

Human demographic insights from paleogenetic studies of American and Iberian skeletal remains

Xavier Roca Rada

Principal Supervisor: Assoc. Prof. Bastien Llamas

Co-supervisors: Dr. João C. Teixeira and Dr. Yassine Souilmi

Australian Centre for Ancient DNA (ACAD)

School of Biological Sciences

The University of Adelaide



THE UNIVERSITY
of ADELAIDE

This thesis is submitted in fulfilment of the requirements for
the degree of Doctor of Philosophy

October 2022

I acknowledge the Kaurna people,
the original custodians of the Adelaide Plains
and the land on which I worked and lived during my PhD.
I extend my respects to Elders past, present, and emerging.
I acknowledge that sovereignty was never ceded.

Table of contents

Thesis Abstract	4
Thesis Declaration	6
Acknowledgements	7
Thesis Introduction	9
Part I: Iberia	24
<u>Chapter I: A 1000-year-old case of Klinefelter’s syndrome diagnosed by integrating morphology, osteology, and genetics</u>	25
Supplementary appendix	31
<u>Chapter II: Paleogenetic studies of Mas d’en Boixos (Catalonia, Spain)</u>	61
1. Introduction	64
2. Methods	67
3. Results	72
4. Discussion	86
5. References	91
Part II: Pre-Columbian Mesoamerica	96
<u>Chapter III: Ancient DNA Studies in Pre-Columbian Mesoamerica</u>	97
<u>Chapter IV: The Genetic History of Pre-Columbian Mesoamerica</u>	118
1. Introduction	121
2. Methods	121
3. Results	131
4. Discussion	153
5. References	158
Part III: Southern Cone of South America	163
<u>Chapter V: Ancient mitochondrial genomes from the Argentinian Pampas inform the early peopling of the Southern Cone of South America</u>	164
STAR Methods	181
Supplemental information	186
Thesis Discussion	187
Appendix	203

Thesis Abstract

The study of the past through the examination of preserved genetic material from ancient remains has revolutionised the fields of archaeology, anthropology, and evolutionary biology. Ancient DNA can be used as a tool to tackle several questions from a narrow to a broad view, i.e., from a specific archaeological context to global demographic processes. Indeed, in this thesis I covered a large range of projects following this perspective, grouping them according to their respective geographical area: Iberia, Mesoamerica, and the Southern Cone of South America.

Part I focuses on the Iberian Peninsula, the westernmost edge of Eurasia and probably one of the most distant regions during the peopling of the continent. In Chapter I, I describe a 1,000-year-old case of Klinefelter's Syndrome, from an individual unearthed at the mediaeval archaeological site of Torre Velha (Portugal). This work provides an unprecedented integration of morphological, osteological, and genetic evidence to confirm the clinical diagnosis. It also introduces a novel statistical implementation that allows to probabilistically assign individuals to karyotypes based on the number of reads mapping to the X, Y, and autosomal chromosomes. In Chapter II, I investigate the genetic context of the Mas d'en Boixos (MDB) archaeological site—located in the central Pre-littoral depression of Catalonia, Spain—analysing two individuals from the Middle Neolithic, nine from the Early Bronze Age and three previously published from the Iron Age. I infer kinship relationships between individuals buried in close proximity, determine diachronic genetic changes at MDB, contextualise MDB within the rest of Iberia and investigate the Bronze Age transition in Iberia.

Part II focuses on the study of pre-Columbian Mesoamerica, a historical and cultural region in Central America whose people probably carried an invaluable genetic diversity partly lost during the Spanish conquest and the subsequent colonial period. In Chapter III, I review

previously published ancient DNA studies in pre-Columbian Mesoamerica. Despite a rich archaeological record and Indigenous cultural diversity, aDNA studies are limited and have thus far focused on the study of mitochondrial DNA regions. In Chapter IV, I investigate the past genetic history of pre-Columbian Mesoamerica by analysing 25 newly sequenced ancient genomes from five different archaeological sites across Mesoamerica spanning the time periods between 320 and 1,400 CE. Despite the small sample size, this chapter offers the first (and ongoing) Mesoamerican aDNA genetic screening that includes full mitogenomes and nuclear data.

Part III focuses on the Southern Cone of South America (SCSA), a key study region to build a comprehensive picture of the peopling of the Americas as it is geographically the most distant location from the entry point, Beringia. In Chapter V, I investigate the early peopling of the SCSA analysing eighteen ancient human mitochondrial genomes—seven of which are novel—from three Early to Late Holocene (10,000–1,500 years ago) archaeological sites from the understudied Argentinian Pampas and assessing its influence in the southernmost regions.

Thesis Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

The author acknowledges that copyright of published works contained within the thesis resides with the copyright holder(s) of those works.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Acknowledgements

I am grateful to all the people who have come before me and made this research possible.

To my supervisors João Teixeira, Yassine Souilmi and Bastien Llamas. Thank you for all your support throughout this process. You have been extremely supportive both scientifically and emotionally. Without you all, I would have not made it through. João, thank you for your training, your critical thinking, your interesting conversations and all your proposals about cool projects that are impossible to say no to. Thank you for showing me around Portugal. Yassine, thank you for all your training and help in bioinformatics and for always going beyond to all the questions that I have, it really helped me to broaden my horizon. Bastien, thank you for trusting me from the beginning. It's been a long journey, from twitter and sightseeing in Barcelona to all the way to Australia. A dream you made possible. Thank you for all your training and help in the lab, in bioinformatics, and editing. Thanks for all your emotional support and always having time for me, for being so caring and approachable and always having a smile on your face and sweet cakes for all of us. You were the best principal supervisor I could have asked for.

To everyone that was and still is part of ACAD for the great conversations and the fun breaks. Specially, to Holly Heiniger, Yichen Liu, Matthew Williams, Alex Salis, and Vilma Perez for training me in the lab and having the patience with all my mess-ups as well as for all the laughs we had. To Evelyn Collen, Gina Guzzo, Caitlin Mudge, Gludhug Purnomo, Matilda Handsley-Davis, Raph Eisenhofer, Adrien Oliva, and Olivia Johnson for all the great conversations and all the fun in and out of the office. To Roberta Davidson for being my non-official supervisor and being essential to my life from now on.

To all my collaborators in Australia, US, Mexico, Argentina, Portugal, France, Spain, Germany, and all around the globe. You made all these projects possible, and I will be thankful for evermore.

To Iñigo Olalde and Rui Martiniano for their excellent work examining my thesis and their kind words.

A la Cristina Santos, l'Assumpció Malgosa i la Diana Vinueza que em van obrir les portes encantades per aprendre i treballar amb elles. Gràcies per introduir-me al món de l'ADN antic i per continuar treballant junts.

Als meus pares, Pere i Antònia, i al meu germà, Adrià, perquè sense el vostre suport res hagués sigut possible. Gràcies per estar sempre presents malgrat la distància i la pandèmia. Per ser els més atents del món, per donar-me totes les oportunitats de poder estudiar el que volia, per confiar en mi i per recolzar-me incondicionalment. Espero que estigueu orgullosos. A tots els meus avis per donar-me tant d'amor, suport i felicitat. A la meva iaia Teresina, que siguis on siguis, et trobo molt a faltar, espero que estiguis orgullosa. A tota la meva família materna i paterna pels ànims.

A tots els meus amics de casa que malgrat la distància han estat sempre presents.

To all my friends in Adelaide, thanks for all your support, trips, hikes, and parties. Thanks for giving me the opportunity to have a great work-life balance.

To all my favorite artists, specially to Taylor Swift, for accompanying me with their music during all those long hours in the lab, running analysis and writing.

Thesis Introduction

1. The advent of molecular anthropology

Learning about the past enables us to develop a better understanding of the world we currently live in. The study of human origins and evolution is central to various academic disciplines, such as philosophy, archaeology, palaeontology, anthropology, or linguistics, each of which brought us closer to a holistic understanding of who we are.

In 1962, at the Burg Wartenstein symposium on “Classification and Human Evolution” in Austria, Emile Zuckerkandl coined the term “molecular anthropology” to designate the study of primate phylogeny and human evolution through the study of the genetic information contained in proteins and nucleic acids^{1,2}. The pioneering work on mitochondrial genetic variation in worldwide human populations by Rebecca Cann and colleagues in the late 1980s was among the most important landmark studies that launched the field of molecular anthropology. Their findings of a recent African origin of our species revolutionised our understanding of human evolution and led to the spread of the popular concept of “Mitochondrial Eve”, regarded as the matrilineal most recent common genetic ancestor of all living humans³.

2. Global genetic diversity

In the early 1990s, the Human Genome Diversity Project (HGDP) led by Luigi Luca Cavalli-Sforza and collaborators marked the transition of the field from genetics to genomics through the study of genetic information contained in the cell nucleus. The HGDP aimed to explore human demographic history by looking at genomes from over fifty geographically, linguistically, and culturally diverse human populations across the world⁴. Despite the significant contributions to human population genetics, molecular anthropology, and heritable

disease research, the HGDP raised some ethical, philosophical, and transcultural concerns that affected the trust in genetic studies focusing on present-day Indigenous populations⁵.

Nowadays, molecular anthropology is moving towards being a more inclusive and ethically aware field and uses novel tools and techniques from the fields of molecular biology and bioinformatics to answer questions about the geographical origins of human populations, the spread of people around the globe, our relationship to other great apes, and the molecular adaptations to novel environments⁶. However, pre-historical and historical demographic events—including the European colonial expansion and more recent globalisation—resulted in admixture, population collapse and migrations that can confound human genetic history inferences⁷.

3. Ancient DNA to travel back in time

Confounding demographic events can be circumvented by using ancient DNA (aDNA) retrieved from archaeological specimens. Indeed, aDNA provides accurate genetic snapshots of the genetic variation of past populations at specific times and places, enabling e.g., the identification of splits between populations, the identification of past migrations and the timing of these events.

The premise of aDNA research came long before the establishment of the field, as attested in 1963 when Linus Carl Pauling and Emile Zuckerkandl coined the term “paleogenetics” to designate the study of the past through the examination of preserved genetic material from ancient remains⁸. The field was formally initiated in 1984 with the publication of the molecular cloning of an authentic short mitochondrial DNA (mtDNA) sequence from the quagga, a zebra-like equid that has been extinct since the 1880s⁹.

During the early days of paleogenetics, laboratory methods rapidly integrated PCR-based approaches in which specific loci were individually targeted¹⁰. The main PCR targets were uniparentally inherited genetic markers (mtDNA and Y chromosome DNA). Particularly, mtDNA had a higher relevance due to its molecular characteristics. The human mtDNA is a circular double-stranded DNA molecule¹¹ that is: (i) small (16,569 bp)^{12,13}; (ii) maternally inherited¹⁴; (iii) does not recombine¹⁵; (iv) has a higher mutation rate compared to the nuclear genome¹⁶; and (v) high copy number of molecules in cells which increases the odds of successfully obtaining ancient mtDNA from archaeological remains^{10,17}. These characteristics made and still make mtDNA a suitable tool for maternal phylogenetic and phylogeographic inferences¹⁸.

Nevertheless, it took a few decades until technological developments allowed to circumvent limitations associated with the sensitivity of PCR to contamination derived from environmental microorganisms, human handling, and/or the molecular biology reagents, consumables, and laboratories¹⁹. Another limitation was the degraded nature of aDNA, which undergoes post-mortem modifications of the original DNA sequence over time. The most common types of aDNA damage are: (i) fragmentation of DNA, with fragments typically shorter than 100 base pairs (bp)²⁰; (ii) DNA strand breaks at purine [adenine (A) and guanine (G)] sites^{21,22}; and (iii) random conversion of cytosines (C) to uracils (U), which are amplified and sequenced as thymine (T) and result in random C-to-T (and complementary G-to-A) substitutions^{22,23}.

4. From a few DNA fragments to complete genomes

Paleogenetics was propelled forward during the middle 2000s by the application of novel laboratory techniques to successfully retrieve endogenous DNA from archaeological remains, as well as the emergence of high-throughput sequencing (HTS) platforms^{17,24} that can generate data from billions of short DNA fragments simultaneously²⁵. These improvements allowed the

study of not only uniparentally inherited genetic markers but also complete genomes of past humans up to population scales⁷, and even draft genomes of extinct humans like Neanderthals²⁶ and Denisovans²⁷. In fact, molecular anthropology broadened its applications not only to demographic and adaptive inferences but also to evolutionary medicine investigating health and disease²⁸.

While aDNA is a powerful tool to understand our past, it is crucial that its implementation includes complementary paleontological and archaeological data to harness the power of interdisciplinary research², in a discipline coined archaeogenetics by the British archaeologist Colin Renfrew²⁹. Whilst it was initially received with promise, archaeogenetics generated, in some cases, a conflict between archaeologists and geneticists about data interpretation due to cross-disciplinary misunderstandings³⁰. Nevertheless, since the middle 2000s, aDNA has supported or rejected many hypotheses previously laid out by archaeologists, anthropologists, and linguists^{31–34}, and is an essential part of the development and testing of new hypotheses, including some about the peopling of the world^{7,31,35}.

5. Brief peopling history of Europe

As previously suggested by Cann and colleagues³, current studies of present-day global genetic variation support an expansion of modern humans out of Africa ~65–55 thousand years ago (65–55 ka)^{36–39}. The earliest split among non-Africans happened ~60–45 ka between one lineage colonising mainland Eurasia and the other colonising New Guinea and Australia⁴⁰. The genomes of early Eurasian individuals such as Ust’Ishim (45 ka)⁴¹, Kostenki (37 ka)⁴², and Tianyuan (40 ka)⁴³ further demonstrate a split between western Eurasia and East Asia/Oceania around 55–45 ka.

Anatomically modern humans were widely distributed in Europe as early as 45 ka, but this early Palaeolithic European ancestry is virtually absent in contemporary populations⁴⁴. Three subsequent migration waves contributed to the modern-day European genetic makeup: i) Palaeolithic populations were replaced by Mesolithic hunter-gatherers coming from southern European and central Eurasian refugia ~14 ka after the Last Glacial Maximum (LGM; 28–18 ka)^{45,46}; ii) the Neolithic farmers of Anatolia, the Levant, and northern Iran expanded throughout Europe ~8.5 ka and admixed with Mesolithic hunter-gatherers^{47–51}. The Neolithic lifestyle (animal husbandry, agriculture and sedentarism) helped increase population size, but health became poorer⁵²; and iii) Bronze Age herders migrated from the Eurasian steppe to central Europe ~5 ka. These Yamnaya and Afanasievo people already had a mixed ancestry related to various Russian and Caucasus hunter-gatherers^{53,54}. The Steppe migration revolutionised European technology with the notable introduction of the wheel and spread all ancient Indo-European^{53,55}. Interestingly, the location connecting the Proto-Indo-European-speaking Yamnaya with the speakers of Anatolian languages was in the highlands of West Asia, the ancestral region that they both shared⁵⁵.

Iberia—a peninsula that comprises Spain, Portugal, Andorra, Gibraltar, and a small area of Southern France—is the westernmost edge of Eurasia and probably one of the most distant regions during the peopling of the continent as well as having a crucial location for interactions with North Africa. Iberia provides an excellent context to build a comprehensive European and Mediterranean genetic history^{56–61}. Recent studies analysed the impact of the previously described European cultural shifts in Iberia and supported the existence of a clear Neolithic transition and a nuanced Bronze Age transition in Iberia^{56,57,59–62}.

6. Brief peopling history of the Americas

The location that has been suggested to have had a decisive role in the initial peopling of the Americas is Beringia, a land bridge that connected present-day north-eastern Siberia and Alaska and was exposed by lowered sea levels around 34–11 ka^{63,64}. During the Last Glacial Maximum (LGM), an isolated population in south-eastern Beringia with an Ancestral North Eurasian-related ancestry admixed with East Asian-related populations that were expanding northwards. This admixture resulted in two different lineages that diverged ~20 ka: the Ancient Paleo-Siberians and the Ancestral Native Americans^{65–68}. Ancestral Native Americans were isolated for 2,400–9,000 years before entering North America ~25–15 ka when ice sheets covering the continent started to melt, suffering a bottleneck^{67,69–71}. This population gave rise to most of the non-Arctic Native American ancestry and split into two lineages either in Beringia or in North America ~17–15 ka^{66,68,72–74}: Northern Native Americans and Southern Native Americans (represented by the Anzick individual—~11 ka—associated with the Clovis culture⁷²). The peopling of America was swift⁷⁵ and humans reached southern South America as early as ~14.3 ka according to the archaeological record (Monte Verde, Chile⁷⁶).

Interestingly, recent studies provided new evidence of human presence in Chiquihuite Cave (Zacatecas, Mexico)⁷⁷ and White Sands National Park (New Mexico, United States)⁷⁸ during the Last Glacial Maximum (26.5–19 ka) suggesting that humans were present in the Americas earlier than previously thought⁷⁹.

Within the Americas, the genetic history of the Arctic, northern North America, and South America have been recently studied using aDNA^{65,71,74,75}. However, aDNA research in populations from southern North America and Central America is extremely limited, despite a rich archaeological record and Indigenous cultural diversity^{80,81}. Concretely, Mesoamerica—a historically and culturally defined geographic region comprising current central and south

Mexico, Belize, Guatemala, El Salvador, and border regions of Honduras, western Nicaragua, and north-western Costa Rica—was composed by several cultures connected by commerce and farming. In fact, its peoples probably carried an invaluable genetic diversity that has never been explored and was partly lost during the Spanish conquest and the subsequent colonial period⁸².

Furthermore, another region of interest in the Americas is the Southern Cone of South America, a subcontinental area composed by Argentina, Chile, Uruguay, and Southern Brazil. The Southern Cone of South America is a key study region to build a comprehensive picture of the peopling of the Americas as it is geographically the most distant location from Beringia. A growing body of genetic research addressing the population history of the Southern Cone of South America has focused mostly on the southernmost archaeological and extant groups from Patagonia and Tierra del Fuego^{83–90}. However, only a few recent studies included populations from the understudied eastern sector, the Argentinian Pampas^{67,75,91,92}.

7. Thesis structure

My thesis consists in three parts that are divided according to two principles: (i) the studied geographic area (Iberia, Mesoamerica, and Southern Cone of South America); and (ii) the use of aDNA as a tool to tackle questions from a narrow to a broad view (a medical condition diagnosed from a single individual, the genetic changes of one archaeological site through time, the genetic history of a historical and cultural region, and the peopling of a subcontinental area).

Part I focuses on the Iberian Peninsula. In Chapter I, I describe a 1,000-year-old case of Klinefelter's Syndrome, from an individual unearthed at the mediaeval archaeological site of Torre Velha (Portugal). This work provides an unprecedented integration of morphological, osteological, and genetic evidence to confirm the clinical diagnosis. In Chapter II, I investigate the temporal genetic changes at the Mas d'en Boixos archaeological site—the greatest

prehistorical site of the central pre-littoral depression in Catalonia (Spain)—analysing two individuals from the Middle Neolithic, nine from the Early Bronze Age and three previously published individuals from the Iron Age⁵⁹.

Part II focuses on the study of pre-Columbian Mesoamerica. In Chapter III, I review previously published ancient DNA studies in pre-Columbian Mesoamerica and evaluate how aDNA research helped discerning population dynamics patterns, in the light of archaeological, linguistic, and anthropological information, offering new working hypotheses. In Chapter IV, I investigate the past genetic history of pre-Columbian Mesoamerica by analysing 25 newly sequenced ancient genomes from five different archaeological sites across Mesoamerica spanning the time periods between 320 and 1,400 CE.

In Part III: Chapter V, I investigate the early peopling of the Southern Cone of South America. I analyse eighteen ancient human mitochondrial genomes—seven of which are novel—from three Early to Late Holocene (10–1.5 ka) archaeological sites from the understudied Argentinian Pampas and assess its influence in the southernmost regions: Patagonia and Tierra del Fuego.

8. References

1. Morgan GJ. Emile Zuckerkandl, Linus Pauling, and the Molecular Evolutionary Clock, 1959-1965. *J Hist Biol* 1998;31(2):155–78.
2. Destro-Bisol G, Jobling MA, Rocha J, et al. Molecular Anthropology in the Genomic Era. *Journal of Anthropological Sciences* 2010;88:93–112.
3. Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. *Nature* 1987;325(1):31–6.
4. Cavalli-Sforza LL. The Human Genome Diversity Project: past, present and future. *Nat Rev Genet* 2005;6:333–40.
5. Ilklic I, Paul NW. Ethical aspects of genome diversity research: Genome research into cultural diversity or cultural diversity in genome research? *Med Health Care Philos* 2009;12(1):25–34.
6. Bergstrom A, McCarthy SA, Hui R, et al. Insights into human genetic variation and population history from 929 diverse genomes. 2019;1–23.
7. Llamas B, Roca-Rada X, Collen E. Ancient DNA helps trace the peopling of the world. *Biochem (Lond)* 2020;42(1):18–22.
8. Pauling L, Zuckerkandl E. Molecular “Restoration Studies” of Extinct Forms of Life. *Acta Chem Scand* 1963;17:9–16.
9. Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. DNA sequences from the quagga, an extinct member of the horse family. *Nature* 1984;312(5991):282–4.
10. Krings M, Stone AC, Schmitz RW, Krainitzki H, Stoneking M, Paabo S. Neandertal DNA sequences and the origin of modern humans. *Cell Press* 1997;90(1):19–30.
11. Anderson S, Bankier AT, Barrell BG, et al. Sequence and organization of the human mitochondrial genome. *Nature* 1981;290(5806):457–65.
12. 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. *Nat Genet* 1999;23(2):147.
13. Behar DM, van Oven M, Rosset S, et al. A “Copernican” Reassessment of the Human Mitochondrial DNA Tree from its Root. *Am J Hum Genet* 2012;90(4):675–84.
14. Giles RE, Blanc H, Cann HM, Wallace DC. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* 1980;77(11):6715–9.
15. Brown WM, George M, Wilson AC. Rapid evolution of animal mitochondrial DNA. 1979.
16. Sigurðardóttir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P. The Mutation Rate in the Human mtDNA Control Region. *Am J Hum Genet* 2000;66:1599–609.
17. Llamas B, Willerslev E, Orlando L, Orlando L. Human evolution: a tale from ancient genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2017;372(1713).
18. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 2009;30(2):386–94.
19. Rizzi E, Lari M, Gigli E, de Bellis G, Caramelli D. Ancient DNA studies: New perspectives on old samples. *Genetics Selection Evolution*. 2012;44(1).

20. Allentoft ME, Collins M, Harker D, et al. The half-life of DNA in bone: Measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society B: Biological Sciences* 2012;279(1748):4724–33.
21. Briggs AW, Stenzel U, Johnson PLF, et al. Patterns of damage in genomic DNA sequences from a Neandertal . 2007.
22. Bokelmann L, Glocke I, Meyer M. Reconstructing double-stranded DNA fragments on a single-molecule level reveals patterns of degradation in ancient samples. *Genome Res* 2020;1449–57.
23. Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. 2001.
24. Orlando L, Gilbert MTP, Willerslev E. Reconstructing ancient genomes and epigenomes. *Nat Rev Genet* 2015;16(7):395–408.
25. Metzker ML. Sequencing technologies the next generation. *Nat Rev Genet*. 2010;11(1):31–46.
26. Green RE, Krause J, Briggs AW, et al. A draft sequence of the neandertal genome. *Science (1979)* 2010;328(5979):710–22.
27. Meyer M, Kircher M, Gansauge MT, et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science (1979)* 2012;338(6104):222–6.
28. Benton ML, Abraham A, LaBella AL, Abbot P, Rokas A, Capra JA. The influence of evolutionary history on human health and disease. *Nat Rev Genet*. 2021;22(5):269–83.
29. Renfrew C. From molecular genetics to archaeogenetics. 1997.
30. Booth TJ. A stranger in a strange land: a perspective on archaeological responses to the palaeogenetic revolution from an archaeologist working amongst palaeogeneticists. *World Archaeol* 2019;51(4):586–601.
31. Skoglund P, Mathieson I. Ancient genomics: a new view into human prehistory and evolution. *Annu Rev Genomics Hum Genet* 2018;19:381–404.
32. Marciniak S, Perry GH. Harnessing ancient genomes to study the history of human adaptation. *Nat Rev Genet* 2017;18(11):659–74.
33. Pääbo S. The human condition - A molecular approach. *Cell* 2014;157(1):216–26.
34. Pääbo S. The diverse origins of the human gene pool. *Nat Rev Genet* 2015;16(6):313–4.
35. Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. Tracing the peopling of the world through genomics. *Nature* 2017;541(7637):302–10.
36. Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences* 2005;102(44):15942–7.
37. Jakobsson M, Scholz SW, Scheet P, et al. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 2008;451(7181):998–1003.
38. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature* 2011;475(7357):493–6.

39. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet* 2014;46(8):919–25.
40. Malaspinas AS, Westaway MC, Muller C, et al. A genomic history of Aboriginal Australia. *Nature*. 2016;538(7624):207–14.
41. Fu Q, Li H, Moorjani P, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 2014;514(7253):445–9.
42. Seguin-Orlando A, Korneliussen TS, Sikora M, et al. Genomic structure in Europeans dating back at least 36,200 years. *Science (1979)* 2014;346(6213):1113–8.
43. Fu Q, Meyer M, Gao X, et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci U S A* 2013;110(6):2223–7.
44. Günther T, Jakobsson M. Genes mirror migrations and cultures in prehistoric Europe—a population genomic perspective. *Curr Opin Genet Dev* 2016;41:115–23.
45. Fu Q, Posth C, Hajdinjak M, et al. The genetic history of Ice Age Europe. *Nature* 2016;534(7606):200–5.
46. Posth C, Renaud G, Mittnik A, et al. Pleistocene mitochondrial genomes suggest a single major dispersal of non-africans and a late glacial population turnover in Europe. *Current Biology* 2016;26(6):827–33.
47. Broushaki F, Thomas MG, Link V, et al. Early Neolithic genomes from the eastern Fertile Crescent. *Science (1979)* 2016;353(6298):499–503.
48. Skoglund P, Malmström H, Raghavan M, et al. Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science (1979)* 2012;336(6080):466–9.
49. Lazaridis I, Nadel D, Rollefson G, et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* 2016;536(7617):419–24.
50. Mathieson I, Lazaridis I, Rohland N, et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 2015;528(7583):499–503.
51. Omrak A, Günther T, Valdiosera C, et al. Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. *Current Biology* 2016;26(2):270–5.
52. Childe VG. *The dawn of European civilization*. London: Kegan Paul, Trench, Trübner & Co.; New York: Alfred A. Knopf; 1925.
53. Haak W, Lazaridis I, Patterson N, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 2015;522(7555):207–11.
54. Allentoft ME, Sikora M, Sjögren KG, et al. Population genomics of Bronze Age Eurasia. *Nature* 2015;522(7555):167–72.
55. Lazaridis I, Alpaslan-Roodenberg S, Acar A, et al. The genetic history of the Southern Arc: A bridge between West Asia and Europe. *Science (1979)* 2022;377(6609).
56. Villalba-Mouco V, Oliart C, Rihuete-Herrada C, et al. Genomic transformation and social organization during the Copper Age–Bronze Age transition in southern Iberia. *Sci Adv* 2021;7:21.
57. Martiniano R, Cassidy LM, Ó'Maoldúin R, et al. The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genet* 2017;13(7).

