distribution of tribes in the Iberian Peninsula during 300 BCE. Figure from Alcides Pinto - Based on the map done by Portuguese Archeologist Luís Fraga, from the "Campo Arqueológico de Tavira".*

The question of the origin and evolution of Iberian society has been one of the most debated topics since the earliest research on this culture. The scarcity of written sources and the indirect nature of much of the available documentation, often ambiguous, have led to shifting theories on how Iberian society emerged. One of the first hypotheses proposed that Iberian culture originated from a local substrate with Neolithic African roots, later shaped by interactions with Greek colonies in the southern Iberian Peninsula (Bosch-Gimpera, 1938). This idea contributed to the concept of the Ibero-Tartessian culture, which was thought to have expanded from the southeast to northern territories (Sanmartí, 2013). However, in the 60s, several authors such as Joan Maluquer de Motes and Miquel Tarradell challenged the notion of migration from the southeast (Sanmartí, 2011, 2013). Instead proposed a model of simultaneous local development across the Iberian territory (Maluquer de Motes, 1966, 1982; Sanmartí, 2011, 2013; Tarradell, 1962). According to Maluquer de Motes (1982), Iberian society emerged as an evolution of pre-Iberian communities influenced by colonial interactions, particularly from the 7th century BCE onward (Maluquer de Motes, 1982).

In the Catalonian territory (Northeast of the Iberian Peninsula), several archaeological indicators suggest a shift toward Iberian culture after the 6th century BCE. One of the most significant changes occurred in urbanization: pre-Iberian settlements, consisting of small clusters of detached houses (1100–550 BCE), evolved into larger, hierarchical entities with organized territorial structures without significant external influence (Bosch & Santacana, 2009; Junyent, 1996a, 1996b, 2023; Sanmartí, 2003). Settlements became classified into different categories based on their function and regional influence (Asensio et al., 2001; Sanmartí et al., 1998). Social transformations are also evident from the late 6th century BCE, marked by the emergence of aristocratic elites with clear displays of power (Belarte, 2010; Bendala, 1998).

Throughout the first millennium BCE, the Iberian Peninsula was a dynamic crossroads of cultural interactions, shaped by the arrival of external civilizations that influenced local societies. Phoenician and Greek colonization played a crucial role in transforming economic, social, and political structures within Iberian territories. These foreign settlements not only introduced new trade networks but also facilitated the exchange of goods, technologies, and ideas between indigenous Iberian groups and the broader Mediterranean world. Phoenician colonies were established in the southern and southeastern regions of the Iberian Peninsula as early as the 9th century BCE (Gitin & Bierling, 2002). Notable examples include Gadir, Malaka, and Abdera (Figure 4). The colony of Abdera (8th century) was situated in the territory of the Iberian tribe known as Basetani, which corresponded to present-day Granada, Almería,

and parts of Málaga, Jaén, Albacete, and Murcia (Keay & Deamos, 1998). This colony became a vital trade center, facilitating exchanges between the Iberians of the whole Iberian Peninsula and the Phoenicians. In the year 575 BCE, Emporion was founded by the Phocean Greeks and became the primary trade hub for the northeast Iberians (Figure 4), connecting them with various Mediterranean cultures, including the Greeks, Phoenicians, and Etruscans (Dietler, 2009).

Founded in North Africa around 814 BCE by Phoenician settlers, Carthage gained independence from its parent cities around the 7th century BCE and gradually expanded its influence across the western Mediterranean. This included maintaining and strengthening the trading links established by earlier Phoenician colonies in southern Iberia. Over the following centuries, Carthage emerged as a dominant power, thriving as a commercial and maritime hub. By the 4th century BCE, it had established itself as the center of a vast empire, controlling trade routes and exerting influence across North Africa, southern Iberia, and the islands of the western Mediterranean. This expansion brought the Carthaginians (also known as Punics) into conflict with the emerging power of Rome. Initially a monarchy, Rome transitioned to a republic in 509 BCE and embarked on a period of rapid territorial expansion (Cornell, 1995). Its consolidation of the Italian Peninsula involved conflicts such as the Samnite Wars (343–341 BCE, 326–304 BCE, 298–290 BCE), the Latin War (340–338 BCE), and the Pyrrhic War (280–275 BCE) (Cornell, 1989, 1995; Franke, 1989; Grant, 1978; Oakley, 1998; Schullard, 1989). Having established dominance over the Italian territory, Rome turned its ambitions towards the Mediterranean, which conflicted with Carthaginian expansionist policies. This rivalry culminated in the three Punic Wars (264–241 BCE, 218–201 BCE, and 149–146 BCE), a series of conflicts that defined the struggle for supremacy in the western Mediterranean (Bagnall, 1999; Carey et al., 2007; Cornell, 1995; Edwell, 2011; Feeney, 2017; Goldsworthy, 2012; Grant, 1978; Richard, 2011; Schullard, 1989; Zimmermann, 2011).

In 236 BCE, following the defeat of Carthage in the First Punic War, a Carthaginian army composed of Numidians, mercenaries, and another infantry entered the Iberian Peninsula. They aimed to establish a new empire for Carthage to compensate for the territories lost in recent conflicts with Rome and to create a base for future revenge against the Romans (López Castro, 2019). In eight years, through military might and diplomatic efforts, the Carthaginians secured a territory that covered nearly half of the Iberian Peninsula (Feeney, 2017). Several cities were founded during this time, including Qart-Hadasht (Cartago Nova or New Carthage) (Cutillas-Victoria et al., 2024). This city was located in present-day Murcia (Figure 4), within the territory of the Iberian tribes known as the Mastieni (Ballester

Tormo, 1952). In 226 BCE, a treaty between the Romans and Carthaginians established the Ebro River as the northern boundary of Carthaginians in the Iberian Peninsula, except for Saguntum, which was located under the Ebro River (Goldsworthy, 2012) and was part of the Iberian territory of the Edetani tribe (present-day Valencia). The Ebro Treaty implied that Iberian tribes from the Northeast of the Iberian Peninsula (Cessetani, Layetani, Lacetani, Ilergetae, Ausetani, Indiketes, Bergistani, Ceretani, Andosini, and Arenosios) did not come directly into contact with the Carthaginian army. In the spring of 218 BCE, Hannibal assembled a Carthaginian army in New Carthage, leading it northwards along the Iberian coast to attack Rome through an inland route from the North of the Italian Peninsula, avoiding Roman allies from the South of that peninsula (Mahaney, 2013). In the fall of that year, the Roman fleet landed in the northeastern part of the Iberian Peninsula at Emporion, gaining the support of the local tribes. Over the following years, groups of Iberians and Celtiberians defected from the Carthaginians to Rome, strengthening the Roman armies as they moved south (Zimmermann, 2011).

In 206 BC, at the Battle of Ilipa (present-day Sevilla), an army equally composed of Italians and Iberians defeated a Carthaginian force, sealing the Carthaginians' fate in Iberia (Edwell, 2011; Zimmermann, 2011). This defeat led to the defection of the last Carthaginian-held city in Iberia, Gadir (present-day Cádiz), to the Romans (Richard, 2011). After the Carthaginians were expelled, Roman troops remained on the peninsula, which sparked various mutinies and uprisings among different communities in the Iberian Peninsula (Bagnall, 1999; Carey et al., 2007). Some mercenaries from Iberia even participated in Carthaginian offensives (Edwell, 2011). In 197 BCE, the Roman Republic divided the Iberian Peninsula into two provinces: Hispania Citerior ('Nearer Spain') and Hispania Ulterior ('Further Spain') to formalize its control over the region (Keay, 1988; Richardson, 1986). Hispania Citerior initially covered the northeastern coastal areas and parts of the interior, including the territories of modern Valencia, Catalonia, and part of Aragon, with Tarraco (modern Tarragona) as its capital (Ruiz de Arbulo, 2006). The region faced significant resistance from local Iberian tribes and Celtiberians, leading to prolonged conflicts such as the Celtiberian Wars (181–133 BCE), which ended with the Siege of Numantia (134-133 BCE), a decisive victory that consolidated Roman dominance over Citerior (Curchin, 1991; Jimeno Martínez, 1996). Hispania Citerior became a crucial stage for Roman generals seeking political and military glory during the late Republic. Great names, such as Gaius Marius, Quintus Sertorius, and Julius Caesar, conducted campaigns in the region, using its strategic position for broader ambitions (Morillo Cerdán, 2014). At the beginning of the 1st century BCE,

a civil war known as the Sertorian War (80–72 BCE) highlighted the local resistance against Roman central authority, with the province of Citerior briefly serving as a base for anti-Sullan forces (Spann, 1987). Under Augustus, Hispania Citerior was reorganized into the province of Tarraconensis during the early Imperial period (27 BCE) (Alföldy, 1996). This transformation marked the region's integration into the Roman administrative system, focusing on infrastructure development, including roads, aqueducts, and the expansion of Tarraco (Macias & Rodà, 2015).

As can be seen, different Iberian tribes engaged with a variety of cultures, shaping their development and identity. While material evidence from archaeological contexts offers valuable insights into the cultural traits of each site and the trade networks connecting them, it falls short of revealing the nature of these interactions or the extent to which individuals from other cultures integrated into local communities and traditions (or vice versa). To address these questions, we must turn to a complementary field: paleogenomics. This powerful tool can provide a new dimension of evidence, shedding light on the genetic makeup of Iberians and offering deeper insights into their interactions and migrations. These genetic insights should complement archaeological evidence, providing a more comprehensive understanding of the demographic processes that shaped populations from the Iberian Peninsula over time.

Despite extensive research on the prehistory of the Iberian Peninsula, genetic studies focusing on the Iron Age remain limited, particularly regarding Iberian culture. As previously mentioned, the main funerary practice of the Iberians was based on cremation (Oliver Foix, 2004), which has resulted in a significant lack of anthropological remains. This absence has made comprehensive genetic and anthropological research on these populations particularly challenging. Nevertheless, some intramural newborn burials have been documented within houses and productive areas (Afonso et al., 2019; Crespo et al., 2011; Oliver Foix, 2004; Sandoval-Ávila et al., 2025; Subirà & Molist, 2007), and these remains may represent valuable sources of biological information.

The first paleogenetic research on Iberian culture groups focused on studying mitochondrial DNA, specifically the hypervariable region I. These early paleogenetic investigations aimed to identify the prevalent mitochondrial haplogroups among Iberians in comparison to Chalcolithic, Bronze Age, and Roman individuals (Gamba et al., 2008) and with different cultural groups, including the Basques and Etruscans (Sampietro et al., 2005). These studies analyzed the remains of 27 adults from nine archaeological sites, including bone

fragments found in regions historically occupied by Iberians but with uncertain Iberian origins. The results revealed that the haplogroup composition of ancient Iberians was remarkably similar to that of modern Iberian Peninsula populations, albeit with less variability, suggesting long-term genetic continuity in the region since pre-Roman times (Sampietro et al., 2005). Moreover, no close genetic relationship was found between the Iberians and the Etruscans, indicating considerable genetic heterogeneity in pre-Roman Western Europe (Sampietro et al., 2005). The predominant haplogroup in these populations was H, with additional lineages of J, T, K, U4, U5, and W also present (Gamba et al., 2008; Sampietro et al., 2005). Notably, an unusually high frequency of haplogroup V was observed, which has been considered a marker of Paleolithic populations in Europe (Gamba et al., 2008). Advancements in genetic research methods, particularly the adoption of NGS and capture techniques in paleogenomic studies, have significantly improved the recovery of mitochondrial and nuclear data from 15 individuals (3 newborns from Valencia and 12 adults from Catalonia) from seven Iberian archaeological sites (Olalde et al., 2019). The results revealed that Iron Age individuals from the Iberian Peninsula had increased ancestry related to Northern and Central European populations compared to Bronze Age individuals (Olalde et al., 2019), as previously mentioned.

1.7. Iberian Archaeological sites analyzed in this thesis

In this thesis, paleogenomic cutting-edge methods were used to investigate the newborns recovered so far and buried under the pavement of six key Iberian archaeological sites, each representing different periods and Iberian tribes (Figure 5).

The fortress of Vilars (1 in Figure 5), located near Arbeca in the province of Lleida (Catalonia, Spain), represents a significant archaeological site from the Early Iron Age. It was founded around 775 BCE and was continuously occupied until approximately 325 BCE. The site is stratigraphically divided into five distinct phases, reflecting its historical and cultural evolution. These include two phases from the Early Iron Age: Vilars 0 (775–675/650 BCE) and Vilars I (675/650–550/525 BCE). The subsequent phases correspond to the Ancient Iberian period: Vilars II (550/525–450/425 BCE), and the Iberian period: Vilars III (450/425–375 BCE) and Vilars IV (375–350/325 BCE). Constructed by the Ilergetae tribe, this fortified settlement is notable for its impressive defensive architecture, including a massive wall measuring 5 meters in thickness and height, reinforced by twelve semicircular towers and a floodable moat that enhances its fortification. The fortress was strategically situated on an alluvial

plain, allowing its inhabitants to control vital water resources while facilitating agricultural practices centered on livestock and grain cultivation. The layout of the fortress is characterized by an oval shape, with residential structures arranged around a central water cistern, indicative of advanced planning and community organization. The site's defensive features — such as the chevaux-de-frise barrier designed to impede access — rendered it nearly impregnable and unique within Iberian and broader European contexts. The significance of Vilars extends beyond its architectural prowess; it serves as a critical point for understanding the socio-economic transformations during the Iron Age in northeastern Iberia. The ongoing research at this site has provided valuable insights into the emergence of segmentary societies and the development of early aristocracies in the region. (Agustí et al., 2000; Junyent et al., 1998; Junyent & López, 2017; Junyent & Moya, 2011; López et al., 2020). Regarding the material culture found in the site, during the earlier phases (Vilars 0-II), most of the pottery corresponded to pre-Iberian or Iberian tradition, but in phase Vilars I already appeared evidence of Phoenician influence, including jewelry and pottery. By phase Vilars II, Punic- and Greek-influenced Iberian pottery became prominent, alongside Mediterranean or Mediterranean-influenced ivory artifacts and additional Phoenician jewelry (Alonso et al., 2004; Junyent & López, 2017). More than thirty newborns were found buried under the pavement of living rooms, storage rooms, or metallurgical workshops along the different phases.

The archaeological sites of Illa d'en Reixac and Puig de Sant Andreu are integral components of the Iberian settlement of Ullastret (2 in Figure 5), located in the province of Girona (Catalonia, Spain), which is recognized as one of the largest and most significant Iberian sites in Catalonia. It dated to the 6th century BCE and had two urban centers. One settlement is named after the hill on which it is located, Puig de Sant Andreu, and features a complex defensive system that includes imposing walls and towers, reflecting Mediterranean architectural influences acquired through contact with Greek colonies. The site encompasses a well-preserved urban layout, including residential areas, public spaces, and religious temples, which offer insights into the social organization and daily life of the Indiketes tribe. Illa d'en Reixac, located approximately 500 meters northeast of Puig de Sant Andreu, on the plain, is believed to have been a vital part of this dual settlement. This site features similar architectural elements, including fortifications and domestic structures, indicating that it may have served as an important administrative center. Excavations at Illa d'en Reixac have revealed significant findings, such as storage pits for grains and evidence of metallurgical activities, which highlight the economic practices of the Iberians. Together,

these sites illustrate the complexity of Iberian urbanism and provide a rich context for understanding cultural interactions in the region (López i Melcion et al., 1989; Martín i Ortega, 1985; Martín i Ortega et al., 2016). Several cut-off heads had been found, opening a debate about whether they were enemies kept as trophies or venerated ancestors. However, new insights into the mobility of three individuals have revealed that what previously could have been considered enemies were, in fact, local and non-local individuals from the surrounding Iberian territory (de la Fuente-Seoane et al., 2025). Several newborns buried under the pavement of the houses have already been found and studied from an anthropological point of view (Martirosyan et al., 2024; Sandoval-Ávila et al., 2025).

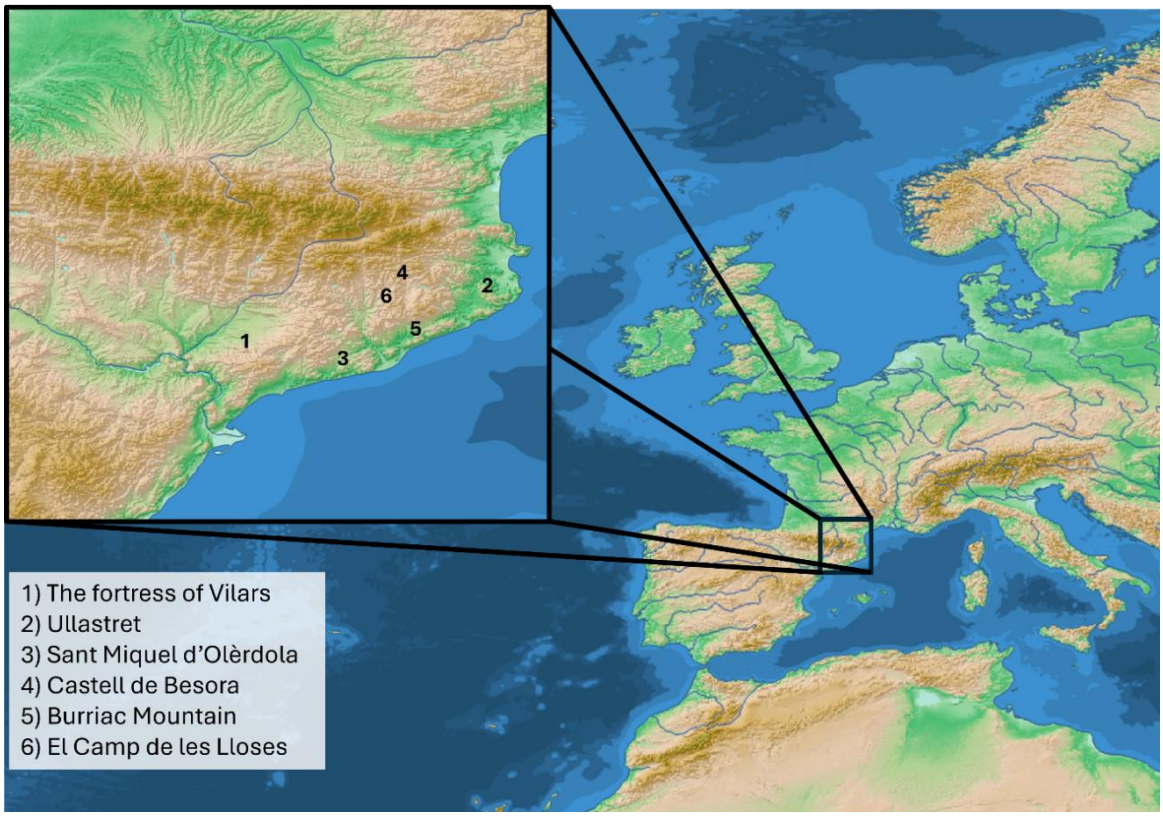


Figure 5. *Location of each of the Iberian sites studied in this thesis.*

The archaeological site of Sant Miquel d'Olèrdola (3 in Figure 5) is a significant historical landmark situated on a strategic hill overlooking the Penedès plain, in the province of Barcelona (Catalonia, Spain). This site has been continuously inhabited from the Bronze Age through the High Middle Ages, reflecting its long-standing importance in the region. The earliest fortified structures date to the Iron Age (8th-7th centuries BCE), when the Cessetani tribe established a walled oppidum that served as a central hub for trade and defense. The site features impressive remnants of its defensive architecture, including a Roman wall approximately 145 meters long, which includes a central gateway and several towers. These

fortifications were later reused during the Roman period (2nd-1st centuries BCE) and the medieval era (9th-12th centuries CE), illustrating the site's adaptability and continued significance throughout various historical phases. Excavations have uncovered various elements from different periods, including residential structures, silos, and a cistern carved into the rock, highlighting the advanced engineering skills of its inhabitants. (Crespo et al., 2011; Subirà & Molist, 2007). In addition to the typical Iberian and pre-Iberian material culture already expected on this site, several Mediterranean objects have been also found in the Iron Age stratigraphy, such as Punic amphorae and ceramics, from Ebusus (present-day Ibiza) and Central Mediterranean, and Italian and Greek ceramics (Molist Capella & Bosch Casadevall, 2002). Regarding the anthropological remains, the most characteristic finding of this site is the simultaneous burial of two newborns, that have been interpreted as twins, despite no successful genetic analysis available (Crespo et al., 2011).

The Castell de Besora (4 in Figure 5), located at an altitude of approximately 1,025 meters in the municipality of Santa Maria de Besora, province of Barcelona (Catalonia, Spain), is a historically significant site that has played a crucial role in the region's development from the Iron Age through the modern period. The site has been occupied since at least the Iberian era (by the Bergistani tribe), with archaeological evidence suggesting continuous habitation and strategic use throughout various historical phases. The earliest documented references to the castle date back to the late 9th century, indicating its importance as a defensive structure during the early medieval period. Two Iberian newborns have been found in the Iron Age stratigraphy, which to date has been poorly investigated (Busquets et al., 2017).

The archaeological site of El poblat Ibèric de Burriac, located at the foot of Burriac Mountain (5 in Figure 5), in the province of Barcelona (Catalonia, Spain), is a significant Iberian settlement that showcases the rich history of the Layetani, an ancient Iberian people who inhabited the northeastern region of the Iberian Peninsula. The site began to be occupied in the 6th century BCE, but it was during the 4th century BCE that it gained political significance, marked by the construction of fortifications and a network of streets. Covering an area of approximately 7 to 10 hectares, Cabrera de Mar emerged as one of the largest urban centers of its time, serving as a political and economic hub for the surrounding agricultural territories. The strategic location of Cabrera de Mar allowed it to control vital resources in the Vallés plain, particularly wheat production, which contributed to its economic power until the end of the 3rd century BCE. Following the Second Punic War, significant changes occurred in the region's political landscape, leading to the decline of Iberian settlements dependent on

Cabrera and prompting Roman urban development in the area. Excavations have revealed important public and private buildings from this period, including a Roman settlement where minted coins inscribed with the Iberian legend "ILTURO", indicating its continued importance during Roman rule. Recent archaeological discoveries at Cabrera de Mar have also included thermal baths and a wine production center from the late Republican period, providing further insights into the daily life and economic activities of its inhabitants. The remains of these structures illustrate a blend of Iberian and Roman influences that characterize this transitional period in Iberian history. To date, only one newborn has been found in the Iron Age stratum (Martín i Méndez, 1998; Sinner, 2015; Sinner & Ferrer i Jané, 2016).

El Camp de les Lloses, located in Tona, province of Barcelona (Catalonia, Spain) (6 in Figure 5) and territory of the Ausetani tribe, is a significant archaeological site that provides crucial insights into the Romanization process in the Northeast of the Iberian Peninsula during the late Republican period. Dating from the 2nd to the 1st century BCE, this site was strategically situated at a crossroads of ancient trade routes, facilitating interactions between indigenous populations and Roman settlers. The settlement is characterized by its complex layout, which includes residential structures, workshops, and evidence of agricultural activities, indicating a vibrant community engaged in both local production and trade. Excavations at El Camp de les Lloses have revealed a wealth of artifacts, including Iberian coins, bronze objects, and ceramics that reflect the cultural exchanges occurring during this period. The presence of imported goods alongside local products underscores the site's role as a center for economic and cultural interactions between the indigenous Ausetani tribe and Roman influences. Notably, the discovery of a section of the oldest Roman road in the Iberian Peninsula highlights the importance of this site in facilitating military and commercial movements across the region. Architecturally, this site consists of a series of buildings that represent a mix of indigenous and Roman construction techniques: rectangular houses, with rationalized layouts, stone foundations, walls made of rammed earth, and roofs constructed using Roman tiles (Duran i Caixal & Mestres i Santacreu, 2011; Duran et al., 2015, 2017; Mestres Santacreu, 2003). Twelve newborns have been found in single burials, although some of them were in the same houses, being buried both in living rooms and metallurgical workshops. The age of the newborns was determined by Martirosyan et al. (2024) using dental histology, and Afonso et al. (2019) performed the anthropological study of the remains and also provided the genetic sex for some of them.



2.

THE RATIONALE AND OBJECTIVES OF THE THESIS

2. THE RATIONALE AND OBJECTIVES OF THE THESIS

2.1. The rationale of the thesis

The study of Iberian genetics is hindered by the scarcity of anthropological remains, primarily due to the prevalent funerary practice of cremation among the Iberians. This has resulted in a limited number of genetic studies, with few samples being analyzed and authenticated as belonging to Iberian culture. Moreover, with the advance of paleogenomic techniques and the possibility to retrieve genetic information from limited samples, the remains of newborns found buried under the houses of Iberian sites now represent valuable sources of biological information and could be used as a proxy of the Iberian population. However, to date, there are no systematic paleogenomic studies focused on Iberian sites from different tribes that include individuals that span the Early Iron Age to the Roman Conquest period.

Due to various constraints, such as sample availability and quality, conducting paleogenomic analyses on all individuals recovered from an archaeological site is not always feasible. However, whole-genome analysis can be combined with mtDNA analysis, which *per se* can provide valuable insights on social organization and female mobility. Since the number of mtDNA copies per cell is much higher than the observed for nuclear regions, it is more easily recoverable than nuclear DNA, and its analysis is faster and cheaper, allowing for the inclusion of more individuals in genetic analyses.

Nonetheless, analyzing mitochondrial DNA, especially from ancient samples with low-depth coverage, is complicated by several factors. These include the potential for contamination, the presence of heteroplasmy, and *post-mortem* DNA damage. While profiles

of mutational hotspots and heteroplasmy are well-established, a comprehensive profile of damage specific to aDNA samples sequenced using Next-Generation Sequencing (NGS) technologies is currently lacking.

Moreover, mtDNA characterization can be performed using different approaches. including shotgun sequencing, but also cheaper methods that rely on amplicon-based NGS implemented in commercial kits and further optimized for degraded samples. However, bioinformatic tools available are not fully adapted to handle the specificities of aDNA and amplicon-based NGS sequencing.

2.2. Objectives of the thesis

This thesis aims to provide new tools for ancient mitochondrial DNA analysis and to contribute significantly to our understanding of the Northeastern Iberian communities' genetic origins, evolution, and diversity.

The main goals of the thesis are:

1) To generate a pattern of damage for mitochondrial DNA to help interpret ancient mtDNA sequences.
2) To create an easy-to-use pipeline that enables the analysis of mtDNA sequences from degraded samples obtained through amplicon-based NGS methodologies.
3) To investigate the genetic origin and evolution of Northeastern Iberian communities using the newborns found buried under the houses as a proxy of the Iberian populations, more specifically:
    a. To provide a paleogenomic characterization of different Iberian sites to evaluate the genetic evolution of the local population over time, from the Early Iron Age to the Roman Conquest period, providing insights into genetic ancestry, admixture, and demographic dynamics in the Northeastern region of the Iberian Peninsula.
    b. To compare the mitochondrial diversity of Northeastern Iberian populations from different Iberian tribes with earlier and later groups from the same region, to trace the origins and distribution of specific mitochondrial haplogroups identified in the Iberian populations to understand their evolutionary trajectories, and to investigate maternal lineage diversity over time, with particular attention to signals of long-distance female mobility.



3.

MATERIAL AND METHODS

3. MATERIAL AND METHODS

The base material and the laboratory, bioinformatic, and data analysis methodologies used in each of the publications included in this doctoral thesis are described in detail in the corresponding Articles 1–4.



4.

RESULTS

STRUCTURE OF THE RESULTS

The results of this thesis have been structured into two main parts and four papers published or under consideration for publication in scientific journals:

Part 1: Development of New Tools for Ancient Mitochondrial DNA Analysis. Due to their methodological nature, the two crucial works on mtDNA tools are placed before the results of the main analysis.

- **Article 1** concentrates on generating a DNA damage profile for ancient human mitochondrial DNA. This chapter will involve the analysis of compiled mtDNA sequences from 427 ancient samples obtained by shotgun or capture techniques and sequenced with NGS technology. Present-day and UDG-treated ancient samples were also included and used to establish levels of damaged reads. The aim is to identify and characterize patterns of post-mortem damage, which will be crucial for authenticating and interpreting ancient mtDNA data.

Cuesta-Aguirre, D. R., Amor-Jimenez, C., Malgosa, A., & Santos, C. (2025). A *Post-Mortem* Molecular Damage Profile in the Ancient Human Mitochondrial DNA. Molecular Ecology Resources, 25(4), e14061. https://doi.org/10.1111/1755-0998.14061

- **Article 2** focuses on creating an amplicon-based pipeline for analyzing the mtDNA non-coding region from degraded samples. This chapter will detail the development and validation of the pipeline, including comparisons of existing tools and parameters to ensure its efficacy and user-friendliness. 148 high-quality ancient Spanish human samples encompassing various chronologies were used for that purpose. Moreover, 15 present-day samples were also included as a control for low-damage samples.

Cuesta-Aguirre, D. R., Malgosa, A., & Santos, C. (2024). An easy-to-use pipeline to analyze amplicon-based Next Generation Sequencing results of human mitochondrial DNA from degraded samples. PLOS ONE, 19(11), e0311115. https://doi.org/10.1371/journal.pone.0311115

Part 2: Genetic Analysis of Iberian Communities

- **Article 3** uses samples from 54 newborns from three northeastern Iberian sites to obtain both nuclear and mitochondrial DNA and tries to infer the population origin and evolution of the Iberians from the Early Iron Age to the Roman Conquest period. This chapter will integrate genomic data, generated by shotgun sequencing, with archaeological and historical contexts to provide a comprehensive understanding of Iberian population dynamics.

Cuesta-Aguirre, D. R., Roca-Rada, X., Di Biase, B., Vinueza-Espinosa, D. C., Sandoval-Ávila, C., Ravishankar, S., Molist Capella, N., Duran i Caixal, M., Mestres Santacreu, I., Junyent, E., Alonso, N., Martinez Majoral, J., Knudsen, A. K., Lanigan, L. T., Schroeder, H., Llamas, B., Malgosa, A., & Santos, C. The Genetic Landscape of Northeastern Iberian Communities from the Early to Late Iron Age. Under review in iScience since 09/04/2025 (ISCIENCE-D-25-05072). Available at SSRN: http://dx.doi.org/10.2139/ssrn.5211463

- **Article 4** focuses on the mitochondrial DNA diversity of Iberians during the Iron Age and explores the possible origins of the haplogroups identified. Mitochondrial genomic data was obtained for 20 additional Iberian newborns, distinct from those presented in Article 3, and for a Roman newborn. These data were merged with 41 previously published mtDNA profiles from unrelated Iberian individuals, including 27 Iberian newborns described in Article 3. Additionally, haplogroup information from other ancient individuals from the Northeast of the Iberian Peninsula, ranging from the Neolithic to the Visigothic periods, was used for comparative purposes. This chapter compares the mitochondrial diversity in different Iberian tribes and between prehistoric and early historic periods in the region, based on 256 individuals, and traces the origins of the haplogroups identified in the Iron Age Iberians. In doing so, it aims to investigate patterns of female mobility during the Iron Age in the Iberian territory.

Cuesta-Aguirre, D. R., Campoy-Caballero, M. R., Sandoval-Ávila, C., Busquets i Costa, C., Fàbregas i Espadaler, M., Sinner, A. G., de Prado, G., Molist Capella, N., Duran i Caixal, M., Mestres Santacreu, I., Alonso, N. Aluja, M. P., Malgosa, A., & Santos, C. Mitochondrial DNA diversity in Northeast Iberians during the Iron Age.

4.1. Article 1: A *Post-Mortem* Molecular Damage Profile in the Ancient Human Mitochondrial DNA

Authors

Daniel R. Cuesta-Aguirre, Cristina Amor-Jimenez, Assumpció Malgosa, Cristina Santos

Journal

Molecular Ecology Resources

Impact factor (JCR) and quartile ranking

5.5 and Q1

Date of publication

07 January 2025

DOI

https://doi.org/10.1111/1755-0998.14061

Addressed objectives

This first publication addresses the previously defined objective 1:

1) To generate a pattern of damage for mitochondrial DNA to help interpret ancient mtDNA sequences.

Supplementary data is available in "Article 1" at the following link: https://uab-my.sharepoint.com/:f:/g/personal/2103711_uab_cat/EhLQT7GxLs1GoowdsaXF7QYBF7DKCwofUOyqhbjFw4XWRw?e=3ddWqy

Check for updates

Molecular Ecology Resources

WILEY

MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE OPEN ACCESS

A *Post-Mortem* Molecular Damage Profile in the Ancient Human Mitochondrial DNA

Daniel R. Cuesta-Aguirre[1] | Cristina Amor-Jimenez[1,2] | Assumpció Malgosa[1] | Cristina Santos[1]

[1]Research Group in Biological Anthropology, Biological Anthropology Unit, Department of Animal Biology, Vegetal Biology and Ecology, Universitat Autònoma de Barcelona, Barcelona, Spain | [2]Centre de Recerca Ecològica i d'Aplicacions Forestals, Universitat Autònoma de Barcelona, Barcelona, Spain

Correspondence: Cristina Santos (cristina.santos@uab.cat)

Received: 9 January 2024 | **Revised:** 9 July 2024 | **Accepted:** 1 October 2024

Handling Editor: Brent Emerson

Funding: This research was supported by the Ministerio de Ciencia e Innovación (Ref. PGC2018-096666-B-100), and by the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya and the Fons Social Europeu. D.R.C.-A. is a PhD fellow from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) (2020 FI_B1 00641). C.S. and A.M. are members of the GREAB supported by the Generalitat de Catalunya (Ref. 2017 SGR 1630; 2021 SGR 00186).

Keywords: ancient DNA (aDNA) | damage hotspot | mitochondrial DNA (mtDNA) | molecular damage profile | post-mortem DNA damage

ABSTRACT

Mitochondrial DNA (mtDNA) analysis is crucial for understanding human population structure and genetic diversity. However, *post-mortem* DNA damage poses challenges, that make analysis difficult. DNA preservation is affected by environmental conditions which, among other factors, complicates the differentiation of endogenous variants from artefacts in ancient mtDNA mix profiles. This study aims to develop a molecular damage profile for ancient mtDNA that can become a useful tool in analysing mtDNA from ancient remains. A dataset of 427 whole genomes or capture of mtDNA sequences from individuals representing different historical periods and climatic regions was compiled from the ENA database. Present-day and UDG-treated ancient samples were also included and used to establish levels of damaged reads. Results indicated that samples from cold regions exhibited the lowest percentage of damaged reads, followed by arid, cold, tropical and temperate regions, with significant differences observed between cold and temperate regions. A global damage profile was generated, identifying 2933 positions (25% of the positions considered) with damage in more than 23.8% of the samples analysed, deemed as damage hotspots. Notably, 2856 of these hotspots had never been reported as damage or mutational hotspots, or heteroplasmic positions. Damage hotspot frequency by position was slightly higher in the non-coding region compared with the coding region. In conclusion, this study provides a molecular damage profile for ancient mtDNA analysis that is expected to be a valuable tool in the interpretation of mtDNA variation in ancient samples.

1 | Introduction

In living cells, the damage produced in DNA molecules is repaired by enzymatic mechanisms, thus maintaining the integrity of the genome (Lindahl 1993). However, these mechanisms stop working when the organism dies, exposing the genome to factors threatening its integrity through time, such as intracellular nucleases, microorganisms or environmental conditions.

When analysing ancient DNA (aDNA) samples, three types of damage can be observed: fragmentation of DNA molecules, which leads to a substantial reduction in size; lesions that block DNA replication; and nucleotide modifications that cause misincorporations during DNA polymerase reading (Dabney, Meyer, and Paabo 2013). The fragmentation and decay of endogenous DNA can quickly reduce fragments to lengths ranging from 40 to 500 bp (Pääbo 1989). Moreover, in skeletal remains, the average

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

fragment of really ancient specimens can be lower than 60 bp (Glocke and Meyer 2017; Noonan et al. 2006). Regarding miscoding lesions, hydrolytic deamination of cytosine (C) to uracil (U) is the major mechanism, being cytosine to thymine (T) the large majority of nucleotide substitutions in aDNA (Brotherton et al. 2007; Pääbo 1989; Stiller et al. 2006). Deamination of C to U leads polymerases to incorporate an adenine across from the uracil during DNA replication and, in turn, a thymine across from the adenine, causing a high frequency of C-to-T substitutions (Briggs et al. 2007; Dabney, Meyer, and Paabo 2013; Meyer et al. 2012). These substitutions tend to accumulate on both ends of aDNA fragments. In contrast, they decrease exponentially along the DNA molecule, where uracils rarely occur (Briggs et al. 2007). As aDNA is sensitive to Uracil-DNA-glycosylase (UDG), this has been commonly used as a treatment to remove such substitutions and improve DNA quality substantially (Briggs et al. 2009; Rohland et al. 2015). However, C-to-T substitutions are not the only kind of misincorporations that can be found, despite being the most common (Noonan et al. 2006).

DNA preservation is affected by environmental conditions such as temperature, humidity, amount of water percolation, salinity, pH, microbial growth and processes such as hydrolysis or oxidation (Dabney, Meyer, and Paabo 2013; Gutiérrez-García et al. 2014; Molak and Ho 2011; Olivieri et al. 2010; Poinar et al. 1996; Smith et al. 2001). In some favourable environmental conditions, certain processes such as tissue freezing or rapidly drying after death can inhibit the destruction of the organism's endogenous DNA (Dabney, Meyer, and Paabo 2013). In vitro experiments have estimated that DNA would survive a few hundred thousand years (Lindahl 1993), but there are several examples of DNA preservation in more ancient samples (Dabney, Meyer, and Paabo 2013; Herbig et al. 2016; Kjær et al. 2022; Meyer et al. 2014, 2016; Orlando et al. 2013; van der Valk et al. 2021). When studying DNA features indicative of damage, a strong positive correlation was found between cytosine deamination and age (Sawyer et al. 2012). This characteristic has been used since then as an indicator of antiquity. However, later was described that deamination depends on the thermal age, a combination of age and temperature, instead of just the age (Kistler et al. 2017).

For decades, the analysis of mitochondrial DNA (mtDNA) has become key in aDNA studies and population genetics research. Being more abundant than the nuclear genome, as there are hundreds to thousands of copies of mtDNA in each cell, makes mtDNA more easily recovered from ancient remains than nuclear DNA. Studies focused on mtDNA provide useful information to understand the appearance and dispersion of species and how social, cultural, and environmental factors have shaped human populations and their genetic diversity (Ho and Gilbert 2010; Juras et al. 2020; Modi et al. 2019; Poma et al. 2019). The mutational rate of mitochondrial regions has also been studied, resulting in the description of mutational hotspots (Soares et al. 2009)—positions that are more likely to present mutations than would be expected. Furthermore, the study of hotspots has also been extended to heteroplasmy, a condition in which an individual carries mtDNA molecules with different haplotypes (Li et al. 2015). However, *post-mortem* molecular DNA damage remains a challenge in mitochondrial aDNA data analysis due to the difficulty distinguishing it from mtDNA heteroplasmy, contamination, and DNA amplification and sequencing errors.

Taking all this into account, evaluating a mixed profile in ancient mtDNA becomes a very complex issue due to the difficulty of distinguishing true endogenous variants from artefacts. This is why the identification of *post-mortem* damage-derived hotspots is crucial for the precise interpretation of the base mixtures in low-depth coverage ancient samples. Although several studies have focused on identifying the amount or kind of post-mortem damage (Briggs et al. 2007; Brotherton et al. 2007; Olivieri et al. 2010; Sawyer et al. 2012), research has also been performed to identify *post-mortem* damage hotspots in mtDNA from plants (Staats et al. 2011) and animals, such as bison (Gilbert et al. 2005), Neanderthals (Vives et al. 2008) and humans (Gilbert et al. 2003; Lamers, Hayter, and Matheson 2009). Those studies performed on ancient humans (Gilbert et al. 2003; Lamers, Hayter, and Matheson 2009) were based on cloning technology to generate different sequences of the same region, and generally focused on mtDNA Hypervariable Region I. Gilbert et al. (2003) also analysed Cytochrome C oxidase subunit III genes. Using samples from < 50 individuals in each study, they demonstrated that *post-mortem* damage is not randomly distributed, finding some hotspots that had not been observed as hypermutable positions in modern datasets (Gilbert et al. 2003; Gilbert 2006). More recent studies have focused on the analysis of the non-coding region (NCR) of the mtDNA amplified by PCR and sequenced by Next Generation Sequencing (NGS) from artificially degraded samples (Holland et al. 2021; Rathbun et al. 2017). Rathbun et al. (2017) demonstrated that DNA damage was induced both passively through storage conditions and actively through accelerated hydrolysis. They found 27 position changes that appeared more than once, suggesting that these sites may be more susceptible to damage. Holland et al. (2021) evaluated the amount of damage generated in free-diluted aDNA and highlighted the importance of distinguishing between damage and heteroplasmy. In that study, 35 potential damage hotspots were found. Both studies considered all kinds of base changes instead of just C-to-T or G-to-A misincorporations.

As far as we are aware, there is no study focused on finding molecular damage hotspots throughout the whole mitogenome using NGS technology in humans. The main goal of this work is to develop a molecular damage profile of ancient human mtDNA, identifying damage hotspots, that can become a useful tool in the analysis of human mtDNA from ancient remains. All kinds of base changes are going to be included. Moreover, samples from different climatic regions will be analysed to understand if different conditions could affect the damage profile generated.

2 | Material and Methods

2.1 | Threshold of Damage and Sample Selection

Twenty-two present-day samples from the 1000 genomes project (The 1000 Genomes Project Consortium 2015), 89 ancient samples with partial-UDG-treatment (Antonio et al. 2019; Lipson et al. 2020; Olalde et al. 2018) and 12 ancient samples with total-UDG-treatment (Haak et al. 2015) were used as negative controls to set the threshold of damaged reads rate that may be observed in a non-UDG-treated ancient sample (Table S1). The frequency of damaged reads was determined by extracting the

17550998, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.14061 by Readcube (Labtiva Inc.), Wiley Online Library on [07/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

damaged reads with PMDtools v0.60 (Skoglund et al. 2017) as mentioned in the '*Post-mortem* Molecular Damage detection' section. PMDtools is a tool that uses the PMD score (PMDS). The PMDS value indicates the likelihood that a change is because of DNA degradation. For negative values of PMDS, PMDtools identifies and selects reads affected by *post-mortem* molecular damage, PCR errors, and sequencing misreads. However, for positive PMDS values, only changes caused by DNA degradation are identified. Higher values of PMDS imply fewer available reads.

A Kruskal–Wallis analysis was performed to compare the percentage of damaged reads in the negative controls. Moreover, the highest percentage of damaged reads was used as a minimum frequency of damage for a non-UDG-treated ancient sample to be considered authentic. This damage threshold using UDG-treated samples is conservative and some authentic samples may be excluded. However, this should not substantially affect the results of the present study.

To analyse the damage profile, 427 non-UDG-treated ancient samples, all derived from double-strand libraries, that were whole-genome sequenced or mtDNA-targeted captured (Table S2), were selected from previous publications (Allentoft et al. 2015; Damgaard et al. 2018; Ebenesersdóttir et al. 2018; Fregel et al. 2018; Lipson et al. 2018; Malaspinas et al. 2014; Margaryan et al. 2020; Mccoll et al. 2018; Moreno-Mayar et al. 2018; Sikora et al. 2019; Skoglund et al. 2017; Villa-Islas et al. 2023; Wright et al. 2018). All samples had at least 10000 positions with a depth coverage over 30, presented a mean frequency of contamination estimated with Schmutzi (Renaud et al. 2015) lower than 0.05, and reached the established damage threshold.

2.2 | *Post-Mortem* Molecular Damage Detection

Adapters were trimmed and forward and reverse reads were merged in the case of paired-end data, when the overlapped sequence was higher than 10 bases, using fastp v0.23.2 (Chen et al. 2018). *Post-mortem* molecular damage events—or damage hits, considered for all kinds of substitutions—were estimated independently for the non-coding region (NCR, 15988–16569 and 1–670 bp) and the mtDNA coding region (CR, 671–15987 bp). For CR, reads were aligned to a fusion of the whole human sequence (hg19, GenBank assembly GCA_000001405.14) with the revised Cambridge Reference Sequence (rCRS, GenBank accession NC_012920.1). For NCR, reads were aligned to a mix of the hg19 sequence and a linearised rCRS fragment from base 15901 to base 700. In both cases, alignment was made using BWA v0.7.17 (Li and Durbin 2009). By aligning reads with a mixture of the whole human genome and the mitogenome, nuclear mitochondrial DNA segments (NUMTs), which are mtDNA insertions into the nuclear chromosomes, were excluded. Reads with a mapping quality lower than 30 were removed with SAMtools v1.13 (Danecek et al. 2021). Duplicated reads were removed with MarkDuplicates, Picard v2.26.2 (Deharvengt et al. 2020).

The *post-mortem molecular* damage rate was estimated with PMDtools v0.60 (Skoglund et al. 2017) with a threshold (PMDS) of 1 just for the reads aligned with the whole rCRS after the duplicates were removed. Over the same file, the contamination detection with contDeam.pl. script from Schmutzi (Renaud et al. 2015) was performed. Samples with a contaminant frequency mean higher than 0.05 were discarded. For the included samples, the contaminant profile was estimated with schmutzi.pl. script (Renaud et al. 2015).

Freebayes v0.9.21.7 (Garrison and Marth 2012) and vcftools v0.1.16 (Danecek et al. 2011) software were used to identify changes to the reference genome. Only bases with a quality higher than 30 were considered. A change was considered a hit when it was detected in only one read out of a minimum of 30 non-duplicate reads. However, in those cases where the hit coincided with the sequence of the contaminant profile, hits were discarded, as probably belonged to the contaminant sequence instead of being damage. As a result of it, just damage hits were annotated.

Damage hits from the CR (671–15987 bp) were extracted from the file that aligned with the rCRS (1–16569 bp) and merged with the damage hits from the NCR (15988–16569 and 1–670 bp), obtained from the file that aligned just with the linearised fragment of the rCRS (15901–700).

Positions that could not be observed with a depth coverage over 30 in, at least, 75% of samples, were excluded from the analysis.

2.3 | mtDNA Damage by Climate Region

Non-UDG-treated ancient samples were classified according to their localisation to a climatic region (Figure 1), according to Köppen-Geiger climate classification (Peel, Finlayson, and McMahon 2007). Five main groups were considered: Polar, Cold, Temperate, Arid, and Tropical. A principal component analysis (PCA) by sample was computed with *prcomp* in R (R Core Team 2022) to analyse each sample damage behaviour and the possible clustering of samples based on climatic regions (Peel, Finlayson, and McMahon 2007) and age. Other characteristics, such as sample type, sequencing method, type of sequencer, country, region based on the United Nations Statics Division (UN), and the paper of precedence were also analysed. The presence (1) or absence (0) of a damage hit was indicated for each sample and all analysed mtDNA positions. For a position missing in a sample but present in at least 75% of all samples, the mean across all samples was used as the position's value for that sample.

Jamovi 2.2.5 software (The Jamovi project 2021) was used to perform statistical analysis and a significance level of 0.05 was considered.

2.4 | Profile of *Post-Mortem* Molecular Damage in the Samples and Damage Hotspots Determination

A global damage profile was obtained using all the non-UDG-treated ancient samples included in the analysis. The frequency of damage by position was computed by dividing the number of damage hits by the number of samples with a specific position available to analyse the distribution of damage in the samples.

17550998, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.14061 by Readcube (Labtiva Inc.), Wiley Online Library on [07/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

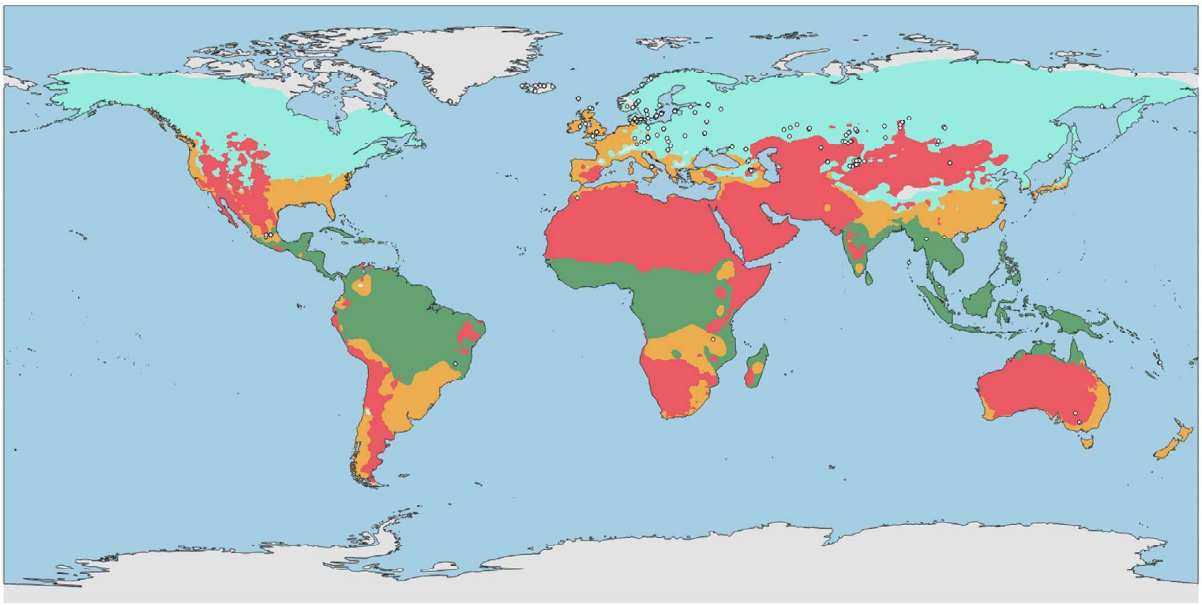


FIGURE 1 | Distribution of compiled samples and their climatic classification according to Peel, Finlayson, and McMahon (2007) based on Köppen–Geiger climate classification made with QGIS 3.30 (QGIS.org 2023). Sample locations are represented by white circles on the map, which is colour-coded to correspond with climatic classification: grey for Polar, blue for Cold, orange for Temperate, red for Arid and green for Tropical.

The 95% confidence interval (95% CI) for each position was determined with *exactci* R package (R Core Team 2022). Positions with a frequency of damage in the samples higher than the value of the 75th percentile were considered damage hotspots.

The results of the mtDNA non-coding region were analysed and compared with the positions previously reported as damage hotspots by other authors (Gilbert et al. 2003; Holland et al. 2021; Lamers, Hayter, and Matheson 2009; Rathbun et al. 2017). For the present study, all positions were included, but the frequency of damage in the samples and 95% CI were not calculated for positions present in < 75% of the samples.

Following Gilbert et al. (2003) (16055-16410) criteria, only positions with a 'Standardised Damage Rate' of 3 or 4 were considered as damage hotspots (Gilbert et al. 2003).

For positions reported by Lamers, Hayter, and Matheson 2009 (15950-429), the number of occurrences was counted and divided between the number of samples analysed to obtain the frequency of damage in the samples.

The authors' criteria were used for positions described by Rathbun et al. (2017) (16024-576). Those positions with more than one independent event were considered damage hotspots. The authors' criteria were also used for the positions described by Holland et al. (2021) (16024-570), where those positions with at least three independent damage occurrences, were considered hotspots.

Finally, for the whole mitochondrial molecule, mutational hotspots were compiled according to Soares et al. (2009), which considered a position as mutational hotspots if eight or more hits on the mtDNA phylogeny were observed. In addition, mtDNA positions presenting heteroplasmy were also recorded based on the results of Li et al. (2015).

Positions considered as damage hotspots in the present study were investigated to find if they were previously reported as damage hotspots (Gilbert et al. 2003; Holland et al. 2021; Rathbun et al. 2017), mutational hotspots (Soares et al. 2009) or heteroplasmic positions (Li et al. 2015). The rate of hotspots by mitochondrial region (non-coding region and coding region) was estimated by dividing the number of damage hotspots determined in the present study by the number of available positions. The 95% Confidence Intervals were estimated with *exactci* in R (R Core Team 2022).

3 | Results

3.1 | Threshold of Damage

Descriptive statistics for the frequency of the damaged reads based on PMDtools v0.60 for present-day samples, ancient samples treated with total-UDG-treatment and ancient samples treated with partial-UDG-treatment are shown in Table 1. Overall, there were no statistically significant differences in distributions between the sample categories for PMDS = 0 (Kruskal–Wallis, $\chi^2 = 0.298$; $p = 0.861$). For PMDS = 1, the highest median of damaged reads was found in samples with partial-UDG-treatment, while the lowest median was observed in present-day samples. In this case, partially treated ancient samples also had the highest value of all samples (12.45%). The distributions of the categories of samples were statistically significantly different (Kruskal–Wallis, $\chi^2 = 15.552$; $p < 0.001$). Post hoc tests on pairwise comparisons showed significant differences between present-day and total-UDG-treated ancient

17550998, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.14061 by Readcube (Labtiva Inc.), Wiley Online Library on [07/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | Descriptive statistics for the frequency of damaged reads based on PMDtools v0.60 for present-day samples, ancient samples with total-UDG-treatment, and ancient samples with partial-UDG-treatment.

					Percentiles			Shapiro–Wilk	
	N	Median	Minimum	Maximum	2.5th	25th	75th	97.5th	*W*	*p*
PMDS = 0										
Present-day	23	6.06	0.57	11.75	1.38	3.26	7.53	10. 61	0.97	0.735
Total-UDG-treatment	12	5.52	4.17	9.27	4.22	4.95	7.39	8.82	0.93	0.347
Partial-UDG-treatment	89	5.51	1.02	17.77	1.33	3.72	8.69	13.45	0.96	0.006
PMDS = 1										
Present-day	23	2.50	0.36	5.45	0.53	1.37	3.10	4.75	0.96	0.533
Total-UDG-treatment	12	3.41	2.03	5.43	2.16	2.54	3.93	5.11	0.95	0.619
Partial-UDG-treatment	89	4.02	0.71	12.45	0.78	2.41	6.32	10.33	0.95	0.002

TABLE 2 | Descriptive statistics for the frequency of damaged reads with a PMDS = 1 for the non-UDG-treated ancient samples divided by climatic region.

					Percentiles			Shapiro–Wilk	
Climatic region	*N*	Median	Minimum	Maximum	2.5th	25th	75th	97.5th	*W*	*p*
Polar	29	45.3	27.0	66.4	28.3	39.3	51.7	65.9	0.97	0.612
Cold	272	43.2	22.9	68.1	26.4	33.8	49.6	59.1	0.98	< 0.001
Temperate	89	47.7	26.9	84.8	28.6	39.7	54.0	62.7	0.96	0.009
Arid	29	44.4	28.9	60.2	30.9	38.6	50.1	57.2	0.99	0.981
Tropical	8	45.7	30.8	53.9	31.5	38.1	47.5	53.1	0.95	0.740

samples (Dwass–Steel–Critchlow–Fligner, $W = 3.59$; $p = 0.030$) and between present-day and partial-UDG-treated ancient samples (Dwass–Steel–Critchlow–Fligner, $W = 5.35$; $p < 0.001$).

Although significant differences between present-day and ancient UDG-treated samples were observed with PMDS = 1, this threshold was further used at PMDtools v0.60 to estimate the amount of damage of the non-UDG-treated ancient samples included to establish the damage profile. Moreover, a non-UDG-treated sample must have a percentage of damage higher than the maximum value detected for partial-UDG-treated samples (12.45%) to be considered authentic. Applying this criterion and the level of contamination estimated by Schmutzi, we finally included 427 samples, divided by climatic region, to establish the mtDNA damage profile (Tables S2 and S3, and referenced therein).

3.2 | mtDNA Damage by Climate Region

Table S3 displays the levels of damage and contamination for each sample included. Table 2 and Figure 2 summarise the percentage of damaged reads by climatic region. Overall, there were statistically significant differences in the distributions between climatic regions (Kruskal–Wallis, $\chi^2 = 14.4$; $p = 0.006$), presenting Cold samples (43.2%) the lowest median of damaged reads; followed by Arid samples (44.4%), Polar samples (45.3%), Tropical samples (45.7%) and Temperate samples (47.7%). Post hoc tests on pairwise comparisons showed significant differences only between Cold and Temperate samples (Dwass–Steel–Critchlow–Fligner, $W = 4.98$; $p = 0.004$).

The PCA based on positions presenting damage by sample did not show clusters of samples according to climatic region or age (Figure S1), indicating that the damage pattern is similar by type of climate region. When the rest of the variables that appear in Table S2 were studied, no pattern was observed either. Therefore, it seemed appropriate to obtain a damage profile based on all samples without any kind of distinction.

3.3 | *Post-Mortem* Molecular Damage Profile Generation and Damage Hotspot Determination

From the 16569 bp of the mitogenome, 11735 positions were present with a minimum depth coverage of 30× in, at least, 75% of the samples analysed and were the ones included in the damage profile. Table S4 displays, for each position, the number of

17550998, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.14061 by Readcube (Labtiva Inc.), Wiley Online Library on [07/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

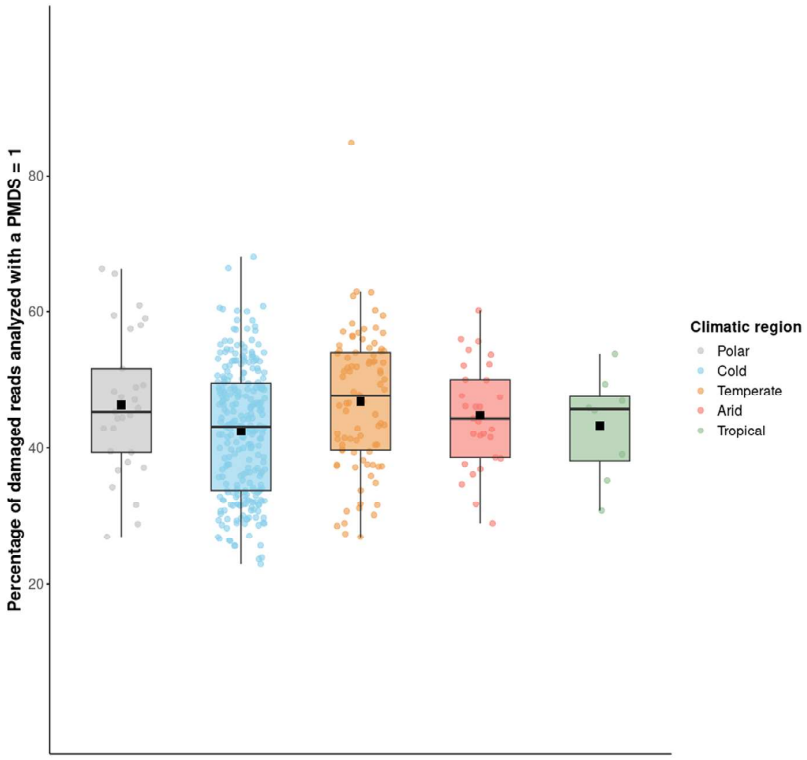


FIGURE 2 | Percentage of damaged reads analysed with a PMDS = 1 for each climatic region. Squares represented the mean. Each climatic region is represented by a distinct colour as indicated in the legend. The black square represents the mean.

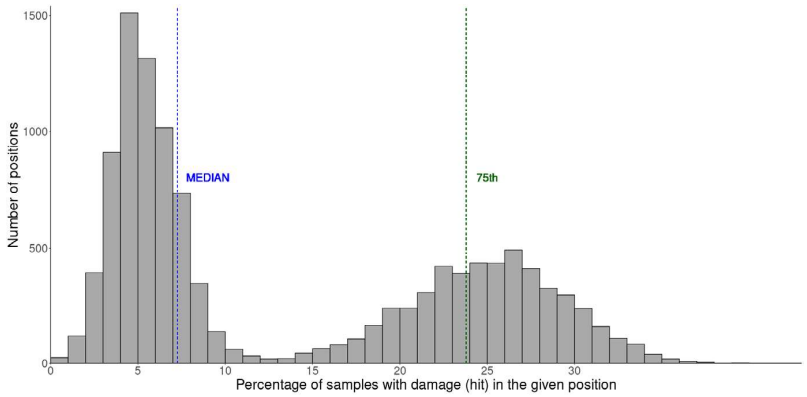


FIGURE 3 | Distribution of positions with a given percentage of damage in the samples (%). The plot shows the frequency of positions with varying percentages of damage. The median is indicated in blue, and the 75th percentile is highlighted in green.

damage hits detected in the samples analysed, the frequency of damage by position, and the respective 95% Confidence Intervals. The information is summarised in Figure 3, which displays the distribution of positions according to the percentage of samples with damage. Half of the positions presented a frequency of samples with damage under 7.5%. The distribution is bimodal, with a first peak between 3% and 7% of the samples with damage and a second peak between 19% and 30% of the samples with damage. Twenty-five per cent of the positions are damaged in more than 23.8% of the samples (Figure 3).

Considering only the NCR of the mtDNA, 87 positions were previously reportedly as potential damage hotspots by other authors (Gilbert et al. 2003; Holland et al. 2021; Rathbun et al. 2017) (Table S4). The majority of them were exclusive from a specific study; although some of them, such as positions 67, 189, 16265, 16390 and 16519 (Holland et al. 2021; Rathbun et al. 2017) were shared between studies. Of those 87 positions, 2 were absent from our final profile because of not reaching a depth coverage of 30× in at least 321 samples (16558 and 16566).

17550998, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.14061 by Readcube (Labtiva Inc.), Wiley Online Library on [07/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Out of the 11735 evaluated positions, 2933 of them (25%) presented damage in more than 23.8% of the samples and were classified as damage hotspots in the present study (Table S4). Among these positions, 2781 (Table S5) had not been previously reported as damage hotspots (Gilbert et al. 2003; Holland et al. 2021; Rathbun et al. 2017), mutational hotspots (Soares et al. 2009) or heteroplasmic positions (Li et al. 2015). However, 152 positions, indicated in Table 3, were previously documented. Specifically, 25 positions were previously reported as damage hotspots, 34 positions were already recognised as mutational hotspots, 64 positions were previously reported as heteroplasmic positions, 12 positions had been documented as both mutational hotspot and heteroplasmic positions; 9 positions were reported as both damage hotspot and mutational hotspot; 2 positions had been noted as both a damage hotspot and a heteroplasmic position; and 6 positions were previously documented as damage hotspot, mutational hotspot and heteroplasmic position (Figure 4).

Figure 5 displays the distribution of the 2933 damage hotspots reported in the present study across the mitogenome. 352 hotspots fit in the non-coding region (1182 bp available), representing a frequency of hotspots by base pair of 0.298 [95% CI: 0.272, 0.325]; and 2581 were in the coding region (10,553 bp available), being a frequency of 0.245 [95% CI: 0.236, 0.253]. Thus, it seems that the frequency of damage hotspots in the non-coding region is slightly higher.

4 | Discussion

Working with aDNA can be challenging for researchers, even when analysing mtDNA due to the samples' low depth coverage. Additionally, chemical changes in nucleotides increase the difficulty of analysis as molecular damage can hardly be distinguished from other processes like mutation, contamination or even sequencing errors. To overcome this challenge, several

TABLE 3 | Description of the damage hotspots found in the present study that were previously presented as damage hotspots, mutational hotspots, or heteroplasmic positions.

Definition	Number of positions	Positions
Considered a damage hotspot in the present study and by Gilbert et al. (2003), Holland et al. (2021) or Rathbun et al. (2017)	25	81, 107, 113, 162, 187, 389, 412, 496, 497, 569, 16072, 16085, 16118, 16179, 16218, 16232, 16262, 16328, 16348, 16401, 16414, 16460, 16472, 16473, 16520
Considered a damage hotspot in the present study and by Gilbert et al. (2003), Holland et al. (2021) or Rathbun et al. (2017) and considered a heteroplasmic position by Li et al. (2015)	2	67, 16496
Considered a damage hotspot in the present study and by Gilbert et al. (2003), Holland et al. (2021) or Rathbun et al. (2017) and considered a mutational hotspot by Soares et al. (2009)	9	16187, 16192, 16223, 16242, 16270, 16278, 16290, 16291, 16319
Considered a damage hotspot in the present study and by Gilbert et al. (2003), Holland et al. (2021) or Rathbun et al. (2017); considered a heteroplasmic position by Li et al. (2015); and considered a mutational hotspot by Soares et al. (2009)	6	143, 16148, 16261, 16294, 16327, 16519
Considered a damage hotspot in the present study and considered a heteroplasmic position by Li et al. (2015)	64	76, 94, 114, 186, 225, 343, 438, 627, 945, 957, 1263, 1339, 1393, 1428, 1858, 2128, 2697, 2698, 2735, 2819, 2826, 3002, 3016, 3068, 3070, 3438, 3454, 3496, 6243, 6382, 6384, 6564, 8141, 8533, 9300, 9921, 10360, 10401, 10589, 10971, 11018, 11536, 11852, 12618, 12818, 12980, 13192, 13230, 13466, 13531, 13711, 13718, 14259, 14459, 14855, 14930, 15173, 15575, 16083, 16108, 16201, 16296, 16337, 16537
Considered a damage hotspot in the present study and considered a mutational hotspot by Soares et al. (2009)	34	103, 150, 151, 198, 1719, 3010, 3531, 3591, 3915, 5460, 9053, 11151, 12406, 12771, 13708, 14560, 15043, 15314, 15355, 15884, 15928, 16145, 16169, 16176, 16184, 16185, 16193, 16234, 16239, 16260, 16287, 16295, 16320, 16526
Considered a damage hotspot in the present study; considered a heteroplasmic position by Li et al. (2015); and considered a mutational hotspot by Soares et al. (2009)	12	182, 185, 709, 11914, 12007, 16111, 16148, 16256, 16292, 16294, 16355, 16527

17550998, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.14061 by Readcube (Labtiva Inc.), Wiley Online Library on [07/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

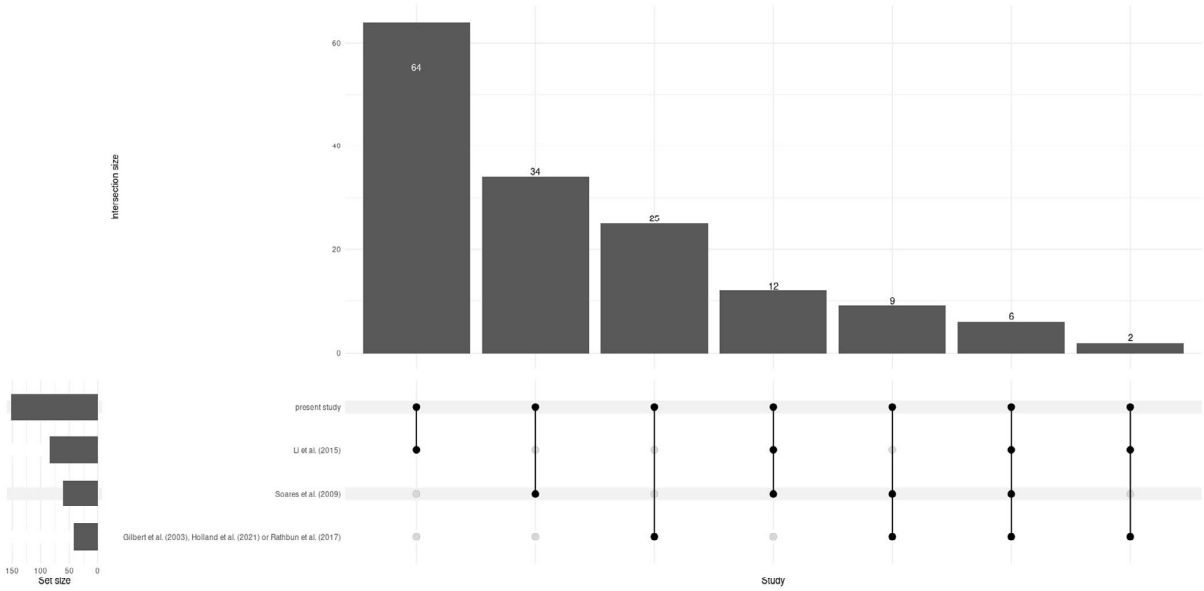


FIGURE 4 | Summary of the damage hotspots presented in the present study that overlapped with the hotspots of heteroplasmy described by Li et al. (2015), the mutational hotspots presented by Soares et al. (2009), and the previous damage hotspots reported by Gilbert et al. (2003), Holland et al. (2021) and Rathbun et al. (2017). The number of hotspots shared within each comparison can be seen in the vertical bar plot.

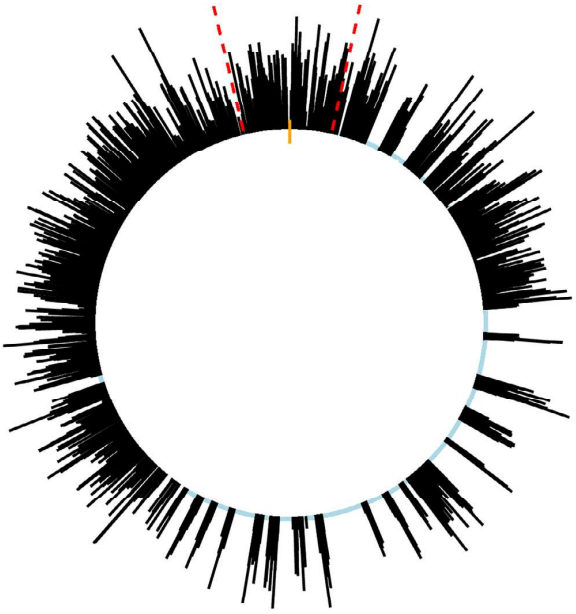


FIGURE 5 | Distribution of the 2933 hotspots identified in the present study. These hotspots exhibit a percentage exceeding 23.8% and collectively represent the top 25% of the positions with the highest values. The non-coding region is highlighted in red, while an orange line denotes the origin of replication. Region of the mitogenome with <30 reads covering them in at least 75% (321 samples) of those analysed are shaded in blue to indicate their absence in our profile.

studies have been conducted focused on the specific type of damage that occurs in such samples (Briggs et al. 2007; Brotherton et al. 2007; Olivieri et al. 2010; Sawyer et al. 2012). However, few studies have centred on mtDNA damage hotspots, and all were devoted to the mtDNA non-coding region (Gilbert et al. 2003; Holland et al. 2021; Lamers, Hayter, and Matheson 2009; Rathbun et al. 2017) or the cytochrome C oxidase subunit III gene (Gilbert et al. 2003). Thus, to date, no studies have analysed the molecular damage in all the mitogenomes obtained by Next-Generation Sequencing techniques of ancient samples. Our study provides a comprehensive view of mtDNA damage in the whole mtDNA molecule along with the probability of mutation and heteroplasmy. This is the first time such a comprehensive view has been presented.

Related to previous treatment, comparisons between partial-UDG-treated ancient samples, total-UDG-treated ancient samples and present-day samples showed, as expected, a higher amount of damage in partial-UDG-treated ancient samples than in total-UDG-treated samples, while the lowest amount of damage was observed in present-day samples. Changes found in present-day samples may have resulted from various factors such as unrepaired damage at the time of sample collection (Sun, Osterman, and Li 2019), oxidative reactions generated during extraction or other lab procedures (Xavier et al. 2018), storage conditions (Holland et al. 2021; Rathbun et al. 2017), or sequencing errors (Stoler and Nekrutenko 2021). Differences between total-UDG-treated ancient samples and present-day samples could be derived from the inability of total-UDG-treatment to eliminate the damage that is not at the end of the fragments (Rohland et al. 2015).

With reference to the physical and biological environment, DNA is better preserved in cool, dry, dark, anaerobic and alkaline conditions (Bollongino, Tresset, and Vigne 2008). In accordance, when we compared differences between the frequency of damaged reads by climatic region, the expected trend can

17550998, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.14061 by Readcube (Labtiva Inc.), Wiley Online Library on [07/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

be observed, with the lowest damage in cold and arid regions and the highest values of damage in tropical and temperate samples. However, differences were significant only between Temperate and Cold samples. This could be explained by the sample size (Faber and Fonseca 2014) as we had more than 100 samples for Cold regions, but we barely had 30 samples for Polar and Arid and <10 for Tropical regions. It also must be considered that we simplified the Köppen–Geiger climate classification to have more samples in each group (Peel, Finlayson, and McMahon 2007).

In addition to the climatic region, other factors can affect aDNA preservation and the amount of DNA damage (Bollongino, Tresset, and Vigne 2008; Dabney, Meyer, and Paabo 2013; Gutiérrez-García et al. 2014; Molak and Ho 2011; Olivieri et al. 2010; Poinar et al. 1996; Smith et al. 2001), namely, age (Kistler et al. 2017), climatic causes (climatic characteristics of the region, season when the body was buried and climatic conditions when the body was buried), physicochemical causes (type of ground, amount of humidity and amount of oxygen), biological causes (amount of post-mortal organic substances in the soil classified as humic and fulvic acids, modified by the size of the body or the presence of more than one individual in the same burial and degree of microorganisms presents in the soil) (Burger et al. 1999), anthropic causes (type of burial, presence of grave goods, spatial location of grave in the site), or post-excavation causes, as storage of the sample (Pruvost et al. 2007), or choice of bone (Parker et al. 2020; Vinueza-Espinosa et al. 2020).

When we compared our results for the NCR with previously reported damage hotspots, two positions were absent in our profile (16558 and 16566). Positions from 16551 to 16569 and from 1 to 18 are absent in our profile since when aligning a circular molecule (Nass 1966), some reads can align with the first and last bases, but without enough length to pass quality controls in any of the two regions (Ayad and Pissis 2017). We tried to solve this problem by doing a double alignment: a common linearised alignment with the whole mitogenome and a small alignment with a region between bases 15988 and 670 to substitute these results for the obtained in the whole mitogenome alignment. However, as we compiled data from different sources, we were not always able to access the raw data that allowed us to succeed in the double alignment.

Out of the 87 positions previously reported as damage hotspots (Gilbert et al. 2003; Holland et al. 2021; Rathbun et al. 2017) that were available on our profile, only 42 positions were considered damage hotspots in the present study. However, two of those positions had also been reported as heteroplasmic positions (Li et al. 2015), nine had been previously reported as mutational hotspots (Soares et al. 2009) and six had been previously reported as mutational hotspots and heteroplasmic positions (Li et al. 2015; Soares et al. 2009). There could be several explanations for these differences. First, the studies performed by Gilbert et al. (2003) and Lamers, Hayter, and Matheson (2009) were carried out with cloning technology and used different depth coverages. By contrast, the study performed by Rathbun et al. (2017), carried out with NGS technology after amplification based on PCR, only considered damage positions with depth coverage of 4000× (minimum alternate frequency of 1% present, at least, on 40 reads). The same criteria were used by Holland et al. (2021), which employed the PowerSeq CRM Nested System kit from Promega (Madison, WI). However, those two studies were based on modern DNA and did not remove either the PCR or the optical duplicates, reaching higher depth coverage compared to aDNA. Even with enrichment or capture techniques, it is not possible to obtain high-depth coverage for mtDNA with aDNA. For example, enrichment was used by Olalde et al. (2019) obtaining a mean depth coverage for the mtDNA of 141.58× and a median of 76.95×, with a minimum of 0.71× and a maximum of 1430×. Secondly, these two studies compared the damage in extracted modern DNA by storing it at different temperatures, for different times (weeks), and resuspending it in different buffers (water and low-TE). Notwithstanding, the authors did not take into account that the great amount of damage in ancient samples happens in the first weeks when the body is decomposing due to inner and outer microorganisms, which generates an acid environment (Latham and Miller 2019). Moreover, the amount of oxygen in the buried site affects the oxidative processes (Janaway, Percival, and Wilson 2009). These differences in the environment surrounding the DNA could explain the differences obtained in our study. Finally, the differences reported in various studies could also be explained by the sample size. While we used at least 321 samples for all positions, Gilbert et al. (2003) used 43 teeth from 34 archaeological human skeletons; Lamers, Hayter, and Matheson (2009) used between 6 and 40 samples; Rathbun et al. (2017) used a set of 54 samples from different extractions from only 3 donors; and Holland et al. (2021) used 240 samples from dilutions from only 10 donors.

Some of the damage hotspots that we reported had previously been identified as mutational hotspots, heteroplasmic positions, or both (Li et al. 2015; Soares et al. 2009). Gilbert et al. (2003) found similar results when comparing with modern datasets (Excoffier and Yang 1999; Meyer, Weiss, and von Haeseler 1999), and eight out of the 27 positions reported by Rathbun et al. (2017) were known to be heteroplasmy hotspots (Irwin et al. 2009). Nonetheless, our results show that 2781 positions that we classified as damage hotspots had not been previously described as mutational hotspots or as heteroplasmic positions. These changes probably originated from damage that affected the whole mtDNA molecule, although we cannot discard sequencing errors (Peck et al. 2018). The differences between non-coding and coding regions can be related to the GC content variation, which is 45.73% in the non-coding region against 44.24% of GC in the coding region.

In summary, we generated a *post-mortem* molecular damage profile indicating if positions had ever been reported as damage hotspots, mutational hotspots, or heteroplasmic positions. However, we were unable to control sequencing errors, which means that some of the positions reported as damage hotspots may be a result of sequencing errors. This profile can be useful to solve doubts on ultra-low depth coverage positions when analysing mtDNA from ancient samples.

Author Contributions

C.S. and A.M. conceived the research project, that was designed by C.S. and D.R.C.-A.; and performed and analysed by D.R.C.-A. and C.A.-J. D.R.C.-A. and C.A.-J. compiled data from the ENA database and the

17550998, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.14061 by Readcube (Labtiva Inc.), Wiley Online Library on [07/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1000 Genomes Project. The manuscript was written by D.R.C.-A. with substantial contributions from C.S., C.A.-J., and A.M. C.A.-J. contributed significantly to the introduction. All authors actively participated in the manuscript review, providing comments and feedback, and subsequently approved the final manuscript.

Acknowledgements

This research was supported by the Ministerio de Ciencia e Innovación (Ref. PGC2018-096666-B-100), and by the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya and the Fons Social Europeu. D.R.C.-A. is a PhD fellow from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) (2020 FI_B1 00641). C.S. and A.M. are members of the GREAB supported by the Generalitat de Catalunya (Ref. 2017 SGR 1630; 2021 SGR 00186).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All samples used in the present study were compiled from bibliography. All information regarding the samples is provided in Tables S1 and S2. BAM files for ancient samples can be accessed through the European Nucleotide Archive (ENA) using the identifiers listed in the Supporting Information. Data for present-day samples can be retrieved from the International Genome Sample Resource (IGSR) using the corresponding identifiers provided.

References

Allentoft, M. E., M. Sikora, K. G. Sjögren, et al. 2015. "Population Genomics of Bronze Age Eurasia." *Nature* 522, no. 7555: 167–172. https://doi.org/10.1038/nature14507.

Antonio, M. L., Z. Gao, H. M. Moots, et al. 2019. "Ancient Rome: A Genetic Crossroads of Europe and the Mediterranean." *Science* 366, no. 6466: 708–714. https://doi.org/10.1126/science.aay6826.

Ayad, L. A. K., and S. P. Pissis. 2017. "MARS: Improving Multiple Circular Sequence Alignment Using Refined Sequences." *BMC Genomics* 18, no. 1: 86. https://doi.org/10.1186/s12864-016-3477-5.

Bollongino, R., A. Tresset, and J.-D. Vigne. 2008. "Environment and Excavation: Pre-Lab Impacts on Ancient DNA Analyses." *Comptes Rendus Palevol* 7, no. 2–3: 91–98. https://doi.org/10.1016/j.crpv.2008.02.002.

Briggs, A. W., U. Stenzel, P. L. F. Johnson, et al. 2007. "Patterns of Damage in Genomic DNA Sequences From a Neandertal." *Proceedings of the National Academy of Sciences of the United States of America* 104, no. 37: 14616–14621.

Briggs, A. W., U. Stenzel, M. Meyer, J. Krause, M. Kircher, and S. Pääbo. 2009. "Removal of Deaminated Cytosines and Detection of In Vivo Methylation in Ancient DNA." *Nucleic Acids Research* 38, no. 6: 1–12. https://doi.org/10.1093/nar/gkp1163.

Brotherton, P., P. Endicott, J. J. Sanchez, et al. 2007. "Novel High-Resolution Characterization of Ancient DNA Reveals C > U-Type Base Modification Events as the Sole Cause of Post Mortem Miscoding Lesions." *Nucleic Acids Research* 35, no. 17: 5717–5728. https://doi.org/10.1093/nar/gkm588.

Burger, J., S. Hummel, B. Herrmann, and W. Henke. 1999. "DNA Preservation: A Microsatellite-DNA Study on Ancient Skeletal Remains." *Electrophoresis* 20, no. 8: 1722–1728. https://doi.org/10.1002/(SICI)1522-2683(19990101)20:8<1722::AID-ELPS1722>3.0.CO;2-4.

Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. "Fastp: An Ultra-Fast All-In-One FASTQ Preprocessor." *Bioinformatics* 34, no. 17: i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Dabney, J., M. Meyer, and S. Paabo. 2013. "Ancient DNA Damage." *Cold Spring Harbor Perspectives in Biology* 5, no. 7: a012567. https://doi.org/10.1101/cshperspect.a012567.

Damgaard, P. D. B., N. Marchi, S. Rasmussen, et al. 2018. "137 Ancient Human Genomes From Across the Eurasian Steppes." *Nature* 557, no. 7705: 369–374. https://doi.org/10.1038/s41586-018-0094-2.

Danecek, P., A. Auton, G. Abecasis, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27, no. 15: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Danecek, P., J. K. Bonfield, J. Liddle, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10, no. 2: 1–4. https://doi.org/10.1093/gigascience/giab008.

Deharvengt, S. J., L. M. Petersen, H.-S. Jung, and G. J. Tsongalis. 2020. "Nucleic Acid Analysis in the Clinical Laboratory." In *Contemporary Practice in Clinical Chemistry*, 215–234. London, UK: Elsevier. https://doi.org/10.1016/B978-0-12-815499-1.00013-2.

Ebenesersdóttir, S. S., M. Sandoval-Velasco, E. D. Gunnarsdóttir, et al. 2018. "Ancient Genomes From Iceland Reveal the Making of a Human Population." *Science* 360, no. 6392: 1028–1032. https://doi.org/10.1126/science.aar2625.

Excoffier, L., and Z. Yang. 1999. "Substitution Rate Variation Among Sites in Mitochondrial Hypervariable Region I of Humans and Chimpanzees." *Molecular Biology and Evolution* 16, no. 10: 1357–1368. https://doi.org/10.1093/oxfordjournals.molbev.a026046.