58. Villalba-Mouco V, van de Loosdrecht MS, Posth C, et al. Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula. *Current Biology* 2019;29(7):1169-1177.e7.
59. Olalde I, Mallick S, Patterson N, et al. The genomic history of the Iberian Peninsula over the past 8000 years. *Science* (1979) 2019;1234(March):1230–4.
60. Valdiosera C, Günther T, Vera-Rodríguez JC, et al. 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* 2018;115(13):3428–33.
61. Szécsényi-Nagy A, Roth C, Brandt G, et al. The maternal genetic make-up of the Iberian Peninsula between the Neolithic and the Early Bronze Age. *Nature* 2017;1–15.
62. Olalde I, Haak W, Barnes I, Lalueza-Fox C, Reich D. The Beaker Phenomenon and the Genomic Transformation of Northwest Europe. *Nature* 2018;555(7695):190–6.
63. Hoffecker JF, Elias SA, O'Rourke DH, Scott GR, Bigelow NH. Beringia and the global dispersal of modern humans. *Evol Anthropol* 2016;25(2):64–78.
64. Hu A, Meehl GA, Otto-Bliesner BL, et al. Influence of Bering Strait flow and North Atlantic circulation on glacial sea-level changes. *Nat Geosci* 2010;3(2):118–21.
65. Sikora M, Pitulko V v., Sousa VC, et al. The population history of northeastern Siberia since the Pleistocene. *Nature* 2019;570:182–8.
66. Raghavan M, Steinrücken M, Harris K, et al. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* (1979) 2015;349(6250).
67. Llamas B, Fehren-Schmitz L, Valverde G, et al. Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Sci Adv* 2016;2(4):e1501385.
68. Moreno-Mayar JV, Potter BA, Vinner L, et al. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* 2018;553(7687):203–7.
69. Skoglund P, Reich D. A genomic view of the peopling of the Americas. *Curr Opin Genet Dev* 2016;41:27–35.
70. Tamm E, Kivisild T, Reidla M, et al. Beringian standstill and spread of native American founders. *PLoS One* 2007;2(9):1–6.
71. Moreno-Mayar JV, Vinner L, de Barros Damgaard P, et al. Early human dispersals within the Americas. *Science* (1979) 2018;362(6419).
72. Rasmussen M, Anzick SL, Waters MR, et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 2014;506(7487):225–9.
73. Reich D, Patterson N, Campbell D, et al. Reconstructing Native American population history. *Nature* 2012;488(7411):370–4.
74. Scheib CL, Li H, Desai T, et al. Ancient human parallel lineages within North America contributed to a coastal expansion. *Science* (1979) 2018;360(6392):1024–7.
75. Posth C, Nakatsuka N, Lazaridis I, et al. Reconstructing the Deep Population History of Central and South America. *Cell* 2018;175(5):1185–97.
76. Dillehay TD, Ramírez C, Pino M, Collins MB, Rossen J, Pino-Navarro JD. Monte Verde: Seaweed, Food, Medicine, and the Peopling of South America. *Science* (1979) 2008;320(5877):784–6.

77. Ardelean CF, Becerra-valdivia L, Pedersen MW, et al. Evidence of human occupation in Mexico around the Last Glacial Maximum. *Nature* 2020;
78. Bennett MR, Bustos D, Pigati JS, et al. Evidence of humans in North America during the Last Glacial Maximum. *Science* (1979) 2021;373(6562):1528–31.
79. Becerra-Valdivia L, Higham T. The timing and effect of the earliest human arrivals in North America. *Nature* 2020;584.
80. McCafferty G, Amador E, Salgado-González S, Dennet C. Archaeology on Mesoamerica's Southern Frontier. In DL Nichols & Pool C A (Eds), *The Oxford Handbook of Mesoamerican Archaeology* (pp 83–105) New York, NY: Oxford University Press 2012;
81. Kirchhoff P. Mesoamérica: sus límites geográficos, composición étnica y caracteres culturales. *Acta Americana* 1943;
82. Alba-Hernandez F. La población de México. Centro de estudios Economicos y Demograficos, El Colegio de México; 1976.
83. Crespo CM, Favier Dubois C, Russo MG, Lanata JL, Dejean CB. First analysis of ancient mtDNA genetic diversity in Northern coast of Argentinean Patagonia. *J Archaeol Sci Rep* 2017;12:91–8.
84. Crespo CM, Lanata JL, Cardozo DG, Avena SA, Dejean CB. Ancient maternal lineages in hunter-gatherer groups of Argentinean Patagonia. Settlement, population continuity and divergence. *J Archaeol Sci Rep* 2017;18(November 2017):689–95.
85. de la Fuente C, Galimany J, Kemp BM, Judd K, Reyes O, Moraga M. Ancient marine hunter-gatherers from Patagonia and Tierra Del Fuego: Diversity and differentiation using uniparentally inherited genetic markers. *Am J Phys Anthropol* 2015;158(4):719–29.
86. de la Fuente C, Ávila-Arcos MC, Galimany J, et al. Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. *Proceedings of the National Academy of Sciences* 2018;115(17):E4006–12.
87. Lalueza C, Pérez-Pérez A, Prats E, Cornudella L, Turbón D. Lack of founding Amerindian mitochondrial DNA lineages in extinct Aborigines from Tierra del Fuego–Patagonia. *Hum Mol Genet* 1997;6(1):41–6.
88. de Saint Pierre M, Gandini F, Perego UA, et al. Arrival of Paleo-Indians to the Southern Cone of South America: New Clues from Mitogenomes. *PLoS One* 2012;7(12):1–9.
89. Nakatsuka N, Luisi P, Motti JMB, et al. Ancient genomes in South Patagonia reveal population movements associated with technological shifts and geography. *Nat Commun* 2020;11(1).
90. García-Bour J, Pérez-Pérez A, Álvarez S, et al. Early Population Differentiation in Extinct Aborigines from Tierra del Fuego-Patagonia: Ancient mtDNA Sequences and Y-Chromosome STR Characterization. *Am J Phys Anthropol* 2004;123(4):361–70.
91. Perez SI, Bernal V, Gonzalez PN, Sardi M, Politis GG. Discrepancy between cranial and DNA data of early Americans: Implications for American peopling. *PLoS One* 2009;4(5).

92. Postillone MB, Martínez G, Flensburg G, Dejean CB. First analysis of mitochondrial lineages from the eastern Pampa–Patagonia transition during the final late Holocene. *Am J Phys Anthropol* 2020;171(4):659–70.


Part I: Iberia

Chapter I: A 1000-year-old case of Klinefelter's syndrome diagnosed by integrating morphology, osteology, and genetics

Statement of Authorship

Title of Paper	A 1000-year-old case of Klinefelter's syndrome diagnosed by integrating morphology, osteology, and genetics
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	Roca-Rada X, Tereso S, Rohrlach AB, et al. A 1000-year-old case of Klinefelter's syndrome diagnosed by integrating morphology, osteology, and genetics. The Lancet. 2022;400(10353):691-2.


Principal Author


Name of Principal Author (Candidate)	Xavier Roca Rada
Contribution to the Paper	Conceptualisation, performed laboratory experiments, bioinformatic processing, data analyses, results interpretation and drafted the manuscript.
Overall percentage (%)	60%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	 Date 05/10/2022


Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- the candidate's stated contribution to the publication is accurate (as detailed above);
- permission is granted for the candidate to include the publication in the thesis; and
- the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Sofia Tereso
Contribution to the Paper	Responsible for the archaeological excavation of Torre Velha necropolis and for the anthropological study (morphological and osteological analyses) with other three co-authors (Cláudia Umbelino, Francisco Curate and André Brito).
Signature	 Date 05/10/2022

Name of Co-Author	Adam B Rohrlach
Contribution to the Paper	I designed, tested and implemented the Bayesian method for detecting chromosomal aneuploidies
Signature	 Date 05/10/2022

Name of Co-Author	ANDRÉ BRITO
Contribution to the Paper	I CONDUCTED THE ARCHAEOLOGICAL EXCAVATIONS, AS WELL AS THE MORPHOLOGICAL AND OSTEOLOGICAL ANALYSES.
Signature	 Date 06/10/2022

Name of Co-Author	Matthew Williams		
Contribution to the Paper	ancient DNA lab work.		
Signature		Date	05.Oct.2022

Name of Co-Author	Cláudia Umbelino		
Contribution to the Paper	conducted morphological and osteological analyses and participated in the writing process		
Signature		Date	06/10/2022


Name of Co-Author	Ira Deveson		
Contribution to the Paper	Data generation		
Signature		Date	06 October 2022

Name of Co-Author	Dr Yassine Souilmi		
Contribution to the Paper	Co supervised the Mr Roca Rada, helped interpret the results, and provided input on the manuscript.		
Signature		Date	5/10/2022

Name of Co-Author	Antonio Amorim		
Contribution to the Paper	Discussion with JCT the design of the project; contribution to manuscript draft revision and writing. Assinado por: António Manuel Amorim dos Santos Num. de Identificação: 02704624 Data: 2022.10.05 12:44:02+01'00'		
Signature		Date	05/10/2022



Name of Co-Author	FRANCISCO CURATE		
Contribution to the Paper	PERFORMED MORPHOLOGICAL AND OSTEOLOGICAL ANALYSES; GAVE INPUT TO THE FINAL DRAFT OF THE TEXT.		
Signature		Date	6-10-2022

Name of Co-Author	Pedro C. Carvalho		
Contribution to the Paper	Assinado por: PEDRO JORGE CARDOSO DE CARVALHO Num. de Identificação: 07636805 Data: 2022.10.05 12:18:21+01'00'	Conducted the archaeological excavations and morphological/osteological analyses	
Signature		Date	05/10/2022

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Wet and dry lab genetic analyses, edited manuscript, provided technical/analytical support and critical feedback		
Signature		Date	06/10/2022

Name of Co-Author	João Teixeira		
Contribution to the Paper	Designed and supervised the research.		
Signature		Date	14/10/2022

A 1000-year-old case of Klinefelter's syndrome diagnosed by integrating morphology, osteology, and genetics



Xavier Roca-Rada, Sofia Tereso, Adam B Rohrlach, André Brito, Matthew P Williams, Cláudia Umbelino, Francisco Curate, Ira W Deveson, Yasmine Souilmi, António Amorim, Pedro C Carvalho, Bastien Llamas, João C Teixeira

The skeleton of a human being buried over 1000 years ago, in the medieval archaeological site of Torre Velha, in north-eastern Portugal, was referred to us for analysis and investigation (appendix).

On examination, we found the outstandingly well preserved skeleton of an adult who was likely to have been more than 25 years old at the time of death. The person was approximately 1.80 m tall. Notably, out of the six skeletons examined, where height could be ascertained, this individual was the tallest (appendix). Based on pelvic morphology and metric analysis of the skeleton, we concluded that the individual was a man. However, the bi-iliac width of our person was 289 mm, considerably larger than the average width previously reported for ancient Portuguese men (mean 261.8 mm; SD=13.4; Student's *t* test: $t=22.6944$; $df=124$; $p<0.001$). Additionally, the individual's teeth were worn asymmetrically, indicating probable malocclusion of the jaw and maxillary prognathism (figure; appendix). Bone densitometry analysis showed normal bone mineral density values (0.792–0.999 g/cm²).

Genetic analysis was undertaken using a variety of different methodologies—including the calculation of X chromosome to Y chromosome autosome ratios (0.902 and 0.298, respectively), X chromosome and Y chromosome dosages (approximately 2 and 1, respectively), and X chromosome heterozygosity (approximately 0.2). We also determined the Y chromosome haplogroup to be R1b-P310 (R1b1a1b1a1), a frequent western-European and pan-European lineage, in accordance with the individual's Iberian genetic ancestry (appendix).

Furthermore, using a novel Bayesian method, allowing us to probabilistically assign individuals to

karyotypes based on the number of sequencing reads—or DNA fragments sequenced—mapping to the X, Y, or autosomal chromosomes, we concluded our individual's karyotype to be 47,XXY, and rejected models of contamination of XXY and XX or XY (figure; appendix). The closeness of the observed position of the studied individual and theoretical position for an XXY karyotype (figure) strongly agrees with our posterior probability of approximately 1 for this individual having a Klinefelter's syndrome karyotype. Considering the morphological findings—specifically the height, the bi-iliac width, the possible jaw malocclusion, and the maxillary prognathism—the genetic findings—indicating a karyotype of 47,XXY—we concluded that the studied individual had Klinefelter's syndrome.

The prevalence of Klinefelter's syndrome in the general population has previously been reported to be 0.1–0.2%, but many patients remain undiagnosed.

Typically, people with Klinefelter's syndrome are tall, have broad hips, sparse body hair, small testes, and gynaecomastia; they may have mandibular prognathism. Obesity, low glucose tolerance, and diabetes are observed, as well as osteoporosis because of androgen deficiency. Notably, the absence of osteoporosis in our case does not preclude a diagnosis of Klinefelter's syndrome; osteoporosis is only present in approximately 10–40% of cases. Genetically, about 80% of cases are 47,XXY.

DNA from osteological remains is often scarce, degraded, fragmented, and unsuitable for analysis using commonly available techniques for clinical diagnosis of chromosomal aneuploidies. To overcome these limitations, we have devised a novel Bayesian method,

Lancet 2022; 400: 691–92
 Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia (X Roca-Rada MSc, M P Williams PhD, Y Souilmi PhD, B Llamas PhD, J C Teixeira PhD); Research Centre for Anthropology and Health, Department of Life Sciences, University of Coimbra, Coimbra, Portugal (S Tereso MSc, A Brito MSc, Prof C Umbelino PhD, F Curate PhD); Institute for Medieval Studies, Faculty of Social and Human Sciences, NOVA University of Lisbon, Lisbon, Portugal (S Tereso); Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany (A B Rohrlach PhD); School of Mathematical Sciences, University of Adelaide, Adelaide, SA, Australia (A B Rohrlach); Interdisciplinary Center for Archaeology and Evolution of Human Behaviour, University of Algarve, Faculdade das Ciências Humanas e Sociais, Universidade do Algarve, Faro, Portugal (Prof C Umbelino); Genomics Pillar, Garvan Institute of Medical Research, Sydney, NSW, Australia (IW Deveson PhD); Centre for Population Genomics, Garvan

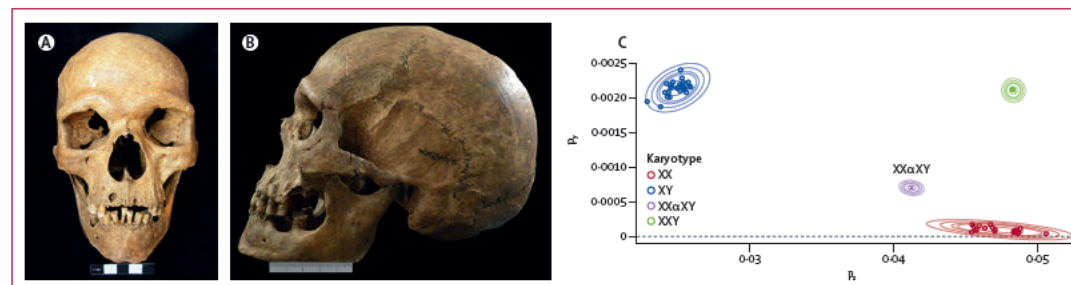


Figure: Diagnosing Klinefelter's syndrome in a 1000-year-old skeleton

Photographs of the skull show anterior (A) and lateral (B) views with a probable malocclusion and maxillary prognathism denoted by atypical dental wear. (C) Graph shows the observed proportion of reads mapping to the X chromosome (x-axis) and Y chromosome (y-axis), coloured by karyotype: XX (red), XY (blue), XXY (green), and the highest likelihood contaminated XXaXY (purple). Real data are indicated by circles; theoretical positions (for XXY and contaminated) are indicated by crosses. Ellipses indicate 50%, 90%, 95%, 99%, and 99.9% confidence.


Clinical Picture

Institute of Medical Research and Murdoch Children's Research Institute, Parkville, Vic, Australia (J W Deveson); School of Clinical Medicine, Faculty of Medicine and Health, University of South Wales, Sydney, NSW, Australia (I W Deveson); National Centre for Indigenous Genomics, Australian National University, Canberra, ACT, Australia (Y Souilmi, B Llamas); Environment Institute, University of Adelaide, Adelaide, SA, Australia (Y Souilmi, B Llamas); Institute of Investigation and Innovation in Health, University of Porto, Porto, Portugal (Prof A Amorim PhD); Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal (Prof A Amorim); Faculty of Sciences, University of Porto, Porto, Portugal (Prof A Amorim); Faculty of Arts and Humanities, University of Coimbra, Coimbra, Portugal (Prof P C Carvalho PhD); Centre of Interdisciplinary Studies, University of Coimbra, Coimbra, Portugal (Prof P C Carvalho); Telethon Kids Institute, Indigenous Genomics Research Group, Adelaide, SA, Australia (B Llamas); Centre of Excellence for Australian Biodiversity and Heritage, University of Adelaide, Adelaide, SA, Australia (B Llamas, J C Teixeira); Evolution of Cultural Diversity Initiative, Australian National University, Canberra, ACT, Australia (J C Teixeira)

Correspondence to:
Dr João C Teixeira, Evolution of Cultural Diversity Initiative, School of Culture, History and Language, Australian National University, Canberra ACT 2601, Australia
joao.teixeira@anu.edu.au
See Online for appendix

which, we believe, is a potentially efficient statistical way of analysing fragmented DNA from a variety of sources (eg, ancient DNA, cell-free DNA, and DNA from forensic cases).

Contributors

XR-R and JCT designed the project. ST, AB, and PCC conducted the archaeological excavations. ST, CU, and FC conducted morphological and osteological analyses. XR-R, MPW, BL, and JCT performed the ancient DNA experiments. IWD generated the DNA sequencing data. XR-R, YS, BL, and JCT analysed the data. ABR designed and implemented the Bayesian method to the genetic data. XR-R and JCT wrote the manuscript with input from all authors. Written consent was not available from the patient.

Declaration of interests

We declare no competing interests.

Data sharing

The genetic data presented in this Clinical Picture will be made available upon publication.

Acknowledgments

We thank the Municipality of Bragança, the University of Coimbra, the University of Adelaide, the Max Planck Society, and the Calouste Gulbenkian Foundation for the support provided. ST is supported by the Fundação para a Ciência e a Tecnologia (SFRH/BD/116363/2016). BL (FT170100448) and JCT (DE210101235) are supported by the Australian Research Council. ABR is supported by the European Research Council (771234-PALEORIDER).

Copyright © 2022 Elsevier Ltd. All rights reserved.

THE LANCET

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed.
We post it as supplied by the authors.

Supplement to: Roca-Rada X, Tereso S, Rohrlach AB, et al. A 1000-year-old case of Klinefelter's syndrome diagnosed by integrating morphology, osteology, and genetics. *Lancet* 2022; **400**: 691–92.

Appendix

A thousand-year-old clinical case of Klinefelter syndrome: integrating morphology, osteology, and genetics

Introduction

Despite the availability of genome-wide data from more than 3,000 ancient humans¹, so far only four putative cases²⁻⁵ of Klinefelter's syndrome (KS)^{6,7} have been reported from ancient DNA analyses. Importantly, none of these studies included a thorough morphological and osteological analysis of the skeletal remains or allowed for a rigorous and in-depth statistical assessment of the ploidy level of the sex chromosomes^{2-5, 2-5}.

Archaeological context of Torre Velha (Castro de Avelãs)

The archaeological site is in north-eastern Portugal. This region was first inhabited by the *Zoelae* in the Iron Age and was later crossed by an important Roman road in north-western Iberian Peninsula – known as *via XVII* –, which connected *Asturica Augusta* to *Bracara Augusta*⁸⁻¹⁰. Historical evidence from the *Parochiale Suevum* shows that the *ciuitas* of the *Zoelae* in Castro de Avelãs was later home to a Suebi diocese known as *pagus Brigantia*, which predated the Mediaeval occupation of the region.

Archaeological excavations at the Torre Velha necropolis (2012-2015) have so far revealed 59 graves directly excavated in the schist rock, which is typical of mediaeval sites in Iberia, with a minimum number of individuals estimated at 57, of which 39 individuals were exhumed from individual graves (33 adults; 6 non-adults) and 18 from ossuaries (17 adults; 1 non-adult)¹¹. Radiocarbon dating on a total of 16 individuals demonstrated these individuals were buried between the 6th and 13th century CE (Cal AD ~500-1300).

The individual presented in this work (referenced as 11/[15] during excavations) was buried in an oval in shaped individual tomb, without a cover or associated grave goods. The skeleton was placed in a supine position with the arms crossed over the chest, with the typical West-East orientation of the Christian burials (Figure S1).



Figure S1: Burial site of individual 11/[15].

Radiocarbon dating

Accelerator Mass Spectrometry (AMS) radiocarbon dating was performed at the Beta Analytic Testing Laboratory (USA), using bone collagen extracted with alkali from the sample's left ulna fragment. The obtained radiocarbon age was calibrated with the Oxcal software (version 4.4.2; © Bronk Ramsey 2021) used with calibration dataset Intcal20¹². The calibrated dating indicates a Mediaeval chronology between the 11th and 12th centuries CE (Cal AD 1020 to 1160 (95.4%); Table S1).

Table S1: Radiocarbon dating of individual 11/[15].

Laboratory reference	Conventional Radiocarbon Age	2 Sigma calibrated result*	$\delta^{13}\text{C}$ (‰)
BETA: 334443	970 \pm 30	Cal AD 1022 to 1159 (95.4%)	-19.0

*Calibration Oxcal 4.4.2 software (© Bronk Ramsey 2021), Intcal20 calibration curve¹².

Morphology

Individual 11/[15] presents a robust skeletal morphology, a long but low cranium, pronounced supraorbital arcades, rectangular orbits with a large interorbital area, a large nasal cavity, and a hyperbolic palate (Figure S2). The sex and age at death were estimated according to standard physical anthropology methods^{13,14}. Individual 11/[15] was assigned a biological male sex based on the morphological traits of the skull and the analysis of the innominate bones. We estimate an age at death between 25 and 40 years.

The stature of 6 individuals from the same archaeological site was estimated based on the regression equations developed by De Mendonça (2000)¹⁵, which consider the physiological length of the femur or the full length of the humerus, and by Cordeiro et al. (2009)¹⁶, which uses the maximum length of the 2nd metatarsal bone (Figure S3 and Table S2). Both methods were developed for, and applied to, the Portuguese population^{17,18}. We preferentially utilised the left bones for the different analysis. The height of individual 11/[15] was highly unusual and estimated at ~1.80 m. Hence, this individual was approximately 10 to 12 cm taller than the five other individuals for which it was possible to estimate stature, despite the relatively small sample size due to poor preservation (Table S2).



Figure S2: Anterior and right lateral views of the skull from individual 11/[15]. The skull presents a robust morphology: a long but low cranium, pronounced supraorbital arcades, rectangular orbits with a large interorbital area, a large nasal cavity, and a hyperbolic palate.

Table S2: Recorded stature of six individuals unearthed at Torre Velha.

Burial	Sex	<i>De Mendonça (2000)</i> ¹⁵		<i>Cordeiro et al. (2009)</i> ¹⁶
		Humerus (cm)	Femur (cm)	2 nd Metatarsal (cm)
11/[15]	Male	169·90±8·44	178·99±6·90	174·85±47·5
12/[02]	Male	NO	169·15±6·90	NO
16/[38]	Male	165·65±8·44	NO	NO
17/[43]	Male	162·71±8·44	167·55±6·90	NO
18/[46]	Female	160·50±7·70	166·59±5·92	167·84±4·75
19/[50]	Male	166·63±8·44	165·15±6·90	161·41±4·75

NO: Non-observable

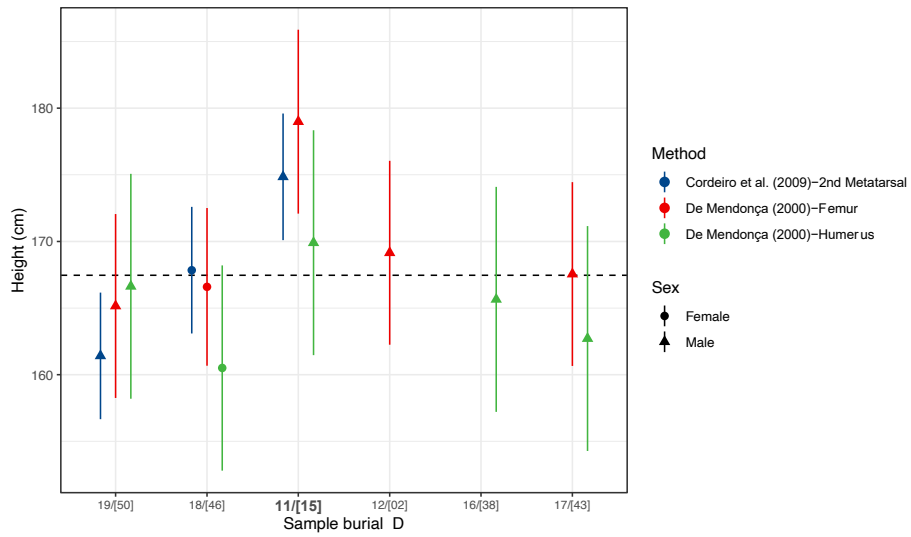


Figure S3: Height estimates of the six individuals unearthed at Torre Velha using three different methods. Dashed line represents the mean across all estimates. Individual 11/[15] is an outlier.

Osteology

The bi-iliac width is an anatomical term referring to the widest measure of the pelvis between the outer edges of the upper iliac bones and the increase of this morphological feature has previously been observed in individuals with KS¹⁹. Accordingly, individual 11/[15] had a bi-iliac width of 289 mm, which is considerably larger than the previously average width reported for Portuguese male individuals from the late 19th to the early 20th centuries²⁰.

In the oral cavity, we found an atypical dental wear for the estimated age of the individual that is probably associated with malocclusion and maxillary prognathism (Figure S4). While both morphological characteristics have recently been associated with KS^{21,22}, other typical signatures of the disease, such as congenital absence of permanent teeth, increased expression of the Carabelli's trait and increased teeth size^{21,22}, were either not observed or not possible to measure due to the significant dental wear and teeth loss present in this individual.

Due to the high prevalence of osteoporosis in Klinefelter Syndrome (KS), the left femur of the studied individual was screened for osteoporosis with a Hologic Horizon© DXA System (Hologic, Inc., Marlborough, MA) at the Department of Nuclear Medicine, University of Coimbra Hospitals, following Curate et al. (2013)²³. The analysis showed normal bone mineral density values at the regions of interest of the proximal femur, following the WHO classification of osteoporosis (Table S3), which is not consistent with a diagnostic of osteoporosis associated with KS. However, we emphasise that only ~10-40% of adult KS cases are accompanied by osteopenia and osteoporosis, whereby our findings are not unexpected²⁴. Incidentally, we have not found any indication of premature fusion and/or excessive calcification of the coronal sutures, as suggested in previous reports²⁵.

We observed other bone pathologies not commonly associated with KS in individual 11/[15], including: a) an old traumatic lesion in the right parietal bone that might have resulted from an accidental episode or interpersonal violence; b) signs of osteochondritis dissecans (Figure S5) on the right side of three diarthrodial joint areas (the iliac acetabulum, the articular surface of the calcaneus and the proximal joint of the first metatarsal), which is usually associated with trauma resulting from intense activity or an insufficient blood supply²⁶⁻²⁸, and more common in young males²⁹; c) an irregular border in the anterior talar articular surface of the right calcaneus (Figure S5), indicating a possible non-osseous unilateral tarsal talocalcaneal coalition, which is an asymptomatic rare congenital condition in clinical and archaeological contexts of autosomal dominant inheritance^{30,31}; and d) periapical and periodontal diseases in the maxilla and mandible (Figure S4), which may have led to several mouth problems.

Table S3: DXA results summary for individual 11/[15].

Region	Area (cm²)	BMC (g)	BMD (g/cm²)	T-score	Z-score
Neck	5.38	4.26	0.792	-1.0	-0.7
Total Hip	54.25	54.18	0.999	-0.2	-0.1



Figure S4: Atypical dental wear and probable malocclusion and maxillary prognathism of individual 11/[15]. Signs of periapical and periodontal diseases are illustrated by the yellow and orange arrows, respectively.



Figure S5: Left and right calcaneus. Right calcaneus: yellow arrow: irregular border of the anterior talar articular surface; blue arrow: focal necrotic lesion in the posterior talar articular surface.

Ancient DNA laboratory work

Ancient DNA analysis was conducted in clean-room facilities at the Australian Centre for Ancient DNA (ACAD), University of Adelaide, following strict precautions to minimise contamination³². DNA was extracted from the petrous bone using a commonly applied extraction method optimised to retrieve degraded ancient DNA fragments³³. A partially UDG-treated double-stranded DNA library was generated³⁴ and shotgun sequenced on an Illumina NovaSeq 6000 platform at the Kinghorn Centre for Clinical Genomics (Sydney, NSW, Australia).

Bioinformatics

Raw sequencing reads (i.e., sequenced DNA fragments) were processed and mapped against the human reference genome (hg19) using the nf-core/eager³⁵ pipeline and the following execution command:

```
nextflow -c /hpcfs/users/a1782219/nextflow/eager_bwa.config run nf-core/eager \
-r 2.3.5 \
-with-singularity \
-with-tower \
-profile singularity \
-name PT_aDNA_1 \
--input /hpcfs/users/a1782219/Sequencing/Portugal_210720/eager_input_table.tsv \
--fasta /hpcfs/users/a1782219/ref/human_g1k_v37_decoy.fasta \
--outdir /hpcfs/users/a1782219/Sequencing/Portugal_210720/eager_results \
-w /hpcfs/users/a1782219/Sequencing/Portugal_210720/work \
--complexity_filter_poly_g \
--mergedonly \
--bwaaln 0.01 \
--bwaalno 2 \
--run_bam_filtering \
--bam_mapping_quality_threshold 25 \
--bam_unmapped_type fastq \
--dedupper dedup \
--run_trim_bam \
--bamutils_clip_half_udg_left 2 \
--bamutils_clip_half_udg_right 2 \
--run_genotyping \
--genotyping_tool pileupcaller \
--genotyping_source trimmed \
--pileupcaller_bedfile /hpcfs/users/a1782219/ref/snps_sites_files/dbsnp_138.b37.pos \
--pileupcaller_snpfile /hpcfs/users/a1782219/ref/snps_sites_files/dbsnp_138.b37.snp \
--pileupcaller_method randomHaploid \
--run_mtnucratio \
--run_sexdeterrmine \
--sexdeterrmine_bedfile /hpcfs/users/a1782219/ref/snps_sites_files/dbsnp_138.b37.pos \
--run_nuclear_contamination
```

This allowed us to generate a total of 9,874,070 reads, of which 1,739,542 unique reads were retained after filtering and deduplication (17.62% of endogenous DNA) (Table S4). Of these, 162,970 also showed post-mortem damage scores above 0 from PMDtools³⁶. Ancient DNA authenticity was evaluated measuring the rate of cytosine deamination—a hallmark of ancient DNA—in the first nucleotides of all sequencing reads and the length of the DNA fragments³⁷ (Figures S6 and S7).

Table S4: Total number of reads in each step of the processing.

Sequenced reads	9,874,070
Reads filtered by read length and sequencing quality	9,248,613
Mapped reads	2,568,801
Reads filtered by mapping quality	1,898,607
Deduplicated reads	1,739,542
Reads showing post-mortem damage scores above 0 from PMDtools³⁶	162,970

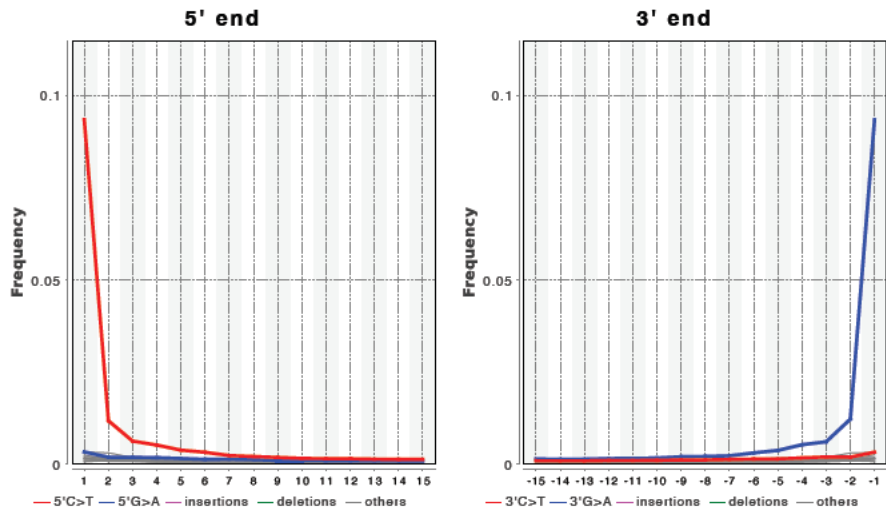


Figure S6: Position-specific substitutions from the 5' (left) to the 3' (right) ends of the reads. The red and blue curves show an accumulation of 5' C-to-T and 3' G-to-A substitutions, respectively, which is characteristic of ancient DNA damage.

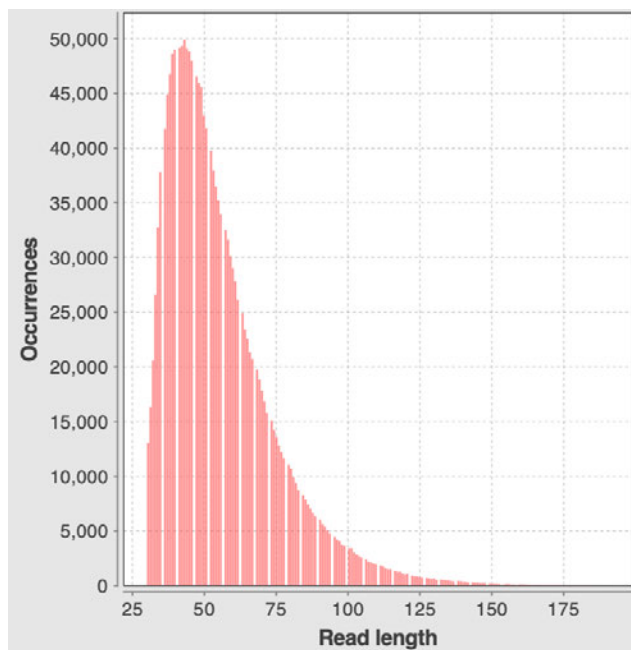


Figure S7: Read length histogram, showing a high number of short reads, which is characteristic of ancient DNA. The distribution was identical for the + and – strands.

Genetic analyses

We were able to retrieve a total of 1,739,542 unique sequencing reads, of which 83,553 and 3,677 mapped to the X and Y chromosomes, respectively. The presence of KS was firstly determined by calculating the number of reads mapping to each of the sex chromosomes relative to the autosomes³⁸. An X-ratio of 1.0 is considered indicative of the presence of two X chromosomes and 0.5 is considered indicative of a single copy of the X chromosome. A Y-ratio of 0.5 indicates the presence of a single copy of the Y chromosome. The X-ratio observed in individual 11/[15] was 0.902, while the Y-ratio was 0.298, indicating a high likelihood for the presence of two X chromosomes and one Y chromosome in the sample. This finding was further confirmed using the method described by Gower et al. (2019)³⁹ (Table S5). This method uses counts of reads mapping to the X chromosome and autosomes to determine the genetic sex of an individual. Under the assumption that the number of reads obtained should reflect the chromosome copy numbers and chromosome lengths, two binomial models are constructed, one for males and one for females. A likelihood ratio test is then used to distinguish between the two models. A read dosage on the X chromosome (M_x) close to 0.5 or 1 means that the individual is assigned as male or female, respectively. Therefore, individual 11/[15] had a M_x of ~ 0.96 . A previous study has reliably assigned cases with KS using sex chromosome dosages in modern DNA data⁴⁰. Finally, the X-chromosome heterozygosity of our sample was observed to be in the range expected for female individuals⁴¹ (Tables S6 and S7).

Table S5: Read dosage estimated using Gower et al. (2019)³⁹ confirm the presence of two copies of the X chromosomes.

Read dosage on the X chromosome (M_x)	Reads mapped to the X chromosome (N_x)	Reads mapped to the autosomes (N_a)
0.95624308	83,543	1,644,852

Table S6: ANGSD contamination estimates (Method 1).

Number of SNPs	Contamination Estimate (Method 1 MOM)	Estimate Error (Method 1 MOM)	Contamination Estimate (Method 1 ML)	Estimate Error (Method 1 ML)
42	0.2	0.1	0.2	0

Table S7: ANGSD contamination estimates (Method 2).

Number of SNPs	Contamination Estimate (Method 2 MOM)	Estimate Error (Method 2 MOM)	Contamination Estimate (Method 2 ML)	Estimate Error (Method 2 ML)
42	0.4	0.2	0.4	0

Bayesian Karyotype Estimation

To probabilistically assign individuals to karyotypes, we developed a novel method based on the number of reads mapping to the X, Y, and autosomal chromosomes, denoted n_x , n_y , and n_a respectively, out of a total N post-processed reads. We denote the vector of observed read counts $n = (n_x, n_y, n_a)$.

Consider the karyotypes XX, XY and XXY. Given a karyotype, $k \in \{XX, XY, XXY\}$, there exists a unique vector of probabilities of reads mapping to either the X, Y, or autosomes due to the difference in prevalence of the X and Y copy numbers for each karyotype. We denote these probability vectors $p^k = (p_x^k, p_y^k, p_a^k)$.

We can confidently estimate p^{XX} and p^{XY} from individuals with confident karyotype assignments from the same sequencing platform (i.e., shotgun sequencing or capture assays). We then define $p^{XXY} = (2p_x^{XXY}, p_y^{XXY}, p_a^{XXY}) = (2p_x^{XY}/c, p_y^{XY}/c, p_a^{XY}/c)$, where $c = 2p_x^{XY} + p_y^{XY} + p_a^{XY}$, such that the relative probability of a read mapping to the X chromosome is twice that for an XY individual, but also so that the probability vector still sums to one.

Finally, we also define a sample to be a (contaminated) mixture of individuals carrying karyotypes k_1 and k_2 , at a proportion α , as karyotype $k_1\alpha k_2$, which can be shown to have associated probability vector

$$p^{k_1\alpha k_2} = \alpha(p_x^{k_1}, p_y^{k_1}, p_a^{k_1}) + (1 - \alpha)(p_x^{k_2}, p_y^{k_2}, p_a^{k_2}).$$

If we assume that reads map independently to the X, Y, and autosomal chromosomes, then for the event that an individual carries karyotype k , denoted $K = k$, we have that $n \sim MN(N, p^k)$, i.e., the vector of read counts follows a multinomial distribution. Hence,

$$P(n | K = k) = \frac{N!}{n_x! n_y! n_a!} (p_x^k)^{n_x} (p_y^k)^{n_y} (p_a^k)^{n_a}.$$

Assuming that the relative incidence rates of the karyotypes are known from independent sources, and that the probability of contamination is known *a priori* (although this parameter can be explored for effect), we can use Bayes' theorem to show that for an uncontaminated sample,

$$P(K = k | n) = P(C) \frac{P(n | K=k)P(K=k)}{P(n)},$$

or

$$P(K = k_1\alpha k_2 | n) = [1 - P(C)] \frac{P(n | K=k_1\alpha k_2)P(K=k_1)P(K=k_2)}{P(n)}$$

if the sample is uncontaminated.

Note that to find the optimal value of α , we numerically optimise the function

$$\begin{aligned} P(n | K = k_1\alpha k_2) &= \frac{N!}{n_x! n_y! n_a!} (\alpha p_x^{k_1} + [1 - \alpha]p_x^{k_2})^{n_x} (\alpha p_y^{k_1} + [1 - \alpha]p_y^{k_2})^{n_y} (\alpha p_a^{k_1} + [1 - \alpha]p_a^{k_2})^{n_a} \end{aligned}$$

with respect to α .

Hence, we estimated the probability of reads mapping to the X, Y, and autosomal chromosomes for XX and XY individuals using read proportions calculated from 9 and 21 individuals, respectively. Using leave-one-out cross validation, our method either correctly identified the input karyotypes (23/30) or identified the correct haplotype with additionally 1.22-3.81% contamination (7/30) from another karyotype.

Assuming a probability that approximately 10% of our samples are contaminated, we found that the individual 11/[15] carried karyotype XXY with a probability approximately equal to 1 (to machine precision). To make sure that the prior probability of contamination was not driving the XXY karyotype assignment, we allowed this value to vary from between 0% to 100%. Only in the case where the probability of contamination is exactly 100% is the probability of Klinefelter no longer effectively 1, thus, showing overwhelming support for assigning karyotype XXY to individual 11/[15]. Finally, we were also able to show that the probability that individual 11/[15] was a contaminated sample containing both XX and XY sequence data was effectively 0.

Since ancient examples of Klinefelter individuals have been identified as carrying an XXY karyotype using related concepts of reads mapping to the X and Y chromosome, and no such whole-genome sequencing modern samples exist, it is difficult to apply the method to an independently identified XXY individual (although this approach, instead using Approximate Bayesian Computation, was previously published⁵). Instead, we used simulations of the observed data from this study to verify the reliability of our XXY assignment.

For each simulation, we randomly sampled the proportions of reads mapped to the X, Y, and autosomal chromosomes for an XX and an XY individual, denoted \mathbf{p}^{XX} and \mathbf{p}^{XY} respectively, and a total number of mapped reads, N , from one of the observed individuals. We also randomly sampled a level of contamination $\alpha \in (0,1)$ from a uniform distribution, and hence defined an observed "contaminated" read mapping probability vector, defined

$$\mathbf{p}^\alpha = \alpha \mathbf{p}^{XX} + (1 - \alpha) \mathbf{p}^{XY}.$$

We then simulated reads mapping to each of the chromosomes by sampling a vector $\mathbf{n} \sim MN(N, \mathbf{p}^\alpha)$. For each simulated vector we applied our method of assigning karyotypes as described above, and recorded the estimated level of contamination, denoted $\hat{\alpha}$, as well as the posterior probability of an XXY karyotype or contamination, denoted $P(XXY|\mathbf{n})$ and $P(C|\mathbf{n})$, respectively.

We found that our method correctly estimated the contamination level in $X\%$ of simulations with a high degree of accuracy (mean square error = 0.00117, $r^2 = 99.47$, $p \leq 2.2 \times 10^{-16}$). More importantly though, we found that our method correctly assigned contamination between an XX and an XY individual in 98.9% of simulations, and critically, incorrectly assigned an XXY karyotype in 0% of simulations (Figure S8).

Hence, we found through a thorough simulation study that, given the observed parameters of our observed sequencing data, the Klinefelter karyotype assignment has a vanishingly low probability of having come from a contaminated source.

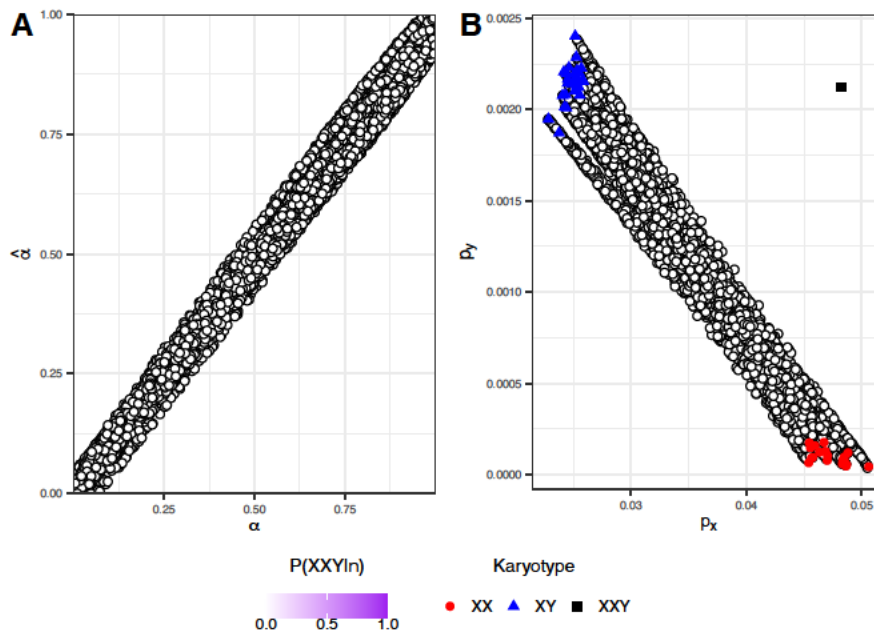


Figure S8: (A) a scatterplot of the true simulated contamination level (α) vs the estimated contamination level ($\hat{\alpha}$), and (B) X and Y chromosome mapping probabilities for the observed data (solid points coloured by karyotype) and for the simulated data (filled circles). In both panels, the filled circles represent simulated data, coloured by the probability of incorrectly assigning the XXY karyotype.

Ancestry

- Y-chromosome haplogroup

The Y-chromosome haplogroup was assigned according to Rohrlach et al. (2021)⁴². We were able to confidently assign this sample a basal Y-chromosome haplogroup R1b-P310 (R1b1a1b1a1). However, due to the low coverage obtained from the shotgun sequencing, we were unable to determine a more derived haplogroup branch within this lineage.

- Mitochondrial haplogroup

The library was reamplified and subsequently captured using in-solution target hybridisation to enrich sequences that overlap the mitochondrial genome⁴³. The captured library was then sequenced on an Illumina NovaSeq sequencer at the Kinghorn Centre for Clinical Genomics (Sydney, NSW, Australia). Raw reads from both sequencing runs were processed and mapped against the mitochondrial revised Cambridge Reference Sequence (rCRS)⁴⁴ using nf-core/eager³⁵ as described previously.

The sequencing read pileups were visualised in Geneious v2022.1.1 (Biomatters; <https://www.geneious.com>) and genetic polymorphisms called with a minimum coverage of 3x and a minimum frequency 0.5. The assembly and the resulting list of SNPs were verified manually and compared to HaploGrep2 (v2.4.; <https://haplogrep.uibk.ac.at/>)^{45,46}, which uses Phylotree 17 version (<http://www.phylotree.org/index.htm>)⁴⁷. The mitochondrial haplogroup was determined as V (Table S8), which is a relatively rare European haplogroup that has been observed at higher-than-average levels in northern Iberia^{48–50}.

- Nuclear SNPs

We performed a principal components analysis (PCA) using the *smartpca* program v10210 (EIGENSOFT) on 573,742 SNPs from 211 modern Western Europeans of the Human Origins dataset⁵¹ and including individual 11/15. Secondly, we performed a PCA on 573,239 SNPs from the same 211 modern individuals⁵¹ and projected individual 11/15 with 311 ancient Iberians⁵²⁻⁶¹ ("Allen Ancient DNA Resource <https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>", version 50.0) onto the modern genetic variation. Our results showed individual 11/[15] locates within the modern and ancient Iberian genetic ancestry (Figure S9 and S10).

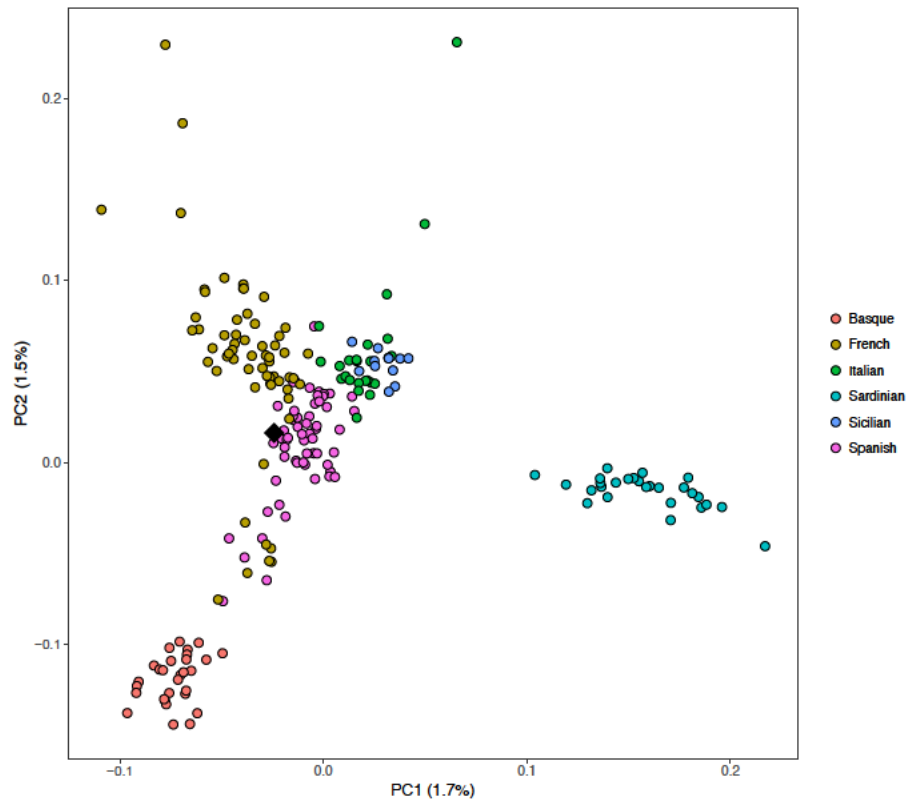


Figure S9: PCA of 211 present-day Western Europeans and individual 11/15 (represented as a black diamond).

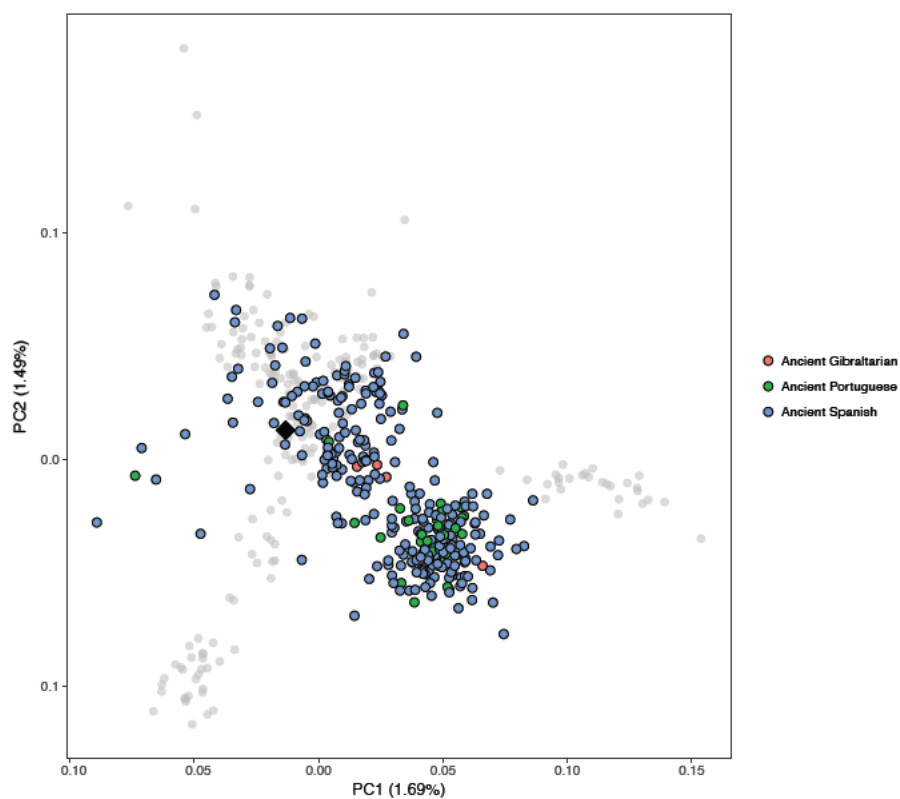


Figure S10: PCA of 211 present-day Western Europeans (as in Figure S9) with 312 projected ancient Iberians (including individual 11/15 represented as a black diamond).

References

1. Racimo F, Sikora M, Vander Linden M, Schroeder H, Lalueza-Fox C. Beyond broad strokes: sociocultural insights from the study of ancient genomes. *Nature Reviews Genetics* 2020;21(6):355–66.
2. Sunna Ebenesersdóttir S, Sandoval-Velasco M, Gunnarsdóttir ED, et al. Ancient genomes from Iceland reveal the making of a human population. *Science* (1979) 2018;360(6392):1028–32.
3. Margaryan A, Lawson DJ, Sikora M, et al. Population genomics of the Viking world. *Nature* 2020;585(7825):390–6.
4. Rivollat M, Jeong C, Schiffels S, et al. Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. *Sci Adv* 2020;6(22):eaaz5344.
5. Moilanen U, Kirkinen T, Saari NJ, et al. A Woman with a Sword?-Weapon Grave at Suontaka Vesitorinmäki, Finland. *European Journal of Archaeology* 2021;(June):1–19.
6. Lanfranco F, Kamischke A, Zitzmann M, Nieschlag E. Klinefelter's syndrome. *Institute of Reproductive Medicine of the University of Münster* 2004;364(11):273–83.
7. Klinefelter HF. Klinefelter's syndrome: historical background and development. *South Med J* 1986;79(9):1089–93.
8. Alarcão J de. *Roman Portugal, vol. II, fasc 1. Warministe. 1988.*
9. Lemos F de S. *O povoamento romano de Trás-os-Montes Oriental. 1993;*
10. Redentor, A.; André, C.; Costa, M.C.; Carvalho, P.C.; Tereso S. Torre Velha de Castro de Avelãs (Bragança). Resultados Arqueológicos e Novidades Epigráficas. In: Soares, C; Brandão, J. L.; Carvalho PC, editor. *História Antiga: Relações Interdisciplinares. Paisagens Urbanas, Rurais & Sociais. Universidade de Coimbra: Hvmnitas Svpplementvm-Estudos Monográficos; 2018. p. 321–44.*
11. Tereso S, Brito A, Umbelino C, Cipriano M, André C, Carvalho P. Arqueologia funerária alto medieval da Torre Velha (Castro de Avelãs, Bragança). In: Quirós Castillo JA, Castellanos García S, editors. *Identidad y etnicidad en Hispania. Propuestas teóricas y cultura material en los siglos V- VIII (DAM 8). Universidad del País Vasco; 2015. p. 145–60.*
12. Reimer PJ, Austin WEN, Bard E, et al. The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). *Radiocarbon* 2020;62(4):725–57.
13. Buikstra J, Ubelaker D. Standards for Data Collection from Human Skeletal remains. *Fayetteville: Arkansas Archaeological Survey Research Series* 1994;44.
14. Bruzek J. A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology* 2002;117(2):157–68.
15. De Mendonça MC. Estimation of height from the length of long bones in a Portuguese adult population. *American Journal of Physical Anthropology* 2000;112(1):39–48.
16. Cordeiro C, Muñoz-Barús JI, Wasterlain S, Cunha E, Vieira DN. Predicting adult stature from metatarsal length in a Portuguese population. *Forensic Science International* 2009;193(1):131.e1-131.e4.
17. José A, Tomé L, Coelho C, Cunha E, Umbelino C, Ferreira MT. The Unidentified Skeletal Collection of Capuchos Cemetery (Santarém) housed at the University of Coimbra. *Antropologia Portuguesa* 2021;38:79–98.
18. Wasterlain SN, Alves R V., Garcia SJ, Marques A. Ovarian teratoma: A case from 15th–18th century Lisbon, Portugal. *International Journal of Paleopathology* 2017;18(February):38–43.
19. Chang S, Skakkebaek A, Trolle C, et al. Anthropometry in Klinefelter syndrome - Multifactorial influences due to CAG length, testosterone treatment and possibly intrauterine hypogonadism. *Journal of Clinical Endocrinology and Metabolism* 2015;100(3):E508–17.
20. d'Oliveira Coelho J, Curate F. CADOES: An interactive machine-learning approach for sex estimation with the pelvis. *Forensic Science International* 2019;302.