Faber, J., and L. M. Fonseca. 2014. "How Sample Size Influences Research Outcomes." *Dental Press Journal of Orthodontics* 19, no. 4: 27–29. https://doi.org/10.1590/2176-9451.19.4.027-029.ebo.

Fregel, R., F. L. Méndez, Y. Bokbot, et al. 2018. "Ancient Genomes From North Africa Evidence Prehistoric Migrations to the Maghreb From Both the Levant and Europe." *Proceedings of the National Academy of Sciences of the United States of America* 115, no. 26: 6774–6779. https://doi.org/10.1073/pnas.1800851115.

Garrison, E., and G. Marth. 2012. "Haplotype-Based Variant Detection From Short-Read Sequencing." http://arxiv.org/abs/1207.3907.

Gilbert, M. T. P. 2006. "Postmortem Damage of Mitochondrial DNA." In *Human Mitochondrial DNA and the Evolution of Homo sapiens*, vol. 18, 91–115. Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-31789-9_5.

Gilbert, M. T. P., B. Shapiro, A. Drummond, and A. Cooper. 2005. "Post-Mortem DNA Damage Hotspots in Bison (*Bison bison*) Provide Evidence for Both Damage and Mutational Hotspots in Human Mitochondrial DNA." *Journal of Archaeological Science* 32, no. 7: 1053–1060. https://doi.org/10.1016/j.jas.2005.02.006.

Gilbert, M. T. P., E. Willerslev, A. J. Hansen, et al. 2003. "Distribution Patterns of Postmortem Damage in Human Mitochondrial DNA." *American Journal of Human Genetics* 72, no. 1: 32–47. https://doi.org/10.1086/345378.

Glocke, I., and M. Meyer. 2017. "Extending the Spectrum of DNA Sequences Retrieved From Ancient Bones and Teeth." *Genome Research* 27, no. 7: 1230–1237. https://doi.org/10.1101/gr.219675.116.

Gutiérrez-García, T., E. Vázquez-Domínguez, J. Arroyo-Cabrales, et al. 2014. "Evolutionary Biology Ancient DNA and the Tropics: A Rodent's Tale." *Biology Letters* 10, no. 6: 20140224. https://doi.org/10.1098/rsbl.2014.0224.

Haak, W., I. Lazaridis, N. Patterson, et al. 2015. "Massive Migration From the Steppe Was a Source for Indo-European Languages in Europe." *Nature* 522, no. 7555: 207–211. https://doi.org/10.1038/nature14317.

Herbig, A., F. Maixner, K. I. Bos, A. Zink, J. Krause, and D. H. Huson. 2016. "MALT: Fast Alignment and Analysis of Metagenomic DNA Sequence Data Applied to the Tyrolean Iceman." *BioRxiv*. https://doi.org/10.1101/050559.

17550998, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.14061 by Readcube (Labtiva Inc.), Wiley Online Library on [07/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Ho, S. Y. W., and M. T. P. Gilbert. 2010. "Ancient Mitogenomics." *Mitochondrion* 10, no. 1: 1–11. https://doi.org/10.1016/j.mito.2009.09.005.

Holland, C. A., J. A. McElhoe, S. Gaston-Sanchez, and M. M. Holland. 2021. "Damage Patterns Observed in mtDNA Control Region MPS Data for a Range of Template Concentrations and When Using Different Amplification Approaches." *International Journal of Legal Medicine* 135, no. 1: 91–106. https://doi.org/10.1007/s00414-020-02410-0.

Irwin, J. A., J. L. Saunier, H. Niederstätter, et al. 2009. "Investigation of Heteroplasmy in the Human Mitochondrial DNA Control Region: A Synthesis of Observations From More Than 5000 Global Population Samples." *Journal of Molecular Evolution* 68, no. 5: 516–527. https://doi.org/10.1007/s00239-009-9227-4.

Janaway, R. C., S. L. Percival, and A. S. Wilson. 2009. "Decomposition of Human Remains." In *Microbiology and Aging*, edited by S. L. Percival, 313–334. Totowa, New Jersey: Humana Press. https://doi.org/10.1007/978-1-59745-327-1_14.

Juras, A., P. Makarowicz, M. Chyleński, et al. 2020. "Mitochondrial Genomes From Bronze Age Poland Reveal Genetic Continuity From the Late Neolithic and Additional Genetic Affinities With the Steppe Populations." *American Journal of Physical Anthropology* 172, no. 2: 176–188.

Kistler, L., R. Ware, O. Smith, M. Collins, and R. G. Allaby. 2017. "A New Model for Ancient DNA Decay Based on Paleogenomic Meta-Analysis." *Nucleic Acids Research* 45, no. 11: 6310–6320. https://doi.org/10.1093/nar/gkx361.

Kjær, K. H., M. Winther Pedersen, B. De Sanctis, et al. 2022. "A 2-Million-Year-Old Ecosystem in Greenland Uncovered by Environmental DNA." *Nature* 612, no. 7939: 283–291. https://doi.org/10.1038/s41586-022-05453-y.

Lamers, R., S. Hayter, and C. D. Matheson. 2009. "Postmortem Miscoding Lesions in Sequence Analysis of Human Ancient Mitochondrial DNA." *Journal of Molecular Evolution* 68, no. 1: 40–55. https://doi.org/10.1007/s00239-008-9184-3.

Latham, K. E., and J. J. Miller. 2019. "DNA Recovery and Analysis From Skeletal Material in Modern Forensic Contexts." *Forensic Sciences Research* 4, no. 1: 51–59. https://doi.org/10.1080/20961790.2018.1515594.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform." *Bioinformatics* 25, no. 14: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, M., R. Schröder, S. Ni, B. Madea, and M. Stoneking. 2015. "Extensive Tissue-Related and Allele-Related mtDNA Heteroplasmy Suggests Positive Selection for Somatic Mutations." *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 8: 2491–2496. https://doi.org/10.1073/pnas.1419651112.

Lindahl, T. 1993. "Instability and Decay of the Primary Structure of DNA." *Nature* 362, no. 6422: 709–715. https://doi.org/10.1038/362709a0.

Lipson, M., I. Ribot, S. Mallick, et al. 2020. "Ancient West African Foragers in the Context of African Population History." *Nature* 577, no. 7792: 665–670. https://doi.org/10.1038/s41586-020-1929-1.

Lipson, M., P. Skoglund, M. Spriggs, et al. 2018. "Population Turnover in Remote Oceania Shortly After Initial Settlement." *Current Biology* 28, no. 7: 1157–1165.e7. https://doi.org/10.1016/j.cub.2018.02.051.

Malaspinas, A. S., O. Lao, H. Schroeder, et al. 2014. "Two Ancient Human Genomes Reveal Polynesian Ancestry Among the Indigenous Botocudos of Brazil." *Current Biology* 24, no. 21: R1035–R1037. https://doi.org/10.1016/j.cub.2014.09.078.

Margaryan, A., D. J. Lawson, M. Sikora, et al. 2020. "Population Genomics of the Viking World." *Nature* 585, no. 7825: 390–396. https://doi.org/10.1038/s41586-020-2688-8.

Mccoll, H., F. Racimo, L. Vinner, et al. 2018. "The Prehistoric Peopling of Southeast Asia." *Science* 361, no. 6397: 88–92. https://doi.org/10.1126/science.aat3628.

Meyer, M., J.-L. Arsuaga, C. de Filippo, et al. 2016. "Nuclear DNA Sequences From the Middle Pleistocene Sima de los Huesos Hominins." *Nature* 531, no. 7595: 504–507. https://doi.org/10.1038/nature17405.

Meyer, M., Q. Fu, A. Aximu-Petri, et al. 2014. "A Mitochondrial Genome Sequence of a Hominin From Sima de los Huesos." *Nature* 505, no. 7483: 403–406. https://doi.org/10.1038/nature12788.

Meyer, M., M. Kircher, M.-T. Gansauge, et al. 2012. "A High-Coverage Genome Sequence From an Archaic Denisovan Individual." *Science* 338, no. 6104: 222–226. https://doi.org/10.1126/science.1224344.

Meyer, S., G. Weiss, and A. von Haeseler. 1999. "Pattern of Nucleotide Substitution and Rate Heterogeneity in the Hypervariable Regions I and II of Human mtDNA." *Genetics* 152, no. 3: 1103–1110. https://doi.org/10.1093/genetics/152.3.1103.

Modi, A., D. Nesheva, S. Sarno, et al. 2019. "Ancient Human Mitochondrial Genomes From Bronze Age Bulgaria: New Insights Into the Genetic History of Thracians." *Scientific Reports* 9, no. 1: 1–11. https://doi.org/10.1038/s41598-019-41945-0.

Molak, M., and S. Y. W. Ho. 2011. "Evaluating the Impact of Post-Mortem Damage in Ancient DNA: A Theoretical Approach." *Journal of Molecular Evolution* 73, no. 3–4: 244–255. https://doi.org/10.1007/s00239-011-9474-z.

Moreno-Mayar, J. V., L. Vinner, P. De Barros Damgaard, et al. 2018. "Early Human Dispersals Within the Americas." *Science* 362, no. 6419: eaav2621. https://doi.org/10.1126/science.aav2621.

Nass, M. M. 1966. "The Circularity of Mitochondrial DNA." *Proceedings of the National Academy of Sciences of the United States of America* 56, no. 4: 1215–1222. https://doi.org/10.1073/pnas.56.4.1215.

Noonan, J. P., G. Coop, S. Kudaravalli, et al. 2006. "Sequencing and Analysis of Neanderthal Genomic DNA." *Science* 314, no. 5802: 1113–1118. https://doi.org/10.1126/science.1131412.

Olalde, I., S. Brace, M. E. Allentoft, et al. 2018. "The Beaker Phenomenon and the Genomic Transformation of Northwest Europe." *Nature* 555, no. 7695: 190–196. https://doi.org/10.1038/nature25738.

Olalde, I., S. Mallick, N. Patterson, et al. 2019. "The Genomic History of the Iberian Peninsula Over the Past 8000 Years." *Science* 363, no. 6432: 1230–1234. https://doi.org/10.1126/science.aav4040.

Olivieri, C., L. Ermini, E. Rizzi, et al. 2010. "Characterization of Nucleotide Misincorporation Patterns in the Iceman's Mitochondrial DNA." *PLoS One* 5, no. 1: e8629. https://doi.org/10.1371/journal.pone.0008629.

Orlando, L., A. Ginolhac, G. Zhang, et al. 2013. "Recalibrating Equus Evolution Using the Genome Sequence of an Early Middle Pleistocene Horse." *Nature* 499, no. 7456: 74–78. https://doi.org/10.1038/nature12323.

Pääbo, S. 1989. "Ancient DNA: Extraction, Characterization, Molecular Cloning, and Enzymatic Amplification." *Proceedings of the National Academy of Sciences of the United States of America* 86: 1939–1943.

Parker, C., A. B. Rohrlach, S. Friederich, et al. 2020. "A Systematic Investigation of Human DNA Preservation in Medieval Skeletons." *Scientific Reports* 10, no. 1: 18225. https://doi.org/10.1038/s41598-020-75163-w.

Peck, M. A., K. Sturk-Andreaggi, J. T. Thomas, R. S. Oliver, S. Barritt-Ross, and C. Marshall. 2018. "Developmental Validation of a Nextera XT Mitogenome Illumina MiSeq Sequencing Method for High-Quality Samples." *Forensic Science International: Genetics* 34, no. January: 25–36. https://doi.org/10.1016/j.fsigen.2018.01.004.

Peel, M. C., B. L. Finlayson, and T. A. McMahon. 2007. "Updated World Map of the Köppen-Geiger Climate Classification." *Hydrology and*

17550998, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.14061 by Readcube (Labtiva Inc.), Wiley Online Library on [07/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Earth System Sciences 11, no. 5: 1633–1644. https://doi.org/10.5194/hess-11-1633-2007.

Poinar, H. N., M. Höss, J. L. Bada, and S. Pääbo. 1996. "Amino Acid Recemization and the Preservation of Ancient DNA." *Science* 22, no. 5263: 864–866.

Poma, A., P. Cesare, A. Bonfigli, et al. 2019. "Analysis of Ancient mtDNA From the Medieval Archeological Site of Amiternum (L'Aquila), central Italy." *Heliyon* 5, no. 10: e02586. https://doi.org/10.1016/j.heliyon.2019.e02586.

Pruvost, M., R. Schwarz, V. B. Correia, et al. 2007. "Freshly Excavated Fossil Bones Are Best for Amplification of Ancient DNA." *Proceedings of the National Academy of Sciences of the United States of America* 104, no. 3: 739–744. https://doi.org/10.1073/pnas.0610257104.

QGIS.org. 2023. *QGIS 3.30 Geographic Information System*. QGIS Association. http://www.qgis.org.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org.

Rathbun, M. M., J. A. McElhoe, W. Parson, and M. M. Holland. 2017. "Considering DNA Damage When Interpreting mtDNA Heteroplasmy in Deep Sequencing Data." *Forensic Science International: Genetics* 26: 1–11. https://doi.org/10.1016/j.fsigen.2016.09.008.

Renaud, G., V. Slon, A. T. Duggan, and J. Kelso. 2015. "Schmutzi: Estimation of Contamination and Endogenous Mitochondrial Consensus Calling for Ancient DNA." *Genome Biology* 16, no. 1: 1–19. https://doi.org/10.1186/s13059-015-0776-0.

Rohland, N., E. Harney, S. Mallick, S. Nordenfelt, and D. Reich. 2015. "Partial Uracil – DNA – Glycosylase Treatment for Screening of Ancient DNA." *Philosophical Transactions of the Royal Society B* 370: 20130624. https://doi.org/10.1098/rstb.2013.0624.

Sawyer, S., J. Krause, K. Guschanski, V. Savolainen, and S. Pääbo. 2012. "Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA." *PLoS One* 7, no. 3: e34131. https://doi.org/10.1371/journal.pone.0034131.

Sikora, M., V. V. Pitulko, V. C. Sousa, et al. 2019. "The Population History of Northeastern Siberia Since the Pleistocene." *Nature* 570, no. 7760: 182–188. https://doi.org/10.1038/s41586-019-1279-z.

Skoglund, P., J. C. Thompson, M. E. Prendergast, et al. 2017. "Reconstructing Prehistoric African Population Structure." *Cell* 171, no. 1: 59–71.e21. https://doi.org/10.1016/j.cell.2017.08.049.

Smith, C. I., A. T. Chamberlain, M. S. Riley, A. Cooper, C. B. Stringer, and M. J. Collins. 2001. "Neanderthal DNA Not Just Old but Old and Cold?" *Nature* 410, no. April: 771–772.

Soares, P., L. Ermini, N. Thomson, et al. 2009. "Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock." *American Journal of Human Genetics* 84, no. 6: 740–759. https://doi.org/10.1016/j.ajhg.2009.05.001.

Staats, M., A. Cuenca, J. E. Richardson, et al. 2011. "DNA damage in plant herbarium tissue." *PLoS One* 6, no. 12: e28448. https://doi.org/10.1371/journal.pone.0028448.

Stiller, M., R. E. Green, M. Ronan, et al. 2006. "Patterns of Nucleotide Misincorporations During Enzymatic Amplification and Direct Large-Scale Sequencing of Ancient DNA." *Proceedings of the National Academy of Sciences of the United States of America* 103, no. 3: 1358–13584.

Stoler, N., and A. Nekrutenko. 2021. "Sequencing Error Profiles of Illumina Sequencing Instruments." *NAR Genomics and Bioinformatics* 3, no. 1: 1–9. https://doi.org/10.1093/nargab/lqab019.

Sun, S., M. D. Osterman, and M. Li. 2019. "Tissue Specificity of DNA Damage Response and Tumorigenesis." *Cancer Biology & Medicine* 16, no. 3: 396–414. https://doi.org/10.20892/j.issn.2095-3941.2019.0097.

The 1000 Geomes Project Consortium. 2015. "A Global Reference for Human Genetic Variation." *Nature* 526, no. 7571: 68–74. https://doi.org/10.1038/nature15393.

van der Valk, T., P. Pečnerová, D. Díez-del-Molino, et al. 2021. "Million-Year-Old DNA Sheds Light on the Genomic History of Mammoths." *Nature* 591, no. 7849: 265–269. https://doi.org/10.1038/s41586-021-03224-9.

Villa-Islas, V., A. Izarraras-Gomez, M. Larena, et al. 2023. "Demographic History and Genetic Structure in Pre-Hispanic Central Mexico." *Science* 380, no. 6645: eadd6142. https://doi.org/10.1126/science.add6142.

Vinueza-Espinosa, D. C., C. Santos, C. Martínez-Labarga, and A. Malgosa. 2020. "Human DNA Extraction From Highly Degraded Skeletal Remains: How to Find a Suitable Method?" *Electrophoresis* 41, no. 24: 2149–2158. https://doi.org/10.1002/elps.202000171.

Vives, S., M. T. Gilbert, C. Arenas, E. Gigli, O. Lao, and C. Lalueza-Fox. 2008. "Statistical Analysis of Post Mortem DNA Damage-Derived Miscoding Lesions in Neandertal Mitochondrial DNA." *BMC Research Notes* 1: 1–6. https://doi.org/10.1186/1756-0500-1-40.

Wright, J. L., S. Wasef, T. H. Heupink, et al. 2018. "Ancient Nuclear Genomes Enable Repatriation of Indigenous Human Remains." *Science Advances* 4, no. 12: 1–12. https://doi.org/10.1126/sciadv.aau5064.

Xavier, M. J., B. Nixon, S. D. Roman, and R. J. Aitken. 2018. "Improved Methods of DNA Extraction From Human Spermatozoa That Mitigate Experimentally-Induced Oxidative DNA Damage." *PLoS One* 13, no. 3: 1–23. https://doi.org/10.1371/journal.pone.0195003.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

17550998, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.14061 by Readcube (Labtiva Inc.), Wiley Online Library on [07/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

MOLECULAR ECOLOGY

Supplemental Information for:

A *post-mortem* molecular damage profile in the ancient human mitochondrial DNA

Daniel R. Cuesta Aguirre, Cristina Amor-Jimenez, Assumpció Malgosa, Cristina Santos

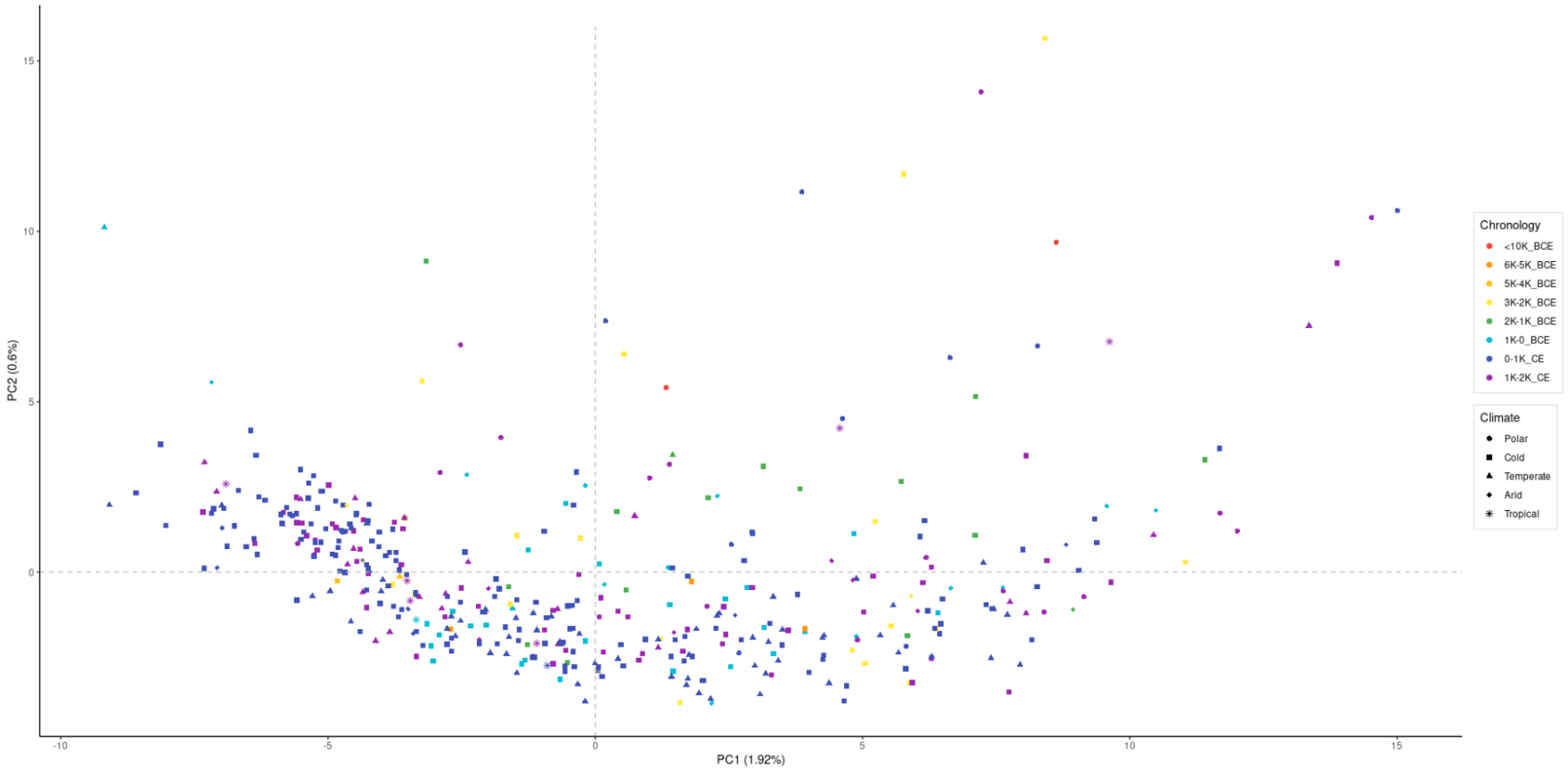


Figure S1. Principal Component Analysis (PCA) of non-UDG-treated ancient samples based on climatic region and chronology. The samples are shaped according to their climatic region (Peel et al., 2007) and colored based on their chronology. Each data point represents a sample, and its position in the PCA plot is determined by the presence (1) or absence (0) of DNA damage hit. In cases where a position was unavailable for an individual, but present in 75% of the individuals, the value was imputed as the mean presence score for that position across all individuals who had it.

4.2. Article 2: An easy-to-use pipeline to analyze amplicon-based Next Generation Sequencing results of human mitochondrial DNA from degraded samples

Authors

Daniel R. Cuesta-Aguirre, Assumpció Malgosa, Cristina Santos

Journal

PLOS One

Impact factor (JCR) and quartile ranking

2.9 and Q1

Date of publication

21 November 2024

DOI

https://doi.org/10.1371/journal.pone.0311115

Addressed objectives

This second publication addresses the previously defined objective 2:

2) To create an easy-to-use pipeline that enables the analysis of mtDNA sequences from degraded samples obtained through amplicon-based NGS methodologies.

Supplementary data is available in "Article 2" at the following link: https://uab-my.sharepoint.com/:f:/g/personal/2103711_uab_cat/EhLQT7GxLs1GoowdsaXF7QYBF7DKCwofUOyqhbjFw4XWRw?e=3ddWqy

PLOS ONE

RESEARCH ARTICLE

An easy-to-use pipeline to analyze amplicon-based Next Generation Sequencing results of human mitochondrial DNA from degraded samples

Daniel R. Cuesta-Aguirre, Assumpció Malgosa, Cristina Santos*

Research Group in Biological Anthropology, Biological Anthropology Unit, Department of Animal Biology, Vegetal Biology and Ecology, Universitat Autònoma de Barcelona, Bellaterra, Spain

* cristina.santos@uab.cat

Check for updates

OPEN ACCESS

Citation: Cuesta-Aguirre DR, Malgosa A, Santos C (2024) An easy-to-use pipeline to analyze amplicon-based Next Generation Sequencing results of human mitochondrial DNA from degraded samples. PLoS ONE 19(11): e0311115. https://doi.org/10.1371/journal.pone.0311115

Editor: Nihad A. M. Al-Rashedi, Al Muthanna University, IRAQ

Received: May 31, 2024

Accepted: September 12, 2024

Published: November 21, 2024

Copyright: © 2024 Cuesta-Aguirre et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All FastQ files are available at the CORA RDR, an internal repository of the Universitat Autònoma de Barcelona (https://doi.org/10.34810/data1541).

Funding: This research was supported by the Ministerio de Ciencia e Innovación (Ref. PGC2018-096666-B-100; PID2022-136748NB-I00), by the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya and the Fons Social Europeu. DRCA is a PhD fellow from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) (2020 FI_B1

Abstract

Genome and transcriptome examinations have become more common due to Next-Generation Sequencing (NGS), which significantly increases throughput and depth coverage while reducing costs and time. Mitochondrial DNA (mtDNA) is often the marker of choice in degraded samples from archaeological and forensic contexts, as its higher number of copies can improve the success of the experiment. Among other sequencing strategies, amplicon-based NGS techniques are currently being used to obtain enough data to be analyzed. There are some pipelines designed for the analysis of ancient mtDNA samples and others for the analysis of amplicon data. However, these pipelines pose a challenge for non-expert users and cannot often address both ancient and forensic DNA particularities and amplicon-based sequencing simultaneously. To overcome these challenges, a user-friendly bioinformatic tool was developed to analyze the non-coding region of human mtDNA from degraded samples recovered in archaeological and forensic contexts. The tool can be easily modified to fit the specifications of other amplicon-based NGS experiments. A comparative analysis between two tools, MarkDuplicates from Picard and dedup parameter from fastp, both designed for duplicate removal was conducted. Additionally, various thresholds of PMDtools, a specialized tool designed for extracting reads affected by post-mortem damage, were used. Finally, the depth coverage of each amplicon was correlated with its level of damage. The results obtained indicated that, for removing duplicates, dedup is a better tool since retains more non-repeated reads, that are removed by MarkDuplicates. On the other hand, a PMDS = 1 in PMDtools was the threshold that allowed better differentiation between present-day and ancient samples, in terms of damage, without losing too many reads in the process. These two bioinformatic tools were added to a pipeline designed to obtain both haplotype and haplogroup of mtDNA. Furthermore, the pipeline presented in the present study generates information about the quality and possible contamination of the sample. This pipeline is designed to automatize mtDNA analysis, however, particularly for ancient samples, some manual analyses may be required to fully validate results since the amplicons that used to be

00641). CS and AM are members of the GREAB supported by the Generalitat de Catalunya (Ref. 2021 SGR 00183). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

more easily recovered were the ones that had fewer reads with damage, indicating that special care must be taken for poor recovered samples.

Introduction

Next-Generation Sequencing (NGS) techniques are commonly used in studies that require the sequencing of the whole genome or transcriptome, or some parts of it, in any organism, whether living or dead. NGS has many advantages over traditional capillary electrophoresis-based Sanger sequencing, such as increased throughput and depth of coverage, while reducing costs and time [1–3]. However, there are situations where the quality or quantity of the template is insufficient to obtain satisfactory results. In such cases, one solution is to use amplification-based enrichment methods. Amplicon-based NGS methods use Polymerase Chain Reaction (PCR) to increase the number of molecules for a specific region before preparing sequencing libraries. This approach can achieve higher depth coverages than common NGS techniques for a region with little off-target sequence, as the start and stop coordinates of each amplicon are predetermined [4].

The use of amplicons with NGS can greatly benefit the analysis of human mitochondrial DNA (mtDNA) non-coding region (NCR) in both forensic and population genetics. The NCR is a region of almost 1120 base pairs (bp) that exhibits high levels of variation in human populations, especially in the two hypervariable regions HVRI (16024–16383 bp) and HVRII (57–372 bp). Traditionally, these two hypervariable regions were amplified by PCR and sequenced using Sanger technology [5]. Nonetheless, amplicons with over 300 bp may not be reliable in samples of poor quality, so shorter amplicons should be used instead. This approach requires many PCR amplification reactions and Sanger sequencing reactions, which can become very time-consuming and demanding [6].

As previously mentioned, NGS technology provides a larger amount of genetic information more efficiently than electrophoresis-based methods [7, 8]. As a result, some commercial kits that allow mtDNA-NCR analysis using NGS have been developed. In the specialized field of forensic genetics, three distinct kits have been developed: The ForenSeq mtDNA Control region kit (Verogen), the Precision ID mtDNA control region Panel Kit (Thermo Fisher Scientific), and the PowerSeq™ CRM Nested System kit (Promega Corporation). Both the ForenSeq mtDNA Control Region kit (Verogen) and the Precision ID mtDNA control region Panel Kit (ThermoFisher Scientific) require two PCR reactions to amplify the entire mtDNA-NCR. Additionally, another PCR step is needed to add adapters and index the samples. In contrast, the PowerSeq™ CRM Nested System kit (Promega Corporation) streamlines the process by using a single-multiplex PCR reaction to both amplify the mtDNA-NCR and add indexed adapters for Illumina MiSeq™ (Illumina, INC., San Diego, CA, USA). This approach significantly reduces overall work time, minimizes amplicon manipulation, and lowers the risk of contamination. Traditionally, the sequencing data obtained from these three kits has been analyzed with the MiSeq FGx and UAS software (Verogen), the Torrent Variant Caller software (ThermoFisher Scientific), and the GeneMarker® HTS software (Promega Corporation), respectively [6, 9–17]. However, these software packages lack specific tools for dealing with particularities of genetic data derived from ancient and forensic samples, that are presented afterward.

Ancient DNA (aDNA) refers to the study of DNA extracted from preserved or fossilized biological material of specimens that perished long ago [18]. It is essential to keep in mind that these samples undergo post-mortem DNA damage due to environmental conditions and

degradation reactions. For instance, they may experience extensive fragmentation, abasic sites, single-strand breaks, interstrand cross-links, and some chemical modifications of nucleotide bases [19–22].

Post-mortem molecular damage (PMD) in DNA can hinder the study of aDNA as undamaged DNA molecules can outcompete ancient templates in enzymatic reactions [23]. To authenticate aDNA, researchers often analyze sequence length and specific nucleotide misincorporation patterns. These patterns include both the short length of the sequences and an excess of cytosine to thymine (C-to-T) misincorporations at 5' ends and its complementary guanine to adenine (G-to-A) misincorporations at 3' ends [24]. Various tools are available for detecting patterns of misincorporation occurring at the ends of molecules. Among these tools are MapDamage2 [25] and Shmutzi [26], which can help to estimate contamination levels and authenticate aDNA by using these misincorporations at the fragment ends. Additionally, PMDtools is a bioinformatic tool that uses the PMD score (PMDS), a log-likelihood ratio, to analyze damage in sequencing data. When the PMDS is negative, PMDtools identifies and selects reads affected by PMD, PCR errors, and sequencing misreads. On the other hand, when the PMDS is positive, it focuses only on changes caused by DNA degradation. The PMDS value indicates the likelihood that a change is due to DNA degradation, with higher values implying fewer available reads. Skoglund et al. demonstrated the effectiveness of PMDS in analyzing ancient samples from different periods. For example, samples over 38,000 years old and around 5,000 years old were well-suited to a PMDS higher than 2. However, for more recent samples, approximately 100 years old, a high PMDS value removed almost all reads and lower values of PMDS must be used [27].

When employing amplicon-based techniques for sequencing analysis of damaged DNA, there is a replication of the original molecules, leading to a surplus of identical sequences that can overshadow some polymorphic variants. Hence, it is essential to remove duplicates in all NGS analyses that involves ancient or degraded samples [28]. For present-day samples, it has been stated that deduplication has a minimal effect on the accuracy of variant calls, so it is not necessary to carry it out [29]. Duplicate reads encompass PCR duplicates (reads derived from the same original fragment in the DNA library) or optical duplicates (reads from clusters read twice in Illumina platforms). There are available numerous tools to eliminate duplicate reads. Samtools' 'markdup' command [30], DeDup software [31], and Picard's 'MarkDuplicates' command [32] operate by eliminating duplicates according to template coordinates. Conversely, the 'dedup' parameter from fastp [33] identifies duplication only when all base pairs match between reads. This functionality was introduced in fastp v0.22.0.

There are pipelines available to analyze aDNA data generated by NGS technologies, such as the nf-core/eager pipeline [34]. However, the nf-core/eager pipeline (v2.5.0) does not yet have a tool to eliminate duplicates based on reads with identical coordinates. Specific pipelines to analyze ancient mtDNA data obtained by NGS are also available, such as the pipeline presented by Diroma et al. [35], but it is not designed to analyze amplicon-based data. As far as we know, no free software is available to analyze data from ancient and forensic mtDNA-NCR specifically generated by amplicon-based NGS methods. Therefore, our main goal is to develop a free and easy-to-use bioinformatic tool to analyze data from ancient or degraded samples obtained by an amplicon-based Next Generation Sequencing method. Also, we will determine the best deduplication method to work with amplicons and establish the best PMDS from PMDtools that can differentiate mtDNA-NCR from present-day and ancient samples.

Material and methods

Samples

Following the processing of numerous samples in our laboratory, 148 high-quality ancient Spanish human samples were carefully selected, encompassing various chronologies. No permits were required for the described study, which complied with all relevant regulations. Moreover, 15 samples from members of the research group (all members provided confirmed consent), which were used as contaminant control, were also selected. All samples exhibited a recovery rate of over 80% for the mtDNA-NCR, and none displayed mixed bases exceeding the designated threshold of 30% of the reads.

Sample preparation and DNA extraction

The external surface of all skeletal remains from ancient samples was polished using a sterile tungsten tip placed into a micro drill to remove the outermost millimeters. Then, a 2–3 cm section was cut using a tungsten sterile disk and pulverized with a hammer. After that, 100–150 mg of sample was used for DNA extraction.

DNA was extracted in the ancient DNA Laboratory of the Biological Anthropology Unit of the Universitat Autònoma de Barcelona (UAB), Spain, using the High Pure Viral Nucleic Acid Large Volume kit (Roche), which is a silica-based method on a HE-membrane. The extraction was carried out following the specifications of Vinueza-Espinosa et al. [36]. Blank controls were processed with every extraction to ensure accuracy.

For present-day individuals, buccal swab samples were taken. DNA was extracted using the QIAamp DNA Investigator Kit (Qiagen) in the pre-PCR area of the modern DNA Laboratory of the Biological Anthropology Unit of the UAB.

Amplification, libraries preparation, and sequencing

All the samples were amplified and indexed using the PowerSeq™ CRM Nested System kit (Promega Corporation). The NCR was amplified from base 16013 to base 592, with 10 overlapped amplicons, as follows: Amplicon 1 (16013–16126 bp), Amplicon 2 (16116–16225 bp), Amplicon 3 (16223–16408 bp), Amplicon 4 (16387–16486 bp), Amplicon 5 (16474–30), Amplicon 6 (16555–152 bp), Amplicon 7 (136–257 bp), Amplicon 8 (246–364 bp), Amplicon 9 (342–436 bp), and Amplicon 10 (429–592 bp).

For the ancient samples, the modified protocol proposed by Vinueza-Espinosa et al. [37] for archaeological samples was used. The PCR conditions were as follows: 96˚C for 15 min; 35 cycles of 96˚C for 15 seconds, 60˚C for 35 seconds and 72˚C for 30 seconds; and 60˚C for 2 min. For present-day samples, the standard conditions proposed by the manufacturer were followed.

After amplification/indexation, the libraries were purified using the GeneRead™ Size Selection kit (Qiagen), following the manufacturer's protocol. Quantifications of the libraries were performed with Qubit™ ds HS Assay kit (ThermoFisher Scientific), and the libraries' concentrations were normalized to 4 nM. Runs were performed on the Illumina MiSeq instrument, using standard flow cells, MiSeq® Reagent kit v2 Nano 2x150bp.

Bioinformatic tool selection for duplicate read identification

Two similar workflows, each with the same filters, were created to compare which deduplication (identification and elimination of duplicates) method was the best to remove duplicates correctly from ancient amplicon-based NGS sequences. Two tools based on different duplicates detection were compared. On the one hand, fastp v0.23.2 [33] was used, which is a tool

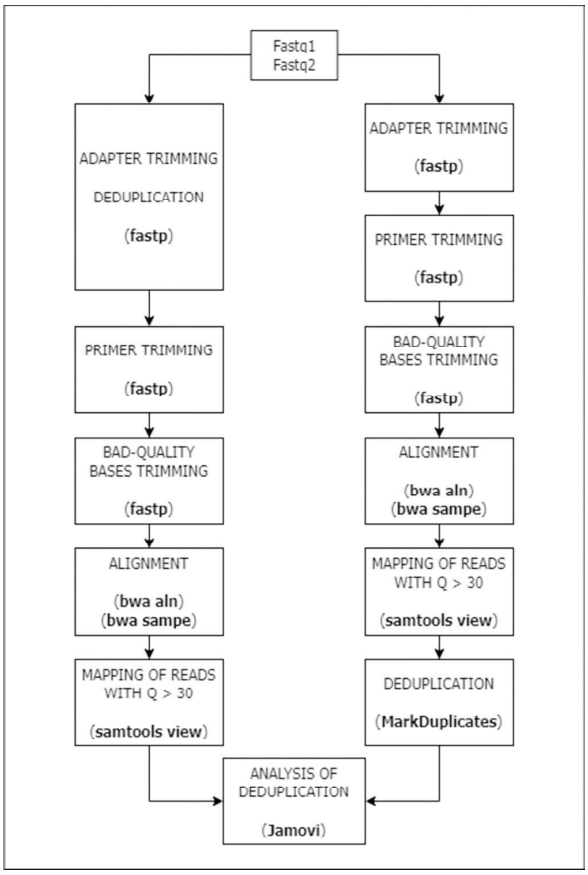


Fig 1. Workflow followed to study what method to remove duplicates brings better results.

https://doi.org/10.1371/journal.pone.0311115.g001

that identifies duplicates when two sequences are identical, meaning that they have the same sequence throughout the whole molecule. On the other hand, 'MadkDuplicates' from Picard v2.26.2 [32], was employed as a proxy of those tools that remove duplicates according to the template's coordinates. The deduplication tool was placed for each workflow as shown in Fig 1. For one workflow, deduplication was performed using fastp v0.23.2 [33] after adapting trimming (the process of adapters' elimination). For the other workflow, deduplication was performed using 'MadkDuplicates' from Picard v2.26.2 [32] after selecting reads mapped with a quality score over 30. The number of useful reads, defined as the reads that passed all filters and allowed for variant calling, and the duplication rate, which was calculated as one minus the number of reads without duplicates divided by the number of reads with duplicates (equivalent to the percentage of duplicates), were both annotated.

The Integrative Genome Viewer (IGV) v2.11.2 [38] was employed to analyze which tool removed more true duplicates and retained more non-duplicated sequences.

To compare the number of useful reads and the duplication rate between different deduplication methods normality was checked using the Shapiro-Wilk test and the differences between the deduplication methods were studied using a Paired Samples T-Test.

PMDS selection for aDNA authentication

To determine if ancient samples can be authenticated—distinguishing them from present-day samples—based on their percentage of reads with PMD, five PMD thresholds (PMDS: -1, 0, 1, 2, and 3) were used in PMDtools 0.23.2 [27], and the percentage of reads with PMD identified for each threshold. An initial exploratory analysis was conducted, and the 95% Confidence Intervals for each percentage were estimated. To assess the variations in the percentage of reads with PMD between chronological age and PMDS thresholds, a Repeated Measures ANOVA test was performed, along with their respective Post Hoc Comparisons, employing Bonferroni's correction. Ancient samples were deduplicated prior PMDS selection analysis, while present-day samples were not.

Study of differences by amplicon

The depth coverage and percentage of damage for each amplicon were estimated, including Amplicon 1 (16013–16126 bp), Amplicon 2 (16116–16225 bp), Amplicon 3 (16223–16408 bp), Amplicon 4 (16387–16486 bp), Amplicon 5 (16474–30), Amplicon 6 (16555–152 bp), Amplicon 7 (136–257 bp), Amplicon 8 (246–364 bp), Amplicon 9 (342–436 bp), and Amplicon 10 (429–592 bp). A PMDS of 1 was used for both ancient and present-day samples to determine the level of damage in each amplicon. The mean and median of depth coverage and percentage of damage were computed. Furthermore, for both ancient and present-day samples, a correlation between the depth coverage and the percentage of damage in amplicons was performed.

All the statistical analyses were carried out using Jamovi 2.3 [39, 40] and considered a 5% significance.

Results

Selection of bioinformatic tools to remove duplicated reads and thresholds of PMDtools to authenticate aDNA

After applying the established filters, the analysis of deduplication and PMD was performed including 148 ancient samples and 15 present-day samples (S1 File).

Upon conducting an exploratory analysis of both the number of useful reads and the duplication rate of the data a few values were identified as outliers. However, they were considered acceptable after thorough analysis. Consequently, all the data was used in subsequent analyses.

When the samples were analyzed with fastp v0.23.2, there were a higher number of useful reads and a lower duplication rate compared to the results obtained with MarkDuplicates (Table 1), being the differences between the two methods significant (Wilcoxon: $p < 0.001$).

By visualization of deduplicated samples with IGV v2.11.2, one the one hand, it was seen that MarkDuplicates removed sequences due to the identical coordinates, despite the differences at the sequence of the different reads (S1 Fig). Meaning that MarkDuplicates remove

Table 1. Comparative descriptive statistics of useful reads count and duplication rates across the two deduplication strategies.

	Deduplication method	Mean	Median	IQR	Percentile 2.5th	Percentile 97.5th	Minimum	Maximum	Sapiro-Wilk	
									W	p
Number of useful reads	**fastp**	18,932.5	17,108	15,696.5	3,626.5	42,425.5	1950	47226	0.96	< 0.001
	MarkDuplicates	107.6	102	50	45.75	207.0	28	225	0.96	< 0.001
Duplication rate	**fastp**	28.5	29.8	20.63	2.75	60.3	0.48	67.7	0.98	0.010
	MarkDuplicates	99.6	99.7	0.15	99.1	99.8	98.63	99.9	0.76	< 0.001

https://doi.org/10.1371/journal.pone.0311115.t001

Table 2. Descriptive statistics of the percentage of recovered reads with damage when different post-mortem molecular damage scores (PMDS) were used in present-day and ancient samples.

	PMDS in PMDtools	Chronology	Mean	Median	IQR	Minimum	Maximum	Percentile 2.5th	Percentile 97.5th
Percentage of recovered reads with damage	**-1**	**Present-day**	74.91	75.76	8.15	67.30	83.18	75.33	85.08
		Ancient	94.92	95.23	5.67	86.62	99.87	87.40	99.59
	0	**Present-day**	8.74	10.51	9.73	2.18	15.76	2.28	15.12
		Ancient	32.41	31.79	15.15	8.76	67.45	13.82	56.35
	1	**Present-day**	1.65	1.35	1.31	0.78	2.89	0.82	2.84
		Ancient	26.57	25.86	12.59	7.49	62.92	9.17	47.74
	2	**Present-day**	0.56	0.31	0.57	0.18	1.35	0.19	1.26
		Ancient	12.28	11.53	8.06	2.01	31.53	3.33	23.32
	3	**Present-day**	0.33	0.11	0.38	0.08	1.14	0.08	1.06
		Ancient	7.44	7.01	5.36	1.06	19.02	1.69	15.74

https://doi.org/10.1371/journal.pone.0311115.t002

non-duplicate sequences. On the other hand, fastp toke sequences into account and allowed to retain more original molecules.

Table 2 presents descriptive statistics for the percentage of recovered reads with damage in ancient and present-day samples when different PMDS values were considered. It is evidenced that the higher the PMDS, the lower the percentages of recovered reads. In accordance with what is expected for the relatively recent chronology of the samples (some samples classified as ancient belong to the 20th century), values of PMDS of 2 or 3 identify a very low percentage of reads with damage, limiting subsequent analysis based only on damaged reads. However, for PMDS 1, the percentage is higher, with a median value reaching 25% for ancient samples, and of important relevance, clearly different between present-day and ancient samples.

When the chronology was considered, in each PMDS analyzed, ancient samples had a higher percentage of reads with damage than present-day samples. The 2.5th and 97.5th percentiles between present-day and ancient samples overlapped only for PMDS = 0 (Table 2 and Fig 2).

The significance of differences in the percentage of recovered reads with PMD was confirmed for both chronology (Repeated Measures ANOVA: $F = 127$; p-value < 0.001) and PMDS (Repeated Measures ANOVA: $F = 2291.4$; p-value < 0.001). Moreover, pairwise comparisons for the chronology in each PMDS also showed significant differences (p-value < 0.001).

Differences by amplicon

Considering the parameters previously validated (fastp to eliminate duplicated reads and PMDS = 1), the results for depth coverage and the percentage of damaged reads for each amplicon were calculated and are available in S1 File for both present-day and ancient samples. The median and mean values, for ancient samples, are shown in Table 3 and Fig 3. In ancient samples, amplicons with the lowest depth coverage have the highest damage values (Amplicon 1: median depth coverage = 235.75, median percentage of damage = 44.41; Amplicon 2: median depth coverage = 301.60, median percentage of damage = 46.10). In contrast, amplicons with the highest amount of reads have the lowest levels of damage (Amplicon 4: median depth coverage = 911.55, median percentage of damage = 20.29; Amplicon 5: median depth

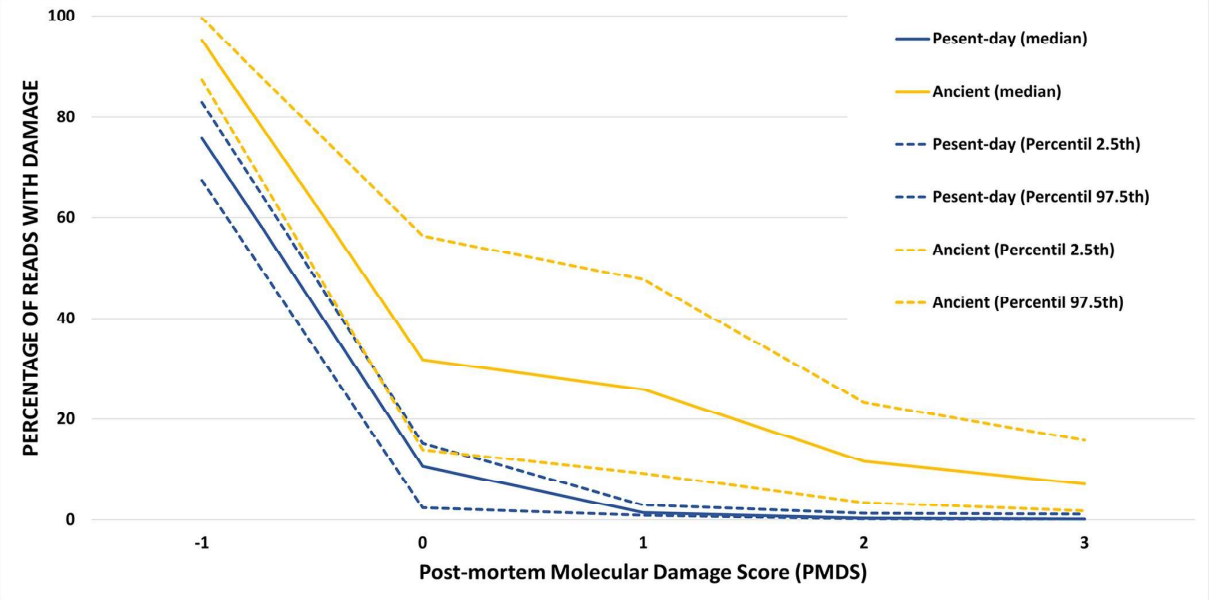


Fig 2. Percentage of damage with different thresholds from PMDtools. The percentage of reads with post-mortem molecular damage (PMD) extracted by PMDtools (Y axis) with five different PMD scores (PMDS) (X axis). These percentages have been estimated by dividing the number of reads with damage given by PMD (reads with post-mortem molecular damage) by the number of reads aligned with the reference sequence. Medians, 2.5th, and 97.5th percentiles are indicated for ancient and present-day samples. For both ancient and Present-day samples, the higher the PMDS the lower the percentage of reads recovered with damage. Ancient reads presented higher percentages of damage than Present-day samples.

https://doi.org/10.1371/journal.pone.0311115.g002

coverage = 1074.21, median percentage of damage = 21.17). For present-day samples, the relationship between depth coverage and the percentage of damaged reads was less clear when analyzing the descriptives. However, for both ancient and present-day samples together there is a significant negative correlation between the percentage of damaged reads and the median depth coverage (Sperman's rho = -0.187, p-value < 0.001), as it can be seen in Fig 4; as well as when ancient and present-day samples were analyzed separately (ancient Spearman's rho = -0.174, p-value < 0.001; present-day Spearman's rho = -0.224, p-value = 0.006).

Table 3. Mean and median depth coverage and damage percentage (PMDS = 1) per amplicon for ancient samples.

Amplicon	Depth coverage: mean (SD)	Percentage of damage: mean (SD)	Depth coverage: median (IQR)	Percentage of damage: median (IQR)
1	257.86 (168.15)	43.89 (16.67)	235.75 (212.60)	44.41 (20.29)
2	332.24 (228.92)	45.78 (18.41)	301.60 (265.84)	46.10 (21.75)
3	443.70 (254.27)	31.54 (17.34)	413.12 (318.19)	28.90 (13.03)
4	985.39 (532.46)	22.18 (11.51)	911.55 (736.21)	20.29 (8.55)
5	1225.41 (651.26)	20.97 (7.33)	1074.21 (941.31)	21.17 (9.77)
6	931.64 (501.42)	30.23 (14.05)	853.22 (601.34)	28.15 (15.75)
7	725.81 (454.60)	30.96 (23.34)	671.35 (554.41)	23.42 (17.96)
8	531.57 (368.32)	40.11 (16.61)	479.47 (467.24)	39.03 (22.28)
9	948.08 (759.52)	33.19 (13.44)	821.32 (832.29)	34.96 (17.20)
10	984.11 (810.61)	32.48 (13.39)	817.35 (809.48)	33.87 (15.85)

https://doi.org/10.1371/journal.pone.0311115.t003

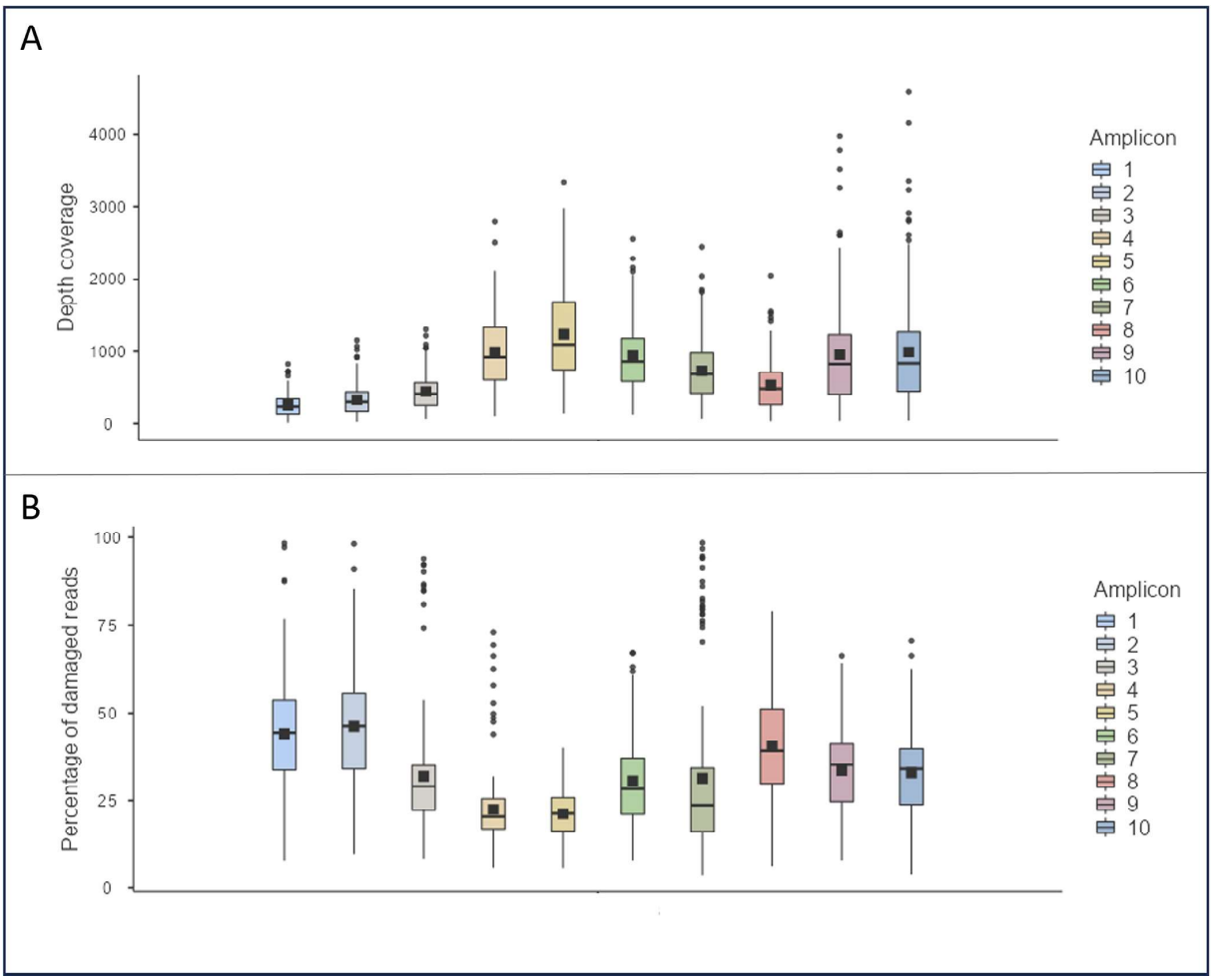


Fig 3. Comparison between the 10 amplicons amplified with the PowerSeq™ CRM Nested System kit (Promega Corporation) in ancient samples. Boxplot of (A) the depth coverage (number of unique reads aligned with that amplicon) and (B) the percentage of damaged reads (percentage of reads that presented post-mortem molecular damage, according to PMDtools). They are estimated from the 148 ancient samples in each one of the ten amplicons amplified with the PowerSeq™ CRM Nested System kit (Promega Corporation) (marked with different colors and sorted from 1 to 10). Points represent outliers and square marks mean. The higher the depth coverage, the lower the percentage of damaged reads and vice-versa.

https://doi.org/10.1371/journal.pone.0311115.g003

Pipeline development

Deduplication using fastp and a threshold of 1 in PMDtools were included in an aDNA pipeline to analyze the NCR of mtDNA reads generated from PCR-based libraries (Fig 5).

The preprocessing of the reads involved quality-check analysis and processing of both 5' and 3' ends of each read to obtain better quality mapped reads. Raw paired-end FastQ files were first subjected to quality check using FastQC v0.11.9 [41]. Duplicates, adapters, primers, and bad quality bases in 3' ends were eliminated by three consecutively fastp v0.23.2 [33] analyses. Initially, a fastp v0.23.2 [33] analysis was performed to trim the adapters and remove the duplicates, restricting the minimum length of the reads to 30 bp and trimming the poliX of the tails. Another fastp v0.23.2 [33] analysis was used to eliminate primers. Since these sequences are unknown and unidentified by fastp v0.23.2, 25 bp from the 5' end and 5 bp from the 3' end

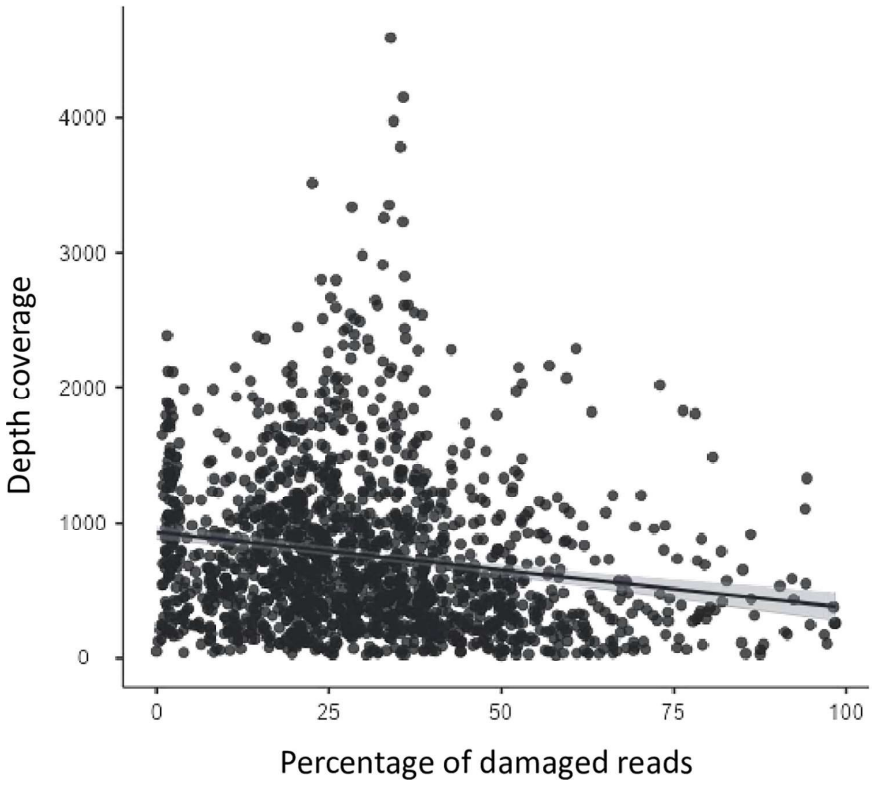


Fig 4. Relation between depth coverage and percentage of damage. Scatterplot that relates the depth coverage (Y axis) against the percentage of reads that presented post-mortem molecular damage when a PMDS of 1 was used in PMDtools (Y axis) for all samples and all the regions of the non-coding region of the mtDNA.

https://doi.org/10.1371/journal.pone.0311115.g004

were trimmed for both paired-end reads. Finally, the last fastp v0.23.2 [33] was used to eliminate bases with a quality lower than 20 from the 3' end, if any. Additionally, in this last analysis, overlapped paired-end reads were merged if the overlapping region was higher than 10 bp. Another FastQC v0.11.9 [41] was used to check the quality of the trimmed FastQ files.

The reads were aligned to the non-coding region of the revised Cambridge Reference Sequence (rCRS, GenBank accession NC_012920.1) [42]. This alignment was done using BWA v0.7.17 [43]. To avoid issues caused by the circularity of the mtDNA and the reads aligned in the replication origin, the sequence was linearized from base 15901 to base 700. Later, position renumbering was made as position 15901 from the mtDNA was deemed position 1 in the new reference sequence.

To generate a sorted BAM file that only contained useful reads, SAMtools v1.16.1 [30] was used, and reads with a mapping quality of over 30 were included. Then, the samtools view -c command was applied to estimate the number of useful reads, those that would be used to determine the haplotype and haplogroup. Next, reads that presented PMD were extracted using PMDtools v0.60 [27] and saved in a separate BAM file. The number of reads with PMD was then estimated using the samtools view -c command.

QualiMap v2.2.2a [44] was used to quality-checking of BAM files, allowing both mean depth coverage (*mean coverageData* in the output of QualiMap) and mean mapping quality (*mean mapping quality* in the output of QualiMap) to be extracted by command *grep*.

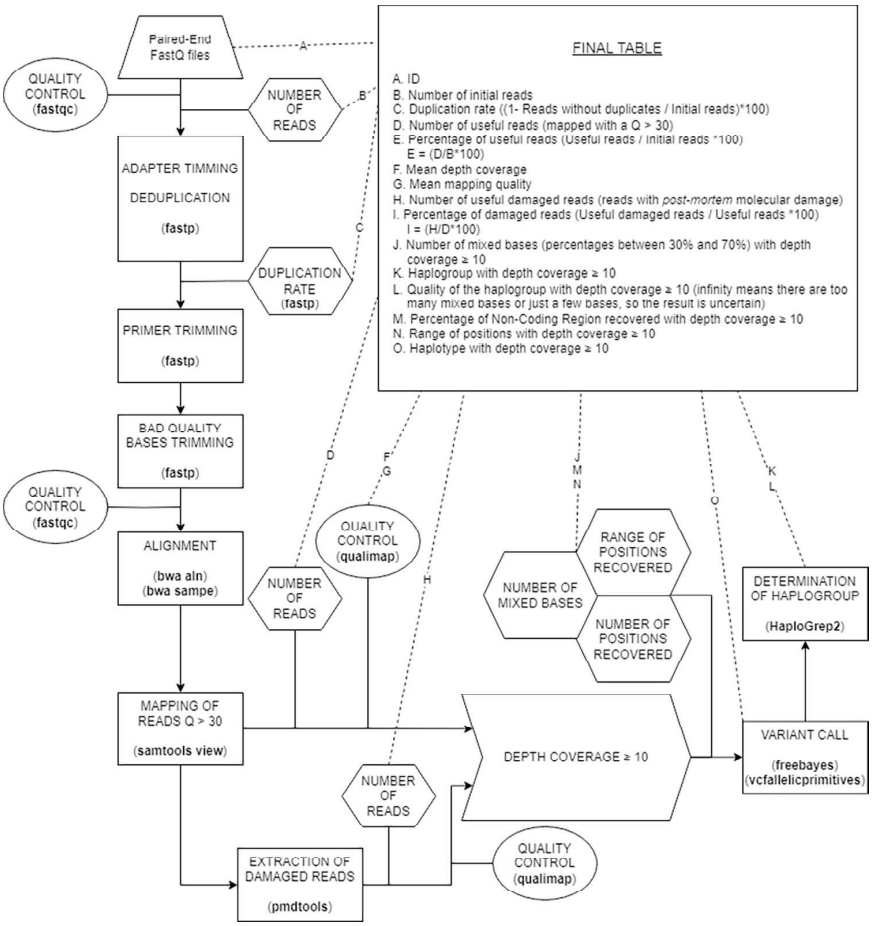


Fig 5. Pipeline followed to generate results from paired-end reads obtained by amplicon-based NGS methods. Tools are marked in bold while the processes are in capital letters. The minimum depth coverage can be easily changed at the top of the script.

https://doi.org/10.1371/journal.pone.0311115.g005

From this point forward, the same procedures were applied to both BAM files containing all reads and BAM files containing only those reads exhibiting *post-mortem* molecular damage that were extracted by PMDtools. The range of positions with more than 10 reads was determined using the 'samtools depth' command and a Python script. Variant calling was performed with FreeBayes v1.3.6 [45] and vcfallelicprimitives from VCFlib v1.0.3 [46]. In this analysis, the minimum allele depth was set at 10 reads (parameter: min-coverage = 10) and the threshold for minor variant frequency call was established at 30% (parameter: F = 0.30). Indels, multi-nucleotide polymorphisms, and complex events were excluded from the analysis, as well as the region between bases 303 and 315 due to their alignment complexity. Positions with more than one nucleotide variant with a frequency ranging between 30% and 70% were classified as mixed bases, and each instance was recorded for further examination. In such cases, the International Union of Pure and Applied Chemistry (IUPAC) nomenclature was employed in variant call.

An HSD file was generated, containing the sample name, the range of the mitogenome covered by more than 10 reads, an interrogation mark (representing the unknown haplogroup),

and the haplotype; all these were separated by tabs. Finally, Haplogrep2 v2.4.0 [47] was employed locally to predict mtDNA haplogroups.

Several outputs are provided (summarized in S1 File). The first output is a table that includes the ID, number of initial reads, duplication rate, number of useful reads, percentage of useful reads, mean depth coverage, mean mapping quality, number of useful damaged reads, percentage of damaged reads, and number of mixed bases. It also includes the most probable haplogroup, quality of that haplogroup, range of positions, and haplotype for a depth coverage $\geq$ 10. Secondly, an HSD file is generated for uploading to Haplogrep3, offering more specific information not available from the local analysis. This file contains all pertinent details for the two types of BAM files, including those with all reads and those composed only by the reads with post-mortem molecular damage extracted by PMDtools for depth coverage $\geq$ 10. Finally, a TXT file containing information on the mixed bases is provided. This information includes the sample ID, position, alternative base, depth coverage of the position, number of appearances of the alternative base, and percentage of the alternative base. A mixed base will only be reported if the percentage of the mixture falls between 30% and 70%.

Some parameters, such as the reference sequence, the minimum allele depth, and the minor variant frequency can be changed at the top of the script.

The pipeline and its usage description are available at https://github.com/DanielRCA/NCR-mtDNA_ampliconbasedngs.

Discussion

Amplicon-based NGS methods offer the advantage of yielding a high number of reads from regions with few template sequences. However, this process generates a significant number of duplicated reads, which must be eliminated to analyze just unique reads. There are numerous tools for removing duplicates. Some are based on query or template coordinates (the positions where reads start and/or finish), such as the 'markdup' command from SAMtools [30], DeDup software [31], and 'MarkDuplicates' from Picard [32]. Nonetheless, since amplicons are reads with identical coordinates, these tools may not be suitable. For this purpose, the 'dedup' parameter in fastp, implemented in v0.22.0, should be the best option because it identifies duplicates only when all base pairs between reads are identical [33]. The number and percentage of reads remaining after duplicate elimination were analyzed using two different tools: MarkDuplicates and fastp. Using MarkDuplicates, over 99% of reads were removed. In contrast, fastp removed an average of 30% of the reads as duplicates. These differences were significant, indicating that fastp removes fewer duplicates in amplicon-based studies, as we expected for the characteristics of each tool. Although both tools retained some true duplicates, MarkDuplicates also removed some molecules with the same coordinates and differences in sequence, as showed with the manual analysis with Integrative Genome Viewer (IGV) v2.11.2 [38]. Overall, it is evidenced that making the deduplication with fastp is a better approach when amplicons derived from degraded or damaged samples are involved.

The analysis of ancient or degraded samples presents the issue of *post-mortem* molecular damage (PMD) misincorporations, specifically C-to-T at 5' ends and their complementary G-to-A misincorporations at 3' ends [24]. However, when using amplicons, misincorporations can occur throughout the entire sequence. This is because amplicons may not extend to the molecule ends due to their predetermined coordinates. Consequently, tools like MapDamage2 [25] or Schmutzi [26] become less effective for amplicon analysis, as they primarily detect misincorporation patterns at the sequence ends, making this kind of sample appear as if they were not ancient samples. Unlike these tools, PMDtools does not limit DNA damage patterns to the

sequence ends [27], calculating the ratio of reads with damage, a viable method for authenticating ancient DNA (aDNA) sequenced by NGS from amplicons.

The variation of the percentage of PMD in the NCR of mtDNA between ancient and present-day samples was investigated by using different Post-mortem Molecular Damage Scores (PMDS) to establish a threshold for distinguishing ancient from present-day samples. This approach could also aid in identifying contaminations. A PMDS of -1, which ignores any damage, was used as control. Skoglund et al. [27] already stated that high thresholds of PMDS would work great for really ancient samples (> 5,000 years old), but it would remove almost all sequences for more recent ancient samples (100 years old). As both recent ancient samples and more ancient samples were analyzed, only a PMDS of 1 or 2 should be used. Considering that higher PMDS results in fewer reads recovered [27], we opted for a PMDS of 1 to ensure enough reads for downstream analysis. With this PMDS, we determined that an ancient sample with 80% of the NCR recovered could be considered ancient with a 95% confidence if the percentage of damage is higher than 9.17%. For lower values, caution is advised, due to possible present-day contamination.

Some authors have described that PMD is not randomly distributed [48–51], so the frequency of damage could vary depending on the region being analyzed. Therefore, differences between amplicons must be considered. This is particularly important for samples for which not all amplicons are recovered, as the value obtained with PMDtools could be biased. It can be observed that amplicons with the highest depth coverage (amplicons 4 and 5) tend to accumulate shorter reads (20–40 bp) with less damage. Additionally, a negative correlation between depth coverage and the percentage of damage in both ancient samples and present-day samples was found. These findings could be explained by the possibility that damage accumulates at the binding sites of some primers, potentially acting as a hotspot of damage and hindering primer binding [37, 48–51]. The results in present-day samples are nearly non-significant, which could be due to the low amount of damage typically found in these samples [49, 51–54].

It is important to consider that for samples with poor quality–for example, those that only amplify two amplicons (4 and 5)–the percentage of damage is likely to be underestimated and may fall below the threshold we have presented in this study for ancient samples (9.17%). Nonetheless, we offer another method to detect potential contamination, which is also useful in cross-contamination: the examination of mixed bases. When mixed bases appear in a low-coverage mitochondrial genome, several factors could be the cause: *post-mortem* molecular damage, heteroplasmy, or contamination. Working with amplicons, depth coverages tend to be high, so it would be difficult for PMD to be a source of mixed bases [37]. Heteroplasmy–a condition in which an individual has different types of mtDNA–is a natural phenomenon that varies among individuals from the same and different populations and, even within tissues from the same individual [55–57]. The only study that used bones for the analysis of the NCR identified that 19% of their sample (17 out of 87 individuals) exhibited at least one heteroplasmy: fourteen individuals had one heteroplasmy, two individuals had 2 heteroplasmies, and one individual had three heteroplasmies [55]. In samples where more than one mixed base is observed, a more detailed study should be conducted, as it is quite likely to indicate contamination. Haplotypes that are generated exclusively from BAM files containing only the damaged reads can help to identify the presence of contamination since these BAM files include only ancient reads. By comparing both the haplogroup and haplotype of BAM files that include only damaged reads with the results obtained for the BAM file with all the reads, it is possible to determine the presence of contamination. Moreover, using the IGV program [38] to visualize the alignments can improve the analysis of regions that contain multiple mixed bases, as particular patterns can be observed, and multiple mitochondrial profiles may be detected.

Using all the previously mentioned considerations we propose the implementation of a new pipeline that consists of a group of tools designed to extract information about the mtDNA-NCR for ancient or degraded samples that have been sequenced using an amplicon-based NGS method. Virtually the tool can be applied to mtDNA-NCR data generated with different commercial kits or other custom amplicon panels. The pipeline produces a final table that includes the sample identifier, total number of reads obtained from sequencing, number of deduplicated reads, duplication rate, number of useful reads, percentage of useful reads, mean depth coverage, number of reads with PMD, percentage of reads with PMD, number of mixed bases, haplogroup, quality of the haplogroup according to Haplogrep2 [47], percentage of NCR recovered, range of positions with coverage over 10 reads, and haplotype. Additionally, an HSD file with the haplotypes for both BAM files with useful reads and BAM files with only reads with PMD is also generated. Finally, a description of the mixed bases is also provided in a different TXT file.

The circular nature of the mtDNA molecule generates an alignment problem in the region where the start (base pair 1) and the end (base pair 16569) merge [58]. The alignment in this region has less quality and some reads are depreciated due to alignment being made through a lineal molecule. To solve these problems, other authors used tools such as CircularMapper [59] or had the first 500 bp at the end of the reference mtDNA molecule [35]. In the present work, to avoid the circularity of the mtDNA molecule the NCR sequence was linearized, allowing all reads to align over it without the circularity problem. Alignment was performed with bwa aln and samse/sampe [43], as it works better for aDNA than other aligners [60].

For the variant calling, indels, multi-nucleotide polymorphisms (MNPs), and complex events were ignored. When *vcfallelicprimitives* from vcflib [46] is set, MNPs are transformed into single-nucleotide polymorphisms, allowing the pipeline and the rest of the programs as HaploGrep2, to work correctly. Moreover, the region between bases 303 and 315 was excluded due to its alignment complexity. This region involves a PolyC tract that can be interrupted by one T [61, 62]. Additionally, the possibility of both length and point heteroplasmy, makes the automatic analysis more difficult and the presence of damage increases the problems that this region represents. The combination of all these facts generates a challenge to the automatic interpretation of this complex region, so the pipeline avoids this region. However, the analysis of this region can be done manually, by visualizing the alignment with the IGV program [38].

This pipeline has been proven to work for data obtained by PowerSeq^{TM} CRM Nested System kit (Promega Corporation) (coverage: 16013–16569; 1–592). It should work without any modification with the ForenSeq mtDNA Control region kit (Verogen, San Diego, USA) (coverage: 16008–16569; 1–595) and the precision ID mtDNA Control Region Panel (Thernmo-Fisher Scientific, Waltham, USA) (coverage: 15954–16569;1–610) since the reference sequence employed covered from position 15901 to position 700. The pipeline should be modified for different amplicons from different regions: new reference sequence, new starting coordinates, and removing the coordinates restructuration based on the origin of replication. For non-overlapped amplicons, a major adaptation of the pipeline should be done, and future versions of the pipeline may explore these aspects.

In summary, we confirmed that the better tool to remove duplicates from NGS mtDNA data derived from amplicon-based methods would be fastp v0.23.2 and we established that PMDS = 1 from PMDtools v0.23.2 is the best alternative to analyze ancient samples. To be considered authentic, the sequencing of mtDNA-CR from ancient samples must have a minimum of 9.17% of damaged reads. Furthermore, we developed an easy-to-use pipeline to analyze ancient mtDNA data generated with NGS technology, providing information about possible contamination of the sample and the mtDNA haplogroup and haplotype. However, particularly for ancient samples, some manual analyses may be required to fully validate results

since the amplicons that used to be more easily recovered were the ones that had fewer reads with damage, indicating that special care must be taken for poor recovered samples.

Supporting information

S1 File. Supporting information. Summary of deduplication analysis and Post-mortem Molecular Damage (PMD) analysis in genomic samples across different centuries (S1.1 File), mean depth coverage and percentage of damage for the ten analyzed amplicons (S1.2 File), and information about the outputs provided by the script (S1.3 File).
(XLSX)

S1 Fig. Screenshots from the IGV program for sample 29. Black frames delimit each stage of the sample: A) with duplicated reads; B) without duplicated reads, removed by fastp software; C) without duplicated reads, removed by MarkDuplicates software. Grey arrows symbolize the aligned reads and coloured rectangles indicate a change of base from the reference sequence (red for Thymine, green for Adenine, blue for Cytosine, and brown for Guanine).
(TIF)

Author Contributions

Conceptualization: Daniel R. Cuesta-Aguirre, Assumpció Malgosa, Cristina Santos.

Data curation: Daniel R. Cuesta-Aguirre, Cristina Santos.

Formal analysis: Daniel R. Cuesta-Aguirre.

Funding acquisition: Assumpció Malgosa, Cristina Santos.

Investigation: Daniel R. Cuesta-Aguirre.

Methodology: Daniel R. Cuesta-Aguirre, Cristina Santos.

Project administration: Cristina Santos.

Resources: Assumpció Malgosa, Cristina Santos.

Supervision: Assumpció Malgosa, Cristina Santos.

Validation: Daniel R. Cuesta-Aguirre, Assumpció Malgosa, Cristina Santos.

Visualization: Daniel R. Cuesta-Aguirre, Cristina Santos.

Writing – original draft: Daniel R. Cuesta-Aguirre.

Writing – review & editing: Assumpció Malgosa, Cristina Santos.

References

1. King JL, Larue BL, Novroski NM, Stoljarova M, Seo SB, Zeng X, et al. High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. Forensic Science International: Genetics. 2014; 12:128–35. https://doi.org/10.1016/j.fsigen.2014.06.001 PMID: 24973578
2. Prieto L, Alonso A, Alves C, Crespillo M, Montesino M, Picornell A, et al. 2006 GEP-ISFG collaborative exercise on mtDNA: reflections about interpretation, artefacts, and DNA mixtures. Forensic Science International: Genetics. 2008; 2(2):126–33. https://doi.org/10.1016/j.fsigen.2007.10.010 PMID: 19083807
3. Peck MA, Sturk-Andreaggi K, Thomas JT, Oliver RS, Barritt-Ross S, Marshall C. Developmental validation of a Nextera XT mitogenome Illumina MiSeq sequencing method for high-quality samples. Forensic Science International: Genetics. 2018; 34(July 2017):25–36. https://doi.org/10.1016/j.fsigen.2018.01.004 PMID: 29413633

4. Deharvengt SJ, Petersen LM, Jung HS, Tsongalis GJ. Nucleic acid analysis in the clinical laboratory. In: Clarke W, Marzinke MA, editors. Contemporary Practice in Clinical Chemistry. Elsevier; 2020. p. 215–34. https://doi.org/10.1016/B978-0-12-815499-1.00013–2
5. Wilson MR, DiZinno JA, Polanskey D, Replogle J, Budowle B. Validation of mitochondrial DNA sequencing for forensic casework analysis. International Journal of Legal Medicine. 1995; 108(2):68–74. https://doi.org/10.1007/BF01369907 PMID: 8547161
6. Gabriel MN, Huffine EF, Ryan JH, Holland MM, Parsons TJ. Improved MtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy. Journal of forensic sciences. 2001; 46 (2):247–53. PMID: 11305426
7. Mardis ER. Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics. 2008; 9:387–402. https://doi.org/10.1146/annurev.genom.9.081307.164359 PMID: 18576944
8. Yang Y, Xie B, Yan J. Application of next-generation sequencing technology in forensic science. Genomics, Proteomics and Bioinformatics. 2014; 12(5):190–7. https://doi.org/10.1016/j.gpb.2014.09.001 PMID: 25462152
9. Ambers A, Elwick K, Cropper ER, Brandhagen MD, Jones B, Durst J, et al. Mitochondrial DNA analysis of the putative skeletal remains of Sieur de Marle: Genetic support for anthropological assessment of biogeographic ancestry. Forensic Science International. 2021; 320:110682. https://doi.org/10.1016/j.forsciint.2021.110682 PMID: 33461005
10. Brandhagen MD, Just RS, Irwin JA. Validation of NGS for mitochondrial DNA casework at the FBI Laboratory. Forensic Science International: Genetics. 2020; 44(October 2019):102151. https://doi.org/10.1016/j.fsigen.2019.102151 PMID: 31629185
11. Buś MM, Lembring M, Kjellström A, Strobl C, Zimmermann B, Parson W, et al. Mitochondrial DNA analysis of a Viking age mass grave in Sweden. Forensic Science International: Genetics. 2019; 42 (March):268–74. https://doi.org/10.1016/j.fsigen.2019.06.002 PMID: 31442669
12. Cuenca D, Battaglia J, Halsing M, Sheehan S. Mitochondrial sequencing of missing persons dna casework by implementing thermo fisher's precision id mtdna whole genome assay. Genes. 2020; 11(11):1–17. https://doi.org/10.3390/genes11111303 PMID: 33158032
13. Gutierrez R, Roman MG, Harrel M, Hughes S, LaRue B, Houston R. Assessment of the ForenSeq mtDNA control region kit and comparison of orthogonal technologies. Forensic Science International: Genetics. 2022; 59(May). https://doi.org/10.1016/j.fsigen.2022.102721 PMID: 35576791
14. Holland MM, Pack ED, McElhoe JA. Evaluation of GeneMarker® HTS for improved alignment of mtDNA MPS data, haplotype determination, and heteroplasmy assessment. Forensic Science International: Genetics. 2017; 28:90–8. https://doi.org/10.1016/j.fsigen.2017.01.016 PMID: 28193506
15. Holt CL, Stephens KM, Walichiewicz P, Fleming KD, Forouzmand E, Wu SF. Human mitochondrial control region and mtgenome: Design and forensic validation of ngs multiplexes, sequencing and analytical software. Genes. 2021; 12(4). https://doi.org/10.3390/genes12040599 PMID: 33921728
16. Strobl C, Eduardoff M, Bus MM, Allen M, Parson W. Evaluation of the precision ID whole MtDNA genome panel for forensic analyses. Forensic Science International: Genetics. 2018; 35(March):21–5. https://doi.org/10.1016/j.fsigen.2018.03.013 PMID: 29626805
17. Pereira V, Longobardi A, Børsting C. Sequencing of mitochondrial genomes using the Precision ID mtDNA Whole Genome Panel. ELECTROPHORESIS. 2018; 39(21):2766–75. https://doi.org/10.1002/elps.201800088 PMID: 30058717
18. Brown TA, Brown KA. Ancient DNA: Using molecular biology to explore the past. BioEssays. 1994; 16 (10):719–26. https://doi.org/10.1002/bies.950161006 PMID: 7980476
19. Pääbo S, Poinar H, Serre D, Jaenicke-Després V, Hebler J, Rohland N, et al. Genetic analyses from ancient DNA. Annual Review of Genetics. 2004; 38:645–79. https://doi.org/10.1146/annurev.genet.37.110801.143214 PMID: 15568989
20. Pääbo S. Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci USA. 1989; 86:1939–43. https://doi.org/10.1073/pnas.86.6.1939 PMID: 2928314
21. Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993; 362(6422):709–15. https://doi.org/10.1038/362709a0 PMID: 8469282
22. Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS ONE. 2012; 7(3). https://doi.org/10.1371/journal.pone.0034131 PMID: 22479540
23. Hansen AJ, Mitchell DL, Wiuf C, Paniker L, Brand TB, Binladen J, et al. Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. Genetics. 2006; 173 (2):1175–9. https://doi.org/10.1534/genetics.106.057349 PMID: 16582426

24. Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, et al. Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences. 2007; 104 (37):14616–21. https://doi.org/10.1073/pnas.0704665104 PMID: 17715061

25. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 2013; 29(13):1682–4. https://doi.org/10.1093/bioinformatics/btt193 PMID: 23613487

26. Renaud G, Slon V, Duggan AT, Kelso J. Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biology. 2015; 16(1):1–19. https://doi.org/10.1186/s13059-015-0776-0 PMID: 26458810

27. Skoglund P, Northoff BH, Shunkov M V., Derevianko AP, Pääbo S, Krause J, et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proceedings of the National Academy of Sciences. 2014; 111(6):2229–34. https://doi.org/10.1073/pnas.1318934111 PMID: 24469802

28. Hofreiter M, Sneberger J, Pospisek M, Vanek D. Progress in forensic bone DNA analysis: Lessons learned from ancient DNA. Forensic Science International: Genetics. 2021; 54(March):102538. https://doi.org/10.1016/j.fsigen.2021.102538 PMID: 34265517

29. Ebbert MTW, Wadsworth ME, Staley LA, Hoyt KL, Pickett B, Miller J, et al. Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. BMC Bioinformatics. 2016; 17(S7):239. https://doi.org/10.1186/s12859-016-1097-3 PMID: 27454357

30. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021; 10(2):1–4. https://doi.org/10.1093/gigascience/giab008 PMID: 33590861

31. Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, et al. EAGER: efficient ancient genome reconstruction. Genome Biology. 2016; 17(1):60. https://doi.org/10.1186/s13059-016-0918-z PMID: 27036623

32. Picard Toolkit. Broad Institute. 2019. Retrieved from https://broadinstitute.github.io/picard/.

33. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018; 34(17):i884–90. https://doi.org/10.1093/bioinformatics/bty560 PMID: 30423086

34. Fellows Yates JA, Lamnidis TC, Borry M, Andrades Valtueña A, Fagernäs Z, Clayton S, et al. Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. PeerJ. 2021; 9: e10947. https://doi.org/10.7717/peerj.10947 PMID: 33777521

35. Diroma MA, Modi A, Lari M, Sineo L, Caramelli D, Vai S. New Insights Into Mitochondrial DNA Reconstruction and Variant Detection in Ancient Samples. Frontiers in Genetics. 2021; 12(February):1–20. https://doi.org/10.3389/fgene.2021.619950 PMID: 33679884

36. Vinueza-Espinosa DC, Santos C, Martínez-Labarga C, Malgosa A. Human DNA extraction from highly degraded skeletal remains: How to find a suitable method? ELECTROPHORESIS. 2020; 41(24):2149–58. https://doi.org/10.1002/elps.202000171 PMID: 33002215

37. Vinueza-Espinosa DC, Cuesta-Aguirre DR, Malgosa A, Santos C. Mitochondrial DNA control region typing from highly degraded skeletal remains by single-multiplex next-generation sequencing. ELECTROPHORESIS. 2023; 44(17–18):1423–34. https://doi.org/10.1002/elps.202200052 PMID: 37379235

38. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature Biotechnology. 2011; 29(1):24–6. https://doi.org/10.1038/nbt.1754 PMID: 21221095

39. R Core Team. R: A Language and environment for statistical computing. (Version 4.1) [Computer software]. 2021. Retrieved from https://cran.r-project.org.