21. Maier C, Dumančić J, Brkić H, et al. Tooth crown morphology in Turner and Klinefelter syndrome individuals from a Croatian sample. *Acta Stomatologica Croatica* 2019;53(2):106–18.
22. Pentinpuro R, Lähdesmäki R, Pesonen P, Alvesalo L. Crown heights in the permanent teeth of 47,XXY males and 47,XXX females. *Acta Odontologica Scandinavica* 2021;0(0):1–8.
23. Curate JFT, Albuquerque A, Correia J, Ferreira I, de Lima JP, Cunha EM. A glimpse from the past: Osteoporosis and osteoporotic fractures in a portuguese identified skeletal sample. *Acta Reumatologica Portuguesa* 2013;38(1):20–7.
24. Breuil V, Euller-Ziegler L. Gonadal dysgenesis and bone metabolism. *Revue du Rhumatisme (Edition Francaise)* 2001;68(1):32–9.
25. Kosowicz J, Owecki M, el Ali Z, Sowiński J. Premature fusion and excessive calcification of coronal sutures in patients with Klinefelter syndrome. *Neuro Endocrinol Lett* 2006;27(1–2):137–41.
26. Ortner DJ. Identification of pathological conditions in human skeletal remains. 2nd ed. London: London Academic Press; 2003.
27. Waldron T. Paleopathology. Cambridge: Cambridge University Press; 2009.
28. Vikatou I, Hoogland MLP, Waters-Rist AL. Osteochondritis Dissecans of skeletal elements of the foot in a 19th century rural farming community from The Netherlands. *International Journal of Paleopathology* 2017;19:53–63.
29. Schindler OS. Osteochondritis dissecans of the knee. *Current Orthopaedics* 2007;21(1):47–58.
30. Silva AM, Silva AL. Unilateral non-osseous calcaneonavicular coalition: Report of a Portuguese archeological case. *Anthropological Science* 2010;118(1):61–4.
31. Albee ME. Diagnosing tarsal coalition in medieval Exeter. *International Journal of Paleopathology* 2020;28(April 2019):32–41.
32. Llamas B, Willerslev E, Orlando L, Orlando L. Human evolution: a tale from ancient genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2017;372(1713).
33. Brotherton P, Haak W, Templeton J, et al. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nature Communications* 2013;4(1764).
34. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* 2010;5(6).
35. Fellows Yates JA, Lamnidis TC, Borry M, et al. Reproducible, portable, and efficient ancient genome reconstruction with ncore/eager. *PeerJ* 2021;9:1–25.
36. Skoglund P, Northoff BH, Shunkov M V., et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc Natl Acad Sci U S A* 2014;111(6):2229–34.
37. Neukamm J, Peltzer A, Nieselt K. DamageProfiler: fast damage pattern calculation for ancient DNA. *Bioinformatics* 2021;37(April):3652–3.
38. Lamnidis TC, Majander K, Jeong C, et al. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nature Communications* 2018;9(1):1–12.
39. Gower G, Fenderson LE, Salis AT, et al. Widespread male sex bias in mammal fossil and museum collections. *Proc Natl Acad Sci U S A* 2019;116(38):19019–24.
40. Zhao Y, Gardner EJ, Tuke MA, et al. Detection and characterization of male sex chromosome abnormalities in the UK Biobank study. *Genetics in Medicine [Internet]* 2022; Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1098360022007778>
41. Rasmussen M, Guo X, Wang Y, et al. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science (1979)* 2011;334(6052):94–8.
42. Rohrlach AB, Papac L, Childebayeva A, et al. Using Y-chromosome capture enrichment to resolve haplogroup H2 shows new evidence for a two-path Neolithic expansion to Western Europe. *Scientific Reports [Internet]* 2021;11(1):1–11. Available from: <https://doi.org/10.1038/s41598-021-94491-z>

43. Green RE, Malaspina AS, Krause J, et al. A Complete Neandertal Mitochondrial Genome Sequence Determined by High-Throughput Sequencing. *Cell* 2008;134(3):416–26.
44. 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.
45. Weissensteiner H, Pacher D, Kloss-Brandstätter A, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res* 2016;44.
46. Kloss-Brandstätter A, Pacher D, Schönher S, et al. HaploGrep: A fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Human Mutation* 2011;32(1):25–32.
47. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 2009;30(2):386–94.
48. Maca-Meyer N, Sánchez-Velasco P, Flores C, et al. Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). *Annals of Human Genetics* 2003;67(4):329–39.
49. Torroni A, Bandelt HJ, D’Urbano L, et al. mtDNA analysis reveals a major late paleolithic population expansion from southwestern to northeastern Europe. *American Journal of Human Genetics* 1998;62(5):1137–52.
50. Izagirre N, De La Rúa C. An mtDNA analysis in ancient Basque populations implications for haplogroup V as a marker for a major paleolithic expansion from southwestern Europe. *American Journal of Human Genetics* 1999;65(1):199–207.
51. Lazaridis I, Nadel D, Rollefson G, et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* 2016;536(7617):419–24.
52. Olalde I, Haak W, Barnes I, Lalueza-Fox C, Reich D. The Beaker Phenomenon and the Genomic Transformation of Northwest Europe. *Nature* 2018;555(7695):190–6.
53. Martiniano R, Cassidy LM, Ó’Maoldúin R, et al. The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genetics* 2017;13(7).
54. Fu Q, Posth C, Hajdinjak M, et al. The genetic history of Ice Age Europe. *Nature* 2016;534(7606):200–5.
55. Günther T, Valdiosera C, Malmström H, Ureña I, Rodríguez-Varela R. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc Natl Acad Sci U S A* 2015;112(38):11917–22.
56. Valdiosera C, Günther T, Vera-Rodríguez JC, et al. 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* 2018;115(13):3428–33.
57. Mathieson I, Lazaridis I, Rohland N, et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 2015;528(7583):499–503.
58. Olalde I, Mallick S, Patterson N, et al. The genomic history of the Iberian Peninsula over the past 8000 years. *Science* (1979) 2019;1234(March):1230–4.
59. Lipson M, Szécsényi-Nagy A, Mallick S, et al. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* 2017;551(7680):368–72.
60. Villalba-Mouco V, van de Loosdrecht MS, Posth C, et al. Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula. *Current Biology* 2019;29(7):1169–1177.e7.
61. González-Fortes G, Tassi F, Trucchi E, et al. A western route of prehistoric human migration from Africa into the Iberian Peninsula. *Proceedings of the Royal Society B: Biological Sciences* 2019;286(1895).

Chapter II: Paleogenetic studies of Mas d'en Boixos (Catalonia, Spain)

Statement of Authorship

Title of Paper	Paleogenetic studies of Mas d'en Boixos (Catalonia, Spain)
Publication Status	<input type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input checked="" type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	

Principal Author

Name of Principal Author (Candidate)	Xavier Roca Rada
Contribution to the Paper	Conceptualisation, performed laboratory experiments, bioinformatic processing, data analyses, results interpretation and drafted the manuscript.
Overall percentage (%)	75%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	
	Date 12/10/2022

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Xavier Esteve Gràcia
Contribution to the Paper	Archaeological context
Signature	
	Date 23-10-2022

Name of Co-Author	Dr Yassine Souilmi
Contribution to the Paper	Co supervised the Mr Roca Rada, helped interpret the results, and provided input on the manuscript.
Signature	
	Date 20/10/2022

Name of Co-Author	Jão Teixeira
Contribution to the Paper	Supervision, technical and general guidance, edited the manuscript
Signature	
	Date 14/10/2022

Name of Co-Author	Núria Armentano Oller		
Contribution to the Paper	Sample selection and extraction, historical contextualization.		
Signature		Date	20/10/2022

Name of Co-Author	Diana C. Vinuesa-Espinosa		
Contribution to the Paper	Samples preparation and DNA extractions		
Signature		Date	27/10/2022

Name of Co-Author	Assumpció Malgosa		
Contribution to the Paper	Provide samples, conceptualization and discussion of results		
Signature		Date	27/10/2022

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Conceptualisation, technical and general guidance, edited the manuscript		
Signature		Date	18/10/2022

Name of Co-Author	Cristina Santos		
Contribution to the Paper	Provide samples, conceptualization, technical guidance, and discussion of results		
Signature		Date	24/10/2022

1. Introduction

The Mas d'en Boixos (MDB) archaeological site is the greatest prehistoric site of the Catalan Pre-littoral depression in the Penedès region (Barcelona, Catalonia, Spain) (Figure 1)¹⁻³. Several excavation seasons have been undertaken since 1997, uncovering more than 450 structures¹⁻³. Stratigraphic layers range from the Early Neolithic to the Ancient Iberian Period (~7,500-2,200 years ago; ~7.5-2.2 ka)¹⁻³.

Two major cultural shifts took place in Europe during this time: (i) the Neolithic (N) transition from hunting and gathering to farming driven by population movements from the Middle East (~8.5 ka)⁴; and (ii) the Bronze Age (BA) transition driven by migrations of Western Steppe herders into North and Central Europe (~5 ka)⁵. Recent studies analysing these significant events support the existence of a clear N transition and a more nuanced BA transition in Iberia⁶⁻

10.

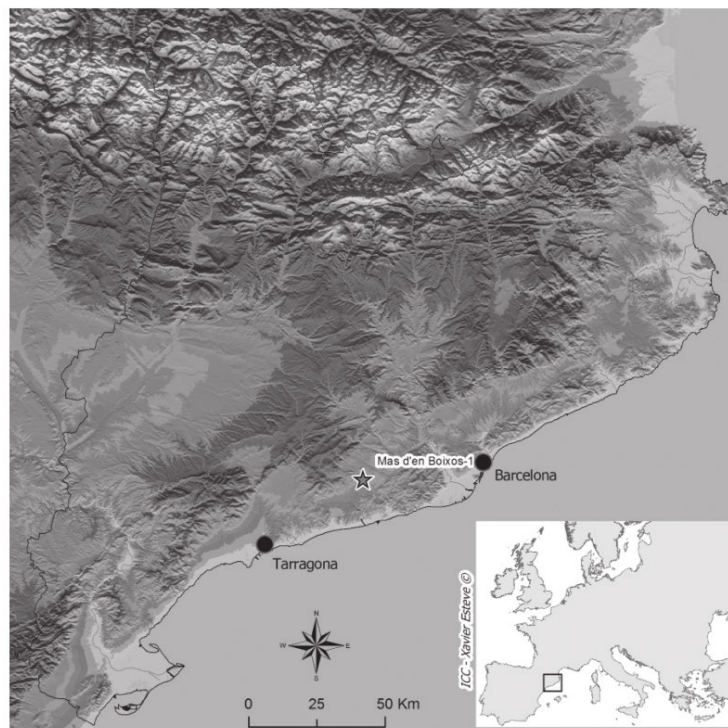


Figure 1. Location of Mas d'en Boixos.

In this study, I genetically analysed 25 new individuals from MDB. Two of them date back to the Middle Neolithic (MN) and were found in the same burial (E88). Anthropologically, both individuals were classified as females, with an estimated age at death of 25-35 years. Archaeologically, they were associated to the Post-Cardial culture with radiocarbon dates of $4,908 \pm 33$ BP and $4,994 \pm 33$ BP.

Additionally, the other 23 individuals date back to the Early Bronze Age (EBA). All these individuals were unearthed from the same archaeological structure (E35): a hypogeum (Figure 2). A hypogeum is formed by one cavity with lateral access through a vertical pit that connects the entry with the outside. In this case, it had an exclusively funeral use with successive burials¹. Only a few archaeological objects were associated with the burial which hindered the archaeological culture determination. Interestingly, a small piece of bronze was found, which was consistent with the associated time period and the two radiocarbon dated individuals ($3,095 \pm 50$ BP and $3,350 \pm 60$ BP). As most of the skeletal remains were intermingled and their preservation is poor, different bones were used for the determination of the minimum number of individuals. A total of 24 skulls and one-sided humerus bones were found. Based on the morphological traits of the skulls, 9 and 8 individuals were assigned as male and female, respectively, whereas the rest were undetermined. Only 23 mandibles were retrieved, 18 of which being identified as adults and 5 as subadults.

Interestingly, the relationship between individuals buried in close proximity during the BA in Catalonia has generated debates within the scientific community. The main hypothesis supports nuclear family ties^{11,12}, while other studies suggested a cultural connection in addition to kinship, proposing an extended family structure concept¹²⁻¹⁶.



Figure 2. The EBA hypogeum (E35) from MDB.

Moreover, I also included 3 individuals from the Iron Age (IA; ~5-2 ka) that were previously genetically analysed by Olalde et al. (2019)⁸. While the human remains retrieved from the IA are scarce at MDB, these three specimens are quite exceptional as they represent inhumation burials instead of cremation, the common IA funerary practice¹⁷. Olalde et al. (2019)⁸ analysed two individuals from one burial (E448): a male with an estimated age at death of 15-25 (515–375 cal BCE - $2,350 \pm 30$ BP, Beta-495153) and a 6-year-old whose biological sex was not determined (515–375 BCE). Olalde and colleagues also analysed another female (507–366 cal BCE - $2,340 \pm 30$ BP, Beta-495155) with an estimated age at death of 35-45 from another burial (E449).

The main aims of this project are to: (i) infer kinship relationships between individuals buried in close proximity; (ii) determine diachronic genetic changes at MDB; (iii) contextualise MDB within the rest of Iberia; and (iv) investigate the BA transition in Iberia. It is worth noting that this chapter is still work in progress, with further analyses proposed at the end of the discussion.

2. Methods

2.1. Ancient DNA laboratory work

Ancient DNA analysis was conducted in two different clean-room facilities—University of Adelaide’s Australian Centre for Ancient DNA (ACAD), Australia; and Autonomous University of Barcelona, Spain—following strict precautions to minimise contamination¹⁸. Samples were processed while wearing face masks, visors, hooded coverall, hair net, and gloves, in laboratories exclusively dedicated to ancient DNA analysis with positive air flow pressure. Standard precautions to avoid contaminations were employed¹⁹.

A total of 25 well-preserved teeth were included in this study. Bone powder was extracted from cementum, which is richer in DNA than dentin²⁰, and DNA was extracted using a method optimised to retrieve degraded ancient DNA fragments²¹ at the Autonomous University of Barcelona. Partially UDG-treated double-stranded DNA libraries were generated²² at ACAD, University of Adelaide. Paired-end shotgun sequencing using an Illumina NovaSeq (200 cycles) was performed by service providers at the Kinghorn Centre for Clinical Genomics (Sydney, Australia).

All post-shotgun sequencing lab work was performed in ACAD’s post-PCR facilities. Libraries were reamplified and subjected to in-solution target hybridisation using an in-house protocol to capture mitochondrial fragments¹⁹. Subsequently, all libraries were captured again using the myBaits Expert Human Affinities Prime Plus Kit by DAICEL Arbor Biosciences (Ann Harbor, MI, USA), which targets 1.24 million nuclear single nucleotide polymorphisms (SNPs) (hereafter 1240k) used in population genetics²³, as well as the mitochondrial genome and 46 thousand SNPs on the Y chromosome. After each capture, libraries were paired-end sequenced at the Kinghorn Centre for Clinical Genomics (Sydney, NSW, Australia) using an Illumina NovaSeq (200 cycles).

2.2. Sequencing data processing

Raw reads from the shotgun sequencing run were processed and mapped against the human reference genome (GRCh37d5) using the nf-core/eager²⁴ v2.4.0 pipeline. To determine data authenticity, I evaluated the endogenous DNA percentage, determined the fragment size distribution, and calculated the post-mortem damage rate at the read termini using DamageProfiler²⁵.

Raw reads from all other sequencing runs were processed and mapped against the human reference genome (hg19) using the nf-core/eager²⁴ v2.4.0 pipeline. A subset of libraries passed authenticity quality thresholds to undergo downstream analyses (See Results, Table 1, and Table 2). Retained reads were trimmed 2 bp from each end using the trimBam function of bamUtil (<https://github.com/statgen/bamUtil>). I applied read q30 and Q30 base quality filters through samtools²⁶ version 1.10 mpileup function and I then performed pseudohaploidisation on the ancient samples by randomly selecting one allele at heterozygous sites and subsetted the genome-wide data to match the number of SNPs in the 1240k capture panel using pileupCaller (<https://github.com/stschiff/sequenceTools>). Pseudohaploidisation is a standard step in aDNA studies due to the highly degraded nature of aDNA meaning that it is often not possible to call diploid genotypes and even if a site presents more than one allele it could derive from actual heterozygosity, sequencing errors or deamination, whereby a single allele is randomly chosen to represent each locus.

2.3. Sex determination

Using shotgun data, genetic sex determination was assigned using the method described by Gower et al. (2019)²⁷. This method uses counts of reads mapping to the X chromosome and autosomes to determine the genetic sex of an individual. Under the assumption that the number

of reads obtained should reflect the chromosome copy numbers and chromosome lengths, two binomial models are constructed, one for males and one for females. A likelihood ratio test is then used to distinguish between the two models. A read dosage on the X chromosome (M_x) close to 0.5 or 1 means that the individual is assigned as male or female, respectively.

To compare genetic sex results, I also used SexDetERRmine (<https://github.com/nf-core/modules/tree/master/modules/nf-core/sexdeterminer>) with default quality cut-off values for -q30 and -Q30. SexDetERRmine calculates X- and Y- ratios, corresponding to the number of reads mapping to each of the sex chromosomes relative to the autosomes.

2.4. Uniparental markers

2.4.1. Y Chromosome

To obtain Y-chromosome haplotypes for the male individuals, I converted the Eigenstrat files into Packedped files and then into a VCF file selecting only the Y chromosome sites. The VCF file was inputted into yhaplo v1.1.2²⁸ to automatically determine the haplotypes.

2.4.2. Mitochondrial DNA

Raw reads from all the sequencing runs were processed and mapped against the mitochondrial revised Cambridge Reference Sequence (rCRS)²⁹ using nf-core/eager²⁴ v2.4.0 pipeline. The sequencing read pileups were visualised in Geneious v2022.1.1 (Biomatters; <https://www.geneious.com>) and polymorphisms were called with minimum a coverage of 3 and a minimum variant frequency 0.5. The assembly and the resulting list of SNPs were manually verified and compared to HaploGrep2 v2.4.0 (online implementation) from the Division of Genetic Epidemiology of the Medical University of Innsbruck (<https://haplogrep.uibk.ac.at/>)^{30 31}, which uses Phylotree v17

(<http://www.phylotree.org/index.htm>)³².

In addition, I calculated mitochondrial contamination using the mitoverse Haplocheck cloud service³³. This method checks for the contamination status by determining if multiple mitochondrial haplotypes are present within a single sample. Multiple haplotypes are detected through polymorphic variants and the resulting sequence is split into two components that then undergo haplotype assessment using haplogrep to determine contamination based on polymorphic site frequencies.

2.5. Kinship analysis

I used READ v3.0 to determine kinship between pairs of individuals³⁴. In the default settings, READ performs a normalisation step using the median of all pairwise differences as an estimate of the expected distance between two unrelated individuals.

2.6. Principal Component Analysis (PCA)

I performed a principal components analysis (PCA) using the smartpca program v10210 (EIGENSOFT) on 616,427 SNPs from 211 modern Western Europeans of the HGDP³⁵, and projected³⁶ the studied ancient samples along 245 ancient Iberians^{6 8 10 23 37 41} ("Allen Ancient DNA Resource <https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>", v50.0) onto the modern-day genetic variation. Of note, this version does not include the latest genetic dataset coming from south-eastern Iberia by Villalba-Mouco et al. (2021)⁹.

2.7. ADMIXTURE

Three populations were selected as fixed source groups of ancestry ($k=3$) implementing a supervised ADMIXTURE model³⁴: Western Hunter-Gatherers (WHG)^{8,23,37,39,40,42}, N individuals from Turkey^{23,43–46} representing the Anatolian Farmers ancestry, and BA Eurasians from the Yamnaya culture^{5,10,23,47–50} as representatives of the Steppe ancestry. A total of 239 previously published ancient Iberians from the N to the IA^{6 8 10 23 37 41} were used and 10 replicates were run. Common modes among the different runs were identified and clusters were aligned using pong (<https://github.com/ramachandran-lab/pong/>; v1.4.9).

2.8. f_3 -statistics

To explore the genetic affinities among the studied ancient individuals from MDB, I performed a series of outgroup f_3 -statistics in the form (Mbuti; Ind1, Ind2) using the qp3Pop function implemented in ADMIXTOOLS v7.0.2⁵¹. Since the Mbuti population is an outgroup to all the analysed samples, the resultant f_3 -statistic measures the shared drift between Ind1 and Ind2. I then computed the distance matrices from the outgroup f_3 -statistic results by subtracting the f_3 values from 1. I subsequently performed a classical (metric) Multi-Dimensional Scaling (MDS) from such distance matrices, which is a means of visualising the level of similarity of individual cases of a dataset.

To explore the Anatolian and Steppe ancestries during the BA transition in Europe, I performed various iterations of a configuration of outgroup f_3 -statistics of the form (Mbuti; Pop, X). Mbuti is the outgroup, Pop refers to either N individuals from Turkey to represent the Anatolian Farmers ancestry or BA Eurasians from the Yamnaya culture as representatives of the Steppe ancestry, and X refers to either the studied EBA individuals from MDB or previously published BA Eurasians.

2.9. f_4 -ratio statistic

To estimate the Steppe-like ancestry proportion in ancient populations, I performed a series of f_4 -ratios in the form of $f_4[(\text{Yamnaya1}, \text{Mbuti}; \text{Test}, \text{WHG})/(\text{Yamnaya1}, \text{Mbuti}; \text{Yamnaya2}, \text{WHG})]$ using the `D4RatioTest` function implemented in `ADMIXTOOLS`⁵¹. The f_4 -ratio estimation allows inference of the mixing proportions of an admixture event, even without access to accurate surrogates for the ancestral populations.

3. Results

3.1. Ancient genomic data screening

I genetically screened 25 archaeological samples from 2 different time periods: MN (n=2), and EBA (n=23). For each sample, I evaluated the percentage of endogenous aDNA via shotgun sequencing. Out of the 25 samples, only 3 samples had an endogenous content above 1% (Table 1). As all samples had at least an endogenous content above 0.05% which I personally observed it to be the capture success threshold, I subsequently performed mtDNA and 1240k captures. As expected, hybridisation captures efficiently enriched for DNA targets, resulting in increased endogenous DNA proportions and a drop in the complexity of the libraries (Table 1). After filtering individuals with: (i) more than 10,000 SNPs covered by at least one read of the 1240k SNPs panel (Table 1); (ii) a significant amount of reads mapping and covering the whole mitochondrial rCRS²⁹ (Table 1); (iii) the presence of the misincorporation patterns characteristic of aDNA (>3%) (Table 1); and (iv) absence of contamination (Table 2); I retained 11 samples: MN (n=2) and EBA (n=9).

3.2. Genetic sex, uniparental markers, kinship analysis and genetic affinities between individuals

The two individuals of the MN burial were genetically identified as one male and one female. They were first-degree relatives and shared the same mitochondrial haplotype (K1a4a1). The male individual carried a Y chromosome haplogroup I (Table 2 and Figure 3). These results suggest these individuals were either mother and son or siblings.

Seven males and two females were genetically determined from the EBA hypogeum (Table 2). All nine individuals had different mitochondrial haplogroups (H1bm, U2e1, R0a, H1e1a, U5b1, H1c, H3, HV0+195, and X2b+226) and all the Y chromosome subhaplogroups were part of the R1b lineage. Of note, I could not assign a more derived Y haplogroup branch within the basal F lineage for one individual (24193) due to missing diagnostic SNP calls. While most of the individuals were unrelated based on kinship analyses, a male (24218) had a second-degree relationship (i.e., grandchild-grandparent, half-siblings, or uncle-aunt to nephew-niece) with a female (24215) and another male (24197) (Figure 3). Regarding the male-female second-degree relationship, as they both have different mitochondrial haplogroups, their only possible relationship on the maternal side would be grandfather-granddaughter, but on the paternal side could be any of the following: grandparent-grandchild, half-siblings, or uncle-aunt to nephew-niece. Regarding the relationship between the two males, further analyses are needed. On the one hand, if 24218 was found to be ancestral for the SNPs determining R1b1a2a, R1b1a2a1, R1b1a2a1a, R1b1a2a1a2, R1b1a2a1a2a and R1b1a2a1a2a1, which are derived in 24197, the two Y chromosome haplogroups would be different. As their mitochondrial haplogroups are also different, their only possible relationship would be maternal grandfather-grandson. On the other side, if 24218 had data missingness on the definitory SNPs determining R1b1a2a, R1b1a2a1, R1b1a2a1a, R1b1a2a1a2, R1b1a2a1a2a and

R1b1a2a1a2a1, which is the most likely scenario given the low number of recovered SNPs, it would open the possibility for other paternal-side relationships, such as grandfather-grandson, half-brothers, and uncle-nephew.

Finally, the IA individuals previously analysed by Olalde et al. (2019)⁸ were genetically identified as two males and one female. They had different mitochondrial haplogroups (H, H3, and J1c1) and were unrelated. However, the two males shared the same Y chromosome haplogroup (R1b1a2a1a2) (Table 2 and Figure 3).

Of note, the P0 values computed on READ are higher than previously published P0 values on 1240k SNPs for ancient post-Mesolithic Europeans. Moreover, most comparisons involving one or two IA individuals have the highest P0 values. Generally, BA groups are genetically closer to IA groups than to MN and comparisons involving one MN individual and one BA or IA individual should have higher P0 values. This observation could suggest some possible biases stemming from different capture methods (original 1240k for IA and Arbor Biosciences for MN and BA). Otherwise, different data processing steps could be driving this signal.

Finally, in order to explore the genetic affinities between the MDB individuals across time, I performed various iterations of a configuration of outgroup f_3 -statistics of the form (Mbuti; Ind1, Ind2) and a MDS analysis (Figure 4) that suggested a clear genetic grouping of the MN individuals away from the rest. While the EBA individuals were quite diverse, two of the three IA individuals were relatively more distinct from the EBA individuals as well as from each other.

Table 1. Summary of the studied archaeological samples and their paleogenetic results for quality assessment. Successful samples are highlighted in light grey.

Time period	ID	Endogenous DNA (%) Post-Mapping (Shotgun)	Endogenous DNA (%) Post-Dedup (Shotgun)	Endogenous DNA (%) Post-Mapping (Shotgun + Captures)	Endogenous DNA (%) Post-Dedup (Shotgun + Captures)	#SNPs	#Reads Mapped to rCRS ²⁹	5' C→T 1 st base (%)	3' G→A 1 st base (5)	Mean read length (bp)
MN	24227	0.1	0.06	16.81	7.18	34,263	3,755	9.70	9.80	57.62
	24228	2.56	1.78	79.9	24.86	335,985	157,447	20.40	19.70	48.96
EBA	24193	0.08	0.04	15.7	9.49	24,635	35,683	8.90	7.80	59.41
	24194	0.93	0.67	49.73	42.45	186,504	10,074	16.10	14.60	60.55
	24195	5.45	4.11	64.36	54.4	498,694	61,745	11.90	8.80	60.98
	24196	0.07	0.03	20.96	2.78	2,589	669	1.80	1.90	58.66
	24197	2.47	1.71	42.16	29.21	276,125	44,141	15.20	14.60	58.54
	24212	0.3	0.2	49.25	37.13	153,562	8,454	12.90	12.90	60.49
	24213	0.11	0.07	19.08	13.09	33,224	3,232	8.90	7.50	62.19
	24214	0.05	0.02	21.37	1.7	1,944	208	2.90	2.20	57.98
	24215	0.27	0.19	40.3	27.01	188,652	115,753	8.60	7.40	62.13
	24216	0.33	0.12	14.64	2	3,126	1,206	3.00	2.20	58.25
	24217	0.06	0.03	1.6	0.38	1,141	203	4.60	3.50	63.03
	24218	0.09	0.04	50.06	24.15	13,088	22,744	6.10	5.00	57.59
	24219	0.06	0.03	28.03	0.76	1,388	223	3.50	3.60	57.51
	24220	0.05	0.02	13.53	4.49	8,809	486	1.60	1.60	59.79
	24221	0.13	0.09	42.7	34.02	69,263	5,226	1.00	0.90	60.64
	24222	0.06	0.03	16.84	1.1	629	80	2.10	2.10	58.29
	24223	0.08	0.03	29.34	0.99	659	199	2.10	2.20	58.04
	24224	0.07	0.03	8.27	1.05	3,651	470	2.00	1.70	60.87
	24225	0.05	0.03	15.72	8.67	10,404	476	1.20	1.80	61.85
24230	0.07	0.03	60.47	13.21	24,926	1,787	7.00	5.80	58.03	
24231	0.08	0.03	35.29	5.7	11,404	918	3.20	3.50	59.03	
24232	0.08	0.03	17.19	2.05	4,160	388	2.40	2.20	58.14	
24233	0.55	0.4	59.51	53.74	218,398	13,489	11.30	9.10	61.92	

Table 2. Summary of the studied archaeological samples, their genetic sex determination, and uniparental markers haplogroup assignment. ND stands for “Not Detectable”. The IA samples were previously analysed by Olalde et al. (2019)⁸.

Time period	ID	Genetic Sex Determination	Mitochondrial Haplogroup	Mitochondrial Contamination Level	Y Chromosome Haplogroup
MN	24227	M	K1a4a1	ND	I
	24228	F	K1a4a1	ND	-
EBA	24193	M	H1bm	ND	<i>F</i>
	24194	M	U2e1	ND	R1b1a2a1a
	24195	M	R0a	ND	R1b1a2a1a2a1
	24197	M	H1e1a	ND	R1b1a2a1a2a1
	24212	F	U5b1	ND	-
	24213	M	H1c	ND	R1b1a2
	24215	F	H3	ND	-
	24218	M	HV0+195	ND	R1b1a2
	24233	M	X2b+226	ND	R1b1a2a1a2c1f2a1a
IA	I12410	M	H	-	R1b1a2a1a2
	I12877	M	J1c1	-	R1b1a2a1a2
	I12878	F	H3	-	-

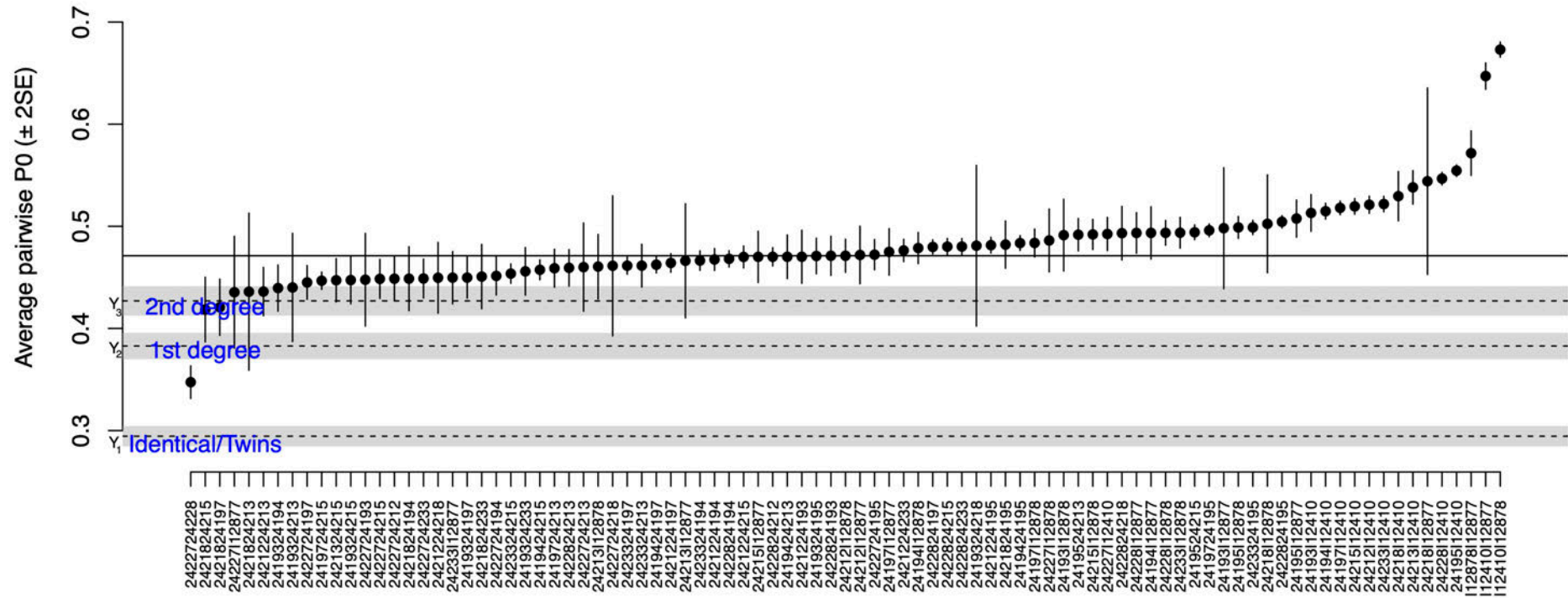


Figure 3. Kinship analysis using READ. Sorted non-normalised average P0 values for all pairwise comparisons between individuals. Error bars show two standard errors. The solid horizontal line indicates the median value used for normalisation. Dashed lines show the cut-offs used to classify the related individuals. The grey areas indicate 95% confidence intervals for the cut-offs, accounting for the uncertainty in estimating the average P0 in the pair used to set the baseline for unrelated individuals. $P_0 < Y_1$ indicates two individuals are same individuals or twins; $Y_1 \leq P_0 \leq Y_2$ indicates two individuals are first degree relatives (parent-offspring or siblings); $Y_2 \leq P_0 \leq Y_3$ indicates two individuals are second-degree relatives (i.e., grandchild-grandparent, half-siblings, or uncle-aunt to nephew-niece); and if $P_0 > Y_3$ the two individuals can be considered unrelated.



Figure 4. Multidimensional scaling ($1-f_3$) for the MDB individuals. Outgroup pairwise f_3 computed with Mbuti as outgroup.

3.3. Genetic affinities between MDB and ancient Iberians

I performed a PCA by projecting ancient samples onto PC axes computed using present-day south-western Europeans from the HGDP³⁵ (Figure 5). The two MN fall in a cluster of Iberian individuals from the N and the Chalcolithic (C) located between present-day Basques and Sardinians, whereas all the EBA and IA samples fall in a cluster of Iberian individuals from the BA and IA shifting towards present-day Spanish.

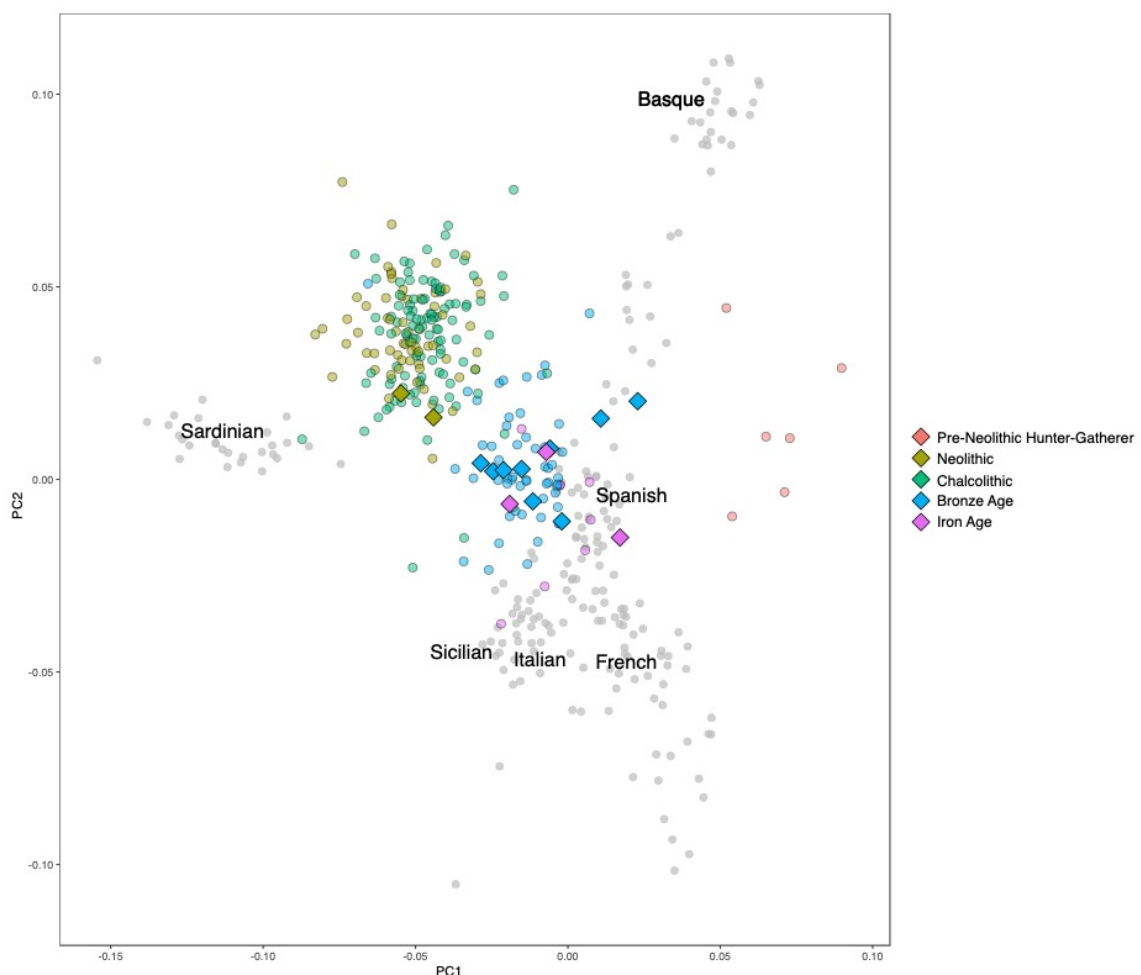


Figure 5. PCA of 211 present-day Western Europeans (light grey dots) with 259 projected ancient Iberians as coloured circles, including the MDB individuals as coloured diamonds.

The ADMIXTURE analysis ($k=3$; Figure 6) suggests that during the N and C, all Iberians as well as the two MN individuals from MDB had an admixed genetic ancestry made of a WHG component (red) and another one based on N individuals from Turkey (green), representing the Anatolian ancestry. During the BA, the Yamnaya ancestry (blue) representing the Steppe Ancestry was present in most Iberians and the EBA samples from MDB. In fact, MDB had a slightly higher component than the rest of their contemporary individuals. Finally, IA individuals had, in general, a higher Steppe component than BA individuals.

3.4. BA transition in Iberia

I performed several analyses to test the distinctness of the BA transition in Iberia. Firstly, a comparison with Europeans and Middle Easterners was explored. Two values of outgroup f_3 -statistics were computed per individual, one being $f_3(\text{Mbuti}; \text{Anatolia_N}, \text{Test})$ on the x-axis representing the shared genetic drift with Anatolia, and the other being $f_3(\text{Mbuti}; \text{Yamnaya_BA}, \text{Test})$ on the y-axis representing the shared genetic drift with the Steppe (Figure 7). While most individuals had relatively similar values of shared genetic drift with Anatolia, the shared genetic drift with the Steppe was quite diverse and formed different clusters. A cluster including individuals from the British Isles, Scandinavia, and Eastern Europe had the highest shared genetic drift with the Steppe (represented in shades of blue). I also uncovered a cluster of Central Europeans with a very wide distribution, going from the previous cluster all the way to the following ones (represented in light grey). Next, the analysis showed an extended cluster with a more pronounced cline formed by Iberians, Middle Easterners and South-eastern Europeans (represented in shades of pink). The MDB individuals are part of the Iberian subcluster, being quite diverse and some shifting towards the Central Europeans cluster. The last cluster, being closer to the Iberian subcluster, includes individuals from the Western Mediterranean Islands and Italy (represented in shades of amber).

I further investigated the Steppe ancestry component in ancient Iberians from the N to the IA computing f_4 -ratios and discarding individuals whose absolute Z-scores were higher than two times their standard error (Figures 8 and 9). While the Steppe ancestry proportion in Spain increases with time and peaks during the IA, it remains absent in BA Portugal and totally unknown during the Portuguese IA. All the BA values were plotted spatially (Figure 10), and the Steppe ancestry proportion decreases south-westwards in Iberia, being higher in areas close to the Pyrenees and lower in the extreme south of Portugal, potentially suggesting a non-homogenous BA transition in Iberia.

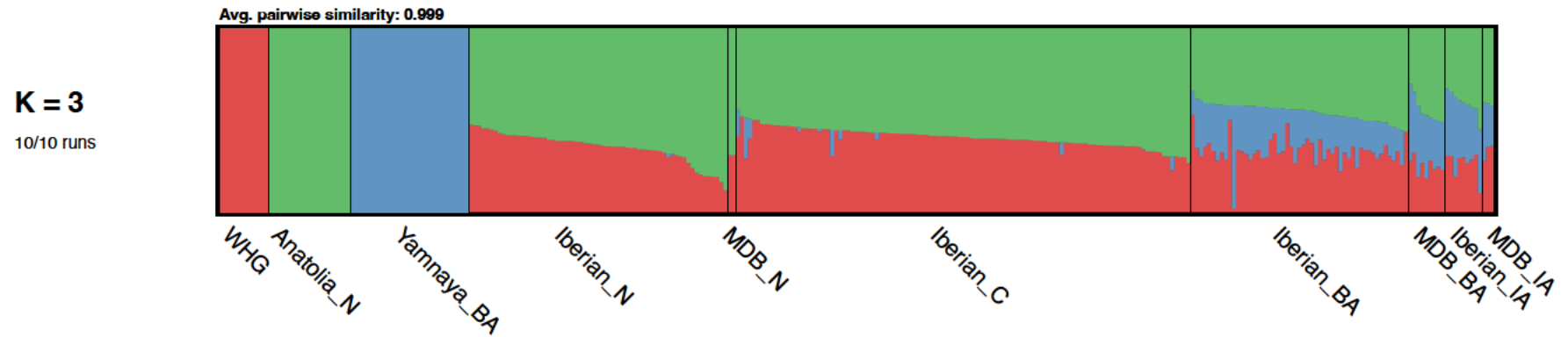


Figure 6. Ancestry proportions estimated with a supervised ADMIXTURE at $k=3$, with the three predefined ancestry being Western Hunter-Gatherers (WHG; red), Turkish individuals from the Neolithic representing the Anatolian Farmers ancestry (Anatolia_N; green) and Bronze Age Eurasians from the Yamnaya culture representing the Steppe ancestry (Yamnaya_BA; blue) with ancient Iberians from the Neolithic (N), Chalcolithic (C), Bronze Age (BA) and Iron Age (IA).

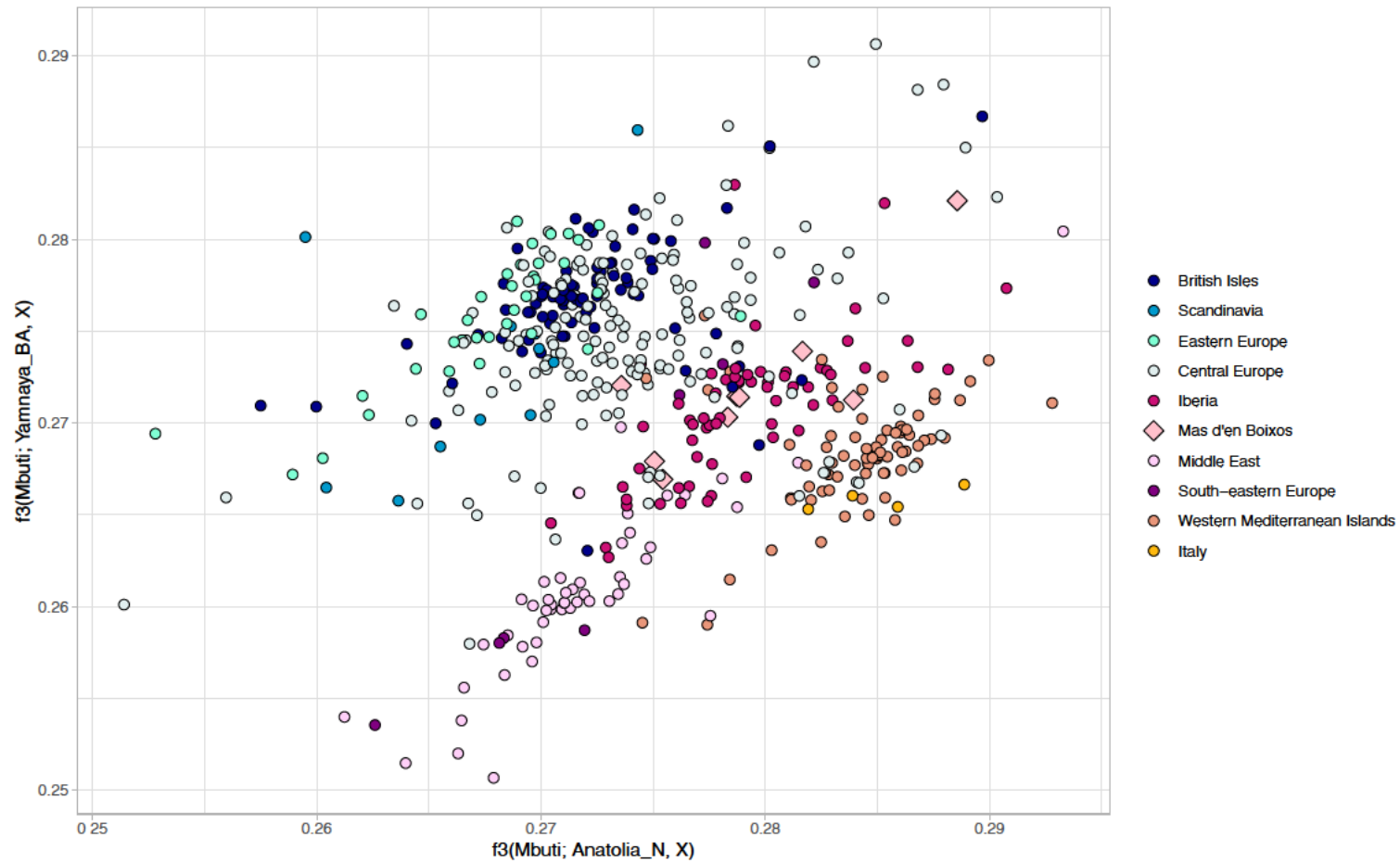


Figure 7. Outgroup f_3 -statistics of the form $f_3(\text{Mbuti}; \text{Pop}, X)$ showing the amount of shared drift between the ancient European individual (X) with Turkish individuals from the Neolithic representing the Anatolian Farmers ancestry (Anatolia_N; x-axis) and Bronze Age Eurasians from the Yamnaya culture representing the Steppe ancestry (Yamnaya_BA; y-axis). Clusters of individuals are represented in shades of blue, pink, and amber from the top-left to the bottom-right. The MDB individuals are represented as coloured diamonds.

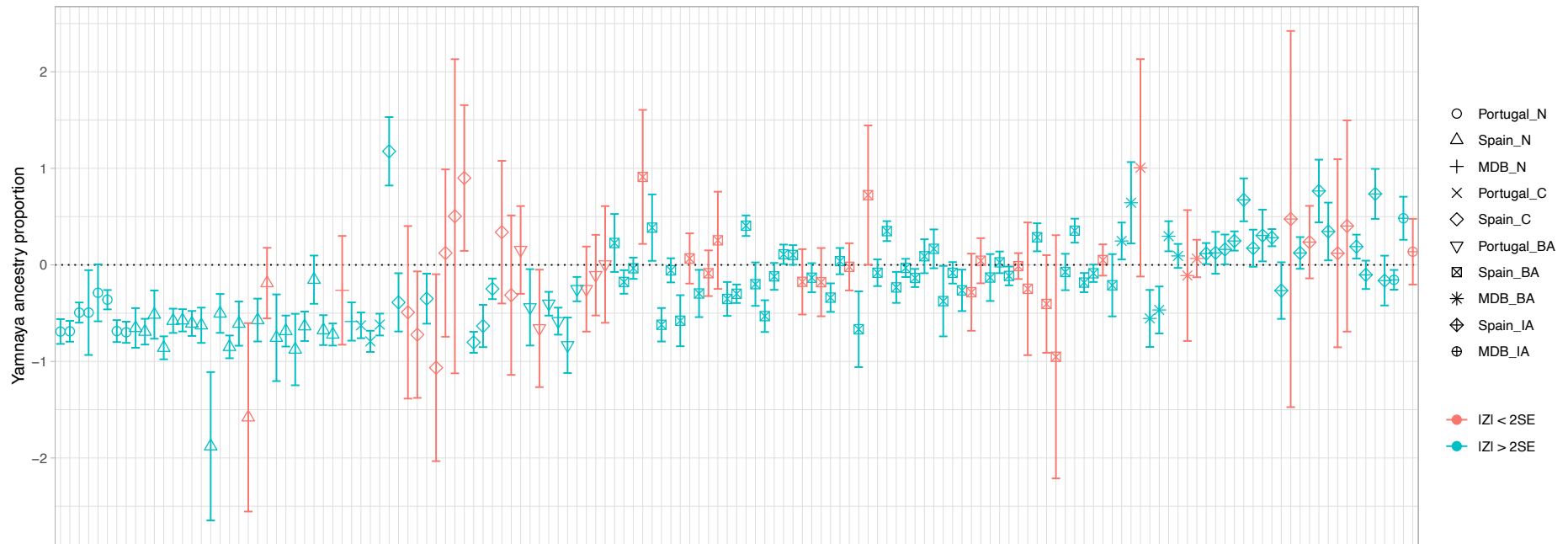


Figure 8. Ancient Iberians distributed along the x-axis according to their geographical location (Portugal, Spain, and MDB) and temporal period (Neolithic, Chalcolithic, Bronze Age and Iron Age - N, C, BA, and IA, respectively) and their f_4 -ratios values along the y-axis representing the Yamnaya ancestry proportion. Discarded individuals are coloured in red as their absolute Z-scores were lower than two times the standard error.

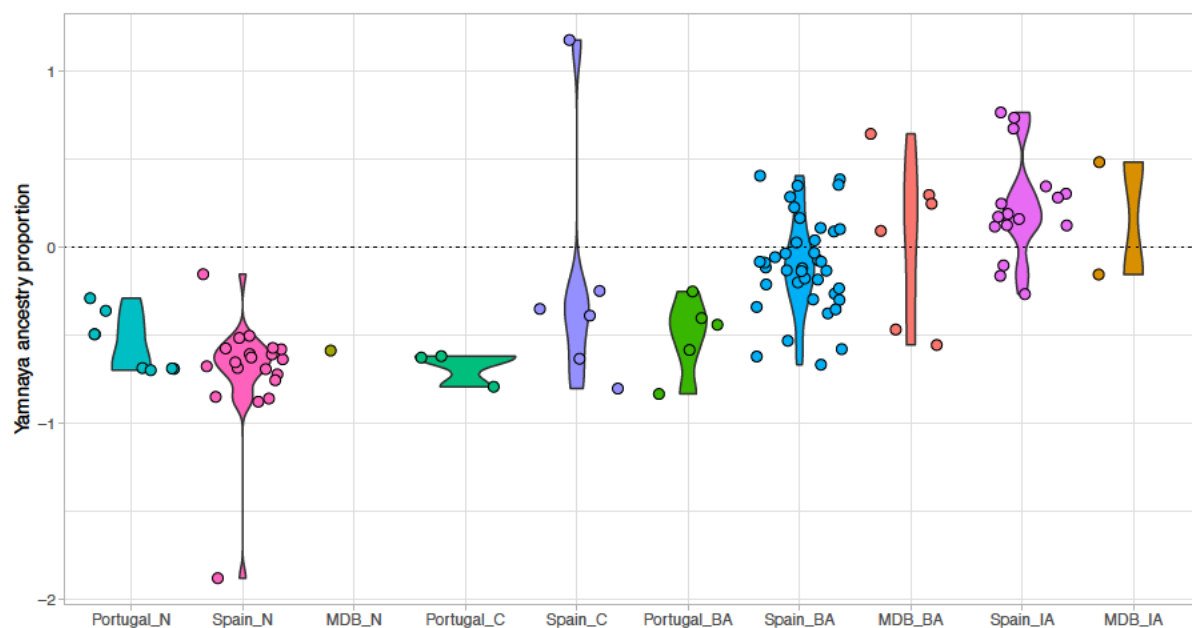


Figure 9. Ancient Iberian populations along the x-axis according to their geographical location (Portugal, Spain, and MDB) and temporal period (Neolithic, Chalcolithic, Bronze Age and Iron Age - N, C, BA, and IA, respectively) and the Yamnaya ancestry proportion of each individual computed with f_4 -ratio on the y-axis.