40. The jamovi project. jamovi. (Version 2.3) [Computer Software]. 2022. Retrieved from https://www.jamovi.org.

41. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. [Online]. 2010. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

42. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature Genetics. 1999; 23 (2):147–147. https://doi.org/10.1038/13779 PMID: 10508508

43. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009; 25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

44. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016; 32(2):292–4. https://doi.org/10.1093/bioinformatics/btv566 PMID: 26428292

45. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. 2012. arXiv preprint arXiv:1207.3907 [q-bio.GN].

46. Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. Schneidman-Duhovny D, editor. PLOS Computational Biology. 2022; 18(5):e1009123. https://doi.org/10.1371/journal.pcbi.1009123 PMID: 35639788

47. Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt HJ, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Research. 2016; 44(W1):W58–63. https://doi.org/10.1093/nar/gkw233 PMID: 27084951

48. Gilbert MTP, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, Lynnerup N, et al. Distribution patterns of postmortem damage in human mitochondrial DNA. American Journal of Human Genetics. 2003; 72 (1):32–47. https://doi.org/10.1086/345378 PMID: 12489041

49. Holland CA, McElhoe JA, Gaston-Sanchez S, Holland MM. Damage patterns observed in mtDNA control region MPS data for a range of template concentrations and when using different amplification approaches. International Journal of Legal Medicine. 2021; 135(1):91–106. https://doi.org/10.1007/s00414-020-02410-0 PMID: 32940843

50. Lamers R, Hayter S, Matheson CD. Postmortem miscoding lesions in sequence analysis of human ancient mitochondrial DNA. Journal of Molecular Evolution. 2009; 68(1):40–55. https://doi.org/10.1007/s00239-008-9184-3 PMID: 19067027

51. Rathbun MM, McElhoe JA, Parson W, Holland MM. Considering DNA damage when interpreting mtDNA heteroplasmy in deep sequencing data. Forensic Science International: Genetics. 2017; 26:1–11. https://doi.org/10.1016/j.fsigen.2016.09.008 PMID: 27718383

52. Sun S, Osterman MD, Li M. Tissue specificity of DNA damage response and tumorigenesis. Cancer Biology & Medicine. 2019; 16(3):396–414. https://doi.org/10.20892/j.issn.2095-3941.2019.0097 PMID: 31565474

53. Xavier MJ, Nixon B, Roman SD, Aitken RJ. Improved methods of DNA extraction from human spermatozoa that mitigate experimentally-induced oxidative DNA damage. PLoS ONE. 2018; 13(3):1–23. https://doi.org/10.1371/journal.pone.0195003 PMID: 29579126

54. Stoler N, Nekrutenko A. Sequencing error profiles of Illumina sequencing instruments. NAR Genomics and Bioinformatics. 2021; 3(1):1–9. https://doi.org/10.1093/nargab/lqab019 PMID: 33817639

55. Naue J, Hörer S, Sänger T, Strobl C, Hatzer-Grubwieser P, Parson W, et al. Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA. Mitochondrion. 2015; 20(1):82–94. https://doi.org/10.1016/j.mito.2014.12.002 PMID: 25526677

56. Li M, Schröder R, Ni S, Madea B, Stoneking M. Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112(8):2491–6. https://doi.org/10.1073/pnas.1419651112 PMID: 25675502

57. Irwin JA, Saunier JL, Niederstätter H, Strouss KM, Sturk KA, Diegoli TM, et al. Investigation of Heteroplasmy in the Human Mitochondrial DNA Control Region: A Synthesis of Observations from More Than 5000 Global Population Samples. Journal of Molecular Evolution. 2009; 68(5):516–27. https://doi.org/10.1007/s00239-009-9227-4 PMID: 19407924

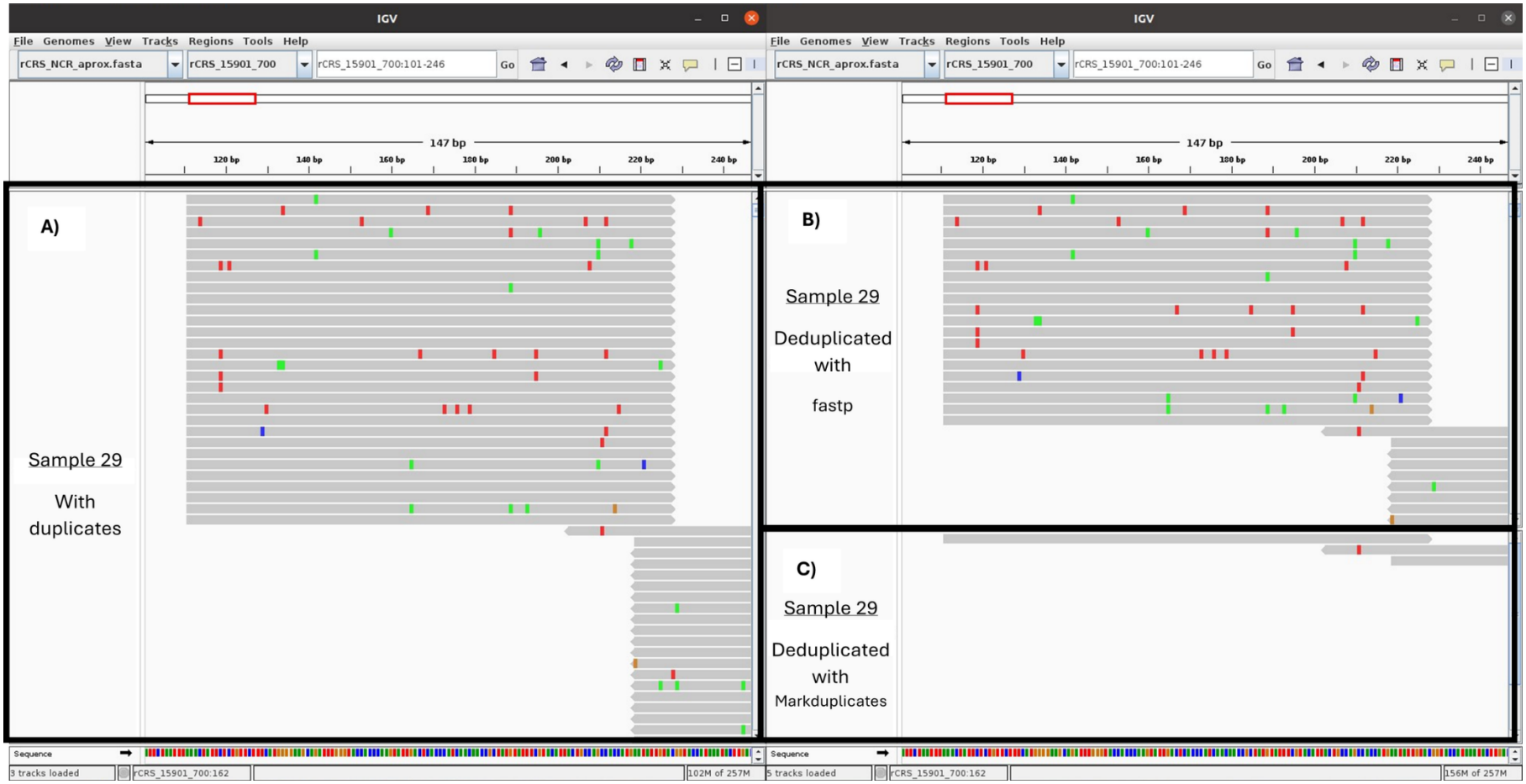
58. Ayad LAK, Pissis SP. MARS: improving multiple circular sequence alignment using refined sequences. BMC Genomics. 2017; 18(1):86. https://doi.org/10.1186/s12864-016-3477-5 PMID: 28088189

59. Peltzer A. CircularMapper (v1.93.5) [Computer software]. 2016. Retrieved from https://github.com/apeltzer/CircularMapper.

60. Oliva A, Tobler R, Llamas B, Souilmi Y. Additional evaluations show that specific BWA-aln settings still outperform BWA-mem for ancient DNA data alignment. Ecology and Evolution. 2021; 11(24):18743–8. https://doi.org/10.1002/ece3.8297 PMID: 35003706

61. Ramos A, Santos C, Mateiu L, Gonzalez M del M, Alvarez L, Azevedo L, et al. Frequency and Pattern of Heteroplasmy in the Complete Human Mitochondrial Genome. Mishmar D, editor. PLoS ONE. 2013; 8(10):e74636. https://doi.org/10.1371/journal.pone.0074636 PMID: 24098342

62. Santos C, Sierra B, Álvarez L, Ramos A, Fernández E, Nogués R, et al. Frequency and Pattern of Heteroplasmy in the Control Region of Human Mitochondrial DNA. Journal of Molecular Evolution. 2008; 67(2):191–200. https://doi.org/10.1007/s00239-008-9138-9 PMID: 18618067



S1 Fig. Screenshots from the IGV program for sample 29.

Black frames delimit each stage of the sample: A) with duplicated reads; B) without duplicated reads, removed by fastp software; C) without duplicated reads, removed by MarkDuplicates software. Grey arrows symbolize the aligned reads and coloured rectangles indicate a change of base from the reference sequence (red for Thymine, green for Adenine, blue for Cytosine, and brown for Guanine).

4.3. Article 3: The Genetic Landscape of Northeastern Iberian Communities from the Early to Late Iron Age

Authors

Daniel R. Cuesta-Aguirre, Xavier Roca-Rada, Beatrice Di Biase, Diana C. Vinueza-Espinosa, Carolina Sandoval-Ávila, Shyamsundar Ravishankar, Nuria Molist Capella, Montserrat Duran i Caixal, Imma Mestres Santacreu, Emili Junyent, Natalia Alonso, Jordi Martinez Majoral, Anna Kjær Knudsen, Liam T. Lanigan, Hannes Schroeder, Bastien Llamas, Assumpció Malgosa, Cristina Santos

Journal

iScience

Impact factor (JCR) and quartile ranking

4.6 and Q1

Date of publication

ISCIENCE-D-25-05072 Submitted 08/04/2025

DOI

Preprint available at http://dx.doi.org/10.2139/ssrn.5211463

Addressed objectives

This third publication investigates the origin and evolution of Northeastern Iberian communities and addresses the previously defined objective 3a:

3) To investigate the genetic origin and evolution of Northeastern Iberian communities using the newborns found buried under the houses as a proxy of the Iberian populations, more specifically:

    a. To provide a paleogenomic characterization of different Iberian sites to evaluate the genetic evolution of the local population over time, from the Early Iron Age to the Roman Conquest period, providing insights into genetic ancestry, admixture, and demographic dynamics in the Northeastern region of the Iberian Peninsula.

Supplementary data is available in "Article 3" at the following link: https://uab-my.sharepoint.com/:f:/g/personal/2103711_uab_cat/EhLQT7GxLs1GoowdsaXF7QYBF7DKCwofUOyqhbjFw4XWRw?e=3ddWqy

The Genetic Landscape of Northeastern Iberian Communities from the Early to Late Iron Age

Daniel R. Cuesta-Aguirre[1], Xavier Roca-Rada[2,3,4,5], Beatrice Di Biase[1], Diana C. Vinueza-Espinosa[6], Carolina Sandoval-Ávila[1], Shyamsundar Ravishankar[2], Nuria Molist Capella[7], Montserrat Duran i Caixal[8], Imma Mestres Santacreu[8], Emili Junyent[9], Natalia Alonso[9], Jordi Martinez Majoral[9], Anna Kjær Knudsen[10], Liam T. Lanigan[10], Hannes Schroeder[10], Bastien Llamas[2,11,12,13], Assumpció Malgosa[1], Cristina Santos[1]*

[1]Grup de Recerca en Antropologia Biològica (GREAB), Unitat d'Antropologia Biològica, Departament of Aniam Biology, Plant Biology and Ecology, Faculty of Biosciences, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

[2]Australian Centre for Ancient DNA, The Environment Institute, School of Biological Sciences, The University of Adelaide, Adelaide, South Australia, Australia.

[3]Faculty of Arts and Humanities, University of Coimbra, Largo da Porta Férrea, 3004-530 Coimbra, Portugal

[4]Department of Ecology, Evolution, and Organismal Biology, Brown University, Providence, RI, USA

[5]Center for Computational Molecular Biology, Brown University, Providence, RI, USA

[6]Universidad de Granada-Laboratorio de Identificación Genética (LABIGÉN), Spain

[7]Museu d'Arqueologia de Catalunya, Barcelona, Barcelona, Spain

[8]Jaciment i Museu El Camp de les Lloses, Tona, Barcelona, Spain

[9]Universitat de Lleida, Lleida, Spain

[10]Section for Molecular Ecology and Evolution, Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[11]Australian Research Council Centre of Excellence for Australian Biodiversity and Heritage (CABAH), School of Biological Sciences, The University of Adelaide, South Australia, Australia

[12]National Centre for Indigenous Genomics, John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia

[13]Indigenous Genomics, The Kids Research Institute Australia, Adelaide, Australia.

***Corresponding Author**

Cristina Santos, PhD

Research Group in Biological Anthropology

Unitat Antropologia Biològica

Dep. Biologia Animal, Biologia Vegetal i Ecologia

Facultat Biociències Edifici C

Universitat Autònoma de Barcelona

08193 Cerdanyola del Vallès

Barcelona (SPAIN)

Tel.: +34 93 581 13 17

Fax: +34 93 581 13 21

https://orcid.org/0000-0002-7014-8014 cristina.santos@uab.cat

Information about the authors: ORCID and mail

- Daniel R. Cuesta-Aguirre; 0000-0001-7483-8541; daniel.ruizdelacuesta@uab.cat
- Xavier Roca-Rada; 0000-0001-7502-6270; xrocarada@gmail.com
- Beatrice Di Biase; 0009-000-4914-031X; beatrice_dibiase@yahoo.com
- Diana C. Vinueza-Espinosa; 0000-0001-9931-2108; diana.vinueza@outlook.com
- Carolina Sandoval-Ávila; 0000-0001-7678-1390; carolina.sandoval@uab.cat
- Shyamsundar Ravishankar; 0000-0003-3006-6134; shyamsundar.ravishankar@adelaide.edu.au
- Nuria Molist Capella; nmolist@gencat.cat
- Montserrat Duran i Caixal; montserratduranicaixal@gmail.com
- Imma Mestres Santacreu; mestressi@tona.cat
- Emili Junyent; ejunyent@historia.udl.cat
- Natalia Alonso; natalia.alonso@udl.cat
- Jordi Martinez Majoral; jordi.martinez@udl.cat
- Anna Kjær Knudsen; 0009-0007-8743-554X; anna.kjaer.knudsen@sund.ku.dk
- Liam T. Lanigan; 0000-0001-6415-3506; liam.lanigan@sund.ku.dk
- Hannes Schroeder; 0000-0002-6743-0270; hschroeder@sund.ku.dk
- Bastien Llamas; 0000-0002-5550-9176; bastien.llamas@adelaide.edu.au
- Assumpció Malgosa; 0000-0003-1723-3671; assumpcio.malgosa@uab.cat

Abstract

During the Iron Age (800-100 BCE), the Iberian Peninsula was home to diverse cultural groups. Indigenous Iberians in the eastern coastal regions and Celtic-speaking populations in central and western regions interacted with Mediterranean civilizations, including the Phoenicians and Greeks during the Early Iron Age, and later Punics/Carthaginians and Romans. Iberian culture developed along the Mediterranean coast around the 6th century BCE, flourishing between 450-200 BCE. While archaeological evidence is abundant, genetic studies remain scarce due to predominant cremation practices. However, the tradition of burying newborns beneath houses provides a unique opportunity for paleogenomic analysis. This study investigates the genetic profile of northeastern Iberians from the Early to Late Iron Age by analyzing 54 newborns from three sites: the fortress of Vilars (775–325 BCE), Sant Miquel d'Olèrdola (350–200 BCE), and El Camp de les Lloses (125 BCE–50 CE). More than 20k SNPs from the 1240k panel are reported for 22 individuals, and mitochondrial data for a further 9 individuals. Our findings support the archaeological hypothesis that Iberians emerged from local Bronze Age groups, evolving gradually through the Iron Age, with increased Mediterranean ancestry. This genetic continuity persisted until the Late Iron Age and the arrival of the Romans. From then on, the new genetic influences led to a more diverse Iberian-Roman population alongside significant social and political transformations. This study provides the first paleogenomic insights into the enduring genetic legacy of Iberian communities, highlighting the complex interplay between cultural and genetic continuity in the northeastern Iberian Peninsula.

Keywords

Iberians; Iron Age (IA); paleogenomics; ancient DNA (aDNA); Iberia; Iberian Peninsula; population genomics; molecular anthropology

Introduction

The Iberian Peninsula, like many other territories worldwide, is characterized by its continual integration of diverse populations. Whether through historical conflicts and conquests or peaceful economic transitions, a complex cultural mosaic has been woven into this southwestern part of the Eurasian continent[1,2]. By exploring the genetic landscape of this region, recent studies have elucidated the ancient populations that inhabited different parts of the peninsula.

The Iberian Peninsula served as a refuge during the Last Glacial Maximum (LGM, 26,000-19,000 years ago), with genetic continuity observed from pre-LGM to post-LGM populations[3,4]. In the Mesolithic, northern hunter-gatherers from the Iberian Peninsula exhibited stronger genetic influence from Western Hunter-Gatherers (WHG) compared to those in the south, reflecting dynamic genetic shifts[2]. In the Early Neolithic, Anatolian farmers expanded into Europe, leading to a period of coexistence with local hunter-gatherers. This interaction resulted in higher Anatolian Neolithic ancestry in Early Neolithic populations, which gradually diluted in favor of WHG ancestry by the end of the Neolithic and the beginning of the Chalcolithic[3,5–9]. This effect also took place in the Iberian Peninsula[2,10,11]. The Chalcolithic in the Iberian Peninsula was characterized by continuity from the Neolithic[1,2,12].

During the Bronze Age, different groups carrying Pontic-Caspian Steppe ancestry spread westward throughout Europe. In the Iberian Peninsula, the Early Bronze Age (2200–1500 BCE) was marked by the introduction of Steppe-related ancestry, although its influence varied regionally. Steppe ancestry was more prominent in the northern and northeastern regions of the Iberian Peninsula, while southern regions exhibited less impact, suggesting a gradual and uneven genetic transition during this period. This period also saw changes in Y-chromosome lineages, with R1b replacing I2, G2, and H as the dominant lineage, although these earlier lineages persisted at lower frequencies[1,2,13,14].

Throughout the Iron Age (Figure 1), diverse cultures coexisted in the Iberian Peninsula (800-100 BCE). Indigenous groups included different Iberian tribes (described in more detail afterward), Tartessians, and Celtic-influenced groups like the Celtiberians. There were also Celtic and Indo-European (pre-Celtic) groups such as the Lusitanians, along with other regional groups including the Gallaeci, Astures, Cantabri, Vascones, Vaccaei, Vettones, Turduli, and Turdetanians. It is important to note that the boundaries of these groups in terms of time and location were fluid, and cultural blending was common. Additionally, according to the archaeological record, the establishment of Mediterranean colonies in the Iberian

Peninsula led to the southeast becoming a center of Phoenician culture, while the northeast saw the formation of Greek colonies[15].

During the 3rd century BCE, the Carthaginian Empire (also known as Punics) and the Roman Republic emerged as dominant powers in the western Mediterranean. The Carthaginians, who descended from the Phoenicians, established a vast maritime empire, while Rome became the leader and protector of the Italian Peninsula. Their expansionist ambitions led to the Punic Wars (264–146 BCE). The Second Punic War (218–202 BCE) was particularly significant in the Iberian Peninsula, culminating in the Carthaginians' expulsion from the territory in 206 BCE. The Romans established control over the region and incorporated it into their Empire to prevent potential resurgence. The local population became increasingly Romanized in the process, gradually adopting Roman customs and culture.[16].

During the Iron Age, the term “Iberians” was used by Greeks and Romans to refer to the communities inhabiting the eastern and southern Mediterranean coasts of the Iberian Peninsula. This designation includes a variety of independent tribes, such as the Layetani, Indiketes, Cessetani, Lacetani, Ilercavones, Ilergetae, and Ausetani, among others, each with its governance structures and customs that, while sharing similarities, were not identical[17]. The Iberian culture, which flourished during the Iron Age around the 6th century BCE, reflects the rich history of the Iberian Peninsula, with significant influences from other Mediterranean civilizations such as Phoenicians, Greeks, and Carthaginians[18,19].

The Iberians demonstrated significant advancement, with an economy based on metallurgy, agriculture, livestock, and, in coastal regions, fishing. They developed their own alphabet and currency system. In terms of religious practices, the Iberians adored a pantheon of deities, sharing several elements with other Mediterranean regions. Cremation was their main funeral rite[20]. However, intramural newborn burials in houses and productive areas have been documented[21–23], providing an important source of information for anthropological and paleogenomic studies. While this practice was not exclusive to Iberians, as they have been documented globally from prehistoric to medieval times[24–31], a debate over the ritual significance of infant burials has emerged[32–34]. However, recent evidence suggests that the age profile of newborns aligns well with natural mortality causes[35].

The first paleogenetic research on Iberian culture groups focused on studying mitochondrial DNA (mtDNA), specifically the hypervariable region I. These early studies aimed to identify the prevalent mitochondrial haplogroups among Iberians in comparison to

ancient individuals from different periods (such as the Chalcolithic, Bronze Age, and Roman era)[36] and cultural groups, including the Basques and Etruscans[37]. These studies, which analyzed the remains of 27 adults from 9 archaeological sites, often involved bone fragments found in regions historically occupied by Iberians but sometimes with questionable Iberian origin. The results revealed that the haplogroup composition of ancient Iberians was remarkably similar to that of modern Iberian Peninsula populations (with less variability), suggesting long-term genetic continuity in the region since pre-Roman times[37]. Moreover, they did not find a close genetic relationship between the Iberians and the Etruscans, indicating considerable genetic heterogeneity in pre-Roman Western Europe[37]. The main haplogroup in those populations was H, with lineages of J, T, K, U4, U5, and W also present[36,37]. Additionally, they found an unusually high frequency of haplogroup V, considered a marker of Paleolithic populations in Europe[36]. Advancements in genetic research methods, particularly the adoption of Next Generation Sequencing (NGS) for paleogenomic studies, have provided more detailed insights into both mitochondrial and nuclear DNA obtained from the remains of 15 individuals (3 newborns from Valencia and 12 adults) from 7 Iberian archaeological sites[2], finding that Iron Age individuals from the Iberian Peninsula had increased ancestry related to Northern and Central European populations compared to Bronze Age ancestry[2].

Currently, no systematic paleogenomic study has focused on Iberians that span the Early Iron Age to the Roman Conquest period, mostly due to widespread cremation practices during the Iron Age[38] resulting in a paucity of skeletal remains suitable for ancient DNA analysis. Therefore, in this paper, we will take advantage of newborn burials found in three archaeological sites from the northeast of the Iberian Peninsula characterized by Iberian culture, namely, the fortress of Vilars (VILA), Sant Miquel d'Olèrdola (OLE), and El Camp de les Lloses (CL) (detailed information in Supplementary Material). The fortress of Vilars, located in the territory of the Ilergetae tribe and active from the 8th to the 3rd century BCE, showcases the architectural evolution of a power center in pre-Iberian and Iberian cultures[39]. Sant Miquel d'Olèrdola, located in the territory of the Cessetani tribe and occupied by Iberians from the 4th to the 2nd centuries BCE, highlights artisanal activities and trade, with notable burials of perinatal children[21]. Finally, El Camp de les Lloses, which was located in the territory of the Ausetani tribe and dates from the Late Iron Age (2nd century BCE) to the 1st century CE, was a specialized complex related to Roman-republican army logistics, with metalworking as its main activity[40]. These sites present a unique opportunity to study the genetic characteristics of people living in the Northeastern part of the Iberian Peninsula

during the Iron Age and its changes over time. Our study aims to provide a paleogenomic characterization of the three Iberian sites to evaluate the genetic evolution of the local population over time, from the Early Iron Age to the Roman Conquest period, providing insights into genetic ancestry, admixture, and demographic dynamics in the Northeastern region of the Iberian Peninsula.

Results

Genome-wide data across the Iron Age in Iberian Culture

Thirty-six newborns out of fifty-four reached the necessary amount of extracted DNA for generating libraries (Figure 2B). The main results are displayed in Figure 2C (More extensive results of each library can be seen in Supplementary Material, Table S2). Out of the thirty-six sequenced individuals, VILA35 presented 1.5% of C-to-T misincorporations, so it was considered contaminated and was removed from further analyses. Moreover, fourteen individuals (OLE02, OLE05, OLE07, VILA01, VILA02, VILA03, VILA04, VILA08, VILA10, VILA11, VILA16, VILA19 and VILA21, and VILA29) exhibited fewer than 20,000 autosomal SNPs from the 1240k panel[9] and were removed from population analyses, except for VILA04, which presented almost 20,000 (19,394) autosomal SNPs from the 1240k panel[9] and was included in subsequent analysis. In contrast, four individuals (OLE08, VILA05, VILA15, and VILA18) had more than 100,000 autosomal SNPs from the 1240k panel[9]. The remaining eighteen individuals contained between 20,000 and 100,000 autosomal SNPs from the 1240k panel[9].

Genetic sex was estimated for all 36 individuals (Figure 2C). Almost 60 percent of the individuals (20 out of 35) were female. The highest female frequency was observed in Sant Miquel d'Olèrdola (0.75; 6 out of 8 individuals) and El Camp de les Lloses (0.67; 6 out of 9 individuals), compared to the fortress of Vilars, which had a frequency of 0.44 (8 out of 18 individuals).

The mitochondrial haplogroup was obtained for 31 individuals (Figure 2C and Supplementary Material, Table S4). Haplogroup H was the most common, making up 32.26% of the individuals (H1, H2, H3, H4), followed by haplogroup K at 22.58% (K1a, K1b) and haplogroup U at 16.13% (U4, U5a, U5b). The most frequent subhaplogroup was K1 (K1a3a at the fortress of Vilars, K1a+195 at both the fortress of Vilars and Sant Miquel d'Olèrdola, and K1b1a at El Camp de les Lloses). Significant diversity was observed across all sites, with various sublineages (H1, H3, H4, HV0, T2, K1, U4, U5a, U5b, N1, and M1) present in the

Iberian culture. At the fortress of Vilars, individuals VILA01 and VILA07 shared the same haplotype. Additionally, individuals VILA16, VILA18, and VILA21, which were buried in the same burial, also shared the haplotype. At the El Camp de les Lloses, CL6 and CL7 (H1at1) shared the same haplotype, and CL9 and CL12 (H3+152) also shared the same haplotype. In contrast, no shared haplotypes were detected at Sant Miquel d'Olèrdola.

The Y-chromosome haplogroup was determined in 10 out of 16 males (Figure 2C and Supplementary Material, Table S5). None of the male Y-chromosome results from Sant Miquel d'Olèrdola could be classified. Most individuals belonged to the R1b haplogroup, but VILA14 was classified under the H2a1 haplogroup. With the current sequencing resolution, it is not possible to determine if the R1b lineages are identical across individuals.

The main genetic relatedness results for all individuals with first- or second-degree relationships according to BREADR are presented in Table 1 (for complete results, see Supplementary Material, Table S6). The analysis conducted at the fortress of Vilars did not reveal any first- or second-degree relationships for the analyzed individuals. In Sant Miquel d'Olèrdola, OLE03 and OLE04 showed a second-degree relationship. At El Camp de les Lloses, a first-degree relationship was identified between CL6 and CL7, while second-degree relationships were observed in pairs CL7-CL8, CL2-CL6, and CL2-CL9. Among these, the relationship between CL2 and CL6 appears to be the most plausible, supported by a probability of 0.86 (Table 1).

Individuals OLE03, CL6, and CL9 were removed for *f*-statistics and qpAdm analyses based on populations.

Table 1. Posterior probabilities for each degree of relatedness in non-unrelated pairs with more than 500 shared SNPs. The probability is highlighted in green for first-degree and blue for second-degree when greater than 0.65. The "Pair ID" column corresponds to the pair ID in Table S6.

Pair ID	Pair of comparing individuals	#shared SNPs	Same/Twins	First-degree	Second-degree	Unrelated
38	**OLE04-OLE03**	608	< 0.0001	0.0451	**0.6591**	0.2958
48	**CL7-CL6**	2152	< 0.0001	**1.0000**	< 0.0001	< 0.0001
51	**CL7-CL8**	2695	< 0.0001	< 0.0001	0.5332	0.4668
54	**CL2-CL6**	1164	< 0.0001	0.0074	**0.8619**	0.1308
58	**CL2-CL9**	2025	< 0.0001	< 0.0001	**0.7243**	0.2757

The multidimensional scaling plot of 1-f_3(Ind1, Ind2; Mbuti) indicated no clustering based on chronology or sex when all Iberian populations from the Northeast of the Iberian Peninsula were analyzed together (Supplementary Material, figure S12). In each

archaeological site (Supplementary Material, Figures S13-S15), some clustering along dimensions 1 and 2 could be observed, but again it was not associated with sex, chronology, or disposition in the site.

Genetic continuity between the Bronze Age and the Early/Middle Iron Age in Iberians (775-200 BCE)

To investigate the genetic profile of the Iberians, we merged our dataset with a collection of ancient Mediterranean individuals (see Supplementary Material, Table S3 for labels). These individuals were projected onto the first principal components (PC1 and PC2) of a Principal Component Analysis (PCA) based on present-day populations from Europe, North Africa, the Caucasus, and Southwestern Asia. After filtering, 616,427 autosomal SNPs from the 1240k panel[9] remained for analysis. As indicated in the "Material and Methods" section, in this study, only newborns buried beneath the pavement of houses at Iberian sites will be considered to belong to Iberian culture. Individuals found at Iberian sites—such as partially cremated remains, exceptions to cremation, and decontextualized remains—will be referred to as "probably-Iberians".

Most individuals from the fortress of Vilars and Sant Miquel d'Olèrdola cluster closely with other Iron Age individuals. However, a few individuals exhibit a slight genetic shift towards present-day populations from the Caucasus, Southwestern Asia, and North Africa (Figure 3A). Notably, all Iberians analyzed here fall within the genetic variation established by prehistoric Iberian Peninsula individuals from the Neolithic to the Iron Age. Furthermore, many also align with prehistoric populations from the Italian Peninsula and Sardinia (Supplementary Material, Figures S10 and S11). Among the analyzed individuals, VILA15 is the only individual clustering closely with previously studied Iberians from the present-day territory of the Valencian Community (Figure 3B). Other individuals, such as VILA13, VILA18, and OLE03, fall within the broader Iron Age genetic cluster (Figure 3B).

To further assess these PCA-based observations, we performed outgroup f_3-statistics of the form f_3(IberianSite, Test; Mbuti). This analysis was used to identify the closest genetic matches for the Iberian individuals studied here. Our results indicated a higher shared genetic drift with Bronze Age and Iron Age populations from the Iberian Peninsula, including the Tartessians, Proto-Iberians, Celts, Greeks, Hellenistics, and local Romans from the Northeast of the Iberian Peninsula (territory of present-day Catalonia). In contrast, lower genetic affinity was observed with Roman populations from the West (territory of present-day Portugal), the South (territory of present-day Granada), and the Northeast of the Iberian

Peninsula. These results support genetic continuity from the Bronze Age to the Iron Age in the Northeast of the Iberian Peninsula (Supplementary Material, Figure S16 and Table S7).

To explore genetic relationships and potential admixture events, we applied f_4-statistics of the form f_4(IberianSite, Reference; Test, Chimp). Positive significant values indicate that the *Test* population shares more genetic similarity with the *IberianSite* population than with the *Reference* population. Chimpanzee (Chimp) was selected as the outgroup instead of Mbuti to facilitate the exploration of potential sub-Saharan genetic contributions. When using Bronze Age individuals from the entire Iberian Peninsula as the *Reference* population, no significant values were detected for the Early Iberians from the fortress of Vilars (Supplementary Material, Figure S17 and Table S8). However, these individuals showed a lower affinity to Bell Beaker populations than to preceding Bronze Age populations from the Northeast of the Iberian Peninsula. In contrast, Middle Iron Age individuals from Sant Miquel d'Olèrdola exhibited reduced Neolithic Anatolian ancestry compared to the Bronze Age population of the Iberian Peninsula. When restricting the *Reference* population to the Bronze Age population of the Northeast of the Iberian Peninsula, no significant differences were observed in the f_4-statistics (Supplementary Material, Figure S17 and Table S8).

To test whether Iberians from the Northeast of the Iberian Peninsula show genetic continuity with earlier regional populations, we applied qpAdm to model their ancestry as a mixture of populations from earlier periods in the region. The best models are plotted in Figure 4 (for a comprehensive overview, refer to Supplementary Material: the top models for all analyzed populations in Table S9 and all tested models in Table S10). The p-values differed among the various populations. However, the models for Sant Miquel d'Olèrdola — Chalcolithic Iberian Peninsula (0.776) + Bronze Age Steppe (0.224) with a p-value of 0.99, Chalcolithic Sardinia (0.442) + Bell Beaker Germany (0.558) with a p-value of 0.98, and Chalcolithic Iberian Peninsula (0.431) + Bell Beaker Germany (0.569) with a p-value of 0.94 — and Ullastret — Chalcolithic Iberian Peninsula (0.392) + Bell Beaker Germany (0.608) with a p-value of 0.92 — showed strong support, as their p-values exceed 0.9. The Northeast Bronze Age population and most of the Iberian and probably-Iberian populations could be modeled as a combination of Chalcolithic Iberian ancestry and Steppe-related ancestry (either Steppe BA or Steppe-related Bell Beaker groups from Germany), as shown in Figure 4 and Supplementary Material (Tables S9 and S10). Interestingly, Steppe-related ancestry was lower in the Bronze Age population from the Northeast of the Iberian Peninsula (0.498 ± 0.07) compared to the fortress of Vilars (0.782 ± 0.089), with non-overlapping standard deviations.

In contrast, the Steppe ancestry levels in the Northeast Bronze Age population overlapped with those observed in most other Iberian and probably-Iberian populations, including Ullastret (0.608 ± 0.086), Sant Miquel d'Olèrdola (0.569 ± 0.143), Castellví de la Marca (0.564 ± 0.155), and Mas d'en Boixos (0.474 ± 0.113) (Supplementary Material, Table S9).

Genetic interaction between Iberians and other Mediterranean cultures during the Iron Age in the Northeast of the Iberian Peninsula before Roman arrival (218 BCE)

Despite most Iberian individuals from the fortress of Vilars and Sant Miquel d'Olèrdola clustering within or around the general diversity of the Iron Age population from the Iberian Peninsula in the PCA, two individuals—VILA04 and OLE06—deviate significantly from the probably-Iberian cluster, shifting towards the Caucasus and Southwestern Asia (Figure 3). Additionally, VILA12 and OLE08 exhibited a distinct pattern, being displaced towards the North African cluster (Figure 4). Notably, qpAdm analysis indicated that the best-fitting ancestry models for the fortress of Vilars and Sant Miquel d'Olèrdola populations included a Chalcolithic Sardinian component alongside Steppe-related ancestry (Figure 4). This indicates that some interactions between Iberians and other Mediterranean populations could have occurred.

To explore these genetic connections in greater detail, we tested several qpAdm models for each individual. The results showed that VILA12 carried a mix of Mediterranean Chalcolithic and Steppe ancestries (Figure 5, Supplementary Material, Tables S11 and S12), while the top model for VILA04 does not include Mediterranean ancestry. Conversely, VILA13 exhibited high Mediterranean ancestry, lacking any Steppe component (Figure 5, Supplementary Material, Tables S10 and S11). The genetic landscape of Sant Miquel d'Olèrdola was more complex. OLE01 was the only individual without Mediterranean ancestry (Figure 5 and Supplementary Material, Tables S11 and S12). Meanwhile, OLE03 and OLE04 showed no additional Steppe-related ancestry beyond their Chalcolithic subtract. OLE06 was the only individual with Iranian or North African ancestry, presenting a North African mitochondrial haplogroup (Figure 5 and Supplementary Material, Tables S4, S11, and S12). Connections between Iberians and Mediterranean populations are also evident in other Iberian individuals, including individuals from present-day Valencia and I3327, a probably-Iberian individual from Ullastret (Northeast of the Iberian Peninsula) (Figure 5, Supplementary Material, Tables S11 and S12).

To better contextualize these findings, for all the Northeastern individuals and populations from the Bronze Age to the Roman period and some Mediterranean populations

from Sardinia, Sicily, and the Italian Peninsula, we selected the top qpAdm model for each individual, along with alternative models whose p-values differed by less than 0.1 from the best fit. We then estimated the mean ancestry proportions. The data generated was summarized with a PCA (Figure 6 and Supplementary Material, Figure S18). VILA12 clustered with the Italian Bronze Age population and Late Iberians from the Northeast of the Iberian Peninsula, while VILA13 was positioned near Etruscans (Iron Age population from the Italian Peninsula) and Sardinian Romans. Regarding Sant Miquel d'Olèrdola, OLE03 aligned with Roman and Punic populations from Sardinia and Romans from Empuries (territory of present-day Girona, Northeast of the Iberian Peninsula), OLE04 clustered with the Roman population from Mas Gassol (Territory of present-day Tarragona, Northeast of the Iberian Peninsula), and OLE06 reflected North African ancestry.

The genetic influence of Roman gene flow on Late Iberian populations (125 BCE – 50 CE)

Six out of the nine individuals from El Camp de les Lloses showed a clear deviation toward the North African cluster in the PCA, with only CL2 and CL7 clustering within the Iberian Iron Age perimeter (Figure 3). This pattern suggests a distinct genetic composition at this site, which dates to between 125 BCE and 50 CE—a period when Roman presence in the Northeast of the Iberian Peninsula was already well established.

To investigate genetic affinities, we applied outgroup f_3-statistic of the form f_3(IberianSite, Test; Mbuti), where *Test* were different Bronze Age, Iron Age, and Roman populations from the Iberian Peninsula. As observed in the other Iberian sites, the closest genetic matches for El Camp de les Lloses were populations from the Bronze and Iron Ages, and the local Roman population from the Northeast of the Iberian Peninsula, rather than Roman populations from the rest of the Iberian Peninsula (Supplementary Material, Figure S16 and Table S7). To examine the genetic relationships and potential admixture events between Late Iberians (El Camp de les Lloses) and Iberians, we performed f_4-statistics in the form of f_4(CL, Iberians; Test, Chimp), where *Iberians* consisted of unrelated individuals from the fortress of Vilars and Sant Miquel d'Olèrdola. No significant values were detected, suggesting that the overall genetic makeup of El Camp de les Lloses was still largely consistent with prior Iberian populations (Supplementary Material, Figure S17 and Table S8).

Although *f*-statistics did not reveal strong differentiation between El Camp de les Lloses and earlier Iberian populations, the PCA suggested a distinct genetic composition. To clarify this, we evaluated qpAdm models (Figure 4, supplementary material, Tables S11 and S12). The best-fitting model for El Camp de les Lloses included ancestries from the Sardinian

Chalcolithic (0.61) and Steppe-related Bronze Age populations (0.39), with a p-value of 0.315 (Figure 4 and Supplementary Material, Tables S11 and S12). Unlike the fortress of Vilars and Sant Miquel d'Olèrdola, where qpAdm models frequently included Mediterranean ancestry alongside Iberian sources, the top model for El Camp de les Lloses lacked Chalcolithic Iberian Peninsula ancestry (Figure 4 and Supplementary Material, Tables S11 and S12).

All individuals from El Camp de les Lloses could be modeled with two major ancestral components: Chalcolithic ancestry from either the Iberian Peninsula, Sardinia, or the Italian Peninsula, combined with Steppe-related ancestry from the Russian steppes or Neolithic Iranian ancestry. The exception was CL8, which also required a North African component in addition to Chalcolithic and Steppe ancestries (Figure 5 and Supplementary Material, Tables S11 and S12). Notably, all individuals that had a viable model with Chalcolithic Iberian ancestry also had an alternative model with Sardinian Chalcolithic ancestry at an equivalent p-value (Figure 5 and Supplementary Material, Tables S11 and S12).

In the qpAdm-PCA (Figure 6), the El Camp de les Lloses population and its individuals did not cluster within the Bronze/Iron Age Iberian Peninsula group (Cluster 1), where most individuals of the fortress of Vilars and Sant Miquel d'Olèrdola were located. Instead, they grouped within the Mediterranean cluster (Cluster 2), as seen in Figure 6 and Supplementary Material (Figure S18). A subset of individuals (CL2, CL3, and CL9) clustered near the Roman population from Mas Gassol (present-day Tarragona, Northeast Iberian Peninsula), while CL1, CL6, and CL12 showed affinities with specific Mas Gassol individuals (I6491, I6492, and I6490, respectively), though with some dispersion (Figure 6 and Supplementary Material, Figure S18). Finally, CL4 and CL8 exhibited genetic affinities to Iron Age populations from the Italian Peninsula, including the Etruscans and Daunians, and to a Roman individual from Empuries (present-day Girona, Northeast Iberian Peninsula), reinforcing the evidence for Roman genetic influence in late Iberian groups (Figure 6 and Supplementary Material, Figure S18).

Discussion

Although some individuals from the fortress of Vilars and El Camp de les Lloses could not be evaluated, male and female individuals were identified at all three sites, albeit in varying proportions. This variability is likely influenced by the limited sample size, which restricts the ability to draw definitive conclusions about sex distribution. Nevertheless, the

presence of both sexes at each site, combined with the natural mortality pattern observed by Martirosyan et al. (2024)[35], suggests that we have a representative sample of the population. This was also seen when genetic relatedness was analyzed, where few first- and second-degree relationships were found. No first- or second-degree relationships were found in the fortress of Vilars. However, some shared haplotypes were found among individuals, some buried together. The low number of common SNPs recovered for some pairs of individuals made it impossible to determine the genetic relatedness among them. It is interesting that in Sant Miquel d'Olèrdola, individuals OLE04 and OLE06 had previously been interpreted as potential twins due to their discovery near a wall, positioned in opposite directions with their legs intermingled and no sediment separating the two skeletons[22]. However, genome-wide analysis (Supplementary Material, Table S6) revealed that they were unrelated. Furthermore, they presented different mitochondrial haplogroups: OLE04 had haplogroup N1a1a1a3, and OLE06 had haplogroup M1. These findings conclusively rule out the possibility that they were twins. At El Camp de les Lloses, slightly more genetic relatedness patterns were observed, with a pair of sisters and two second-degree relationships identified.

Across the three analyzed Iberian sites, a high diversity of mitochondrial lineages is evident. Lineages such as H, J, T, K, N, and U are typically associated with Western Europe[41] and are related to the expansion of Neolithic farmers[42–44]. Most of them had already been detected in Iberian territory during the Iron Age[2,36,37] . However, the presence of the N1 subhaplogroup, which has also been linked to Southwest Asia and Northeast Africa[41], and sublineages of M1, characteristic of Northwest Africa[45,46], may suggest a broader genetic influence. Remarkably, this mitochondrial diversity is high despite the low number of individuals analyzed. This pattern is consistent with other Iron Age communities from the Iberian Peninsula, such as northern Celts[47] or northeastern Greek-Hellenistic populations[2], but contrast with British Celts, who exhibited less diversity[48]. Notably, haplogroup U4d2, typically associated with Eastern Europe[49,50] was found in the most recent individual (CL4: 50 BCE – 50 CE). Haplogroup V was not detected in any individual, contrasting with previous genomic studies made in Iberian territory[36]. However, that paper used 16298C as a defining position of haplogroup V, when nowadays it is used to define haplogroup HV0[51], which was frequent in our individuals.

Regarding Y-chromosome haplogroups, the expected predominance of R1b lineages —associated with the arrival of steppe-related individuals during the Bronze Age, which largely replaced earlier lineages in Europe and the Iberian Peninsula— is observed. However, the persistence of Neolithic Y-chromosome haplogroups, such as H2a1 (detected in VILA14),

highlights continuity with earlier populations. This haplogroup has been previously found in two Bronze Age individuals from the Czech Republic[52] and two Neolithic individuals from France[52]. Moreover, H2 sublineage has also been detected during the Late Neolithic and Chalcolithic Iberian Peninsula[2,9,11,13] and other European populations[53].

Our results indicate that the transition from the Bronze Age to the Iron Age in the Northeastern of the Iberian Peninsula occurred without significant external influence, aligning with previous studies[2]. This transition highlights the continuity of local communities, which evolved into the Iberians as proposed by Junyent, who emphasized the transformation of Late Bronze Age societies into hierarchical chiefdoms and eventually into archaic aristocratic states without significant external influence[54–56]. Notably, the Steppe-related ancestry was higher in the Iberian populations than in the Bronze Age population from the Northeast of the Iberian Peninsula, as it was observed by Olalde et al. (2019)[2]. Nonetheless, the six unrelated Bronze Age individuals that formed that Bronze Age population were from 4 different archeological sites and presented a great variability among them. More Bronze Age individuals from the Northeast of the Iberian Peninsula should be analyzed to validate the increase of Steppe-related ancestry during the Iron Age in that territory.

The fortress of Vilars aligns closely with Bronze Age and Iron Age populations, as well as with Iron Age individuals from various cultures across the Northeastern of the Iberian Peninsula. This observation is consistent with the material culture recovered from the site[57,58]. During the earlier phases (775-450 BCE), most of the pottery corresponded to pre-Iberian or Iberian tradition[57,58]. However, as early as 700-550 BCE, evidence of Phoenician influence emerged, including jewelry and pottery[57,58]. By 550-450 BCE, Punic- and Greek-influenced Iberian pottery became prominent, alongside Mediterranean or Mediterranean-influenced ivory artifacts and additional Phoenician jewelry[57,58]. Despite these cultural influences, genetic results suggest limited Mediterranean ancestry in the population overall. Notable exceptions include individuals like VILA12 (775–675 BCE) and VILA13 (550–450 BCE), which exhibit significant Mediterranean ancestry, highlighting individual variability within the broader population. Similar smooth Bronze Age/Iron Age transitions have been documented in the Iberian Peninsula[47] and the territory of present-day France[59].

The continuity from the Bronze Age can also be observed at Sant Miquel d'Olèrdola, despite a higher proportion of Mediterranean ancestry being detected at this site than at the fortress of Vilars. This could be related to the fact that Sant Miquel d'Olèrdola is closer to the coast, connected to the Mediterranean through the Vilafranca stream, which leads to two

potential ports at the Iberian sites of Darró and Sitges, located approximately 15-17 km away from Sant Miquel d'Olèrdola. Almost all individuals (OLE03, OLE04, OLE06, and OLE08) presented Mediterranean ancestry. OLE06 also presented Eastern Mediterranean ancestry (visible from Iran's Neolithic ancestry). This individual presented a mitochondrial haplogroup belonging to M1 (North African Ancestry). This mix could make us think of this individual as an outlier, a non-local individual. However, as a newborn who died at birth[35], the most probable conclusion is that we are facing a mixture of a local individual and a female with close Carthaginian/Punic ancestry. This theory could be supported by the fact that many Punic (Punic-Ebusitans and Central Mediterranean) amphorae and ceramics have been found through the Iberian stratigraphy of this site[60]. Italian and Greek ceramics were also found in this chronolgy[60], which could also be correlated with the Mediterranean ancestry found in OLE03 and OLE04. However, it should be noted that the presence of amphorae, ointment jars, and semi-luxury tableware at Iberian sites was associated to commercial exchanges[60].

During the Iron Age, Iberians exhibited a higher proportion of Bell Beaker North/Central European ancestry, linked to steppe-related ancestry. This pattern has also been identified by Olalde et al. (2021) in decontextualized Iron Age individuals from Iberian sites[2] and by Papac et al. (2023) in Celtic individuals from the North of the Iberian Peninsula[52]. However, neither of these studies identified Mediterranean ancestry in their individuals. Our analyses confirm that Iberians interacted with other Mediterranean cultures, such as the Phoenician, Punic/Carthaginian, and Italian. These interactions, corroborated by archaeological evidence, left a genetic imprint on Iberian populations, but the core genetic identity of Iberians remained intact, like what occurred in many other European Iron Age populations[59,61]. While the absence of Mediterranean ancestry is consistent with the Celtic individuals from the inland territory[52], it is less expected for the Eastern Iberian individuals analyzed by Olalde et al. (2021)[2] because they are located next to the Mediterranean Coast of the Iberian Peninsula. In our analyses, we reexamined these Northeastern probably-Iberian and Eastern Iberian individuals and identified Mediterranean ancestry in just a few individuals when analyzing them individually. This finding suggests that these individuals may not have been part of Iberian culture or may have belonged to isolated tribes or communities with limited interaction with Mediterranean populations. Alternatively, they could be individuals of non-Mediterranean ancestry within a larger population including individuals with Mediterranean ancestry, as observed in Sant Miquel d'Olèrdola.

The arrival of the Romans marked a gradual but transformative shift in Iberian communities, both culturally and genetically. This change is evident in the results from El Camp de les Lloses. Although *f*-statistics did not reveal strong differentiation between El Camp de les Lloses and earlier Iberian populations, the PCA suggested a distinct genetic composition, implying a continuity of the population with some changes. The qpAdm analysis for the El Camp de les Lloses population revealed values more similar to Bronze Age or Iron Age Italian populations. Notably, in the PCA based on the top qpAdm results, no individual clustered with the Bronze Age or Iron Age populations of the Iberian Peninsula. Any of these findings are surprising, given that El Camp de les Lloses is associated with metalworking and logistical activities tied to the Roman military[62]. In fact, a fusion of Iberian and Roman-Italian cultures is evident in the site's architecture, which incorporates Roman-Italian styles, such as rectangular houses with central courtyards, while utilizing indigenous building materials like stone bases and earthen walls[62]. Burial practices at the site align with Iberian traditions (e.g., cremation), and pottery assemblages include a mix of Iberian and Roman-Italian artifacts[62]. Additionally, most of the coins recovered (90%) are of Iberian origin, although Roman and Greek coins from Massaliot (present-day Marseille) have also been found[62].

Individuals CL2 and CL3 were the closest to the Bronze Age/Iron Age Iberian Peninsula cluster, yet all individuals exhibited a higher proportion of Mediterranean ancestry than previous Iberian populations. While CL1 showed the strongest influence on Iranian Neolithic ancestry, CL8 was characterized by a greater North African Neolithic ancestry. Interestingly, both individuals presented mitochondrial lineages typical of the Iberian Peninsula (K1b1a for CL1 and H3 for CL8) and Western Europe[63,64], suggesting that the Mediterranean and North African ancestries might have been introduced through the paternal line. As previously noted, the Roman influence at the site is evident and could have contributed to the introduction of Mediterranean or North African ancestries. However, the presence of pottery from Ebusus (a Phoenician city in present-day Ibiza, Balearic Islands), the Tripoli region (North Africa), and the southern Iberian Peninsula (around the Strait of Gibraltar)[62] suggests that these ancestries could also have been introduced through Phoenician-Punic or Carthaginian contacts. CL2, who carried the mitochondrial haplogroup M1a3a, which is most commonly found in North and East Africa, was a different case. In this individual, no North or East African influence was detected in nuclear DNA. While M1a lineages have been previously identified in the Iberian Peninsula—such as M1a1b1 in a Bell Beaker individual from the central region[2]—the specific M1a3a sublineage had not been documented in ancient individuals from the Iberian Peninsula before CL2. This may suggest

that M1a3a may have been introduced through an earlier migration or contact but remained at very low frequencies. Over time, any associated North African nuclear ancestry may have been diluted through admixture, leaving the mitochondrial haplogroup as the only genetic trace of that ancestry in CL2.

This study provides the first paleogenomic insights into the origin and evolution of Iron Age Iberian communities belonging to Iberian culture, focusing on newborns interred under the pavements of houses and productive areas. Our findings confirm the archaeological theory that Iberian culture emerged from local Bronze Age populations, evolving gradually through the Iron Age. While interactions with other Mediterranean cultures contributed to the development of Iberians, their genetic identity remained largely intact. However, with the arrival of the Romans, new genetic influences were introduced, leading to a more diverse Iberian-Roman population alongside significant social and political transformations. This study highlights the complex and enduring genetic legacy of Iberian communities.

Material and methods

Sample preparation, DNA extraction, quantification, library preprataion and sequencing

Fifty-four newborns were selected from three Iberian sites: 34 from the fortress of Vilars, 8 from Sant Miquel d'Olèrdola, and 12 from El Camp de les Lloses (Figure 1 and Supplementary Material, Table S1).

Fifteen samples from 8 newborns from El Camp de les Lloses were processed in the dedicated clean laboratory facilities at the Globe Institute, University of Copenhagen (indicated as Protocol I in Supplementary Material, Table S2). Around 100-200 mg of bone was used for DNA extraction. Powder bone was predigested for 10 minutes with 1 ml of digestion buffer DB1 (0.45 M EDTA, 0.25 mg/ml Proteinase K). After discarding the supernatant, 4 ml of digestion buffer DB1 were added to the sample, which was incubated at 37ºC overnight. Extracted DNA was purified using a silica-in-solution method. First, 40 mL of binding buffer (5 M Gu-HCl, 100 mM NaOAc, 20 mM NaCl, 30% isopropanol) and 50 µl of silica pellets were mixed with the digestion supernatant. Then, the samples were rotated at room temperature for 1 hour. Following incubation, the samples were spun at 2000 rpm for 5 minutes, and the supernatant was discarded. Next, two cleaning steps were performed using 1 ml of ice-cold 80% ethanol resuspending the pellet, after that the supernatant was removed and the samples were left to air-dry briefly to allow any remaining ethanol to evaporate.

Finally, the pellets were resuspended with 62 µl of EB buffer and incubated on a shaker at 37ºC for 10 minutes. After centrifugation 60 µL purified extracted DNA were stored at -20ºC. Blank controls were processed with each extraction.

For the library preparation, 20 µl of each DNA extract were built into single-stranded libraries using the SCR protocol[65]. For each sample, 50 µl of the library was recovered and purified using SPRI beads. The optimal number of PCR cycles was determined by qPCR (MxPro 3000, Agilent Technologies). The qPCR reactions consisted of 1 µl library template with 19 µl Master Mix containing (12.3 µl H_2O, 2µl Taq Gold buffer, 2µl $MgCl_2$, 0.8 µl BSA, 0.8 µl SYBR green/ROXY, 0.4 µl Forward primer IS4, 0.4 µl Reverse primer P7 index primer #50 (10µM), 0.16 µl dNTPs (25 mM), 0.16 µl Taq Gold polymerase). Library indexing was performed using the KAPA HiFi HotStart Uracil+ kit (Roche). Finally, the amplified libraries were purified using SPRI beads and quantified on a Fragment Analyser (Agilent). The amplified and indexed libraries were then pooled in equimolar amounts and sequenced with Illumina technology.

The remaining samples (Supplementary Material, Table S2) were processed in the dedicated clean laboratory facilities at the Universitat Autònoma de Barcelona (UAB). Around 100-150 mg of bone powder was used for DNA extraction in a 48-hour incubation with a solution of Proteinase-K and EDTA as in Vinueza-Espinosa et al. (2020)[66]. Extracted DNA was purified as described by Vinueza-Espinosa et al. (2020)[66], using the High Pure Viral Nucleic Acid Large Volume Kit (Roche), a Silica in HE-membrane column-based method, and eluted in 60 µl of EBT. Blank controls were processed with each extraction. A quantitative PCR (qPCR) of a small mitochondrial DNA fragment was performed to confirm the absence of contamination. Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific, 2015) was used to quantify each sample, and Bioanalyzer® Agilent High Sensitivity DNA Kit (Agilent Technologies, 2017) was employed to obtain fragmentation profiles.

Only extractions with at least around 0.5 pmol of DNA in the input volume were used to make whole genome libraries using protocols II to IV (described afterward). Protocol V was used to generate a library and sequence the non-coding region (NCR) of mtDNA to some samples with bad mitochondrial DNA recuperation. The type of library preparation protocol for each sample is described in Supplementary Material, Table S2.

Protocol II - Blunt-End-Single-Tube (BEST): 32 µl of each DNA extract were used in the BEST protocol[67]. For each sample, 60 µl of the non-indexed library was recovered and purified by MiniElute Reaction Cleanup Kit (Qiagen). The optimal number of PCR cycles was determined by qPCR (Rotor-Gene SYBR Green, Qiagen). The qPCR reaction consisted of

1 µl of 1:100 diluted sample with 24 µl of Master Mix (12.5 µl of SYBR Green PCR Master Mix, 2.5 µl of each P5 (IS7) and P7 (IS8) primers at 1 µM, and 6.5 µl of RNase-free water. Library indexing was performed using the KAPA HIFI HotStart Uracil+ kit (Roche). The indexed libraries were purified with the MiniElute PCR kit (Qiagen). 3 µl of the purified library were diluted 1:2. From here, 3 µl were quantified with Qubit™ 1X dsDNA HS Assay Kit (Invitrogen), and the other 3 µl were used to study the fragmentation profile with a High Sensitivity DNA Assay in a 2100 Bioanalyzer Instrument (Agilent Technologies) at the Servei de Genòmica i Espectroscòpia de Biomolècules (SGiEB) from the UAB, Barcelona. The amplified and indexed libraries were then pooled in equimolar amounts and sequenced with Illumina technology.

Protocol III - Blunt-End-Single-Tube (BEST) with half-UDG: 25 µl of each DNA extract were partially treated with Uracil-DNA-Glycosylase (UDG) with 7 µl of Master Mix (0.96 µl of Thermolabile USER® II (New England BioLabs), 3.2 µl of CutSmart Buffer (New England BioLabs) and 2.84 µl of EBT (Qiagen) following the incubation times from Rohland et al. (2015)[68]. The resulting 32 µl were used in the BEST protocol as explained in "Protocol II".

Protocol IV: "Protocol II" was followed but with half of the volume for each step.

Protocol V: for VILA16, VILA21, OLE07, and CL3, NCR of mitochondrial DNA was sequenced using the PowerSeq® CRM Nested System (Promega), considering the protocol for archaeological samples described by Vinueza-Espinosa et al. (2023)[69]. Samples were sequenced on an Illumina instrument.

aDNA data processing and authentication of whole genome sequencing

After demultiplexing based on a dual-unique index combination developed by the sequence provider, raw sequencing data was processed with nf-core/eager pipeline v2.4.6[70,71]. Adapters were removed with AdapterRemoval v2.3.2[72] and sequences shorter than 30 bp were removed. Then, sequences were mapped against the Human Reference Genome GRCh37 (hg19, GenBank accession GCA_000001405.1) with BWA v0.7.17-r1188[73] aln and samse/sampe commands. Parameters "-l 1024 -n 0.01 -o 2" were used, following recommendations from Oliva and colleagues[74,75]. SAMtools v1.12[76] was used to generate sorted BAM files with all the reads with a mapping quality of over 20. Duplicates were removed with the MarkDuplicates tool from Picard v2.26.0 software[77]. Damage levels were assessed with PMDTtools v0.50[78]. Finally, libraries were clipped with BamUtils v1.1.2[79] based on the pre-library treatment. In libraries with half-UDG treatment, 2 bases were clipped at the

start and end of each read. However, 7 bases were removed at the start/end of the reads for non-UDG treated libraries. The quality of FASTQ files was evaluated with FastQC v0.11.9[80] while the quality of the BAM file was analyzed with QualiMap v2.2.2-dev[81].

Contamination, sex determination, and genotyping

Both Methods 1 and 2 in ANGSD[82] were used to estimate contamination, with a threshold of 3%. In addition, mtDNA was manually checked for inconsistencies. Samples with a percentage of endogenous DNA without duplicates higher than 1% and a C-to-T misincorporation pattern under 10% were excluded.

Genetic sex was determined with Sex.DetERRmine.py v1.1.2[83], a tool implemented in the nf-core/eager pipeline, based on SNPs. Individuals were classified as male if the frequency of reads assigned to the Y- and X- chromosomes was around 0.5, while they were classified as female if the Y-chromosome frequency was 0 and the X-chromosome was around 1.

Trimmed BAM files were used to call genotypes with pileupCaller, which generates pseudo-haploid genotypes by randomly choosing one allele at every SNP position. The panel with 1240k nuclear SNPs used in ancient population genomics was used[9].

Mitochondrial and Y chromosome haplogroups assignment

Mitochondrial reads that had already been mapped with the mitochondrial DNA for the hg19 genome were realigned against the revised Cambridge Reference Sequence (rCRS, GenBank accession NC_012920.1)[80] with nf-core/eager pipeline v2.4.6[70,71]. CircularMapper v1.0[84] was used to avoid the problems of mtDNA molecule circularity. Genotypes were called with freebayes v1.3.6[85] and vcflib v1.0.3[86]. Haplogrep 3 v3.2.1[87] was used to call mitochondrial DNA haplogroups. SNPs were manually double-checked visualizing alignments with the Integrative Genome Viewer (IGV) v2.11.2[88]. Mitochondrial DNA data from samples sequenced using “protocol V” were analyzed using the pipeline developed by Cuesta-Aguirre et al. (2024)[89], which already provides the haplotype. Manually double-checking of the genotypes was done with IGV, as recommended. Data from NCR and whole genome sequencing were combined. Results from the NCR analysis were considered reliable at a coverage of 10X, while the shotgun sequencing results were considered at 3X after visual confirmation. In the low-coverage regions, the damage, mutational, and heteroplasmic profiles[90–92] were examined to improve credibility.

Based on the merged and trimmed BAM files, the Y-chromosome genotype was determined for each individual using UnifiedGenotyper from Picard v2.26.0 (Broad Institute, 2019). For each individual, a BED file including all SNPs with a base quality greater than 30 and belonging to reads with a mapping quality exceeding 20 was created. The python2 script hFrpr2.py from HaploGrouper[93], alongside the default tree and the SNP file from the Ghr37 reference genome, was used to obtain the Y-chromosome haplogroups (https://gitlab.com/bio_anth_decode/haploGrouper).

Genetic relatedness

Only individuals with more than 20,000 autosomal SNPs from the 1240k panel[9] were used to study genetic relatedness. BREADR v1.0.2[94] was used based on autosomal SNPs to define first- and second-degree relatedness among pairs of individuals. The pairs of individuals were categorized between "Same/Twins", "First-degree", "Second-degree" or "Unrelated" (in the first or second degree), according to their pair mismatch rate (PMR). Only pairs with at least 500 shared SNPs were trusted.

Population genetic analysis

Dataset

Newly reported individuals were selected based on the following criteria: 1) individuals with more than around 20,000 autosomal SNPs from the 1240k panel[9] were included; 2) Among individuals from the same archaeological site sharing the same mitochondrial haplotype, the individual with the higher number of SNPs was selected; 3) For related individuals with a posterior probability of first- and second-degree relatedness higher than 0.6, the individual with the higher number of SNPs was chosen. These selected individuals were merged with previously published datasets of ancient and modern individuals from the Allen Ancient DNA Resource (AADR)[53]. Only one of the individuals was considered when there was first- or second-degree of relatedness. Ancient North African individuals[95] that were not included at the AADR were downloaded from the European Nucleotide Archive and followed the same bioinformatic analysis as our individuals (see the "*aDNA data processing and authentication of whole genome sequencing*" and "*Sex determination, contamination, and genotyping*" sections). individuals from Human Origins[96] were also included for principal component analysis (PCA). Individuals were labeled for each analysis as indicated in Supplementary Material (Table S3).

In this study, only newborns buried beneath the pavement of houses at Iberian sites will be considered to belong to Iberian culture. Individuals found at Iberian sites—such as partially cremated remains, exceptions to cremation, and decontextualized remains—will be referred to as "probably-Iberians" hereafter.

PCA

A principal components analysis (PCA) was performed using the smartpca software from EIGENSOFT v8.0.0[97,98] with the "lsqproject" and "shrinkmode" options activated (YES) and a list of modern populations from Europe, North Africa, Caucasus, and Southwestern Asia (Supplementary Material, Table S3). All ancient individuals were projected onto PC1 and PC2.

f-statistics

For *f*-statistics with ADMIXTOOLS v5.1 (https://github.com/DReichLab/AdmixTools). Local individuals were analyzed both independently (lc) and alongside non-local individuals (all). Thus, populations were represented either by only local individuals (population_lc) or by all available individuals, including local and non-local individuals (population_all). All *f*-statistics were plotted with R v4.4.2[99] using the ggplot2 v3.5.1 package[100].

Within each Iberian site (the fortress of Vilars, Sant Miquel d'Olèrdola, and El Camp de les Lloses), the shared genetic drift between pairs of individuals was examined using the outgroup f3-statistic in the form f_3(Ind1, Ind2; Mbuti) with the qp3Pop function from ADMIXTOOLS v5.1 (https://github.com/DReichLab/AdmixTools). Results were plotted with ggplot2 v3.5.1[100] from R v4.4.2[99] in a multidimensional scaling plot of 1-f3(Ind1, Ind2; Mbuti).

Outgroup f_3-statistics, computed using qp3Pop with the 'inbreed' parameter activated, measured the shared genetic drift between two populations (A and B) relative to an outgroup (C), providing a similarity measure: higher values indicate greater genetic similarity between A and B. In this study, we calculated outgroup f_3-statistics in the form f_3(IberianPopulation, Test; Mbuti), where *IberianPopulation* refers to each of the three Iberian populations—VILA (the fortress of Vilars), OLE (Sant Miquel d'Olèrdola), and CL (El Camp de les Lloses). The *Test* populations were Iron Age and Roman populations from the Iberian Peninsula, with Mbuti, a present-day Sub-Saharan hunter-gatherer population from the Democratic Republic of Congo, serving as the outgroup.

For f_4 statistics, qpDstats was used with “f4mode” and “printsd” parameters activated to examine genetic relationships between ancient populations from the Northeast Iberian Peninsula and ancestral influences. The f_4-statistic in the form f_4(Target, Reference; Test, Chimp) evaluates whether the *Test* population shares more genetic drift with the *Reference* or *Target* population relative to an outgroup. Negative values indicate greater genetic affinity between the *Test* and *Reference* populations, while positive values suggest a closer relationship between the *Test* and *Target* populations. Chimpanzee (Chimp) was selected as the outgroup instead of Mbuti to facilitate the exploration of potential sub-Saharan genetic contributions. The *Test* population included a wide range of populations representing diverse temporal and geographic origins:

- Sub-Saharan populations such as Mbuti and Cameroon_SMA (Stone to Metal Age).
- North African populations including Iberomaurusian (Epipaleolithic individuals from Morocco), Morocco_EN (Early Neolithic), Tetuan_EN (Early Neolithic), Morocco_MN (Middle Neolithic), Morocco_LN (Late Neolithic), Punic_Tunisia_lc (local individuals from Iron Age Tunisia), and Punic_Tunisia_oAfrica (individuals with African ancestry from Iron Age Tunisia).
- Levantine and Southwestern Asian populations including Israel_Natufian (Epipaleolithic), Anatolia_N (Neolithic), Iran_N (Neolithic), Iran_BA (Bronze Age), Iran_IA (Iron Age), Lebanon_MBA (Middle Bronze Age), Lebanon_IA (Iron Age), and Israel_IA (Iron Age).
- Aegean populations such as Minoan (Bronze Age Crete) and Mycenaean (Bronze Age Greece).
- Western Mediterranean populations including Sardinia_Nuragic (Middle Bronze Age) and Punic_Sardinia (Iron Age).
- European populations such as WHG (Western Hunter-Gatherers) and Germany_BellBeaker (Bronze Age).

Two models were tested for the fortress of Vilars and Sant Miquel d’Olèrdola with different *Reference* populations: (I) Bronze Age individuals from the entire Iberian Peninsula and (II) Bronze Age individuals from the Northeast of the Iberian Peninsula (territory of present-day Catalonia). For Late-Iberians (El Camp de les Lloses), the Iberians (unrelated individuals from the fortress of Vilars and Sant Miquel d’Olèrdola) were used as the *Reference* population.

Admixture modelling

Several semiproximal qpAdm analyses using ADMIXTOOLS v5.1 (https://github.com/DReichLab/AdmixTools) were conducted with the "allsnps" parameter set to "YES" for the three Iberian sites (namely, the fortress of Vilars, Sant Miquel d'Olèrdola, and El Camp de les Lloses) and different Mediterranean populations spanning from the Bronze Age to the Roman Period. For Northeastern individuals from the Iberian Peninsula dated between the Bronze Age and the Roman period, the same qpAdm analyses were performed. For all the models tested, one of the sources consistently included Chalcolithic (C) Iberian Peninsula, Italian Peninsula, or Sardinia, which was then combined with the other two, Pontica-Caspian Steppe (BA), Germany (Bell Beaker, BA), Iran (N), Morocco (EN), Tetuan (EN), Morocco (MN), and Morocco (LN), in various two- and three-source models. The populations used as outgroups (or right populations) in these analyses included:

- The Mbuti population.
- An ancient North Italian Hunter-Gatherer individual (Villabruna).
- Ancient Western Hunter-Gatherers (WHG).
- The Onge population, present-day Indigenous hunter-gatherers from the Andaman Islands.
- The Mixe population, present-day Indigenous people from the highlands of Oaxaca, Mexico.
- The Ami population, a present-day Indigenous Austronesian group native to Taiwan.
- The Itelmen population, a present-day Indigenous group from the Kamchatka Peninsula in Russia.
- The Iberomaurusian population.

Models with a p-value below 0.05 or two and three-way models with an ancestry proportion (C) below 0.1 were eliminated. Models with negative values after subtracting two standard deviations from the ancestry proportion (C - 2sd < 0) were also discarded.

For a specific individual or population, top qpAdm models and those with a p-value with a difference lower than 0.1 from the top model were selected, and the mean of each ancestry was estimated. Data generated was summarized with a PCA calculated in R v4.4.2[99]. Principal Components were estimated with the prcomp() function and plotted with ggplot2 v3.5.1[100]. Two main clusters were found with the kmeans() function.

Bibliography

1. Roca-Rada, X., Davidson, R., Williams, M. P., Ravishankar, S., Collen, E., Haarkötter, C., Taufik, L., Carvalho, A. F., Villalba-Mouco, V., Cuesta-Aguirre, D. R., *et al.* (2024). The genetic history of Portugal over the past 5,000 years. *bioRxiv* https://doi.org/10.1101/2024.09.12.612544.

2. Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., Dulias, K., Edwards, C. J., Gandini, F., Pala, M., *et al.* (2019). The genomic history of the Iberian Peninsula over the past 8000 years. *Science (80-.).* 363, 1230–1234 https://doi.org/10.1126/science.aav4040.

3. Posth, C., Yu, H., Ghalichi, A., Rougier, H., Crevecoeur, I., Huang, Y., Ringbauer, H., Rohrlach, A. B., Nägele, K., Villalba-Mouco, V., *et al.* (2023). Palaeogenomics of Upper Palaeolithic to Neolithic European hunter-gatherers. *Nature* 615, 117–126 https://doi.org/10.1038/s41586-023-05726-0.

4. Villalba-Mouco, V., van de Loosdrecht, M. S., Rohrlach, A. B., Fewlass, H., Talamo, S., Yu, H., Aron, F., Lalueza-Fox, C., Cabello, L., Cantalejo Duarte, P., *et al.* (2023). A 23,000-year-old southern Iberian individual links human groups that lived in Western Europe before and after the Last Glacial Maximum. *Nat. Ecol. Evol.* https://doi.org/10.1038/s41559-023-01987-0.

5. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., *et al.* (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207–211 https://doi.org/10.1038/nature14317.

6. Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., *et al.* (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* 551, 368–372 https://doi.org/10.1038/nature24476.

7. Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., *et al.* (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* 5, 5257 https://doi.org/10.1038/ncomms6257.

8. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., *et al.* (2016). The genetic history of Ice Age Europe. *Nature* 534, 200–205 https://doi.org/10.1038/nature17993.

9. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., *et al.* (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528, 499–503 https://doi.org/10.1038/nature16152.

10. Martiniano, R., Cassidy, L. M., Ó'Maoldúin, R., McLaughlin, R., Silva, N. M., Manco, L., Fidalgo, D., Pereira, T., Coelho, M. J., Serra, M., *et al.* (2017). The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLOS Genet.* 13, e1006852 https://doi.org/10.1371/journal.pgen.1006852.

11. Valdiosera, C., Günther, T., Vera-Rodríguez, J. C., Ureña, I., Iriarte, E., Rodríguez-Varela, R., Simões, L. G., Martínez-Sánchez, R. M., Svensson, E. M., Malmström, H., *et al.* (2018). Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. *Proc. Natl. Acad. Sci.* 115, 3428–3433 https://doi.org/10.1073/pnas.1717762115.

12. Szécsényi-Nagy, A., Roth, C., Brandt, G., Rihuete-Herrada, C., Tejedor-Rodríguez, C., Held, P., García-Martínez-de-Lagrán, Í., Arcusa Magallón, H., Zesch, S., Knipper, C., *et al.* (2017). The maternal genetic make-up of the Iberian Peninsula between the Neolithic and the Early Bronze Age. *Sci. Rep.* 7, 15644 https://doi.org/10.1038/s41598-017-15480-9.

13. Villalba-Mouco, V., Oliart, C., Rihuete-Herrada, C., Childebayeva, A., Rohrlach, A. B., Fregeiro, M. I., Celdrán Beltrán, E., Velasco-Felipe, C., Aron, F., Himmel, M., *et al.* (2021). Genomic transformation and social organization during the Copper Age–Bronze Age transition in southern Iberia. *Sci. Adv.* 7, 21 https://doi.org/10.1126/sciadv.abi7038.

14. Olalde, I., Brace, S., Allentoft, M. E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., *et al.* (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* 555, 190–196 https://doi.org/10.1038/nature25738.

15. Arroyo, E., Baeza-Richer, C., Fernandez Dominguez, E. & López-Parra, A. (2007). Genetic history of the Iberian Peninsula. in *Recent Advances in Molecular Biology and Evolution: Applications to Biological Anthropology* 389–411.

16. Curchin, L. A. (Routledge, 1991). *Roman Spain*. https://doi.org/10.4324/9781315815671.

17. Ruiz, A. & Molinos, M. (Cambridge University Press, 1998). *The Archaeology of the Iberians*.

18. Jordá Cerdá, F., Pellicer Catalán, M., Acosta Martínez, P. & Almagro-Gorbea, M. (1986). *Historia de España 1 Prehistoria*.

19. Almagro, M. (2014). *Protohistoria de La Península Ibérica: Del Neolítico a La Romanización*.

20. Oliver Foix, A. (2004). Sacrificios y 'mala muerte' en el Registro Arqueológico de los Yacimientos Ibéricos. *Espac. Tiempo y Forma. Ser. I, Prehist. y Arqueol.* 1, 391 https://doi.org/10.5944/etfi.16-17.2003.4763.

21. Subirà, E. & Molist, N. (2008). Inhumacions perinatals múltiples i espais de treball en els

assentaments ibers. in *Nasciturus: infans, puerulus. Vobis mater terra. La muerte en la infancia* (eds. Gusi, F., Muriel, S. & Olaria, C. M.) 365–385.

22. Crespo, L., Subirà, M. E. & Ruiz, J. (2011). Twins in prehistory: The case from Olèrdola (Barcelona, Spain; s. IV II BC). *Int. J. Osteoarchaeol.* 21, 751–756 https://doi.org/10.1002/oa.1169.

23. Afonso, C., Nociarova, D., Santos, C., Martinez-Labarga, C., Mestres, I., Duran, M. & Malgosa, A. (2019). Sex selection in late Iberian infant burials: Integrating evidence from morphological and genetic data. *Am. J. Hum. Biol.* 31, 1–10 https://doi.org/10.1002/ajhb.23204.

24. Conklin, B. A. & Morgan, L. M. (1996). Babies, bodies, and the production of personhood in North America and a native Amazonian society. *Ethos* 24, 657–694.

25. Halcrow, S. E., Tayles, N. & Livingstone, V. (2008). Infant death in late prehistoric Southeast Asia. *Asian Perspect.* 47, 371–404.

26. Borić, D. & Stefanović, S. (2004). Birth and death: infant burials from Vlasac and Lepenski Vir. *Antiquity* 78, 526–546 https://doi.org/DOI: 10.1017/S0003598X00113201.

27. Carroll, M. (2011). Infant death and burial in Roman Italy. *J. Rom. Archaeol.* 24, 99–120 https://doi.org/DOI: 10.1017/S1047759400003329.

28. Beilke-Voigt, I. (Cambridge Scholars Newcastle, 2008). Burials of children in houses and settlements during the Roman Iron Age and Early Medieval period in northern Germany and Denmark. in *Children, identity and the past* (eds. Dommasnes, L. H. & Wrigglesworth, M.) 83–95.

29. Laes, C., Guimier-Sorbets, A.-M., Morizot, Y., Nenna, M.-D., Hermary, A. & Dubois, C. (2015). Children, Burial and Death. *Antiq. Class.* 84, 259–269.

30. McGeorge, P. J. P. (Istanbul, Turkey, 2011). Intramural infant burials in the Aegean Bronze age. in *2èmes Rencontres d'archéologie de l'IFEA : Le Mort dans la ville Pratiques, contextes et impacts des inhumations intra-muros en Anatolie, du début de l'Age du Bronze à l'époque romaine*. 1–20.

31. Moore, A. (2009). Hearth and Home: The Burial of Infants within Romano-British Domestic Contexts. *Child. Past* 2, 33–54 https://doi.org/10.1179/cip.2009.2.1.33.

32. Eng, R. Y. & Smith, T. C. (1976). Peasant Families and Population Control in Eighteenth-Century Japan. *J. Interdiscip. Hist.* 6, 417 https://doi.org/10.2307/202664.

33. Gowland, R. L., Chamberlain, A. & Redfern, R. C. (2014). 'On the brink of being : re-evaluating infanticide and infant burial in Roman Britain. *J. Rom. Archaeol. Suppl. Ser.* 96, 69–88.

34. Abu-Mandil Hassan, N., Brown, K. A., Eyers, J., Brown, T. A. & Mays, S. (2014). Ancient DNA study of the remains of putative infanticide victims from the Yewden Roman villa site at Hambleden, England. *J. Archaeol. Sci.* 43, 192–197 https://doi.org/10.1016/j.jas.2013.12.017.

35. Martirosyan, A., Sandoval-Ávila, C., Irurita, J., Juanhuix, J., Molist, N., Mestres, I., Durán, M., Alonso, N., Santos, C., Malgosa, A., *et al.* (2024). Reconstructing infant mortality in Iberian Iron Age populations from tooth histology. *J. Archaeol. Sci.* 171, 106088 https://doi.org/10.1016/j.jas.2024.106088.

36. Gamba, C., Fernández, E., Oliver, A., Tirado, M., Baeza, C., López-Parra, A. M. & Arroyo-Pardo, E. (2008). Population genetics and DNA preservation in ancient human remains from Eastern Spain. *Forensic Sci. Int. Genet. Suppl. Ser.* 1, 462–464 https://doi.org/10.1016/j.fsigss.2007.10.037.

37. Sampietro, M. L., Caramelli, D., Lao, O., Calafell, F., Comas, D., Lari, M., Agustí, B., Bertranpetit, J. & Lalueza-Fox, C. (2005). The Genetics of the Pre-Roman Iberian Peninsula: A mtDNA Study of Ancient Iberians. *Ann. Hum. Genet.* 69, 535–548 https://doi.org/10.1111/j.1529-8817.2005.00194.x.

38. López-Cachero, F. J. (2011). Cremation Cemeteries in the Northeastern Iberian Peninsula: Funeral Diversity and Social Transformation during the Late Bronze and Early Iron Ages. *Eur. J. Archaeol.* 14, 116–132 https://doi.org/10.1179/146195711798369382.

39. Junyent Sánchez, E. & Moya i Garra, A. (2011). Els fossats de la fortalesa dels Vilars d'Arbeca (Catalunya, Espanya). *Rev. d'Arqueologia Ponent* 21, 93–120.

40. Duran, M., Mestres, I., Principal, J. & Padrós, C. (2015). El Camp de les Lloses (Tona , Osona): un post avançat amb tallers metal·lúrgics al servei de la logística de l'exèrcit romanorepublicà (125-75 ane). *Rev. d'Arqueologia Ponent* 25, 293–307.

41. Hernández, C. (2023). Mitochondrial DNA in Human Diversity and Health: From the Golden Age to the Omics Era. *Genes (Basel).* 14, 1534 https://doi.org/10.3390/genes14081534.

42. Barral-Arca, R., Pischedda, S., Gómez-Carballa, A., Pastoriza, A., Mosquera-Miguel, A., López-Soto, M., Martinón-Torres, F., Álvarez-Iglesias, V. & Salas, A. (2016). Meta-Analysis of Mitochondrial DNA Variation in the Iberian Peninsula. *PLoS One* 11, e0159735 https://doi.org/10.1371/journal.pone.0159735.

43. Bramanti, B., Thomas, M. G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M. N., Jankauskas, R., Kind, C.-J., *et al.* (2009). Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers. *Science (80-.).* 326, 137–140 https://doi.org/10.1126/science.1176869.