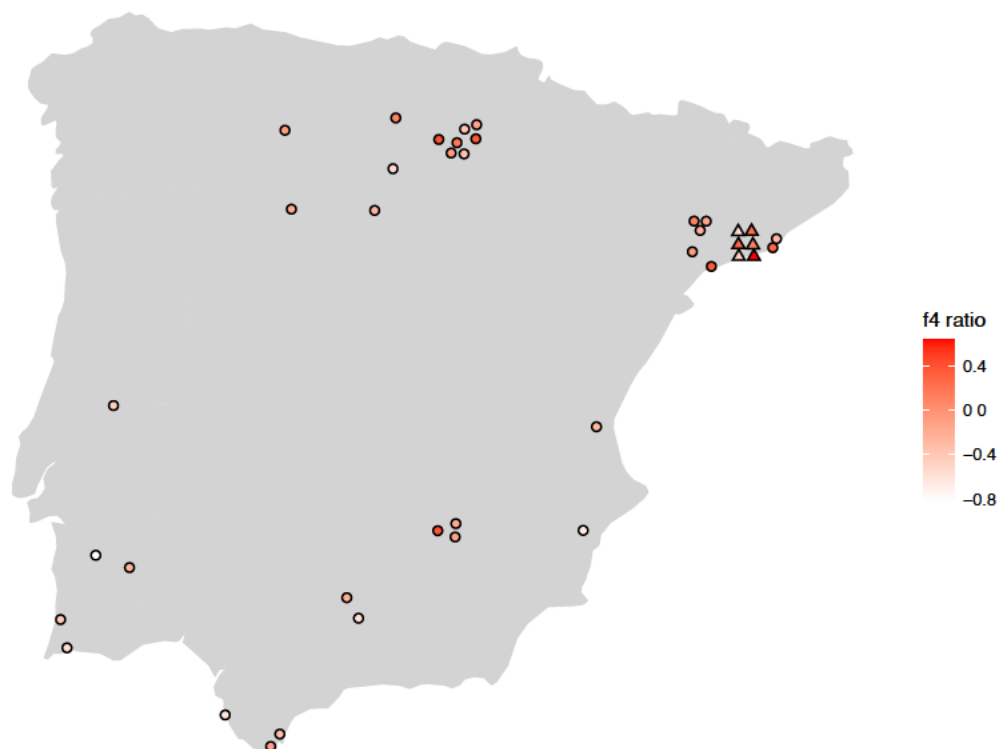


Figure 10. The Yamnaya ancestry proportion in Iberia during the BA computed with f_4 -ratios coloured from highest to lowest using a red-white scale. The MDB individuals are represented as triangles. The Yamnaya ancestry proportion decreases south-westwards.

4. Discussion

4.1. MN burial

While the anthropological study assigned two females to the MN burial associated to the Post-Cardial culture, genetic analyses determined a male and a female. They were first-degree relatives and shared the same mitochondrial haplotype (K1a4a1). Therefore, they probably were either mother and son or siblings.

In the Iberian Peninsula, other N individuals have been found carrying the K1a4a1 subhaplogroup^{8,10,39,52–54}. Interestingly, the same subhaplogroup—but a slightly different haplotype—was found in a C sample from the Paris Street archaeological site in Cerdanyola del Vallès¹⁰, only ~52 km away from MDB. In fact, it has been suggested that the mitochondrial haplogroup K1a probably reached the European continent ~8 ka during migrations of EN people from the Near East through continental and Mediterranean routes^{52,55–57}.

The MN male from MDB carried a Y chromosome haplogroup I. It has been suggested that I lineages were already present in Central European hunter-gatherers^{36,53,58,59} and they were assimilated by the aforementioned wave of East Mediterranean N farmers moving to the west, associated with the mitochondrial and Y chromosome haplogroups K1a and G2a, respectively. In fact, south-eastern and south-western European farmers from the Starčevo culture—which gave rise to the Linear Pottery culture—and Cardial Ware culture carried both I and G2a lineages^{23,36,47,53,55,58,60,61}. Interestingly, Olalde et al. (2015)⁶¹ suggested that populations related to the Cardial and Linear Pottery cultures were descended from a common farming population in the Balkans, which had subsequently migrated further westwards into Europe along the Mediterranean coast and Danube River. The Cardial-related individuals were found to be most

closely related to present-day Sardinians and Basques⁶¹. Indeed, the nuclear genetic information of the studied Post-Cardial individuals from MDB suggested the same affinity.

4.2. EBA hypogeu and the BA transition in Iberia

Archaeological evidence has demonstrated the existence of collective funerary practice differences between coastal sites and mountain areas during the Catalan BA. Coastal sites, like MDB, mainly include megalithic constructions, while mountainous areas, such as Montanissell¹⁴, are characterized by caves, caverns, or dolmens. It has been previously hypothesised that individuals found in these collective burials were part of nuclear family ties^{11,12}. However, an extended family structure unifying kinship and culture in small societies was suggested in Montanissell¹⁴ and might also be the case in MDB. I found seven males and two females carrying different mitochondrial haplogroups and Y chromosome subhaplogroups in the R1b lineage. In addition, my analysis revealed a second-degree relationship between one male and two other individuals (one male and one female).

Most of the EBA individuals from MDB carried maternal lineages (H1, H3, U5, and HV0) that were previously reported for N and C Iberians and probably expanded throughout Europe after the Last Glacial Maximum from the Iberian/Franco-Cantabrian refugium^{62–65}. However, one individual carries a lineage that arrived in Iberia during the N transition (X2b)³⁹, while another one carries a previously unreported lineage in ancient Iberia (U2e1) that has higher affinity to Palaeolithic/Mesolithic Eastern hunter-gatherers and BA Yamnaya culture, but not to N farmers^{37,66}. Moreover, while the haplogroups H1b and H1c are found among Eastern Europeans^{67,68}, little is known about the subhaplogroup H1bm, which has been found in EN Turkmenistan⁴⁸, LN Germany⁴⁸ and BA Spain⁴⁸ and Crete⁶⁹.

Villalba-Mouco et al. (2021)⁹ reported the R0a haplogroup for the first time in ancient Iberia. This haplogroup is presently more frequent in the Arabian Peninsula but has also been found in ancient individuals from the eastern Mediterranean, as well as Ukraine Yamnaya and Sicilian Bell Beakers. Villalba-Mouco and colleagues suggested that the Mediterranean region was a route of dissemination during the C and EBA as R0a has not been found in Central Europe. However, Gandini et al. 2016⁷⁰ suggested that it might represent a relic of Late Glacial or postglacial dispersals around the Mediterranean.

On the other hand, all EBA males had a Y chromosome haplogroup belonging to R1b, which is prevalent in individuals associated with the Steppe Yamnaya culture and associated with the spread of some Indo-European languages⁵³. In Iberia, it has been suggested that there was a Y chromosome turnover during the BA with the arrival of the Steppe ancestry⁸⁻¹⁰, which I observed when comparing the EBA individuals from MDB to their N counterpart.

Nevertheless, the Bronze Age transition must have been different in Portugal. Martiniano et al. (2017)⁷ analysed three MBA men that also had a Y chromosome haplogroup R1b and N Iberian maternal lineages (H1+152, U5b3, X2b+226), but presented a distinctly weaker Steppe ancestry compared to the Centre and North of Iberia, which is also observed in some southeastern Iberians reported by Villalba-Mouco et al. (2021)⁹. Hence, Martiniano et al. (2017)⁷ suggested that Indo-European R1b men entered Iberia around 3.8 ka as a small group of migrants being involved with local women, thus diluting their DNA at each generation, until hardly any Steppe ancestry was left after a few centuries.

These results suggest the BA transition in Iberia was not homogenous, as the Steppe ancestry expansion did not reach the whole peninsula in the same way, having the second lowest Steppe affinity in Europe after Italy and the Mediterranean Islands. Even though this expansion has

been suggested to have been male driven⁸, some mitochondrial lineages associated with Eastern populations have been detected in MDB, suggesting that females were probably migrating as well. In fact, Villalba-Mouco et al. (2021)⁹ did not find any signal for male bias related to Steppe ancestry.

Of note, while in this study most BA Iberians have negative f_4 -ratios (Figure 9), previously published publications reported that BA Iberians generally have 10–20% Yamnaya ancestry, which translates into 20–40% of Central European Bell Beaker-related ancestry, a group that was ~50% Yamnaya ancestry and ~50% Central European N ancestry. These negative values could suggest that the proposed f_4 -ratio is not the best measure of Yamnaya proportions. This would explain why the re-analysed BA Portuguese have negative f_4 -ratios, even though they show a Y chromosome turnover and the same shift in PCA as BA Spanish. In fact, the Steppe ancestry for these individuals was previously found in Olalde et al. (2019)⁸ and Martiniano et al. (2017)⁷, where these samples were initially reported, as well as in Villalba-Mouco et al. (2021)⁹. Hence, Central European Bell Beaker would be a better proxy for the actual population that arrived in Iberia ~2,500–2,220 BCE.

4.3. IA burials

Nowadays, Spaniards and Portuguese carry higher Steppe ancestry proportions than BA Iberians meaning that other posterior migrations took place after the BA. The two males and the female previously analysed by Olalde et al. (2019)⁸ had different mitochondrial haplogroups, were unrelated and the males shared the same R1b lineage. The three individuals were relatively distinct from each other and closer to the EBA individuals than the MN individuals. These three individuals as well as the rest of previously published Spanish IA individuals generally had a higher Steppe ancestry than previous periods. However, little is

known about IA Iberia from a genetic standpoint, as the most common funerary practice was cremation¹⁷, which is not favourable for DNA preservation. Hence, it is still not known if the previously published IA individuals (n=24) are representative of the main IA Iberian population. Moreover, the IA in Portugal is totally unexplored and more samples from that area are needed to discern the genetic history of ancient Iberia.

4.4. Future perspectives

Further analyses could be performed to better explore the genetic context of MDB. Firstly, in order to increase the resolution of the patrilineal analysis, the Y chromosome haplogroups could be better assigned running a more updated software such as HaploGrouper (https://gitlab.com/bio_anth_decode/haploGrouper) and/or calculating and recording the number of derived and ancestral SNP calls as described in Rohrlach et al. (2021)⁷¹. Secondly, other kinship analyses could be further explored. For instance, separate READ runs between individuals with similar ancestry backgrounds could be performed, as it could properly estimate better normalisation values. Furthermore, new paleogenomic methods have recently been developed to increase the resolution of kinship analyses^{72,73}. Thirdly, the studied admixture events could be further tested with *qpAdm*, a statistical tool for studying the ancestry of populations with histories that involve admixture between two or more source populations⁷⁴. Fourthly, a re-exploration of some analyses including the south-eastern individuals reported by Villalba-Mouco et al. (2021)⁹ should be performed. Fifthly, the BA transition has been analysed spatially, but it would be interesting to explore it temporally as the BA is a long period (~2,200–900 BCE). Lastly, further investigations in Portugal are needed as not many individuals have been genetically analysed across time in order to provide a more comprehensive view of the genetic history of the whole Iberian Peninsula.

5. References

1. Bouso M, Esteve X, Farré J, et al. Del Bronze inicial a la depressió prelitoral catalana: Can Roqueta (Sabadell, Vallès Occidental) i Mas d'en Boixos-1 (Pacs del Penedès, Alt Penedès). *CYPSELA* 2004;15:73–101.
2. Pascual MP, Mercadé JM. Mas d'en Boixos (Pacs del Penedès, Alt Penedès). Resultats de la campanya d'intervenció arqueològica de l'any 2008. *Jornades d'Arqueologia del Penedès 2011* 2016;137–48.
3. Farré J, Mestres J, Senabre MR, Feliu JM. El jaciment de Mas d'en Boixos (Pacs del Penedès, Alt Penedès). Un espai utilitzat des del Neolític fins a l'època ibèrica.
4. Skoglund P, Malmström H, Raghavan M, et al. Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science* (1979) 2012;336(6080):466–9.
5. Allentoft ME, Sikora M, Sjögren KG, et al. Population genomics of Bronze Age Eurasia. *Nature* 2015;522(7555):167–72.
6. Valdiosera C, Günther T, Vera-Rodríguez JC, et al. 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* 2018;115(13):3428–33.
7. Martiniano R, Cassidy LM, Ó'Maoldúin R, et al. The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genet* 2017;13(7).
8. Olalde I, Mallick S, Patterson N, et al. The genomic history of the Iberian Peninsula over the past 8000 years. *Science* (1979) 2019;1234(March):1230–4.
9. Villalba-Mouco V, Oliart C, Rihuete-Herrada C, et al. Genomic transformation and social organization during the Copper Age–Bronze Age transition in southern Iberia. *Sci Adv* 2021;7:21.
10. Olalde I, Haak W, Barnes I, Lalueza-Fox C, Reich D. The Beaker Phenomenon and the Genomic Transformation of Northwest Europe. *Nature* 2018;555(7695):190–6.
11. López JB, Malgosa A, Gallart J. Cova de Montanissell (Sallent-Coll de Nargó, Alt Urgell). Operació: «Senyora de les muntanyes». *Cota Zero* 2005;20:27–36.
12. Haak W, Brandt G, Jong HN de, et al. Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. *Proc Natl Acad Sci U S A* 2008;105:18226–18231.
13. Gilbert MTP, Djurhuus D, Melchior L, et al. mtDNA from hair and nail clarifies the genetic relationship of the 15th century Qilakitsoq Inuit mummies. *Am J Phys Anthropol* 2007;133:847–853.
14. Simón M, Jordana X, Armentano N, et al. The Presence of Nuclear Families in Prehistoric Collective Burials Revisited: The Bronze Age Burial of Montanissell Cave (Spain) in the Light of aDNA. *Am J Phys Anthropol* 2011;146(3):406–13.
15. Alt KW, Zesch S, Garrido-pena R, et al. A Community in Life and Death: The Late Neolithic Megalithic Tomb at Alto de Reinoso (Burgos, Spain). *PLoS One* 2016;11(1).
16. Kuhn JMM, Jakobsson M, Günther T. Estimating genetic kin relationships in prehistoric populations. *PLoS One* 2018;13(4).

17. López-Cachero FJ. Cremation Cemeteries in the Northeastern Iberian Peninsula: Funeral Diversity and Social Transformation during the Late Bronze and Early Iron Ages. *Eur J Archaeol* 2011;14(1–2):116–32.
18. Llamas B, Willerslev E, Orlando L, Orlando L. Human evolution: a tale from ancient genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2017;372(1713).
19. Llamas B, Valverde G, Fehren-Schmitz L, Weyrich LS, Cooper A, Haak W. From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *Sci Technol Archaeol Res.* 2017;3(1):1–14.
20. Damgaard PB, Margaryan A, Schroeder H, Orlando L, Willerslev E, Allentoft ME. Improving access to endogenous DNA in ancient bones and teeth. *Sci Rep* 2015;5(11184):1–12.
21. Dabney J, Knapp M, Glocke I, et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences* 2013;110(39):15758–63.
22. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010;5(6).
23. Mathieson I, Lazaridis I, Rohland N, et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 2015;528(7583):499–503.
24. Fellows Yates JA, Lamnidis TC, Borry M, et al. Reproducible, portable, and efficient ancient genome reconstruction with nfcore/eager. *PeerJ* 2021;9:1–25.
25. Neukamm J, Peltzer A, Nieselt K. DamageProfiler: fast damage pattern calculation for ancient DNA. *Bioinformatics* 2021;37(April):3652–3.
26. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *Gigascience* 2021;10(2).
27. Gower G, Fenderson LE, Salis AT, et al. Widespread male sex bias in mammal fossil and museum collections. *Proc Natl Acad Sci U S A* 2019;116(38):19019–24.
28. Poznik GD. Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. *bioRxiv* 2016;088716.
29. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 1999;23(2):147.
30. Weissensteiner H, Pacher D, Kloss-Brandstätter A, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res* 2016;44.
31. Kloss-Brandstätter A, Pacher D, Schönherr S, et al. HaploGrep: A fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat* 2011;32(1):25–32.
32. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 2009;30(2):386–94.
33. Weissensteiner H, Forer L, Fendt L, et al. Contamination detection in sequencing studies using the mitochondrial phylogeny. *Genome Res* 2021;31(2):309–16.

34. Kuhn JMM, Jakobsson M, Günther T. Estimating genetic kin relationships in prehistoric populations. *PLoS One* 2018;13(4).
35. Bergstrom A, Mccarthy SA, Hui R, et al. Insights into human genetic variation and population history from 929 diverse genomes. 2019;1–23.
36. Lazaridis I, Nowicka U, Zhang D, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 2014;513(7518):409–13.
37. Fu Q, Posth C, Hajdinjak M, et al. The genetic history of Ice Age Europe. *Nature* 2016;534(7606):200–5.
38. Günther T, Valdiosera C, Malmström H, Ureña I, Rodriguez-varela R. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc Natl Acad Sci U S A* 2015;112(38):11917–22.
39. Lipson M, Szécsényi-Nagy A, Mallick S, et al. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* 2017;551(7680):368–72.
40. Villalba-Mouco V, van de Loosdrecht MS, Posth C, et al. Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula. *Current Biology* 2019;29(7):1169–1177.e7.
41. González-Fortes G, Tassi F, Trucchi E, et al. A western route of prehistoric human migration from Africa into the Iberian Peninsula. *Proceedings of the Royal Society B: Biological Sciences* 2019;286(1895).
42. Rivollat M, Jeong C, Schiffels S, et al. Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. *Sci Adv* 2020;6(22):eaaz5344.
43. Kılınç GM, Omrak A, Özer F, et al. The Demographic Development of the First Farmers in Anatolia. *Current Biology* 2016;26(19):2659–66.
44. Omrak A, Günther T, Valdiosera C, et al. Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. *Current Biology* 2016;26(2):270–5.
45. Hofmanová Z, Kreutzer S, Hellenthal G, et al. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc Natl Acad Sci U S A* 2016;113(25):6886–91.
46. Gokhman D, Nissim-Rafinia M, Agranat-Tamir L, et al. Differential DNA methylation of vocal and facial anatomy genes in modern humans. *Nat Commun* 2020;11(1).
47. Mathieson I, Alpaslan-Roodenberg S, Posth C, et al. The genomic history of southeastern Europe. *Nature* 2018;555(7695):197–203.
48. Narasimhan VM, Patterson N, Moorjani P, et al. The formation of human populations in South and Central Asia. *Science (1979)* 2019;365(6457).
49. Wang CC, Reinhold S, Kalmykov A, et al. Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nat Commun* 2019;10(1).
50. Järve M, Saag L, Scheib CL, et al. Shifts in the Genetic Landscape of the Western Eurasian Steppe Associated with the Beginning and End of the Scythian Dominance. *Current Biology* 2019;29(14):2430–2441.e10.
51. Patterson N, Moorjani P, Luo Y, et al. Ancient admixture in human history. *Genetics* 2012;192(3):1065–93.

52. Lacan M, Keyser C, Ricaut FX, et al. Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proc Natl Acad Sci U S A* 2011;108(45):18255–9.
53. Haak W, Lazaridis I, Patterson N, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 2015;522(7555):207–11.
54. Hervella M, Izagirre N, Alonso S, Fregel R, Alonso A. Ancient DNA from Hunter-Gatherer and Farmer Groups from Northern Spain Supports a Random Dispersion Model for the Neolithic Expansion into Europe. *PLoS One* 2012;7(4).
55. Lacan M, Keyser C, Ricaut FX, et al. Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc Natl Acad Sci U S A* 2011;108(24):9788–91.
56. Brandt G, Haak W, Adler CJ, et al. Ancient DNA Reveals Key Stages in the Formation of Central European Mitochondrial Genetic Diversity. *Science* (1979) 2013;342(6155):257–61
57. Brandt G, Szécsényi-Nagy A, Roth C, Alt KW, Haak W. Human paleogenetics of Europe - The known knowns and the known unknowns. *J Hum Evol* 2015;79:73–92.
58. Szécsényi-Nagy A, Brandt G, Haak W, et al. Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. *Proceedings of the Royal Society B: Biological Sciences* 2015;282(1805).
59. Skoglund P, Northoff BH, Shunkov M v., et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc Natl Acad Sci U S A* 2014;111(6):2229–34.
60. Gamba C, Jones ER, Teasdale MD, et al. Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun* 2014;5.
61. Olalde I, Schroeder H, Sandoval-Velasco M, et al. A common genetic origin for early farmers from mediterranean cardial and central european LBK cultures. *Mol Biol Evol* 2015;32(12):3132–42.
62. Achilli A, Rengo C, Magri C, et al. 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* 2004;75(5):910–8.
63. Loogväli EL, Roostalu U, Malyarchuk BA, et al. Disuniting uniformity: A pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol* 2004;21(11):2012–21.
64. Pereira L, Richards M, Goios A, et al. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res* 2005;15(1):19–24.
65. García O, Fregel R, Larruga JM, et al. Using mitochondrial DNA to test the hypothesis of a European post-glacial human recolonization from the Franco-Cantabrian refuge. *Heredity (Edinb)* 2011;106(1):37–45.
66. Juras A, Chyleński M, Ehler E, et al. Mitochondrial genomes reveal an east to west cline of steppe ancestry in Corded Ware populations. *Sci Rep* 2018;8(1).
67. Mittnik A, Wang CC, Pfrengle S, et al. The genetic prehistory of the Baltic Sea region. *Nat Commun* 2018;9(1).
68. Saag L, Vasilyev S v, Varul L, et al. Genetic ancestry changes in Stone to Bronze Age transition in the East European plain. 2021.

69. Lazaridis I, Mittnik A, Patterson N, et al. Genetic origins of the Minoans and Mycenaeans. *Nature* 2017;548(7666):214–8.
70. Gandini F, Achilli A, Pala M, et al. Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes. *Sci Rep* 2016;6.
71. Rohrlach AB, Papac L, Childebayeva A, et al. Using Y-chromosome capture enrichment to resolve haplogroup H2 shows new evidence for a two-path Neolithic expansion to Western Europe. *Sci Rep* 2021;11(1):1–11.
72. Fowler C, Olalde I, Cummings V, et al. A high-resolution picture of kinship practices in an Early Neolithic tomb. *Nature* 2022;601(7894):584–7.
73. Hanghøj K, Moltke I, Andersen PA, Manica A, Korneliussen TS. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience* 2019;8(5).
74. Harney É, Patterson N, Reich D, Wakeley J. Assessing the performance of qpAdm: A statistical tool for studying population admixture. *Genetics* 2021;217(4).

Part II: Pre-Columbian Mesoamerica

Chapter III: Ancient DNA Studies in Pre-Columbian Mesoamerica

Statement of Authorship

Title of Paper	Ancient DNA Studies in Pre-Columbian Mesoamerica
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	Roca-Rada, X.; Souilmi, Y.; Teixeira, J.C.; Llamas, B. Ancient DNA Studies in Pre-Columbian Mesoamerica. Genes 2020, 11, 1346. https://doi.org/10.3390/genes11111346

Principal Author

Name of Principal Author (Candidate)	Xavier Roca Rada			
Contribution to the Paper	Conceptualisation, data collection, interpretation and drafted the manuscript.			
Overall percentage (%)	80%			
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.			
Signature	<table border="1" style="width: 100%;"> <tr> <td style="width: 80%;"></td> <td style="width: 10%;">Date</td> <td style="width: 10%;">05/10/2022</td> </tr> </table>		Date	05/10/2022
	Date	05/10/2022		

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Dr Yassine Souilmi			
Contribution to the Paper	Co supervised the Mr Roca Rada, helped interpret the results, and provided input on the manuscript.			
Signature	<table border="1" style="width: 100%;"> <tr> <td style="width: 80%;"></td> <td style="width: 10%;">Date</td> <td style="width: 10%;">5/10/2022</td> </tr> </table>		Date	5/10/2022
	Date	5/10/2022		

Name of Co-Author	João Teixeira			
Contribution to the Paper	Supervision, provided technical/analytical support and critical feedback on the manuscript			
Signature	<table border="1" style="width: 100%;"> <tr> <td style="width: 80%;"></td> <td style="width: 10%;">Date</td> <td style="width: 10%;">14/10/2022</td> </tr> </table>		Date	14/10/2022
	Date	14/10/2022		

Name of Co-Author	Bastien Llamas			
Contribution to the Paper	Conceptualisation, provided technical/analytical support and critical feedback on the manuscript			
Signature	<table border="1" style="width: 100%;"> <tr> <td style="width: 80%;"></td> <td style="width: 10%;">Date</td> <td style="width: 10%;">06/10/2022</td> </tr> </table>		Date	06/10/2022
	Date	06/10/2022		



Review

Ancient DNA Studies in Pre-Columbian Mesoamerica

Xavier Roca-Rada ^{1,*} , Yassine Souilmi ^{1,2,3} , João C. Teixeira ^{1,4} and Bastien Llamas ^{1,2,3,4,*}

¹ Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia; yassine.souilmi@adelaide.edu.au (Y.S.); joao.teixeira@adelaide.edu.au (J.C.T.)

² National Centre for Indigenous Genomics, Australian National University, Canberra, ACT 0200, Australia

³ Environment Institute, University of Adelaide, Adelaide, SA 5005, Australia

⁴ Centre of Excellence for Australian Biodiversity and Heritage (CABAH), School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia

* Correspondence: xavier.rocarada@adelaide.edu.au (X.R.-R.); bastien.llamas@adelaide.edu.au (B.L.); Tel.: +61-8-8313-0262 (X.R.-R. & B.L.)

Received: 1 October 2020; Accepted: 10 November 2020; Published: 13 November 2020



Abstract: Mesoamerica is a historically and culturally defined geographic area comprising current central and south Mexico, Belize, Guatemala, El Salvador, and border regions of Honduras, western Nicaragua, and northwestern Costa Rica. The permanent settling of Mesoamerica was accompanied by the development of agriculture and pottery manufacturing (2500 BCE–150 CE), which led to the rise of several cultures connected by commerce and farming. Hence, Mesoamericans probably carried an invaluable genetic diversity partly lost during the Spanish conquest and the subsequent colonial period. Mesoamerican ancient DNA (aDNA) research has mainly focused on the study of mitochondrial DNA in the Basin of Mexico and the Yucatán Peninsula and its nearby territories, particularly during the Postclassic period (900–1519 CE). Despite limitations associated with the poor preservation of samples in tropical areas, recent methodological improvements pave the way for a deeper analysis of Mesoamerica. Here, we review how aDNA research has helped discern population dynamics patterns in the pre-Columbian Mesoamerican context, how it supports archaeological, linguistic, and anthropological conclusions, and finally, how it offers new working hypotheses.

Keywords: ancient DNA; Mesoamerica; Teotihuacan; mtDNA; Native American founding lineages; Native American genetic history; Native American ancestries

1. Introduction

Genomics, namely the study of the complete set of genetic information carried by an individual, has become a powerful tool to map the peopling of the world in detail. Ancient DNA (aDNA) adds a crucial layer of information by providing a snapshot of genetic diversity in past populations, most of which is often hardly accessible in modern-day populations due to recent demographic events. Therefore, aDNA provides evidence in addition to archaeology, linguistics, and anthropology to better comprehend human expansions and demographic events.

Within the Americas, the genetic history of the Arctic, northern North America, and South America have been recently studied using aDNA [1–4]. However, aDNA research in populations from southern North America and Central America is extremely limited, despite a rich archaeological record and Indigenous cultural diversity. This is mostly due to the highly degraded nature of genetic material found in archaeological remains [5], further exacerbated by tropical environments or burial practices such as covering the dead with cinnabar. Nevertheless, recent advances in the recovery of authentic aDNA and the application of high-throughput DNA sequencing technologies are enabling the reconstruction of human history in tropical environments, such as the Caribbean islands [6–8] and Central America [9], opening the door to a deep exploration of ancient Mesoamerica.

Mesoamerica is a historically and culturally defined geographic area comprising current central and south Mexico, Belize, Guatemala, El Salvador, and border regions of Honduras, western Nicaragua, and northwestern Costa Rica [10] (Figure 1). Mesoamerica was once home to some of the most famous pre-Columbian American cultures, including the Mayas, Aztecs, and Teotihuacanos. These past cultures developed advanced architecture (step pyramids and ball courts with rings), social organisation, commercial cooperation and trade, culture (hieroglyphic writing and a calendrical system), art (pyrite mirrors and ceramics), politics, an intensive agricultural lifestyle, and in some cases, specific rituals such as human sacrifices [11].

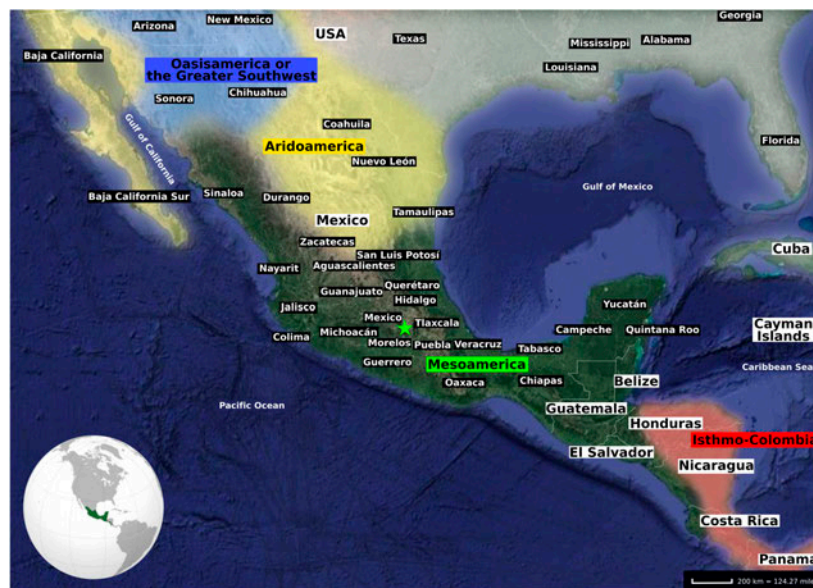


Figure 1. The historical and cultural area of Mesoamerica, as well as Isthmo-Colombia, Aridoamerica, and Oasisamerica (also known as the Greater Southwest) represented with coloured shades. Regions outside these areas are depicted with a white shade. Of note, Oasisamerica is occasionally included in Aridoamerica in the literature. White lines delineate the modern-day borders between countries and white dashed lines the borders between states. The green star is the localisation of present-day Mexico City.

Archaeological evidence indicates a continuous occupation of Mesoamerica since the end of the Late Pleistocene, specifically the Lithic period (~13000–5000 BCE) (Table 1) [11]. However, recent studies provided controversial evidence of human presence in the central-northern region of Mexico, relatively close to Mesoamerica, during the Last Glacial Maximum (26,500–19,000 years ago) [12,13]. Notwithstanding, the first human populations inhabiting this region were nomadic hunter-gatherers. This subsistence strategy lasted until the Archaic (or Proto-Neolithic) period (5000–2500 BCE) when human societies started to transition into semi-sedentary foraging and farming. This lifestyle change was essential for the subsequent development of complex societal structure [11,14], even though the process was not uniform throughout Mesoamerica and differed across time periods and geographical regions.

The permanent settling of Mesoamerica took place in the Preclassic (or Formative) period (2500 BCE–150 CE), with the development of agriculture and pottery manufacturing. Agricultural practices intensified and improved during the Classic and Epiclassic periods (150–900 CE). The number of large population centres rapidly increased and cultural differences incremented between regions. During the

Postclassic period (900–1519 CE), Mesoamerica was inhabited by large intermingled cultural groups connected by trade and agriculture, which fostered migrations that intensified admixture and built up the diversity patterns found across Mesoamerica. The Postclassic period ended with the start of the Spanish conquest (1519–1521 CE) and the subsequent colonial period (1521–1821 CE) [11,15]. The Spanish occupation had extreme cultural and demographic impacts: high mortality rate among Indigenous people, the extinction of languages and cultures, several massive population displacements, and admixture between Indigenous and European, African, and Asian populations [16,17].

Table 1. Mesoamerican chronology: BCE (Before Common Era); CE (Common Era).

Period	Calendar Dates
Lithic	~13000–5000 BCE
Archaic or Proto-Neolithic	5000–2500 BCE
Preclassic or Formative	2500 BCE–150 CE
Classic	150–650 CE
Epiclassic *	650–900 CE
Postclassic	900–1519 † CE
Colonial	1519 †–1821 ‡ CE
Present	1821 ‡ CE–Current Day

* The Epiclassic is also known as Late Classic in parts of Mesoamerica. † Start of the Spanish conquest of Mexico.

‡ Mexican Independence Date.

2. Mesoamerican Genetic Studies

Mesoamerica was home to different human cultures connected by commerce and farming, and whose varied genetic backgrounds created an invaluable source of genetic diversity in the continent. Previous studies have attempted to uncover the peopling process of Mesoamerica using genetic approaches on modern-day Indigenous and cosmopolitan populations, mostly from Mexico and Guatemala. These analyses mainly focused on studying the so-called uniparental markers: the mitochondrial DNA (mtDNA) [17–27] and the Y-chromosome [17,19,28–30], which allow for demographic inference of maternal and paternal ancestries, respectively. Other studies further analysed single-nucleotide polymorphisms (SNPs—variable genomic positions that are present at a frequency $\geq 1\%$ in the population) located on the autosomes (non-sex chromosomes) [31,32], or sequenced complete exomes (genomic regions that encode proteins) in order to describe Mexican genetic variation, population history, adaptation to the environment, and potential implications for biomedical research [33]. Importantly, this body of work contributes valuable data to describe the global human genetic diversity, especially since Indigenous populations from the Americas are among the most underrepresented populations in human genomic studies [34–36].

Most of these genetic studies proposed a single origin for Mesoamerican populations [27–29,33], instead of a dual origin as previously argued by Mizuno et al. (2014) [26]. Moreover, Sandoval and colleagues [22,29] analysed the Indigenous population substructure. They demonstrated that the patterns observed by analysing uniparental markers do not display significant population stratification in relation to linguistics, implying that genetic divergence preceded linguistic diversification. It was also shown that modern-day Mesoamerican-related populations contain higher levels of genetic diversity and lower levels of autozygosity (DNA segments identical by descent) compared to modern populations related to other cultural areas, such as Aridoamerica (Figure 1) [25,27,29,32]. These observations agree with the hypothesised importance of agriculture and trade in pre-Columbian Mesoamerica compared to the Aridoamerican hunter-gatherer lifestyle from the northern regions, to foster population movements and admixture that increase genome-wide levels of genetic diversity.

3. Ancient DNA Studies

Inferring the population history of Mesoamerica through genetic studies that solely incorporate modern-day populations can be challenging due to recent historical events, such as population

admixture with European, African, or Asian populations, that can distort pre-colonisation genetic signals. For instance, present-day Mexicans have, in general, three different ancestries: Native American, European (closely related to Iberians), and African. While the African component is generally minimal, the European component has a much lower proportion in Native Mexicans than what is found in cosmopolitan Mexicans. Notably, the Native American ancestry component can be divided into six separate ancestries [32]. Three of these are restricted to isolated populations of the northwest, southeast, and southwest, while the other three are widely geographically distributed [32]. Hence, it is essential to include pre-colonisation samples from the archaeological record to contextualise the true genetic diversity and population substructure of ancient Mesoamerica and identify genetic ancestry that might have been lost since the arrival of the Spanish.

In spite of the extensive archaeological, linguistic, and anthropological research on pre-Columbian Mesoamerica, there have been surprisingly few aDNA studies. Therefore, there is a current lack of genetic knowledge about the genetic history of the region, probably because of the highly degraded nature of genetic material from tropical environments. Thus, currently available ancient genetic data rely mostly on mtDNA studies. The human mtDNA is a circular double-stranded DNA molecule that is maternally inherited, does not recombine [37,38], and has a higher mutation rate compared to the nuclear genome [39]. The sizes of the human reference mtDNA sequences rCRS [40] and RSRS [41] are small at 16,569 base pairs. These characteristics make mtDNA a suitable tool for maternal phylogenetic and phylogeographic inferences [42]. In addition, the high copy number of mtDNA in cells increases the odds of successfully obtaining ancient mtDNA in archaeological remains [43,44]. Mitochondrial haplogroups are lineages defined by shared genetic variation. They have been used to represent the major branches along the mitochondrial phylogenetic tree, unravelling the evolutionary paths of maternal lineages back to human origins in Africa and the subsequent peopling of the world. The Indigenous haplogroups found in present-day non-Arctic populations from the Americas are A2, B2, C1, D1, and D4h3a. All of them descend from a founding population that split from Northern Eurasian lineages ~25,000 years ago and remained stranded in Beringia for ~6000 years before starting to colonise the Americas ~16,000 years ago [45–47]. However, there are limitations associated with the study of mtDNA. Indeed, mtDNA represents the evolutionary history of the female population at a single locus and sample sizes are often small. Therefore, it is necessary to remain cautious when interpreting mtDNA results mainly when there is disagreement with those obtained from genome-wide analysis. Nevertheless, most pre-Columbian Mesoamerican genetic data (Tables 2 and 3) originate from general diagnosis of mtDNA haplogroup variants in the control and/or coding regions. This early technique tested restriction fragment length polymorphisms (RFLPs) which, in this case, result from the presence or absence of mitochondrial sites cleaved by restriction enzymes in order to characterise the mitochondrial haplogroups. Although RFLPs were reported in earlier studies, the partial or complete sequencing of the mitochondrial Hypervariable Regions 1 and 2 (HVR1 and HVR2, respectively) rapidly became the most common strategy to study mtDNA history. Only a few studies in Mesoamerica have been recently able to reconstruct complete ancient mitogenomes [9,48,49] and only one has also generated and analysed genomic data [3].

Table 2. Ancient DNA studies in the Basin of Mexico.

Site	n	mtDNA Sequence	Genetic Sex Assign	Nuclear Data	mtDNA Haplogroup Frequency (%)				Date	Period	Location	Reference
					A	B	C	D				
Teotihuacan (Moon Pyramid)	1	Imputed WMG	No	No	100	0	0	0	~500 CE	Classic	State of Mexico, Mexico	Mizuno et al. (2014) [26]
Teotihuacan (Tlalotlacan)	8	GD and HVRI	No	No	50	0	0	50	300–500 CE	Classic	State of Mexico, Mexico	Herrera Salazar (2007) [50]
Teotihuacan (La Ventilla, San Sebastián Xolalpan and San Francisco Mazapa)	36	GD	No	No	58	25	14	3	300–700 CE	Classic–Epiclassic	State of Mexico, Mexico	Aguirre-Samudio et al. (2017) [51]
Teotihuacan (Teopanazco)	29	HVR1 and HRM	Yes	No	55	21	17	7	200–550 CE	Classic	State of Mexico, Mexico	Álvarez-Sandoval et al. (2014) and (2015) [52,53]
Xaltocan (Pre-Aztec conquest)	10	GD and HVR1	No	No	30	30	0	40	1240–1541 CE	Postclassic	Mexico City, Mexico	Mata-Míguez et al. (2012) [54]
Xaltocan (Post-Aztec conquest)	15				60	20	7	13			Mexico City, Mexico	
Tlatelolco	23	GD	No	No	65	13	4	18	1325 CE		Mexico City, Mexico	Kemp et al. (2005) [55]
	30	GD and HVR1	No	No	46	37	7	10	1350–1457 CE	Postclassic	Mexico City, Mexico	Solorzano Navarro (2006) [56]
	14	GD	Yes	No	57	21	7	14	1454–1457 CE		Mexico City, Mexico	De La Cruz et al. (2008) [57]
	11	HVR1, HVR2, and WMG *	Yes	Yes *	55	18	9	18	1350–1519 CE		Mexico City, Mexico	Morales-Arce et al. (2019) [48]
Cholula	12				42	42	16	0	240–1400 CE	Classic–Postclassic	State of Puebla, Mexico	
	9	GD and HVRI	No	No	100	0	0	0	1100–1500 CE	Postclassic	State of Puebla, Mexico	Juárez Martín (2002) [58]

HVR1—Hypervariable Region 1 (or a segment); HVR2—Hypervariable Region 2 (or a segment); GD—general diagnostics of mtDNA haplogroup variants in the control and/or coding regions (restriction fragment length polymorphism (RFLP) analyses, sometimes followed by Sanger sequencing); HRM—haplogroup characterisation by high resolution melting analysis; WMG—whole mitogenome. * Some individuals.

Table 3. Ancient DNA studies in the Yucatán Peninsula, Maya Civilisation, and Greater Nicoya.

Region	Population/ Site	n	mtDNA Sequence	Genetic Sex Assign	Nuclear Data	mtDNA Haplogroup Frequency (%)				Date	Period	Location	Reference
						A	B	C	D				
Yucatán Peninsula	Hoyo Negro	1	GD	No	No	0	0	0	100	~10500 BCE	Lithic	State of Quintana Roo, Mexico	Chatters et al. (2014) [59]
	Saki Tzul	2				0	0	0	100	~5300 BCE			
	Mayahak Cab Pek	1	WMG	Yes	Yes	0	0	0	100	~7300 BCE		Toledo District, Belize	Posth et al. (2018) [3]
Maya Civilisation	Mayas	38	HVR1	No	No	61	0	34	5	250–1500 CE	Classic–Postclassic	States of Quintana Roo, Yucatán, Chiapas and Tabasco, Mexico	Ochoa-Lugo et al. (2016) [60]
	Midnight Terror Cave	17	WMG	Yes	No	82	12	6	0	550–900 CE	Classic–Epiclassic *	Cayo District, Belize	Verdugo et al. (2020) [49]
	Xcaret	24	GD	No	No	88	4	8	0	600–1521 CE	Epiclassic *–Postclassic	State of Quintana Roo, Mexico	González-Oliver et al. (2001) [61]
	Copán	9	GD	No	No	0	0	88	12	700–1300 CE		Copán Department, Honduras	Merriwether, Reed, and Ferrell (1997) [62]
Greater Nicoya	Jicaro	3	WMG	No	No	0	100	0	0	800–1250 CE	Epiclassic *–Postclassic	Guanacaste Province, Costa Rica	Morales-Arce et al. (2017) [9]

HVR1—Hypervariable Region 1 (or a segment); GD—general diagnostics of mtDNA haplogroup variants in the control and/or coding regions (restriction fragment length polymorphism (RFLP) analyses, sometimes followed by Sanger sequencing); WMG—whole mitogenome. * The Epiclassic is also known as Late Classic in parts of Mesoamerica.

Interestingly, existing aDNA data come from the Central and Southeast areas of Mesoamerica, corresponding to the Basin of Mexico and the Yucatán Peninsula and its nearby territories, respectively. Therefore this review will focus on the research that has been performed in these two geographical areas. Nonetheless, it should be noted that Mesoamerica contains other regions with extensive archaeological and anthropological records that have not been investigated in aDNA studies yet. These regions include the Gulf Coast and Oaxaca. The Gulf Coast was initially peopled by the Olmecs (the earliest known major Mesoamerican civilisation) and was later home to the city of El Tajín. Oaxaca comprises the current Mexican state of Oaxaca as well as the border regions of Guerrero, Puebla, and Veracruz. It was home to the Mixtec and Zapotec civilisations, the Preclassic–Classic city of Monte Albán being the most important archaeological site. In contrast, the northern and western regions of Mesoamerica are largely understudied from an archaeological and anthropological perspective and have been severely looted. The North was the transition area between Mesoamerica and Aridoamerica inhabited by both hunter-gatherers and farmers and it was nearly abandoned during the Postclassic period. The West was a geographically diverse area that led to the rise of highly diverse societies characterised by the shaft tomb tradition in the Preclassic and Classic periods and the Tarascan state during the Postclassic [11].

4. Basin of Mexico

The Basin of Mexico is a highland plateau that presently includes Mexico City and parts of the states of Mexico, Hidalgo, Tlaxcala, Morelos, and Puebla. The Basin of Mexico previously contained five lakes drained by Spanish colonists in order to expand Mexico City and prevent flooding (Figure 2). This area used to be the homeland of several Mesoamerican civilisations, including the Classic city of Teotihuacan, the Epiclassic city of Cholula, and the Postclassic Toltec, Tepanec, and Aztec Empires [11]. Some aDNA studies have focused on this region, including individuals from the Postclassic period carrying haplogroups A and B at higher frequencies (Table 2) [48,50–58,63].

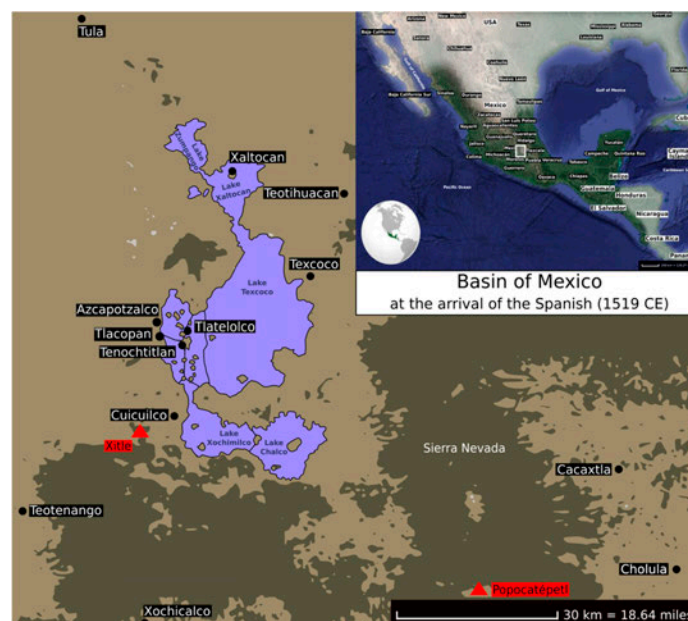


Figure 2. Approximate locations of the archaeological sites (black dots) and volcanoes (red triangles) in the Basin of Mexico at the arrival of the Spanish (1519 CE), as discussed in the present review.

4.1. Preclassic Period

Mesoamerica's settling took place during the Early Preclassic period with the emergence of agrarian villages with egalitarian societies. The Middle Preclassic was defined by the development of agriculture, the increase in population sizes in settlements with hierarchical societies, and a strong influence from the Gulf Coast. This influence later disappeared during the Late Preclassic, when some settlements (e.g., Cuicuilco) became capital towns with strong economic and political power. At the end of this period, Teotihuacan rose as the undisputed power in Mesoamerica [11].

4.2. Classic Period

Teotihuacan was a powerful political and military centre that controlled trade routes and attracted large masses of people from distant regions, coordinating them into an effective multi-ethnic system. The metropolis grew into the largest city of Mesoamerica and one of the world's largest and most urbanised population centres of the first millennium, reaching 80,000–100,000 inhabitants by approximately 150 CE [64]. Teotihuacan's art and architecture styles are widespread in Mesoamerica, suggesting that the city had extensive influence and acted as the principal trade metropolis in the region. Obsidian tools were the main export (along with ceramics and other luxury goods) due to the presence close to the city of the most important deposit of obsidian in Mesoamerica [11].

Teotihuacan was first occupied ~500 BCE and consisted of villages of local people as the valley offered several advantages: an abundance of freshwater springs, volcanic and raw materials for construction and technology, and a crucial location along the route from the Gulf Coast to the Basin of Mexico, skirting the Trans-Mexican Volcanic Belt also known as Sierra Nevada [64].

During the first century CE, Teotihuacan was the destination of numerous migrations from Puebla and Tlaxcala, probably due to the Popocatepetl volcano eruption [65]. It has been proposed that the metropolis was either controlled by powerful rulers and military elites [66,67] or coordinated as a corporate co-rulership system, where each one of the four seats of the state was occupied by one of the four administrative sectors of the city [64].

Around 350 CE, a new episode of population displacement from Cuicuilco and surrounding regions likely took place when the Xitle volcano erupted [68]. This migration coupled with a possible rivalry between two of the seats of the state may have generated a political crisis that climaxed with the destruction and burning of the Temple of the Feathered Serpent [64] located in the Avenue of the Dead, the city's main avenue. The Moon Pyramid is located on the same avenue and underwent a reconstruction approximately during this period. Interestingly, ancient DNA was extracted from a human archaeological sample excavated from the Moon Pyramid. Because of poor DNA preservation, and despite using advanced high-throughput sequencing techniques, investigators had to rely on imputation methods to assign the mitochondrial subhaplogroup A2 to the individual. A major limitation of this study is the unavailability of a panel from a population both closely related to and contemporaneous with the Moon Pyramid individual to perm imputations with the highest level of confidence [63].

Following the political crisis, Teotihuacan became a multi-ethnic city inhabited by locals, returnee migrants, people from the Tlaxcala–Hidalgo–Puebla corridor, and individuals from distant regions, such as the Gulf Coast, Oaxaca, and Chiapas. This pattern has been sustained by archaeological studies [64], trace element analysis [69], stable isotopes [70], and strontium isotopes [71]. The city's internal structure consisted of various neighbourhoods, each one specialised in producing specific goods and supervised by the intermediate elites whose power rapidly increased, changing Teotihuacan's political organisation [64]. On the one hand, most neighbourhoods were segregated by the province of origin of its inhabitants, such as Tlailotlacan, home to migrants with an Oaxaca origin [64]. Genetic analyses on eight individuals from Tlailotlacan revealed only the mtDNA haplogroups A and D [50]. On the other hand, some neighbourhoods were multi-ethnic settlements. Aguirre-Samudio and colleagues [51] examined 36 individuals from three different multi-ethnic neighbourhoods (San Sebastián Xolalpan, La Ventilla, and San Francisco Mazapa). They found that the most frequent mtDNA haplogroup

was A (58%), followed by haplogroups B (25%), C (14%), and D (3%). Furthermore, the genetic diversity of the analysed mitochondrial sequences was higher compared to the Tlailotlacan individuals, which supports the multiethnicity hypothesis. Additionally, two studies by Álvarez-Sandoval and colleagues [52,53] also supported this hypothesis using 29 individuals from Teopanazco, another multi-ethnic neighbourhood [64]. They also found the four autochthonous mtDNA haplogroups (A, B, C, and D), but reported no statistically significant differences between haplogroup frequencies across periods. Although the levels of genetic diversity between periods were not significant either, they suggested that the population from Teopanazco was heterogeneous since the beginning of its settlement and that diversity increased over time [52,53,64].

Around 550 CE, Teotihuacan's population size declined by half, and the city collapsed without any clear explanation. Although some of the earliest theories postulated that Teotihuacan was sacked by outside invaders [66], the most recent archaeological studies have proposed alternative explanations. Manzanilla [64] suggests that disagreements between the corporate structure of the state and the exclusionary structure of the strong intermediate elites eventually reached a breaking point that tore the city's social and political organisation apart. It has also been proposed that Teotihuacan's influence in Mesoamerica began to weaken, as many of the conquered cities became increasingly autonomous, which could have caused an economic crisis and led to a revolt [72]. The archaeological evidence for both of these theories is the deliberate burning of major buildings and the destruction of artworks and religious sculptures [64]. However, the specific destruction of effigies of Tlaloc (the God of the Rain) [73,74] has been argued to instead represent the abandonment feeling that Teotihuacanos might have had towards their God. It has been suggested that during the end of the Classic period, a severe drought caused population loss and abandonment of large urban centres and villages throughout all Mesoamerica [11,74,75]. Teotihuacan utilised freshwater springs for irrigation and domestic consumption, whereby a severe drought would have affected the agriculturalist lifestyle, preventing the growth of staple food crops, such as maize, beans, avocados, peppers, and squash, subsequently affecting the breeding of farm animals such as chickens and turkeys [64,69]. This situation might have caused famine and probably the spread of epidemic diseases, such as haemorrhagic fevers, as hypothesised by Acuna-Soto and colleagues [75] after demonstrating a synchronicity of climatological and demographic events during the Classic period and the 16th century in Mesoamerica [76]. Lastly, it has also been proposed that the agriculturalist lifestyle was instead affected by a possible depletion of cultivated soils and the overexploitation of nearby wooded areas [11]. Insofar, the fall of Teotihuacan and the collapse of the Classic period in Mesoamerica remain a mystery.

4.3. Epiclassic Period

The Epiclassic, also known as Late Classic in other parts of Mesoamerica, was marked by the decadence of Teotihuacan and extensive population movements. Various ephemeral centres of power emerged throughout the Basin of Mexico, such as Xochicalco, Teotenango, Cacaxtla, and Cholula [11]. The latter was a sacred city in the Puebla–Tlaxcala Valley that had been continuously occupied since ~100 BCE and was heavily affected by the Spanish conquest and colonisation. After Teotihuacan's collapse, the population is thought to have migrated to Cholula, with which they were culturally associated since the Classic period [77]. Even though Cholula gained power in the Basin of Mexico during the Epiclassic period, it was highly influenced by the Gulf Coast culture until the Postclassic period [11]. Morales-Arce and colleagues (2019) [48] analysed 12 individuals from Cholula, dating from the Classic to the Middle Postclassic to assess the population dynamics around this time. The study determined that haplogroups A and B were equally the most frequent (42%), haplogroup C was the rarest (16%), and haplogroup D was absent. Moreover, a previous study by Juárez Martín [58] analysed nine individuals from the Postclassic that all carried haplogroup A. Interestingly, the genetic distance between the Cholula [48] and Teopanazco individuals [53] was higher than any other Central Mesoamerica population, which challenges the hypothesised migration from Teotihuacan to Cholula based on the archaeological record.

4.4. Postclassic Period

The study of the Postclassic includes not only archaeology and anthropology, but also linguistic information from accounts written in Spanish and different Indigenous languages. While the Classic period was a peaceful epoch of cultural climax, the Postclassic period was marked by political instability and war. This period emerges with the apogee of Tula (also known as Tollan-Xicocotitlan), the capital of the Toltec Empire. Even though the political system and area of influence of the Toltec culture remain a point of discussion among scholars, especially in the city of Chichen Itza, inhabited by the Mayas in the Yucatán Peninsula, it has been proposed that its influence was widespread and persisted through time in central Mesoamerica [11].