44. Kristjansson, D., Schurr, T. G., Bohlin, J. & Jugessur, A. (2023). Phylogeographic history of

mitochondrial haplogroup J in Scandinavia. *Am. J. Biol. Anthropol.* 180, 298–315 https://doi.org/10.1002/ajpa.24666.

45. González, A. M., Larruga, J. M., Abu-Amero, K. K., Shi, Y., Pestano, J. & Cabrera, V. M. (2007). Mitochondrial lineage M1 traces an early human backflow to Africa. *BMC Genomics* 8, 223 https://doi.org/10.1186/1471-2164-8-223.

46. Winters, C. (2010). The African Origin of mtDNA Haplogroup M1. *Curr. Res. J. Biol. Sci.* 2, 380–389.

47. Papac, L., de Miguel Ibáñez, P., Rohrlach, A. B., Armendáriz, J., Peres, M., Lamnidis, T. C., Mötsch, A., Schiffels, S. & Risch, R. (2023). Intramural child burials in Iron Age Navarra: How ancient DNA can contribute to household archaeology . in *Kinship, Sex, and Biological Relatedness : The contribution of archaeogenetics to the understanding of social and biological relations. 15. Mitteldeutscher Archäologentag vom 6. bis 8. Oktober 2022 in Halle (Saale)* (eds. Meller, H., Krause, J., Haak, W. & Risch, R.) 263–295 https://doi.org/10.11588/propylaeum.1280.c18012.

48. Cassidy, L. M., Russell, M., Smith, M., Delbarre, G., Cheetham, P., Manley, H., Mattiangeli, V., Breslin, E. M., Jackson, I., McCann, M., *et al.* (2025). Continental influx and pervasive matrilocality in Iron Age Britain. *Nature* https://doi.org/10.1038/s41586-024-08409-6.

49. Rusu, I., Modi, A., Radu, C., Mircea, C., Vulpoi, A., Dobrinescu, C., Bodolică, V., Potârniche, T., Popescu, O., Caramelli, D., *et al.* (2019). Mitochondrial ancestry of medieval individuals carelessly interred in a multiple burial from southeastern Romania. *Sci. Rep.* 9, 961 https://doi.org/10.1038/s41598-018-37760-8.

50. Dryomov, S. V., Nazhmidenova, A. M., Starikovskaya, E. B., Shalaurova, S. A., Rohland, N., Mallick, S., Bernardos, R., Derevianko, A. P., Reich, D. & Sukernik, R. I. (2021). Mitochondrial genome diversity on the Central Siberian Plateau with particular reference to the prehistory of northernmost Eurasia. *PLoS One* 16, e0244228 https://doi.org/10.1371/journal.pone.0244228.

51. van Oven, M. & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* 30, E386–E394 https://doi.org/10.1002/humu.20921.

52. Papac, L., Ernée, M., Dobeš, M., Langová, M., Rohrlach, A. B., Aron, F., Neumann, G. U., Spyrou, M. A., Rohland, N., Velemínský, P., *et al.* (2021). Dynamic changes in genomic and social structures in third millennium BCE central Europe. *Sci. Adv.* 7, https://doi.org/10.1126/sciadv.abi6941.

53. Mallick, S., Micco, A., Mah, M., Ringbauer, H., Lazaridis, I., Olalde, I., Patterson, N. & Reich, D.

(2024). The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Sci. Data* 11, 182 https://doi.org/10.1038/s41597-024-03031-7.

54. Junyent, E. (1996). Els origens de la cultura ilergeta. in *Indíbil i mandoni : reis i guerrers* 13–16.

55. Junyent, E. (1996). La societat ilergata: un món de pagesos i aristòcrates. in *Indíbil i mandoni : reis i guerrers* 17–25.

56. Junyent, E. (2023). The times of Indivilis and Mandonius. in *Romans in the West* 21–33.

57. Natàlia Alonso M. Gené, E. Junyent, A. Lafuente, J. B. López, A. Moya, E. Tartera, A. Vidal, A. C. https://calaix.gencat.cat/bitstream/handle/10687/454193/qmem4437.pdf?sequence=1&isAllowed=y (2004). *Els Vilars 2001. Memòria de La XVII[a] Campanya d'excavacions Arqueològiques*.

58. Universitat de Lleida. at https://vilars.udl.cat/export/sites/vilars/ca/.galleries/Documents/Cataleg-Vilars-Reduit2.pdf (2020). La Fortalesa del Vilars d'Arbeca. Terra, aigua i poder en el món iber. 166.

59. Brunel, S., Bennett, E. A., Cardin, L., Garraud, D., Barrand Emam, H., Beylier, A., Boulestin, B., Chenal, F., Ciesielski, E., Convertini, F., *et al.* (2020). Ancient genomes from present-day France unveil 7,000 years of its demographic history. *Proc. Natl. Acad. Sci.* 117, 12791–12798 https://doi.org/10.1073/pnas.1918034117.

60. Molist Capella, N. & Bosch Casadevall, J. M. http://hdl.handle.net/10687/23910 (2002). *Memòria de La Intervenció Arqueològica Al Conjunt Històric d'Olèrdola (2000-2001)*.

61. Antonio, M. L., Weiß, C. L., Gao, Z., Sawyer, S., Oberreiter, V., Moots, H. M., Spence, J. P., Cheronet, O., Zagorc, B., Praxmarer, E., *et al.* (2024). Stable population structure in Europe since the Iron Age, despite high mobility. *Elife* 13, https://doi.org/10.7554/eLife.79714.

62. Duran i Caixal, M., Mestres i Santacreu, I., Padrós, C. & Principal, J. (2017). El Camp de les Lloses, un exemple d'implantació militar al territori d'època romanorepublicana. *Trib. d'Arqueologia* Nº 2014-20, 11–45.

63. Costa, M. D., Pereira, J. B., Pala, M., Fernandes, V., Olivieri, A., Achilli, A., Perego, U. A., Rychkov, S., Naumova, O., Hatina, J., *et al.* (2013). A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. *Nat. Commun.* 4, 2543 https://doi.org/10.1038/ncomms3543.

64. Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E., Carelli, V., *et al.* (2004). The Molecular Dissection of mtDNA Haplogroup H Confirms That the Franco-Cantabrian Glacial Refuge Was a Major Source for the European Gene Pool. *Am. J. Hum. Genet.* 75, 910–918 https://doi.org/10.1086/425590.

65. Kapp, J. D., Green, R. E. & Shapiro, B. (2021). A Fast and Efficient Single-stranded Genomic Library Preparation Method Optimized for Ancient DNA. *J. Hered.* 112, 241–249 https://doi.org/10.1093/jhered/esab012.

66. Vinueza-Espinosa, D. C., Santos, C., Martínez-Labarga, C. & Malgosa, A. (2020). Human DNA extraction from highly degraded skeletal remains: How to find a suitable method? *Electrophoresis* 41, 2149–2158 https://doi.org/10.1002/elps.202000171.

67. Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., Wales, N., Sicheritz-Pontén, T. & Gilbert, M. T. P. (2018). Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* 9, 410–419 https://doi.org/10.1111/2041-210X.12871.

68. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. B Biol. Sci.* 370, 20130624 https://doi.org/10.1098/rstb.2013.0624.

69. Vinueza-Espinosa, D. C., Cuesta-Aguirre, D. R., Malgosa, A. & Santos, C. (2023). Mitochondrial DNA control region typing from highly degraded skeletal remains by single-multiplex next-generation sequencing. *Electrophoresis* 44, 1423–1434 https://doi.org/10.1002/elps.202200052.

70. Fellows Yates, J. A., Lamnidis, T. C., Borry, M., Andrades Valtueña, A., Fagernäs, Z., Clayton, S., Garcia, M. U., Neukamm, J. & Peltzer, A. (2021). Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. *PeerJ* 9, e10947 https://doi.org/10.7717/peerj.10947.

71. Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P. & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* 38, 276–278 https://doi.org/10.1038/s41587-020-0439-x.

72. Schubert, M., Lindgreen, S. & Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9, 88 https://doi.org/10.1186/s13104-016-1900-2.

73. Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760 https://doi.org/10.1093/bioinformatics/btp324.

74. Oliva, A., Tobler, R., Cooper, A., Llamas, B. & Souilmi, Y. (2021). Systematic benchmark of ancient DNA read mapping. *Brief. Bioinform.* 22, 1–12 https://doi.org/10.1093/bib/bbab076.

75. Oliva, A., Tobler, R., Llamas, B. & Souilmi, Y. (2021). Additional evaluations show that specific BWA-aln settings still outperform BWA-mem for ancient DNA data alignment. *Ecol. Evol.* 11, 18743–18748 https://doi.org/10.1002/ece3.8297.

76. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., *et al.* (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, 1–4 https://doi.org/10.1093/gigascience/giab008.

77. Picard Toolkit. Broad Institute. 2019. Retrieved from https://broadinstitute.github.io/picard/.

78. Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J. & Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci.* 111, 2229–2234 https://doi.org/10.1073/pnas.1318934111.

79. Jun, G., Wing, M. K., Abecasis, G. R. & Kang, H. M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* 25, 918–925 https://doi.org/10.1101/gr.176552.114.

80. Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 23, 147–147 https://doi.org/10.1038/13779.

81. Okonechnikov, K., Conesa, A. & García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32, 292–294 https://doi.org/10.1093/bioinformatics/btv566.

82. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15, 356 https://doi.org/10.1186/s12859-014-0356-4.

83. Lamnidis, T. C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., Khartanovich, V., Balanovsky, O., Ongyerth, M., Weihmann, A., *et al.* (2018). Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nat. Commun.* 9, 5018 https://doi.org/10.1038/s41467-018-07483-5.

84. Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J. & Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. *Genome Biol.* 17, 60 https://doi.org/10.1186/s13059-016-0918-z.

85. Garrison, E. & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing.

86. Garrison, E., Kronenberg, Z. N., Dawson, E. T., Pedersen, B. S. & Prins, P. (2022). A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. *PLOS Comput. Biol.* 18, e1009123 https://doi.org/10.1371/journal.pcbi.1009123.

87. Schönherr, S., Weissensteiner, H., Kronenberg, F. & Forer, L. (2023). Haplogrep 3 - an interactive haplogroup classification and analysis platform. *Nucleic Acids Res.* 51, W263–

W268 https://doi.org/10.1093/nar/gkad284.

88. Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. & Mesirov, J. P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26 https://doi.org/10.1038/nbt.1754.

89. Cuesta-Aguirre, D. R., Malgosa, A. & Santos, C. (2024). An easy-to-use pipeline to analyze amplicon-based Next Generation Sequencing results of human mitochondrial DNA from degraded samples. *PLoS One* 19, e0311115 https://doi.org/10.1371/journal.pone.0311115.

90. Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V. & Richards, M. B. (2009). Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *Am. J. Hum. Genet.* 84, 740–759 https://doi.org/10.1016/j.ajhg.2009.05.001.

91. Li, M., Schröder, R., Ni, S., Madea, B. & Stoneking, M. (2015). Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proc. Natl. Acad. Sci.* 112, 2491–2496 https://doi.org/10.1073/pnas.1419651112.

92. Cuesta-Aguirre, D. R., Amor-Jimenez, C., Malgosa, A. & Santos, C. (2025). A Post-Mortem Molecular Damage Profile in the Ancient Human Mitochondrial DNA. *Mol. Ecol. Resour.* e14061 https://doi.org/10.1111/1755-0998.14061.

93. Jagadeesan, A., Ebenesersdóttir, S. S., Guðmundsdóttir, V. B., Thordardottir, E. L., Moore, K. H. S. & Helgason, A. (2021). HaploGrouper: a generalized approach to haplogroup classification. *Bioinformatics* 37, 570–572 https://doi.org/10.1093/bioinformatics/btaa729.

94. Rohrlach, A. B., Tuke, J., Popli, D. & Haak, W. (2023). BREADR: An R Package for the Bayesian Estimation of Genetic Relatedness from Low-coverage Genotype Data. *bioRxiv* https://doi.org/10.1101/2023.04.17.537144.

95. Simões, L. G., Günther, T., Martínez-Sánchez, R. M., Vera-Rodríguez, J. C., Iriarte, E., Rodríguez-Varela, R., Bokbot, Y., Valdiosera, C. & Jakobsson, M. (2023). Northwest African Neolithic initiated by migrants from Iberia and Levant. *Nature* 618, 550–556 https://doi.org/10.1038/s41586-023-06166-6.

96. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D. C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., *et al.* (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature* 536, 419–424 https://doi.org/10.1038/nature19310.

97. Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909 https://doi.org/10.1038/ng1847.

98. Patterson, N., Price, A. L. & Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genet*. 2, e190 https://doi.org/10.1371/journal.pgen.0020190.

99. R Core Team. at https://www.r-project.org/ (2024). R: A Language and Environment for Statistical Computing.

100. Wickham, H. (Springer-Verlag New York, 2016). *Ggplot2: Elegant Graphics for Data Analysis*.

Resource availability

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Cristina Santos (cristina.santos@uab.cat)

Materials availability

This study did not generate new unique reagents.

Data and code availability

The datasets generated and analyzed during the current study will become publicly available upon publication.

Acknowledgments

This research was supported by the Ministerio de Ciencia e Innovaciòn (Ref. PGC2018-096666-B-I00), by the Fundación Palarq (PALARQ23_55), and by the Departament de Recerca i Universitats de la Generalitat de Catalunya and the European Social Fund (Fons Social Europeu). D.R.C.-A. is a PhD fellow from the Agència de Gestiò d'Ajuts Universitaris i de Recerca (AGAUR) (2020 FI_B1 00641). GREAB is supported by the Generalitat de Catalunya (2021 SGR 00186).

Author contributions

Conceptualization: D.R.C-A., A.M., C.S.

Data analysis: D.R.C-A., X.R-R., C.S.

Funding acquisition: A.M., C.S.

Interpretation: All authors

Lab work: D.R.C-A., B.DB., D.C.V-E., C.S-A., A.K.K., L.T.L.

Project administration: A.M., C.S.

Sample provision: N.M.C., M.D.C., I.M.S., E.J., N.A., J.M.M.

Software: D.R.C-A., X.R-R., S.R.

Supervision: A.M., C.S., H.S., B.L.

Visualization: D.R.C-A., C.S.

Writing - original draft: D.R.C-A., C.S.

Writing - review & editing: All authors

Declaration of interests

The authors declare no competing interests.

Declaration of generative AI and AI-assisted technologies

During the preparation of this work, the author used ChatGPT and Grammarly to improve the structure, clarity, and order of some paragraphs. Moreover, ChatGPT was also employed to improve the scripts that were used on R to generate the plots. After using these tools, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.



Figure 1. Main cultures and population movements during the Iron Age (750-250 BCE) around the Mediterranean Sea**.** Modified from the "Map of Early Iron Age migration ca. 750-250 BC" (https://indo-european.eu/maps/iron-age/).

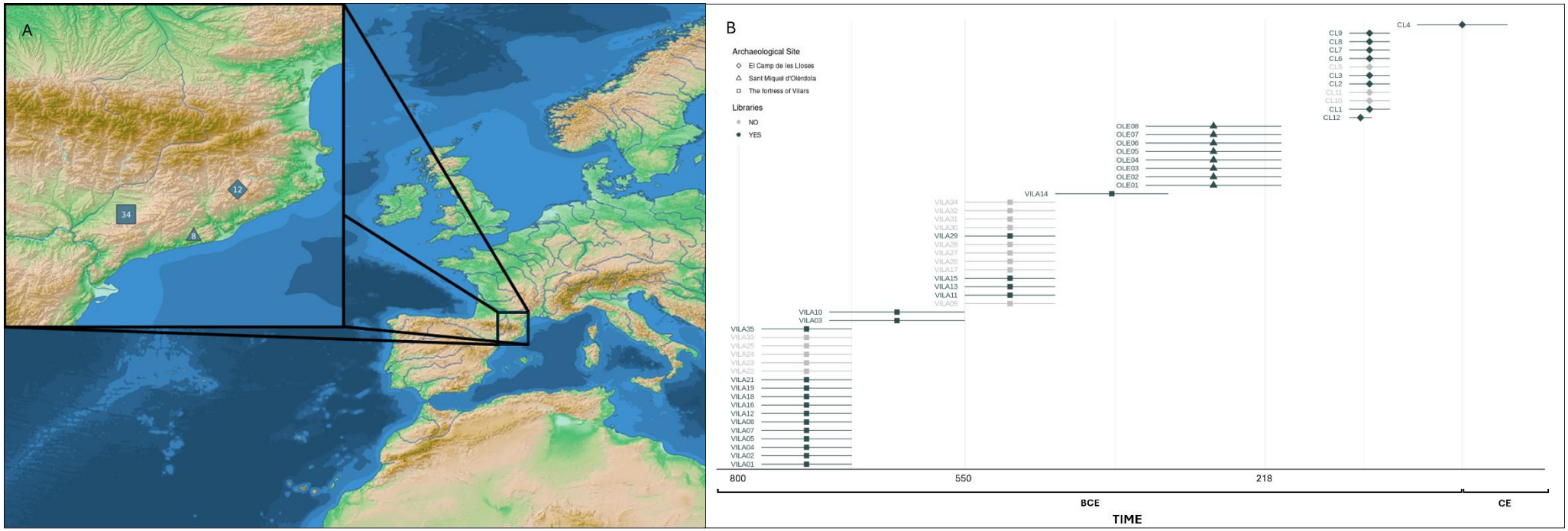


Figure 2. A) Location and number of the newly analyzed individuals, and B) the chronology of each one, indicating if libraries could be done with the extracted DNA.

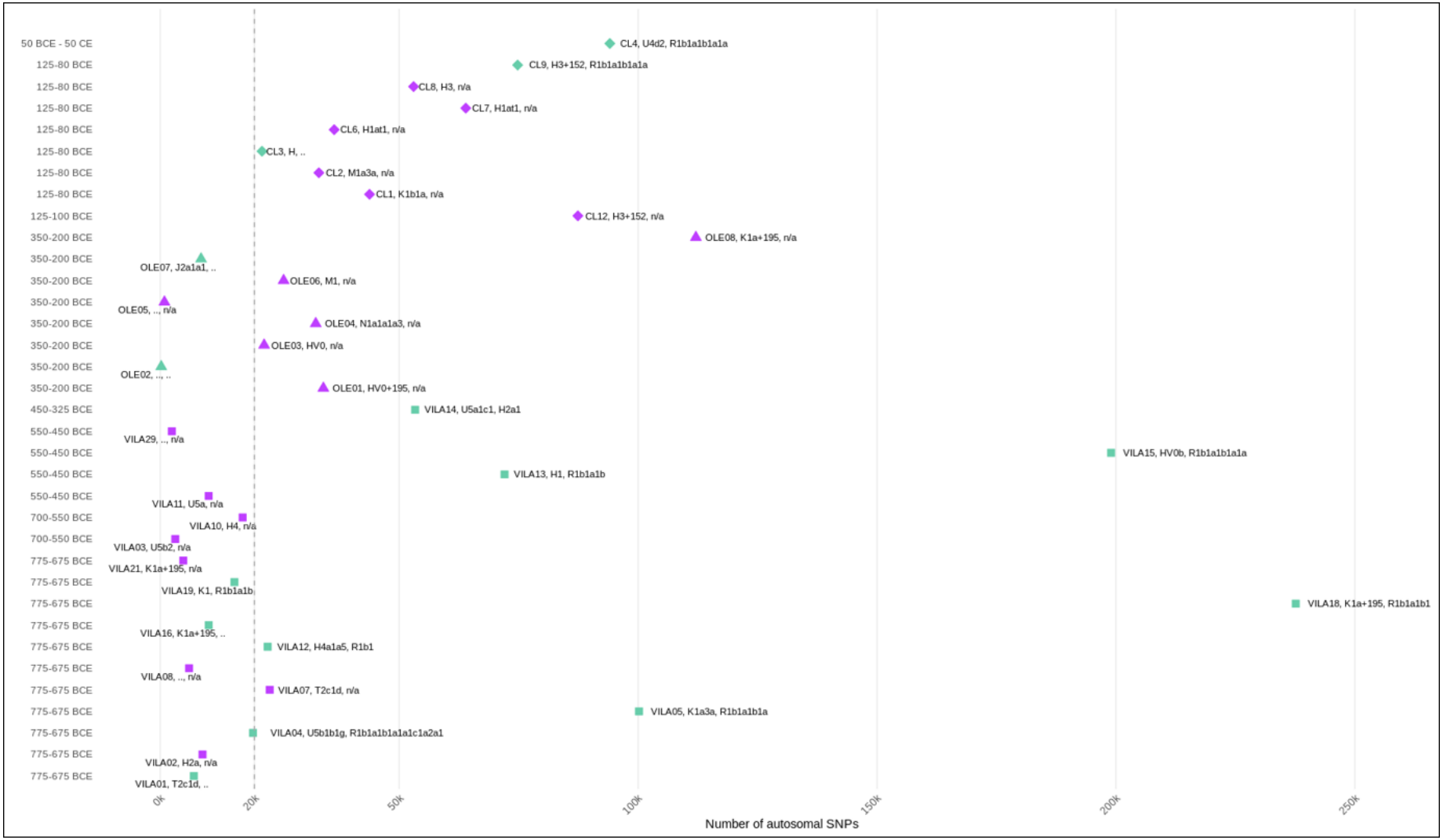


Figure 2C. Summary of sequencing results for each individual. The X-axis represents the number of autosomal SNPs from the 1240k panel per individual, with the 20,000 SNP threshold marked. The Y-axis shows the chronological age of each individual. Point colors indicate biological sex estimated by Sex.DetERRmine.py v1.1.256, while point shapes represent the archaeological site. Each point is labeled with the individual's identifier, followed by the mitochondrial haplogroup and Y-chromosome haplogroup. If a haplogroup was not determined, it is represented by two dots ("..."). For females, "n/a" denotes the absence of a Y-chromosome.

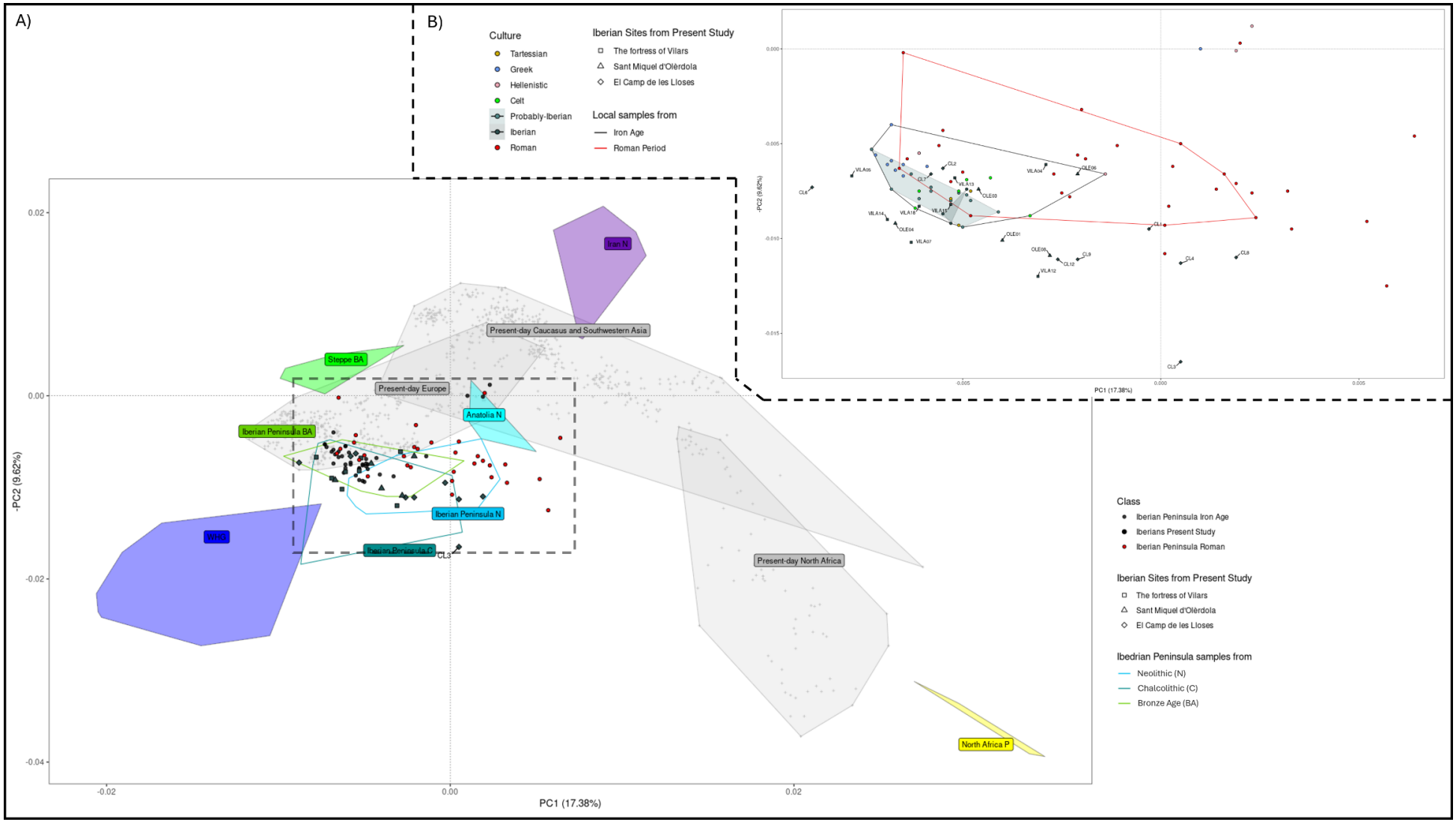


Figure 3. A) Principal Component Analysis (PCA) of present-day European, Caucasus, Southwestern Asian, and North African populations, with ancient individuals projected onto PC1 and PC2. Prehistoric individuals from the Iberian Peninsula, spanning the Neolithic to the Iron Age, are outlined with empty polygons. Iberians analyzed in this study, along with previously published Iron Age and Roman individuals from the Iberian Peninsula, are represented as colored dots (see legend for details). Dot shapes for the individuals from this study indicate their respective archaeological sites. Iberians located outside the prehistoric Iberian cluster are labeled. B) Close-up view of the region enclosed on the dashed square from Figure 5A, focusing on the Iron Age and Roman periods of the Iberian Peninsula. Filled polygons represent previously published Iberian and probably-Iberian individuals, while empty polygons highlight local Iron Age (black) and Roman (red) individuals. Dot shapes for the Iberians analyzed in this study indicate their archaeological sites, and all are labeled with their identifier.

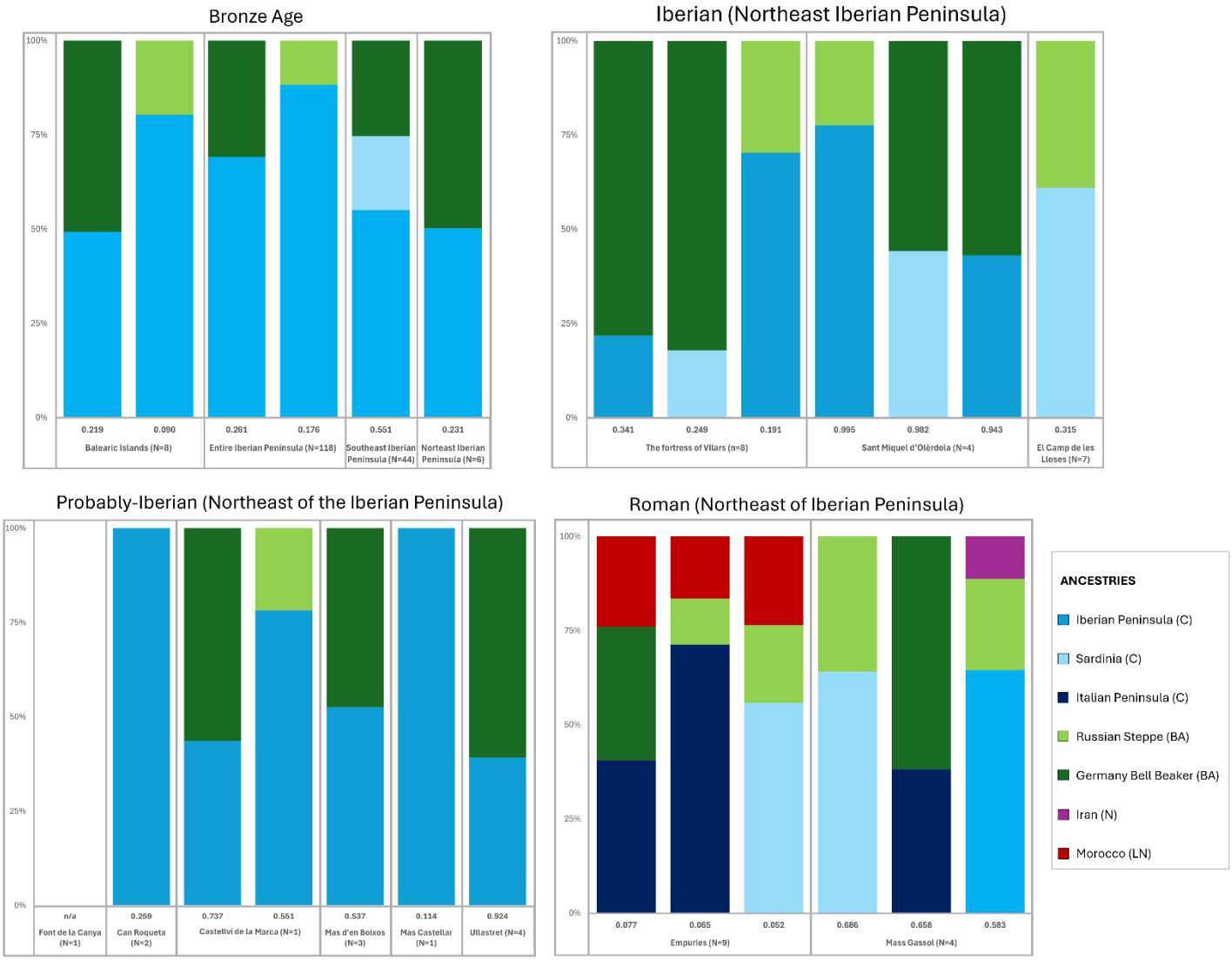


Figure 4. Top qpAdm models for different Bronze Age populations from the Iberian Peninsula, as well as Iberian, Probably-Iberian, and Roman populations from the Northeast of the Iberian Peninsula. The p-value associated with each model is indicated under the bars, which are colored to represent the different ancestries. The number of unrelated individuals in each population is next to the population's name. If no model provided a suitable fit, the corresponding space is left blank, with 'n/a' indicating the absence of a valid p-value.



Figure 5. Top qpAdm models for individuals from the Northeast of the Iberian Peninsula spanning the Bronze Age to the Roman period, as well as three Iberian individuals from present-day Valencian Community (noted in the title). The p-value associated with each model is displayed under the bars, which are colored to represent the different ancestries. Mitochondrial DNA haplogroups are indicated for Iberian individuals from the Northeast of the Iberian Peninsula. If no model provided a suitable fit, the corresponding space is left blank, with 'n/a' indicating the absence of a valid p-value.

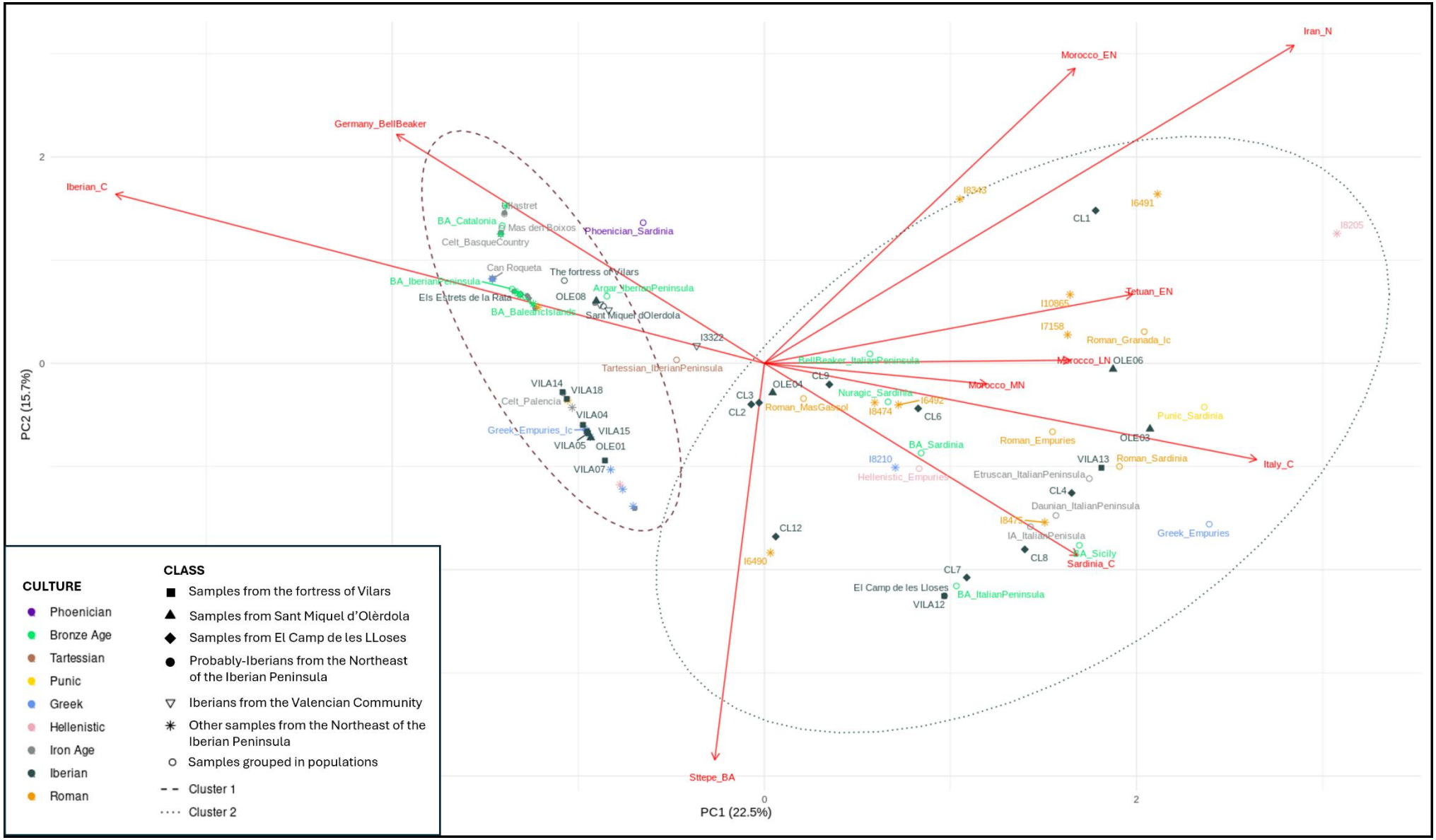


Figure 6. Close-up of the PCA with the top qpAdm models showing the mean ancestry proportions for each individual and population analyzed. Colors indicate culture, and shapes represent the class. All populations are labeled, as well as all new Iberians from the present study. Individuals from the Iberian Peninsula are labeled only if they fall outside Cluster 1. Results are available in Supplementary Material (Tables S9-S11). Outliers can be seen in Figure S16.

Supplementary Material

INDEX

ARCHAEOLOGICAL SITES 128

THE FORTRESS OF VILARS (775-325 BCE) 128
Figure S1. *Ground plan of the fortress of Vilars phase 0 (775-675 BCE)* *129*
Figure S2. *Ground plan of the fortress of Vilars phase I (700-550 BCE)* *130*
Figure S3. *Ground plan of the fortress of Vilars phase II (550-450 BCE)* *131*
Figure S4. *Ground plan of the fortress of Vilars phase III (450-325 BCE)* *132*
SANT MIQUEL D'OLÈRDOLA (IBERIAN PHASE: 350-200 BCE) 133
Figure S5. *Ground plan of the Sant Miquel d'Olèrdola's Iberian phase (350-200 BCE)* *134*
Figure S6. *Ground plan of room 2 (UH 2) in Sant Miquel d'Olèrdola* *135*
Figure S7. *Ground plan of room 3 (UH 3) in Sant Miquel d'Olèrdola* *136*
Figure S8. *Ground plan of room 4 (UH 4) in Sant Miquel d'Olèrdola* *137*
EL CAMP DE LES LLOSES (125 BCE – 50 CE) 138
Figure S9. *Ground plan of El Camp de les Lloses archaeological site* *139*

SUPPLEMENTARY FIGURES FOR "RESULTS" SECTION 140

PCA 140
Figure S10. *PCA of Iron Age and Roman samples from the Iberian Peninsula and ancient Italian samples* *140*
Figure S11. *PCA of Iron Age and Roman samples from the Iberian Peninsula and ancient Sardinian samples* *141*
MULTIDIMENSIONAL SCALING FOR 1-f_3(IND1,IND2;MBUTI) 142
Figure S12. *MDS of the Iberians analyzed in the present study* *142*
Figure S13. *MDS of the Iberians from the fortress of Vilars* *143*
Figure S14. *MDS of the Iberians from Sant Miquel d'Olèrdola* *144*
Figure S15. *MDS of the Iberians from El Camp de les Lloses* *145*
OUTGROUP f_3-STATISTICS 146
Figure S16. *Outgroup f_3-statistic in the form f_3(IberianPopulation, test; Mbuti)* *146*
f_4-STATISTICS 147
Figure S17. *Results of the f_4-statistic analyses* *147*
PCA MADE WITH QPADM TOP MODELS 148
Figure S18. *PCA with the top qpAdm models showing the mean ancestry proportions for each sample and population analyzed* *148*

BIBLIOGRAPHY 149

Archaeological sites

The fortress of Vilars (775-325 BCE)

The fortress of Vilars, situated in Arbeca (Lleida, Catalonia, Spain), represents a remarkable archaeological site. It is an almost completely excavated site that allows to know the process of Iberization and the development of the Eary Iberian and Iberian stages during the 4th century. Continuity is present, without any break, in the stratigraphy, the urban and architectural evolution and in the material culture. The site's abandonment around 300 BCE remains a mystery, as there is no evidence of a violent event or destruction. This chronological sequence provides valuable insights into the evolution of the Iron Age and Iberian fortifications, as well as the technological and social developments of the communities that inhabited Els Vilars over its 400-year history.

The fortress underwent several distinct phases during its occupation, spanning from the Early Iron Age to the Iberian period. The chronological sequence began with Vilars 0 (775-700 BCE), marking the establishment of the first fortress with a towered wall, two gates, radial urbanism, and rectangular houses (Figure S1). This was followed by Vilars I (700-550 BCE), during which the walls and towers were reinforced, and a chevaux-de-frise was constructed (Figure S2). The Iberian period commenced with Vilars II (550-450 BCE), divided into three subphases. This Early Iberian phase saw significant developments, including the introduction of rotary mills, with some of the oldest models dating to this time (Figure S3). Vilars III (425-325 BCE) (Figure S4) and Vilars IV (325-300 BCE) represented the Middle Iberian period, with the latter being the final phase before the fortress's abandonment. Dating has been possible thanks to stratigraphy and ceramics found in the different phases.

Newborns have been found in almost all phases (Vilars 0-III). Most of the cases corresponded to single burials, but some of the burials found were reused and more than one newborn can be found there. Some of them were buried in living rooms, others in storage rooms and others in metallurgical workshops. Not all of them had a tomb, but they were buried directly.

The histological age of the newborns was obtained from Martirosyan et al. (2024)[1] or applying their methodology (VILA16) (Supplementary Material, Table S1). For those samples that did not provide a result with that histological methodology or did not have teeth, different morphometric techniques were employed, depending on the available bones[2–9] (Supplementary Material, Table S1).

For more information, refer to published articles[10–14] or Natalia Alonso, Emili Junyent, and Jordi Martinez Majoral.

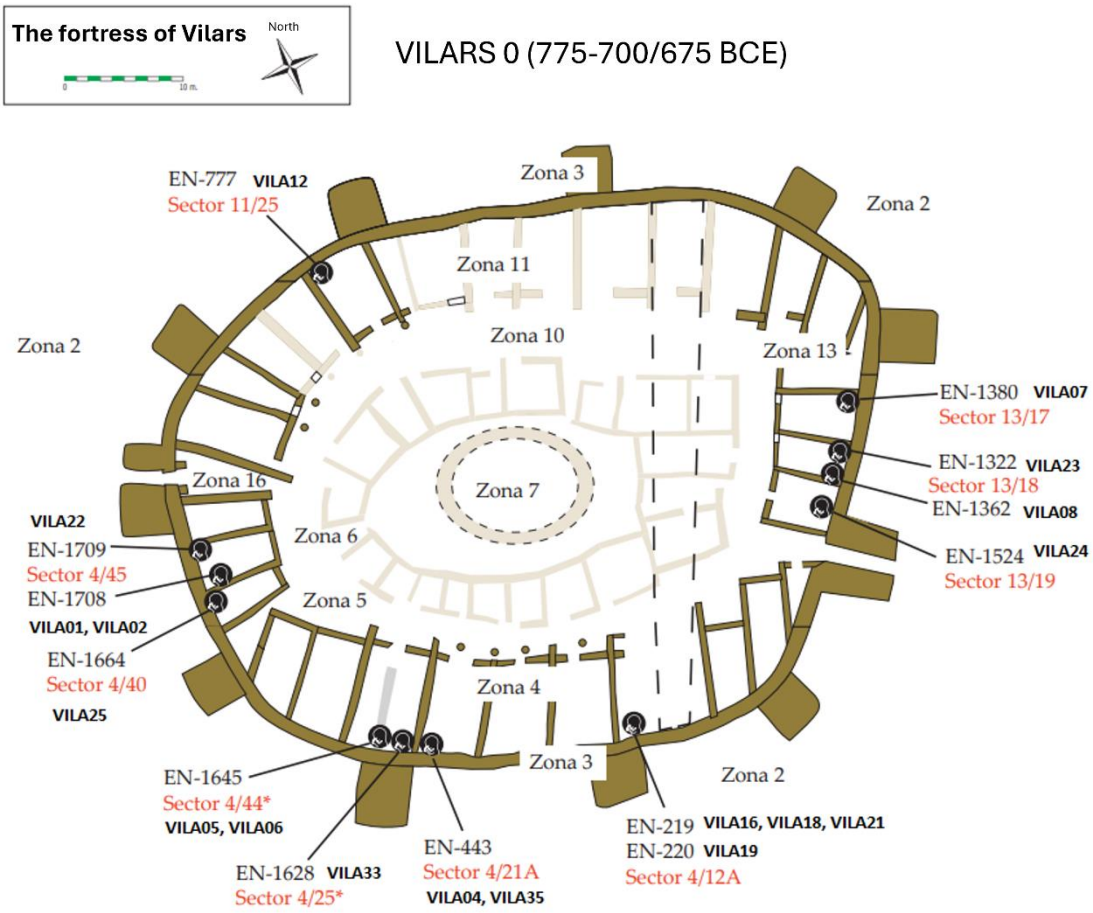


Figure S1. Ground plan of the fortress of Vilars phase 0 (775-675 BCE) located in Arbeca (Lleida, Catalonia, Spain). Site is divided into different zones ("*Zona*") and each zone into sectors, highlighted in red. Burials are indicated as "EN-", and the samples found in each one are also indicated as "VILA". When more than one newborn was found within the same burial, they are separated by commas. A total of 18 newborns were found in 13 different burials. Figure modified from Nieto Espinet et al. (2013)[14].

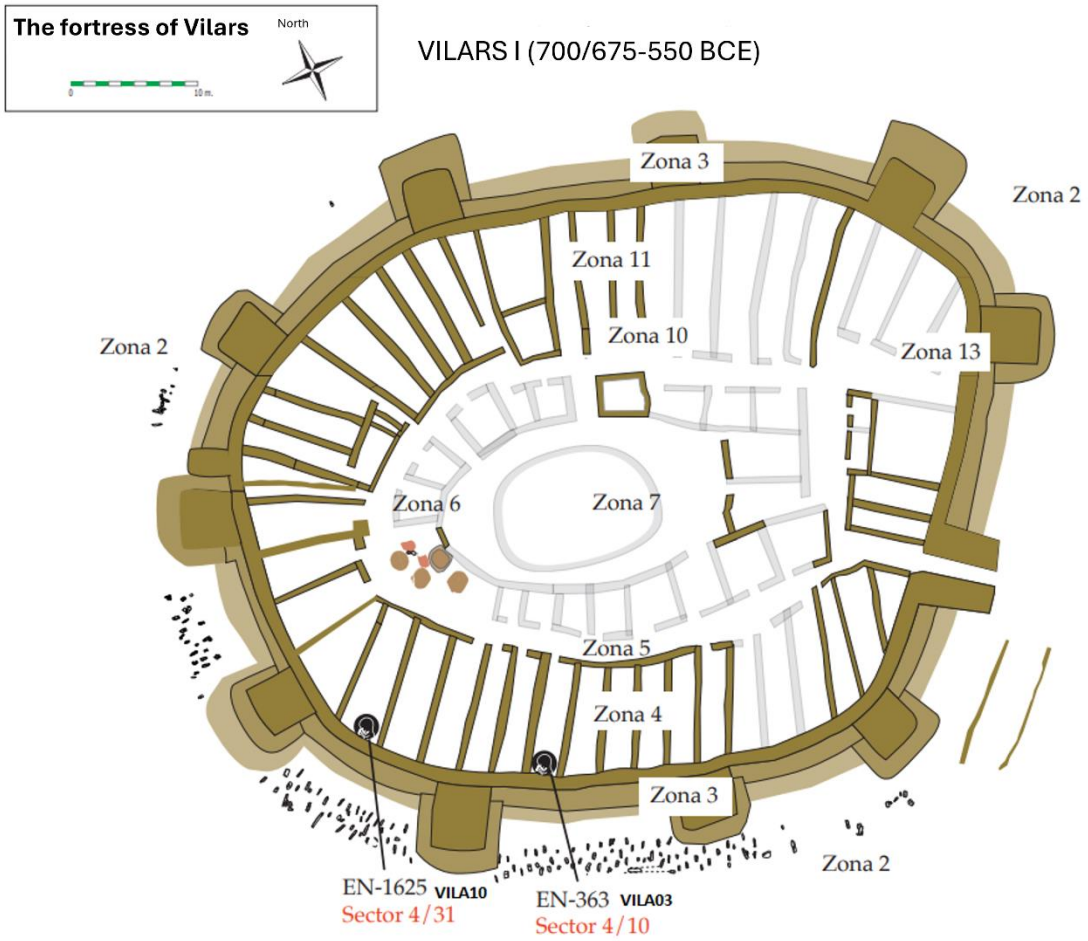


Figure S2. Ground plan of the fortress of Vilars phase I (700-550 BCE) located in Arbeca (Lleida, Catalonia, Spain). Site is divided into different zones ("*Zona*") and each zone into sectors, highlighted in red. Burials are indicated as "EN-", and the samples found in each one are also indicated as "VILA". A total of 2 newborns were found in 2 different burials. Figure modified from Nieto Espinet et al. (2013)[14].

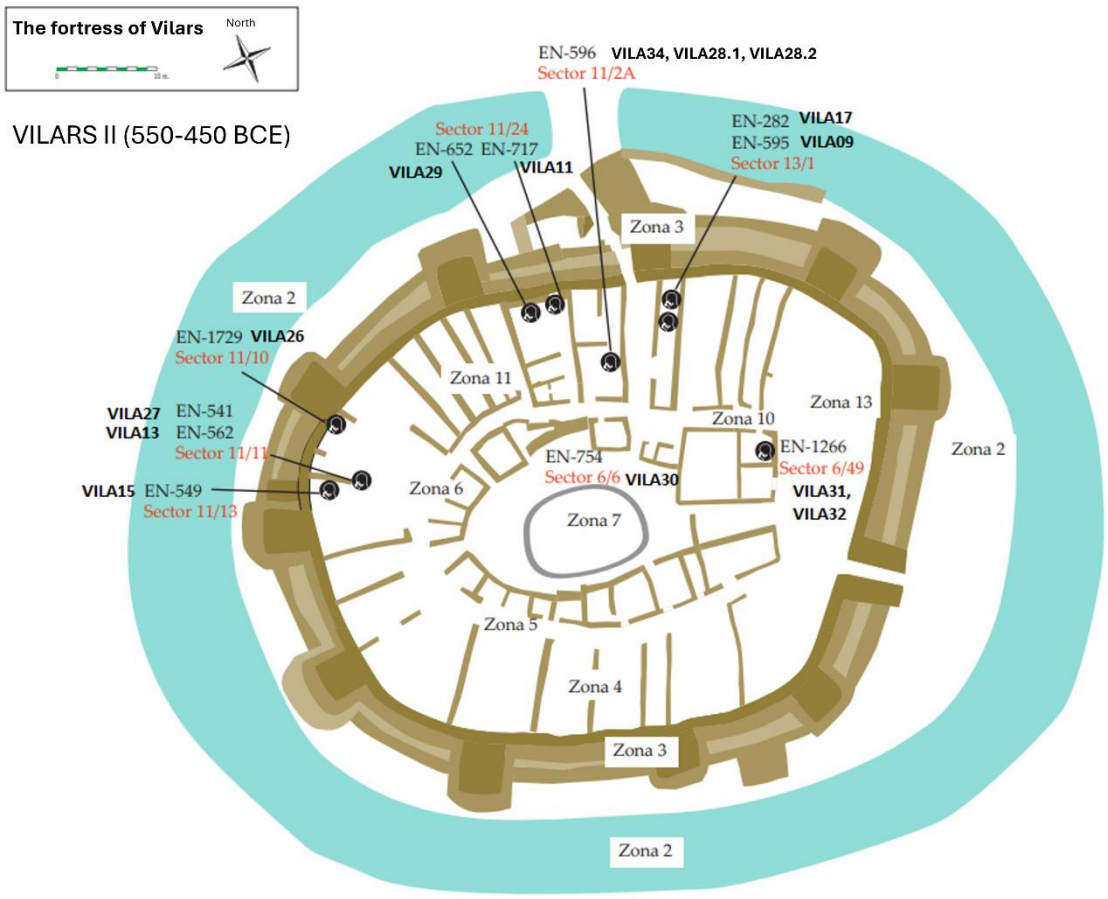


Figure S3. Ground plan of the fortress of Vilars phase II (550-450 BCE) located in Arbeca (Lleida, Catalonia, Spain). Site is divided into different zones ("*Zona*") and each zone into sectors, highlighted in red. Burials are indicated as "EN-", and the samples found in each one are also indicated as "VILA". When more than one newborn was found within the same burial, they are separated by commas. A total of 14 newborns were found in 11 different burials. Figure modified from Nieto Espinet et al. (2013)[14].

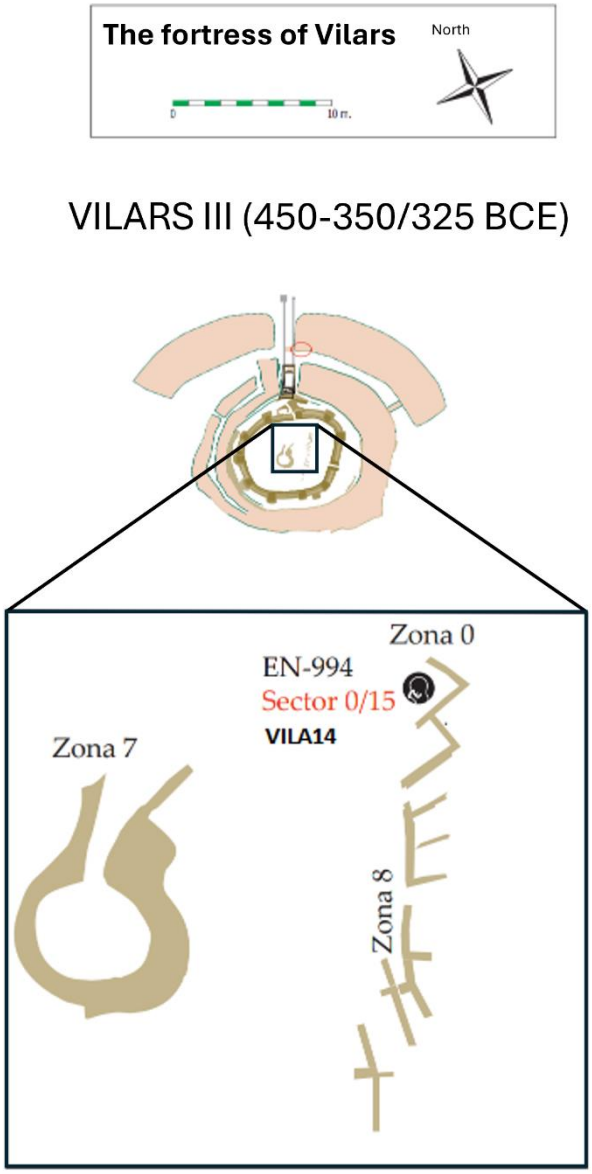


Figure S4. Ground plan of the fortress of Vilars phase III (450-325 BCE) located in Arbeca (Lleida, Catalonia, Spain). The imagen is zoomed into the zones (*"Zona"*) and sector (highlighted in red) where the newborn VILA14 was found buried (EN-994). Figure modified from Nieto Espinet et al. (2013)[14].

Sant Miquel d'Olèrdola (Iberian phase: 350-200 BCE)

The archaeological site of Santa Miquel d'Olèrdola, located in the Alt Penedès region of Catalonia (Spain), is a fortified settlement with evidence of occupation spanning from the Bronze Age to the medieval period. The site is strategically located on a hill, offering natural defensive advantages and overlooking key communication routes. Additionally, it provides a direct path to the coastline via the Vilafranca stream, leading to two potential ports where the Darró and Sitges Iberian sites are located, approximately 15-17 km away. This prime location allowed the site to play a crucial role in various historical phases, from the Iberian occupation to Roman military control and later medieval repurposing.

During the Iberian period (5th–3rd centuries BCE), Olèrdola was an important settlement featuring extensive fortifications. Excavations have uncovered a variety of structures, including domestic areas, workshops, and spaces associated with craft production. Notably, remains of a tannery/dyeing workshop and a metallurgy workshop have been documented, both dating between the mid-4th and early 2nd centuries BCE (Figure S5). These artisan zones suggest economic specialization and integration into regional trade networks. Since the beginning of the 2nd century, this space was restructured, the productive facilities disappeared, and houses were built. Dating the site has been possible thanks to the ceramics found.

A particularly remarkable aspect of the site is the discovery of multiple perinatal inhumations within these workshop areas (Figures S5-S8). Archaeological evidence indicates that at least eleven newborns were buried. Some of these burials were individual, while others were multiple, including cases of possible twin burials. The presence of these inhumations has led to discussions regarding their social and ritual significance, including potential symbolic or protective functions.

The histological age of the newborns was obtained from Martirosyan et al. (2024)[1] (Supplementary Material, Table S1). For OLE02, due to the absence of teeth, different morphometric techniques were employed, depending on the available bones[2–7] (Supplementary Material, Table S1).

For more information, refer to published articles[15–17] or to Nuria Molist Capella.

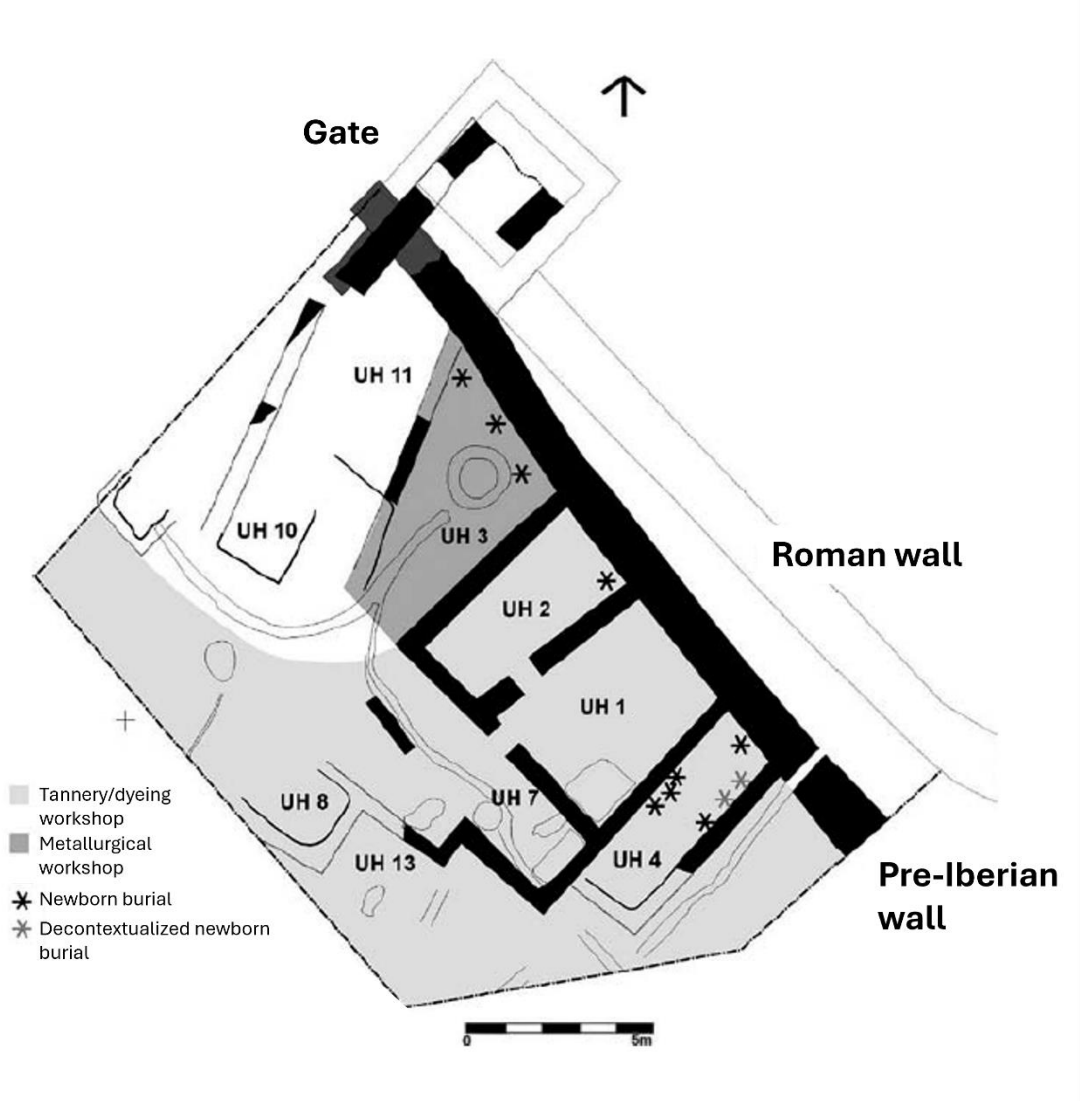


Figure S5. Ground plan of the Sant Miquel d'Olèrdola's Iberian phase (350-200 BCE), located in Olèrdola (Alt Penedès, Barcelona, Catalonia, Spain). Site is divided into different rooms ("UH"). Tannery/dyeing and metallurgical workshops are indicated in grey. Newborn burials and decontextualized newborn burials are indicated with an asterisk. Arrow indicates the North. Figure modified from Subirà and Molist (2008)[17].

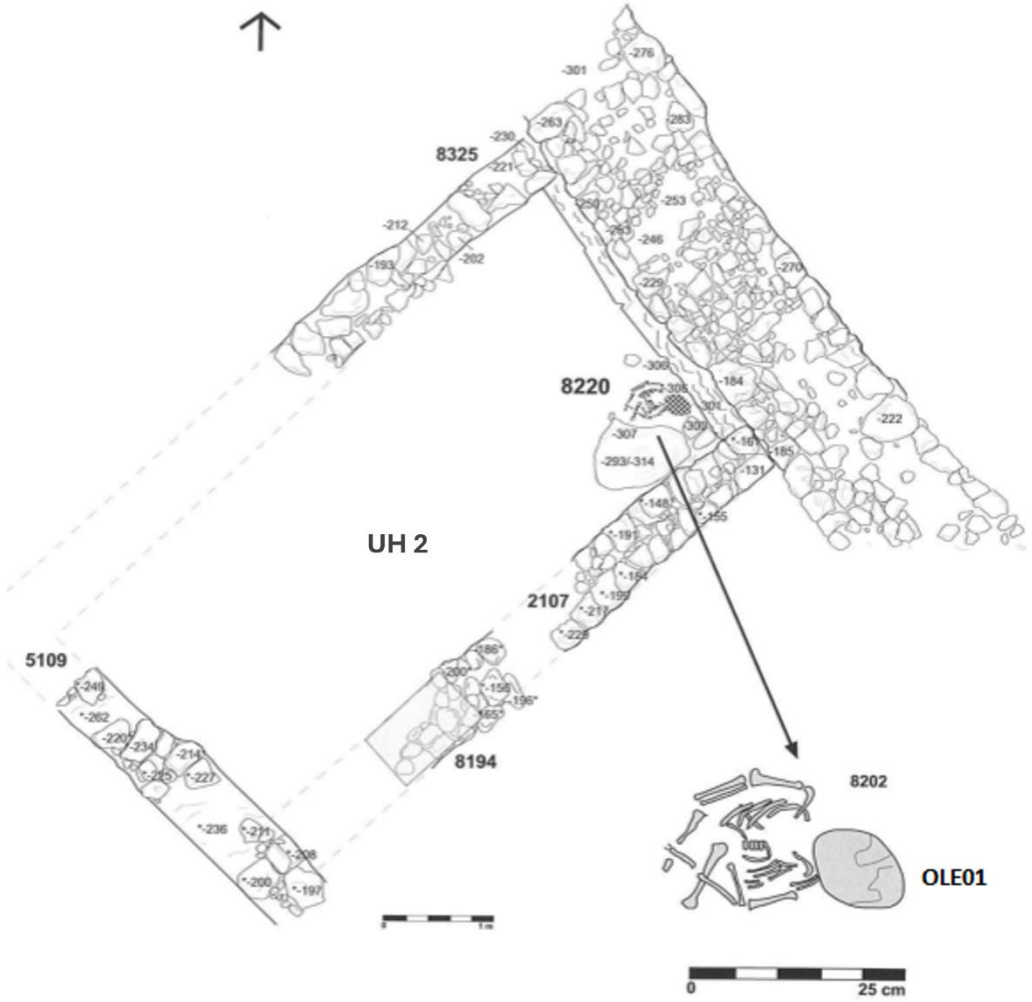


Figure S6. Ground plan of room 2 (UH 2) in Sant Miquel d'Olèrdola's Iberian phase (350-200 BCE), located in Olèrdola (Alt Penedès, Barcelona, Catalonia, Spain). It corresponded to a tannery/dyeing workshop. One newborn was found in a single burial (8202). Arrow indicates the North. Figure modified from Subirà and Molist (2008)[17].



Figure S7. Ground plan of room 3 (UH 3) in Sant Miquel d'Olèrdola's Iberian phase (350-200 BCE), located in Olèrdola (Alt Penedès, Barcelona, Catalonia, Spain). It corresponded to a metallurgical workshop. Two newborns were found in two single burials (8440 and 8448). An extra newborn was found under the medieval wall, but no data is available. Arrow indicates the North. Figure modified from Subirà and Molist (2008)[17].

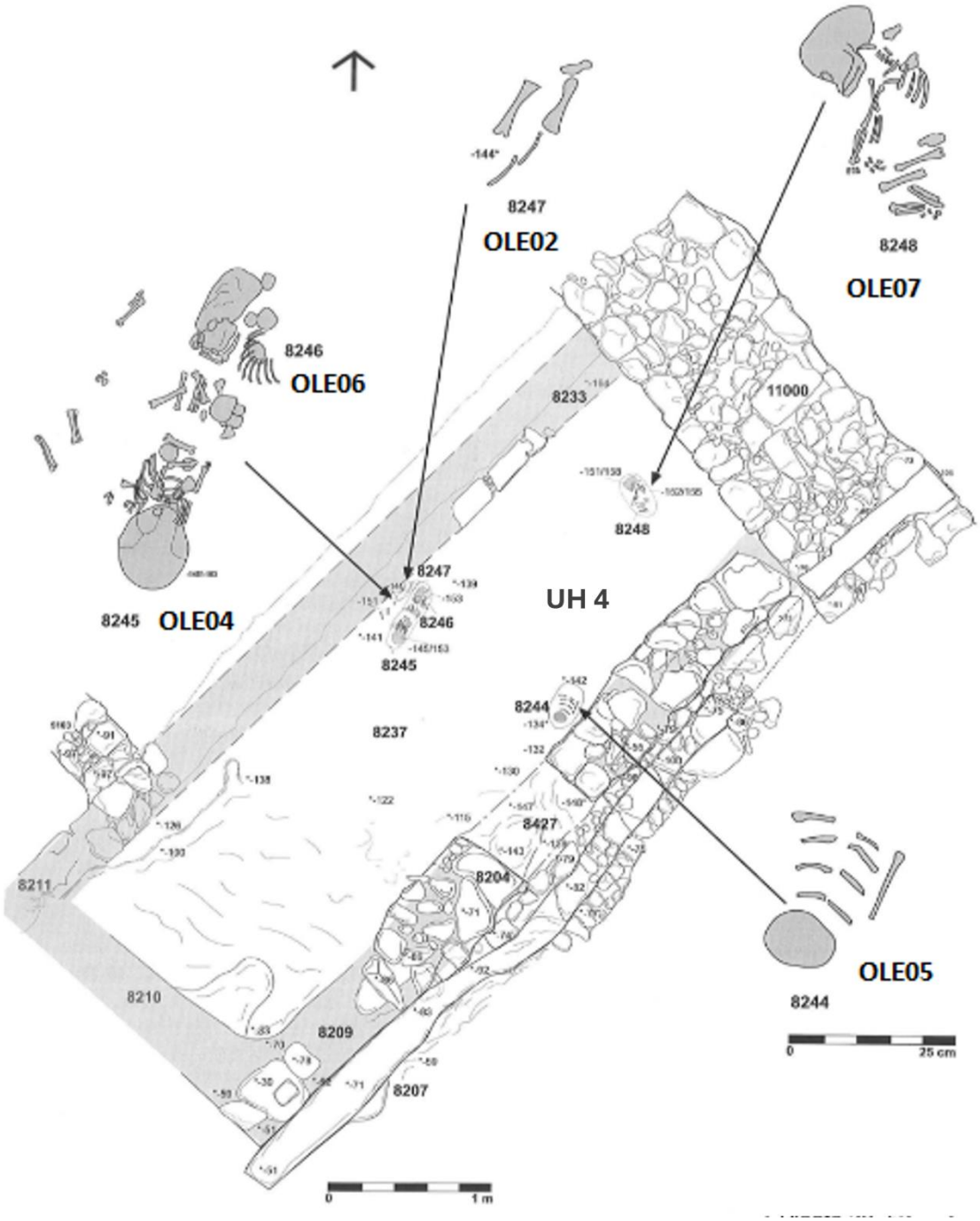


Figure S8. Ground plan of room 4 (UH 4) in Sant Miquel d'Olèrdola's Iberian phase (350-200 BCE), located in Olèrdola (Alt Penedès, Barcelona, Catalonia, Spain). It corresponded to a tannery/dyeing workshop. Five newborns were found in four different burials, three of them were individual burials (8244, 8247, and 8248), and one was a double simultaneous one (8245 and 8246). Arrow indicates the North. Figure modified from Subirà and Molist (2008)[17].

El Camp de les Lloses (125 BCE – 50 CE)

El Camp de les Lloses is an archaeological site located in the municipality of Tona (Barcelona, Catalonia, Spain), in the transitional zone between the Ter and Congost river basins, south of the Plana de Vic. The site holds significant historical and heritage value because it is one of the few well-preserved archaeological examples in Catalonia of a Roman-Republican settlement linked to military occupation. The site's foundation is closely associated with the construction of a Roman road infrastructure, which played a crucial role in the integration of the territory into the Roman world. The settlement functioned as a logistical hub, facilitating the movement and supply of troops. This is supported by the discovery of metalworking workshops dedicated to iron, bronze, and lead production, which likely served military needs. These workshops contained furnaces, slag from metallurgical processes, ingots, and tools related to metalworking, including anvils, molds, crucibles, and scales.

Archaeological evidence indicates that the settlement was occupied between approximately 125 and 75 BCE. The community that inhabited the site exhibited strong signs of Romanization, visible in both public and private spheres. This is reflected in various aspects of daily life, including clothing, personal ornaments, hygiene practices, and medical care, as well as in the architectural design of houses and domestic spaces. The use of written communication, the presence of wax tablets, and the adoption of Roman-style leisure activities further illustrate this cultural transformation. The ceramic assemblage, including amphorae and fine tableware, suggests an active involvement in commercial exchanges. The settlement was eventually abandoned towards the end of the 1st century BCE. However, evidence suggests that a hiatus occurred before a later reoccupation, during which metallurgical activities resumed. This pattern of occupation and reoccupation reflects broader transformations in the region during the Romanization of the Iberian Peninsula. Dating the site has been possible thanks to the ceramics found.

Architecturally, the site consists of a series of buildings that showcase a mix of indigenous and Roman construction techniques. The houses are rectangular (Figure S9), with rationalized layouts, stone foundations, and walls made of rammed earth (tapial). The roofs, constructed using Roman tiles (tegulae and imbrices), represent an adaptation of indigenous building styles to Roman techniques. Some buildings feature paved courtyards, storage areas, and drainage systems designed to channel rainwater efficiently.

Today, El Camp de les Lloses is an archaeological park with an interpretation center that exhibits collections from the site, including artifacts from both indigenous Iberian and Roman traditions. Ongoing excavations and research continue to provide insights into the settlement's role in the military, economic, and cultural landscape of Roman Republican Hispania.

Twelve newborns have been found in single burials, but some of them in the same houses (Figure S9). Newborns were buried both in living rooms and metallurgical workshops (Figure S9). The histological age of the newborns was obtained from Martirosyan et al. (2024)[1] or Afonso et al. (2019)[18] (Supplementary Material, Table S1).

For more information, refer to published articles[18–20] or to Montserrat Duran i Caixal and Imma Mestres Santaceu.



Figure S9. Ground plan of El Camp de les Lloses archaeological site located in Tona (Osona, Barcelona, Catalonia, Spain) with the twelve buildings (edifici A-M) colored according to its chronology. Newborns from CL1 to CL12 are indicated with dots. Figure modified from the General plan of the archaeological structures and perinatal graves provided by Montserrat Duran i Caixal and Imma Mestres Santaceu.

Supplementary figures for “Results” section

PCA

Figure S10. PCA of Iron Age and Roman samples from the Iberian Peninsula and ancient Italian samples. Present-day European and Caucasus and Southwestern Asian populations are indicated in gray. Ancient samples from the Italian Peninsula, spanning the Neolithic to Roman period, are outlined with empty polygons. Iberian samples analyzed in this study, along with previously published Iron Age and Roman samples from the Iberian Peninsula, are represented as colored dots (see legend for details). Dot shapes for the samples from this study indicate their respective archaeological sites. Iberian samples located outside the Italians polygons are labeled.

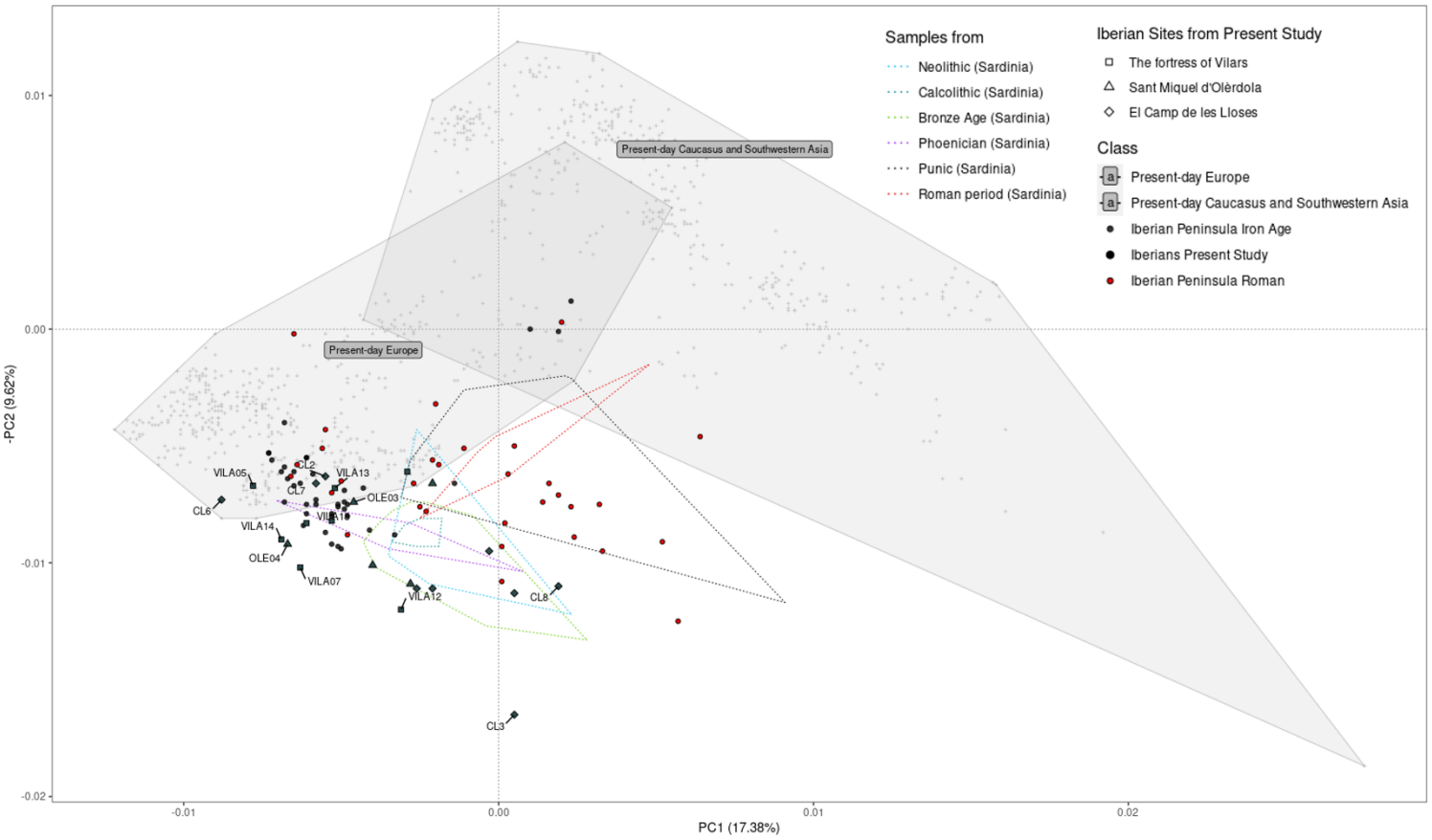


Figure S11. PCA of Iron Age and Roman samples from the Iberian Peninsula and ancient Sardinian samples. Present-day European and Caucasus and Southwestern Asian populations are indicated in gray. Ancient samples from Sardinia, spanning the Neolithic to Roman period, are outlined with empty polygons. Iberian samples analyzed in this study, along with previously published Iron Age and Roman samples from the Iberian Peninsula, are represented as colored dots (see legend for details). Dot shapes for the samples from this study indicate their respective archaeological sites. Iberian samples located outside the Sardinian polygons are labeled.

Multidimensional scaling for 1-f_3(Ind1,Ind2;Mbuti)

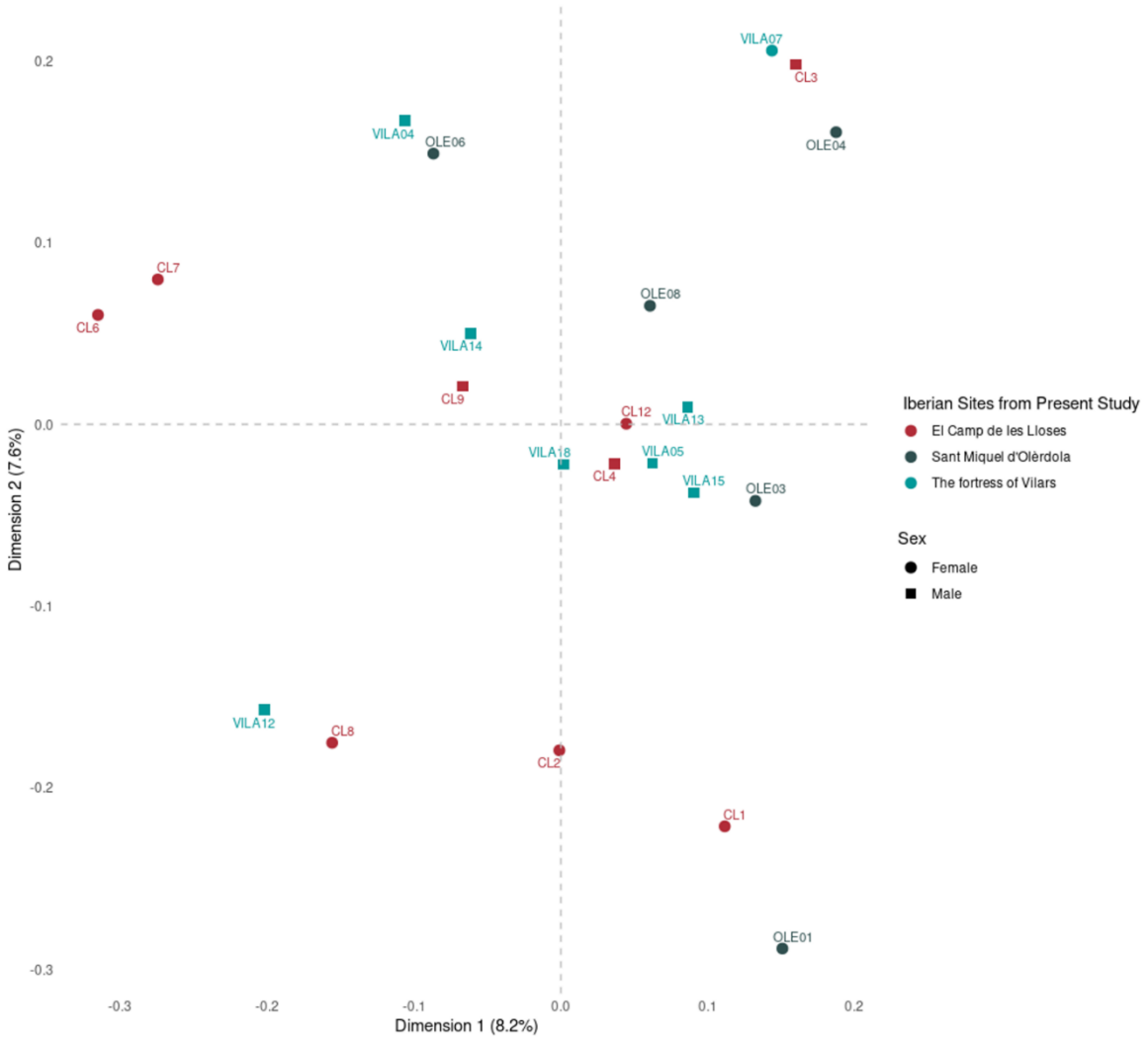


Figure S12. MDS of the Iberians analyzed in the present study. All samples presented around 20K autosomal SNPs from the 1240K panel[21]. Colors indicate the Iberian site of the samples and shapes indicate their biological sex.

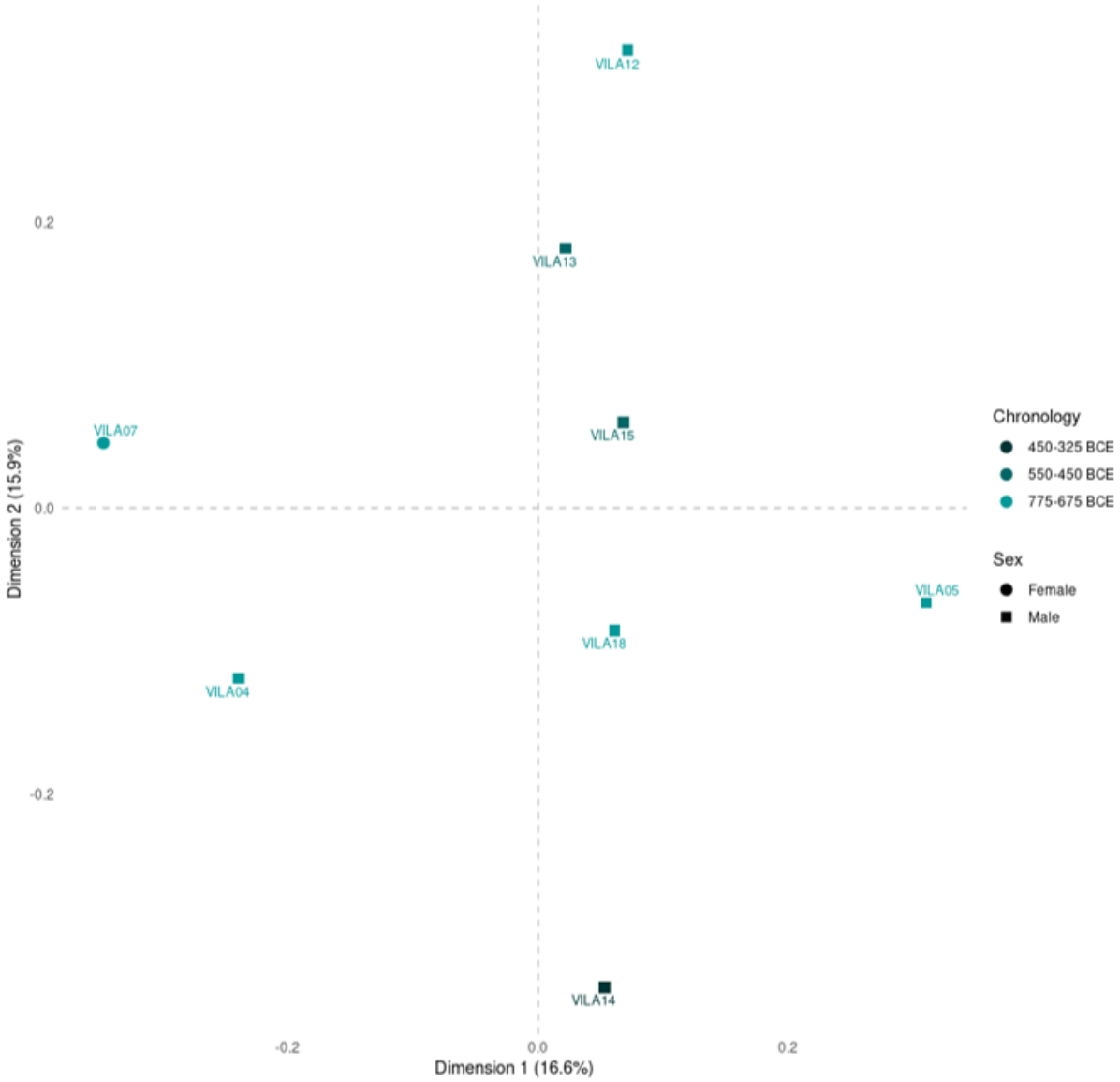


Figure S13. MDS of the Iberians from the fortress of Vilars. All samples presented around 20K autosomal SNPs from the 1240K panel[21]. Colors indicate the chronology of each sample, which is associated with their archaeological phase, and shapes indicate their biological sex.

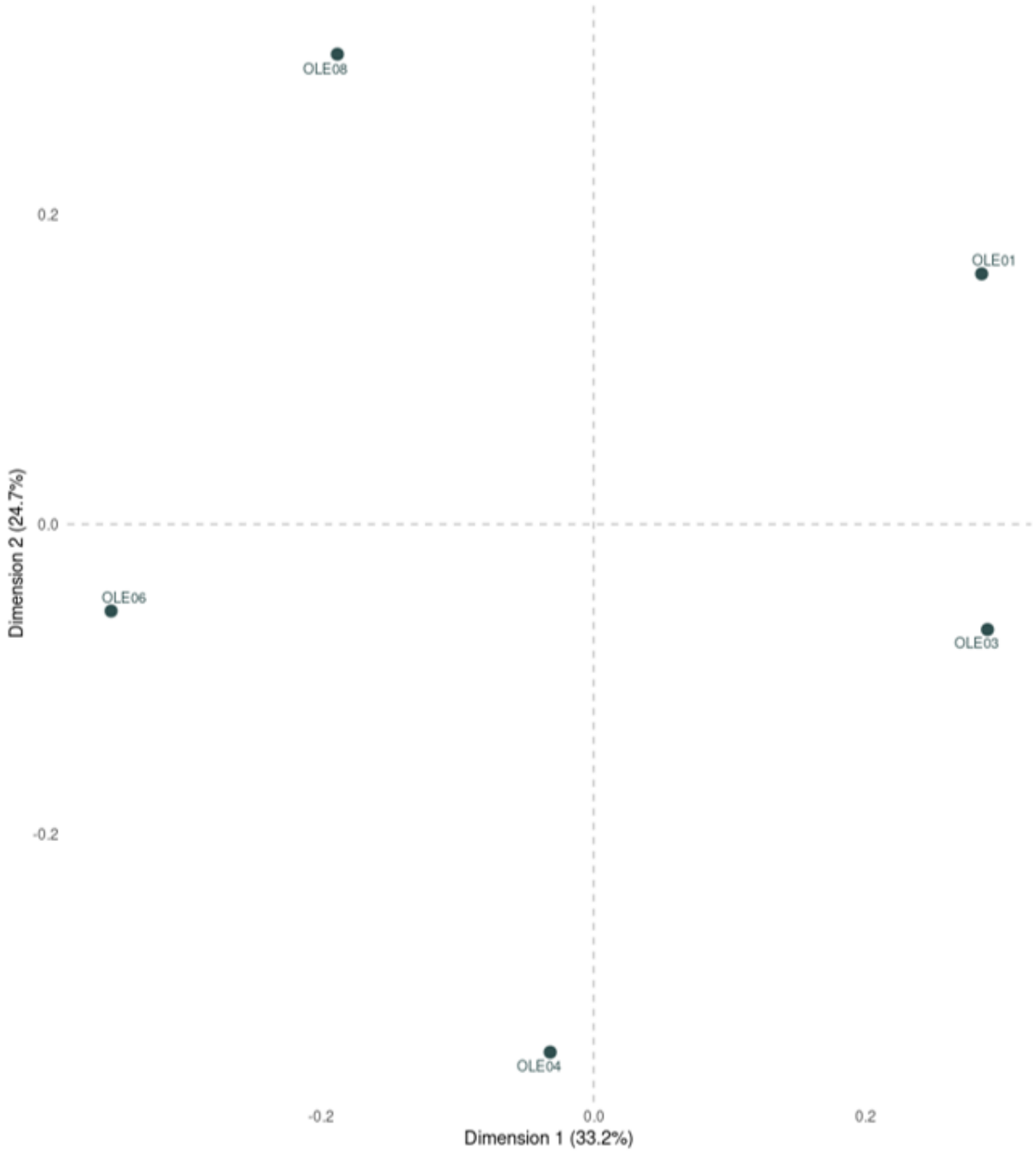


Figure S14. MDS of the Iberians from Sant Miquel d'Olèrdola. All samples presented around 20K autosomal SNPs from the 1240K panel[21]. All samples were female and belonged to the same archaeological phase (350-200).

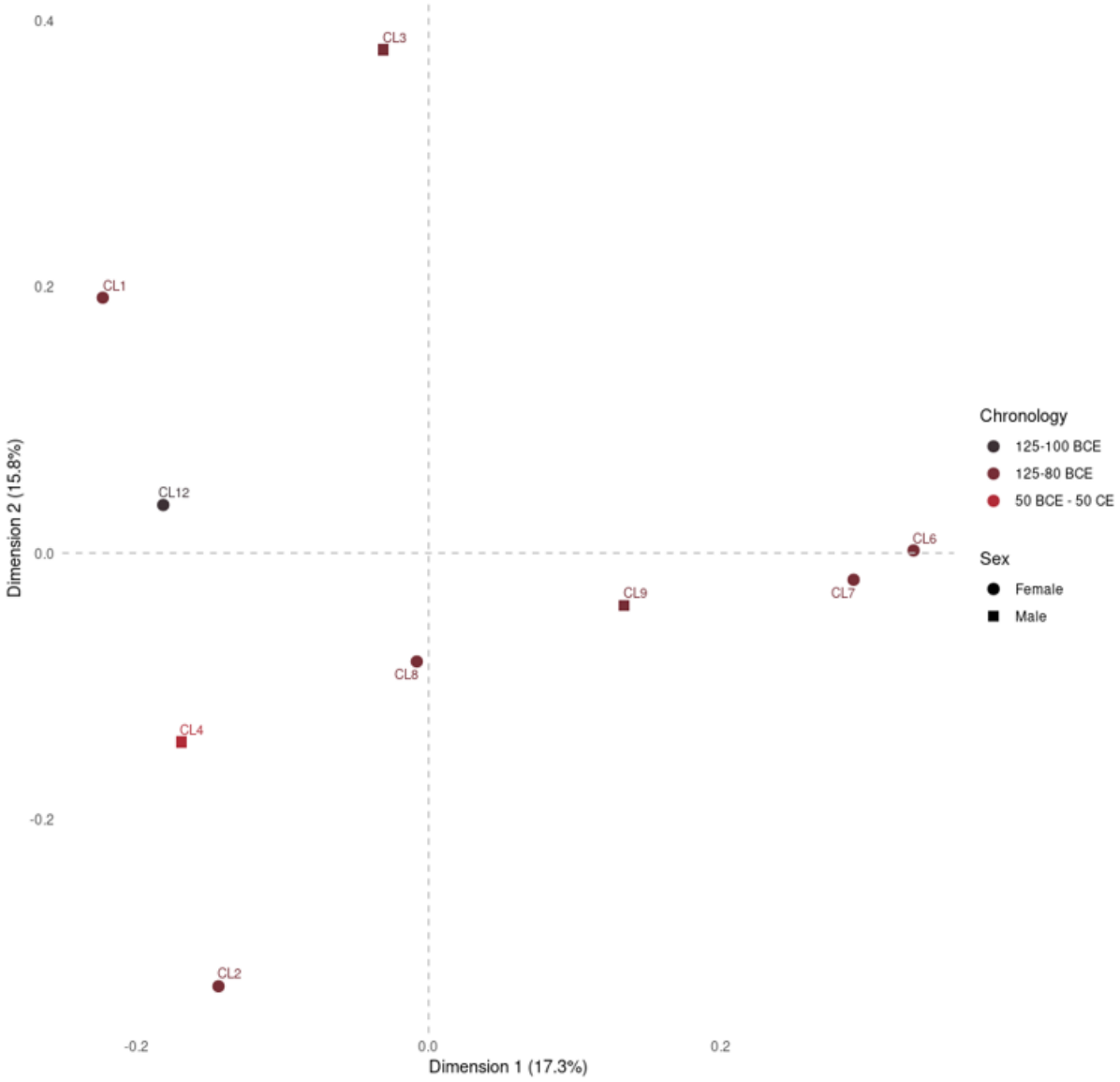


Figure S15. MDS of the Iberians from El Camp de les Lloses. All samples presented around 20K autosomal SNPs from the 1240K panel[21]. Colors indicate the chronology of each sample, which is associated with their archaeological phase, and shapes indicate their biological sex.

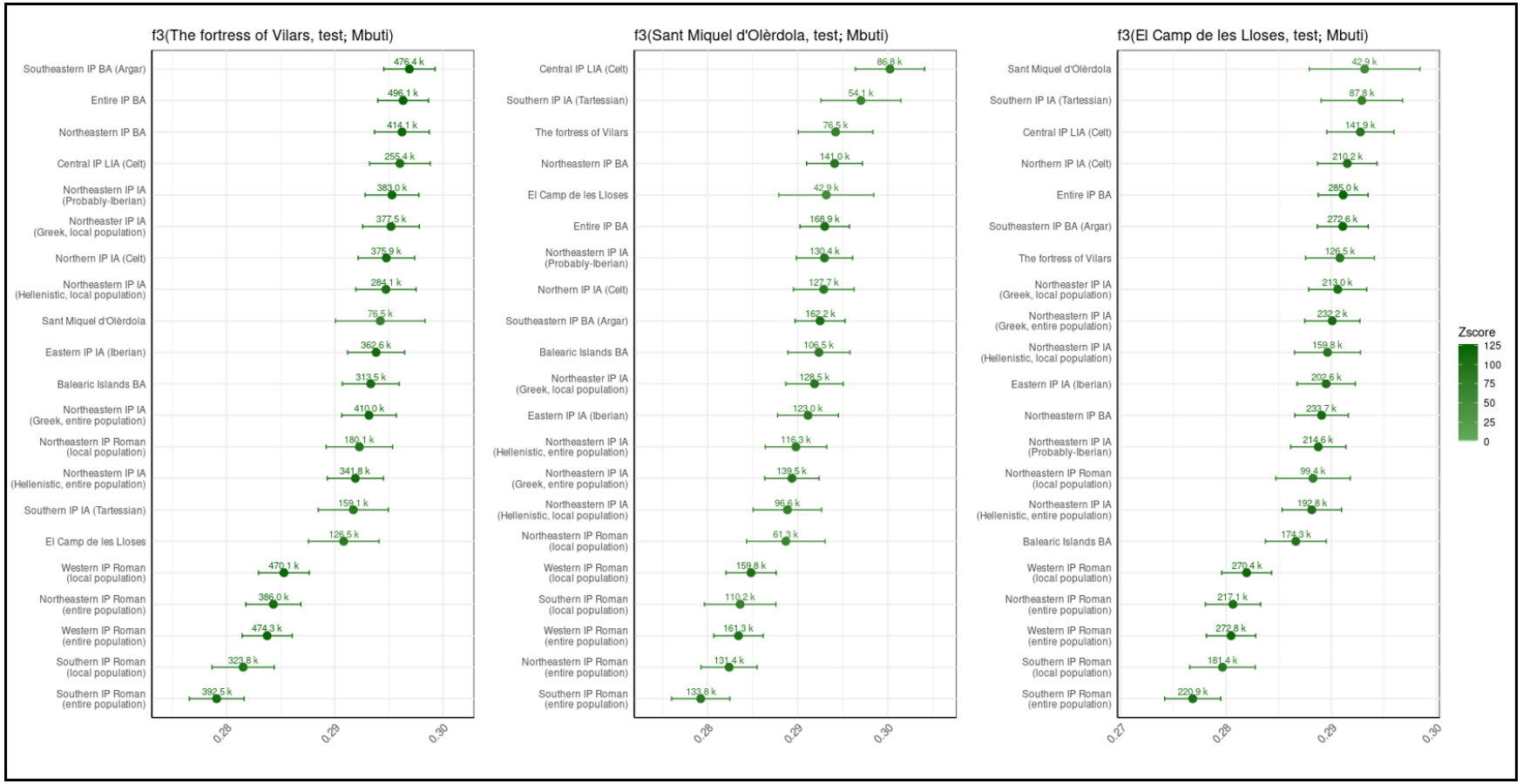


Figure S16. Outgroup f_3-statistic in the form f_3(IberianPopulation, test; Mbuti) for the fortress of Vilars, Sant Miquel d'Olèrdola, and El Camp de les Lloses. Color represents the Z-score values, with a minimum Z-score of 65.173 and a maximum Z-score of 127.095. The number of SNPs is displayed in a shortened format, where values are expressed in thousands. The analysis reveals a consistent pattern across all three Iberian sites: two distinct groups - one consisting of Iron Age populations and local Roman individuals from the region of present-day Catalonia, and the other including Romans from the broader Iberian Peninsula and the entire Roman population of the region from present-day Catalonia. All results are displayed in Supplementary Material (Table S7).

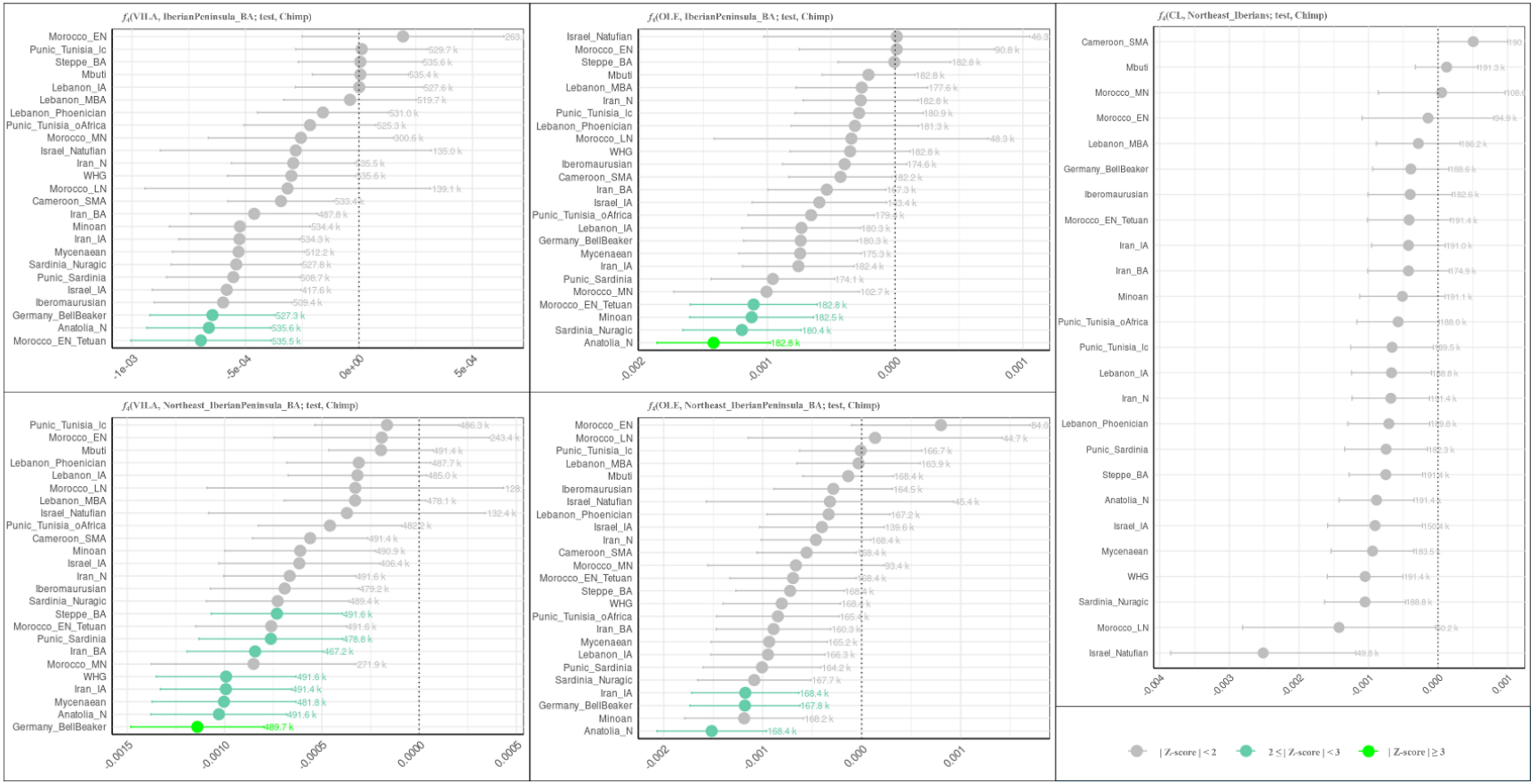


Figure S17. Results of the f_4-statistic analyses in the form of f_4(Target, BA-IberianPeninsula; Test, Chimp) for the Fortress of Vilars (VILA) and Sant Miquel d'Olèrdola (OLE), and in the form of f_4(Target, Northeast_Iberians; test, Chimp) for El Camp de les Lloses. Colors represent the Z-score values: grey indicates absolute values lower than 2, aquamarine represents values between 2 and 3, and green indicates values higher than 3. The number of SNPs is in a shortened format, with values expressed in thousands. All results can be found in the Supplementary Material (Table S8).

PCA made with qpAdm top models

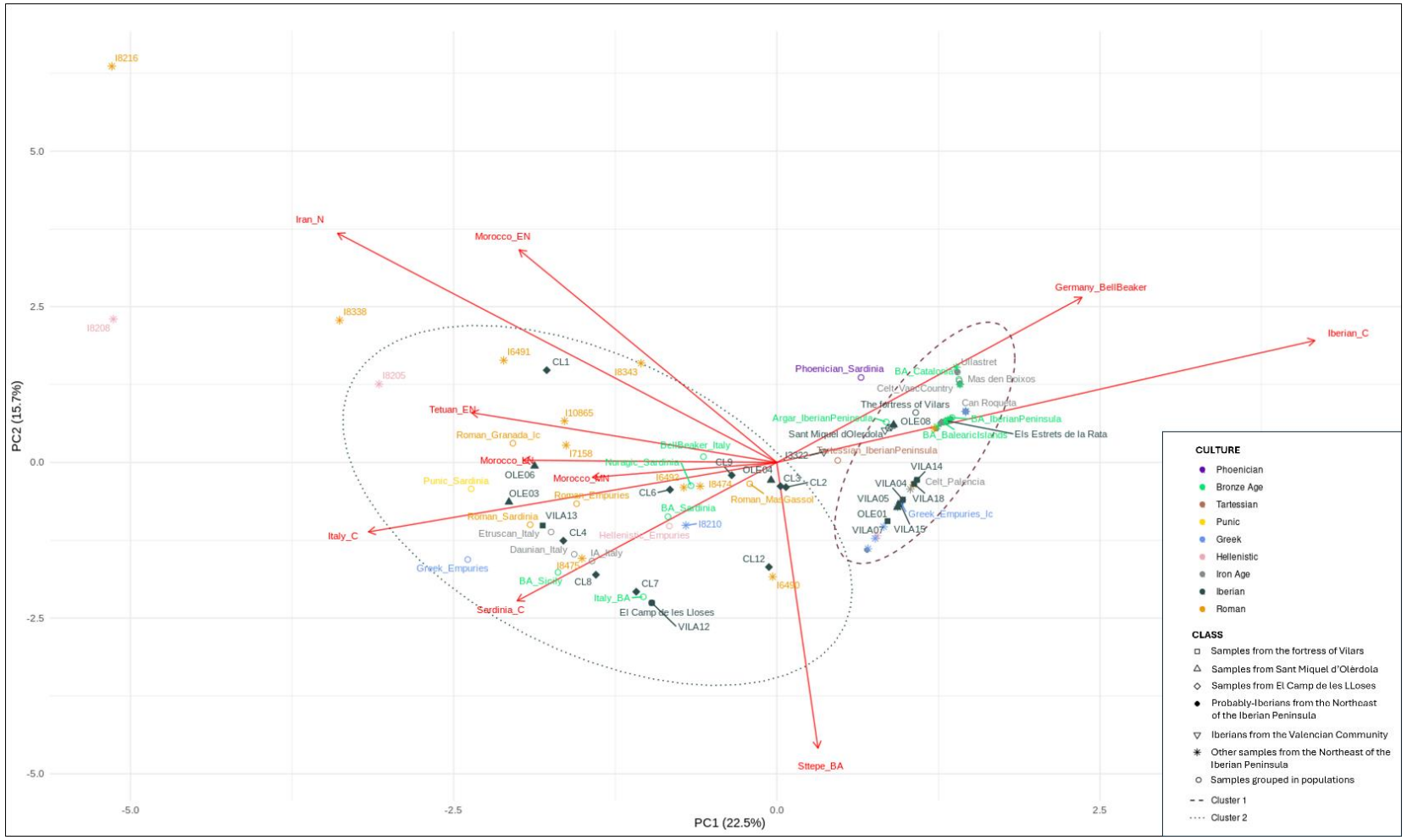


Figure S18. PCA with the top qpAdm models showing the mean ancestry proportions for each sample and population analyzed. Colors indicate culture, and shapes represent sample type. All populations are labeled, as well as all new Iberian samples from the present study. Samples from the Iberian Peninsula are labeled only if they fall outside Cluster 1. Results are available in Supplementary Material, Tables S11 and S12. A close-up can be seen in Figure 9.

Bibliography

1. Martirosyan, A., Sandoval-Ávila, C., Irurita, J., Juanhuix, J., Molist, N., Mestres, I., Durán, M., Alonso, N., Santos, C., Malgosa, A., *et al.* (2024). Reconstructing infant mortality in Iberian Iron Age populations from tooth histology. *J. Archaeol. Sci.* **171**, 106088.

2. Sherer, D. M., Sokolovski, M., Dalloul, M., Khoury-Collado, F., Osho, J. A., Lamarque, M. D. & Abulafia, O. (2006). Fetal clavicle length throughout gestation: a nomogram. *Ultrasound Obstet. Gynecol.* **27**, 306–310.

3. Irurita Olivares, J. & Alemán Aguilera, I. (2017). Proposal of new regression formulae for the estimation of age in infant skeletal remains from the metric study of the pars basilaris. *Int. J. Legal Med.* **131**, 781–788.

4. Scheuer, L. & MacLaughlin-Black, S. (1994). Age estimation from the pars basilaris of the fetal and juvenile occipital bone. *Int. J. Osteoarchaeol.* **4**, 377–380.

5. Carneiro, C., Curate, F., Borralho, P. & Cunha, E. (2013). Radiographic fetal osteometry: Approach on age estimation for the portuguese population. *Forensic Sci. Int.* **231**, 397.e1-397.e5.

6. Jeanty, P. (1983). Fetal limb biometry. *Radiology* **147**, 601–602.

7. Scheuer, J. L., Musgrave, J. H. & Evans, S. P. (1980). The estimation of late fetal and perinatal age from limb bone length by linear and logarithmic regression. *Ann. Hum. Biol.* **7**, 257–265.

8. Irurita Olivares, J., Alemán Aguilera, I., Viciano Badal, J., De Luca, S. & Botella López, M. C. (2014). Evaluation of the maximum length of deciduous teeth for estimation of the age of infants and young children: proposal of new regression formulas. *Int. J. Legal Med.* **128**, 345–352.

9. AlQahtani, S. J., Hector, M. P. & Liversidge, H. M. (2010). Brief communication: The London atlas of human tooth development and eruption. *Am. J. Phys. Anthropol.* **142**, 481–490.

10. Universitat de Lleida. at (2020). La Fortalesa del Vilars d'Arbeca. Terra, aigua i poder en el món iber. 166.

11. Natàlia Alonso M. Gené, E. Junyent, A. Lafuente, J. B. López, A. Moya, E. Tartera, A. Vidal, A. C. (2004). *Els Vilars 2001. Memòria de La XVII[a] Campanya d'excavacions Arqueològiques*.

12. Junyent Sánchez, E. & Moya i Garra, A. (2011). Els fossats de la fortalesa dels Vilars d'Arbeca (Catalunya, Espanya). *Rev. d'Arqueologia Ponent* **21**, 93–120.

13. Agustí, B., Alonso, N., Garcés, I., Junyent, E., Lafuente, Á. & López, J. B. (2000). Una inhumación múltiple de perinatales en la fortaleza de Els Vilars (Arbeca, Lleida) y las prácticas de enterrami-ento en hábitat durante la primera Edad del Hierro en el valle del Segre (Cataluña). in *Archéologie de la mort, Archéologie de la tombe au Premier Âge du Fer. Actes du XXIe Colloque International de l'AFEAF tenu à Conques-Montrozier le 8-11 mai 1997.* 305–324.

14. Nieto Espinet, A. (2013). Porcs, cavalls, ovelles i infants. Noves aportacions a les pràctiques rituals de la fortalesa dels Vilars (Arbeca, les Garrigues). *Rev. d'arqueologia Ponent* **23**, 127–162.

15. Crespo, L., Subirà, M. E. & Ruiz, J. (2011). Twins in prehistory: The case from Olèrdola (Barcelona, Spain; s. IV II BC). *Int. J. Osteoarchaeol.* **21**, 751–756.

16. Molist Capella, N. & Bosch Casadevall, J. M. (2002). *Memòria de La Intervenció Arqueològica Al Conjunt Històric d'Olèrdola (2000-2001)*.

17. Subirà, E. & Molist, N. (2008). Inhumacions perinatals múltiples i espais de treball en els assentaments ibers. in *Nasciturus: infans, puerulus. Vobis mater terra. La muerte en la infancia* (eds. Gusi, F., Muriel, S. & Olaria, C. M.) 365–385.

18. Afonso, C., Nociarova, D., Santos, C., Martinez-Labarga, C., Mestres, I., Duran, M. & Malgosa, A. (2019). Sex selection in late Iberian infant burials: Integrating evidence from morphological and genetic data. *Am. J. Hum. Biol.* **31**, 1–10.

19. Duran i Caixal, M., Mestres i Santacreu, I., Padrós, C. & Principal, J. (2017). El Camp de les Lloses, un exemple d'implantació militar al territori d'època romanorepublicana. *Trib. d'Arqueologia* **Nº 2014-20**, 11–45.

20. Duran, M., Mestres, I., Principal, J. & Padrós, C. (2015). El Camp de les Lloses (Tona , Osona): un post avançat amb tallers metal·lúrgics al servei de la logística de l'exèrcit romanorepublicà (125-75 ane). *Rev. d'Arqueologia Ponent* **25**, 293–307.

21. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., *et al.* (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503.

4.4. Article 4: Mitochondrial DNA diversity in Northeast Iberians during the Iron Age

Authors

Daniel R. Cuesta-Aguirre, M. Rosa Campoy-Caballero, Carolina Sandoval-Ávila, Cesc Busquets i Costa, Marta Fàbregas i Espadaler, Alejandro G. Sinner, Gabriel de Prado, Nuria Molist Capella, Montserrat Duran i Caixal, Imma Mestres Santacreu, Natalia Alonso, Maria Pilar Aluja, Assumpció Malgosa, Cristina Santos

Journal

Journal of Archaeological Science

Impact factor (JCR) and quartile ranking

2.6 and Q1

Date of publication

It will be submitted soon

DOI

Not available yet

Addressed objectives

This fourth publication investigates the origin and evolution of Northeastern Iberian communities and addresses the previously defined objective 3a:

3) To investigate the genetic origin and evolution of Northeastern Iberian communities using the newborns found buried under the houses as a proxy of the Iberian populations, more specifically:

    b. To compare the mitochondrial diversity of Northeastern Iberian populations from different Iberian tribes with earlier and later groups from the same region, to trace the origins and distribution of specific mitochondrial haplogroups identified in the Iberian populations to understand their evolutionary trajectories, and to investigate maternal lineage diversity over time, with particular attention to signals of long-distance female mobility.

Supplementary data is available in "Article 4" at the following link: https://uab-my.sharepoint.com/:f:/g/personal/2103711_uab_cat/EhLQT7GxLs1GoowdsaXF7QYBF7DKCwofUOyqhbjFw4XWRw?e=3ddWqy

Mitochondrial DNA diversity in Northeast Iberians during the Iron Age

*Daniel R. Cuesta-Aguirre[a], M. Rosa Campoy-Caballero[a], Carolina Sandoval-Ávila[a], Cesc Busquets i Costa[b,c,d], Marta Fàbregas i Espadaler[b,c], Alejandro G. Sinner[e], Gabriel de Prado[f], Nuria Molist Capella[g], Montserrat Duran i Caixal[h], Imma Mestres Santacreu[h], Natalia Alonso[i], Maria Pilar Aluja[a], Assumpció Malgosa[a], Cristina Santos[a]**

[a]Grup de Recerca en Antropologia Biològica (GREAB), Unitat d'Antropologia Biològica, Departament of Animal Biology, Plant Biology and Ecology, Faculty of Biosciences, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

[b]Fundació Privada Conjunt Monumental Castell de Besora

[c]Atics SL

[d]Departament de Ciències de l'Antiguitat i de l'Edat Mitjana, Àrea d'Arqueologia, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

[e]Greek & Roman Studies, University of Victoria, Canada

[f]Museu d'Arqueologia de Catalunya, Ullastret, Giorona, Spain

[g]Museu d'Arqueologia de Catalunya, Barcelona, Barcelona, Spain

[h]Jaciment i Museu El Camp de les Lloses, Tona, Barcelona, Spain

[i]Universitat de Lleida, Lleida, Spain

Corresponding Author

Cristina Santos, PhD

Research Group in Biological Anthropology

Unitat Antropologia Biològica

Dep. Biologia Animal, Biologia Vegetal i Ecologia

Facultat Biociències Edifici C

Universitat Autònoma de Barcelona

08193 Cerdanyola del Vallès

Barcelona (Spain)

Tel.: +34 93 581 13 17

Fax: +34 93 581 13 21

https://orcid.org/0000-0002-7014-8014 cristina.santos@uab.cat

Information about the authors: ORCID and email

- Daniel R. Cuesta-Aguirre; 0000-0001-7483-8541; daniel.ruizdelacuesta@uab.cat
- M. Rosa Campoy-Caballero; 0009-0004-3387-2005; mariarosa.campoy@uab.cat
- Carolina Sandoval-Ávila; 0000-0001-7678-1390; carolina.sandoval@uab.cat
- Cesc Busquets i Costa; 0000-0001-5029-2919; cesc@atics.org
- Marta Fàbregasi Espalader; marta@atics.org
- Alejandro G. Sinner; 0000-0002-5816-5794; agsinner@uvic.ca
- Gabriel de Prado; 0000-0002-1408-1361; gdeprado@gencat.cat
- Nuria Molist Capella; 0000-0003-2048-6987; nmolist@gencat.cat
- Montserrat Duran i Caixal; 0000-0002-8098-8156; montserratduranicaixal@gmail.com
- Imma Mestres Santacreu; 0000-0003-4527-2678; mestressi@tona.cat
- Natalia Alonso; 0000-0003-4081-0262; natalia.alonso@udl.cat
- Maria Pilar Aluja; 0000-0002-0407-9356; mariapilar.aluja@uab.cat
- Assumpció Malgosa; 0000-0003-1723-3671; assumpcio.malgosa@uab.cat

Abstract

Iberian culture emerged along the Mediterranean coast of the Iberian Peninsula around the 6th century BCE, flourishing between 400-200 BCE until the Roman conquest. Iberians engaged in metallurgy, agriculture, and livestock, and actively participated in Mediterranean commercial trade networks. Although cremation was the predominant funerary practice, advances in ancient DNA techniques have enabled the recovery of mitochondrial DNA (mtDNA) from an increasing number of Iberian individuals. Here, we analyzed 31 newborns, successfully obtaining mtDNA profiles for 21 individuals (20 Iberians and 1 Late Roman). These data were merged with 41 previously published mtDNA profiles from unrelated Iberians across different tribes of the Northeast of the Iberian Peninsula. Additional prehistoric data were compiled to contextualize Iberian haplogroups. We investigated maternal lineage diversity between tribes, temporal shifts in haplogroup composition, and signatures of long-distance female mobility. Our results revealed subtle differences in mtDNA haplogroup frequencies between tribes, although genetic differentiation was not statistically significant. Mitochondrial diversity remained relatively high across all tribes, consistent with patrilocal mating systems and small-distance female migration that may have prevented strong differentiation among tribes. A predominance of haplogroups H, J, K, HV0, and U was observed, most of which were already present in the Iberian Peninsula before the Iron Age. Haplogroup diversity remained stable over time, without population differentiation, suggesting maternal genetic continuity from the Bronze Age. However, the presence of some haplogroups pointed to occasional female-mediated gene flow from North Africa, the Near East, and Central Europe. Overall, this study provides the most comprehensive NGS-based assessment of maternal ancestry in Iron Age Iberians to date, revealing a genetic landscape shaped by local continuity alongside some long-distance female mobility linked to commercial trade and cultural interaction.

Keywords

Mitochondrial DNA (mtDNA); Iberian culture; Paleogenetic; Female mobility; Iron Age (IA)

1. INTRODUCTION

Iberian culture emerged in the 6th century BCE in the Iberian Peninsula along the eastern and southern coast of the Mediterranean Sea. It reached its peak between 450 and 200 BCE. This culture developed a distinct identity marked by its alphabet, coinage, iron metallurgy, and economic activities like agriculture, livestock, fishing, and Mediterranean trade (Almagro, et al., 2001; Jordá Cerdá et al., 1986). Far from being a homogenous entity, the term "Iberians" (later coined by the Romans) encompassed multiple groups or tribes — Ausetani, Cessetani, Ilergetae, Indiketes, Lacetani, Layetani, etc. — each with its own sociopolitical trajectory.

Their funeral practices were primarily based on cremation. However, cremation necropolises did not represent all types of Iberian burial (Oliver Foix, 2004), as some newborns have been found buried under the pavement of houses (Afonso et al., 2019; Crespo et al., 2011; Subirà & Molist, 2007). This funerary rite has been widely discussed in different archaeological contexts (Eng & Smith, 1976; Gowland et al., 2014; Hassan et al., 2014), but recent evidence suggests that the age profile of Iberian newborns aligns well with natural mortality causes (Martirosyan et al., 2024), and both male and female individuals have been identified (Cuesta-Aguirre et al., 2025; Sandoval-Ávila et al., 2025).

The origins of Iberian society have been broadly debated. Early theories proposed a Neolithic African substrate influenced by Greek colonists, leading to the idea of an expansive Ibero-Tartessian culture (Bosch-Gimpera, 1938). However, Maluquer de Motes and Tarradell later proposed a model of simultaneous local development, with colonial influence acting as a catalyst rather than a cause (Maluquer de Motes, 1966, 1982; Sanmartí, 2011, 2013; Tarradell, 1962). In the Northeast of the Iberian Peninsula, a clear shift toward Iberian cultural traits occurred after the 6th century BCE, seen in increased urbanization, territorial organization, and the rise of aristocratic elites (Asensio et al., 2001; Belarte, 2010; Bendala, 1998; Bosch & Santacana, 2009; Junyent, 1996a, 1996b, 2023; Sanmartí, 2003; Sanmartí et al., 1998).

From the 9th to 6th centuries BCE, Phoenician and Greek colonization reshaped Iberian territories through the foundation of trade hubs such as Gadir, Malaka, Abdera, and Emporion. These settlements connected the Iberians to wider Mediterranean economic and cultural exchanges (Dietler, 2009; Gitin & Bierling, 2002; Keay & Deamos, 1998). Later, Carthage, founded in the 9th century BCE and expanding from the 7th century BCE, reinforced these

connections. By 236 BCE, Carthaginian military campaigns brought much of the southern Iberian Peninsula under their control, establishing cities like Qart-Hadasht (Cartago Nova). However, the Roman expansion soon clashed with Carthaginian ambitions, leading to the Punic Wars (Bagnall, 1999; Carey et al., 2007; Cornell, 1995; Edwell, 2011; Feeney, 2017; Goldsworthy, 2012; Grant, 1978; Richard, 2011; Schullard, 1989; Zimmermann, 2011). In 218 BCE, Rome landed in Emporion, gaining support from northeastern Iberian tribes (Zimmermann, 2011). By 206 BCE, the Battle of Ilipa (near present-day Seville) secured Roman dominance (Edwell, 2011; Zimmermann, 2011), initiating a gradual process of cultural and sociopolitical integration known as Romanization.

Different Iberian tribes interacted with a range of external cultures, shaping their identities in diverse ways. While archaeological evidence reveals cultural traits and trade networks, it often cannot clarify the nature or depth of these interactions, particularly whether individuals from foreign groups integrated into local communities. Paleogenetics provides a crucial complementary perspective to explore such dynamics. It offers deeper insights into processes that shaped the Iberian Peninsula over time by uncovering the genetic makeup of ancient populations. Among genetic markers, mitochondrial DNA (mtDNA) is especially valuable in ancient DNA studies. Its robustness makes it well-suited for degraded samples where whole-genome sequencing is sometimes unfeasible (Ehler et al., 2019). Unlike nuclear DNA, which exists only two copies per cell and is more prone to post-mortem degradation, mtDNA is present in hundreds to thousands of copies, greatly increasing the chance of successful recovery from ancient remains (Gilbert et al., 2007; Pääbo et al., 2004). Its circular structure also provides stability against enzymatic degradation, aiding its long-term preservation (Krause et al., 2010). Moreover, mtDNA's strict maternal inheritance and lack of recombination simplify phylogenetic analyses (Nesheva, 2014), allowing researchers to accurately reconstruct maternal lineages. This enables the tracking of genetic continuity, migration events, and population admixture (Översti et al., 2019). It has also proven instrumental in identifying female mobility and affinities between ancient and modern populations, shedding light on long-term demographic patterns (Achilli et al., 2008).

The study of female mobility has revealed diverse kinship and residence systems across prehistoric Europe, illustrating how social structures evolved. In Neolithic and Bronze Age contexts, many societies followed patrilocal models with widespread female exogamy. For example, in Corded Ware groups of Bohemia (2,900-2,400 BCE), genetic and isotopic data showed that non-local women were integrated into communities that maintained stable Y-

chromosome lineages but diverse mitochondrial DNA, indicating patrilocality and female mobility (Papac et al., 2021). A patrilocal mating system during the Bronze Age, where women moved to the husband's community after marriage, was also hypothesized in the Iberian Peninsula (Simón et al., 2011; Villalba-Mouco et al., 2022). In contrast, evidence for matrilocal or female-centered systems appeared in other contexts. In Neolithic Greece, a marked decline in mitochondrial DNA diversity during the later phases of the Neolithic (by the 4th millennium BCE) suggested increasing female sedentism, likely associated with matrilocal residence (Silva et al., 2022). In Iron Age southern England (100 BCE - 100 CE), mitochondrial data from a Celtic community revealed that 64% of individuals across both sexes shared a single founding maternal lineage, consistent with matrilocality. Additionally, non-local grave goods were found in male burials, suggesting that men moved into established female kin groups through marriage (Cassidy et al., 2025). Beyond residence systems, female mobility also played a central role in cultural transmission and demographic integration. In Phoenician settlements across the Mediterranean (700-400 BCE), mitochondrial genomes from southern Sardinia and Lebanon revealed the presence of indigenous and non-local haplogroups, such as Near Eastern N1b1a5 and European W5, in Sardinia and some European lineages in Lebanon. These findings indicated bidirectional female migration between North Africa, the Levant, and Europe, demonstrating active female participation in trade and settlement networks, challenging assumptions of male-dominated migration (Matisoo-Smith et al., 2018). Together, these examples highlight how institutionalized sex-biased mobility shaped the genetic and social landscapes of past European societies, with patrilocality reinforcing male continuity and matrilocality fostering female-centered kinship systems.

Despite extensive research on the prehistory of the Iberian Peninsula (Gamba et al., 2014; Olalde et al., 2019; Roca-Rada et al., 2024; Villalba-Mouco et al., 2021, 2023), genetic studies focusing specifically on the Iron Age remain limited, particularly for Iberian culture. This scarcity is largely due to the limited availability of anthropological remains, as cremation was the predominant funerary practice among Iberian groups. The earliest paleogenetic research on Iberian populations focused on mitochondrial DNA, particularly the hypervariable region I. These initial studies aimed to identify the main mitochondrial haplogroups present in Iron Age Iberians and to compare them with those from the Chalcolithic, Bronze Age, and Roman periods (Gamba et al., 2008), as well as with culturally distinct groups like the Basques and Etruscans (Sampietro et al., 2005). The analysis based on the remains of 27 adults from nine archaeological sites — including bone fragments from

areas historically associated with Iberians, though not always of confirmed Iberian origin — revealed several key insights. The haplogroup composition of these individuals was strikingly similar to that of present-day Iberian Peninsula populations, albeit with reduced variability, suggesting a degree of genetic continuity since pre-Roman times (Sampietro et al., 2005). Importantly, no close genetic relationship was observed between the Iberians and the Etruscans, underscoring the genetic diversity of pre-Roman Western Europe. Haplogroup H was the most frequent, followed by J, T, K, U4, U5, and W (Gamba et al., 2008; Sampietro et al., 2005). An unusually high frequency of haplogroup V, a lineage often associated with Paleolithic populations in Europe (Torroni et al., 1998, 2001), was also detected (Gamba et al., 2008). Recent methodological advances, particularly in Next-Generation Sequencing (NGS) and targeted capture techniques, have significantly improved the retrieval of ancient mitochondrial data. These newer approaches enabled the recovery of mtDNA from 16 individuals (including three newborns from Valencia and 13 adults from Catalonia) across seven Iberian archaeological sites (Olalde et al., 2019), and one additional individual from Catalonia (Patterson et al., 2022). The results again showed a predominance of haplogroups H and J, along with lineages from HV0, K1, and X2b (Olalde et al., 2019; Patterson et al., 2022).

No systematic mitochondrial study has been conducted on Iberians that spans the Early Iron Age to the Roman Conquest period. In this article, we focus on high-resolution mitochondrial DNA variation identified by us (Cuesta-Aguirre et al., 2025) and other groups (Olalde et al., 2019; Patterson et al., 2022) in Iberian individuals from different tribes, and we expand our dataset by analyzing new Iberian samples. Our main objectives are to compare the mitochondrial diversity of Northeastern Iberian tribes with that of earlier and later population groups from the same region, to trace the origins and distribution of specific mitochondrial haplogroups identified in the Iberian populations to understand their evolutionary trajectories, and to investigate maternal lineage diversity over time, with particular attention to signals of long-distance female mobility.

2. MATERIAL AND METHODS

2.1. Sample selection

Thirty Iberian individuals from the fortress of Vilars, Sant Miquel d'Olèrdola, El Camp de les Lloses, Ullastret - Illa d'en Reixac, Ullastret - Puig de Sant Andreu, Castell de Besora, and El poblat Ibèric de Burriac, located in Catalonia, Spain, that span the Early Iron Age to the Roman Conquest period, were analyzed. Additionally, a Late Roman individual from the

fortress of Vilars was also analyzed, and 4 Iberian individuals from Ullastret - Illa d'en Reixac c were reanalyzed. Furthermore, information from 49 published Iberians from the Northeast of the Iberian Peninsula has been collected. All Iberian individuals involved in this study are represented in Figure 1 and Table 1. In addition, 172 individuals from the Neolithic to Visigoth era (Supplementary Material, Table S1) were compiled from previous works to build the chronological frame.

Summary and most relevant references for the archaeological context of the fortress of Vilars, Sant Miquel d'Olèrdola, and El Camp de les Lloses can be found in Cuesta-Aguirre et al. (2025), for Can-Roqueta/Can Revella, Hort d'en Grimau, and Mas d'en Boixos can be found in Olalde et al. (2019), for Illa d'en Reixac and Puig de Sant Andreu can be found in Sandoval-Ávila et al. (2025), and for Mas Castellar in Patterson et al. (2022).

Castell de Besora (Bergistani tribe) is located at approximately 1,025 meters above sea level in the municipality of Santa Maria de Besora, province of Barcelona (Catalonia, Spain). This strategically positioned site has played a crucial role in the region's history, with occupation traces dating back to the Iron Age, associated with the Ausetani tribe. Although the Iron Age strata have been poorly explored, the discovery of two Iberian newborns highlights the archaeological potential (Busquets i Costa et al., 2018). The site continued to be used through various historical periods, with the first written reference to the castle dating to the late 9th century, when it served as a defensive stronghold during the early medieval era.

El poblat Ibèric de Burriac (Layetani tribe) lies at the foot of Burriac Mountain in the province of Barcelona (Catalonia, Spain). This key settlement of the Layetani people reflects the political and economic complexity of Iberian societies in northeastern Iberia. Initially settled in the 6th century BCE, the site flourished in the 4th century BCE with the construction of fortifications and an urban street network. Spanning 7-10 hectares, it became one of the largest Iberian urban centers, exerting control over surrounding agricultural zones until the late 3rd century BCE. Following the Second Punic War, Iberian dominance declined, and Roman urban development took over. Archaeological finds include public and private buildings, thermal baths, a wine production center from the late Republican period, and a Roman mint issuing coins with the Iberian legend "ILTURO", attesting to the site's continued significance under Roman rule. So far, the remains of only one newborn from the Iron Age levels has been available (Martín i Méndez, 1998; Sinner, 2015; Sinner & Ferrer i Jané, 2016).

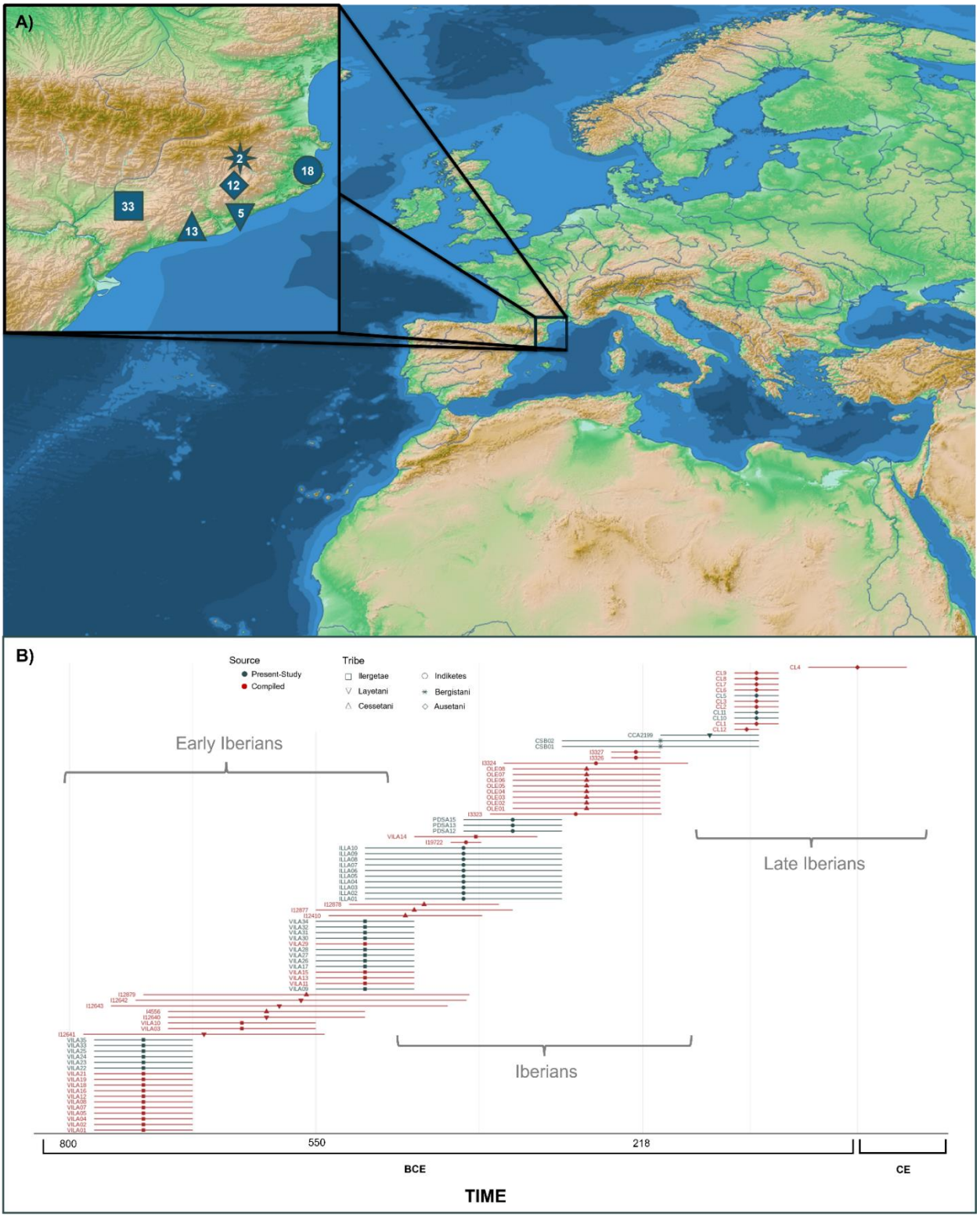


Figure 1. *Geographic (A) and temporal (B) distribution of the Iberians included in this study. The number of samples is indicated in white in panel A, and shapes correspond to the tribe each individual is associated with, as defined in panel B. Additional information on the individuals is provided in Supplementary Table S1.*

Table 1. Summary of Iberian archaeological sites from the northeastern Iberian Peninsula included in this study. Sites are grouped by associated tribe, with the number of individuals analyzed, the number of mitochondrial DNA (mtDNA) sequences retained, and the corresponding references. The number of mtDNA sequences may be lower than the number of individuals due to failed recovery or genetic relatedness

Tribe	Site	Number of Iberians available	Number of mtDNA employed	Reference
Ilergetae	The fortress of Vilar	15	10	Present-study
		18	13	Cuesta-Aguirre et al. (2025)
Cessetani	Font de la Canya	1	1	Olalde et al. (2019)
	Hort d'en Grimau	1	1	Olalde et al. (2019)
	Mas d'en Boixos	3	3	Olalde et al. (2019)
	Sant Miquel d'Olèrdola	8	6	Cuesta-Aguirre et al. (2025)
Layetani	Can-Roqueta/Can Revella	4	4	Olalde et al. (2019)
	El poblat Ibèric de Burriac	1	0	Present-Study
Indiketes	Ullastret - Illa d'en Reixac	6	2	Present-study
		4	1	Sandoval-Ávila et al. (2025)
	Ullastret - Puig de Sant Andreu	3	2	Present-study
		4	4	Olalde et al. (2019)
	Mas Castellar	1	1	Patterson et al. (2022)
Bergistani	Castell de Besora	2	2	Present-Study
Ausetani	El Camp de les Lloses	3	3	Present-Study
		9	8	Cuesta-Aguirre et al. (2025)

2.2. Sample preparation, DNA extraction, and quantification

In this study, new mtDNA data was generated for newborns from various Iberian archaeological sites across the Northeast of the Iberian Peninsula (Table 1 and Supplementary Table S1). All processing steps further described were conducted in the Anthropology Unit's dedicated clean ancient DNA laboratory at the Universitat Autònoma de Barcelona (UAB).

When available, the petrous bone was prioritized for analysis. Vertebral hemiarches or long bones (humerus, femur, or tibia) were used for samples lacking the petrous bone. The outer surfaces of all bones were initially cleaned to remove dirt and then polished with a milling cutter to eliminate the outermost layers. Petrous bones were cut transversely near the internal acoustic opening to expose the cochlea, and a small section was selected and pulverized using a hammer for subsequent DNA extraction. Similarly, a section of long bones was pulverized, while two vertebral hemiarches were directly pulverized after light polishing.

DNA was extracted from all samples using a protocol that involved mixing 25 µl of proteinase K with 1 ml of extraction buffer, followed by purification with the High Pure Viral Nucleic Acid Large Volume Assembly Kit (Roche). This kit employs a silica-in-HE-membrane column-based method, as described in Vinueza-Espinosa et al. (2020). Blank controls were included in each extraction batch to monitor and ensure the absence of contamination, which was further confirmed by performing quantitative PCR (qPCR) targeting the Hypervariable Region I (HVR-I) of mitochondrial DNA (mtDNA). The DNA concentration of each sample was quantified using the Qubit® 1x dsDNA High Sensitivity (HS) Assay Kit (Thermo Fisher Scientific, 2015). The DNA fragmentation profile was assessed using the Bioanalyzer® Agilent High Sensitivity DNA Kit (Agilent Technologies, 2017).

2.3. Library preparation and sequencing

When the samples presented more than 0.05 pmol/ul of DNA (CSB01 and VILA20), Whole Genome Sequencing (WGS) was conducted. Double-strand libraries were made following the Blunt-End-Single-Tube (BEST) protocol (Carøe et al., 2018), with or without partial-UDG treatment (Rohland et al., 2015), as described in Protocol II or III from Cuesta-Aguirre et al. (2025) (see Supplementary Material, Table S2) and sequenced using Illumina technology.

For the rest of the samples, where WGS was not feasible, NGS libraries targeting the non-coding region (NCR) of human mtDNA were prepared using the PowerSeq CRM Nested System Custom kit (Promega) following the M3 method described by Vinueza-Espinosa et al. (2023). These libraries were sequenced on an Illumina MiSeq instrument using standard flow cells and the MiSeq Reagent Kit v2 Nano.

2.4. Bioinformatic analysis and haplogroup determination

For WGS nf-core/eager pipeline (v2.4.6) (Ewels et al., 2020; Fellows Yates et al., 2021) was used for pre-processing the reads and aligning them to the Human Reference Genome GRCh37 (hg19, GenBank accession GCA_000001405.1). Reads mapping to mitochondrial DNA were subsequently realigned to the Cambridge Reference Sequence (rCRS, GenBank accession NC_012920.1) (Andrews et al., 1999) using CircularMapper (v1.0) (Peltzer et al., 2016). Haplotypes were called using freebayes (v1.3.6) (Garrison & Marth, 2012) and processed with vcflib (v1.0.3) (Garrison et al., 2022). The C-to-T misincorporation rate in the 1st base of the 5' end was analyzed to discard contamination (a threshold higher than 10% was considered as indicative of no contamination).

For samples in which only the NCR was sequenced, the pipeline developed by Cuesta-Aguirre et al. (2024) was applied following their recommendations. Samples with a mean mapping quality lower than 25, less than 10% of damaged reads, or mixed bases were discarded, and neither their haplotype nor haplogroup were estimated.

Visual inspection of all samples was conducted using the Integrative Genome Viewer (IGV, v2.16.1) (Robinson et al., 2011) to verify the called positions. In the low-coverage regions, the damage, mutational, and heteroplasmic profiles (Cuesta-Aguirre et al., 2025; Li et al., 2015; Soares et al., 2009) were examined to improve the credibility of call variants. Finally, haplogroups were assigned using Haplogrep 3 (v3.2.1) (Schönherr et al., 2023), with the PhyloTree 17 - Forensic Update 1.2 (Dür et al., 2021; Weissensteiner et al., 2016).

2.5. Haplogroup Diversity and Genetic Structure

Haplogroups were categorized according to frequency and phylogenetic significance (Supplementary Material, Table S1). Individuals were grouped according to cultural and chronological criteria (Cultural-Chronological groups) and by Iberian tribes (Supplementary Material, Table S1). Based on these groupings, haplogroup frequency tables were generated.

To investigate associations between mitochondrial haplogroups and different cultural-chronological contexts or tribes, a Correspondence Analysis (CA) was carried out. This multivariate technique enables the visualization of relationships in categorical data by reducing its dimensionality while retaining the main patterns of variation. The analysis was performed in R v4.4.2 (R Core Team, 2024) using the CA() function from the FactoMineR package (Lê et al., 2008), applied to the contingency table with the frequencies of mtDNA haplogroups across the defined cultural-chronological groups and Iberian tribes.

Nei's gene diversity (Nei, 1987) was estimated for each cultural-chronological group and for each tribe with more than 10 individuals using Arlequin v3.5.2.2 (Excoffier & Lischer, 2010). Moreover, genetic structure analyses were also conducted with Arlequin, incorporating pairwise population comparisons using F_{ST} and an exact test of population differentiation based on haplogroup frequencies (Goudet et al., 1996; Raymond & Rousset, 1995). The significance level was considered 0.05 in all the analyses.

2.6. Distribution of specific haplogroups

Haplogroups present in Iberian individuals were investigated to determine their occurrence and geographical distribution. Samples belonging to haplogroups reported in Iberian individuals were compiled from the AADR database (Mallick et al., 2024) and several recent paleogenomic studies focusing on the Iberian Peninsula (Carrión et al., 2024; Papac et al., 2023; Roca-Rada et al., 2024). Only samples marked as "PASS" in the AADR database and not indicated as possibly contaminated were used. Samples with the same haplogroup and phylogenetically close haplogroups were compiled. For haplogroups that could be precisely classified and that were not exclusively from the Iberian Peninsula, heatmaps were made with QGIS v3.30 (QGIS Development Team, 2025). For this, only unrelated samples older than the Iberians were used to determine if the lineage was already present on the Iberian Peninsula before the Iron Age. Moreover, maternal relatives were removed from the analysis to avoid frequency inflation due to kinship. For cases when the haplogroup was found only in the Iberian Peninsula or when it was rare in ancient populations, the heatmaps from the present-day population were extracted from EMPOP v5/R12 (Huber et al., 2018).

3. RESULTS

The Supplementary Material (Table S2) presents information about each library for the samples analyzed in the present study. All 38 NCR-mtDNA libraries, obtained from 33 individuals (two libraries were made for some of the individuals), presented a percentage of damage over 10%. However, thirteen libraries belonging to nine individuals were discarded due to mixed bases on their profiles. Moreover, three additional libraries were removed from downstream analysis due to low mean mapping quality (lower than 25). All 4 WGS BEST libraries presented a level of damage over 10% and were retained for further analysis.

A total of 21 new mtDNA sequences were generated, 20 from the Iberian period (Table 1), one of them a reanalyzed individual from Sandoval-Ávila et al. (2025), and 1 from the Roman period. The Supplementary Material (Table S3) provides more information about haplotypes and haplogroups for present-study individuals and those from Sandoval-Ávila et al. (2025).

Additionally, mtDNA information from 41 unrelated Iberian individuals (Figure 1 and Table 1) and 172 unrelated individuals from the Neolithic to Visigoth era, all from the Northeast of the Iberian Peninsula, was used in further analysis.

3.1. Haplogroup Diversity and Genetic Structure

Mitochondrial haplogroups were categorized into 17 groups based on frequency and phylogenetic relevance (Supplementary Table S1).

Table 2 presents the frequency of haplogroups for each cultural-chronological group. The most common lineages across all periods are H and U, although their frequencies fluctuate over time. Lineages of H are present in 22.73% of Neolithic-Bronze Age individuals and increase throughout the Iron Age, reaching their highest frequency among the Late Iberians (72.73%). The U lineage (excluding U6), which encompasses several haplogroups, including U5a, U5b, and K, is particularly frequent in the Neolithic-Bronze Age (43.18%) and remains present in subsequent periods. The U5b subclade is more common in earlier populations, while U5a increases during the Iron Age. An elevated frequency of HV0 lineages is observed among the Early Iberians and Iberians, a pattern also reflected in the Greek-Hellenistic group. Values of Nei Gene diversity can be considered high and were similar between populations from different periods, overlapping in all cases.

The Correspondence Analysis (Figure 2) revealed that the first dimension (31.94% explained) separated prehistoric populations chronologically, while dimension 2 (23.25% explained) separated all populations chronologically, except for the Greek-Hellenistic population, which presented a wide chronological range. Along the first dimension, we observed the decrease of subhaplogroup U5b and haplogroup X, and an increase in H3, HV0, and African haplogroups (L, M1, and U6). In the second dimension, the frequency of X and U5b haplogroups decreased and increased for haplogroups T, J2, U5a, H5, V, and Others (R0 group, C4, and W). Although some chronological structuring can be deduced, globally, a small degree of differentiation is evidenced, with only pairwise F_{ST} between Neolithic-Bronze Age compared to Greek-Hellenistic being significant (F_{ST}=0.05185; p-value=0.02734+/-0.0051), and the exact test of population differentiation not showing significant differences.

Table 3 presents the haplogroup frequencies for the 61 Iberian individuals analyzed, including Early Iberians, Iberians, and Late Iberians, and the frequency for each Iberian tribe. Among the Iberian individuals, the most frequent haplogroup was H, accounting for 39.34% of the total. This was followed by haplogroup U (excluding U6) lineages, with a frequency of 22.95%. Other haplogroups were present at lower frequencies, such as J, HV0, and T. N1 and African lineages, specifically M1, were identified in two individuals. Finally, haplogroup X appeared in only one Iberian (Table 3).

Table 2. Haplogroup frequency (%) for all the cultural-chronological groups. The total sample size is indicated in parentheses next to the population's name. Nei's gene diversity is also indicated for each group.

	Neolithic-Bronze Age (5500-1400 BCE) (N=44)	Early Iberians (775-450 BCE) (N=28)	Iberians (450-200 BCE) (N=22)	Late Iberians (200 BCE – 50 CE) (N=11)	Greek-Hellenistic (750-50 BCE) (N=13)	Romans (50 BCE – 700 CE) (N=37)	Visigoths (200-800 CE) (N=58)
H*	22.73	35.71	27.27	72.73	61.54	29.73	34.48
H1	13.64	7.14	4.55	18.18	7.69	2.70	12.07
H3	0.00	3.57	4.55	27.27	15.38	5.41	1.72
H5	2.27	0.00	0.00	9.09	0.00	2.70	5.17
Others H	6.82	25.00	18.18	18.18	38.46	18.92	15.52
V	2.27	0.00	0.00	0.00	0.00	5.41	5.17
HV0	2.27	10.71	13.64	0.00	15.38	2.70	3.45
J1	13.64	7.14	9.09	0.00	0.00	8.11	6.90
J2	2.27	3.57	9.09	0.00	0.00	5.41	5.17
T	4.55	10.71	9.09	0.00	0.00	10.81	6.90
L/M1/U6	0.00	0.00	4.55	9.09	0.00	5.41	0.00
N1	2.27	0.00	9.09	0.00	0.00	2.70	1.72
U*	43.18	32.14	13.64	18.18	23.08	29.73	27.59
U5a	2.27	3.57	4.55	0.00	0.00	5.41	6.90
U5b	13.64	10.71	0.00	0.00	7.69	0.00	3.45
Others U	9.09	0.00	0.00	9.09	7.69	10.81	3.45
K	18.18	17.86	9.09	9.09	7.69	13.51	13.79
X	6.82	0.00	4.55	0.00	0.00	0.00	0.00
Others (R0, C4, W)	0.00	0.00	0.00	0.00	0.00	0.00	8.62
Nei gene diversity	0.9091 +/- 0.0186	0.8889 +/ 0.0328	0.9394 +/- 0.0277	0.9091 +/- 0.0656	0.8462 +/- 0.0854	0.9234 +/- 0.0209	0.9244 +/- 0.0134

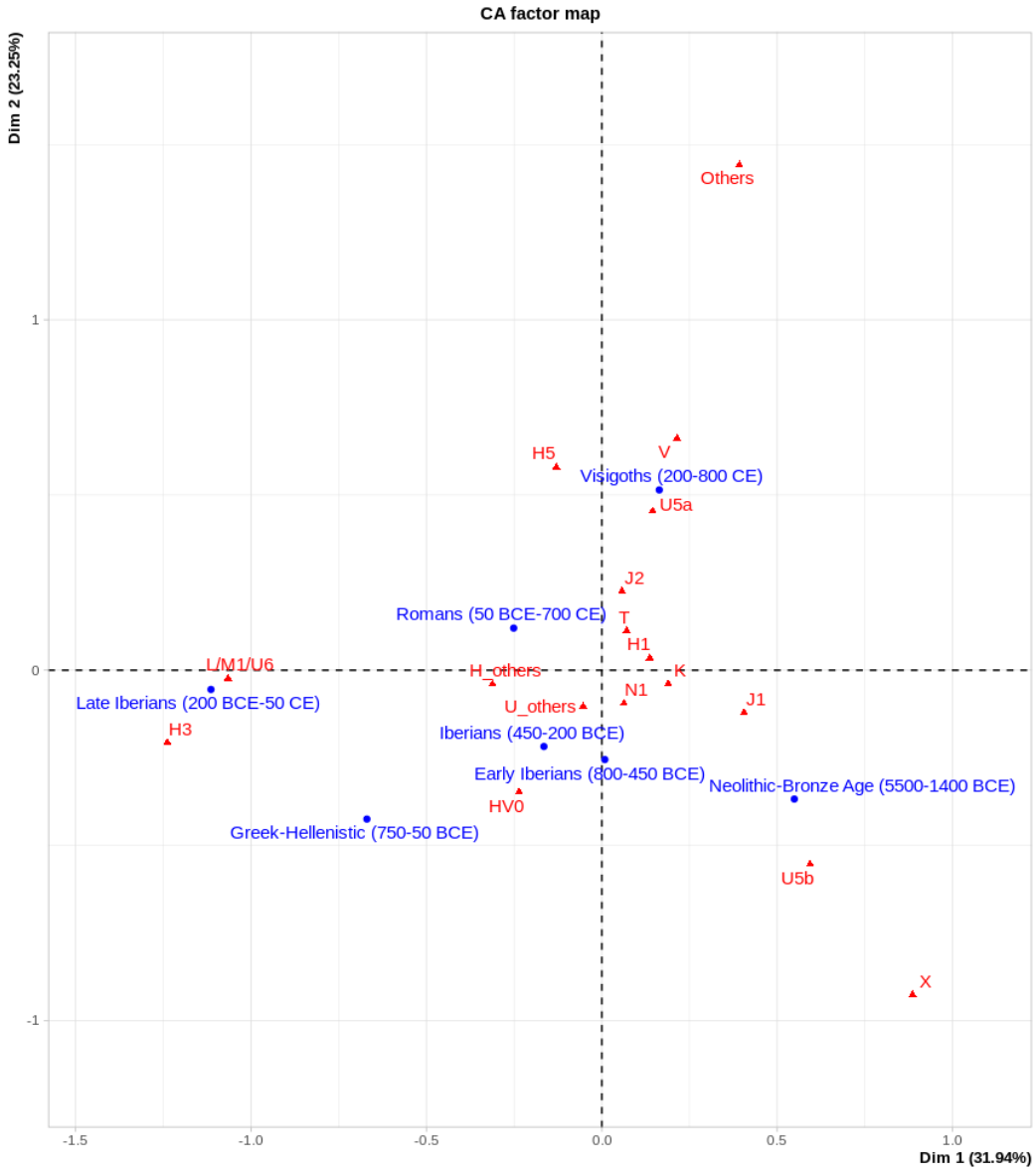


Figure 2. *Correspondence Analysis (CA) based on mitochondrial haplogroup frequencies across cultural-chronological groups. Blue labels represent population groups, while red labels correspond to mitochondrial haplogroups. Dimensions 1 and 2 explain 31.94% and 22.35% of the total variance, respectively.*

Figure 3 displays the results of the Correspondence Analysis based on tribes. The Ilergetae are notable for carrying U5 sublineages, which are absent in the other tribes. The Cessetani exhibit an elevated frequency of haplogroup HV0 (27.27%) compared to the other groups with more than 10 individuals. Among the Indiketes, haplogroups T and X are more frequent than in other tribes. The Ausetani showed an increased prevalence of haplogroup H (72.73%), particularly subclades H1 (18.18%) and H3 (27.27%). Despite these variations in haplogroup composition, no significant genetic differentiation was detected. Pairwise F_{ST} values were not significant, and the exact test of population differentiation did not reveal

statistically significant differences. Nei's gene diversity values are high, comparable across tribes, and overlap in all cases.

Table 3. Haplogroup frequencies across Iberian tribal groups and the whole culture. Sample sizes are shown in parentheses next to each group's name. Tribes with fewer than 10 individuals are shown in *italics*. For each tribe, the haplogroup frequency that most differs from the other groups is highlighted in grey. Nei's gene diversity is reported for groups with more than 10 individuals.

	Iberians (N=61)	**Ilergetae (N=23)**	**Cessetani (N=11)**	***Layetani (N=4)***	**Indiketes (N=10)**	***Bergistani (N=2)***	**Ausetani (N=11)**
H*	39.34	34.78	18.18	*50.00*	30.00	*50.00*	72.73
H1	8.20	4.35	0.00	*25.00*	10.00	*0.00*	18.18
H3	8.20	4.35	9.09	*0.00*	0.00	*0.00*	27.27
H5	1.64	0.00	0.00	*0.00*	0.00	*0.00*	9.09
Others	21.31	26.09	9.09	*25.00*	20.00	*50.00*	18.18
HV0	9.84	4.35	27.27	*25.00*	0.00	*50.00*	0.00
J1	6.56	8.70	9.09	*0.00*	10.00	*0.00*	0.00
J2	4.92	0.00	9.09	*25.00*	10.00	*0.00*	0.00
T	8.20	13.04	0.00	*0.00*	20.00	*0.00*	0.00
L/M1/U6	3.28	0.00	9.09	*0.00*	0.00	*0.00*	9.09
N1	3.28	0.00	9.09	*0.00*	10.00	*0.00*	0.00
U*	22.95	39.13	18.18	*0.00*	10.00	*0.00*	18.18
U5a	3.28	8.70	0.00	*0.00*	0.00	*0.00*	0.00
U5b	4.92	13.04	0.00	*0.00*	0.00	*0.00*	0.00
Others	1.64	0.00	0.00	*0.00*	0.00	*0.00*	9.09
K	13.11	17.39	18.18	*0.00*	10.00	*0.00*	9.09
X	1.64	0.00	0.00	*0.00*	10.00	*0.00*	0.00
Nei Gene Diversity	0.9093 +/- 0.0176	0.8854 +/- 0.0379	0.9273 +/- 0.0665	*NA*	0.9556 +/- 0.0594	*NA*	0.9091 +/- 0.0656

3.2. Distribution of Iberian haplogroups before and during the Iron Age

Most haplogroups found in Iberians were already present in the Iberian Peninsula and Europe before the Iron Age. However, some haplogroups were either rare or absent in the Iberian Peninsula during earlier periods, while others were exclusive to the region before the Iron Age (Table 4).

For the H lineages observed in the Iberians, haplogroups H1, H1t, H3, H3+152, H4a1, and H5 (Supplementary Material, Tables S4-S8) were already present in the Iberian Peninsula before the Iron Age. Interestingly, haplogroup H1t was exclusively found in the Iberian

Peninsula until 50 CE (Supplementary Material, Table S5). Haplogroups H3am, H4a1a5, H6a1b2, and H33c (Figure 4) had not been previously detected in the Iberian Peninsula, but their parent lineages (H3, H4a1a, H6ba1b, and H33) were already present during the Iron Age (Supplementary Material, Tables S6, S7, S9, and S10). On the other hand, haplogroup H1at1 has only been found in Iron Age individuals from Italy and England, with its parent haplogroup only found in the Bronze Age in Croatia and Montenegro (Figure 5 and Supplementary, Table S11). Additionally, haplogroup H1n2* had not been previously detected, though its parent lineage, H1n+146, was found in an Iron Age individual from Macedonia (Supplementary Material, Table S12).



Figure 3. *Correspondence Analysis (CA) based on mitochondrial haplogroup frequencies across Iberian tribes. Blue labels represent tribes, while red labels correspond to mitochondrial haplogroups. Dimensions 1 and 2 explain 44.13% and 32.86% of the total variance, respectively.*

Table 4. Presence of detected haplogroups in the Iberian Peninsula during or before the Iron Age for each tribe.

Tribe	Present in the Iberian Peninsula by the Iron Age	Absent in the Iberian Peninsula, but ancestral and/or sublineages present by the Iron Age	Absent in the Iberian Peninsula by the Iron Age	Never reported by the Iron Age
Illergetae	H1, HV0b, J1c, T2b3+151, T2c1d, K1a+195, U5b1b1g, U5b1f1a, K1a3a, K1a13	H, H3am, H4, H4a1a5, J1c3e, U5a, U5a1c1, K		
Cessetani	H3, HV0, HV0+195, J1c1, J2a1a1, K1a+195	H	K1a12a, N1a1a1a3, M1b	
Layetani	H1t, HV0d	H, J2a1a1c		
Indiketes	H1, H4a1, J1c, J2b1a, K1a, I1a1, X2b	H, T2		T1a1'3
Bergistani	HV0	H6a1b2		
Ausetani	H3, H3+152, H5, K1b1a	H, H33c	H1at1, U4d2	H1n2*, M1a3a

All HV0 lineages (HV0, HV0+195, HV0b, and HV0d) and the X2b haplogroup were present in the Iberian Peninsula before the Iron Age (Supplementary Material, Tables S13 and S14).

Regarding the JT lineages, haplogroups J1c, J1c1, J2a1a1, J2b1a, T2b3+151, and T2c1d were already present in the Iberian Peninsula before the Iron Age (Supplementary Material, Tables S15-S19). Haplogroups J1c3e and J2a1a1c had not been previously identified in the Iberian Peninsula, but their parent haplogroups (J1c3 and J2a1a1) were already present (Supplementary Material, Tables S15 and S16). Haplogroup T2 had not been detected, but the T2+16189 sublineage was found in a Bronze Age sample from the Iberian Peninsula (Supplementary Material, Table S20). The T1a1'3 lineage had never been observed in the Iberian Peninsula, although the T1a haplogroup was detected in France by the Bronze Age (Figure 5 and Supplementary Material, Table S21).

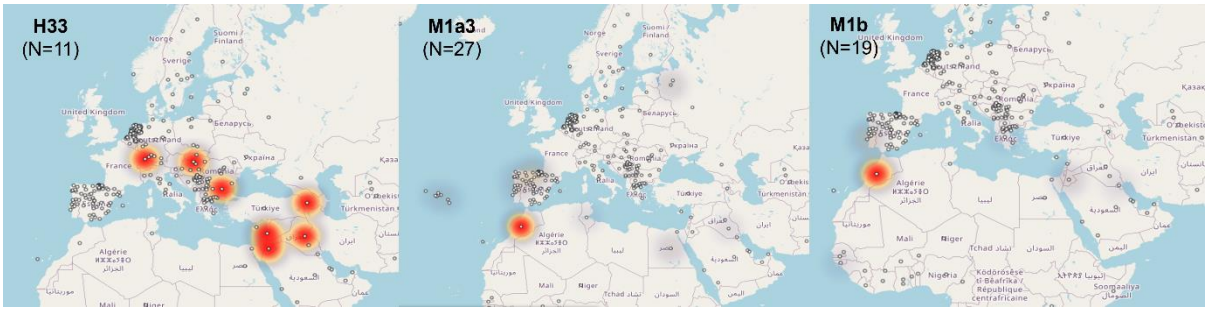


Figure 4. *Present-day distribution of haplogroups H33, M1a, and M1b based on data from EMPOP v5/R12 (N= 48,572; Huber et al., 2018). The map is centered on Europe and the Mediterranean regions. The number of individuals is indicated in parentheses, although a few individuals with haplogroups H33 and M1a3 are also found in America.*

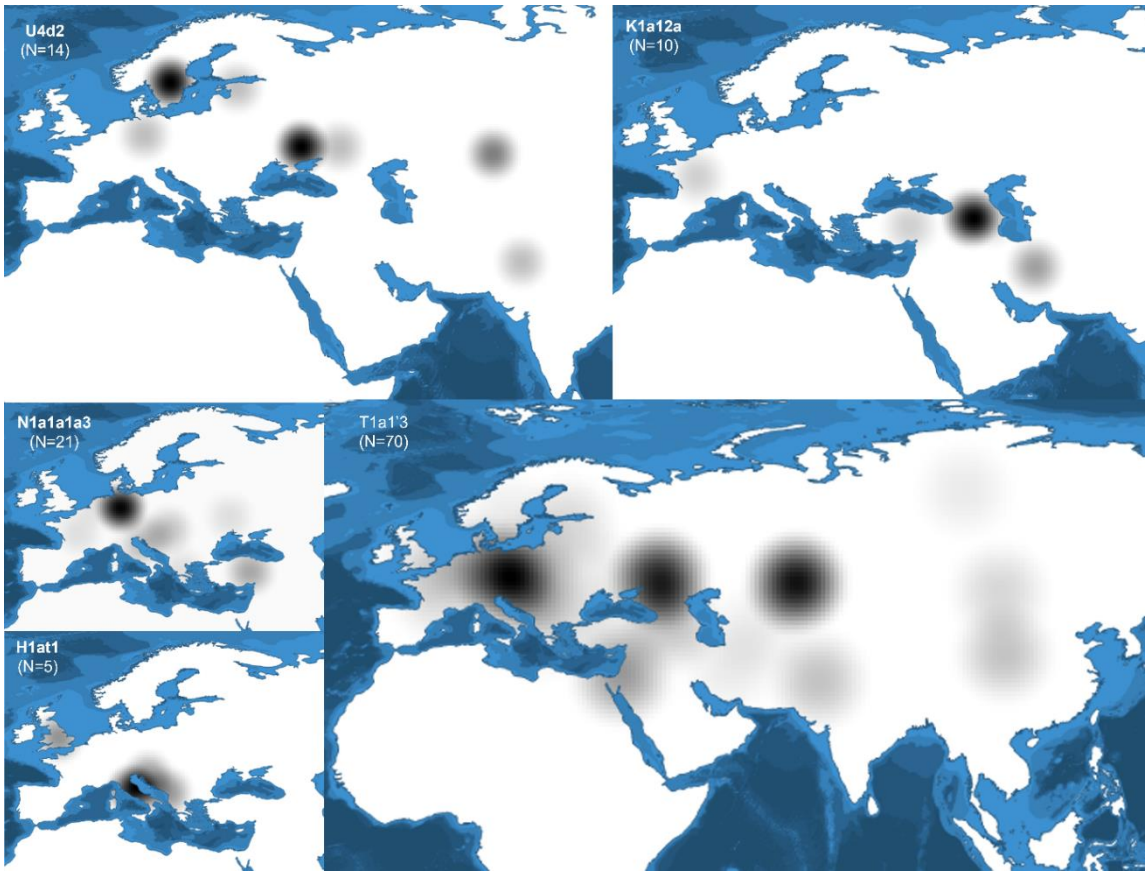


Figure 5. *Heatmaps showing the distribution of ancient samples, predating the Iron Age, carrying haplogroups U4d2 (including U4d), K1a12a (including K1a12), N1a1a1a3 (including N1a1a1a), H1at1 (including H1at), and T1a1'3 (including both its sublineages and ancestral lineage, T1a). Ancient DNA data were obtained from the Ancient Allen database, Carrión et al. (2024), Papac et al. (2023), and Roca-Rada et al. (2024). Maps were generated using QGIS v3.30. Details on the samples used are available in the Supplementary Material (Tables S11, S21, S29, S30, and S32), where the employed samples are highlighted in italics.*

For the U group (including K), haplogroups K1a, K1a+195, K1a3a, K1a13, and K1b1a were present in the Iberian Peninsula by the Iron Age (Supplementary Material, Tables S22-24). U5a1c1 had not been detected in the region, but its ancestral haplogroup (U5a1c) and subhaplogroup (U5a1c1a) were identified in Chalcolithic samples from the southeastern Iberian Peninsula (Supplementary Material, Table S25). Haplogroups U5b1f1a and U5b1b1g were found exclusively in the Iberian Peninsula, in very low frequencies (Supplementary Material, Tables S26 and S27). While K, U5a, and U5b2 haplogroups have not been detected in the Iberian Peninsula, several of their sublineages have been found in prehistoric individuals from the Iberian Peninsula (Supplementary Material, Tables S25 and S28). U4d2, a haplogroup with no known presence in Western Europe, was present in Eastern and Northern Europe during prehistory (Figure 5 and Supplementary Material, Table S29). Similarly, K1a12 was previously found only in Iran, Armenia, and Turkey since the Neolithic, though a sample was also detected in France during the early Neolithic (Figure 5 and Supplementary Material, Table S30).

Despite its higher frequency in Central Europe, haplogroup I1a1 was already present in the Iberian Peninsula by the Bronze Age (Supplementary Material, Table S31). N1a1a1a3, while also common in Central Europe, has not been detected in the Iberian Peninsula outside of our sample (Figure 4 and Supplementary Material, Table S32).

Two M lineages were identified in Iberians: M1a3a and M1b. M1a3a was found in a Roman sample from the Iberian Peninsula but had not been detected before (Figure 4 and Supplementary Material, Table S33). M1b lineages, however, were found in Morocco (Northwest Africa) before the Neolithic and in an Islamic sample from the Iberian Peninsula (Figure 4 and Supplementary Material, Table S34).

4. DISCUSSION

The analyses of haplogroup diversity and genetic structure suggested a gradual shift in maternal lineages throughout pre-Roman times, consistent with the smooth transition between the Bronze and Iron Ages proposed for the northeastern Iberian Peninsula (Cuesta-Aguirre et al., 2025; Olalde et al., 2019). This trend continued until the Romans arrived, introducing more marked genetic changes. The Correspondence Analysis supports this interpretation, with the first dimension reflecting a chronological gradient in haplogroup composition, showing a decline in U5b and X (typically associated with earlier periods) and

an increase in H3, HV, and African lineages (such as L, M1, and U6), which become more prominent in later phases. Despite these gradual changes, population differentiation was generally low, aligning with nuclear DNA results for Iberians (Cuesta-Aguirre et al., 2025), and overall genetic diversity (Nei's diversity) remained relatively stable across periods.

Our results revealed a high frequency of haplogroups H and U (including K but excluding U6), alongside moderate representation of J, T, and HV0, and low frequencies of M1, N1, and X. These findings broadly align with those of Sampietro et al. (2005), who examined 17 individuals from the Northeast of the Iberian Peninsula. In their study, H was found in 52.9% of individuals, U (including K) in 23.5%, and J, T, and Pre-HV in smaller proportions (11.8%, 5.9%, and 5.9%, respectively). It is important to note that their samples consisted exclusively of adult males, analyzed using cloning technology on fragmentary cranial remains not associated with funerary contexts. In contrast, Gamba et al. (2008) reported different haplogroup frequencies among Iberians from the central-eastern Iberian Peninsula (modern-day Valencian Community), with H (36.4%), V (27.3%), and W (18.2%) being the most frequent, alongside additional instances of T1 and K. We did not detect any individuals belonging to haplogroup V in our dataset. However, the study of Gamba et al. (2008) is before the redefinition of some haplogroup nomenclature, and the mutation at position 16298C previously used to define V is now considered characteristic of HV0 (van Oven & Kayser, 2009). The HV0 frequency we observed in the Cessetani tribe (St. Miquel d'Olérdola), geographically the closest tribe to the area studied by Gamba et al. (2008), is comparable to their reported frequency of haplogroup V.

Although no statistically significant genetic differentiation was detected between the tribes in our study, differences in haplogroup frequencies suggest some degree of biological distinction, and the absence of significant results could be related to the still limited sample size available for each tribe. This observation is consistent with archaeological interpretations of tribal diversity in the region (Aranegui, 2012; González Reyero, 2012). Moreover, values of mtDNA diversity are slightly higher, supporting a patrilocal mating system during the Iron Age, as suggested for Bronze Age communities of the same region (Simón et al., 2011), and small distance female migration that could have prevented high differentiation among tribes.

Most of the haplogroups identified in Ilergetae individuals from the fortress of Vilars were already present in the Iberian Peninsula before the onset of the Iron Age. These include

H1, H2a2a1, HV0b, J1c, T2b3+151, T2c1d, U5b1b1g, U5b1f1a, K1a+195, K1a3a, and K1a13. In some cases, the lineages observed in Iberian individuals, such as H3am, H4a1a5, J1c3e, U5a1c1, and U5b1b1g1, had not previously been detected in the region; however, their ancestral haplogroups were already present before the Iron Age. These local haplogroups and the mtDNA slightly higher diversity support a scenario of genetic continuity from the Bronze Age (Cuesta-Aguirre et al., 2025; Olalde et al., 2019), suggesting that female mobility occurred largely within the Iberian Peninsula, as happened during the Bronze Age (Simón et al., 2011; Villalba-Mouco et al., 2022).

Interestingly, an ancestral lineage of haplogroup H4a1a5 (VILA12 775-675 BCE) was found in the Indiketes individual from Mas Castellar (I19722, 413-382 calBCE). The other site analyzed within the Indiketes territory was the city of Ullastret, which was composed of two urban centers, both of which reflected Mediterranean architectural influences (López i Melcion et al., 1989; Martín i Ortega, 1985; Martín i Ortega et al., 2016). In Illa d'en Reixac, the haplogroups of three individuals could be determined: H2a2a1, I1a1, and T1a1'3. H2a1a1 has already been detected in the Iberian Peninsula with a much higher frequency than I1a1. The haplogroup I originated in the Near East 19,000-15,000 years ago as part of the N1a1b lineage, diverging into I1 — a haplogroup linked to postglacial re-peopling, including Indo-European associated groups — around 16,000 years ago (Olivieri et al., 2013). I1a1, which is restricted to Europe, emerged 7,000 years ago and is associated with the Neolithic agricultural expansion (Olivieri et al., 2013). Notably, haplogroup I1 has been detected in an Iberian newborn from the central-eastern of the Iberian Peninsula (Olalde et al., 2019).

Haplogroup T is estimated to have originated approximately 28,000 years ago in the Near East, followed by a significant post-glacial expansion (Pala et al., 2012), like J and I (Olivieri et al., 2013). T1a1, the most common lineage within T1a* (comprising about 80% of samples), is linked to early agricultural communities and emerged around 15,500 years ago. Its subclade T1a1a1, dating back roughly 6,800 years, represents approximately 70% of T1a1 samples and is prevalent in regions with high frequencies of Y-haplogroup R1a, suggesting a possible connection to Indo-European migrations during the Bronze Age (Pala et al., 2012). Determining whether ILLA04 belongs to a sublineage associated with Bronze Age migrations is challenging, as it requires unavailable coding region mutations for our sample. However, analyzing the compiled dataset of T1a lineages may offer insights. T1a has been found as early as the Neolithic in Southeastern Europe (Bulgaria, 5723-5571 calBCE) (Lazaridis et al., 2022), while the first appearance of the T1a1 lineage was in Chalcolithic Croatia (4300-3900

BCE) (Lazaridis et al., 2022). During the Bronze Age (3500-1600 BCE), T1a1 frequency was elevated in Central Europe, the Pontiac Steppe, and Central Asia. The first appearance in Western Europe was in France (1891-1743 calBCE) (Patterson et al., 2022), after which a few samples appeared in England and France during the Bronze Age. The relatively late appearance of these lineages in Western Europe raises questions about their origin and dispersal. It is possible that T1a1'3 may have arrived in Western Europe during Neolithic expansion but remained at low frequencies, making it difficult to detect in previous archaeological samples. However, it is also possible that it was introduced in Europe during the Neolithic, and later migrations from the Steppe during the Bronze Age may have increased the frequency of T1a1'3 sublineages across Europe, ultimately leading to their spread into the Iberian Peninsula. The absence of earlier detections in the Iberian Peninsula could be due to several factors: limited sampling, localized founder effects, or the influx of the lineage during the Iron Age from regions with stronger Steppe influences. It's also important to consider that T1a has a low overall frequency in present-day samples, rising to about 5% in the South Caucasus, 6% in northeastern Iran, 8% in Tunisia, and 9% in Romania (Pala et al., 2012). T1a1a1 is present-day shared between geographically distant individuals from Scandinavia, the Baltic, the North Caucasus, Anatolia, and Morocco (Pala et al., 2012). This wide European and Mediterranean distribution in the present day could be explained by later migrations, such as Roman, Germanic, or Viking expansions.

In the other settlement of Ullastret, Puig de Sant Andreu, several haplogroups commonly found in the Iberian Peninsula were identified, including H1, J1c, J2b1a, K1a, and X2b. One exception is haplogroup T2b4+152, detected in the individual PDSA12, which is rare in prehistoric contexts. Its presence in a Neolithic individual from France (Rivollat et al., 2020) suggests it may have arrived following the same Neolithic migration routes as haplogroups like T1a1'3 or I1a1. Overall, the haplogroup profiles identified at Ullastret (including both Puig de Sant Andreu and Illa d'en Reixac) suggest that the female individuals were local. Among the four individuals with available nuclear data (I3323, I3324, I3326, and I3327) (Olalde et al., 2019), only I3327 showed minor traces of Mediterranean ancestry (Cuesta-Aguirre et al., 2025).

For the Layetani tribe, at Can Roqueta-Can Revella, haplogroups H, H1t, HV0d, and J2a1a1c were identified in Iberian individuals dated between 786-416 calBCE. Haplogroup H, due to its wide distribution across Western Eurasia, offers limited information on maternal origin (Hernández et al., 2017; Roostalu et al., 2007). In contrast, HV0d and H1t appear to be

local, with H1t detected exclusively within the Iberian Peninsula until the Iron Age. Notably, this haplogroup was detected in an Iberian newborn from the central-eastern coast of the Iberian Peninsula (Olalde et al., 2019). Although J2a1a1c has not previously been reported in ancient samples, its ancestral haplogroup (J2a1a1) was present at high frequency in the Iberian Peninsula since the Neolithic. Taken together, these findings suggest a predominantly local maternal ancestry, consistent with the local genetic affinities inferred from nuclear genome data (Cuesta-Aguirre et al., 2025; Olalde et al., 2019).

In the Cessetani site of Font de la Canya (700-500 BCE), the female individual analyzed was carrying haplogroup K1a12a. This lineage had not previously been reported in the Iberian Peninsula, but has been observed in regions such as Turkey, Iran, Azerbaijan, and Armenia from the Neolithic through the Iron Age (Kılınç et al., 2016; Lazaridis et al., 2016, 2022; Skourtanioti et al., 2020). Its ancestral lineage was also identified in a Neolithic individual from France, although it was later classified as K1a2 when capture-based methods were applied (Brunel et al., 2020). While nuclear data placed this individual within the genetic variability of other Iron Age Iberians in Principal Component Analyses (Cuesta-Aguirre et al., 2025; Olalde et al., 2019), and showed similar admixture proportions to other contemporary samples (Olalde et al., 2019), her ancestry could not be successfully modeled using the same qpAdm frameworks applied to other Iberian individuals (Cuesta-Aguirre et al., 2025). Several factors complicate the interpretation of her origins. First, the published study provides limited information about the context and condition of the remains analyzed (Olalde et al., 2019). Second, given that cremation was the dominant funerary practice among the Iberians, the preservation of her remains does not align with local traditions, raising questions about her cultural affiliation. Third, Font de la Canya maintained early and sustained contact with Phoenician groups since its foundation (Olalde et al., 2019), making the presence of non-local individuals plausible. These lines of evidence suggest that this individual may have been non-local or had mixed ancestry. In any case, the presence of K1a12a may indicate the introduction of this haplogroup into the Iberian Peninsula via Phoenician female mobility during the Late Bronze Age or Early Iron Age.

In contrast, other sites within the Cessetani territory, such as Horta d'en Grimau (725-394 BCE) and Mas d'en Boixos (550-350 BCE), revealed only local maternal lineages (e.g., H, H3, HV0+195, and J1c1), and nuclear DNA data supported a local genetic profile (Cuesta-Aguirre et al., 2025; Olalde et al., 2019). However, the small number of individuals analyzed may influence the apparent absence of non-local ancestry at these sites.

In the most recent Cessetani site, Sant Miquel d'Olèrdola (350-200 BCE), four haplogroups, namely HV0, HV0+195, J2a1a1, and K1a+195, out of the six determined haplogroups, were most probably local. However, the other two could have a non-local origin, since they have never been detected in the Iberian Peninsula. OLE04 presented haplogroup N1a1a1a3. Limited information is available regarding N1a lineages, a relatively rare group in present-day Europeans (approximately 0.2%) that is believed to have originated in the Near East, potentially the Arabian Peninsula, between 12,000 and 32,000 years ago (Palanichamy et al., 2010). During the Neolithic, N1a presented a low frequency (3.7%) compared to Central Europe (9.4%) (Szécsényi-Nagy et al., 2017). Today, N1a1a1 reaches its highest frequency in Eastern Europe and Central Asia, decreasing towards Central and Northern Europe, Anatolia, and the Caucasus (Palanichamy et al., 2010). Furthermore, the subclade N1a1a1a, with a proposed expansion from the Near East to Europe through Eastern Europe around 9,000-11,000 years ago, has a restricted distribution in Central Asia (Kazakhstan, the Altai region, and the Buryat Republic) and the Pontiac Steppe (Palanichamy et al., 2010). The ancient dataset compiled for N1a1a1a and N1a1a1a3 supports this origin and expansion theory. Early samples are found around 6000 BCE in Anatolia (Lazaridis et al., 2022), followed by a presence in Germany between 5500-4700 BCE (Childebayeva et al., 2022; Lipson et al., 2017; Rivollat et al., 2020). From the Middle Neolithic to the Chalcolithic, this haplogroup was detected in a wide range of European regions, such as France (Rivollat et al., 2020), Italy (Antonio et al., 2019), Croatia (Freilich et al., 2021; Novak et al., 2021), Hungary (Gamba et al., 2014; Mathieson et al., 2015), Ukraine (Gelabert et al., 2022), and Turkey (Skourtanioti et al., 2020). Notably, this haplogroup appears absent from the archaeological record between 3000 and 1200 BCE until its detection in a Greek sample from the Mycenaean culture (Lazaridis et al., 2022). The haplogroup disappeared again until the Middle Iron Age when it appeared in our Iberian sample OLE04 (350-200 BCE) and a Netherlandish individual (356-57 calBCE) (Patterson et al., 2022). Gaps in the archaeological record may suggest the low frequency of this haplogroup, while the presence of its ancestral haplogroup (N1a1a1a) in France could indicate the arrival of this lineage in the Iberian Peninsula during the Neolithic.

The other probably non-local haplogroup found in Sant Miquel d'Olèrdola was M1b, which was identified in OLE06, and its origin was previously proposed in Cuesta-Aguirre et al. (2025). This haplogroup, primarily found in North Africa (González et al., 2007; Winters, 2010), was associated with an individual exhibiting high Mediterranean ancestry, potentially linked to North Africa or the eastern Mediterranean. We suggested that a woman with Punic/Carthaginian ancestry may have introduced it to Sant Miquel d'Olèrdola during the Iron

Age. Ancient DNA evidence for M1b was limited to three records: two from Moroccan Paleolithic (van de Loosdrecht et al., 2018) and Early Neolithic (Fregel et al., 2018) samples and one from an Islamic-period individual in the Iberian Peninsula (Olalde et al., 2019). Today, according to EMPOP v5/R12 (Huber et al., 2018) (N=19), its distribution is largely restricted to Morocco. The rarity of this lineage should be noted. However, its presence in an Iron Age individual from the northeastern Iberian Peninsula may reflect previously underappreciated female mobility linked to Mediterranean exchange networks, further highlighting the complex genetic landscape of the region during this period.

Two local haplogroups were detected in Castell de Besora (Bergistani tribe): HV0 and H6a1b2. By visualizing the heatmap, haplogroup H6a1b is a lineage that migrated to Western Europe during the Bronze Age, and, despite presenting low frequency, its ancestral haplogroup H6a1b was present at the Iberian Peninsula before Iberian culture was developed.

Out of the ten different haplogroups detected in the Ausetani site of El Camp de les Lloses (125 BCE – 50 CE), half of them, namely H, H3, H3+152, H5, and K1b1a, were already present in the Iberian Peninsula before the Iron Age started. However, several rare haplogroups were detected, including H33c, in CL11 (125-80 BCE). The H33 lineage was previously identified in an Aegean Hellenistic individual from Empuries (Girona, Northeast Iberian Peninsula) dating to approximately 376-201 BCE (Olalde et al., 2019) and is also present in modern-day populations across Central Europe, Southeast Europe, and Southwest Asia. Today, H33 remains extremely rare, with a broad yet low-frequency distribution, according to EMPOP v5/R12 (Huber et al., 2018) (N=11). Given this information, it is tempting to associate the arrival of this haplogroup with the Aegean populations. This raises an important question: Was the H33 lineage introduced into the Iberian world through trade with the Greeks, or did it arrive later with the Romans? The Roman-influenced Iberian city of El Camp de les Lloses has yielded Greek and Roman coins, underscoring the region's strong Mediterranean connections (Duran i Caixal et al., 2017). Yet, without nuclear DNA data, determining the exact source remains speculative, and even with such data, fully resolving this question may prove impossible. One thing, however, is clear: this newborn's maternal lineage provides undeniable evidence of Mediterranean influence.

Another rare haplogroup found in El Camp de les Lloses was M1a3a, detected in CL2. In contrast to M1b, M1a lineages are typical of Northeast Africa and have been identified in Neolithic individuals from Kenya (Prendergast et al., 2019) and Mediterranean regions, including the Iberian Peninsula (Olalde et al., 2019), Sardinia (Fernandes et al., 2020), and

Egypt (Fernandes et al., 2020). However, all these individuals carried M1a1 lineages. The present-day distribution of M1a3, according to EMPOP v5/R12 (Huber et al., 2018) (N=27), differs from the M1a distribution and closely resembles M1b's distribution, but with a slightly higher frequency. Interestingly, the exact M1a3a haplogroup was later identified by Roca-Rada et al. (2024) in a Roman-era female from the western Iberian Peninsula (ID_25522; 257-531 CE). Nuclear DNA analysis of the CL2 individual did not indicate North African ancestry (Cuesta-Aguirre et al., 2025), suggesting that Haplogroup M1a3a may have entered the Iberian Peninsula long before the foundation of El Camp de les Lloses, remaining at such a low frequency that it had not been detected previously.

The sisters from El Camp de les Lloses (CL6 and CL7, 125-80 BCE) shared the same haplotype, belonging to haplogroup H1at1 (Cuesta-Aguirre et al., 2025). The information about this haplogroup is scarce. It has only been identified in two Iron Age Etruscans from Italy (775-201 calBCE) (Posth et al., 2021) and a Middle Iron Age individual from England (372-197 calBCE) (Patterson et al., 2022), despite its ancestral haplogroup being present in two individuals from Croatia and Montenegro during the Middle Bronze Age (1500-1000 BCE) (Lazaridis et al., 2022). Its low frequency makes it difficult to trace its exact origin, but the most probable source is the Italian Peninsula. Whether it was introduced into the Iberian Peninsula by Etruscans or later by Romans remains uncertain.

Finally, the U4d2 haplogroup was found in CL4 (50 BCE - 50 CE) (Cuesta-Aguirre et al., 2025). U4d lineages had previously been detected in the Pontiac Steppe during the Neolithic and in North and Central Europe during the Bronze Age (Mittnik et al., 2018; Olalde et al., 2018; Skoglund et al., 2014). Since the Pontiac Steppe was the source of Steppe-related expansions into Western Europe during the Bronze Age, these migrations likely introduced the haplogroup at a low frequency, making it undetectable until the Iron Age. Given the high proportion of Italian ancestry in CL4's nuclear DNA, it is plausible that Romans introduced this haplogroup to El Camp de les Lloses. However, the possibility that U4d2 was already present in the Iberian Peninsula since the Bronze Age, remaining undetected, cannot be ruled out.

Although genetic evidence revealed a mixture of Mediterranean and local ancestries in the nuclear DNA of the Ausetani individuals of El Camp de les Lloses (Cuesta-Aguirre et al., 2025), the most probable scenario is that almost all mitochondrial haplogroups detected were already present in the region before the arrival of the Romans. This aligns with the interpretation of the site as a Roman military vicus associated with metallurgical workshops and logistical support for the army (Duran i Caixal et al., 2017; Duran et al., 2015, 2017).

Within the context of El Camp de les Lloses' military function, this pattern supports the hypothesis that while male soldiers or artisans likely came from outside the Iberian Peninsula, perhaps from other parts of the Mediterranean, the women connected to the settlement were likely local inhabitants.

In summary, the gradual shift in haplogroup frequencies observed until the Roman period supports a smooth transition between the Bronze and Iron Ages, highlighting the complex interplay of migration, commercial trade, and cultural interactions that shaped the genetic heritage of the region. While no statistically significant genetic differentiation was detected between Iberian tribes, subtle differences in haplogroup frequencies suggest that these groups maintained a certain degree of autonomy. The mitochondrial DNA analysis of Iberians from the Northeast of the Iberian Peninsula reveals a dynamic genetic landscape influenced by both local continuity and short-distance women migration, probably linked to patrilocal mating systems, and external inputs evidenced by non-local mtDNA lineages. Most of the detected haplogroups, or their ancestral lineages, were already present in the Iberian Peninsula during the Iron Age, suggesting that most Iberian women were of local origin. However, signs of Mediterranean female influence are evident during the Iron Age, pointing to potential gene flow from other regions and the role of women in commercial trade and cultural interaction. The genetic analysis of the Ausetani individuals from El Camp de les Lloses showed a mix of Mediterranean and local ancestries in their nuclear DNA, with almost all mitochondrial haplogroups likely pre-dating the Roman arrival, suggesting they were already present in the region. This supports the idea that while male soldiers or artisans may have come from outside the Iberian Peninsula, the women from this site were likely local inhabitants.

AUTHORS CONTRIBUTIONS

Conceptualization: D.R.C-A., A.M., C.S.

Data analysis: D.R.C-A., C.S.

Funding acquisition: A.M., C.S.

Interpretation: All authors

Lab work: D.R.C-A.

Project administration: A.M., C.S.

Sample provision: C.B.C., M.F.E., A.G.S., G.dP., N.M.C., M.D.C., I.M.S., N.A.

Software: D.R.C-A., C.S.

Supervision: A.M., C.S.

Visualization: D.R.C-A., M.R.C-C., C.S.

Writing - original draft: D.R.C-A., M.R.C-C., C.S.

Writing - review & editing: All authors

DECLARATION OF INTEREST

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI IN SCIENTIFIC WRITING

During the preparation of this work, the authors used ChatGPT and Grammarly in order to improve the structure, clarity, and order of some paragraphs. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

DATA AVAILABILITY

The datasets generated and analyzed during the current study will become publicly available upon publication.

ACKNOWLEDGMENTS

This research was supported by the Ministerio de Ciencia e Innovaciòn (Ref. PGC2018-096666-B-I00), by the Fundación Palarq (PALARQ23_55), and by the Departament de Recerca i Universitats de la Generalitat de Catalunya and the European Social Fund (Fons Social Europeu). D.R.C.-A. is a PhD fellow from the Agència de Gestiò d'Ajuts Universitaris i de Recerca (AGAUR) (2020 FI_B1 00641). GREAB is supported by the Generalitat de Catalunya (2021 SGR 00186).

BIBLIOGRAPHY

Achilli, A., Perego, U. A., Bravi, C. M., Coble, M. D., Kong, Q.-P., Woodward, S. R., Salas, A., Torroni, A., & Bandelt, H.-J. (2008). The Phylogeny of the Four Pan-American MtDNA Haplogroups: Implications for Evolutionary and Disease Studies. *PLoS ONE*, *3*(3), e1764. https://doi.org/10.1371/journal.pone.0001764

Afonso, C., Nociarova, D., Santos, C., Martinez-Labarga, C., Mestres, I., Duran, M., & Malgosa, A. (2019). Sex selection in late Iberian infant burials: Integrating evidence from morphological and genetic data. *American Journal of Human Biology*, *31*(1). https://doi.org/10.1002/ajhb.23204

Almagro, M., Blech, M., Ruia Marta, D., & Arteaga, O. (2001). *Protohistoria de la Península Ibérica*. Ariel.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics*, *23*(2), 147–147. https://doi.org/10.1038/13779

Antonio, M. L., Gao, Z., Moots, H. M., Lucci, M., Candilio, F., Sawyer, S., Oberreiter, V., Calderon, D., Devitofranceschi, K., Aikens, R. C., Aneli, S., Bartoli, F., Bedini, A., Cheronet, O., Cotter, D. J., Fernandes, D. M., Gasperetti, G., Grifoni, R., Guidi, A., La Pastina, F., Loreti, E., Manacorda, D., Matullo, G., Morretta, S., Nava, A., Fiocchi Nicolai, V., Nomi, F., Pavolini, C., Pentiricci, M., Pergola, P., Piranomonte, M., Schmidt, R., Spinola, G., Sperduti, A., Rubini, M., Bondioli, L., Coppa, A., Pinhasi, R., & Pritchard, J. K. (2019). Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science*, *366*(6466), 708–714. https://doi.org/10.1126/science.aay6826

Aranegui, C. (2012). *Los iberos ayer y hoy. Arqueologías y culturas*.

Asensio, D., Morer de Llorens, J., Rigo Jovells, A., & Sanmartí, J. (2001). Les formes d'organització social i econòmica a la cossetània ibèrica noves dades sobrel'evolució i tipologia dels assentaments entre els segles VII-I a.c. *Territori Polític i Territori Rural Durant l'edat Del Ferro a La Mediterrània Occidental : Actes de La Taula Rodona Celebrada a Ullastret Del 25 Al 27 de Maig de 2000*, 253–272.

Bagnall, N. (1999). *The Punic Wars: Rome, Carthage and the Struggle for the Mediterranean* (Pimilico (ed.)).

Belarte, M. C. (2010). Los individuos en el espacio doméstico en la protohistoria de Cataluña [Article]. *Arqueología Espacial*, *28*, 109–134.

Bendala, M. (1998). La ciudad entre los iberos, espacio de poder. *SAGVNTVM Extra*, *1*, 25–34.

Bosch-Gimpera, P. (1938). Superestructures de la història d'Espanya. *Revista de Catalunya*, *84*, 313–322.

Bosch, J., & Santacana, J. (2009). *Blat, metalls, cabdills. Catalunya del neolític a la iberització*.

Brunel, S., Bennett, E. A., Cardin, L., Garraud, D., Barrand Emam, H., Beylier, A., Boulestin, B., Chenal, F., Ciesielski, E., Convertini, F., Dedet, B., Desbrosse-Degobertiere, S.,

Desenne, S., Dubouloz, J., Duday, H., Escalon, G., Fabre, V., Gailledrat, E., Gandelin, M., Gleize, Y., Goepfert, S., Guilaine, J., Hachem, L., Ilett, M., Lambach, F., Maziere, F., Perrin, B., Plouin, S., Pinard, E., Praud, I., Richard, I., Riquier, V., Roure, R., Sendra, B., Thevenet, C., Thiol, S., Vauquelin, E., Vergnaud, L., Grange, T., Geigl, E.-M., & Pruvost, M. (2020). Ancient genomes from present-day France unveil 7,000 years of its demographic history. *Proceedings of the National Academy of Sciences*, *117*(23), 12791–12798. https://doi.org/10.1073/pnas.1918034117

Busquets i Costa, F., Fàbregas i Espadaler, M., Corts, A., & Capdevila, N. (2018). El conjunt monumental del castell de Besora: primeres aproximaciones. *AUSA*, *XXVIII*(179), 111–129.

Carey, B. T., Allfree, J. B., & Cairns, J. (2007). *Hannibal's Last Battle: Zama & The Fall of Carthage* (Pen & Sword Books (ed.)).

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., Wales, N., Sicheritz-Pontén, T., & Gilbert, M. T. P. (2018). Single-tube library preparation for degraded <scp>DNA</scp>. *Methods in Ecology and Evolution*, *9*(2), 410–419. https://doi.org/10.1111/2041-210X.12871

Carrión, P., Olalde, I., Jiménez-Arenas, J. M., Coromina, N., Vivó, D., Vergès, J. M., Costa, A., Botella, D., Bustamante-Álvarez, M., Heras-Mora, J., Ortega-Ruiz, R., Chaves, C., García-Collado, M. I., Quirós-Castillo, J. A., Roig, J., Suárez-Padilla, J., Navarro-Luengo, I., Cuadrado, M. Á., Aguilera, I., Morera, J., Catalán, R., Cerdeño, M. L., Roig-Pérez, J. F., Díaz-García, M., Chirosa-Cañavate, P., Piza-Ruiz, T., Vallejo-Casas, E., Vidal-Álvarez, S., Burch, J., Sagrera, J., Vivo, J., Cubo-Córdoba, A., Martínez-Enamorado, V., Rengel-Castro, F., García-Entero, V., Rodero, A., Viguera, E., Rohland, N., Morales, J. I., Soto, M., Mallick, S., Cebrià, A., García-Borja, P., Calduch-Bardoll, P., Ulloa-Eres, P., Carretero, A., Espinosa, T., Campderá-Gutiérrez, B., Pagés-Alonso, P., Vara-Izquierdo, C., Martínez-Peñarroya, J., Sardà-Seuma, S., Castaño-Aguilar, J. M., López-Chamizo, S., Pinhasi, R., Lalueza-Fox, C., & Reich, D. (2024). *Disparate demographic impacts of the Roman Colonization and the Migration Period in the Iberian Peninsula*. https://doi.org/10.1101/2024.09.23.614606

Cassidy, L. M., Russell, M., Smith, M., Delbarre, G., Cheetham, P., Manley, H., Mattiangeli, V., Breslin, E. M., Jackson, I., McCann, M., Little, H., O'Connor, C. G., Heaslip, B., Lawson, D., Endicott, P., & Bradley, D. G. (2025). Continental influx and pervasive matrilocality in Iron Age Britain. *Nature*, *637*(8048), 1136–1142. https://doi.org/10.1038/s41586-024-08409-6

Childebayeva, A., Rohrlach, A. B., Barquera, R., Rivollat, M., Aron, F., Szolek, A., Kohlbacher, O., Nicklisch, N., Alt, K. W., Gronenborn, D., Meller, H., Friederich, S., Prüfer, K., Deguilloux, M.-F., Krause, J., & Haak, W. (2022). Population Genetics and Signatures of Selection in Early Neolithic European Farmers. *Molecular Biology and Evolution*, *39*(6). https://doi.org/10.1093/molbev/msac108

Cornell, T. J. (1995). *The beginnings of Rome : Italy and Rome from the Bronze Age to the Punic Wars (c. 1000-264 BC)*. Routledge.

Crespo, L., Subirà, M. E., & Ruiz, J. (2011). Twins in prehistory: The case from Olèrdola (Barcelona, Spain; s. IV II BC). *International Journal of Osteoarchaeology*, *21*(6), 751–756. https://doi.org/10.1002/oa.1169

Cuesta-Aguirre, D. R., Malgosa, A., & Santos, C. (2024). An easy-to-use pipeline to analyze amplicon-based Next Generation Sequencing results of human mitochondrial DNA from degraded samples. *PLOS ONE*, *19*(11), e0311115. https://doi.org/10.1371/journal.pone.0311115

Cuesta-Aguirre, D. R., Roca-Rada, X., Di Biase, B., Vinueza-Espinosa, D. C., Sandoval-Ávila, C., Ravishankar, S., Molist, N., Durán, M., Mestres, I., Junyent, E., Alonso, N., Martinez Majoral, J., Kjær Knudsen, A., Lanigan, L. T., Schroeder, H., Llamas, B., Malgosa, A., & Santos, C. (2025). The Genetic Landscape of Northeastern Iberian Communities from the Early to Late Iron Age. *SSRN*. https://doi.org/10.2139/ssrn.5211463

Cuesta-Aguirre, D. R., Amor-Jimenez, C., Malgosa, A., & Santos, C. (2025). A Post-Mortem Molecular Damage Profile in the Ancient Human Mitochondrial DNA. *Molecular Ecology Resources*, e14061. https://doi.org/10.1111/1755-0998.14061

Dietler, M. (2009). Colonial Encounters in Iberia and the Western Mediterranean: An Exploratory Framework. In M. Dietler & C. López-Ruiz (Eds.), *Colonial encounters in ancient Iberia : Phoenician, Greek and indigenous relations*. The University of Chicago Press.

Dür, A., Huber, N., & Parson, W. (2021). Fine-Tuning Phylogenetic Alignment and Haplogrouping of mtDNA Sequences. *International Journal of Molecular Sciences*, *22*(11), 5747. https://doi.org/10.3390/ijms22115747

Duran i Caixal, M., Mestres i Santacreu, I., Padrós, C., & Principal, J. (2017). El Camp de les Lloses, un exemple d'implantació militar al territori d'època romanorepublicana. *Tribuna d'Arqueologia*, *Nº 2014-20*, 11–45.

Duran, M., Mestres, I., Padrós, C., & Principal, J. (2017). El Camp de les Lloses (Tona, Barcelona): evolución y significado del vicus romanorrepublicano. In *Roma en la Península Ibérica presertoriana. Escenarios de implantación militar provincial* (Vol. 56, pp. 153–189).

Duran, M., Mestres, I., Principal, J., & Padrós, C. (2015). El Camp de les Lloses (Tona , Osona): un post avançat amb tallers metal·lúrgics al servei de la logística de l'exèrcit romanorepublicà (125-75 ane). *Revista d'Arqueologia de Ponent*, *25*(May 2017), 293–307.

Edwell, P. (2011). War Abroad: Spain, Sicily, Macedon, Africa [Bookitem]. In *A Companion to the Punic Wars* (pp. 320–338). Wiley. https://doi.org/10.1002/9781444393712.ch18

Ehler, E., Novotný, J., Juras, A., Chyleński, M., Moravčík, O., & Pačes, J. (2019). AmtDB: a database of ancient human mitochondrial genomes. *Nucleic Acids Research*, *47*(D1), D29–D32. https://doi.org/10.1093/nar/gky843

Eng, R. Y., & Smith, T. C. (1976). Peasant Families and Population Control in Eighteenth-

Century Japan. *Journal of Interdisciplinary History*, *6*(3), 417. https://doi.org/10.2307/202664

Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, *38*(3), 276–278. https://doi.org/10.1038/s41587-020-0439-x

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Feeney, D. (2017). Carthage and Rome: Introduction. *Classical Philology*, *112*(3), 301–311. https://doi.org/10.1086/692440

Fellows Yates, J. A., Lamnidis, T. C., Borry, M., Andrades Valtueña, A., Fagernäs, Z., Clayton, S., Garcia, M. U., Neukamm, J., & Peltzer, A. (2021). Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. *PeerJ*, *9*, e10947. https://doi.org/10.7717/peerj.10947

Fernandes, D. M., Mittnik, A., Olalde, I., Lazaridis, I., Cheronet, O., Rohland, N., Mallick, S., Bernardos, R., Broomandkhoshbacht, N., Carlsson, J., Culleton, B. J., Ferry, M., Gamarra, B., Lari, M., Mah, M., Michel, M., Modi, A., Novak, M., Oppenheimer, J., Sirak, K. A., Stewardson, K., Mandl, K., Schattke, C., Özdoğan, K. T., Lucci, M., Gasperetti, G., Candilio, F., Salis, G., Vai, S., Camarós, E., Calò, C., Catalano, G., Cueto, M., Forgia, V., Lozano, M., Marini, E., Micheletti, M., Miccichè, R. M., Palombo, M. R., Ramis, D., Schimmenti, V., Sureda, P., Teira, L., Teschler-Nicola, M., Kennett, D. J., Lalueza-Fox, C., Patterson, N., Sineo, L., Coppa, A., Caramelli, D., Pinhasi, R., & Reich, D. (2020). The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. *Nature Ecology & Evolution*, *4*(3), 334–345. https://doi.org/10.1038/s41559-020-1102-0

Fregel, R., Méndez, F. L., Bokbot, Y., Martín-Socas, D., Camalich-Massieu, M. D., Santana, J., Morales, J., Ávila-Arcos, M. C., Underhill, P. A., Shapiro, B., Wojcik, G., Rasmussen, M., Soares, A. E. R., Kapp, J., Sockell, A., Rodríguez-Santos, F. J., Mikdad, A., Trujillo-Mederos, A., & Bustamante, C. D. (2018). Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proceedings of the National Academy of Sciences*, *115*(26), 6774–6779. https://doi.org/10.1073/pnas.1800851115

Freilich, S., Ringbauer, H., Los, D., Novak, M., Pavičić, D. T., Schiffels, S., & Pinhasi, R. (2021). Reconstructing genetic histories and social organisation in Neolithic and Bronze Age Croatia. *Scientific Reports*, *11*(1), 16729. https://doi.org/10.1038/s41598-021-94932-9

Gamba, C., Fernández, E., Oliver, A., Tirado, M., Baeza, C., López-Parra, A. M., & Arroyo-Pardo, E. (2008). Population genetics and DNA preservation in ancient human remains from Eastern Spain. *Forensic Science International: Genetics Supplement Series*, *1*(1), 462–464. https://doi.org/10.1016/j.fsigss.2007.10.037

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., Whittle, A., Dani, J., Raczky, P., Higham, T. F. G., Hofreiter, M., Bradley, D. G., & Pinhasi, R. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications*, *5*(1), 5257. https://doi.org/10.1038/ncomms6257

Garrison, E., Kronenberg, Z. N., Dawson, E. T., Pedersen, B. S., & Prins, P. (2022). A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. *PLOS Computational Biology*, *18*(5), e1009123. https://doi.org/10.1371/journal.pcbi.1009123

Garrison, E., & Marth, G. (2012). *Haplotype-based variant detection from short-read sequencing*.

Gelabert, P., Schmidt, R. W., Fernandes, D. M., Karsten, J. K., Harper, T. K., Madden, G. D., Ledogar, S. H., Sokhatsky, M., Oota, H., Kennett, D. J., & Pinhasi, R. (2022). Genomes from Verteba cave suggest diversity within the Trypillians in Ukraine. *Scientific Reports*, *12*(1), 7242. https://doi.org/10.1038/s41598-022-11117-8

Gilbert, M. T. P., Djurhuus, D., Melchior, L., Lynnerup, N., Worobey, M., Wilson, A. S., Andreasen, C., & Dissing, J. (2007). mtDNA from hair and nail clarifies the genetic relationship of the 15th century Qilakitsoq Inuit mummies. *American Journal of Physical Anthropology*, *133*(2), 847–853. https://doi.org/10.1002/ajpa.20602

Gitin, S., & Bierling, M. R. (2002). *The Phoenicians in Spain* (S. (Sy) Gitin (ed.)) [Book]. Penn State University Press. https://doi.org/10.1515/9781575065298

Goldsworthy, A. (2012). *The Fall of Carthage: The Punic Wars 265-146BC*. Orion.

González, A. M., Larruga, J. M., Abu-Amero, K. K., Shi, Y., Pestano, J., & Cabrera, V. M. (2007). Mitochondrial lineage M1 traces an early human backflow to Africa. *BMC Genomics*, *8*(1), 223. https://doi.org/10.1186/1471-2164-8-223

González Reyero, S. (2012). *Iberos: sociedades y territorios del occidente mediterráneo*.

Goudet, J., Raymond, M., de Meeüs, T., & Rousset, F. (1996). Testing Differentiation in Diploid Populations. *Genetics*, *144*(4), 1933–1940. https://doi.org/10.1093/genetics/144.4.1933

Gowland, R. L., Chamberlain, A., & Redfern, R. C. (2014). 'On the brink of being : re-evaluating infanticide and infant burial in Roman Britain. *Journal of Roman Archaeology Supplementary Series*, *96*, 69–88.

Grant, M. (1978). *History of Rome*. Prentice Hall.

Hassan, N. A.-M., Brown, K. A., Eyers, J., Brown, T. A., & Mays, S. (2014). Ancient DNA study of the remains of putative infanticide victims from the Yewden Roman villa site at Hambleden, England. *Journal of Archaeological Science*, *43*, 192–197. https://doi.org/10.1016/j.jas.2013.12.017

Hernández, C. L., Dugoujon, J. M., Novelletto, A., Rodríguez, J. N., Cuesta, P., & Calderón, R.

(2017). The distribution of mitochondrial DNA haplogroup H in southern Iberia indicates ancient human genetic exchanges along the western edge of the Mediterranean. *BMC Genetics*, *18*(1), 46. https://doi.org/10.1186/s12863-017-0514-6

Huber, N., Parson, W., & Dür, A. (2018). Next generation database search algorithm for forensic mitogenome analyses. *Forensic Science International: Genetics*, *37*, 204–214. https://doi.org/10.1016/j.fsigen.2018.09.001

Jordá Cerdá, F., Pellecer Catalán, M., Acosta Martínez, P., & Almagro-Gorbea, M. (1986). *Prehistoria*. Gredos.

Junyent, E. (1996a). Els origens de la cultura ilergeta. In *Indíbil i mandoni : reis i guerrers* (pp. 13–16).

Junyent, E. (1996b). La societat ilergata: un món de pagesos i aristòcrates. In *Indíbil i mandoni : reis i guerrers* (pp. 17–25).

Junyent, E. (2023). The times of Indivilis and Mandonius. In *Romans in the West* (pp. 21–33).

Keay, S. J., & Deamos, M. B. (1998). *The archaeology of early Roman Baetica*. Journal of Roman Archaeology.

Kılınç, G. M., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, A. M., Bıçakçı, E., Baird, D., Dönertaş, H. M., Ghalichi, A., Yaka, R., Koptekin, D., Açan, S. C., Parvizi, P., Krzewińska, M., Daskalaki, E. A., Yüncü, E., Dağtaş, N. D., Fairbairn, A., Pearson, J., Mustafaoğlu, G., Erdal, Y. S., Çakan, Y. G., Togan, İ., Somel, M., Storå, J., Jakobsson, M., & Götherström, A. (2016). The Demographic Development of the First Farmers in Anatolia. *Current Biology*, *26*(19), 2659–2666. https://doi.org/10.1016/j.cub.2016.07.057

Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P., & Pääbo, S. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature*, *464*(7290), 894–897. https://doi.org/10.1038/nature08976

Lazaridis, I., Alpaslan-Roodenberg, S., Acar, A., Açıkkol, A., Agelarakis, A., Aghikyan, L., Akyüz, U., Andreeva, D., Andrijašević, G., Antonović, D., Armit, I., Atmaca, A., Avetisyan, P., Aytek, A. İ., Bacvarov, K., Badalyan, R., Bakardzhiev, S., Balen, J., Bejko, L., Bernardos, R., Bertsatos, A., Biber, H., Bilir, A., Bodružić, M., Bonogofsky, M., Bonsall, C., Borić, D., Borovinić, N., Bravo Morante, G., Buttinger, K., Callan, K., Candilio, F., Carić, M., Cheronet, O., Chohadzhiev, S., Chovalopoulou, M.-E., Chryssoulaki, S., Ciobanu, I., Čondić, N., Constantinescu, M., Cristiani, E., Culleton, B. J., Curtis, E., Davis, J., Davtyan, R., Demcenco, T. I., Dergachev, V., Derin, Z., Deskaj, S., Devejyan, S., Djordjević, V., Duffett Carlson, K. S., Eccles, L. R., Elenski, N., Engin, A., Erdoğan, N., Erir-Pazarcı, S., Fernandes, D. M., Ferry, M., Freilich, S., Frînculeasa, A., Galaty, M. L., Gamarra, B., Gasparyan, B., Gaydarska, B., Genç, E., Gültekin, T., Gündüz, S., Hajdu, T., Heyd, V., Hobosyan, S., Hovhannisyan, N., Iliev, I., Iliev, L., Iliev, S., İvgin, İ., Janković, I., Jovanova, L., Karkanas, P., Kavaz-Kındığılı, B., Kaya, E. H., Keating, D., Kennett, D. J., Deniz Kesici, S., Khudaverdyan, A., Kiss, K., Kılıç, S., Klostermann, P., Kostak Boca Negra Valdes, S., Kovačević, S., Krenz-Niedbała, M., Krznarić Škrivanko, M., Kurti, R., Kuzman, P., Lawson, A. M., Lazar, C., Leshtakov, K., Levy, T. E., Liritzis, I., Lorentz, K. O.,

Łukasik, S., Mah, M., Mallick, S., Mandl, K., Martirosyan-Olshansky, K., Matthews, R., Matthews, W., McSweeney, K., Melikyan, V., Micco, A., Michel, M., Milašinović, L., Mittnik, A., Monge, J. M., Nekhrizov, G., Nicholls, R., Nikitin, A. G., Nikolov, V., Novak, M., Olalde, I., Oppenheimer, J., Osterholtz, A., Özdemir, C., Özdoğan, K. T., Öztürk, N., Papadimitriou, N., Papakonstantinou, N., Papathanasiou, A., Paraman, L., Paskary, E. G., Patterson, N., Petrakiev, I., Petrosyan, L., Petrova, V., Philippa-Touchais, A., Piliposyan, A., Pocuca Kuzman, N., Potrebica, H., Preda-Bălănică, B., Premužić, Z., Price, T. D., Qiu, L., Radović, S., Raeuf Aziz, K., Rajić Šikanjić, P., Rasheed Raheem, K., Razumov, S., Richardson, A., Roodenberg, J., Ruka, R., Russeva, V., Şahin, M., Şarbak, A., Savaş, E., Schattke, C., Schepartz, L., Selçuk, T., Sevim-Erol, A., Shamoon-Pour, M., Shephard, H. M., Sideris, A., Simalcsik, A., Simonyan, H., Sinika, V., Sirak, K., Sirbu, G., Šlaus, M., Soficaru, A., Söğüt, B., Sołtysiak, A., Sönmez-Sözer, Ç., Stathi, M., Steskal, M., Stewardson, K., Stocker, S., Suata-Alpaslan, F., Suvorov, A., Szécsényi-Nagy, A., Szeniczey, T., Telnov, N., Temov, S., Todorova, N., Tota, U., Touchais, G., Triantaphyllou, S., Türker, A., Ugarković, M., Valchev, T., Veljanovska, F., Videvski, Z., Virag, C., Wagner, A., Walsh, S., Włodarczak, P., Workman, J. N., Yardumian, A., Yarovoy, E., Yavuz, A. Y., Yılmaz, H., Zalzala, F., Zettl, A., Zhang, Z., Çavuşoğlu, R., Rohland, N., Pinhasi, R., & Reich, D. (2022). The genetic history of the Southern Arc: A bridge between West Asia and Europe. *Science*, *377*(6609). https://doi.org/10.1126/science.abm4247

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D. C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., Connell, S., Stewardson, K., Harney, E., Fu, Q., Gonzalez-Fortes, G., Jones, E. R., Roodenberg, S. A., Lengyel, G., Bocquentin, F., Gasparian, B., Monge, J. M., Gregg, M., Eshed, V., Mizrahi, A.-S., Meiklejohn, C., Gerritsen, F., Bejenaru, L., Blüher, M., Campbell, A., Cavalleri, G., Comas, D., Froguel, P., Gilbert, E., Kerr, S. M., Kovacs, P., Krause, J., McGettigan, D., Merrigan, M., Merriwether, D. A., O'Reilly, S., Richards, M. B., Semino, O., Shamoon-Pour, M., Stefanescu, G., Stumvoll, M., Tönjes, A., Torroni, A., Wilson, J. F., Yengo, L., Hovhannisyan, N. A., Patterson, N., Pinhasi, R., & Reich, D. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature*, *536*(7617), 419–424. https://doi.org/10.1038/nature19310

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR : An R Package for Multivariate Analysis. *Journal of Statistical Software*, *25*(1). https://doi.org/10.18637/jss.v025.i01

Li, M., Schröder, R., Ni, S., Madea, B., & Stoneking, M. (2015). Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proceedings of the National Academy of Sciences*, *112*(8), 2491–2496. https://doi.org/10.1073/pnas.1419651112

Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., Oppenheimer, J., Broomandkhoshbacht, N., Harney, E., Nordenfelt, S., Llamas, B., Gusztáv Mende, B., Köhler, K., Oross, K., Bondár, M., Marton, T., Osztás, A., Jakucs, J., Paluch, T., Horváth, F., Csengeri, P., Koós, J., Sebők, K., Anders, A., Raczky, P., Regenye, J., Barna, J. P., Fábián, S., Serlegi, G., Toldi, Z., Gyöngyvér Nagy, E., Dani, J., Molnár, E., Pálfi, G., Márk, L., Melegh, B., Bánfai, Z., Domboróczki, L., Fernández-Eraso, J., Antonio Mujika-Alustiza, J., Alonso Fernández, C.,

Jiménez Echevarría, J., Bollongino, R., Orschiedt, J., Schierhold, K., Meller, H., Cooper, A., Burger, J., Bánffy, E., Alt, K. W., Lalueza-Fox, C., Haak, W., & Reich, D. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature*, *551*(7680), 368–372. https://doi.org/10.1038/nature24476

López i Melcion, J. B., Martín i Ortega, A., Mataró Pladelasala, M., & Pons i Brun, E. (1989). *Memòria d'excavació 1987-1989 de l'Illa d'en Reixac (Ullastret, Baix Empordà)*.

Mallick, S., Micco, A., Mah, M., Ringbauer, H., Lazaridis, I., Olalde, I., Patterson, N., & Reich, D. (2024). The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Scientific Data*, *11*(1), 182. https://doi.org/10.1038/s41597-024-03031-7

Maluquer de Motes, J. (1966). *El impacto colonial griego y el comienzo de la vida urbana en Cataluña: Discurso leido en la sala de actos del Instituto de Estudios Ilerdenses el d{\'\i}a 8 de mayo*. Consejo Superior de Investigaciones Científicas. Delegación de Barcelona.

Maluquer de Motes, J. (1982). Problemática histórica de la Cultura Ibérica. *XVI Congreso Nacional de Arqueología: Murcia-Cartagena*, 29–49.

Martín i Méndez, A. (1998). *Memòria de les intervencions arqueologiques a Can Mateu (Cabrera de Mar, Maresme) 1997-1998*.

Martín i Ortega, A. (1985). *Memòria de les campanyes d'excavació de 1984 i 1985 en el Puig de Sant Andreu d'Ullastret (Baix Empordà)*.

Martín i Ortega, A., Codina, F., & Prado, G. (2016). La cerámica ática de la necrópolis del Puig de Serra (Serra de Daró-Ullastret, Cataluña). *Vie Quotidienne, Tombes et Symboles Des Sociétés Protohistoriques de Méditerranée Nord-Occidentale: Mélanges Offerts à Bernard Dedet*, *7*(2), 479–496.

Martirosyan, A., Sandoval-Ávila, C., Irurita, J., Juanhuix, J., Molist, N., Mestres, I., Durán, M., Alonso, N., Santos, C., Malgosa, A., Molera, J., & Jordana, X. (2024). Reconstructing infant mortality in Iberian Iron Age populations from tooth histology. *Journal of Archaeological Science*, *171*, 106088. https://doi.org/10.1016/j.jas.2024.106088

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E. R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J. L., de Castro, J. M. B., Carbonell, E., Gerritsen, F., Khokhlov, A., Kuznetsov, P., Lozano, M., Meller, H., Mochalov, O., Moiseyev, V., Guerra, M. A. R., Roodenberg, J., Vergès, J. M., Krause, J., Cooper, A., Alt, K. W., Brown, D., Anthony, D., Lalueza-Fox, C., Haak, W., Pinhasi, R., & Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*(7583), 499–503. https://doi.org/10.1038/nature16152

Matisoo-Smith, E., Gosling, A. L., Platt, D., Kardailsky, O., Prost, S., Cameron-Christie, S., Collins, C. J., Boocock, J., Kurumilian, Y., Guirguis, M., Pla Orquín, R., Khalil, W., Genz, H., Abou Diwan, G., Nassar, J., & Zalloua, P. (2018). Ancient mitogenomes of Phoenicians from Sardinia and Lebanon: A story of settlement, integration, and female

mobility. *PLOS ONE*, *13*(1), e0190169. https://doi.org/10.1371/journal.pone.0190169

Mittnik, A., Wang, C.-C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R., Khartanovich, V., Moiseyev, V., Tõrv, M., Furtwängler, A., Andrades Valtueña, A., Feldman, M., Economou, C., Oinonen, M., Vasks, A., Balanovska, E., Reich, D., Jankauskas, R., Haak, W., Schiffels, S., & Krause, J. (2018). The genetic prehistory of the Baltic Sea region. *Nature Communications*, *9*(1), 442. https://doi.org/10.1038/s41467-018-02825-9

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press. https://doi.org/10.7312/nei-92038

Nesheva, D. V. (2014). Aspects of Ancient Mitochondrial DNA Analysis in Different Populations for Understanding Human Evolution. *Balkan Journal of Medical Genetics*, *17*(1), 5–14. https://doi.org/10.2478/bjmg-2014-0019

Novak, M., Olalde, I., Ringbauer, H., Rohland, N., Ahern, J., Balen, J., Janković, I., Potrebica, H., Pinhasi, R., & Reich, D. (2021). Genome-wide analysis of nearly all the victims of a 6200 year old massacre. *PLOS ONE*, *16*(3), e0247332. https://doi.org/10.1371/journal.pone.0247332

Olalde, I., Brace, S., Allentoft, M. E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., Altena, E., Lipson, M., Lazaridis, I., Harper, T. K., Patterson, N., Broomandkhoshbacht, N., Diekmann, Y., Faltyskova, Z., Fernandes, D., Ferry, M., Harney, E., de Knijff, P., Michel, M., Oppenheimer, J., Stewardson, K., Barclay, A., Alt, K. W., Liesau, C., Ríos, P., Blasco, C., Miguel, J. V., García, R. M., Fernández, A. A., Bánffy, E., Bernabò-Brea, M., Billoin, D., Bonsall, C., Bonsall, L., Allen, T., Büster, L., Carver, S., Navarro, L. C., Craig, O. E., Cook, G. T., Cunliffe, B., Denaire, A., Dinwiddy, K. E., Dodwell, N., Ernée, M., Evans, C., Kuchařík, M., Farré, J. F., Fowler, C., Gazenbeek, M., Pena, R. G., Haber-Uriarte, M., Haduch, E., Hey, G., Jowett, N., Knowles, T., Massy, K., Pfrengle, S., Lefranc, P., Lemercier, O., Lefebvre, A., Martínez, C. H., Olmo, V. G., Ramírez, A. B., Maurandi, J. L., Majó, T., McKinley, J. I., McSweeney, K., Mende, B. G., Modi, A., Kulcsár, G., Kiss, V., Czene, A., Patay, R., Endrődi, A., Köhler, K., Hajdu, T., Szeniczey, T., Dani, J., Bernert, Z., Hoole, M., Cheronet, O., Keating, D., Velemínský, P., Dobeš, M., Candilio, F., Brown, F., Fernández, R. F., Herrero-Corral, A.-M., Tusa, S., Carnieri, E., Lentini, L., Valenti, A., Zanini, A., Waddington, C., Delibes, G., Guerra-Doce, E., Neil, B., Brittain, M., Luke, M., Mortimer, R., Desideri, J., Besse, M., Brücken, G., Furmanek, M., Hałuszko, A., Mackiewicz, M., Rapiński, A., Leach, S., Soriano, I., Lillios, K. T., Cardoso, J. L., Pearson, M. P., Włodarczak, P., Price, T. D., Prieto, P., Rey, P.-J., Risch, R., Rojo Guerra, M. A., Schmitt, A., Serralongue, J., Silva, A. M., Smrčka, V., Vergnaud, L., Zilhão, J., Caramelli, D., Higham, T., Thomas, M. G., Kennett, D. J., Fokkens, H., Heyd, V., Sheridan, A., Sjögren, K.-G., Stockhammer, P. W., Krause, J., Pinhasi, R., Haak, W., Barnes, I., Lalueza-Fox, C., & Reich, D. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature*, *555*(7695), 190–196. https://doi.org/10.1038/nature25738

Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., Dulias, K., Edwards, C. J., Gandini, F., Pala, M., Soares, P., Ferrando-Bernal, M., Adamski, N.,

Broomandkhoshbacht, N., Cheronet, O., Culleton, B. J., Fernandes, D., Lawson, A. M., Mah, M., Oppenheimer, J., Stewardson, K., Zhang, Z., Jiménez Arenas, J. M., Toro Moyano, I. J., Salazar-García, D. C., Castanyer, P., Santos, M., Tremoleda, J., Lozano, M., García Borja, P., Fernández-Eraso, J., Mujika-Alustiza, J. A., Barroso, C., Bermúdez, F. J., Viguera Mínguez, E., Burch, J., Coromina, N., Vivó, D., Cebrià, A., Fullola, J. M., García-Puchol, O., Morales, J. I., Oms, F. X., Majó, T., Vergès, J. M., Díaz-Carvajal, A., Ollich-Castanyer, I., López-Cachero, F. J., Silva, A. M., Alonso-Fernández, C., Delibes de Castro, G., Jiménez Echevarría, J., Moreno-Márquez, A., Pascual Berlanga, G., Ramos-García, P., Ramos-Muñoz, J., Vijande Vila, E., Aguilella Arzo, G., Esparza Arroyo, Á., Lillios, K. T., Mack, J., Velasco-Vázquez, J., Waterman, A., Benítez de Lugo Enrich, L., Benito Sánchez, M., Agustí, B., Codina, F., de Prado, G., Estalrrich, A., Fernández Flores, Á., Finlayson, C., Finlayson, G., Finlayson, S., Giles-Guzmán, F., Rosas, A., Barciela González, V., García Atiénzar, G., Hernández Pérez, M. S., Llanos, A., Carrión Marco, Y., Collado Beneyto, I., López-Serrano, D., Sanz Tormo, M., Valera, A. C., Blasco, C., Liesau, C., Ríos, P., Daura, J., de Pedro Michó, M. J., Diez-Castillo, A. A., Flores Fernández, R., Francès Farré, J., Garrido-Pena, R., Gonçalves, V. S., Guerra-Doce, E., Herrero-Corral, A. M., Juan-Cabanilles, J., López-Reyes, D., McClure, S. B., Merino Pérez, M., Oliver Foix, A., Sanz Borràs, M., Sousa, A. C., Vidal Encinas, J. M., Kennett, D. J., Richards, M. B., Werner Alt, K., Haak, W., Pinhasi, R., Lalueza-Fox, C., & Reich, D. (2019). The genomic history of the Iberian Peninsula over the past 8000 years. *Science*, *363*(6432), 1230–1234. https://doi.org/10.1126/science.aav4040

Oliver Foix, A. (2004). Sacrificios y “mala muerte” en el Registro Arqueológico de los Yacimientos Ibéricos. *Espacio Tiempo y Forma. Serie I, Prehistoria y Arqueología*, *1*(16–17). https://doi.org/10.5944/etfi.16-17.2003.4763

Olivieri, A., Pala, M., Gandini, F., Kashani, B. H., Perego, U. A., Woodward, S. R., Grugni, V., Battaglia, V., Semino, O., Achilli, A., Richards, M. B., & Torroni, A. (2013). Mitogenomes from Two Uncommon Haplogroups Mark Late Glacial/Postglacial Expansions from the Near East and Neolithic Dispersals within Europe. *PLoS ONE*, *8*(7), e70492. https://doi.org/10.1371/journal.pone.0070492

Översti, S., Majander, K., Salmela, E., Salo, K., Arppe, L., Belskiy, S., Etu-Sihvola, H., Laakso, V., Mikkola, E., Pfrengle, S., Putkonen, M., Taavitsainen, J.-P., Vuoristo, K., Wessman, A., Sajantila, A., Oinonen, M., Haak, W., Schuenemann, V. J., Krause, J., Palo, J. U., & Onkamo, P. (2019). Human mitochondrial DNA lineages in Iron-Age Fennoscandia suggest incipient admixture and eastern introduction of farming-related maternal ancestry. *Scientific Reports*, *9*(1), 16883. https://doi.org/10.1038/s41598-019-51045-8

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*, 645–679. https://doi.org/10.1146/annurev.genet.37.110801.143214

Pala, M., Olivieri, A., Achilli, A., Accetturo, M., Metspalu, E., Reidla, M., Tamm, E., Karmin, M., Reisberg, T., Kashani, B. H., Perego, U. A., Carossa, V., Gandini, F., Pereira, J. B., Soares, P., Angerhofer, N., Rychkov, S., Al-Zahery, N., Carelli, V., Sanati, M. H., Houshmand, M.,

Hatina, J., Macaulay, V., Pereira, L., Woodward, S. R., Davies, W., Gamble, C., Baird, D., Semino, O., Villems, R., Torroni, A., & Richards, M. B. (2012). Mitochondrial DNA Signals of Late Glacial Recolonization of Europe from Near Eastern Refugia. *The American Journal of Human Genetics*, *90*(5), 915–924. https://doi.org/10.1016/j.ajhg.2012.04.003

Palanichamy, M. G., Zhang, C.-L., Mitra, B., Malyarchuk, B., Derenko, M., Chaudhuri, T. K., & Zhang, Y.-P. (2010). Mitochondrial haplogroup N1a phylogeography, with implication to the origin of European farmers. *BMC Evolutionary Biology*, *10*(1), 304. https://doi.org/10.1186/1471-2148-10-304

Papac, L., de Miguel Ibáñez, P., Rohrlach, A. B., Armendáriz, J., Peres, M., Lamnidis, T. C., Mötsch, A., Schiffels, S., & Risch, R. (2023). Intramural child burials in Iron Age Navarra: How ancient DNA can contribute to household archaeology . In H. Meller, J. Krause, W. Haak, & R. Risch (Eds.), *Kinship, Sex, and Biological Relatedness : The contribution of archaeogenetics to the understanding of social and biological relations*. *15. Mitteldeutscher Archäologentag vom 6. bis 8. Oktober 2022 in Halle (Saale)* (pp. 263–295). https://doi.org/10.11588/propylaeum.1280.c18012

Papac, L., Ernée, M., Dobeš, M., Langová, M., Rohrlach, A. B., Aron, F., Neumann, G. U., Spyrou, M. A., Rohland, N., Velemínský, P., Kuna, M., Brzobohatá, H., Culleton, B., Daněček, D., Danielisová, A., Dobisíková, M., Hložek, J., Kennett, D. J., Klementová, J., Kostka, M., Krištuf, P., Kuchařík, M., Hlavová, J. K., Limburský, P., Malyková, D., Mattiello, L., Pecinovská, M., Petriščáková, K., Průchová, E., Stránská, P., Smejtek, L., Špaček, J., Šumberová, R., Švejcar, O., Trefný, M., Vávra, M., Kolář, J., Heyd, V., Krause, J., Pinhasi, R., Reich, D., Schiffels, S., & Haak, W. (2021). Dynamic changes in genomic and social structures in third millennium BCE central Europe. *Science Advances*, *7*(35). https://doi.org/10.1126/sciadv.abi6941

Patterson, N., Isakov, M., Booth, T., Büster, L., Fischer, C.-E., Olalde, I., Ringbauer, H., Akbari, A., Cheronet, O., Bleasdale, M., Adamski, N., Altena, E., Bernardos, R., Brace, S., Broomandkhoshbacht, N., Callan, K., Candilio, F., Culleton, B., Curtis, E., Demetz, L., Carlson, K. S. D., Edwards, C. J., Fernandes, D. M., Foody, M. G. B., Freilich, S., Goodchild, H., Kearns, A., Lawson, A. M., Lazaridis, I., Mah, M., Mallick, S., Mandl, K., Micco, A., Michel, M., Morante, G. B., Oppenheimer, J., Özdoğan, K. T., Qiu, L., Schattke, C., Stewardson, K., Workman, J. N., Zalzala, F., Zhang, Z., Agustí, B., Allen, T., Almássy, K., Amkreutz, L., Ash, A., Baillif-Ducros, C., Barclay, A., Bartosiewicz, L., Baxter, K., Bernert, Z., Blažek, J., Bodružić, M., Boissinot, P., Bonsall, C., Bradley, P., Brittain, M., Brookes, A., Brown, F., Brown, L., Brunning, R., Budd, C., Burmaz, J., Canet, S., Carnicero-Cáceres, S., Čaušević-Bully, M., Chamberlain, A., Chauvin, S., Clough, S., Čondić, N., Coppa, A., Craig, O., Črešnar, M., Cummings, V., Czifra, S., Danielisová, A., Daniels, R., Davies, A., de Jersey, P., Deacon, J., Deminger, C., Ditchfield, P. W., Dizdar, M., Dobeš, M., Dobisíková, M., Domboróczki, L., Drinkall, G., Đukić, A., Ernée, M., Evans, C., Evans, J., Fernández-Götz, M., Filipović, S., Fitzpatrick, A., Fokkens, H., Fowler, C., Fox, A., Gallina, Z., Gamble, M., González Morales, M. R., González-Rabanal, B., Green, A., Gyenesei, K., Habermehl, D., Hajdu, T., Hamilton, D., Harris, J., Hayden, C., Hendriks, J., Hernu, B., Hey, G., Horňák, M., Ilon, G., Istvánovits, E., Jones, A. M., Kavur, M. B., Kazek, K., Kenyon, R. A., Khreisheh, A., Kiss, V., Kleijne, J., Knight, M., Kootker, L. M.,

Kovács, P. F., Kozubová, A., Kulcsár, G., Kulcsár, V., Le Pennec, C., Legge, M., Leivers, M., Loe, L., López-Costas, O., Lord, T., Los, D., Lyall, J., Marín-Arroyo, A. B., Mason, P., Matošević, D., Maxted, A., McIntyre, L., McKinley, J., McSweeney, K., Meijlink, B., Mende, B. G., Menđušić, M., Metlička, M., Meyer, S., Mihovilić, K., Milasinovic, L., Minnitt, S., Moore, J., Morley, G., Mullan, G., Musilová, M., Neil, B., Nicholls, R., Novak, M., Pala, M., Papworth, M., Paresys, C., Patten, R., Perkić, D., Pesti, K., Petit, A., Petriščáková, K., Pichon, C., Pickard, C., Pilling, Z., Price, T. D., Radović, S., Redfern, R., Resutík, B., Rhodes, D. T., Richards, M. B., Roberts, A., Roefstra, J., Sankot, P., Šefčáková, A., Sheridan, A., Skae, S., Šmolíková, M., Somogyi, K., Somogyvári, Á., Stephens, M., Szabó, G., Szécsényi-Nagy, A., Szeniczey, T., Tabor, J., Tankó, K., Maria, C. T., Terry, R., Teržan, B., Teschler-Nicola, M., Torres-Martínez, J. F., Trapp, J., Turle, R., Ujvári, F., van der Heiden, M., Veleminsky, P., Veselka, B., Vytlačil, Z., Waddington, C., Ware, P., Wilkinson, P., Wilson, L., Wiseman, R., Young, E., Zaninović, J., Žitňan, A., Lalueza-Fox, C., de Knijff, P., Barnes, I., Halkon, P., Thomas, M. G., Kennett, D. J., Cunliffe, B., Lillie, M., Rohland, N., Pinhasi, R., Armit, I., & Reich, D. (2022). Large-scale migration into Britain during the Middle to Late Bronze Age. *Nature*, *601*(7894), 588–594. https://doi.org/10.1038/s41586-021-04287-4

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., & Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. *Genome Biology*, *17*(1), 60. https://doi.org/10.1186/s13059-016-0918-z

Posth, C., Zaro, V., Spyrou, M. A., Vai, S., Gnecchi-Ruscone, G. A., Modi, A., Peltzer, A., Mötsch, A., Nägele, K., Vågene, Å. J., Nelson, E. A., Radzevičiūtė, R., Freund, C., Bondioli, L. M., Cappuccini, L., Frenzel, H., Pacciani, E., Boschin, F., Capecchi, G., Martini, I., Moroni, A., Ricci, S., Sperduti, A., Turchetti, M. A., Riga, A., Zavattaro, M., Zifferero, A., Heyne, H. O., Fernández-Domínguez, E., Kroonen, G. J., McCormick, M., Haak, W., Lari, M., Barbujani, G., Bondioli, L., Bos, K. I., Caramelli, D., & Krause, J. (2021). The origin and legacy of the Etruscans through a 2000-year archeogenomic time transect. *Science Advances*, *7*(39). https://doi.org/10.1126/sciadv.abi7673

Prendergast, M. E., Lipson, M., Sawchuk, E. A., Olalde, I., Ogola, C. A., Rohland, N., Sirak, K. A., Adamski, N., Bernardos, R., Broomandkhoshbacht, N., Callan, K., Culleton, B. J., Eccles, L., Harper, T. K., Lawson, A. M., Mah, M., Oppenheimer, J., Stewardson, K., Zalzala, F., Ambrose, S. H., Ayodo, G., Gates, H. L., Gidna, A. O., Katongo, M., Kwekason, A., Mabulla, A. Z. P., Mudenda, G. S., Ndiema, E. K., Nelson, C., Robertshaw, P., Kennett, D. J., Manthi, F. K., & Reich, D. (2019). Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. *Science*, *365*(6448). https://doi.org/10.1126/science.aaw6275

QGIS Development Team. (2025). *QGIS Geographic Information System*.

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*.

Raymond, M., & Rousset, F. (1995). An Exact Test for Population Differentiation. *Evolution*, *49*(6), 1280–1283. https://doi.org/10.1111/j.1558-5646.1995.tb04456.x

Richard, M. (2011). *Carthage Must Be Destroyed: The Rise and Fall of an Ancient Civilization*

(Penguin (ed.)).

Rivollat, M., Jeong, C., Schiffels, S., Küçükkalıpçı, İ., Pemonge, M.-H., Rohrlach, A. B., Alt, K. W., Binder, D., Friederich, S., Ghesquière, E., Gronenborn, D., Laporte, L., Lefranc, P., Meller, H., Réveillas, H., Rosenstock, E., Rottier, S., Scarre, C., Soler, L., Wahl, J., Krause, J., Deguilloux, M.-F., & Haak, W. (2020). Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. *Science Advances*, *6*(22). https://doi.org/10.1126/sciadv.aaz5344

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Roca-Rada, X., Davidson, R., Williams, M. P., Ravishankar, S., Collen, E., Haarkötter, C., Taufik, L., Carvalho, A. F., Villalba-Mouco, V., Cuesta-Aguirre, D. R., Tente, C., Monge Calleja, Á. M., MacRoberts, R. A., Melo, L., Purnomo, G. A., Souilmi, Y., Tobler, R., Cunha, E., Tereso, S., Matos, V. M. J., Fernandes, T. M., Mauer, A.-F., Silva, A. M., Carvalho, P. C., Llamas, B., & Teixeira, J. C. (2024). *The genetic history of Portugal over the past 5,000 years*. https://doi.org/10.1101/2024.09.12.612544

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., & Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1660), 20130624. https://doi.org/10.1098/rstb.2013.0624

Roostalu, U., Kutuev, I., Loogväli, E.-L., Metspalu, E., Tambets, K., Reidla, M., Khusnutdinova, E., Usanga, E., Kivisild, T., & Villems, R. (2007). Origin and Expansion of Haplogroup H, the Dominant Human Mitochondrial DNA Lineage in West Eurasia: The Near Eastern and Caucasian Perspective. *Molecular Biology and Evolution*, *24*(2), 436–448. https://doi.org/10.1093/molbev/msl173

Sampietro, M. L., Caramelli, D., Lao, O., Calafell, F., Comas, D., Lari, M., Agustí, B., Bertranpetit, J., & Lalueza-Fox, C. (2005). The Genetics of the Pre-Roman Iberian Peninsula: A mtDNA Study of Ancient Iberians. *Annals of Human Genetics*, *69*(5), 535–548. https://doi.org/10.1111/j.1529-8817.2005.00194.x

Sandoval-Ávila, C., Martirosyan, A., Cuesta-Aguirre, D. R., Jordana, X., Nociarová, D., De Prado, G., Santos, C., & Malgosa, A. (2025). Muerte infantil en época ibérica: el complejo arqueológico de Ullastret (Girona, Cataluña). *Trabajos de Prehistoria*, *81*(2), 959. https://doi.org/10.3989/tp.2024.959

Sanmartí, J. (2003). Els ibers quaranta anys més tard : Un breu balanç de la recerca. *Cota Zero: Revista d'arqueologia i Ciència*, *18*, 147–157.

Sanmartí, J. (2011). L'Escola Catalana d'Arqueologia i l'estudi Dels Ibers. In *Institut d'Estudis Catalans, Secció Històrico-Arqueològica*.

Sanmartí, J. (2013). Joan Maluquer de Motes i els estudis ibèrics. *Revista d'arqueologia de Ponent*, *23*, 343–352.

Sanmartí, J., Santacana, J., Belarte, M. C., & Asensio, D. (1998). Paisatges ibèrics: Tipus d'assentaments i formes d'ocupació del territori a la costa central de Catalunya durant el període ibèric ple. *SAGVNTVM Extra*, *1*, 373–385.

Schönherr, S., Weissensteiner, H., Kronenberg, F., & Forer, L. (2023). Haplogrep 3 - an interactive haplogroup classification and analysis platform. *Nucleic Acids Research*, *51*(W1), W263–W268. https://doi.org/10.1093/nar/gkad284

Schullard, H. H. (1989). Carthage and Rome. In *The Cambridge Ancient History: VII Part 2, The Rise of Rome to 220 BC* (2nd editio, pp. 486–569). Cambridge University Press.

Silva, N. M., Kreutzer, S., Souleles, A., Triantaphyllou, S., Kotsakis, K., Urem-Kotsou, D., Halstead, P., Efstratiou, N., Kotsos, S., Karamitrou-Mentessidi, G., Adaktylou, F., Chondroyianni-Metoki, A., Pappa, M., Ziota, C., Sampson, A., Papathanasiou, A., Vitelli, K., Cullen, T., Kyparissi-Apostolika, N., Lanz, A. Z., Peters, J., Rio, J., Wegmann, D., Burger, J., Currat, M., & Papageorgopoulou, C. (2022). Ancient mitochondrial diversity reveals population homogeneity in Neolithic Greece and identifies population dynamics along the Danubian expansion axis. *Scientific Reports*, *12*(1), 13474. https://doi.org/10.1038/s41598-022-16745-8

Simón, M., Jordana, X., Armentano, N., Santos, C., Díaz, N., Solórzano, E., López, J. B., González-Ruiz, M., & Malgosa, A. (2011). The presence of nuclear families in prehistoric collective burials revisited: The bronze age burial of montanissell cave (Spain) in the light of aDNA. *American Journal of Physical Anthropology*, *146*(3), 406–413. https://doi.org/10.1002/ajpa.21590

Sinner, A. G. (2015). Cultural contacts and identity construction: A colonial context in NE Spain (2nd - Early 1st c. B.C.). *Journal of Roman Archaeology*, *28*(1), 7–37. https://doi.org/10.1017/S1047759415002391

Sinner, A. G., & Ferrer i Jané, J. (2016). Del oppidum de Burriac a las termas de Ca l'Arnau. Una aproximación a la lengua y a la identidad de los habitantes de Ilduro (Cabrera de Mar, Barcelona). *Archivo Español de Arqueología*, *89*(November), 193. https://doi.org/10.3989/aespa.089.016.010

Skoglund, P., Malmström, H., Omrak, A., Raghavan, M., Valdiosera, C., Günther, T., Hall, P., Tambets, K., Parik, J., Sjögren, K.-G., Apel, J., Willerslev, E., Storå, J., Götherström, A., & Jakobsson, M. (2014). Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. *Science*, *344*(6185), 747–750. https://doi.org/10.1126/science.1253448

Skourtanioti, E., Erdal, Y. S., Frangipane, M., Balossi Restelli, F., Yener, K. A., Pinnock, F., Matthiae, P., Özbal, R., Schoop, U.-D., Guliyev, F., Akhundov, T., Lyonnet, B., Hammer, E. L., Nugent, S. E., Burri, M., Neumann, G. U., Penske, S., Ingman, T., Akar, M., Shafiq, R., Palumbi, G., Eisenmann, S., D'Andrea, M., Rohrlach, A. B., Warinner, C., Jeong, C., Stockhammer, P. W., Haak, W., & Krause, J. (2020). Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. *Cell*, *181*(5), 1158-1175.e28. https://doi.org/10.1016/j.cell.2020.04.044

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V., & Richards, M. B. (2009). Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *The American Journal of Human Genetics*, *84*(6), 740–759. https://doi.org/10.1016/j.ajhg.2009.05.001

Subirà, M. E., & Molist, N. (2007). Inhumacions perinatals múltiples i espais de treball en els assentaments ibers. *Nasciturus: Infans, Puerulus. Vobis Mater Terra. La Muerte En La Infancia*, *Uh 4*, 365–385.

Szécsényi-Nagy, A., Roth, C., Brandt, G., Rihuete-Herrada, C., Tejedor-Rodríguez, C., Held, P., García-Martínez-de-Lagrán, Í., Arcusa Magallón, H., Zesch, S., Knipper, C., Bánffy, E., Friederich, S., Meller, H., Bueno Ramírez, P., Barroso Bermejo, R., de Balbín Behrmann, R., Herrero-Corral, A. M., Flores Fernández, R., Alonso Fernández, C., Jiménez Echevarria, J., Rindlisbacher, L., Oliart, C., Fregeiro, M.-I., Soriano, I., Vicente, O., Micó, R., Lull, V., Soler Díaz, J., López Padilla, J. A., Roca de Togores Muñoz, C., Hernández Pérez, M. S., Jover Maestre, F. J., Lomba Maurandi, J., Avilés Fernández, A., Lillios, K. T., Silva, A. M., Magalhães Ramalho, M., Oosterbeek, L. M., Cunha, C., Waterman, A. J., Roig Buxó, J., Martínez, A., Ponce Martínez, J., Hunt Ortiz, M., Mejías-García, J. C., Pecero Espín, J. C., Cruz-Auñón Briones, R., Tomé, T., Carmona Ballestero, E., Cardoso, J. L., Araújo, A. C., Liesau von Lettow-Vorbeck, C., Blasco Bosqued, C., Ríos Mendoza, P., Pujante, A., Royo-Guillén, J. I., Esquembre Beviá, M. A., Dos Santos Goncalves, V. M., Parreira, R., Morán Hernández, E., Méndez Izquierdo, E., Vega y Miguel, J., Menduiña García, R., Martínez Calvo, V., López Jiménez, O., Krause, J., Pichler, S. L., Garrido-Pena, R., Kunst, M., Risch, R., Rojo-Guerra, M. A., Haak, W., & Alt, K. W. (2017). The maternal genetic make-up of the Iberian Peninsula between the Neolithic and the Early Bronze Age. *Scientific Reports*, *7*(1), 15644. https://doi.org/10.1038/s41598-017-15480-9

Tarradell, M. (1962). *Les Arrels de Catalunya*. Vicens-Vives.

Torroni, A., Bandelt, H.-J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M.-L., Bonné-Tamir, B., & Scozzari, R. (1998). mtDNA Analysis Reveals a Major Late Paleolithic Population Expansion from Southwestern to Northeastern Europe. *The American Journal of Human Genetics*, *62*(5), 1137–1152. https://doi.org/10.1086/301822

Torroni, A., Bandelt, H.-J., Macaulay, V., Richards, M., Cruciani, F., Rengo, C., Martinez-Cabrera, V., Villems, R., Kivisild, T., Metspalu, E., Parik, J., Tolk, H.-V., Tambets, K., Forster, P., Karger, B., Francalacci, P., Rudan, P., Janicijevic, B., Rickards, O., Savontaus, M.-L., Huoponen, K., Laitinen, V., Koivumäki, S., Sykes, B., Hickey, E., Novelletto, A., Moral, P., Sellitto, D., Coppa, A., Al-Zaheri, N., Santachiara-Benerecetti, A. S., Semino, O., & Scozzari, R. (2001). A Signal, from Human mtDNA, of Postglacial Recolonization in Europe. *The American Journal of Human Genetics*, *69*(4), 844–852. https://doi.org/10.1086/323485

van de Loosdrecht, M., Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., Nickel, B., Nagel, S., Talbi, E. H., El Hajraoui, M. A., Amzazi, S., Hublin, J.-J., Pääbo, S., Schiffels, S., Meyer, M., Haak, W., Jeong, C., & Krause, J. (2018). Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations.

Science, *360*(6388), 548–552. https://doi.org/10.1126/science.aar8380

van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, *30*(2), E386–E394. https://doi.org/10.1002/humu.20921

Villalba-Mouco, V., Oliart, C., Rihuete-Herrada, C., Childebayeva, A., Rohrlach, A. B., Fregeiro, M. I., Celdrán Beltrán, E., Velasco-Felipe, C., Aron, F., Himmel, M., Freund, C., Alt, K. W., Salazar-García, D. C., García Atiénzar, G., de Miguel Ibáñez, M. P., Hernández Pérez, M. S., Barciela, V., Romero, A., Ponce, J., Martínez, A., Lomba, J., Soler, J., Martínez, A. P., Avilés Fernández, A., Haber-Uriarte, M., Roca de Togores Muñoz, C., Olalde, I., Lalueza-Fox, C., Reich, D., Krause, J., García Sanjuán, L., Lull, V., Micó, R., Risch, R., & Haak, W. (2021). Genomic transformation and social organization during the Copper Age–Bronze Age transition in southern Iberia. *Science Advances*, *7*(47), 21. https://doi.org/10.1126/sciadv.abi7038

Villalba-Mouco, V., Oliart, C., Rihuete-Herrada, C., Rohrlach, A. B., Fregeiro, M. I., Childebayeva, A., Ringbauer, H., Olalde, I., Celdrán Beltrán, E., Puello-Mora, C., Valério, M., Krause, J., Lull, V., Micó, R., Risch, R., & Haak, W. (2022). Kinship practices in the early state El Argar society from Bronze Age Iberia. *Scientific Reports*, *12*(1), 22415. https://doi.org/10.1038/s41598-022-25975-9

Villalba-Mouco, V., van de Loosdrecht, M. S., Rohrlach, A. B., Fewlass, H., Talamo, S., Yu, H., Aron, F., Lalueza-Fox, C., Cabello, L., Cantalejo Duarte, P., Ramos-Muñoz, J., Posth, C., Krause, J., Weniger, G.-C., & Haak, W. (2023). A 23,000-year-old southern Iberian individual links human groups that lived in Western Europe before and after the Last Glacial Maximum. *Nature Ecology & Evolution*. https://doi.org/10.1038/s41559-023-01987-0

Vinueza-Espinosa, D. C., Cuesta-Aguirre, D. R., Malgosa, A., & Santos, C. (2023). Mitochondrial DNA control region typing from highly degraded skeletal remains by single-multiplex next-generation sequencing. *ELECTROPHORESIS*, *44*(17–18), 1423–1434. https://doi.org/10.1002/elps.202200052

Vinueza-Espinosa, D. C., Santos, C., Martínez-Labarga, C., & Malgosa, A. (2020). Human DNA extraction from highly degraded skeletal remains: How to find a suitable method? *ELECTROPHORESIS*, *41*(24), 2149–2158. https://doi.org/10.1002/elps.202000171

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., & Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*, *44*(W1), W58-63. https://doi.org/10.1093/nar/gkw233

Winters, C. (2010). The African Origin of mtDNA Haplogroup M1. *Current Research Journal of Biological Sciences*, *2*(6), 380–389.

Zimmermann, K. (2011). Roman Strategy and Aims in the Second Punic War [Bookitem]. In *A Companion to the Punic Wars* (pp. 280–298). Wiley. https://doi.org/10.1002/9781444393712.ch16



5.

GENERAL DISCUSSION

5. GENERAL DISCUSSION

The remains of 69 newborns from seven Iberian archaeological sites located in the northeast of the Iberian Peninsula were used in this thesis. Namely, 33 individuals from the fortress of Vilars, 10 from Illa d'en Reixac, 3 from Puig de Sant Andreu, 8 from Sant Miquel d'Olèrdola, 2 from Castell de Besora, 1 from El poblat Ibèric de Burriac, and 12 from El Camp de les Lloses. Some of these samples were excavated more than 30 years ago during a period when archaeological techniques were less refined and precautions against DNA contamination were virtually nonexistent. After excavation, the remains typically underwent anthropological examination before being curated in museums, where they remained accessible for future research. While this accessibility enabled important contributions to archaeology and anthropology, standardized protocols for the preservation of genetic material had not yet been adopted. Consequently, some samples may have been exposed to suboptimal conditions that could have compromised DNA preservation.

Today, ancient DNA analyses are conducted under rigorous protocols designed to preserve genetic material and prevent contamination. Ideally, anthropologists extract samples directly at the excavation site using appropriate protective equipment and transfer them to dedicated ancient DNA laboratories. When additional material is required, further sampling is performed in clean laboratory environments under sterile conditions, using gloves and sterilized tools to ensure sample integrity. In contrast, for most of the individuals analyzed in this thesis, such protocols were not in place at the time of excavation, which may have compromised DNA preservation. Nevertheless, a few samples, such as those from El Camp de les Lloses, were taken directly in the field following more recent standards, offering improved chances of genetic recovery.

Concerning samples available for genetic analysis, not all of the individuals used in this thesis presented petrous bone, which is known to yield the highest quality and most abundant aDNA due to its dense protected nature and its high concentrations of osteocytes (Haarkötter et al., 2023; Ibrahim et al., 2022; Parker et al., 2020; Pinhasi et al., 2015; Vinueza-Espinosa et al., 2020; Zupanič Pajnič et al., 2025). While the petrous bone's efficacy for ancient DNA preservation in adults is well-documented, its utility for infant individuals remains underexplored. Neonatal petrous bones exhibit distinct developmental characteristics, including thinner cortical layers and incomplete ossification, which may influence DNA retention differently than in mature individuals. To address this gap, our research group previously conducted parallel analyses on perinatal remains from El Camp de les Lloses and El Poblat Ibèric de Burriac (Vinueza-Espinosa, 2022). As happened with adult remains, the amount of DNA extracted from the petrous bone in newborns was considerably higher than that extracted from vertebrae (Vinueza-Espinosa, 2022). In fact, the endogenous DNA yield from vertebrae was extremely low, making them far less effective for genetic analysis (Vinueza-Espinosa, 2022).

From the beginning of the thesis, our study design explicitly acknowledged the inherent challenges of recovering whole genome information from infant remains, particularly if the petrous bone was not available. However, it would have been a missed opportunity not to report any genetic information from some of the newborns, especially considering the novelty of analyzing such samples in the context of Iberian archaeology. Thus, for most critical samples, we focus our analytical resources on regions with the highest likelihood of successful retrieval, namely on the analysis of mitochondrial DNA (mtDNA).

5.1. Development of New Tools for Ancient Mitochondrial DNA Analysis

Post-mortem damage is one of the major challenges in ancient DNA research, and its impact becomes even more significant when working with low-coverage samples. While the nature of this damage has been extensively studied (Briggs et al., 2007; Brotherton et al., 2007; Olivieri et al., 2010; Sawyer et al., 2012), few studies have specifically investigated damage patterns in human mitochondrial DNA (mtDNA) to determine whether certain positions are more frequently affected than others, so-called damage hotspots. (Gilbert et al., 2003; Holland et al., 2021; Lamers et al., 2009; Rathbun et al., 2017). The few existing studies on damage hotspots have focused solely on the non-coding region (NCR) of the mtDNA. Article 1 of the thesis (Cuesta-Aguirre et al., 2025a) aimed to provide a more

comprehensive view of the human mitochondrial DNA damage profile by analyzing damage patterns across the entire human mitochondrial genome. MtDNA sequences from 427 ancient individuals, obtained through shotgun or capture techniques and sequenced using Next Generation Sequencing (NGS) technologies, were compiled. To establish a baseline for *post-mortem* damage levels, present-day samples and ancient samples treated with uracil-DNA glycosylase (UDG), which removes characteristic aDNA damage, were also compiled.

As expected, we found that the amount of damage was highest in non-UDG-treated samples, followed by partial-UDG-treated samples, and lowest in present-day samples. Unrepaired lesions at the time of collection (Sun et al., 2019), oxidative reactions during DNA extraction or other laboratory procedures (Xavier et al., 2018), suboptimal storage conditions (Holland et al., 2021; Rathbun et al., 2017), or sequencing errors (Stoler & Nekrutenko, 2021) may explain the presence of damage in present-day samples. Additionally, the persistence of some damage in total-UDG-treated ancient samples likely reflects the limitations of this treatment, which cannot remove the damage that occurs away from the ends of DNA fragments (Rohland et al., 2015).

The compiled non-UDG-treated samples were classified into five climatic regions — Polar (N=29), Cold (N=272), Temperate (N=89), Arid (N=29), and Tropical (N=8) — based on the climatic classification proposed by Peel et al. (2007). The goal was to determine whether different climatic regions not only affected the amount of DNA damage but also whether the damage pattern was similar. While the amount of damage followed the expected trend, lowest in Polar and highest in Tropical samples, statistically significant differences were observed only between the Cold and Temperate groups. This limited significance was interpreted as mainly due to sample size disparities: while more than 250 samples were available from Cold regions, the Polar, Arid, and Tropical categories had fewer than 30 samples each. This imbalance likely reflects the same publication bias discussed in the introduction (see *1.2. Bias in the ancient human genomes published*). Given these results, and considering the many additional factors that influence ancient DNA preservation and damage patterns (Bollongino et al., 2008; Dabney et al., 2013b; Gutiérrez-García et al., 2014; Molak & Ho, 2011; Olivieri et al., 2010; Poinar et al., 1996; Smith et al., 2001), all individuals were analyzed together, regardless of climatic origin, to obtain a more robust overall damage profile.

The damage profile included 11,735 base pairs (bp) out of 16,569 bp from the mitochondrial genome. This was significantly determined by the low depth of coverage

observed at the start and end positions of the mtDNA genome, attributed to the circular nature of mtDNA. Although two independent linear alignments were performed and subsequently merged to address this issue, this strategy did not fully resolve the coverage drop at the ends of the mtDNA since the raw sequencing data of the compiled individuals were not always available and, in many cases, only the aligned mtDNA sequences were publicly available.

From the 11,735 base pairs included, 2,933 bp were classified as damage hotspots. As expected, some of these overlapped with previously described mutational hotspots, heteroplasmic positions, or both (Li et al., 2015; Soares et al., 2009). Similar results were reported by Gilbert et al. (2003), who found overlaps between damage hotspots and modern datasets (Excoffier & Yang, 1999; Meyer et al., 1999). Nonetheless, 2,781 of the damage hotspots identified had not been previously described as either mutational or heteroplasmic hotspots. While these are likely due to DNA degradation affecting the entire molecule, sequencing errors cannot be completely ruled out (Peck et al., 2018). Out of 87 positions previously reported as damage hotspots in the literature, only 42 met the threshold to be considered hotspots according to the criteria we applied. Interestingly, many of these were already known as mutational or heteroplasmic hotspots (Li et al., 2015; Soares et al., 2009). A major strength of our analysis lay in the sample size: each position in the profile was supported by data from at least 321 ancient individuals. In contrast, previous studies used much smaller datasets. Gilbert et al. (2003) analyzed 43 teeth from 34 archaeological skeletons; Lamers et al. (2009) worked with only 6 to 40 samples; Rathbun et al. (2017) used 54 samples derived from three modern donors; and Holland et al. (2021) analyzed 240 samples from dilutions of just ten individuals. These differences in sample size may certainly account for some discrepancies in the identified hotspots. In addition, technological differences are likely contributors. Gilbert et al. (2003) and Lamers et al. (2009) used cloning technology, whereas our study was based on NGS. As for Rathbun et al. (2017) and Holland et al. (2021), their work relied on PCR amplification of modern DNA samples artificially damaged by storage under various conditions. Their bioinformatic approach for mtDNA analysis did not remove duplicate reads and failed to replicate the complex decomposition processes that occur naturally in archaeological remains.

For their analysis, Holland et al. (2021) employed the PowerSeq® CRM Nested System kit (Promega). This kit amplifies 10 small fragments covering the non-coding region (NCR) of the mitochondrial genome, including the three hypervariable regions (HVI, HVII, and HVIII), in a single multiplex reaction designed for NGS. The software they used for data analysis, the

GeneMarker® HTS software (SoftGenetics), was not open-access and had been designed for modern DNA, not ancient samples (Holland et al., 2017). As a result, it failed to remove PCR duplicates — an essential step in ancient DNA analysis (Hofreiter et al., 2021) — and, more critically, it did not account for *post-mortem* damage, which meant the authenticity of the samples could not be properly assessed. Our group successfully adapted the lab protocol of the PowerSeq® CRM Nested System kit (Promega) to be employed with degraded samples. This was possible by modifying the number of PCR cycles and the extension times (Vinueza-Espinosa et al., 2023). This kit presents advantages compared to other PCR based systems to analyze damaged mtDNA and we routinely use it in our lab.

There were pipelines available for processing ancient DNA generated by NGS, such as the nf-core/eager pipeline (Ewels et al., 2020; Fellows Yates et al., 2021; Peltzer et al., 2016). However, as of its latest version (v2.5.3), it does not yet offer a tool specifically for removing duplicates based on identical coordinates (resulting from amplicon-based libraries), which is critical for aDNA analysis. A published pipeline had been specifically designed for ancient mtDNA (Diroma et al., 2021), but it was not intended for use with amplicon-based data. Considering all these limitations, a custom, user-friendly pipeline tailored for the analysis of data generated by the PowerSeq® CRM Nested kit was generated for this thesis (Article 2 - Cuesta-Aguirre et al., 2024). This pipeline would allow proper filtering of duplicates and damage-aware authentication of ancient DNA, ensuring the reliability of the results. For this, 148 ancient Spanish human samples were sequenced following the protocol for degraded samples described by Vinueza-Espinosa et al. (2023), and 15 samples from members of the research group, that had been sequenced as contaminant control, were used.

The first step in pipeline design involved deciding how to effectively remove duplicates. Most studies rely on MarkDuplicates from Picard, a tool initially designed to mark the duplicated reads but later updated to also remove them. This tool, along with others such as Samtools' 'markdup' command (Danecek et al., 2021) and DeDup software (Peltzer et al., 2016), removes duplicates based on the coordinates of the reads. In shotgun sequencing for ancient DNA, where there are few molecules available, the likelihood that two molecules will start at the same base is almost nonexistent, so those tools work perfectly. In contrast, when working with amplicons, all reads from the same amplicon start and end at the same base. Therefore, coordinate-based duplication removal is not appropriate. In contrast, fastp is a versatile tool designed for various preprocessing tasks, including adapter removal, read trimming, and duplicate removal based on sequence similarity rather than read coordinates

(Chen et al., 2018). While other tools are known to be more effective at removing adapters and merging reads (Lien et al., 2023), fastp was the only one that allowed for duplicate removal in the manner required for amplicon-based data. Therefore, this tool was selected for both adapter removal and duplicate filtering. However, we have been considering the employment of other tools for removing the adapters in future versions of the pipeline. Using MarkDuplicates resulted in the removal of over 99% of the reads. Instead, fastp removed only about 30% of the reads as duplicates. These differences were statistically significant, as expected given the distinct characteristics of each tool. Although both tools retained some true duplicates, MarkDuplicates also removed some valid molecules that had the same coordinates but differed in sequence, as confirmed by manual analysis using the Integrative Genome Viewer (IGV) v2.11.2 (Robinson et al., 2011).

The next step in the pipeline was to authenticate the results, specifically identifying damage patterns. Tools like MapDamage2 (Jónsson et al., 2013) and Schmutzi (Renaud et al., 2015) are commonly used for this purpose, but they are less effective for amplicon analysis, as they primarily detect misincorporation patterns at the sequence ends. In the case of amplicon sequencing, these methods can misclassify samples as non-ancient, since the ends of the molecules are often not reached, and therefore, C-to-T or G-to-A misincorporation patterns are not observed. PMDtools, however, does not limit DNA damage detection to the sequence ends (Skoglund et al., 2014a), providing a viable method for authenticating aDNA sequenced by NGS from amplicons. To assess damage in the NCR of mtDNA, different *Post-mortem* Molecular Damage Scores (PMDS) — namely -1, 0, 1, 2, and 3 — were tested to distinguish between ancient and present-day samples. A PMDS of "-1" was used as a control since it considered reads affected by *post-mortem* molecular damage, PCR errors, and sequencing misreads as damage (Skoglund et al., 2014a). PMDS values of 0 and 3 were excluded because the percentage of recovered damaged reads overlapped respectively between the maximum observed in present-day samples and the minimum observed in ancient samples. However, PMDS of 1 and 2 were optimal to distinguish between present-day and ancient samples. Skoglund et al. (2014a) had previously suggested that high PMDS thresholds work well for very ancient samples (over 5,000 years old), but these thresholds would remove nearly all sequences from more recent ancient samples (around 100 years old), so the PMDS of 1 was selected to ensure an adequate number of reads for downstream analysis. Using this PMDS value, it was determined that a sample with 80% of the NCR recovered could be considered ancient with 95% confidence if the percentage of

damage was higher than 9.17%. For values lower than this threshold, caution is advised due to the possibility of present-day contamination.

Since some authors (Gilbert et al., 2003; Holland et al., 2021; Lamers et al., 2009; Rathbun et al., 2017), including us (Article 1 - Cuesta-Aguirre et al., 2025a) described that damage is not randomly distributed, differences between amplicons were studied and a negative correlation between depth coverage and the percentage of damage in both ancient samples and present-day samples was found. These findings could be explained by the possibility that damage accumulates at the binding sites of some primers, potentially acting as a hotspot of damage and hindering primer binding (Gilbert et al., 2003; Holland et al., 2021; Lamers et al., 2009; Rathbun et al., 2017; Vinueza-Espinosa et al., 2023).

Another important factor to consider was the presence of mixed bases. When mixed bases appear in a low-coverage mitochondrial genome, they may result from *post-mortem* molecular damage, heteroplasmy, or contamination. In amplicon-based datasets, the depth of coverage tends to be high, making *post-mortem* molecular damage an unlikely cause of mixed bases (Vinueza-Espinosa et al., 2023). Heteroplasmy — the presence of different types of mtDNA within the same individual (Li et al., 2015; Ramos et al., 2013; Santos et al., 2008) — is a natural phenomenon that can vary among individuals, populations, and even among tissues of the same individual (Irwin et al., 2009; Li et al., 2015; Naue et al., 2015; Ramos et al., 2013; Santos et al., 2008). In the only study that analyzed the non-coding region using bone samples, 19% of individuals (17 out of 87) showed at least one heteroplasmic site: fourteen individuals had one, two had two, and one individual had three heteroplasmic positions (Naue et al., 2015). So, when more than one mixed base is observed in a sample, a more detailed investigation is necessary, as this likely indicates contamination. To detect potential contamination, the pipeline was developed to provide the haplotypes generated from BAM files containing only damaged reads. The comparison of haplogroups and haplotypes inferred from this subset to those obtained using the whole BAM file can help assess the presence of contamination. Additionally, visualizing alignments in IGV (Robinson et al., 2011) is recommended because it allows a more detailed examination of regions with multiple mixed bases, where distinct patterns or even multiple mitochondrial profiles may be identified.

The last consideration in the pipeline was related to the circularity of mtDNA. The circular nature of the mtDNA molecule generates an alignment problem in the region where the start (base pair 1) and the end (base pair 16569) merge (Ayad & Pissis, 2017), as we

observed while trying to generate the damage profile (Article 1 - Cuesta-Aguirre et al., 2025a). The alignment in this region is of lower quality, and some reads are depreciated due to alignment being made through a linear molecule. To solve these problems, other authors used tools such as CircularMapper (Peltzer et al., 2016) or had the first 500 bp at the end of the reference mtDNA molecule (Diroma et al., 2021). In this pipeline, an easier approach was taken: the NCR sequence was linearized from base 15901 to base 700 to avoid the circularity of the mtDNA molecule, allowing all reads to align over it without the circularity problem.

This pipeline has been proven to work for data obtained by the PowerSeq® CRM Nested kit (Promega) (coverage: 16013-16569; 1-592). It should work without any modification with the ForenSeq mtDNA Control region kit (Verogen) (coverage: 16008-16569; 1-595) and the precision ID mtDNA Control Region Panel (ThermoFisher Scientific) (coverage: 15954-16569;1-610) since the reference sequence employed covered from position 15901 to position 700. Moreover, for use in another set of amplicons, the pipeline should be modified by changing the reference sequence, altering the starting coordinates, and re-editing the coordinates restructuration based on the origin of replication. For non-overlapped amplicons, a major adaptation of the pipeline should be done, and future versions of the pipeline may explore these aspects.

This pipeline has been used since its development, with consistently good results. As already mentioned, it was applied in Articles 3 and 4 of this thesis to the samples that lacked petrous bones or did not yield enough DNA for whole genome sequencing (WGS). Moreover, it has also been extensively used in our lab to obtain the mtDNA profile from individuals retrieved from mass graves of the Spanish Civil War. Beyond these, it has been applied to various other projects, providing valuable insights into contamination levels and overall sample quality. One limitation, unrelated to the pipeline itself, arises when the DNA is highly degraded, resulting in extremely short fragments that prevent successful amplification.

5.2. Genetic Analysis of Northeast Iberian Communities

As mentioned before, the tools developed in Part 1 (Articles 1 and 2) of the thesis were employed, in addition to state-of-the-art paleogenomic and bioinformatic technologies, to reconstruct the genetic history of Iberian tribes of the Northeast of the Iberian Peninsula from the Early to the Late Iron Age (Articles 3 and 4). As previously mentioned, cremation was the primary funerary rite of the Iberians (Oliver Foix, 2004), resulting in very few skeletal remains

surviving to the present day, which has limited genetic analysis. However, it is known that there were some exceptions to cremation, as some newborns were found buried under the pavement of houses (Afonso et al., 2019; Crespo et al., 2011; Subirà & Molist, 2007). In the last few years, our group has conducted an interdisciplinary study to understand these burials. Morphological and dental histology techniques allowed to establish very precisely the chronological age of the newborns, providing solid evidence for a natural mortality pattern (Martirosyan et al., 2024). Moreover, newborns from both sexes were found in the studied Iberian sites (Article 3 - Cuesta-Aguirre et al., 2025b; Sandoval-Ávila et al., 2025), suggesting that these newborns are a representative sample of the population, allowing the employment of a paleogenomic approach to understand their genetic history.

Despite the few remains excavated from Iberian sites, some genetic studies have been performed using bones of uncertain origin found in some Iberian sites (Gamba et al., 2008; Olalde et al., 2019; Sampietro et al., 2005). However, as mentioned in the introduction (1.5. Overview of paleogenomic studies in the Iberian Peninsula), these analyses have focused on the hypervariable region I of the mtDNA-NCR, such as in the works of Gamba et al. (2008) and Sampietro et al. (2005), or have examined samples without considering their cultural context, as seen in Olalde et al. (2019) and Patterson et al. (2022). In contrast, Part 2 of this thesis (Articles 3 and 4) presents the first comprehensive genetic study of Iberian populations using NGS technology, offering unprecedented paleogenomic insights into the origins and evolution of Iron Age communities associated with the Iberian culture. We obtained 22 individuals with more than 20,000 SNPs from the 1,240k panel (Mathieson et al., 2015) and the mitochondrial DNA haplogroup for 51 Iberians.

The origins and evolution of Iberian communities based on archaeological findings have been debated for decades. The earliest model proposed that the Iberian culture emerged from a local substrate with ancient North African roots dating back to the Neolithic period. This population mixed with local groups from the southern Iberian Peninsula and was further shaped through interactions with the Greek colonies from the South. These people were believed to have later migrated along the Mediterranean coast of the Iberian Peninsula to the north, spreading Iberian culture (Bosch-Gimpera, 1938). However, this theory was refuted in the 60s and substituted by a model that implied the evolution of local populations with the influences of Mediterranean cultures such as Phoenicians and Greeks (Maluquer de Motes, 1966, 1982; Sanmartí, 2011, 2013; Tarradell, 1962). In the Northeast of the Iberian Peninsula, the smooth evolution of urbanization from pre-Iberian settlements to larger hierarchical entities agrees with this last theory (Bosch & Santacana, 2009; Junyent, 1996a,

1996b, 2023; Sanmartí, 2003). Our results support the idea that the transition from the Bronze Age to the Iron Age in the northeastern Iberian Peninsula occurred without significant external genetic influence, as also proposed by Olalde et al. (2019). This has also been documented in the North of the Iberian Peninsula (Papac et al., 2023) and the territory of present-day France (Brunel et al., 2020). Interestingly, we observed an increase in Steppe-related ancestry among Iron Age Iberians compared to the Bronze Age population from this region, a pattern also noted by Olalde et al. (2019). However, the Bronze Age dataset from the Northeast of the Iberian Peninsula was limited, comprising only six individuals from four different sites with considerable genetic variability, highlighting the need for more data from this period and region to confirm this pattern.

In general terms, our results for the Iberians revealed a high frequency of mitochondrial haplogroups H and U (including K but excluding U6), a moderate presence of J, T, and HV0, and low frequencies of M1, N1, and X. These findings broadly align with those of Sampietro et al. (2005), who analyzed 17 adult male samples from the Northeast of the Iberian Peninsula using cloning technology on fragmentary cranial remains not associated with funerary contexts. They also found H and U as the most frequent haplogroups, with smaller proportions of J, T, and Pre-HV. In contrast, Gamba et al. (2008) reported different haplogroup frequencies in Iberians from the central-eastern Iberian Peninsula (modern-day Valencian Community), with H, V, and W being the most frequent, and additional occurrences of T1 and K. We did not detect haplogroup V in our dataset. However, it is worth noting that their study predates the redefinition of certain haplogroups; the mutation at position 16298C, once used to define V, is now considered characteristic of HV0 (van Oven & Kayser, 2009). Interestingly, the frequency of HV0 observed in our study (Article 4) in the Cessetani tribe, which is geographically the closest to the area studied by Gamba et al. (2008), is comparable to their reported frequency of haplogroup V.

Although in our study (Article 4) no statistically significant genetic differentiation was detected between the mtDNAs of the Iberian tribes of the Northeast of the Iberian Peninsula, variations in haplogroup frequencies suggest a certain degree of biological distinctiveness. The lack of significant results is likely due to the still limited sample size available for each tribe. This pattern aligns with archaeological interpretations that emphasize tribal diversity in the region (Aranegui, 2012; González Reyero, 2012). Additionally, the slightly higher values of mtDNA diversity observed support the hypothesis of a patrilocal mating system during the Iron Age, as has been proposed for Bronze Age communities in the same area (Simón et al., 2011).

The genomic nuclear DNA results of the population of the fortress of Vilars (Ilergetae tribe), the oldest and westernmost of the populations studied in this thesis, aligned with other Iron Age populations of the Iberian Peninsula. Moreover, all mitochondrial lineages found in this population, or their ancestral lineages, had already been found in the Iberian Peninsula before the Iron Age. All of this implies continuity since the Bronze Age populations. This is consistent with the material culture recovered from the site, where most pottery corresponded to pre-Iberian or Iberian traditions, despite some Iberian pottery and jewelry showing Phoenician, Punic, and Greek influences (Alonso et al., 2004; Universitat de Lleida, 2020). While the genetic data based on nuclear DNA suggests limited Mediterranean ancestry in the general population, two individuals (VILA12 and VILA13) stand out due to their higher proportions of Mediterranean ancestry.

VILA12 is notably displaced from the Iron Age Iberian cluster in the PCA, instead clustering with Neolithic and Chalcolithic populations of the Iberian Peninsula. In the qpAdm-based PCA, this individual clusters near the Bronze Age populations from the Italian Peninsula and Late Iberians from the northeastern Iberian Peninsula. Interestingly, his mitochondrial haplogroup (H4a1a5) had not been detected in ancient samples before, although its basal haplogroup shows notable frequencies in Neolithic Sardinia (Fernandes et al., 2020; Marcus et al., 2020) and Chalcolithic and Bronze Age samples from the Iberian Peninsula (Olalde et al., 2019; Villalba-Mouco et al., 2021), among other European regions. Notably, haplogroup H4a1 was identified in the Iberian (Indiketes) site of Mas Castellar (I19722, 413-382 calBCE) (Olalde et al., 2019). In contrast, nuclear DNA made VILA13 clustered within the Iron Age Iberian group in the PCA, but in the qpAdm-PCA, he fell near Etruscans, as well as Romans from Sardinia and Empuries. His mitochondrial haplogroup (H1) is widespread in Western Europe, Northwest Africa, and the western Mediterranean islands (Achilli et al., 2004), with high frequencies reported in Phoenicians from Ibiza (Zalloua et al., 2018) and modern Sardinians (Olivieri et al., 2017). In both individuals, the Y-chromosome haplogroup was R1b, which is not particularly informative for identifying geographic origin, as this lineage spread widely across Europe during the Bronze Age (Myres et al., 2011; Olalde et al., 2018).

One particularly notable case is VILA14, the most recent individual recovered from the site (450-325 BCE), whose nuclear genome aligned closely with the Bronze Age individuals. VILA14 carried the mitochondrial haplogroup U5a1c1 and the Y-chromosome haplogroup H2a1, both suggesting genetic continuity from the Bronze Age. U5a1c lineage was found at relatively high frequencies among Balkan Hunter-Gatherers, and during the

Neolithic, it expanded into Central and Western Europe — this lineage has been identified in Late Neolithic Switzerland (Furtwängler et al., 2020), Chalcolithic Iberian Peninsula (Günther et al., 2015; Villalba-Mouco et al., 2021), and Bronze Age individuals from the Netherlands and Scotland (Olalde et al., 2018; Patterson et al., 2022). Similarly, the H2a1 Y-chromosome lineage also points to a Neolithic legacy. Ancient H2 lineages have been reported in a Sicilian individual during the Neolithic (Yu et al., 2022), but with higher frequencies among Early Neolithic individuals in Germany (n=19) (Childebayeva et al., 2022; Rivollat et al., 2020) and in Middle Neolithic France (n=10; 2 H2a1, the rest H2) (Brunel et al., 2020; Rivollat et al., 2020). During the Late Neolithic, H2a1 appeared in two individuals from the Czech Republic (Papac et al., 2021), and H2 appeared later during the Chalcolithic, in a family group from the southeastern Iberian Peninsula (Villalba-Mouco et al., 2021) and in another individual from the central/northern Iberian Peninsula (Valdiosera et al., 2018). Interestingly, individual CDP006 (3300–2300 BCE) from the southeastern Iberian Peninsula (Villalba-Mouco et al., 2021) carried both mtDNA U5a1c1a and Y-chromosome H2. This coincidence suggests a long-standing local presence of these lineages in the Iberian Peninsula.

Thus, the genetic continuity observed at the fortress of Vilars, reflected in both mitochondrial and nuclear DNA, appears to relate not only to the site's greater antiquity but also to its geographic position, farther from the Mediterranean influence that impacted other Iberian populations in the northeast of the Iberian Peninsula.

At the Iberian city of Ullastret (Indiketes tribe), composed of two urban centers, Illa d'en Reixac and Puig de Sant Andreu, both of which with Mediterranean architectural influences (López i Melcion et al., 1989; Martín i Ortega, 1985; Martín i Ortega et al., 2016), genetic continuity during the Iron Age can also be observed. In Illa d'en Reixac, mitochondrial haplogroups were recovered from three individuals: H2a2a1, I1a1, and T1a1'3. H2a2a1 has been previously identified in the Iberian Peninsula, appearing more frequently than I1a1, a haplogroup restricted to Europe that emerged around 7,000 years ago and that is associated with the Neolithic agricultural expansion (Olivieri et al., 2013). In contrast, T1a1'3 had not been detected in the Iberian Peninsula before, but T1a1, the most common lineage within T1a*, which comprises around 80% of the group, is linked to early Neolithic farming migrations (Pala et al., 2012). In Puig de Sant Andreu, several haplogroups commonly found in the Iberian Peninsula were identified, including H1, J1c, J2b1a, K1a, and X2b. An exception is T2, not present in the Iberian Peninsula, but with sublineages reported in a Bronze Age individual (Villalba-Mouco et al., 2021). Overall, the haplogroup profiles recovered at Ullastret suggest that the female individuals were of local origin. Nuclear DNA has been previously

analyzed in only four individuals from the site (Olalde et al., 2019), with only one showing minor traces of Mediterranean ancestry (Article 3 - Cuesta-Aguirre et al., 2025b).

Iron Age continuity is also evident at the site of Sant Miquel d'Olèrdola (Ausetani tribe), although a higher proportion of Mediterranean ancestry was detected when compared to the fortress of Vilars. This difference likely reflects the site's geographic location: Sant Miquel d'Olèrdola is closer to the coast and connected to the Mediterranean via the Vilafranca stream, which leads to two nearby Iberian port sites (Darró and Sitges) located approximately 15–17 km away. This coastal access likely facilitated contact with Mediterranean populations and helps explain the increased Mediterranean genetic influence.

Among the individuals analyzed, OLE1 stood out as the only one without nuclear Mediterranean ancestry. Her mitochondrial haplogroup (HV0+195) is widely represented in the Iberian Peninsula, consistent with a local origin. In contrast, OLE03 and OLE04 both lacked Steppe-related ancestry and showed a high proportion of Mediterranean ancestry. OLE03, in particular, exhibited extremely high levels of Mediterranean ancestry and clustered in the qpAdm-based PCA with Punic and Roman individuals from Sardinia. Despite this, her mitochondrial haplogroup (HV0) is common in the Iberian Peninsula, suggesting potential local maternal ancestry combined with recent Mediterranean genetic input.

Analyzing nuclear DNA, OLE04 fell near the Iberian Peninsula Bronze/Iron Age cluster in the qpAdm-based PCA and carried the mitochondrial haplogroup N1a1a1a3, a rare haplogroup present in a Neolithic French individual (Rivollat et al., 2020), supporting a potential Neolithic arrival to the Iberian Peninsula. However, it could also have arrived later by commercial contacts (Albizuriet al., 2011; Moreno Gil et al., 2025). OLE06 also presented high nuclear Mediterranean ancestry, potentially linked to North Africa or the eastern Mediterranean, and a mitochondrial haplogroup belonging to M1b, primarily found in North Africa (González et al., 2007; Winters, 2010). We suggested that a woman with Punic/Carthaginian ancestry may have introduced it to Sant Miquel d'Olèrdola during the Iron Age. This theory could be supported by the fact that many Punic (Punic-Ebusitans and Central Mediterranean) amphorae and ceramics have been found through the Iberian stratigraphy of this site (Molist Capella & Bosch Casadevall, 2002). Ancient DNA evidence for M1b lineages before the Iron Age was limited to three Moroccan individuals: two Paleolithic individuals (van de Loosdrecht et al., 2018) and the other from the Early Neolithic (Fregel et al., 2018). Today's distribution is largely restricted to Morocco, according to EMPOP v5/R12 (Huber et al., 2018) (N=19). The rarity of this lineage should be noted. Its presence in an Iron Age individual from

the northeastern Iberian Peninsula reflects probable female mobility linked to Mediterranean exchange networks. Interestingly, individuals OLE04 and OLE06 had previously been interpreted as potential twins due to their discovery near a wall, positioned in opposite directions, with their legs intermingled and no sediment separating the two skeletons (Crespo et al., 2011). However, genome-wide analysis (Article 3 - Cuesta-Aguirre et al., 2025b) revealed that they were unrelated. Furthermore, as previously indicated, the two individuals carried different mitochondrial haplogroups (N1a1a1a3 and M1b), effectively ruling out the possibility that they were twins. Notably, both haplogroups likely have non-local origins, suggesting that this joint burial may reflect cultural or ritual practices rather than biological kinship as suggested in other similar cases (Risch et al., 2023; Yaka et al., 2021).

Based on whole genome data, it has been seen that Iberians during the Iron Age showed a notable proportion of Bell Beaker North/Central European ancestry, associated with Steppe-related components, as mentioned in Article 3 (Cuesta-Aguirre et al., 2025b) and by Olalde et al. (2019). This pattern was also observed in Celtic individuals from northern Iberia (Papac et al., 2023). However, our study identifies for the first time Mediterranean ancestry in Iberians, revealing previously undetected gene flow from Mediterranean populations. Our results confirm that Iberians interacted genetically with neighboring Mediterranean cultures, including the Phoenician, Punic/Carthaginian, and Italian groups. These contacts, linked to archaeological evidence, contributed to the genetic diversity of Iron Age Iberian populations. Despite this influx, the core genetic identity of the Iberians remained stable, as seen in other Iron Age populations across Europe (Antonio et al., 2024; Brunel et al., 2020). The absence of Mediterranean ancestry among Celtic individuals from inland Iberia (Papac et al., 2023) aligns with expectations. Mediterranean ancestry was not specifically investigated for the Olalde et al. (2019) samples. We reanalyzed those individuals to specifically detect that component, which we identified in four cases: three newborns from the Valencian Community and one adult from Puig de Sant Andreu (Ullastret).

Nuclear and mtDNA data from other northeastern Iberian sites, with a limited number of samples per site, also provided evidence for continuity from the Bronze Age with some Mediterranean influence. In Castell de Besora (Bergistani tribe), only mtDNA was obtained for two newborns, both belonging to local mtDNA haplogroups (HV0 and H6a1b2). At Can Roqueta-Can Revella (Layetani tribe), all mtDNA haplogroups were identified as local (Article 4, Olalde et al., 2019). The same happened in Cessetani Iberian sites of Horta d'en Grimau and Mas d'en Boixos, which revealed only local maternal lineages (e.g., H, H3, HV0+195, and J1c1). Nuclear DNA data supported a local genetic profile in the last three sites (Article 3 -

Cuesta-Aguirre et al., 2025b; Olalde et al., 2019). It must be noted that the uncontextualized remains employed for whole-genome analysis and the limited number of individuals may influence the apparent absence of non-local ancestry at these sites. By contrast, at Font de la Canya (Cessetani tribe), an Iberian site whose economy centered on cereal storage and distribution, and which also provides the earliest evidence of wine production in the region likely linked to Phoenician contact, the remains of a female individual were analyzed and found to carry mitochondrial haplogroup K1a12a (Olalde et al., 2019). This lineage had not previously been reported in the Iberian Peninsula, but has been documented in regions such as Turkey, Iran, Azerbaijan, and Armenia from the Neolithic through the Iron Age (Kılınç et al., 2016; Lazaridis et al., 2016, 2022; Skourtanioti et al., 2020). Although nuclear data placed this individual within the genetic variability of other Iron Age Iberians in Principal Component Analyses (Article 3 - Cuesta-Aguirre et al., 2025b; Olalde et al., 2019), and her admixture proportions were similar to those of other contemporaneous samples (Olalde et al., 2019), her ancestry could not be successfully modeled using the qpAdm frameworks that worked for other Iberian individuals in (Article 3 - Cuesta-Aguirre et al., 2025b). The presence of K1a12a may reflect the arrival of this lineage via Phoenician female mobility during the Late Bronze Age or Early Iron Age.

The arrival of the Romans marked a gradual yet profound transformation of Iberian communities, both culturally and genetically. At El Camp de les Lloses (125 BCE – 50 CE, Ausetani tribe), this change is reflected in the genomic profiles of the individuals analyzed. Although f-statistics showed no strong differentiation from earlier Iberians, PCA revealed a distinct genetic profile, suggesting continuity with subtle shifts in ancestry. The qpAdm analysis further supports this, indicating increased affinity with Bronze and Iron Age populations from the Italian Peninsula. Notably, none of the Late Iberians from El Camp de les Lloses clustered with the Iberian Peninsula's Iron/Bronze Age cluster in the qpAdm-based PCA, underlining this shift. This genetic transformation aligns well with the archaeological context of El Camp de les Lloses, a site associated with metalworking and logistical activities tied to the Roman military (Duran i Caixal et al., 2017). The fusion of Iberian and Roman-Italian cultures is clear in the site's architecture, presenting Roman-Italian styles (rectangular houses with central courtyards), while employing indigenous building materials (stone bases and earthen walls) (Duran i Caixal et al., 2017). Burial practices at the site align with Iberian traditions (e.g., cremation), and pottery assemblages include a mix of Iberian and Roman-Italian artifacts (Duran i Caixal et al., 2017). Additionally, most of the coins recovered (90%) are of Iberian origin, although Roman and Greek coins from Massaliot (present-day Marseille)

have also been found (Duran i Caixal et al., 2017). The presence of pottery from Ebusus (a Phoenician city in present-day Ibiza, Balearic Islands), the Tripoli region (North Africa), and the southern Iberian Peninsula (around the Strait of Gibraltar) (Duran i Caixal et al., 2017) suggests that North African and Eastern Mediterranean ancestries could also have been introduced through Phoenician-Punic or Carthaginian contacts.

At the individual level, genome-wide data revealed proportions of Iranian Neolithic or North African Neolithic ancestry in all the individuals except CL12, who presented a Mediterranean component without Iranian-like or North African ancestries. Moreover, her haplogroup (H3+152) indicated local ancestry (Achilli et al., 2004; Costa et al., 2013). In contrast, CL1 had the strongest Iranian Neolithic influence, and CL8 carried North African ancestry. Yet, both individuals presented mitochondrial lineages (K1b1a and H3, respectively) typical of ancient and modern populations in the Iberian Peninsula and Western Europe (Achilli et al., 2004; Costa et al., 2013). This discrepancy suggests that the observed Mediterranean and North African genomic components may have been introduced primarily via the paternal line.

The two sisters, CL6 and CL7, carried the mitochondrial haplogroup H1at1, a lineage previously identified only in two Iron Age Etruscans (775–201 calBCE) (Posth et al., 2021) and one Iron Age individual from England (371–97 calBCE) (Patterson et al., 2022). The rarity of this haplogroup makes it difficult to trace its precise origin, but the most likely source appears to be the Italian Peninsula. The qpAdm analysis of CL7, based on nuclear data, revealed strong genetic similarity with Bronze Age populations from Italy and VILA12 (Ilergetae dated between 775–650 BCE), suggesting a potential connection. The qpAdm results for CL6 indicated the presence of Eastern Mediterranean ancestry, raising the possibility that the haplogroup may have been introduced by the Romans. However, since the Mediterranean-related ancestry at El Camp de les Lloses appears to have been introduced mainly via the paternal line (as seen with mtDNA data in CL1, CL10, CL5, CL8, CL9, CL10, and CL12), it seems more plausible that this maternal lineage (H1at1) was brought earlier, possibly by the Etruscans.

CL2 — who carried the mitochondrial haplogroup M1a3a, most commonly found in North and East Africa — did not exhibit any detectable North African ancestry in their nuclear genome. The present-day distribution of M1a3 closely resembles the pattern observed for M1b, with a predominance in Northwest Africa (N=27) according to EMPOP v5/R12 (Huber et al., 2018). Notably, CL2 is the earliest known individual carrying the M1a3a haplogroup. The

only other identified case is a Roman-era female from the western Iberian Peninsula (ID_25522; 257–531 CE), previously reported by Roca-Rada et al. (2024). All of this information suggests that M1a3a may have been introduced in an earlier migratory episode, remaining at extremely low frequencies until its detection in this individual.

Haplogroup H33 — the ancestor of H33c, found in CL11 — has only been reported in a non-local Hellenistic male from Empuries (376-201 BCE) (Olalde et al., 2019) and is present at low frequencies in Central Europe, Southeast Europe, and Southwest Asia, with 11 individuals in EMPOP v5/R12 (Huber et al., 2018). While it is tempting to link this lineage to Aegean migrations, the lack of nuclear data in CL11 prevents a definitive conclusion. Nevertheless, its presence underscores the broader Mediterranean interactions that took place in the region.

The most recent individual from El Camp de les Lloses, CL4 (50 BCE – 50 CE), presented a high proportion of Italian ancestry in his nuclear genome and clustered closely with Etruscans in the qpAdm-based PCA. His mitochondrial haplogroup was U4d2, a lineage not previously reported in Western Europe (N=6 in the AADR, Mallick et al., 2024). U4d lineages have previously been detected in the Pontiac Steppe during the Neolithic and in North and Central Europe during the Bronze Age (Mittnik et al., 2018; Olalde et al., 2018; Skoglund et al., 2014b). The high proportion of Italian ancestry in CL4 suggests that U4d2 may have arrived at El Camp de les Lloses through Roman-mediated gene flow. However, the possibility that this lineage was already present in the Iberian Peninsula since the Bronze Age, or that it was introduced during the Iron Age, albeit at a very low frequency, cannot be excluded.

This part of the thesis provides the first paleogenomic insights into the origin and evolution of Iberian communities during the Iron Age, focusing on newborns buried beneath domestic and productive areas. The genetic data support the archaeological hypothesis that Iberian culture emerged from local Bronze Age populations and evolved gradually throughout the Iron Age. While Mediterranean interactions played a role in shaping Iberian society, the genetic identity, which was not identical for all Iberian tribes, of the populations remained largely local. The analysis of nuclear and mitochondrial DNA haplogroups reveals a genetic landscape marked by continuity, with most lineages — or their ancestral lineages — already present in the Iberian Peninsula during earlier periods. With the arrival of the Romans, however, new genetic components were introduced, contributing to a more diverse Iberian-Roman population and reflecting broader social and political transformations. Together,

these results highlight the enduring yet dynamic genetic legacy of Iberian communities. Genetic influences during the Iron Age appear to have involved both males and females — possibly linked to commercial trade networks — but in the initial phases of Roman conquest, male-mediated gene flow predominated, at least in the case of El Camp de les Lloses.

5.3. Limitations

Out of the 69 Iberian newborns analyzed, no genetic information was recovered for 18 individuals, despite different bones, and a couple of extractions per bone were employed. Namely, 7 from the fortress of Vilars, 5 from Illa d'en Reixac, 1 from Puig de Sant Andreu, 2 from Sant Miquel d'Olèrdola, and 1 from El poblat Ibèric de Burriac. This could be related not only to excavation and conservation conditions, as commented before, but also to site-specific factors such as soil composition, environmental conditions, and microclimate. Unfortunately, no genetic information could be recovered from the only newborn found at El poblat Ibèric de Burriac, despite testing multiple bones and applying various library preparation techniques.

Another limitation was the poor recovery of Y-chromosome haplogroups. No Y-chromosome data was retrieved from Sant Miquel d'Olèrdola, the Iberian site with the highest proportion of Mediterranean ancestry. This information would have been valuable to complement the autosomal and mitochondrial DNA data, potentially offering insights into male-driven ancestry from beyond the Iberian Peninsula. In the other sites where Y-haplogroups could be determined, the low coverage limited our ability to compare them meaningfully. Although most individuals belonged to haplogroup R1b, higher coverage might have allowed us to assess whether they shared common lineages or represented different paternal origins.

As previously described, the research group had been primarily focused on paleogenetic analyses in the years preceding this project. This thesis marked a transition from paleogenetics to paleogenomics, which required substantial investment in infrastructure and the optimization of laboratory protocols. The DNA extraction method had already been optimized (Vinueza-Espinosa et al., 2020), laying the groundwork for subsequent steps. The next critical task involved selecting and optimizing a suitable library preparation method for ancient DNA. At the time, the Blunt-End Single-Tube (BEST) protocol represented the most promising option (Carøe et al., 2018). Although the Santa Cruz Reaction

(SCR) protocol (Kapp et al., 2021) was published during the optimization phase of BEST, the necessary reagents for the BEST protocol and equipment had already been acquired, and the protocol was beginning to yield reliable results. Therefore, the decision was made to continue with the BEST protocol. In 2022, two commercial kits for hybridization-based capture became available (Rohland et al., 2022), and there were initial plans to test the Arbor kit on selected samples during a research stay at the University of Adelaide. However, a bias associated with the Arbor kit was soon identified (Davidson et al., 2023), and the capture strategy was ultimately abandoned in favor of continuing with shotgun sequencing alone.

I hope that the work presented in this thesis contributes to a solid foundation of paleogenomic knowledge, which future PhD generations can build upon. By doing so, they will be able to more easily advance in the field and perform analyses that were not possible within the scope of this project.



6.

CONCLUSIONS

6. CONCLUSIONS

This thesis provides new insights into ancient DNA analysis by combining methodological advancements with paleogenomic interpretations of Iron Age Iberian populations. The main conclusions are:

- A *post-mortem* molecular damage profile on human mitochondrial DNA was obtained from genomic analysis. A total of 2,933 damage hotspots were identified, 2,781 of which had not been previously reported as damage, mutational, or heteroplasmic hotspots. However, some of the positions reported as damage hotspots may be a result of sequencing errors since sequencing errors cannot be precisely identified.

- An easy-to-use bioinformatic pipeline was developed to analyze the non-coding region of mtDNA sequences obtained with amplicon-based Next Generation Sequencing methodologies. This tool provides users with a contamination estimation, the mitochondrial haplogroup, and the haplotype of the samples.

- In the bioinformatic pipeline developed to analyze the non-coding region of mtDNA, fastp v0.23.2 was included since it was confirmed to be the best tool to remove duplicates from Next Generation Sequencing mtDNA data derived from amplicon-based methods.

- In the bioinformatic pipeline developed to analyze the non-coding region of mtDNA, a PMDS threshold of 1 in the bioinformatic tool PMDtools v0.23.2 was validated as an effective cutoff to distinguish ancient from modern DNA in non-coding region mitochondrial datasets based on amplicons, showing that authentic ancient samples can still present as little as 9.17% damaged reads.

- This study provides the first paleogenomic insights into the origin and evolution of Iron Age Northeast Iberian communities belonging to Iberian culture, focusing on newborns interred under the pavements of houses and productive areas. Our findings confirm the archaeological theory that Iberian culture emerged from local Bronze Age populations, evolving gradually through the Iron Age, highlighting the complex interplay of migration, commercial trade, and cultural interactions that shaped the genetic heritage of the region.

- Subtle differences in mitochondrial DNA haplogroup frequencies were observed among Iberian tribes, although genetic differentiation was not statistically significant. Moreover, mitochondrial DNA diversity is relatively high in all the Iberian tribes, being consistent with patrilocal mating systems and small distance female migration that could have prevented high differentiation among tribes.

- Evidence of genetic contact with other Mediterranean populations from regions such as the central/eastern Mediterranean and North Africa was detected through complementary information obtained from nuclear and mitochondrial DNA. While both males and females participate in gene flow, genetic continuity persisted among the Iberians. Most of the detected mitochondrial haplogroups, or their ancestral lineages, were already present in the Iberian Peninsula during the Iron Age, suggesting that most Iberian women had local origin. However, distinct maternal lineages linked to other Mediterranean populations reveal clear signs of long-distance female migration during the Iron Age.

- With the arrival of the Romans, new genetic influences were introduced as evidenced in El Camp de les Lloses, leading to a more diverse Iberian-Roman population alongside significant social and political transformations.

- The genetic analysis of the Late Iberian Ausetani individuals from El Camp de les Lloses showed a mix of Mediterranean (central/eastern Mediterranean and North Africa) and local ancestries in their nuclear DNA, with the majority of mitochondrial haplogroups likely pre-dating the Roman arrival, suggesting they were already present in the region. This supports the idea that while male soldiers or artisans may have come from outside the Iberian Peninsula, the women from this site were likely mainly local inhabitants.



7.

BIBLIOGRAPHY

7. BIBLIOGRAPHY

Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E., Carelli, V., Moral, P., Dugoujon, J.-M., Roostalu, U., Loogväli, E.-L., Kivisild, T., Bandelt, H.-J., Richards, M., Villems, R., Santachiara-Benerecetti, A. S., Semino, O., & Torroni, A. (2004). The Molecular Dissection of mtDNA Haplogroup H Confirms That the Franco-Cantabrian Glacial Refuge Was a Major Source for the European Gene Pool. The American Journal of Human Genetics, 75(5), 910–918. https://doi.org/10.1086/425590

Afonso, C., Nociarova, D., Santos, C., Martinez-Labarga, C., Mestres, I., Duran, M., & Malgosa, A. (2019). Sex selection in late Iberian infant burials: Integrating evidence from morphological and genetic data. American Journal of Human Biology, 31(1). https://doi.org/10.1002/ajhb.23204

Agustí, B., Alonso, N., Garcés, I., Junyent, E., Agusti, B., Alonso, N., Garcés, I., & Junyent, E. (2000). Una inhumación múltiple de perinatales en la fortaleza de Els Vilars (Arbeca, Lleida) y las prácticas de enterramiento en hábitat durante la primera Edad del Hierro en el valle del Segre (Cataluña). Archéologie de La Mort, Archéologie de La Tombe Au Premier Âge Du Fer. Colloque International de l'AFEAF (Conques-Montrozier 8-11 Mai 1997), January, 305–324.

Albizuri, S., Alonso, N., & Cachero, F. J. (2011). Economia i canvi social a Catalunya durant l'edat del bronze i la primera edad del ferro. In S. Valenzuela Lamas, N. Padrós i Font, M. C. Belarte Franco, & J. Sanmartí Grego (Eds.), Economia agropecuària i canvi social a partirde les restes bioarqueològiques.El primer mil·lenni aC ala Mediterrània occidental: Actes de la V Reunió Internacional d'Arqueologia de Calafell (Calafell, 16 al 18 d'abril de 2009) (pp. 11–36).

Alföldy, G. (1996). Spain. In A. K. Bowman, E. Champlin, & A. Lintott (Eds.), The Cambridge Ancient History (2nd ed., Vol. 10) (pp. 449–463). Cambridge University Press.

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., Campos, P. F., Samaniego, J. A., Gilbert, T. P. M., Willerslev, E., Zhang, G., Scofield, R. P., Holdaway, R. N., & Bunce, M. (2012). The half-life of DNA in bone: Measuring decay kinetics in 158 dated fossils. Proceedings of the Royal Society B: Biological Sciences, 279(1748), 4724–4733. https://doi.org/10.1098/rspb.2012.1745

Allentoft, M. E., Sikora, M., Refoyo-Martínez, A., Irving-Pease, E. K., Fischer, A., Barrie, W., Ingason, A., Stenderup, J., Sjögren, K.-G., Pearson, A., Sousa da Mota, B., Schulz Paulsson, B., Halgren, A., Macleod, R., Jørkov, M. L. S., Demeter, F., Sørensen, L., Nielsen, P. O., Henriksen, R. A., Vimala, T., McColl, H., Margaryan, A., Ilardo, M., Vaughn, A., Fischer Mortensen, M., Nielsen, A. B., Ulfeldt Hede, M., Johannsen, N. N., Rasmussen, P., Vinner, L., Renaud, G., Stern, A., Jensen, T. Z. T., Scorrano, G., Schroeder, H., Lysdahl, P., Ramsøe, A. D., Skorobogatov, A., Schork, A. J., Rosengren, A., Ruter, A., Outram, A., Timoshenko, A. A., Buzhilova, A., Coppa, A., Zubova, A., Silva, A. M., Hansen, A. J., Gromov, A., Logvin, A., Gotfredsen, A. B., Henning Nielsen, B., González-Rabanal, B., Lalueza-Fox, C., McKenzie, C. J., Gaunitz, C., Blasco, C., Liesau, C., Martinez-Labarga, C., Pozdnyakov, D. V., Cuenca-Solana, D., Lordkipanidze, D. O., En'shin, D., Salazar-García, D. C., Price, T. D., Borić, D., Kostyleva, E., Veselovskaya, E. V., Usmanova, E. R., Cappellini, E., Brinch Petersen, E., Kannegaard, E., Radina, F., Eylem Yediay, F., Duday, H., Gutiérrez-Zugasti, I., Merts, I., Potekhina, I., Shevnina, I., Altinkaya, I., Guilaine, J., Hansen, J., Aura Tortosa, J. E., Zilhão, J., Vega, J., Buck Pedersen, K., Tunia, K., Zhao, L., Mylnikova, L. N., Larsson, L., Metz, L., Yepiskoposyan, L., Pedersen, L., Sarti, L., Orlando, L., Slimak, L., Klassen, L., Blank, M., González-Morales, M., Silvestrini, M., Vretemark, M., Nesterova, M. S., Rykun, M., Rolfo, M. F., Szmyt, M., Przybyła, M., Calattini, M., Sablin, M., Dobisíková, M., Meldgaard, M., Johansen, M., Berezina, N., Card, N., Saveliev, N. A., Poshekhonova, O., Rickards, O., Lozovskaya, O. V., Gábor, O., Uldum, O. C., Aurino, P., Kosintsev, P., Courtaud, P., Ríos, P., Mortensen, P., Lotz, P., Persson, P., Bangsgaard, P., de Barros Damgaard, P., Vang Petersen, P., Martinez, P. P., Włodarczak, P., Smolyaninov, R. V., Maring, R., Menduiña, R., Badalyan, R., Iversen, R., Turin, R., Vasilyev, S., Wåhlin, S., Borutskaya, S., Skochina, S., Sørensen, S. A., Andersen, S. H., Jørgensen, T., Serikov, Y. B., Molodin, V. I., Smrcka, V., Merts, V., Appadurai, V., Moiseyev, V., Magnusson, Y., Kjær, K. H., Lynnerup, N., Lawson, D. J., Sudmant, P. H., Rasmussen, S., Korneliussen, T. S., Durbin, R., Nielsen, R., Delaneau, O., Werge, T., Racimo, F., Kristiansen, K., & Willerslev, E. (2024).

Population genomics of post-glacial western Eurasia. Nature, 625(7994), 301–311. https://doi.org/10.1038/s41586-023-06865-0

Almagro, M., Blech, M., Ruia Marta, D., & Arteaga, O. (2001). Protohistoria de la Península Ibérica. Ariel.

Alonso, N., Colet, A., Gené, M., Junyent, E., Lafuente, A., López, J. B., Moya, A., Tartera, E., & Vidal, A. (2004). Els Vilars 2001. Memòria de la XVIIa campanya d'excavacions arqueològiques. Generalitat de Catalunya. https://calaix.gencat.cat/bitstream/handle/10687/454193/qmem4437.pdf?sequence=1&isAllowed=y

Antonio, M. L., Gao, Z., Moots, H. M., Lucci, M., Candilio, F., Sawyer, S., Oberreiter, V., Calderon, D., Devitofranceschi, K., Aikens, R. C., Aneli, S., Bartoli, F., Bedini, A., Cheronet, O., Cotter, D. J., Fernandes, D. M., Gasperetti, G., Grifoni, R., Guidi, A., La Pastina, F., Loreti, E., Manacorda, D., Matullo, G., Morretta, S., Nava, A., Fiocchi Nicolai, V., Nomi, F., Pavolini, C., Pentiricci, M., Pergola, P., Piranomonte, M., Schmidt, R., Spinola, G., Sperduti, A., Rubini, M., Bondioli, L., Coppa, A., Pinhasi, R., & Pritchard, J. K. (2019). Ancient Rome: A genetic crossroads of Europe and the Mediterranean. Science, 366(6466), 708–714. https://doi.org/10.1126/science.aay6826

Antonio, M. L., Weiß, C. L., Gao, Z., Sawyer, S., Oberreiter, V., Moots, H. M., Spence, J. P., Cheronet, O., Zagorc, B., Praxmarer, E., Özdoğan, K. T., Demetz, L., Gelabert, P., Fernandes, D., Lucci, M., Alihodžić, T., Amrani, S., Avetisyan, P., Baillif-Ducros, C., Bedić, Ž., Bertrand, A., Bilić, M., Bondioli, L., Borówka, P., Botte, E., Burmaz, J., Bužanić, D., Candilio, F., Cvetko, M., De Angelis, D., Drnić, I., Elschek, K., Fantar, M., Gaspari, A., Gasperetti, G., Genchi, F., Golubović, S., Hukeľová, Z., Jankauskas, R., Vučković, K. J., Jeremić, G., Kaić, I., Kazek, K., Khachatryan, H., Khudaverdyan, A., Kirchengast, S., Korać, M., Kozlowski, V., Krošláková, M., Kušan Špalj, D., La Pastina, F., Laguardia, M., Legrand, S., Leleković, T., Leskovar, T., Lorkiewicz, W., Los, D., Silva, A. M., Masaryk, R., Matijević, V., Cherifi, Y. M. S., Meyer, N., Mikić, I., Miladinović-Radmilović, N., Milošević Zakić, B., Nacouzi, L., Natuniewicz-Sekuła, M., Nava, A., Neugebauer-Maresch, C., Nováček, J., Osterholtz, A., Paige, J., Paraman, L., Pieri, D., Pieta, K., Pop-Lazić, S., Ruttkay, M., Sanader, M., Sołtysiak, A., Sperduti, A., Stankovic Pesterac, T., Teschler-Nicola, M., Teul, I., Tončinić, D., Trapp, J., Vulović, D., Waliszewski, T., Walter, D., Živanović, M., Filah, M. el M., Čaušević-Bully, M., Šlaus, M., Borić, D., Novak, M., Coppa,

A., Pinhasi, R., & Pritchard, J. K. (2024). Stable population structure in Europe since the Iron Age, despite high mobility. ELife, 13. https://doi.org/10.7554/eLife.79714

Aranegui, C. (2012). Los iberos ayer y hoy. Arqueologías y culturas.

Asensio, D., Morer de Llorens, J., Rigo Jovells, A., & Sanmartí, J. (2001). Les formes d'organització social i econòmica a la cossetània ibèrica noves dades sobrel'evolució i tipologia dels assentaments entre els segles VII-I a.c. Territori Polític i Territori Rural Durant l'edat Del Ferro a La Mediterrània Occidental: Actes de La Taula Rodona Celebrada a Ullastret Del 25 Al 27 de Maig de 2000, 253–272.

Austin, J. J., Ross, A. J., Smith, A. B., Fortey, R. A., & Thomas, R. H. (1997). Problems of reproducibility – does geologically ancient DNA survive in amber–preserved insects? Proceedings of the Royal Society of London. Series B: Biological Sciences, 264(1381), 467–474. https://doi.org/10.1098/rspb.1997.0067

Ávila-Arcos, M. C., de la Fuente Castro, C., Nieves-Colón, M. A., & Raghavan, M. (2022). Recommendations for Sustainable Ancient DNA Research in the Global South: Voices From a New Generation of Paleogenomicists. Frontiers in Genetics, 13. https://doi.org/10.3389/fgene.2022.880170

Ayad, L. A. K., & Pissis, S. P. (2017). MARS: improving multiple circular sequence alignment using refined sequences. BMC Genomics, 18(1), 86. https://doi.org/10.1186/s12864-016-3477-5

Bagnall, N. (1999). The Punic Wars: Rome, Carthage and the Struggle for the Mediterranean (Pimilico (ed.)).

Ballester Tormo, I. (1952). Archivo de prehistoria levantina [Journal]. annuario del Servicio de investigación prehistórica de la excelentísma diputación provincial de Valencia, 3, 168.

Barquera, R., Del Castillo-Chávez, O., Nägele, K., Pérez-Ramallo, P., Hernández-Zaragoza, D. I., Szolek, A., Rohrlach, A. B., Librado, P., Childebayeva, A., Bianco, R. A., Penman, B. S., Acuña-Alonzo, V., Lucas, M., Lara-Riegos, J. C., Moo-Mezeta, M. E., Torres-Romero, J. C., Roberts, P., Kohlbacher, O., Warinner, C., & Krause, J. (2024). Ancient genomes reveal insights into ritual life at Chichén Itzá. Nature, 630(8018), 912–919. https://doi.org/10.1038/s41586-024-07509-7

Belarte, M. C. (2010). Los individuos en el espacio doméstico en la protohistoria de Cataluña [Article]. Arqueología Espacial, 28, 109–134.

Bendala, M. (1998). La ciudad entre los iberos, espacio de poder. SAGVNTVM Extra, 1, 25–34. https://turia.uv.es/index.php/saguntumextra/article/view/2742

Bollongino, R., Tresset, A., & Vigne, J.-D. (2008). Environment and excavation: Pre-lab impacts on ancient DNA analyses. Comptes Rendus Palevol, 7(2–3), 91–98. https://doi.org/10.1016/j.crpv.2008.02.002

Boom, R., Sol, C. J., Salimans, M. M., Jansen, C. L., Wertheim-van Dillen, P. M., & van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. Journal of Clinical Microbiology, 28(3), 495–503. https://doi.org/10.1128/jcm.28.3.495-503.1990

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. https://doi.org/10.1038/nature10549

Bosch-Gimpera, P. (1938). Superestructures de la història d'Espanya. Revista de Catalunya, 84, 313–322.

Bosch, J., & Santacana, J. (2009). Blat, metalls, cabdills. Catalunya del neolític a la iberització.

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M., & Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences, 104(37), 14616–14621. https://doi.org/10.1073/pnas.0704665104

Brotherton, P., Endicott, P., Sanchez, J. J., Beaumont, M., Barnett, R., Austin, J., & Cooper, A. (2007). Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. Nucleic Acids Research, 35(17), 5717–5728. https://doi.org/10.1093/nar/gkm588

Brown, T. A., & Brown, K. A. (1994). Ancient DNA: Using molecular biology to explore the past. BioEssays, 16(10), 719–726. https://doi.org/10.1002/bies.950161006

Brown, W. M., George, M., & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences, 76(4), 1967–1971. https://doi.org/10.1073/pnas.76.4.1967

Brunel, S., Bennett, E. A., Cardin, L., Garraud, D., Barrand Emam, H., Beylier, A., Boulestin, B., Chenal, F., Ciesielski, E., Convertini, F., Dedet, B., Desbrosse-Degobertiere, S., Desenne, S., Dubouloz, J., Duday, H., Escalon, G., Fabre, V., Gailledrat, E., Gandelin, M., Gleize, Y., Goepfert, S., Guilaine, J., Hachem, L., Ilett, M., Lambach, F., Maziere, F., Perrin, B., Plouin, S., Pinard, E., Praud, I., Richard, I., Riquier, V., Roure, R., Sendra, B., Thevenet, C., Thiol, S., Vauquelin, E., Vergnaud, L., Grange, T., Geigl, E.-M., & Pruvost, M. (2020). Ancient genomes from present-day France unveil 7,000 years of its demographic history. Proceedings of the National Academy of Sciences, 117(23), 12791–12798. https://doi.org/10.1073/pnas.1918034117

Burbano, H. A., Hodges, E., Green, R. E., Briggs, A. W., Krause, J., Meyer, M., Good, J. M., Maricic, T., Johnson, P. L. F., Xuan, Z., Rooks, M., Bhattacharjee, A., Brizuela, L., Albert, F. W., de la Rasilla, M., Fortea, J., Rosas, A., Lachmann, M., Hannon, G. J., & Pääbo, S. (2010). Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. Science, 328(5979), 723–725. https://doi.org/10.1126/science.1188046

Busquets, F., Fàbregas, M., Corts, A., & Capdevila, N. (2017). El Conjunt Monumental Del Castell De Besora: Primeres Aproximacions a Un Assentament Amb Més De Dos Mil Anys D'Història. Ausa, 28(179), 107–126.

Cano, R. J., Poinar, H. N., Pieniazek, N. J., Acra, A., & Poinar Jr, G. O. (1993). Amplification and sequencing of DNA from a 120–135-million-year-old weevil. Nature, 363(6429), 536–538. https://doi.org/10.1038/363536a0

Cano, R. J., Poinar, H. N., Roubik, D., & Poinar Jr, G. O. (1992). Enzymatic Amplification and Nucleotide Sequencing of Portions of the 18s rRNA gene of the Bee Proplebeia dominicana (Apidae: Hymenoptera) Isolated from 25-40 Million Year Old Dominican Amber. Medical Science Research, 20(17), 619–622.

Carey, B. T., Allfree, J. B., & Cairns, J. (2007). Hannibal's Last Battle: Zama & The Fall of Carthage (Pen & Sword Books (ed.)).

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., Wales, N., Sicheritz-Pontén, T., & Gilbert, M. T. P. (2018). Single-tube library preparation for degraded DNA. Methods in Ecology and Evolution, 9(2), 410–419. https://doi.org/10.1111/2041-210X.12871

Carrión, P., Olalde, I., Jiménez-Arenas, J. M., Coromina, N., Vivó, D., Vergès, J. M., Costa, A., Botella, D., Bustamante-Álvarez, M., Heras-Mora, J., Ortega-Ruiz, R., Chaves, C.,

García-Collado, M. I., Quirós-Castillo, J. A., Roig, J., Suárez-Padilla, J., Navarro-Luengo, I., Cuadrado, M. Á., Aguilera, I., Morera, J., Catalán, R., Cerdeño, M. L., Roig-Pérez, J. F., Díaz-García, M., Chirosa-Cañavate, P., Piza-Ruiz, T., Vallejo-Casas, E., Vidal-Álvarez, S., Burch, J., Sagrera, J., Vivo, J., Cubo-Córdoba, A., Martínez-Enamorado, V., Rengel-Castro, F., García-Entero, V., Rodero, A., Viguera, E., Rohland, N., Morales, J. I., Soto, M., Mallick, S., Cebrià, A., García-Borja, P., Calduch-Bardoll, P., Ulloa-Eres, P., Carretero, A., Espinosa, T., Campderá-Gutiérrez, B., Pagés-Alonso, P., Vara-Izquierdo, C., Martínez-Peñarroya, J., Sardà-Seuma, S., Castaño-Aguilar, J. M., López-Chamizo, S., Pinhasi, R., Lalueza-Fox, C., & Reich, D. (2024). Disparate demographic impacts of the Roman Colonization and the Migration Period in the Iberian Peninsula. https://doi.org/10.1101/2024.09.23.614606

Cassidy, L. M., Russell, M., Smith, M., Delbarre, G., Cheetham, P., Manley, H., Mattiangeli, V., Breslin, E. M., Jackson, I., McCann, M., Little, H., O'Connor, C. G., Heaslip, B., Lawson, D., Endicott, P., & Bradley, D. G. (2025). Continental influx and pervasive matrilocality in Iron Age Britain. Nature, 637(8048), 1136–1142. https://doi.org/10.1038/s41586-024-08409-6

Cerezo, M., Achilli, A., Olivieri, A., Perego, U. A., Gómez-Carballa, A., Brisighelli, F., Lancioni, H., Woodward, S. R., López-Soto, M., Carracedo, Á., Capelli, C., Torroni, A., & Salas, A. (2012). Reconstructing ancient mitochondrial DNA links between Africa and Europe. Genome Research, 22(5), 821–826. https://doi.org/10.1101/gr.134452.111

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Childebayeva, A., Rohrlach, A. B., Barquera, R., Rivollat, M., Aron, F., Szolek, A., Kohlbacher, O., Nicklisch, N., Alt, K. W., Gronenborn, D., Meller, H., Friederich, S., Prüfer, K., Deguilloux, M.-F., Krause, J., & Haak, W. (2022). Population Genetics and Signatures of Selection in Early Neolithic European Farmers. Molecular Biology and Evolution, 39(6). https://doi.org/10.1093/molbev/msac108

Claw, K. G., Anderson, M. Z., Begay, R. L., Tsosie, K. S., Fox, K., Garrison, N. A., Bader, A. C., Bardill, J., Bolnick, D. A., Brooks, J., Cordova, A., Malhi, R. S., Nakatsuka, N., Neller, A., Raff, J. A., Singson, J., TallBear, K., Vargas, T., & Yracheta, J. M. (2018). A framework for enhancing ethical genomic research with Indigenous communities. Nature Communications, 9(1), 2957. https://doi.org/10.1038/s41467-018-05188-3

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research, 38(6), 1767–1771. https://doi.org/10.1093/nar/gkp1137

Cooper, A., Mourer-Chauviré, C., Chambers, G. K., von Haeseler, A., Wilson, A. C., & Pääbo, S. (1992). Independent origins of New Zealand moas and kiwis. Proceedings of the National Academy of Sciences, 89(18), 8741–8744. https://doi.org/10.1073/pnas.89.18.8741

Cooper, A., & Poinar, H. N. (2000). Ancient DNA: Do It Right or Not at All. Science, 289(5482), 1139–1139. https://doi.org/10.1126/science.289.5482.1139b

Cornell, T. J. (1989). The conquest of Italy. In The Cambridge Ancient History: VII Part 2, The Rise of Rome to 220 BC (2nd edition, pp. 309–350). Cambridge University Press.

Cornell, T. J. (1995). The beginnings of Rome: Italy and Rome from the Bronze Age to the Punic Wars (c. 1000-264 BC). Routledge.

Costa, M. D., Pereira, J. B., Pala, M., Fernandes, V., Olivieri, A., Achilli, A., Perego, U. A., Rychkov, S., Naumova, O., Hatina, J., Woodward, S. R., Eng, K. K., Macaulay, V., Carr, M., Soares, P., Pereira, L., & Richards, M. B. (2013). A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. Nature Communications, 4(1), 2543. https://doi.org/10.1038/ncomms3543

Crespo, L., Subirà, M. E., & Ruiz, J. (2011). Twins in prehistory: The case from Olèrdola (Barcelona, Spain; s. IV II BC). International Journal of Osteoarchaeology, 21(6), 751–756. https://doi.org/10.1002/oa.1169

Cuesta-Aguirre, D. R., Malgosa, A., & Santos, C. (2024). An easy-to-use pipeline to analyze amplicon-based Next Generation Sequencing results of human mitochondrial DNA from degraded samples. PLOS ONE, 19(11), e0311115. https://doi.org/10.1371/journal.pone.0311115

Cuesta-Aguirre, D. R., Amor-Jimenez, C., Malgosa, A., & Santos, C. (2025). A *Post-Mortem* Molecular Damage Profile in the Ancient Human Mitochondrial DNA. Molecular Ecology Resources, 25(4), e14061. https://doi.org/10.1111/1755-0998.14061

Cuesta-Aguirre, D. R., Roca-Rada, X., Di Biase, B., Vinueza-Espinosa, D. C., Sandoval-Ávila, C., Ravishankar, S., Molist, N., Durán, M., Mestres, I., Junyent, E., Alonso, N., Martinez Majoral, J., Kjær Knudsen, A., Lanigan, L. T., Schroeder, H., Llamas, B., Malgosa, A., &

Santos, C. (2025). The Genetic Landscape of Northeastern Iberian Communities from the Early to Late Iron Age. SSRN. https://doi.org/10.2139/ssrn.5211463

Curchin, L. A. (1991). Roman Spain: conquest and assimilation. In Roman Spain: conquest and assimilation. Routledge.

Cutillas-Victoria, B., Camino, M. M., & Asensio, S. F. R. (2024). Trade dynamics between Carthage and Iberia at the end of the 3rd century BC: analytical characterisation of central mediterranean amphorae from Qart-Hadasht (Cartagena, Spain). Archaeological and Anthropological Sciences, 16(4), 57. https://doi.org/10.1007/s12520-024-01959-w

Dabney, J., Meyer, M., & Pääbo, S. (2013). Ancient DNA Damage. Cold Spring Harbor Perspectives in Biology, 5(7), a012567–a012567. https://doi.org/10.1101/cshperspect.a012567

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, 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. https://doi.org/10.1073/pnas.1314445110

Dabney, J., & Meyer, M. (2019). Extraction of Highly Degraded DNA from Ancient Bones and Teeth BT - Ancient DNA: Methods and Protocols (B. Shapiro, A. Barlow, P. D. Heintzman, M. Hofreiter, J. L. A. Paijmans, & A. E. R. Soares (eds.); pp. 25–29). Springer New York. https://doi.org/10.1007/978-1-4939-9176-1_4

Dalén, L., Heintzman, P. D., Kapp, J. D., & Shapiro, B. (2023). Deep-time paleogenomics and the limits of DNA survival. Science, 382(6666), 48–53. https://doi.org/10.1126/science.adh7943

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience, 10(2). https://doi.org/10.1093/gigascience/giab008

Davidson, R., Roca-Rada, X., Ravishankar, S., Taufik, L., Haarkötter, C., Collen, E., Williams, M., Webb, P., Irfan Mahmud, M., Djami, E. N. I., Purnomo, G. A., Santos, C., Malgosa, A., Manzanilla, L. R., Silva, A. M., Tereso, S., Matos, V., Carvalho, P. C., Fernandes, T., Maurer, A.-F., Teixeira, J. C., Tobler, R., Fehren-Schmitz, L., & Llamas, B. (2024).

Optimised in-solution enrichment of over a million ancient human SNPs. https://doi.org/10.1101/2024.05.16.594432

Davidson, R., Williams, M. P., Roca-Rada, X., Kassadjikova, K., Tobler, R., Fehren-Schmitz, L., & Llamas, B. (2023). Allelic bias when performing in-solution enrichment of ancient human DNA. Molecular Ecology Resources, 23(8), 1823–1840. https://doi.org/10.1111/1755-0998.13869

de la Fuente-Seoane, R., López-Onaindia, D., Codina Falgas, F., De Prado, G., Álvarez, C. F., Rovira Hortalà, M. C., Díaz-Zorita Bonilla, M., Nieto-Espinet, A., & Subirà, M. E. (2025). Territorialisation and human mobility during the Iron Age in NE Iberia: An approach through Isotope Analyses of the Severed Heads from Puig Castellar (Barcelona, Spain) and Ullastret (Girona, Spain). Journal of Archaeological Science: Reports, 62, 105035. https://doi.org/10.1016/j.jasrep.2025.105035

Department of Agriculture, Water and the Environment, A. G. (2021). Strait Islander Heritage Protection Act 1984: General guide (p. 9).

DeSalle, R., Barcia, M., & Wray, C. (1993). PCR jumping in clones of 30-million-year-old DNA fragments from amber preserved termites (Mastotermes electrodominicus). Experientia, 49(10), 906–909. https://doi.org/10.1007/BF01952607

DeSalle, R., Gatesy, J., Wheeler, W., & Grimaldi, D. (1992). DNA Sequences from a Fossil Termite in Oligo-Miocene Amber and their Phylogenetic Implications. Science, 257(5078), 1933–1936. https://doi.org/10.1126/science.1411508

Dietler, M. (2009). Colonial Encounters in Iberia and the Western Mediterranean: An Exploratory Framework. In M. Dietler & C. López-Ruiz (Eds.), Colonial encounters in ancient Iberia: Phoenician, Greek and indigenous relations. The University of Chicago Press.

Diroma, M. A., Modi, A., Lari, M., Sineo, L., Caramelli, D., & Vai, S. (2021). New Insights Into Mitochondrial DNA Reconstruction and Variant Detection in Ancient Samples. Frontiers in Genetics, 12. https://doi.org/10.3389/fgene.2021.619950

Duran i Caixal, M., & Mestres i Santacreu, I. (2011). Memòria de les intervencions arqueològiques, de restauració, consolidació i museïtzació. El Camp de les Lloses. 2011. In Arxiu del Servei d'Arqueologia de la Generalitat de Catalunya (p. 246). http://calaix.gencat.cat/handle/10687/425409#page=1

Duran i Caixal, M., Mestres i Santacreu, I., Padrós, C., & Principal, J. (2017). El Camp de les Lloses, un exemple d'implantació militar al territori d'època romanorepublicana. Tribuna d'Arqueologia, No 2014-20, 11–45.

Duran, M., Mestres, I., Padrós, C., & Principal, J. (2017). El Camp de les Lloses (Tona, Barcelona): evolución y significado del vicus romanorrepublicano. In Roma en la Península Ibérica presertoriana. Escenarios de implantación militar provincial (Vol. 56, pp. 153–189).

Duran, M., Mestres, I., Principal, J., & Padrós, C. (2015). El Camp de les Lloses (Tona, Osona): un post avançat amb tallers metal·lúrgics al servei de la logística de l'exèrcit romanorepublicà (125-75 ane). Revista d'arqueologia de Ponent, 25(January), 293–307. https://raco.cat/index.php/RAP/article/view/318395

Ebbert, M. T. W., Wadsworth, M. E., Staley, L. A., Hoyt, K. L., Pickett, B., Miller, J., Duce, J., Kauwe, J. S. K., & Ridge, P. G. (2016). Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. BMC Bioinformatics, 17(S7), 239. https://doi.org/10.1186/s12859-016-1097-3

Edwell, P. (2011). War Abroad: Spain, Sicily, Macedon, Africa [Bookitem]. In A Companion to the Punic Wars (pp. 320–338). Wiley. https://doi.org/10.1002/9781444393712.ch18

Eng, R. Y., & Smith, T. C. (1976). Peasant Families and Population Control in Eighteenth-Century Japan. Journal of Interdisciplinary History, 6(3), 417. https://doi.org/10.2307/202664

Essel, E., Zavala, E. I., Schulz-Kornas, E., Kozlikin, M. B., Fewlass, H., Vernot, B., Shunkov, M. V., Derevianko, A. P., Douka, K., Barnes, I., Soulier, M.-C., Schmidt, A., Szymanski, M., Tsanova, T., Sirakov, N., Endarova, E., McPherron, S. P., Hublin, J.-J., Kelso, J., Pääbo, S., Hajdinjak, M., Soressi, M., & Meyer, M. (2023). Ancient human DNA recovered from a Palaeolithic pendant. Nature, 618(7964), 328–332. https://doi.org/10.1038/s41586-023-06035-2

Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. Nature Biotechnology, 38(3), 276–278. https://doi.org/10.1038/s41587-020-0439-x

Excoffier, L., & Yang, Z. (1999). Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Molecular Biology and Evolution, 16(10), 1357–1368. https://doi.org/10.1093/oxfordjournals.molbev.a026046

Feeney, D. (2017). Carthage and Rome: Introduction. Classical Philology, 112(3), 301–311. https://doi.org/10.1086/692440

Fellows Yates, J. A., Lamnidis, T. C., Borry, M., Andrades Valtueña, A., Fagernäs, Z., Clayton, S., Garcia, M. U., Neukamm, J., & Peltzer, A. (2021). Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. PeerJ, 9, e10947. https://doi.org/10.7717/peerj.10947

Fernandes, D. M., Mittnik, A., Olalde, I., Lazaridis, I., Cheronet, O., Rohland, N., Mallick, S., Bernardos, R., Broomandkhoshbacht, N., Carlsson, J., Culleton, B. J., Ferry, M., Gamarra, B., Lari, M., Mah, M., Michel, M., Modi, A., Novak, M., Oppenheimer, J., Sirak, K. A., Stewardson, K., Mandl, K., Schattke, C., Özdoğan, K. T., Lucci, M., Gasperetti, G., Candilio, F., Salis, G., Vai, S., Camarós, E., Calò, C., Catalano, G., Cueto, M., Forgia, V., Lozano, M., Marini, E., Micheletti, M., Miccichè, R. M., Palombo, M. R., Ramis, D., Schimmenti, V., Sureda, P., Teira, L., Teschler-Nicola, M., Kennett, D. J., Lalueza-Fox, C., Patterson, N., Sineo, L., Coppa, A., Caramelli, D., Pinhasi, R., & Reich, D. (2020). The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. Nature Ecology & Evolution, 4(3), 334–345. https://doi.org/10.1038/s41559-020-1102-0

Fleskes, R. E., Bruwelheide, K. S., West, F. L., Owsley, D. W., Griffith, D. R., Barca, K. G., Cabana, G. S., & Schurr, T. G. (2019). Ancient DNA and bioarchaeological perspectives on European and African diversity and relationships on the colonial Delaware frontier. American Journal of Physical Anthropology, 170(2), 232–245. https://doi.org/10.1002/ajpa.23887

Franke, P. R. (1989). Pyrrhus. In The Cambridge Ancient History: VII Part 2, The Rise of Rome to 220 BC (2nd edition, pp. 456–485). Cambridge University Press.

Fregel, R., Cabrera, V., Larruga, J. M., Abu-Amero, K. K., & González, A. M. (2015). Carriers of Mitochondrial DNA Macrohaplogroup N Lineages Reached Australia around 50,000 Years Ago following a Northern Asian Route. PLOS ONE, 10(6), e0129839. https://doi.org/10.1371/journal.pone.0129839

Fregel, R., Méndez, F. L., Bokbot, Y., Martín-Socas, D., Camalich-Massieu, M. D., Santana, J., Morales, J., Ávila-Arcos, M. C., Underhill, P. A., Shapiro, B., Wojcik, G., Rasmussen, M., Soares, A. E. R., Kapp, J., Sockell, A., Rodríguez-Santos, F. J., Mikdad, A., Trujillo-Mederos, A., & Bustamante, C. D. (2018). Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. Proceedings of the National Academy of Sciences, 115(26), 6774–6779. https://doi.org/10.1073/pnas.1800851115

Fu, Q., Hajdinjak, M., Moldovan, O. T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., Viola, B., Prüfer, K., Meyer, M., Kelso, J., Reich, D., & Pääbo, S. (2015). An early modern human from Romania with a recent Neanderthal ancestor. Nature, 524(7564), 216–219. https://doi.org/10.1038/nature14558

Fu, Q., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., Ronchitelli, A. M., Martini, F., Cremonesi, R. G., Svoboda, J., Bauer, P., Caramelli, D., Castellano, S., Reich, D., Pääbo, S., & Krause, J. (2013). A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. Current Biology, 23(7), 553–559. https://doi.org/10.1016/j.cub.2013.02.044

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., Nickel, B., Peltzer, A., Rohland, N., Slon, V., Talamo, S., Lazaridis, I., Lipson, M., Mathieson, I., Schiffels, S., Skoglund, P., Derevianko, A. P., Drozdov, N., Slavinsky, V., Tsybankov, A., Cremonesi, R. G., Mallegni, F., Gély, B., Vacca, E., Morales, M. R. G., Straus, L. G., Neugebauer-Maresch, C., Teschler-Nicola, M., Constantin, S., Moldovan, O. T., Benazzi, S., Peresani, M., Coppola, D., Lari, M., Ricci, S., Ronchitelli, A., Valentin, F., Thevenet, C., Wehrberger, K., Grigorescu, D., Rougier, H., Crevecoeur, I., Flas, D., Semal, P., Mannino, M. A., Cupillard, C., Bocherens, H., Conard, N. J., Harvati, K., Moiseyev, V., Drucker, D. G., Svoboda, J., Richards, M. P., Caramelli, D., Pinhasi, R., Kelso, J., Patterson, N., Krause, J., Pääbo, S., & Reich, D. (2016). The genetic history of Ice Age Europe. Nature, 534(7606), 200–205. https://doi.org/10.1038/nature17993

Furtwängler, A., Rohrlach, A. B., Lamnidis, T. C., Papac, L., Neumann, G. U., Siebke, I., Reiter, E., Steuri, N., Hald, J., Denaire, A., Schnitzler, B., Wahl, J., Ramstein, M., Schuenemann, V. J., Stockhammer, P. W., Hafner, A., Lösch, S., Haak, W., Schiffels, S., & Krause, J. (2020). Ancient genomes reveal social and genetic structure of Late Neolithic

Switzerland. Nature Communications, 11(1), 1915. https://doi.org/10.1038/s41467-020-15560-x

Gamba, C., Fernández, E., Oliver, A., Tirado, M., Baeza, C., López-Parra, A. M., & Arroyo-Pardo, E. (2008). Population genetics and DNA preservation in ancient human remains from Eastern Spain. Forensic Science International: Genetics Supplement Series, 1(1), 462–464. https://doi.org/10.1016/j.fsigss.2007.10.037

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., Whittle, A., Dani, J., Raczky, P., Higham, T. F. G., Hofreiter, M., Bradley, D. G., & Pinhasi, R. (2014). Genome flux and stasis in a five millennium transect of European prehistory. Nature Communications, 5(1), 5257. https://doi.org/10.1038/ncomms6257

Gilbert, M. T. P. (2006). Postmortem Damage of Mitochondrial DNA. In Human Mitochondrial DNA and the Evolution of Homo sapiens (pp. 91–115). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-31789-9_5

Gilbert, M. T. P., Shapiro, B., Drummond, A., & Cooper, A. (2005). Post-mortem DNA damage hotspots in Bison (Bison bison) provide evidence for both damage and mutational hotspots in human mitochondrial DNA. Journal of Archaeological Science, 32(7), 1053–1060. https://doi.org/10.1016/j.jas.2005.02.006

Gilbert, M. T. P., Willerslev, E., Hansen, A. J., Barnes, I., Rudbeck, L., Lynnerup, N., & Cooper, A. (2003). Distribution Patterns of Postmortem Damage in Human Mitochondrial DNA. The American Journal of Human Genetics, 72(1), 32–47. https://doi.org/10.1086/345378

Giles, R. E., Blanc, H., Cann, H. M., & Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. Proceedings of the National Academy of Sciences, 77(11), 6715–6719. https://doi.org/10.1073/pnas.77.11.6715

Gitin, S., & Bierling, M. R. (2002). The Phoenicians in Spain (S. (Sy) Gitin (ed.)) [Book]. Penn State University Press. https://doi.org/10.1515/9781575065298

Gitschier, J. (2008). Imagine: An Interview with Svante Pääbo. PLoS Genetics, 4(3), e1000035. https://doi.org/10.1371/journal.pgen.1000035

Glocke, I., & Meyer, M. (2017). Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. Genome Research, 27(7), 1230–1237. https://doi.org/10.1101/gr.219675.116

Goldsworthy, A. (2012). The Fall of Carthage: The Punic Wars 265-146BC. Orion. https://books.google.es/books?id=jM1sFXcAPvAC

Golenberg, E. M., Giannasi, D. E., Clegg, M. T., Smiley, C. J., Durbin, M., Henderson, D., & Zurawski, G. (1990). Chloroplast DNA sequence from a Miocene Magnolia species. Nature, 344(6267), 656–658. https://doi.org/10.1038/344656a0

González, A. M., Larruga, J. M., Abu-Amero, K. K., Shi, Y., Pestano, J., & Cabrera, V. M. (2007). Mitochondrial lineage M1 traces an early human backflow to Africa. BMC Genomics, 8(1), 223. https://doi.org/10.1186/1471-2164-8-223

González Reyero, S. (2012). Iberos: sociedades y territorios del occidente mediterráneo.

Gowland, R. L., Chamberlain, A., & Redfern, R. C. (2014). 'On the brink of being: re-evaluating infanticide and infant burial in Roman Britain. Journal of Roman Archaeology Supplementary Series, 96, 69–88.

Grant, M. (1978). History of Rome. Prentice Hall.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Ž., Gušic, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R. W., Johnson, P. L. F., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., & Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. Science, 328(5979), 710–722. https://doi.org/10.1126/science.1188021

Günther, T., Valdiosera, C., Malmström, H., Ureña, I., Rodriguez-Varela, R., Sverrisdóttir, Ó. O., Daskalaki, E. A., Skoglund, P., Naidoo, T., Svensson, E. M., Bermúdez de Castro, J. M., Carbonell, E., Dunn, M., Storå, J., Iriarte, E., Arsuaga, J. L., Carretero, J.-M., Götherström, A., & Jakobsson, M. (2015). Ancient genomes link early farmers from

Atapuerca in Spain to modern-day Basques. Proceedings of the National Academy of Sciences, 112(38), 11917–11922. https://doi.org/10.1073/pnas.1509851112

Gutiérrez-García, T. A., Vázquez-Domínguez, E., Arroyo-Cabrales, J., Kuch, M., Enk, J., King, C., & Poinar, H. N. (2014). Ancient DNA and the tropics: a rodent's tale. Biology Letters, 10(6), 20140224. https://doi.org/10.1098/rsbl.2014.0224

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bánffy, E., Economou, C., Francken, M., Friederich, S., Pena, R. G., Hallgren, F., Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov, O., Moiseyev, V., Nicklisch, N., Pichler, S. L., Risch, R., Rojo Guerra, M. A., Roth, C., Szécsényi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D., Cooper, A., Alt, K. W., & Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature, 522(7555), 207–211. https://doi.org/10.1038/nature14317

Haarkötter, C., Vinueza-Espinosa, D. C., Gálvez, X., Saiz, M., Medina-Lozano, M. I., Lorente, J. A., & Álvarez, J. C. (2023). A comparison between petrous bone and tooth, femur and tibia DNA analysis from degraded skeletal remains. ELECTROPHORESIS, 44(19–20), 1559–1568. https://doi.org/10.1002/elps.202300097

Hagelberg, E., Thomas, M. G., Cook, C. E., Sher, A. V., Baryshnikov, G. F., & Lister, A. M. (1994). DNA from ancient mammoth bones. Nature, 370(6488), 333–334. https://doi.org/10.1038/370333b0

Handt, O., Richards, M., Trommsdorff, M., Kilger, C., Simanainen, J., Georgiev, O., Bauer, K., Stone, A., Hedges, R., Schaffner, W., Utermann, G., Sykes, B., & Pääbo, S. (1994). Molecular Genetic Analyses of the Tyrolean Ice Man. Science, 264(5166), 1775–1778. https://doi.org/10.1126/science.8209259

Hassan, N. A.-M., Brown, K. A., Eyers, J., Brown, T. A., & Mays, S. (2014). Ancient DNA study of the remains of putative infanticide victims from the Yewden Roman villa site at Hambleden, England. Journal of Archaeological Science, 43, 192–197. https://doi.org/10.1016/j.jas.2013.12.017

Hedges, S. B., & Schweitzer, M. H. (1995). Detecting Dinosaur DNA. Science, 268(5214), 1191–1192. https://doi.org/10.1126/science.7761839

Hernández, C. (2023). Mitochondrial DNA in Human Diversity and Health: From the Golden Age to the Omics Era. Genes, 14(8), 1534. https://doi.org/10.3390/genes14081534

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. https://doi.org/10.1038/312282a0

Hofreiter, M., Sneberger, J., Pospisek, M., & Vanek, D. (2021). Progress in forensic bone DNA analysis: Lessons learned from ancient DNA. Forensic Science International: Genetics, 54, 102538. https://doi.org/10.1016/j.fsigen.2021.102538

Holland, C. A., McElhoe, J. A., Gaston-Sanchez, S., & Holland, M. M. (2021). Damage patterns observed in mtDNA control region MPS data for a range of template concentrations and when using different amplification approaches. International Journal of Legal Medicine, 135(1), 91–106. https://doi.org/10.1007/s00414-020-02410-0

Holland, M. M., Pack, E. D., & McElhoe, J. A. (2017). Evaluation of GeneMarker® HTS for improved alignment of mtDNA MPS data, haplotype determination, and heteroplasmy assessment. Forensic Science International: Genetics, 28, 90–98. https://doi.org/10.1016/j.fsigen.2017.01.016

Höss, M., & Pääbo, S. (1993). DNA extraction from Pleistocene bones by a silica-based purification method. Nucleic Acids Research, 21(16), 3913–3914. https://doi.org/10.1093/nar/21.16.3913

Huber, N., Parson, W., & Dür, A. (2018). Next generation database search algorithm for forensic mitogenome analyses. Forensic Science International: Genetics, 37, 204–214. https://doi.org/10.1016/j.fsigen.2018.09.001

Hudson, M., Beaton, A., Milne, M., Port, W., Russell, K., Smith, B., Toki, V., Ureata, L., & Wilcox, P. (2016). Te Mata Ira: Guidelines for Genomic Research with Māori (Issue October). https://www.genomics-aotearoa.org.nz/sites/default/files/2019-03/Te-Mata-Ira-Genome-Research-Guidelines.pdf

Ibrahim, J., Brumfeld, V., Addadi, Y., Rubin, S., Weiner, S., & Boaretto, E. (2022). The petrous bone contains high concentrations of osteocytes: One possible reason why ancient DNA is better preserved in this bone. PLOS ONE, 17(10), e0269348. https://doi.org/10.1371/journal.pone.0269348

Illumina. (2024). Adapter trimming: Why are adapter sequences trimmed from only the 3' ends of reads. Illumina Knowledge, #2905. https://knowledge.illumina.com/software/general/software-general-reference_material-list/000002905

Irwin, J. A., Saunier, J. L., Niederstätter, H., Strouss, K. M., Sturk, K. A., Diegoli, T. M., Brandstätter, A., Parson, W., & Parsons, T. J. (2009). Investigation of Heteroplasmy in the Human Mitochondrial DNA Control Region: A Synthesis of Observations from More Than 5000 Global Population Samples. Journal of Molecular Evolution, 68(5), 516–527. https://doi.org/10.1007/s00239-009-9227-4

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., Sikora, M., Rasmussen, S., & Schroeder, H. (2019). A 5700 year-old human genome and oral microbiome from chewed birch pitch. Nature Communications, 10(1), 5520. https://doi.org/10.1038/s41467-019-13549-9

Jimeno Martínez, A. (1996). Numancia: relación necrópolis-poblado. Archivo Español de Arqueología, 69(173–174), 57–76.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics, 29(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Jordá Cerdá, F., Pellecer Catalán, M., Acosta Martínez, P., & Almagro-Gorbea, M. (1986). Prehistoria. Gredos.

Jorde, L. B., Bamshad, M., & Rogers, A. R. (1998). Using mitochondrial and nuclear DNA markers to reconstruct human evolution. BioEssays, 20(2), 126–136. https://doi.org/10.1002/(SICI)1521-1878(199802)20:2<126::AID-BIES5>3.0.CO;2-R

Jun, G., Wing, M. K., Abecasis, G. R., & Kang, H. M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. Genome Research, 25(6), 918–925. https://doi.org/10.1101/gr.176552.114

Junyent, E. (1996a). Els origens de la cultura ilergeta. In Indíbil i mandoni : reis i guerrers (pp. 13–16).

Junyent, E. (1996b). La societat ilergata: un món de pagesos i aristòcrates. In Indíbil i mandoni : reis i guerrers (pp. 17–25).

Junyent, E. (2023). The times of Indivilis and Mandonius. In Romans in the West (pp. 21–33).

Junyent, E., & López, J. B. (2017). La Fortalesa dels Vilars d'Arbeca : terra, aigua i poder en el món iber. Museu de Lleida. https://doi.org/10.21001/fva.2018

Junyent, E., López Melción, J. B., Alonso i Martínez, N., & Lafuente Revuelto, A. (1998). Poder, símbolo y territorio-el caso de la fortaleza de Arbeca. Los Iberos, Príncipes de Occidente. Las Estructuras de Poder En La Sociedad Ibérica, 1, 355–372.

Junyent, E., & Moya, A. (2011). Els fossats de la fortalesa dels Vilars d'Arbeca (Catalunya, Espanya). Revista d'Arqueologia de Ponent, 21, 93–120.

Kaneva, K., Schurr, T. G., Tatarinova, T. V., Buckley, J., Merkurjev, D., Triska, P., Liu, X., Done, J., Maglinte, D. T., Deapen, D., Hwang, A., Schiffman, J. D., Triche, T. J., Biegel, J. A., & Gai, X. (2022). Mitochondrial DNA haplogroup, genetic ancestry, and susceptibility to Ewing sarcoma. Mitochondrion, 67, 6–14. https://doi.org/10.1016/j.mito.2022.09.002

Kapp, J. D., Green, R. E., & Shapiro, B. (2021). A Fast and Efficient Single-stranded Genomic Library Preparation Method Optimized for Ancient DNA. Journal of Heredity, 112(3), 241–249. https://doi.org/10.1093/jhered/esab012

Keay, S. J. (1988). Roman Spain. In Roman Spain. British Museum Publications.

Keay, S. J., & Deamos, M. B. (1998). The archaeology of early Roman Baetica. Journal of Roman Archaeology.

Kılınç, G. M., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, A. M., Bıçakçı, E., Baird, D., Dönertaş, H. M., Ghalichi, A., Yaka, R., Koptekin, D., Açan, S. C., Parvizi, P., Krzewińska, M., Daskalaki, E. A., Yüncü, E., Dağtaş, N. D., Fairbairn, A., Pearson, J., Mustafaoğlu, G., Erdal, Y. S., Çakan, Y. G., Togan, İ., Somel, M., Storå, J., Jakobsson, M., & Götherström, A. (2016). The Demographic Development of the First Farmers in Anatolia. Current Biology, 26(19), 2659–2666. https://doi.org/10.1016/j.cub.2016.07.057

Kjær, K. H., Winther Pedersen, M., De Sanctis, B., De Cahsan, B., Korneliussen, T. S., Michelsen, C. S., Sand, K. K., Jelavić, S., Ruter, A. H., Schmidt, A. M. A., Kjeldsen, K. K.,

Tesakov, A. S., Snowball, I., Gosse, J. C., Alsos, I. G., Wang, Y., Dockter, C., Rasmussen, M., Jørgensen, M. E., Skadhauge, B., Prohaska, A., Kristensen, J. Å., Bjerager, M., Allentoft, M. E., Coissac, E., Alsos, I. G., Coissac, E., Rouillard, A., Simakova, A., Fernandez-Guerra, A., Bowler, C., Macias-Fauria, M., Vinner, L., Welch, J. J., Hidy, A. J., Sikora, M., Collins, M. J., Durbin, R., Larsen, N. K., & Willerslev, E. (2022). A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. Nature, 612(7939), 283–291. https://doi.org/10.1038/s41586-022-05453-y

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics, 15(1), 356. https://doi.org/10.1186/s12859-014-0356-4

Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M., & Pääbo, S. (1997). Neandertal DNA Sequences and the Origin of Modern Humans. Cell, 90(1), 19–30. https://doi.org/10.1016/S0092-8674(00)80310-4

Lamers, R., Hayter, S., & Matheson, C. D. (2009). Postmortem Miscoding Lesions in Sequence Analysis of Human Ancient Mitochondrial DNA. Journal of Molecular Evolution, 68(1), 40–55. https://doi.org/10.1007/s00239-008-9184-3

Larruga, J. M., Marrero, P., Abu-Amero, K. K., Golubenko, M. V, & Cabrera, V. M. (2017). Carriers of mitochondrial DNA macrohaplogroup R colonized Eurasia and Australasia from a southeast Asia core area. BMC Evolutionary Biology, 17(1), 115. https://doi.org/10.1186/s12862-017-0964-5

Lazaridis, I., Alpaslan-Roodenberg, S., Acar, A., Açıkkol, A., Agelarakis, A., Aghikyan, L., Akyüz, U., Andreeva, D., Andrijašević, G., Antonović, D., Armit, I., Atmaca, A., Avetisyan, P., Aytek, A. İ., Bacvarov, K., Badalyan, R., Bakardzhiev, S., Balen, J., Bejko, L., Bernardos, R., Bertsatos, A., Biber, H., Bilir, A., Bodružić, M., Bonogofsky, M., Bonsall, C., Borić, D., Borovinić, N., Bravo Morante, G., Buttinger, K., Callan, K., Candilio, F., Carić, M., Cheronet, O., Chohadzhiev, S., Chovalopoulou, M.-E., Chryssoulaki, S., Ciobanu, I., Čondić, N., Constantinescu, M., Cristiani, E., Culleton, B. J., Curtis, E., Davis, J., Davtyan, R., Demcenco, T. I., Dergachev, V., Derin, Z., Deskaj, S., Devejyan, S., Djordjević, V., Duffett Carlson, K. S., Eccles, L. R., Elenski, N., Engin, A., Erdoğan, N., Erir-Pazarcı, S., Fernandes, D. M., Ferry, M., Freilich, S., Frînculeasa, A., Galaty, M. L., Gamarra, B., Gasparyan, B., Gaydarska, B., Genç, E., Gültekin, T., Gündüz, S., Hajdu, T., Heyd, V., Hobosyan, S., Hovhannisyan, N., Iliev, I., Iliev, L., Iliev, S., İvgin, İ., Janković,

I., Jovanova, L., Karkanas, P., Kavaz-Kındığılı, B., Kaya, E. H., Keating, D., Kennett, D. J., Deniz Kesici, S., Khudaverdyan, A., Kiss, K., Kılıç, S., Klostermann, P., Kostak Boca Negra Valdes, S., Kovačević, S., Krenz-Niedbała, M., Krznarić Škrivanko, M., Kurti, R., Kuzman, P., Lawson, A. M., Lazar, C., Leshtakov, K., Levy, T. E., Liritzis, I., Lorentz, K. O., Łukasik, S., Mah, M., Mallick, S., Mandl, K., Martirosyan-Olshansky, K., Matthews, R., Matthews, W., McSweeney, K., Melikyan, V., Micco, A., Michel, M., Milašinović, L., Mittnik, A., Monge, J. M., Nekhrizov, G., Nicholls, R., Nikitin, A. G., Nikolov, V., Novak, M., Olalde, I., Oppenheimer, J., Osterholtz, A., Özdemir, C., Özdoğan, K. T., Öztürk, N., Papadimitriou, N., Papakonstantinou, N., Papathanasiou, A., Paraman, L., Paskary, E. G., Patterson, N., Petrakiev, I., Petrosyan, L., Petrova, V., Philippa-Touchais, A., Piliposyan, A., Pocuca Kuzman, N., Potrebica, H., Preda-Bălănică, B., Premužić, Z., Price, T. D., Qiu, L., Radović, S., Raeuf Aziz, K., Rajić Šikanjić, P., Rasheed Raheem, K., Razumov, S., Richardson, A., Roodenberg, J., Ruka, R., Russeva, V., Şahin, M., Şarbak, A., Savaş, E., Schattke, C., Schepartz, L., Selçuk, T., Sevim-Erol, A., Shamoon-Pour, M., Shephard, H. M., Sideris, A., Simalcsik, A., Simonyan, H., Sinika, V., Sirak, K., Sirbu, G., Šlaus, M., Soficaru, A., Söğüt, B., Sołtysiak, A., Sönmez-Sözer, Ç., Stathi, M., Steskal, M., Stewardson, K., Stocker, S., Suata-Alpaslan, F., Suvorov, A., Szécsényi-Nagy, A., Szeniczey, T., Telnov, N., Temov, S., Todorova, N., Tota, U., Touchais, G., Triantaphyllou, S., Türker, A., Ugarković, M., Valchev, T., Veljanovska, F., Videvski, Z., Virag, C., Wagner, A., Walsh, S., Włodarczak, P., Workman, J. N., Yardumian, A., Yarovoy, E., Yavuz, A. Y., Yılmaz, H., Zalzala, F., Zettl, A., Zhang, Z., Çavuşoğlu, R., Rohland, N., Pinhasi, R., & Reich, D. (2022). The genetic history of the Southern Arc: A bridge between West Asia and Europe. Science, 377(6609). https://doi.org/10.1126/science.abm4247

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D. C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., Connell, S., Stewardson, K., Harney, E., Fu, Q., Gonzalez-Fortes, G., Jones, E. R., Roodenberg, S. A., Lengyel, G., Bocquentin, F., Gasparian, B., Monge, J. M., Gregg, M., Eshed, V., Mizrahi, A.-S., Meiklejohn, C., Gerritsen, F., Bejenaru, L., Blüher, M., Campbell, A., Cavalleri, G., Comas, D., Froguel, P., Gilbert, E., Kerr, S. M., Kovacs, P., Krause, J., McGettigan, D., Merrigan, M., Merriwether, D. A., O'Reilly, S., Richards, M. B., Semino, O., Shamoon-Pour, M., Stefanescu, G., Stumvoll, M., Tönjes, A., Torroni, A., Wilson, J. F., Yengo, L., Hovhannisyan, N. A., Patterson, N., Pinhasi, R., & Reich, D. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature, 536(7617), 419–424. https://doi.org/10.1038/nature19310

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, M., Schröder, R., Ni, S., Madea, B., & Stoneking, M. (2015). Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. Proceedings of the National Academy of Sciences, 112(8), 2491–2496. https://doi.org/10.1073/pnas.1419651112

Lien, A., Legori, L. P., Kraft, L., Sackett, P. W., & Renaud, G. (2023). Benchmarking software tools for trimming adapters and merging next-generation sequencing data for ancient DNA. Frontiers in Bioinformatics, 3. https://doi.org/10.3389/fbinf.2023.1260486

Lindahl, T. (1993a). Instability and decay of the primary structure of DNA. Nature, 362(6422), 709–715. https://doi.org/10.1038/362709a0

Lindahl, T. (1993b). Recovery of antediluvian DNA. Nature, 365(6448), 700–700. https://doi.org/10.1038/365700a0

Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., Oppenheimer, J., Broomandkhoshbacht, N., Harney, E., Nordenfelt, S., Llamas, B., Gusztáv Mende, B., Köhler, K., Oross, K., Bondár, M., Marton, T., Osztás, A., Jakucs, J., Paluch, T., Horváth, F., Csengeri, P., Koós, J., Sebők, K., Anders, A., Raczky, P., Regenye, J., Barna, J. P., Fábián, S., Serlegi, G., Toldi, Z., Gyöngyvér Nagy, E., Dani, J., Molnár, E., Pálfi, G., Márk, L., Melegh, B., Bánfai, Z., Domboróczki, L., Fernández-Eraso, J., Antonio Mujika-Alustiza, J., Alonso Fernández, C., Jiménez Echevarría, J., Bollongino, R., Orschiedt, J., Schierhold, K., Meller, H., Cooper, A., Burger, J., Bánffy, E., Alt, K. W., Lalueza-Fox, C., Haak, W., & Reich, D. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature, 551(7680), 368–372. https://doi.org/10.1038/nature24476

López Castro, J. L. (2019). The Iberian Peninsula. In B. R. Doak & C. López-Ruiz (Eds.), The Oxford Handbook of the Phoenician and Punic Mediterranean (pp. 583–602). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190499341.013.38

López i Melcion, J. B., Martín i Ortega, A., Mataró Pladelasala, M., & Pons i Brun, E. (1989). Memòria d'excavació 1987-1989 de l'Illa d'en Reixac (Ullastret, Baix Empordà).

López, J. B., Junyent, E., & Alonso, N. (2020). Architecture, Power and Everyday Life in the Iron Age of North-eastern Iberia. Research from 1985 to 2019 on the Tell-like Fortress of Els Vilars (Arbeca, Lleida, Spain). In Current Approaches to Tells in the Prehistoric Old World (pp. 189–208). Oxbow Books. https://doi.org/10.2307/j.ctv13pk5j9.15

Maca-Meyer, N., González, A. M., Larruga, J. M., Flores, C., & Cabrera, V. M. (2001). Major genomic mitochondrial lineages delineate early human expansions. BMC Genetics, 2(1), 13. https://doi.org/10.1186/1471-2156-2-13

Macias, J. M., & Rodà, I. (2015). Tarraco, the first capital. Catalan Historical Review, 8, 9–28.

Mahaney, W. (2013). Hannibal's Odyssey: The Environmental Background to the Alpine Invasion of Italia (1st ed., Vol. 11) [Book]. Gorgias Press. https://doi.org/10.31826/9781463215958

Mallick, S., Micco, A., Mah, M., Ringbauer, H., Lazaridis, I., Olalde, I., Patterson, N., & Reich, D. (2024). The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. Scientific Data, 11(1), 182. https://doi.org/10.1038/s41597-024-03031-7

Maluquer de Motes, J. (1966). El impacto colonial griego y el comienzo de la vida urbana en Cataluña: Discurso leido en la sala de actos del Instituto de Estudios Ilerdenses el día 8 de mayo. Consejo Superior de Investigaciones Científicas. Delegación de Barcelona. https://books.google.es/books?id=ui9exgEACAAJ

Maluquer de Motes, J. (1982). Problemática histórica de la Cultura Ibérica. XVI Congreso Nacional de Arqueología: Murcia-Cartagena, 29–49.

Marcus, J. H., Posth, C., Ringbauer, H., Lai, L., Skeates, R., Sidore, C., Beckett, J., Furtwängler, A., Olivieri, A., Chiang, C. W. K., Al-Asadi, H., Dey, K., Joseph, T. A., Liu, C.-C., Der Sarkissian, C., Radzevičiūtė, R., Michel, M., Gradoli, M. G., Marongiu, P., Rubino, S., Mazzarello, V., Rovina, D., La Fragola, A., Serra, R. M., Bandiera, P., Bianucci, R., Pompianu, E., Murgia, C., Guirguis, M., Orquin, R. P., Tuross, N., van Dommelen, P., Haak, W., Reich, D., Schlessinger, D., Cucca, F., Krause, J., & Novembre, J. (2020). Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia. Nature Communications, 11(1), 939. https://doi.org/10.1038/s41467-020-14523-6

Margaryan, A., Lawson, D. J., Sikora, M., Racimo, F., Rasmussen, S., Moltke, I., Cassidy, L. M., Jørsboe, E., Ingason, A., Pedersen, M. W., Korneliussen, T., Wilhelmson, H., Buś, M. M., de Barros Damgaard, P., Martiniano, R., Renaud, G., Bhérer, C., Moreno-Mayar, J. V., Fotakis, A. K., Allen, M., Allmäe, R., Molak, M., Cappellini, E., Scorrano, G., McColl, H., Buzhilova, A., Fox, A., Albrechtsen, A., Schütz, B., Skar, B., Arcini, C., Falys, C., Jonson, C. H., Błaszczyk, D., Pezhemsky, D., Turner-Walker, G., Gestsdóttir, H., Lundstrøm, I., Gustin, I., Mainland, I., Potekhina, I., Muntoni, I. M., Cheng, J., Stenderup, J., Ma, J., Gibson, J., Peets, J., Gustafsson, J., Iversen, K. H., Simpson, L., Strand, L., Loe, L., Sikora, M., Florek, M., Vretemark, M., Redknap, M., Bajka, M., Pushkina, T., Søvsø, M., Grigoreva, N., Christensen, T., Kastholm, O., Uldum, O., Favia, P., Holck, P., Sten, S., Arge, S. V., Ellingvåg, S., Moiseyev, V., Bogdanowicz, W., Magnusson, Y., Orlando, L., Pentz, P., Jessen, M. D., Pedersen, A., Collard, M., Bradley, D. G., Jørkov, M. L., Arneborg, J., Lynnerup, N., Price, N., Gilbert, M. T. P., Allentoft, M. E., Bill, J., Sindbæk, S. M., Hedeager, L., Kristiansen, K., Nielsen, R., Werge, T., & Willerslev, E. (2020). Population genomics of the Viking world. Nature, 585(7825), 390–396. https://doi.org/10.1038/s41586-020-2688-8

Martín i Méndez, A. (1998). Memòria de les intervencions arqueologiques a Can Mateu (Cabrera de Mar, Maresme) 1997-1998.

Martín i Ortega, A. (1985). Memòria de les campanyes d'excavació de 1984 i 1985 en el Puig de Sant Andreu d'Ullastret (Baix Empordà).

Martín i Ortega, A., Codina, F., & Prado, G. (2016). La cerámica ática de la necrópolis del Puig de Serra (Serra de Daró-Ullastret, Cataluña). Vie Quotidienne, Tombes et Symboles Des Sociétés Protohistoriques de Méditerranée Nord-Occidentale: Mélanges Offerts à Bernard Dedet, 7(2), 479–496.

Martiniano, R., Cassidy, L. M., Ó'Maoldúin, R., McLaughlin, R., Silva, N. M., Manco, L., Fidalgo, D., Pereira, T., Coelho, M. J., Serra, M., Burger, J., Parreira, R., Moran, E., Valera, A. C., Porfirio, E., Boaventura, R., Silva, A. M., & Bradley, D. G. (2017). The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. PLOS Genetics, 13(7), e1006852. https://doi.org/10.1371/journal.pgen.1006852

Martirosyan, A., Sandoval-Ávila, C., Irurita, J., Juanhuix, J., Molist, N., Mestres, I., Durán, M., Alonso, N., Santos, C., Malgosa, A., Molera, J., & Jordana, X. (2024). Reconstructing

infant mortality in Iberian Iron Age populations from tooth histology. Journal of Archaeological Science, 171, 106088. https://doi.org/10.1016/j.jas.2024.106088

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E. R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J. L., de Castro, J. M. B., Carbonell, E., Gerritsen, F., Khokhlov, A., Kuznetsov, P., Lozano, M., Meller, H., Mochalov, O., Moiseyev, V., Guerra, M. A. R., Roodenberg, J., Vergès, J. M., Krause, J., Cooper, A., Alt, K. W., Brown, D., Anthony, D., Lalueza-Fox, C., Haak, W., Pinhasi, R., & Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature, 528(7583), 499–503. https://doi.org/10.1038/nature16152

Matisoo-Smith, E., Gosling, A. L., Platt, D., Kardailsky, O., Prost, S., Cameron-Christie, S., Collins, C. J., Boocock, J., Kurumilian, Y., Guirguis, M., Pla Orquín, R., Khalil, W., Genz, H., Abou Diwan, G., Nassar, J., & Zalloua, P. (2018). Ancient mitogenomes of Phoenicians from Sardinia and Lebanon: A story of settlement, integration, and female mobility. PLOS ONE, 13(1), e0190169. https://doi.org/10.1371/journal.pone.0190169

Merheb, M., Matar, R., Hodeify, R., Siddiqui, S. S., Vazhappilly, C. G., Marton, J., Azharuddin, S., & AL Zouabi, H. (2019). Mitochondrial DNA, a Powerful Tool to Decipher Ancient Human Civilization from Domestication to Music, and to Uncover Historical Murder Cases. Cells, 8(5), 433. https://doi.org/10.3390/cells8050433

Mestres Santacreu, I. (2003). Memòria de la intervenció arqueològica El Camp de les Lloses. In Arxiu del Servei d'Arqueologia de la Generalitat de Catalunya (p. 71). http://calaix.gencat.cat/handle/10687/24042#page=1

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., de Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R. E., Bryc, K., Briggs, A. W., Stenzel, U., Dabney, J., Shendure, J., Kitzman, J., Hammer, M. F., Shunkov, M. V., Derevianko, A. P., Patterson, N., Andrés, A. M., Eichler, E. E., Slatkin, M., Reich, D., Kelso, J., & Pääbo, S. (2012). A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science, 338(6104), 222–226. https://doi.org/10.1126/science.1224344

Meyer, S., Weiss, G., & von Haeseler, A. (1999). Pattern of Nucleotide Substitution and Rate Heterogeneity in the Hypervariable Regions I and II of Human mtDNA. Genetics, 152(3), 1103–1110. https://doi.org/10.1093/genetics/152.3.1103

Mittnik, A., Wang, C.-C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R., Khartanovich, V., Moiseyev, V., Tõrv, M., Furtwängler, A., Andrades Valtueña, A., Feldman, M., Economou, C., Oinonen, M., Vasks, A., Balanovska, E., Reich, D., Jankauskas, R., Haak, W., Schiffels, S., & Krause, J. (2018). The genetic prehistory of the Baltic Sea region. Nature Communications, 9(1), 442. https://doi.org/10.1038/s41467-018-02825-9

Molak, M., & Ho, S. Y. W. (2011). Evaluating the Impact of Post-Mortem Damage in Ancient DNA: A Theoretical Approach. Journal of Molecular Evolution, 73(3–4), 244–255. https://doi.org/10.1007/s00239-011-9474-z

Molist Capella, N., & Bosch Casadevall, J. M. (2002). Memòria de la intervenció arqueològica al conjunt històric d'Olèrdola (2000-2001). Generalitat de Catalunya, Departament de Cultura i Mitjans de Comunicació, Direcció General del Patrimoni Cultural. http://hdl.handle.net/10687/23910

Moots, H. M., Antonio, M., Sawyer, S., Spence, J. P., Oberreiter, V., Weiß, C. L., Lucci, M., Cherifi, Y. M. S., La Pastina, F., Genchi, F., Praxmeier, E., Zagorc, B., Cheronet, O., Özdoğan, K. T., Demetz, L., Amrani, S., Candilio, F., De Angelis, D., Gasperetti, G., Fernandes, D., Gao, Z., Fantar, M., Coppa, A., Pritchard, J. K., & Pinhasi, R. (2023). A genetic history of continuity and mobility in the Iron Age central Mediterranean. Nature Ecology & Evolution, 7(9), 1515–1524. https://doi.org/10.1038/s41559-023-02143-4

Moreno Gil, A., Garrido García, C., Bonora Soriano, B., Gómez-Gras, D., & Risch, R. (2025). Bronze Age Frontiers and Pottery Circulation: Political and Economic Relations at the Northern Fringes of El Argar, Southeast Iberia, ca. 2200–1550 BCE. Journal of Archaeological Method and Theory, 32(2), 36. https://doi.org/10.1007/s10816-025-09702-y

Morillo Cerdán, Á. (2014). The Roman occupation of the north of Hispania: war, military deployment and cultural integration. In F. Cadiou & M. NAvarro Caballero (Eds.), La guerre et ses traces (pp. 49–62). Ausonius Éditions.

Mullis, K. B., & Faloona, F. A. (1987). [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In Recombinant DNA Part F (pp. 335–350). https://doi.org/10.1016/0076-6879(87)55023-6

Myres, N. M., Rootsi, S., Lin, A. A., Järve, M., King, R. J., Kutuev, I., Cabrera, V. M., Khusnutdinova, E. K., Pshenichnov, A., Yunusbayev, B., Balanovsky, O., Balanovska, E.,

Rudan, P., Baldovic, M., Herrera, R. J., Chiaroni, J., Di Cristofaro, J., Villems, R., Kivisild, T., & Underhill, P. A. (2011). A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. European Journal of Human Genetics, 19(1), 95–101. https://doi.org/10.1038/ejhg.2010.146

Naue, J., Hörer, S., Sänger, T., Strobl, C., Hatzer-Grubwieser, P., Parson, W., & Lutz-Bonengel, S. (2015). Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA. Mitochondrion, 20, 82–94. https://doi.org/10.1016/j.mito.2014.12.002

Neukamm, J., Peltzer, A., & Nieselt, K. (2021). DamageProfiler: fast damage pattern calculation for ancient DNA. Bioinformatics, 37(20), 3652–3653. https://doi.org/10.1093/bioinformatics/btab190

Noonan, J. P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Pääbo, S., Pritchard, J. K., & Rubin, E. M. (2006). Sequencing and Analysis of Neanderthal Genomic DNA. Science, 314(5802), 1113–1118. https://doi.org/10.1126/science.1131412

Oakley, S. P. (1998). A Commentary on Livy, Books VI–X, Vol. 2: Books VII–VIII (Vol. 2). Oxford University Press.

Olalde, I., Brace, S., Allentoft, M. E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., Altena, E., Lipson, M., Lazaridis, I., Harper, T. K., Patterson, N., Broomandkhoshbacht, N., Diekmann, Y., Faltyskova, Z., Fernandes, D., Ferry, M., Harney, E., de Knijff, P., Michel, M., Oppenheimer, J., Stewardson, K., Barclay, A., Alt, K. W., Liesau, C., Ríos, P., Blasco, C., Miguel, J. V., García, R. M., Fernández, A. A., Bánffy, E., Bernabò-Brea, M., Billoin, D., Bonsall, C., Bonsall, L., Allen, T., Büster, L., Carver, S., Navarro, L. C., Craig, O. E., Cook, G. T., Cunliffe, B., Denaire, A., Dinwiddy, K. E., Dodwell, N., Ernée, M., Evans, C., Kuchařík, M., Farré, J. F., Fowler, C., Gazenbeek, M., Pena, R. G., Haber-Uriarte, M., Haduch, E., Hey, G., Jowett, N., Knowles, T., Massy, K., Pfrengle, S., Lefranc, P., Lemercier, O., Lefebvre, A., Martínez, C. H., Olmo, V. G., Ramírez, A. B., Maurandi, J. L., Majó, T., McKinley, J. I., McSweeney, K., Mende, B. G., Modi, A., Kulcsár, G., Kiss, V., Czene, A., Patay, R., Endrődi, A., Köhler, K., Hajdu, T., Szeniczey, T., Dani, J., Bernert, Z., Hoole, M., Cheronet, O., Keating, D., Velemínský, P., Dobeš, M., Candilio, F., Brown, F., Fernández, R. F., Herrero-Corral, A.-M., Tusa, S., Carnieri, E., Lentini, L., Valenti, A., Zanini, A., Waddington, C., Delibes, G., Guerra-

Doce, E., Neil, B., Brittain, M., Luke, M., Mortimer, R., Desideri, J., Besse, M., Brücken, G., Furmanek, M., Hałuszko, A., Mackiewicz, M., Rapiński, A., Leach, S., Soriano, I., Lillios, K. T., Cardoso, J. L., Pearson, M. P., Włodarczak, P., Price, T. D., Prieto, P., Rey, P.-J., Risch, R., Rojo Guerra, M. A., Schmitt, A., Serralongue, J., Silva, A. M., Smrčka, V., Vergnaud, L., Zilhão, J., Caramelli, D., Higham, T., Thomas, M. G., Kennett, D. J., Fokkens, H., Heyd, V., Sheridan, A., Sjögren, K.-G., Stockhammer, P. W., Krause, J., Pinhasi, R., Haak, W., Barnes, I., Lalueza-Fox, C., & Reich, D. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. Nature, 555(7695), 190–196. https://doi.org/10.1038/nature25738

Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., Dulias, K., Edwards, C. J., Gandini, F., Pala, M., Soares, P., Ferrando-Bernal, M., Adamski, N., Broomandkhoshbacht, N., Cheronet, O., Culleton, B. J., Fernandes, D., Lawson, A. M., Mah, M., Oppenheimer, J., Stewardson, K., Zhang, Z., Jiménez Arenas, J. M., Toro Moyano, I. J., Salazar-García, D. C., Castanyer, P., Santos, M., Tremoleda, J., Lozano, M., García Borja, P., Fernández-Eraso, J., Mujika-Alustiza, J. A., Barroso, C., Bermúdez, F. J., Viguera Mínguez, E., Burch, J., Coromina, N., Vivó, D., Cebrià, A., Fullola, J. M., García-Puchol, O., Morales, J. I., Oms, F. X., Majó, T., Vergès, J. M., Díaz-Carvajal, A., Ollich-Castanyer, I., López-Cachero, F. J., Silva, A. M., Alonso-Fernández, C., Delibes de Castro, G., Jiménez Echevarría, J., Moreno-Márquez, A., Pascual Berlanga, G., Ramos-García, P., Ramos-Muñoz, J., Vijande Vila, E., Aguilella Arzo, G., Esparza Arroyo, Á., Lillios, K. T., Mack, J., Velasco-Vázquez, J., Waterman, A., Benítez de Lugo Enrich, L., Benito Sánchez, M., Agustí, B., Codina, F., de Prado, G., Estalrrich, A., Fernández Flores, Á., Finlayson, C., Finlayson, G., Finlayson, S., Giles-Guzmán, F., Rosas, A., Barciela González, V., García Atiénzar, G., Hernández Pérez, M. S., Llanos, A., Carrión Marco, Y., Collado Beneyto, I., López-Serrano, D., Sanz Tormo, M., Valera, A. C., Blasco, C., Liesau, C., Ríos, P., Daura, J., de Pedro Michó, M. J., Diez-Castillo, A. A., Flores Fernández, R., Francès Farré, J., Garrido-Pena, R., Gonçalves, V. S., Guerra-Doce, E., Herrero-Corral, A. M., Juan-Cabanilles, J., López-Reyes, D., McClure, S. B., Merino Pérez, M., Oliver Foix, A., Sanz Borràs, M., Sousa, A. C., Vidal Encinas, J. M., Kennett, D. J., Richards, M. B., Werner Alt, K., Haak, W., Pinhasi, R., Lalueza-Fox, C., & Reich, D. (2019). The genomic history of the Iberian Peninsula over the past 8000 years. Science, 363(6432), 1230–1234. https://doi.org/10.1126/science.aav4040

Oliva, A., Tobler, R., Cooper, A., Llamas, B., & Souilmi, Y. (2021). Systematic benchmark of ancient DNA read mapping. Briefings in Bioinformatics, 22(5). https://doi.org/10.1093/bib/bbab076

Oliva, A., Tobler, R., Llamas, B., & Souilmi, Y. (2021). Additional evaluations show that specific BWA-aln settings still outperform BWA-mem for ancient DNA data alignment. Ecology and Evolution, 11(24), 18743–18748. https://doi.org/10.1002/ece3.8297

Oliver Foix, A. (2004). Sacrificios y "mala muerte" en el Registro Arqueológico de los Yacimientos Ibéricos. Espacio Tiempo y Forma. Serie I, Prehistoria y Arqueología, 1(16–17), 391. https://doi.org/10.5944/etfi.16-17.2003.4763

Olivieri, A., Pala, M., Gandini, F., Kashani, B. H., Perego, U. A., Woodward, S. R., Grugni, V., Battaglia, V., Semino, O., Achilli, A., Richards, M. B., & Torroni, A. (2013). Mitogenomes from Two Uncommon Haplogroups Mark Late Glacial/Postglacial Expansions from the Near East and Neolithic Dispersals within Europe. PLoS ONE, 8(7), e70492. https://doi.org/10.1371/journal.pone.0070492

Olivieri, A., Sidore, C., Achilli, A., Angius, A., Posth, C., Furtwängler, A., Brandini, S., Capodiferro, M. R., Gandini, F., Zoledziewska, M., Pitzalis, M., Maschio, A., Busonero, F., Lai, L., Skeates, R., Gradoli, M. G., Beckett, J., Marongiu, M., Mazzarello, V., Marongiu, P., Rubino, S., Rito, T., Macaulay, V., Semino, O., Pala, M., Abecasis, G. R., Schlessinger, D., Conde-Sousa, E., Soares, P., Richards, M. B., Cucca, F., & Torroni, A. (2017). Mitogenome Diversity in Sardinians: A Genetic Window onto an Island's Past. Molecular Biology and Evolution, 34(5), 1230–1239. https://doi.org/10.1093/molbev/msx082

Olivieri, C., Ermini, L., Rizzi, E., Corti, G., Bonnal, R., Luciani, S., Marota, I., De Bellis, G., & Rollo, F. (2010). Characterization of Nucleotide Misincorporation Patterns in the Iceman's Mitochondrial DNA. PLoS ONE, 5(1), e8629. https://doi.org/10.1371/journal.pone.0008629

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., Fu, Q., Krause, J., Willerslev, E., Stone, A. C., & Warinner, C. (2021). Ancient DNA analysis. Nature Reviews Methods Primers, 1(1), 14. https://doi.org/10.1038/s43586-020-00011-0

Ovchinnikov, I. V., Götherström, A., Romanova, G. P., Kharitonov, V. M., Lidén, K., & Goodwin, W. (2000). Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature, 404(6777), 490–493. https://doi.org/10.1038/35006625

Pääbo, S. (1985a). Molecular cloning of Ancient Egyptian mummy DNA. Nature, 314(6012), 644–645. https://doi.org/10.1038/314644a0

Pääbo, S. (1985b). Preservation of DNA in ancient Egyptian mummies. Journal of Archaeological Science, 12(6), 411–417. https://doi.org/10.1016/0305-4403(85)90002-0

Pääbo, S. (1986). Molecular Genetic Investigations of Ancient Human Remains. Cold Spring Harbor Symposia on Quantitative Biology, 51(6012), 441–446. https://doi.org/10.1101/SQB.1986.051.01.053

Pääbo, S. (1989). Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. Proc. Natl. Acad. Sci. USA, 86, 1939–1943.

Pääbo, S., & Wilson, A. C. (1988). Polymerase chain reaction reveals cloning artefacts. Nature, 334(6181), 387–388. https://doi.org/10.1038/334387b0

Pääbo, S., & Wilson, A. C. (1991). Miocene DNA sequences — a dream come true? Current Biology, 1(1), 45–46. https://doi.org/10.1016/0960-9822(91)90125-G

Pala, M., Olivieri, A., Achilli, A., Accetturo, M., Metspalu, E., Reidla, M., Tamm, E., Karmin, M., Reisberg, T., Kashani, B. H., Perego, U. A., Carossa, V., Gandini, F., Pereira, J. B., Soares, P., Angerhofer, N., Rychkov, S., Al-Zahery, N., Carelli, V., Sanati, M. H., Houshmand, M., Hatina, J., Macaulay, V., Pereira, L., Woodward, S. R., Davies, W., Gamble, C., Baird, D., Semino, O., Villems, R., Torroni, A., & Richards, M. B. (2012). Mitochondrial DNA Signals of Late Glacial Recolonization of Europe from Near Eastern Refugia. The American Journal of Human Genetics, 90(5), 915–924. https://doi.org/10.1016/j.ajhg.2012.04.003

Papac, L., de Miguel Ibáñez, P., Rohrlach, A. B., Armendáriz, J., Peres, M., Lamnidis, T. C., Mötsch, A., Schiffels, S., & Risch, R. (2023). Intramural child burials in Iron Age Navarra: How ancient DNA can contribute to household archaeology . In H. Meller, J. Krause, W. Haak, & R. Risch (Eds.), Kinship, Sex, and Biological Relatedness : The contribution of archaeogenetics to the understanding of social and biological relations. 15. Mitteldeutscher Archäologentag vom 6. bis 8. Oktober 2022 in Halle (Saale) (pp. 263–295). https://doi.org/10.11588/propylaeum.1280.c18012

Papac, L., Ernée, M., Dobeš, M., Langová, M., Rohrlach, A. B., Aron, F., Neumann, G. U., Spyrou, M. A., Rohland, N., Velemínský, P., Kuna, M., Brzobohatá, H., Culleton, B.,

Daněček, D., Danielisová, A., Dobisíková, M., Hložek, J., Kennett, D. J., Klementová, J., Kostka, M., Krištuf, P., Kuchařík, M., Hlavová, J. K., Limburský, P., Malyková, D., Mattiello, L., Pecinovská, M., Petriščáková, K., Průchová, E., Stránská, P., Smejtek, L., Špaček, J., Šumberová, R., Švejcar, O., Trefný, M., Vávra, M., Kolář, J., Heyd, V., Krause, J., Pinhasi, R., Reich, D., Schiffels, S., & Haak, W. (2021). Dynamic changes in genomic and social structures in third millennium BCE central Europe. Science Advances, 7(35). https://doi.org/10.1126/sciadv.abi6941

Parker, C., Rohrlach, A. B., Friederich, S., Nagel, S., Meyer, M., Krause, J., Bos, K. I., & Haak, W. (2020). A systematic investigation of human DNA preservation in medieval skeletons. Scientific Reports, 10(1), 18225. https://doi.org/10.1038/s41598-020-75163-w

Patterson, N., Isakov, M., Booth, T., Büster, L., Fischer, C.-E., Olalde, I., Ringbauer, H., Akbari, A., Cheronet, O., Bleasdale, M., Adamski, N., Altena, E., Bernardos, R., Brace, S., Broomandkhoshbacht, N., Callan, K., Candilio, F., Culleton, B., Curtis, E., Demetz, L., Carlson, K. S. D., Edwards, C. J., Fernandes, D. M., Foody, M. G. B., Freilich, S., Goodchild, H., Kearns, A., Lawson, A. M., Lazaridis, I., Mah, M., Mallick, S., Mandl, K., Micco, A., Michel, M., Morante, G. B., Oppenheimer, J., Özdoğan, K. T., Qiu, L., Schattke, C., Stewardson, K., Workman, J. N., Zalzala, F., Zhang, Z., Agustí, B., Allen, T., Almássy, K., Amkreutz, L., Ash, A., Baillif-Ducros, C., Barclay, A., Bartosiewicz, L., Baxter, K., Bernert, Z., Blažek, J., Bodružić, M., Boissinot, P., Bonsall, C., Bradley, P., Brittain, M., Brookes, A., Brown, F., Brown, L., Brunning, R., Budd, C., Burmaz, J., Canet, S., Carnicero-Cáceres, S., Čaušević-Bully, M., Chamberlain, A., Chauvin, S., Clough, S., Čondić, N., Coppa, A., Craig, O., Črešnar, M., Cummings, V., Czifra, S., Danielisová, A., Daniels, R., Davies, A., de Jersey, P., Deacon, J., Deminger, C., Ditchfield, P. W., Dizdar, M., Dobeš, M., Dobisíková, M., Domboróczki, L., Drinkall, G., Đukić, A., Ernée, M., Evans, C., Evans, J., Fernández-Götz, M., Filipović, S., Fitzpatrick, A., Fokkens, H., Fowler, C., Fox, A., Gallina, Z., Gamble, M., González Morales, M. R., González-Rabanal, B., Green, A., Gyenesei, K., Habermehl, D., Hajdu, T., Hamilton, D., Harris, J., Hayden, C., Hendriks, J., Hernu, B., Hey, G., Horňák, M., Ilon, G., Istvánovits, E., Jones, A. M., Kavur, M. B., Kazek, K., Kenyon, R. A., Khreisheh, A., Kiss, V., Kleijne, J., Knight, M., Kootker, L. M., Kovács, P. F., Kozubová, A., Kulcsár, G., Kulcsár, V., Le Pennec, C., Legge, M., Leivers, M., Loe, L., López-Costas, O., Lord, T., Los, D., Lyall, J., Marín-Arroyo, A. B., Mason, P., Matošević, D., Maxted, A., McIntyre, L., McKinley, J., McSweeney, K., Meijlink, B., Mende, B. G., Menđušić, M., Metlička, M., Meyer, S., Mihovilić, K., Milasinovic, L.,

Minnitt, S., Moore, J., Morley, G., Mullan, G., Musilová, M., Neil, B., Nicholls, R., Novak, M., Pala, M., Papworth, M., Paresys, C., Patten, R., Perkić, D., Pesti, K., Petit, A., Petriščáková, K., Pichon, C., Pickard, C., Pilling, Z., Price, T. D., Radović, S., Redfern, R., Resutík, B., Rhodes, D. T., Richards, M. B., Roberts, A., Roefstra, J., Sankot, P., Šefčáková, A., Sheridan, A., Skae, S., Šmolíková, M., Somogyi, K., Somogyvári, Á., Stephens, M., Szabó, G., Szécsényi-Nagy, A., Szeniczey, T., Tabor, J., Tankó, K., Maria, C. T., Terry, R., Teržan, B., Teschler-Nicola, M., Torres-Martínez, J. F., Trapp, J., Turle, R., Ujvári, F., van der Heiden, M., Veleminsky, P., Veselka, B., Vytlačil, Z., Waddington, C., Ware, P., Wilkinson, P., Wilson, L., Wiseman, R., Young, E., Zaninović, J., Žitňan, A., Lalueza-Fox, C., de Knijff, P., Barnes, I., Halkon, P., Thomas, M. G., Kennett, D. J., Cunliffe, B., Lillie, M., Rohland, N., Pinhasi, R., Armit, I., & Reich, D. (2022). Large-scale migration into Britain during the Middle to Late Bronze Age. Nature, 601(7894), 588–594. https://doi.org/10.1038/s41586-021-04287-4

Peck, M. A., Sturk-Andreaggi, K., Thomas, J. T., Oliver, R. S., Barritt-Ross, S., & Marshall, C. (2018). Developmental validation of a Nextera XT mitogenome Illumina MiSeq sequencing method for high-quality samples. Forensic Science International: Genetics, 34, 25–36. https://doi.org/10.1016/j.fsigen.2018.01.004

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. Hydrology and Earth System Sciences, 11(5), 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., & Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biology, 17(1), 60. https://doi.org/10.1186/s13059-016-0918-z

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. https://doi.org/10.1371/journal.pone.0129102

Poinar, H. N., Hofreiter, M., Spaulding, W. G., Martin, P. S., Stankiewicz, B. A., Bland, H., Evershed, R. P., Possnert, G., & Pääbo, S. (1998). Molecular Coproscopy: Dung and Diet

of the Extinct Ground Sloth Nothrotheriops shastensis. Science, 281(5375), 402–406. https://doi.org/10.1126/science.281.5375.402

Poinar, H. N., Höss, M., Bada, J. L., & Pääbo, S. (1996). Amino Acid Racemization and the Preservation of Ancient DNA. Science, 272(5263), 864–866. https://doi.org/10.1126/science.272.5263.864

Poinar, H. N., Schwarz, C., Qi, J., Shapiro, B., MacPhee, R. D. E., Buigues, B., Tikhonov, A., Huson, D. H., Tomsho, L. P., Auch, A., Rampp, M., Miller, W., & Schuster, S. C. (2006). Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. Science, 311(5759), 392–394. https://doi.org/10.1126/science.1123360

Posth, C., Yu, H., Ghalichi, A., Rougier, H., Crevecoeur, I., Huang, Y., Ringbauer, H., Rohrlach, A. B., Nägele, K., Villalba-Mouco, V., Radzeviciute, R., Ferraz, T., Stoessel, A., Tukhbatova, R., Drucker, D. G., Lari, M., Modi, A., Vai, S., Saupe, T., Scheib, C. L., Catalano, G., Pagani, L., Talamo, S., Fewlass, H., Klaric, L., Morala, A., Rué, M., Madelaine, S., Crépin, L., Caverne, J.-B., Bocaege, E., Ricci, S., Boschin, F., Bayle, P., Maureille, B., Le Brun-Ricalens, F., Bordes, J.-G., Oxilia, G., Bortolini, E., Bignon-Lau, O., Debout, G., Orliac, M., Zazzo, A., Sparacello, V., Starnini, E., Sineo, L., van der Plicht, J., Pecqueur, L., Merceron, G., Garcia, G., Leuvrey, J.-M., Garcia, C. B., Gómez-Olivencia, A., Połtowicz-Bobak, M., Bobak, D., Le Luyer, M., Storm, P., Hoffmann, C., Kabaciński, J., Filimonova, T., Shnaider, S., Berezina, N., González-Rabanal, B., González Morales, M. R., Marín-Arroyo, A. B., López, B., Alonso-Llamazares, C., Ronchitelli, A., Polet, C., Jadin, I., Cauwe, N., Soler, J., Coromina, N., Rufí, I., Cottiaux, R., Clark, G., Straus, L. G., Julien, M.-A., Renhart, S., Talaa, D., Benazzi, S., Romandini, M., Amkreutz, L., Bocherens, H., Wißing, C., Villotte, S., de Pablo, J. F.-L., Gómez-Puche, M., Esquembre-Bebia, M. A., Bodu, P., Smits, L., Souffi, B., Jankauskas, R., Kozakaitė, J., Cupillard, C., Benthien, H., Wehrberger, K., Schmitz, R. W., Feine, S. C., Schüler, T., Thevenet, C., Grigorescu, D., Lüth, F., Kotula, A., Piezonka, H., Schopper, F., Svoboda, J., Sázelová, S., Chizhevsky, A., Khokhlov, A., Conard, N. J., Valentin, F., Harvati, K., Semal, P., Jungklaus, B., Suvorov, A., Schulting, R., Moiseyev, V., Mannermaa, K., Buzhilova, A., Terberger, T., Caramelli, D., Altena, E., Haak, W., & Krause, J. (2023). Palaeogenomics of Upper Palaeolithic to Neolithic European hunter-gatherers. Nature, 615(7950), 117–126. https://doi.org/10.1038/s41586-023-05726-0

Posth, C., Zaro, V., Spyrou, M. A., Vai, S., Gnecchi-Ruscone, G. A., Modi, A., Peltzer, A., Mötsch, A., Nägele, K., Vågene, Å. J., Nelson, E. A., Radzevičiūtė, R., Freund, C.,

Bondioli, L. M., Cappuccini, L., Frenzel, H., Pacciani, E., Boschin, F., Capecchi, G., Martini, I., Moroni, A., Ricci, S., Sperduti, A., Turchetti, M. A., Riga, A., Zavattaro, M., Zifferero, A., Heyne, H. O., Fernández-Domínguez, E., Kroonen, G. J., McCormick, M., Haak, W., Lari, M., Barbujani, G., Bondioli, L., Bos, K. I., Caramelli, D., & Krause, J. (2021). The origin and legacy of the Etruscans through a 2000-year archeogenomic time transect. Science Advances, 7(39). https://doi.org/10.1126/sciadv.abi7673

Raff, J. (2024). Engaged research with indigenous communities in North America: an example from the central and southern plains. EMBO | EMBL Symposium: Reconstructing the Human Past: Using Ancient and Modern Genomics.

Ramos, A., Santos, C., Mateiu, L., Gonzalez, M. del M., Alvarez, L., Azevedo, L., Amorim, A., & Aluja, M. P. (2013). Frequency and Pattern of Heteroplasmy in the Complete Human Mitochondrial Genome. PLoS ONE, 8(10), e74636. https://doi.org/10.1371/journal.pone.0074636

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J. S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., Bertalan, M., Nielsen, K., Gilbert, M. T. P., Wang, Y., Raghavan, M., Campos, P. F., Kamp, H. M., Wilson, A. S., Gledhill, A., Tridico, S., Bunce, M., Lorenzen, E. D., Binladen, J., Guo, X., Zhao, J., Zhang, X., Zhang, H., Li, Z., Chen, M., Orlando, L., Kristiansen, K., Bak, M., Tommerup, N., Bendixen, C., Pierre, T. L., Grønnow, B., Meldgaard, M., Andreasen, C., Fedorova, S. A., Osipova, L. P., Higham, T. F. G., Ramsey, C. B., Hansen, T. v. O., Nielsen, F. C., Crawford, M. H., Brunak, S., Sicheritz-Pontén, T., Villems, R., Nielsen, R., Krogh, A., Wang, J., & Willerslev, E. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature, 463(7282), 757–762. https://doi.org/10.1038/nature08835

Rathbun, M. M., McElhoe, J. A., Parson, W., & Holland, M. M. (2017). Considering DNA damage when interpreting mtDNA heteroplasmy in deep sequencing data. Forensic Science International: Genetics, 26, 1–11. https://doi.org/10.1016/j.fsigen.2016.09.008

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: advances and challenges in museomics. Trends in Ecology & Evolution, 36(11), 1049–1060. https://doi.org/10.1016/j.tree.2021.07.009

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., Stoneking, M., Richards, M., Talamo,

S., Shunkov, M. V., Derevianko, A. P., Hublin, J.-J., Kelso, J., Slatkin, M., & Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature, 468(7327), 1053–1060. https://doi.org/10.1038/nature09710

Renaud, G., Slon, V., Duggan, A. T., & Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biology, 16(1), 224. https://doi.org/10.1186/s13059-015-0776-0

Richard, M. (2011). Carthage Must Be Destroyed: The Rise and Fall of an Ancient Civilization (Penguin (ed.)).

Richardson, J. S. (1986). Hispaniae. Cambridge University Press. https://doi.org/10.1017/CBO9780511518591

Risch, R., Haak, W., Krause, J., & Meller, H. (2023). Kinship, Sex, and Biological Relatedness – The contribution of archaeogenetics to the understanding of social relations. In H. Meller, J. Krause, W. Haak, & R. Risch (Eds.), Kinship, Sex, and Biological Relatedness : The contribution of archaeogenetics to the understanding of social and biological relations. 15. Mitteldeutscher Archäologentag vom 6. bis 8. Oktober 2022 in Halle (Saale) (pp. 9–25). https://doi.org/10.11588/propylaeum.1280.c18044

Rivollat, M., Jeong, C., Schiffels, S., Küçükkalıpçı, İ., Pemonge, M.-H., Rohrlach, A. B., Alt, K. W., Binder, D., Friederich, S., Ghesquière, E., Gronenborn, D., Laporte, L., Lefranc, P., Meller, H., Réveillas, H., Rosenstock, E., Rottier, S., Scarre, C., Soler, L., Wahl, J., Krause, J., Deguilloux, M.-F., & Haak, W. (2020). Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. Science Advances, 6(22). https://doi.org/10.1126/sciadv.aaz5344

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. Nature Biotechnology, 29(1), 24–26. https://doi.org/10.1038/nbt.1754

Roca-Rada, X., Davidson, R., Williams, M. P., Ravishankar, S., Collen, E., Haarkötter, C., Taufik, L., Carvalho, A. F., Villalba-Mouco, V., Cuesta-Aguirre, D. R., Tente, C., Monge Calleja, Á. M., MacRoberts, R. A., Melo, L., Purnomo, G. A., Souilmi, Y., Tobler, R., Cunha, E., Tereso, S., Matos, V. M. J., Fernandes, T. M., Mauer, A.-F., Silva, A. M., Carvalho, P. C., Llamas, B., & Teixeira, J. C. (2024). The genetic history of Portugal over the past 5,000 years. BioRxiv. https://doi.org/10.1101/2024.09.12.612544

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., & Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1660), 20130624. https://doi.org/10.1098/rstb.2013.0624

Rohland, N., & Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. Nature Protocols, 2(7), 1756–1762. https://doi.org/10.1038/nprot.2007.247

Rohland, N., Mallick, S., Mah, M., Maier, R., Patterson, N., & Reich, D. (2022). Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs. Genome Research, 32(11–12), 2068–2078. https://doi.org/10.1101/gr.276728.122

Rollo, F., Amici, A., Salvi, R., & Garbuglia, A. (1988). Short but faithful pieces of ancient DNA. Nature, 335(6193), 774–774. https://doi.org/10.1038/335774a0

Ruiz de Arbulo, J. (2006). "Scipionum opus" and something more. In L. Abad Casal, S. J. Keay, & S. F. Ramallo Asensio (Eds.), Early roman towns in Hispania Tarraconensis (pp. 32–43).

Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K., & Erlich, H. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science, 239(4839), 487–491. https://doi.org/10.1126/science.239.4839.487

Sampietro, M. L., Caramelli, D., Lao, O., Calafell, F., Comas, D., Lari, M., Agustí, B., Bertranpetit, J., & Lalueza-Fox, C. (2005). The Genetics of the Pre-Roman Iberian Peninsula: A mtDNA Study of Ancient Iberians. Annals of Human Genetics, 69(5), 535–548. https://doi.org/10.1111/j.1529-8817.2005.00194.x

Sandoval-Ávila, C., Martirosyan, A., Cuesta-Aguirre, D. R., Jordana, X., Nociarová, D., De Prado, G., Santos, C., & Malgosa, A. (2025). Muerte infantil en época ibérica: el complejo arqueológico de Ullastret (Girona, Cataluña). Trabajos de Prehistoria, 81(2), 959. https://doi.org/10.3989/tp.2024.959

Sanmartí, J. (2003). Els ibers quaranta anys més tard : Un breu balanç de la recerca. Cota Zero: Revista d'arqueologia i Ciència, 18, 147–157. https://raco.cat/index.php/CotaZero/article/view/67358

Sanmartí, J. (2011). L'Escola Catalana d'Arqueologia i l'estudi Dels Ibers. In Institut d'Estudis Catalans, Secció Històrico-Arqueològica.

Sanmartí, J. (2013). Joan Maluquer de Motes i els estudis ibèrics. Revista d'arqueologia de Ponent, 23, 343–352. https://raco.cat/index.php/RAP/article/view/283579

Sanmartí, J., Santacana, J., Belarte, M. C., & Asensio, D. (1998). Paisatges ibèrics: Tipus d'assentaments i formes d'ocupació del territori a la costa central de Catalunya durant el període ibèric ple. SAGVNTVM Extra, 1, 373–385. https://turia.uv.es/index.php/saguntumextra/article/view/2762

Santos, C., Montiel, R., Sierra, B., Bettencourt, C., Fernandez, E., Alvarez, L., Lima, M., Abade, A., & Aluja, M. P. (2005). Understanding Differences Between Phylogenetic and Pedigree-Derived mtDNA Mutation Rate: A Model Using Families from the Azores Islands (Portugal). Molecular Biology and Evolution, 22(6), 1490–1505. https://doi.org/10.1093/molbev/msi141

Santos, C., Sierra, B., Álvarez, L., Ramos, A., Fernández, E., Nogués, R., & Aluja, M. P. (2008). Frequency and Pattern of Heteroplasmy in the Control Region of Human Mitochondrial DNA. Journal of Molecular Evolution, 67(2), 191–200. https://doi.org/10.1007/s00239-008-9138-9

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Pääbo, S. (2012). Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. PLoS ONE, 7(3), e34131. https://doi.org/10.1371/journal.pone.0034131

Schullard, H. H. (1989). Carthage and Rome. In The Cambridge Ancient History: VII Part 2, The Rise of Rome to 220 BC (2nd edition, pp. 486–569). Cambridge University Press.

Schwarz, C., Debruyne, R., Kuch, M., McNally, E., Schwarcz, H., Aubrey, A. D., Bada, J., & Poinar, H. (2009). New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. Nucleic Acids Research, 37(10), 3215–3229. https://doi.org/10.1093/nar/gkp159

Silva, N. M., Kreutzer, S., Souleles, A., Triantaphyllou, S., Kotsakis, K., Urem-Kotsou, D., Halstead, P., Efstratiou, N., Kotsos, S., Karamitrou-Mentessidi, G., Adaktylou, F., Chondroyianni-Metoki, A., Pappa, M., Ziota, C., Sampson, A., Papathanasiou, A., Vitelli, K., Cullen, T., Kyparissi-Apostolika, N., Lanz, A. Z., Peters, J., Rio, J., Wegmann, D., Burger, J., Currat, M., & Papageorgopoulou, C. (2022). Ancient mitochondrial diversity reveals population homogeneity in Neolithic Greece and identifies population dynamics along the Danubian expansion axis. Scientific Reports, 12(1), 13474. https://doi.org/10.1038/s41598-022-16745-8

Simões, L. G., Günther, T., Martínez-Sánchez, R. M., Vera-Rodríguez, J. C., Iriarte, E., Rodríguez-Varela, R., Bokbot, Y., Valdiosera, C., & Jakobsson, M. (2023). Northwest African Neolithic initiated by migrants from Iberia and Levant. Nature, 618(7965), 550–556. https://doi.org/10.1038/s41586-023-06166-6

Simón, M., Jordana, X., Armentano, N., Santos, C., Díaz, N., Solórzano, E., López, J. B., González-Ruiz, M., & Malgosa, A. (2011). The presence of nuclear families in prehistoric collective burials revisited: The bronze age burial of montanissell cave (Spain) in the light of aDNA. American Journal of Physical Anthropology, 146(3), 406–413. https://doi.org/10.1002/ajpa.21590

Sinner, A. G. (2015). Cultural contacts and identity construction: A colonial context in NE Spain (2nd - Early 1st c. B.C.). Journal of Roman Archaeology, 28(1), 7–37. https://doi.org/10.1017/S1047759415002391

Sinner, A. G., & Ferrer i Jané, J. (2016). Del *oppidum* de Burriac a las termas de Ca l'Arnau. Una aproximación a la lengua y a la identidad de los habitantes de *Ilduro* (Cabrera de Mar, Barcelona). Archivo Español de Arqueología, 89(November), 193. https://doi.org/10.3989/aespa.089.016.010

Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J., & Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proceedings of the National Academy of Sciences, 111(6), 2229–2234. https://doi.org/10.1073/pnas.1318934111

Skoglund, P., Malmström, H., Omrak, A., Raghavan, M., Valdiosera, C., Günther, T., Hall, P., Tambets, K., Parik, J., Sjögren, K.-G., Apel, J., Willerslev, E., Storå, J., Götherström, A., & Jakobsson, M. (2014). Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. Science, 344(6185), 747–750. https://doi.org/10.1126/science.1253448

Skourtanioti, E., Erdal, Y. S., Frangipane, M., Balossi Restelli, F., Yener, K. A., Pinnock, F., Matthiae, P., Özbal, R., Schoop, U.-D., Guliyev, F., Akhundov, T., Lyonnet, B., Hammer, E. L., Nugent, S. E., Burri, M., Neumann, G. U., Penske, S., Ingman, T., Akar, M., Shafiq, R., Palumbi, G., Eisenmann, S., D'Andrea, M., Rohrlach, A. B., Warinner, C., Jeong, C., Stockhammer, P. W., Haak, W., & Krause, J. (2020). Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. Cell, 181(5), 1158-1175.e28. https://doi.org/10.1016/j.cell.2020.04.044

Smith, C. I., Chamberlain, A. T., Riley, M. S., Cooper, A., Stringer, C. B., & Collins, M. J. (2001). Not just old but old and cold? Nature, 410(6830), 771–772. https://doi.org/10.1038/35071177

Soares, P., Alshamali, F., Pereira, J. B., Fernandes, V., Silva, N. M., Afonso, C., Costa, M. D., Musilova, E., Macaulay, V., Richards, M. B., Cerny, V., & Pereira, L. (2012). The Expansion of mtDNA Haplogroup L3 within and out of Africa. Molecular Biology and Evolution, 29(3), 915–927. https://doi.org/10.1093/molbev/msr245

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V., & Richards, M. B. (2009). Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. The American Journal of Human Genetics, 84(6), 740–759. https://doi.org/10.1016/j.ajhg.2009.05.001

Spann, P. O. (1987). Sertorius - Philip O. Spann: Quintus Sertorius and the Legacy of Sulla.

Staats, M., Cuenca, A., Richardson, J. E., Vrielink-van Ginkel, R., Petersen, G., Seberg, O., & Bakker, F. T. (2011). DNA Damage in Plant Herbarium Tissue. PLoS ONE, 6(12), e28448. https://doi.org/10.1371/journal.pone.0028448

Stiller, M., Green, R. E., Ronan, M., Simons, J. F., Du, L., He, W., Egholm, M., Rothberg, J. M., Keats, S. G., Ovodov, N. D., Antipina, E. E., Baryshnikov, G. F., Kuzmin, Y. V., Vasilevski, A. A., Wuenschell, G. E., Termini, J., Hofreiter, M., Jaenicke-Després, V., & Pääbo, S. (2006). Patterns of Nucleotide Misincorporations during Enzymatic Amplification and Direct Large-scale Sequencing of Ancient DNA. Proceedings of the National Academy of Sciences, 103(3), 1358–13584.

Stoler, N., & Nekrutenko, A. (2021). Sequencing error profiles of Illumina sequencing instruments. NAR Genomics and Bioinformatics, 3(1). https://doi.org/10.1093/nargab/lqab019

Subirà, M. E., & Molist, N. (2007). Inhumacions perinatals múltiples i espais de treball en els assentaments ibers. Nasciturus: Infans, Puerulus. Vobis Mater Terra. La Muerte En La Infancia, Uh 4, 365–385.

Sun, S., Osterman, M. D., & Li, M. (2019). Tissue specificity of DNA damage response and tumorigenesis. Cancer Biology & Medicine, 16(3), 396–414. https://doi.org/10.20892/j.issn.2095-3941.2019.0097

Szécsényi-Nagy, A., Roth, C., Brandt, G., Rihuete-Herrada, C., Tejedor-Rodríguez, C., Held, P., García-Martínez-de-Lagrán, Í., Arcusa Magallón, H., Zesch, S., Knipper, C., Bánffy, E., Friederich, S., Meller, H., Bueno Ramírez, P., Barroso Bermejo, R., de Balbín Behrmann, R., Herrero-Corral, A. M., Flores Fernández, R., Alonso Fernández, C., Jiménez Echevarria, J., Rindlisbacher, L., Oliart, C., Fregeiro, M.-I., Soriano, I., Vicente, O., Micó, R., Lull, V., Soler Díaz, J., López Padilla, J. A., Roca de Togores Muñoz, C., Hernández Pérez, M. S., Jover Maestre, F. J., Lomba Maurandi, J., Avilés Fernández, A., Lillios, K. T., Silva, A. M., Magalhães Ramalho, M., Oosterbeek, L. M., Cunha, C., Waterman, A. J., Roig Buxó, J., Martínez, A., Ponce Martínez, J., Hunt Ortiz, M., Mejías-García, J. C., Pecero Espín, J. C., Cruz-Auñón Briones, R., Tomé, T., Carmona Ballestero, E., Cardoso, J. L., Araújo, A. C., Liesau von Lettow-Vorbeck, C., Blasco Bosqued, C., Ríos Mendoza, P., Pujante, A., Royo-Guillén, J. I., Esquembre Beviá, M. A., Dos Santos Goncalves, V. M., Parreira, R., Morán Hernández, E., Méndez Izquierdo, E., Vega y Miguel, J., Menduiña García, R., Martínez Calvo, V., López Jiménez, O., Krause, J., Pichler, S. L., Garrido-Pena, R., Kunst, M., Risch, R., Rojo-Guerra, M. A., Haak, W., & Alt, K. W. (2017). The maternal genetic make-up of the Iberian Peninsula between the Neolithic and the Early Bronze Age. Scientific Reports, 7(1), 15644. https://doi.org/10.1038/s41598-017-15480-9

Tarradell, M. (1962). Les Arrels de Catalunya. Vicens-Vives.

Thomas, R. H., Schaffner, W., Wilson, A. C., & Pääbo, S. (1989). DNA phylogeny of the extinct marsupial wolf. Nature, 340(6233), 465–467. https://doi.org/10.1038/340465a0

Torres, J. B. (2016). A history of you, me, and humanity: mitochondrial DNA in anthropological research. AIMS Genetics, 03(02), 146–156. https://doi.org/10.3934/genet.2016.2.146

Universitat de Lleida. (2020). La Fortalesa del Vilars d'Arbeca. Terra, aigua i poder en el món iber (p. 166). Universitat de Lleida. https://vilars.udl.cat/export/sites/vilars/ca/.galleries/Documents/Cataleg-Vilars-Reduit2.pdf

Valdiosera, C., Günther, T., Vera-Rodríguez, J. C., Ureña, I., Iriarte, E., Rodríguez-Varela, R., Simões, L. G., Martínez-Sánchez, R. M., Svensson, E. M., Malmström, H., Rodríguez, L., Bermúdez de Castro, J.-M., Carbonell, E., Alday, A., Hernández Vera, J. A., Götherström, A., Carretero, J.-M., Arsuaga, J. L., Smith, C. I., & Jakobsson, M. (2018). Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the

far end of Eurasia. Proceedings of the National Academy of Sciences, 115(13), 3428–3433. https://doi.org/10.1073/pnas.1717762115

van de Loosdrecht, M., Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., Nickel, B., Nagel, S., Talbi, E. H., El Hajraoui, M. A., Amzazi, S., Hublin, J.-J., Pääbo, S., Schiffels, S., Meyer, M., Haak, W., Jeong, C., & Krause, J. (2018). Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. Science, 360(6388), 548–552. https://doi.org/10.1126/science.aar8380

van der Valk, T., Pečnerová, P., Díez-del-Molino, D., Bergström, A., Oppenheimer, J., Hartmann, S., Xenikoudakis, G., Thomas, J. A., Dehasque, M., Sağlıcan, E., Fidan, F. R., Barnes, I., Liu, S., Somel, M., Heintzman, P. D., Nikolskiy, P., Shapiro, B., Skoglund, P., Hofreiter, M., Lister, A. M., Götherström, A., & Dalén, L. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. Nature, 591(7849), 265–269. https://doi.org/10.1038/s41586-021-03224-9

van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Human Mutation, 30(2), E386–E394. https://doi.org/10.1002/humu.20921

Villa-Islas, V., Izarraras-Gomez, A., Larena, M., Campos, E. M. P., Sandoval-Velasco, M., Rodríguez-Rodríguez, J. E., Bravo-Lopez, M., Moguel, B., Fregel, R., Garfias-Morales, E., Medina Tretmanis, J., Velázquez-Ramírez, D. A., Herrera-Muñóz, A., Sandoval, K., Nieves-Colón, M. A., Zepeda García Moreno, G., Villanea, F. A., Medina, E. F. V., Aguayo-Haro, R., Valdiosera, C., Ioannidis, A. G., Moreno-Estrada, A., Jay, F., Huerta-Sanchez, E., Moreno-Mayar, J. V., Sánchez-Quinto, F., & Ávila-Arcos, M. C. (2023). Demographic history and genetic structure in pre-Hispanic Central Mexico. Science, 380(6645). https://doi.org/10.1126/science.add6142

Villalba-Mouco, V., Oliart, C., Rihuete-Herrada, C., Childebayeva, A., Rohrlach, A. B., Fregeiro, M. I., Celdrán Beltrán, E., Velasco-Felipe, C., Aron, F., Himmel, M., Freund, C., Alt, K. W., Salazar-García, D. C., García Atiénzar, G., de Miguel Ibáñez, M. P., Hernández Pérez, M. S., Barciela, V., Romero, A., Ponce, J., Martínez, A., Lomba, J., Soler, J., Martínez, A. P., Avilés Fernández, A., Haber-Uriarte, M., Roca de Togores Muñoz, C., Olalde, I., Lalueza-Fox, C., Reich, D., Krause, J., García Sanjuán, L., Lull, V., Micó, R., Risch, R., & Haak, W. (2021). Genomic transformation and social organization during

the Copper Age–Bronze Age transition in southern Iberia. Science Advances, 7(47). https://doi.org/10.1126/sciadv.abi7038

Villalba-Mouco, V., Oliart, C., Rihuete-Herrada, C., Rohrlach, A. B., Fregeiro, M. I., Childebayeva, A., Ringbauer, H., Olalde, I., Celdrán Beltrán, E., Puello-Mora, C., Valério, M., Krause, J., Lull, V., Micó, R., Risch, R., & Haak, W. (2022). Kinship practices in the early state El Argar society from Bronze Age Iberia. Scientific Reports, 12(1), 22415. https://doi.org/10.1038/s41598-022-25975-9

Villalba-Mouco, V., van de Loosdrecht, M. S., Rohrlach, A. B., Fewlass, H., Talamo, S., Yu, H., Aron, F., Lalueza-Fox, C., Cabello, L., Cantalejo Duarte, P., Ramos-Muñoz, J., Posth, C., Krause, J., Weniger, G.-C., & Haak, W. (2023). A 23,000-year-old southern Iberian individual links human groups that lived in Western Europe before and after the Last Glacial Maximum. Nature Ecology & Evolution. https://doi.org/10.1038/s41559-023-01987-0

Vinueza-Espinosa, D. C. (2022). Methodological Challenges in Ancient and Forensic DNA Analysis : Improvements in DNA Extraction and Geneti Characterization from Human Skeletal Remains. Universitat Autònoma de Barcelona.

Vinueza-Espinosa, D. C., Cuesta-Aguirre, D. R., Malgosa, A., & Santos, C. (2023). Mitochondrial DNA control region typing from highly degraded skeletal remains by single-multiplex next-generation sequencing. ELECTROPHORESIS, 44(17–18), 1423–1434. https://doi.org/10.1002/elps.202200052

Vinueza-Espinosa, D. C., Santos, C., Martínez-Labarga, C., & Malgosa, A. (2020). Human DNA extraction from highly degraded skeletal remains: How to find a suitable method? ELECTROPHORESIS, 41(24), 2149–2158. https://doi.org/10.1002/elps.202000171

Vives, S., Gilbert, M. T., Arenas, C., Gigli, E., Lao, O., & Lalueza-Fox, C. (2008). Statistical analysis of post mortem DNA damage-derived miscoding lesions in Neandertal mitochondrial DNA. BMC Research Notes, 1(1), 40. https://doi.org/10.1186/1756-0500-1-40

Walden, K. K., & Robertson, H. M. (1997). Ancient DNA from amber fossil bees? Molecular Biology and Evolution, 14(10), 1075–1077. https://doi.org/10.1093/oxfordjournals.molbev.a025713

Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D. A., & Cooper, A. (2003). Diverse Plant and Animal Genetic Records from Holocene and Pleistocene Sediments. Science, 300(5620), 791–795. https://doi.org/10.1126/science.1084114

Winters, C. (2010). The African Origin of mtDNA Haplogroup M1. Current Research Journal of Biological Sciences, 2(6), 380–389.

Witas, H. W., & Zawicki, P. (2004). Mitochondrial DNA and human evolution: A review. Anthropol. Rev, 67, 97–110.

Woodward, S. R., Weyand, N. J., & Bunnell, M. (1994). DNA Sequence from Cretaceous Period Bone Fragments. Science, 266(5188), 1229–1232. https://doi.org/10.1126/science.7973705

Xavier, M. J., Nixon, B., Roman, S. D., & Aitken, R. J. (2018). Improved methods of DNA extraction from human spermatozoa that mitigate experimentally-induced oxidative DNA damage. PLOS ONE, 13(3), e0195003. https://doi.org/10.1371/journal.pone.0195003

Yaka, R., Mapelli, I., Kaptan, D., Doğu, A., Chyleński, M., Erdal, Ö. D., Koptekin, D., Vural, K. B., Bayliss, A., Mazzucato, C., Fer, E., Çokoğlu, S. S., Lagerholm, V. K., Krzewińska, M., Karamurat, C., Gemici, H. C., Sevkar, A., Dağtaş, N. D., Kılınç, G. M., Adams, D., Munters, A. R., Sağlıcan, E., Milella, M., Schotsmans, E. M. J., Yurtman, E., Çetin, M., Yorulmaz, S., Altınışık, N. E., Ghalichi, A., Juras, A., Bilgin, C. C., Günther, T., Storå, J., Jakobsson, M., de Kleijn, M., Mustafaoğlu, G., Fairbairn, A., Pearson, J., Togan, İ., Kayacan, N., Marciniak, A., Larsen, C. S., Hodder, I., Atakuman, Ç., Pilloud, M., Sürer, E., Gerritsen, F., Özbal, R., Baird, D., Erdal, Y. S., Duru, G., Özbaşaran, M., Haddow, S. D., Knüsel, C. J., Götherström, A., Özer, F., & Somel, M. (2021). Variable kinship patterns in Neolithic Anatolia revealed by ancient genomes. Current Biology, 31(11), 2455-2468.e18. https://doi.org/10.1016/j.cub.2021.03.050

Yu, H., van de Loosdrecht, M. S., Mannino, M. A., Talamo, S., Rohrlach, A. B., Childebayeva, A., Villalba-Mouco, V., Aron, F., Brandt, G., Burri, M., Freund, C., Radzeviciute, R., Stahl, R., Wissgott, A., Fewlass, H., Tagliacozzo, A., Piperno, M., Tusa, S., Collina, C., Schimmenti, V., Di Salvo, R., Prüfer, K., Posth, C., Hublin, J.-J., Gronenborn, D., Binder, D., Jeong, C., Haak, W., & Krause, J. (2022). Genomic and dietary discontinuities during

the Mesolithic and Neolithic in Sicily. IScience, 25(5), 104244. https://doi.org/10.1016/j.isci.2022.104244

Zalloua, P., Collins, C. J., Gosling, A., Biagini, S. A., Costa, B., Kardailsky, O., Nigro, L., Khalil, W., Calafell, F., & Matisoo-Smith, E. (2018). Ancient DNA of Phoenician remains indicates discontinuity in the settlement history of Ibiza. Scientific Reports, 8(1), 17567. https://doi.org/10.1038/s41598-018-35667-y

Zhu, K., He, H., Tao, L., Ma, H., Yang, X., Wang, R., Guo, J., & Wang, C.-C. (2024). Protocol for a comprehensive pipeline to study ancient human genomes. STAR Protocols, 5(2), 102985. https://doi.org/10.1016/j.xpro.2024.102985

Zimmermann, K. (2011). Roman Strategy and Aims in the Second Punic War [Bookitem]. In A Companion to the Punic Wars (pp. 280–298). Wiley. https://doi.org/10.1002/9781444393712.ch16

Zupanič Pajnič, I., Jeromelj, T., & Leskovar, T. (2025). Petrous bones versus tooth cementum for genetic analysis of aged skeletal remains. International Journal of Legal Medicine, 139(1), 41–48. https://doi.org/10.1007/s00414-024-03346-5



8.

SUPPLEMENTARY MATERIAL

8. SUPPLEMENTARY MATERIAL

The Supplementary Material for each article is available in the directory, numbered according to the chapters of this thesis, at the following link: https://uab-my.sharepoint.com/:f:/g/personal/2103711_uab_cat/EhLQT7GxLs1GoowdsaXF7QYBF7DKCwofUOyqhbjFw4XWRw?e=3ddWqy