After the collapse of the Toltec Empire, numerous city-states emerged and fought one another for the control of political power. Even though Azcapotzalco (capital of the Tepanec state) initially ruled most of the Basin of Mexico, the Triple Alliance of the city-states of Tenochtitlan, Texcoco, and Tlacopan reduced Tepanec influence and allowed the rise of the Aztec Empire. The Aztecs conquered a large number of cities across Mesoamerica over the following years and established a large Empire that lasted until the Spanish conquest [78]. For example, the city of Xaltocan, founded by Otomi-speaking people around 1100 CE, was initially conquered by the Tepanec state and later integrated into the Aztec Empire [79]. Colonial documents clearly state that the Tepanec conquest ended with the fleeing of the original Otomi population and that the Aztec conquest led to population replacements from various places in central Mesoamerica and the Gulf Coast to Xaltocan [80]. Nevertheless, archaeological evidence suggests a population continuity after the first defeat and even during the incorporation into the Aztec empire [81]. A single study has so far focused on the genetic impact of Aztec Imperialism in the archaeological site of Xaltocan. Mata-Míguez and colleagues [54] analysed mtDNA haplogroups in 10 pre- and 15 post-Aztec conquest individuals and found that haplogroups A2 and C1 increased over time (from 30% to 60% and from 0% to ~7%, respectively), whereas haplogroups B2 and D1 decreased over time (from 30% to 20% and from 40% to ~13%, respectively). The study evidenced that some haplotypes were found in multiple individuals from the same time period, indicating close maternal relationships. However, the distinct haplotypes found across time periods are unlikely to have resulted from genetic drift or the occurrence of new mutations, given the short evolutionary time across samples. Instead, results are more consistent with a replacement of maternal lines in the sampled Xaltocan individuals after the Aztec conquest.

In contrast, various studies have focused on the archaeological site of Tlatelolco, which is buried beneath modern-day Mexico City and was founded around 1337 CE. Even though Tlatelolco was a sister city to Tenochtitlan (the capital of the Aztec Empire), religious and cultural practices were distinct between the two cities. Tlatelolco was the biggest and most diversified marketplace in Mesoamerica during the Late Postclassic [11]. The most recent aDNA study was performed by Morales-Arce et al. (2019) [48] on 11 subadults found in a sacrificial ritual context. Results show that 55% of the individuals harboured the haplogroup A, 18% both the haplogroups B and D, and 9% the haplogroup C. Similar frequencies were reported in three previous studies conducted by Kemp et al. (2005) [55] with 23 adults from a burial at the ceremonial centre of Tlatelolco, Solorzano Navarro (2006) [56] with 30 individuals from the same burial site, and De La Cruz et al. (2008) [57] with 14 subadults from a sacrificial context dedicated to Tlaloc, the God of the Rain.

5. Yucatán Peninsula

The Yucatán Peninsula is located in southeastern Mexico and separates the Caribbean Sea from the Gulf of Mexico. Currently, it comprises the Mexican states of Yucatán, Campeche, and Quintana Roo, as well as part of Guatemala and almost the whole of Belize. The first humans inhabiting this area were hunter-gatherers. In fact, the oldest individual (~10500 BCE) with published genetic data in Mesoamerica comes from this region. This individual was found in Hoyo Negro (Figure 3 and Table 3), a submerged collapsed chamber in the Sac Actun cave system, and harboured the autochthonous mitochondrial haplogroup D1 [59,82]. However, this result has been highly debated as contamination

was observed and post-mortem damage patterns (a characteristic property of aDNA) were not evident in the mitochondrial sequences belonging to the haplogroup D1 [83].



Figure 3. Locations of the archaeological sites in the Yucatán Peninsula and nearby territories, as discussed in the present review. Coloured dots represent the archaeological sites included in each aDNA study. Black dots represent archaeological sites mentioned in the present review with no aDNA data so far. The historical and cultural areas of the Maya civilisation and the Greater Nicoya are within the thick white dashed lines. The shaded area represents regions outside of Mesoamerica. White lines delineate current borders between countries, and thin white dashed lines borders between states.

Haplogroup D1 was actually found in two genetically related individuals (~5300 BCE) from the archaeological site of Saki Tzul in the Maya Mountains (Belize) [3]. Another individual (~7300 BCE) from the same region, found in the archaeological site of Mayahak Cab Pek (Belize), carried the native haplogroup D4h3a [3]. The genomic data of these individuals from the Maya Mountains revealed that they are genetically closer to present-day groups elsewhere in Central and South America rather than from the region near Belize [3]. This result suggests that they formed part of a homogenous population associated with the Clovis culture [84] that expanded southwards carrying the primary source of the ancestry of Central and South Americans.

6. Maya Civilisation

The entire Yucatán Peninsula together with the Mexican states of Tabasco and Chiapas, the rest of Guatemala and Belize, as well as the border regions of Honduras and El Salvador, were the home to the Maya civilisation (Figure 3), one of the most advanced and highly developed societies in ancient Mesoamerica [11]. Most aDNA studies performed in this region include individuals from the Postclassic period and the mtDNA haplogroup A is the most frequent (Table 3).

The geographical diversity of the Yucatán Peninsula was one of the most important traits for the Mayan settling in the Preclassic period, allowing the emergence of different cultural focuses that led to the rise of various linguistic families. Despite the existence of a cultural influence from the Gulf Coast during the Middle Preclassic, it disappeared in the Late Preclassic when cities started to urbanise with architectural gigantism [11]. The Classic period in the Maya civilization was defined by the development of the Mayan tradition and the influence from Teotihuacan. For instance, archaeological research has demonstrated how Teotihuacan decisively intervened in Tikal (Guatemala), imposing an ally government [85]. Tikal and Copán (Honduras) were two of the most powerful cities in the Mayan Classic period. During the Late Classic (equivalent to the Epiclassic in the Basin of Mexico), several Mayan cities, especially in the south, collapsed after the fall of Teotihuacan. Despite this circumstance, Teotihuacan's culture lingered in Mayan civilisation. The Classic collapse caused various population displacements that led to huge demographic events in other cities, most of them in the north, such as Chichen Itza. This population relocation enabled the splendour of the Mayan economy, politics, and culture. The apogee of the northern Mayas climaxed in the Early Postclassic. However, this period was shaped by the arrival of invaders, probably from the Toltec Empire. In the northern Mayan area, the Late Postclassic was defined by the fall of Chichen Itza and the rise of Mayapan, whereas in the southern Mayan area, the Late Postclassic was characterised by the brief power of the K'iche' (Quiché), Q'umarcaaj being the most famous capital [11]. This period ended with the Spanish conquest and the onset of colonialism.

Regarding genetic studies of the region, Ochoa-Lugo and colleagues [60] performed a transitional aDNA study analysing 38 Mayan individuals from the Classic until the Postclassic periods across different archaeological sites in Mexico. Haplogroups A and C were the most frequent (61% and 34%, respectively), while haplogroup D was the least frequent (5%). Importantly, the complete absence of haplogroup B in this study, along with its low frequency amongst contemporary Mayan populations found in previous work, seemingly supports the hypothesis that haplogroup B independently entered the Americas through a later, separate migration into the continent [18,60,86,87]. However, this is disputed by genetic evidence showing that all Indigenous mtDNA haplogroups found in present-day Mesoamerica descend from a single founding population [45–47].

Additionally, other aDNA studies just focused on the analysis of mtDNA haplogroups in one archaeological site during a specific period of time. For example, the largest mtDNA collection from Mesoamerica to date comes from the Classic–Epiclassic site of Midnight Terror Cave (Belize), which is believed to have been a site of human sacrifice [49]. A total of 17 complete mitogenomes were reconstructed, including haplogroups A2 (82%), B2 (12%), C1 (6%), but not D. Moreover, a total of 24 individuals were genetically sex assigned and more than 60% were female, which raises questions about the hypothesis that young males often captured in battle by the Mayans were the most common sacrificial victims. Nonetheless, De La Cruz and colleagues [57] analysed 14 subadults in an Aztec sacrificial context in the Postclassic Tlatelolco (Basin of Mexico) and molecular sex identification revealed that among nine individuals analysed, eight were male and one was a female.

Finally, a further study focused on the Epiclassic–Postclassic site of Xcaret (Mexico), one of the most active ports in the eastern Mayan coast [61]. This research included 24 individuals, most of which carried haplogroup A (88%). While haplogroups B and C were found at low frequencies, no haplogroup D was found. These individuals were more related to modern-day Mayans and other contemporary populations of Mesoamerican origin than to the nine Epiclassic–Postclassic individuals from Copán (Honduras), from whom only haplogroups C and D have been reported [62].

7. Greater Nicoya

The Greater Nicoya was an ancient Central American cultural area currently located in the Pacific Nicaragua and northwestern Costa Rica. It was situated between southern Mesoamerica and northern Isthmo-Colombia (Figure 3). This region was first occupied by Chibchan-speaking groups that shared a cultural and linguistic background with ancient and current populations from Isthmo-Colombia [88]. However, it has been postulated that Mesoamerican populations from the Basin of Mexico migrated southwards and settled in this region in the Epiclassic period (around 800 CE), displacing the previous Isthmo-Colombian populations [11]. The Greater Nicoyan culture climaxed in the Postclassic period and received a strong influence from Mesoamerica, evidenced in the archaeological record through ceramic iconography and religious iconology [88]. At the time of the Spanish arrival, it was documented that the populations from the Greater Nicoya spoke Mesoamerican-related languages [89]. In particular, Morales-Arce and colleagues [9] reconstructed three complete ancient mitogenomes from the Epiclassic-Postclassic archaeological site of Jícaro (Figure 3 and Table 3). Although all of them belonged to the same mitochondrial subhaplogroup (B2d), each individual had a distinct sequence, indicating these individuals were not maternally related in the recent past. The study also suggested that these individuals had a greater degree of maternal genetic affinity with contemporary Isthmo-Colombians, than with contemporary Mesoamericans.

8. Future Perspectives

Recent improvements in the field of aDNA have revolutionised the study of the history of past populations, providing additional information highly valuable for archaeologists, linguists, and anthropologists. Given that Mesoamerica is the geographical link between the North and South American continents, aDNA studies from this region may be essential to understand the migration patterns of populations across the region, as well as their territorial expansions and interactions. For example, the aDNA study performed by Moreno-Mayar and colleagues [2] described a putative northward expansion out of Mesoamerica around 8700 years ago and a second one southward that contributed to the ancestry of most South American groups, except Patagonians. However, this study only included archaeological remains from northern North America and South America, and no Mesoamerican records were considered. Moreover, a recent study proposed a putative contact between Central Americans and Oceanian populations (~1200 CE) using modern DNA and some ancient individuals across the Americas (including the Saki Tzul individuals from the Maya Mountains) [90]. Integrating additional ancient genomes through time and space in Mesoamerica could help to clarify these hypotheses.

Archaeological evidence strongly suggests that after the settling of Mesoamerica, most cultures remained connected by trade, agriculture, culture, politics, empires, and conquests. Therefore, it is essential to include genetic analysis to examine the patterns of population structure, biological adaptation, cultural impacts, and natural phenomena (e.g., volcanos) on the evolutionary and demographic history of Mesoamerican populations. Crucially, aDNA can help to discern these patterns in pre-Columbian Mesoamerica and provide further support to archaeological studies as well as offer new insights and working hypotheses. This is particularly important for the highly understudied Preclassic period, which could benefit from interdisciplinary approaches that help to unravel the demographic and genetic architecture of the settling process of Mesoamerica.

Regarding the Basin of Mexico, a detailed study focused on samples from the Classic city of Teotihuacan, concretely in multi-ethnic neighbourhoods such as Teopanaczo, could not only help to determine aspects of population demography of Teotihuacan, but also be useful to infer and contextualize the population history of Mesoamerica during the Classic period. This should be aided by similar studies focusing on the understudied Epiclassic and the Postclassic periods, which could help to explain how people were connected after the fall of Teotihuacan and during the empires that emerged before the Spanish conquest, respectively.

Although Mayan culture remained in the Yucatán peninsula throughout all these cultural changes, it was never isolated and was also affected by the Classic collapse. A genetic study focusing on the Mayas could further expand our knowledge of this important civilisation and reveal connections with other regions of Mesoamerica. In fact, a knowledge gap currently exists in several Mesoamerican areas that could help to define population interactions in the region. These include the Gulf Coast, the North, the West, and Oaxaca regions, as well as the cultural and historical areas of Isthmo-Colombia, Aridoamerica, and Oasisamerica.

9. Conclusions

Mesoamerica represents a critical tropical transition in the southward journey of the First Peoples of the Americas. A thorough description of genetic variation in ancient Mesoamerican populations through space and time may enable us to identify mechanisms of adaptation to changing environments, climates, lifestyles, and social structures, with potential consequence for present-day people's health. For instance, this kind of research could target the Mesoamerican Classic collapse and investigate a possible genetic link to severe droughts or epidemic diseases, and potential consequences to the genetic make-up of contemporary Mesoamerican populations. Thus, it is highly important to reconstruct the genetic past of Mesoamerica before and after European arrival and estimate the genetic impact of European colonisation in the demographic and adaptive history of these populations, which currently remains unknown.

Author Contributions: Conceptualisation: X.R.-R. and B.L. Original draft preparation: X.R.-R. Review and editing: Y.S., J.C.T. and B.L. All authors have read and agreed to the published version of the manuscript.

Funding: X.R.-R is supported by a Molecular Anthropology PhD scholarship from the University of Adelaide. Y.S. is supported by an Australian Research Council Discovery Project (ARC DP190103705). J.C.T. is supported by an ARC Indigenous Discovery Grant (IN180100017). B.L. is supported by an ARC Future Fellowship (FT170100448).

Acknowledgments: We thank Linda R. Manzanilla and two anonymous reviewers for their constructive feedback on this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Scheib, C.L.; Li, H.; Desai, T.; Link, V.; Kendall, C.; Dewar, G.; Griffith, P.W.; Mörseburg, A.; Johnson, J.R.; Potter, A.; et al. Ancient human parallel lineages within North America contributed to a coastal expansion. *Science* **2018**, *360*, 1024–1027. [[CrossRef](#)]
2. Moreno-Mayar, J.V.; Vinner, L.; de Damgaard, P.B.; de la Fuente, C.; Chan, J.; Spence, J.P.; Allentoft, M.E.; Vimala, T.; Racimo, F.; Pinotti, T.; et al. Early human dispersals within the Americas. *Science* **2018**, *362*. [[CrossRef](#)]
3. Posth, C.; Nakatsuka, N.; Lazaridis, I.; Skoglund, P.; Mallick, S.; Lamnidis, T.C.; Rohland, N.; Nägele, K.; Adamski, N.; Bertolini, E.; et al. Reconstructing the deep population history of central and South America. *Cell* **2018**, *175*, 1185–1197. [[CrossRef](#)]
4. Sikora, M.; Pitulko, V.V.; Sousa, V.C.; Allentoft, M.E.; Vinner, L.; Rasmussen, S.; Margaryan, A.; de Barros Damgaard, P.; de la Fuente, C.; Renaud, G.; et al. The population history of northeastern Siberia since the Pleistocene. *Nature* **2019**, *570*, 182–188. [[CrossRef](#)]
5. Orlando, L.; Gilbert, M.T.P.; Willerslev, E. Reconstructing ancient genomes and epigenomes. *Nat. Rev. Genet.* **2015**, *16*, 395–408. [[CrossRef](#)]
6. Nägele, K.; Posth, C.; Orbegozo, M.I.; De Armas, Y.C.; Teresita, S.; Godoy, H.; Herrera, U.M.G.; Nieves-colón, M.A.; Sandoval-velasco, M. Genomic insights into the early peopling of the Caribbean. *Science* **2020**, *8697*. [[CrossRef](#)]
7. Nieves-Colón, M.A.; Pestle, W.J.; Reynolds, A.W.; Llamas, B.; De La Fuente, C.; Fowler, K.; Skerry, K.M.; Crespo-Torres, E.; Bustamante, C.D.; Stone, A.C.; et al. Ancient DNA reconstructs the genetic legacies of precontact puerto rico communities. *Mol. Biol. Evol.* **2020**, *37*, 611–626. [[CrossRef](#)]
8. Fernandes, D.M.; Sirak, K.A.; Ringbauer, H.; Sedig, J.; Cheronet, O.; Mah, M.; Mallick, S.; Olalde, I.; Brendan, J.; Callan, K.; et al. A Genetic History of the Pre-Contact Caribbean. *bioRxiv* **2020**. [[CrossRef](#)]

9. Morales-Arce, A.Y.; Hofman, C.A.; Duggan, A.T.; Benfer, A.K.; Katzenberg, M.A.; McCafferty, G.; Warinner, C. Successful reconstruction of whole mitochondrial genomes from ancient Central America and Mexico. *Sci. Rep.* **2017**, *7*, 18100. [[CrossRef](#)]
10. Kirchhoff, P. Mesoamérica: Sus límites geográficos, composición étnica y caracteres culturales. *Acta Am.* **1943**, *1*, 92–107.
11. López-Austin, A.; López-Luján, L. *El Pasado Indígena*; El Colegio de México, Fondo de Cultura Económica: Mexico City, Mexico, 1996; ISBN 978-607-16-2136-8.
12. Ardelean, C.F.; Becerra-valdivia, L.; Pedersen, M.W.; Schwenninger, J.; Oviatt, C.G.; Macías-quintero, J.I.; Arroyo-cabrales, J.; Sikora, M.; Ocampo-díaz, Y.Z.E.; La Rosa-díaz, J.J.D.; et al. Evidence of human occupation in Mexico around the Last Glacial Maximum. *Nature* **2020**. [[CrossRef](#)]
13. Becerra-Valdivia, L.; Higham, T. The timing and effect of the earliest human arrivals in North America. *Nature* **2020**, *584*. [[CrossRef](#)]
14. Lorenzo, J.L. *La Etapa Lítica de México*; Instituto Nacional de Antropología e Historia: Mexico City, Mexico, 1967.
15. Weaver, M.P. *The Aztecs, Maya, and Their Predecessors: Archaeology of Mesoamerica*; Academic Press: San Diego, CA, USA, 1993.
16. Alba-Hernandez, F. *La Población de México*; Centro de estudios Económicos y Demográficos, El Colegio de México: Mexico City, Mexico, 1976.
17. González-Sobrino, B.Z.; Pintado-Cortina, A.P.; Sebastián-Medina, L.; Morales-Mandujano, F.; Contreras, A.V.; Aguilar, Y.E.; Chávez-Benavides, J.; Carrillo-Rodríguez, A.; Silva-Zolezzi, I.; Medrano-González, L. Genetic diversity and differentiation in urban and indigenous populations of Mexico: Patterns of mitochondrial DNA and Y-chromosome lineages. *Biodemogr. Soc. Biol.* **2016**, *62*, 53–72. [[CrossRef](#)]
18. Torroni, A.; Schurr, T.G.; Cabell, M.F.; Brown, M.D.; Neel, J.V.; Larsen, M.; Smith, D.G.; Vullo, C.M.; Wallace, D.C. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* **1993**, *53*, 563–590. [[PubMed](#)]
19. Söchtig, J.; Álvarez-Iglesias, V.; Mosquera-Miguel, A.; Gelabert-Besada, M.; Gómez-Carballa, A.; Salas, A. Genomic insights on the ethno-history of the Maya and the “Ladinos” from Guatemala. *BMC Genom.* **2015**, *16*. [[CrossRef](#)] [[PubMed](#)]
20. Green, L.D.; Derr, J.N.; Knight, A. mtDNA affinities of the peoples of north-central Mexico. *Am. J. Hum. Genet.* **2000**, *66*, 989–998. [[CrossRef](#)] [[PubMed](#)]
21. Perego, U.A.; Achilli, A.; Angerhofer, N.; Accetturo, M.; Pala, M.; Olivieri, A.; Kashani, B.H.; Ritchie, K.H.; Scozzari, R.; Kong, Q.P.; et al. Distinctive paleo-indian migration routes from beringia marked by two rare mtDNA haplogroups. *Curr. Biol.* **2009**, *19*, 1–8. [[CrossRef](#)]
22. Sandoval, K.; Buentello-Malo, L.; Peñaloza-Espinosa, R.; Avelino, H.; Salas, A.; Calafell, F.; Comas, D. Linguistic and maternal genetic diversity are not correlated in Native Mexicans. *Hum. Genet.* **2009**, *126*, 521–531. [[CrossRef](#)]
23. Guardado-Estrada, M.; Juárez-Torres, E.; Medina-Martínez, I.; Wegier, A.; Macías, A.; Gomez, G.; Cruz-Talonia, F.; Roman-Bassaure, E.; Piñero, D.; Kofman-Alfaro, S.; et al. A great diversity of Amerindian mitochondrial DNA ancestry is present in the Mexican mestizo population. *J. Hum. Genet.* **2009**, *54*, 695–705. [[CrossRef](#)]
24. Kemp, B.M.; Gonzalez-Oliver, A.; Malhi, R.S.; Monroe, C.; Schroeder, K.B.; McDonough, J.; Rhett, G.; Resendez, A.; Penaloza-Espinosa, R.L.; Buentello-Malo, L.; et al. Evaluating the farming/language dispersal hypothesis with genetic variation exhibited by populations in the southwest and Mesoamerica. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 6759–6764. [[CrossRef](#)]
25. Gorostiza, A.; Acunha-Alonzo, V.; Regalado-Liu, L.; Tirado, S.; Granados, J.; Sámano, D.; Rangel-Villalobos, H.; González-Martín, A. Reconstructing the history of mesoamerican populations through the study of the mitochondrial DNA control region. *PLoS ONE* **2012**, *7*, e44666. [[CrossRef](#)] [[PubMed](#)]
26. Mizuno, F.; Gojobori, J.; Wang, L.; Onishi, K.; Sugiyama, S.; Granados, J.; Gomez-Trejo, C.; Acuña-Alonzo, V.; Ueda, S. Complete mitogenome analysis of indigenous populations in Mexico: Its relevance for the origin of Mesoamericans. *J. Hum. Genet.* **2014**, *59*, 359–367. [[CrossRef](#)] [[PubMed](#)]
27. González-Martín, A.; Gorostiza, A.; Regalado-Liu, L.; Arroyo-Peña, S.; Tirado, S.; Nuño-Arana, I.; Rubi-Castellanos, R.; Sandoval, K.; Coble, M.D.; Rangel-Villalobos, H. Demographic history of indigenous populations in Mesoamerica based on mtDNA sequence data. *PLoS ONE* **2015**, *10*, e0131791. [[CrossRef](#)] [[PubMed](#)]

28. Páez-Riberos, L.A.; Muñoz-Valle, J.F.; Figuera, L.E.; Nuño-Arana, I.; Sandoval-Ramírez, L.; González-Martín, A.; Ibarra, B.; Rangel-Villalobos, H. Y-linked haplotypes in Amerindian chromosomes from Mexican populations: Genetic evidence to the dual origin of the Huichol tribe. *Leg. Med.* **2006**, *8*, 220–225. [[CrossRef](#)]
29. Sandoval, K.; Moreno-Estrada, A.; Mendizabal, I.; Underhill, P.A.; Lopez-Valenzuela, M.; Peñaloza-Espinosa, R.; Lopez-Lopez, M.; Buentello-Malo, L.; Avelino, H.; Calafell, F.; et al. Y-chromosome diversity in native Mexicans reveals continental transition of genetic structure in the Americas. *Am. J. Phys. Anthropol.* **2012**, *148*, 395–405. [[CrossRef](#)] [[PubMed](#)]
30. Perez-Benedico, D.; La Salvia, J.; Zeng, Z.; Herrera, G.A.; Garcia-Bertrand, R.; Herrera, R.J. Mayans: A y chromosome perspective. *Eur. J. Hum. Genet.* **2016**, *24*, 1352–1358. [[CrossRef](#)] [[PubMed](#)]
31. Silva-Zolezzi, I.; Hidalgo-Miranda, A.; Estrada-Gil, J.; Fernandez-Lopez, J.C.; Uribe-Figueroa, L.; Contreras, A.; Balam-Ortiz, E.; Del Bosque-Plata, L.; Velazquez-Fernandez, D.; Lara, C.; et al. Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 8611–8616. [[CrossRef](#)] [[PubMed](#)]
32. Moreno-Estrada, A.; Gignoux, C.R.; Fernández-López, J.C.; Zakharia, F.; Sikora, M.; Contreras, A.V.; Acuña-Alonzo, V.; Sandoval, K.; Eng, C.; Romero-Hidalgo, S.; et al. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* **2014**, *344*, 1280–1285. [[CrossRef](#)]
33. Ávila-Arcos, M.C.; McManus, K.F.; Sandoval, K.; Rodríguez-Rodríguez, J.E.; Martin, A.R.; Luisi, P.; Villa-Islas, V.; Peñaloza-Espinosa, R.I.; Eng, C.; Huntsman, S.; et al. Population history and gene divergence in Native Mexicans inferred from 76 human exomes. *bioRxiv* **2019**, 534818. [[CrossRef](#)]
34. Bustamante, C.D.; De La Vega, F.M.; Burchard, E.G. Genomics for the world. *Nature* **2011**, *475*, 163–165. [[CrossRef](#)]
35. Sirugo, G.; Williams, S.M.; Tishkoff, S.A. The missing diversity in human genetic studies. *Cell* **2019**, *177*, 26–31. [[CrossRef](#)] [[PubMed](#)]
36. Tsois, K.S.; Begay, R.L.; Fox, K.; Garrison, N.A. Generations of genomes: Advances in paleogenomics technology and engagement for Indigenous people of the Americas. *Curr. Opin. Genet. Dev.* **2020**, *62*, 91–96. [[CrossRef](#)] [[PubMed](#)]
37. Anderson, S.; Bankier, A.T.; Barrell, B.G.; de Bruijn, M.H.L.; Coulson, A.R.; Drouin, J.; Eperon, I.C.; Nierlich, D.P.; Roe, B.A.; Sanger, F.; et al. Sequence and organization of the human mitochondrial genome. *Nature* **1981**, *290*, 457–465. [[CrossRef](#)] [[PubMed](#)]
38. Giles, R.E.; Blanc, H.; Cann, H.M.; Wallace, D.C. Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 6715–6719. [[CrossRef](#)]
39. Sigurðardóttir, S.; Helgason, A.; Gulcher, J.R.; Stefansson, K.; Donnelly, P. The mutation rate in the human mtDNA control region. *Am. J. Hum. Genet.* **2000**, *66*, 1599–1609. [[CrossRef](#)]
40. Andrews, R.M.; Kubacka, I.; Chinnery, P.F.; Lightowlers, R.N.; Turnbull, D.M.; Howell, N. Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **1999**, *23*, 147. [[CrossRef](#)]
41. Behar, D.M.; van Oven, M.; Rosset, S.; Metspalu, M.; Loogväli, E.-L.; Silva, N.M.; Kivisild, T.; Torroni, A.; Villems, R. A 'copernican' reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **2012**, *90*, 675–684. [[CrossRef](#)]
42. van Oven, M.; Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **2009**, *30*, 386–394. [[CrossRef](#)]
43. Krings, M.; Stone, A.C.; Schmitz, R.W.; Krainitzki, H.; Stoneking, M.; Paabo, S. Neandertal DNA sequences and the origin of modern humans. *Cell* **1997**, *90*, 19–30. [[CrossRef](#)]
44. Llamas, B.; Willerslev, E.; Orlando, L.; Orlando, L. Human evolution: A tale from ancient genomes. *Philos. Trans. R. Soc. B Biol. Sci.* **2017**, *372*. [[CrossRef](#)]
45. Tamm, E.; Kivisild, T.; Reidla, M.; Metspalu, M.; Smith, D.G.; Mulligan, C.J.; Bravi, C.M.; Rickards, O.; Martinez-Labarga, C.; Khusnutdinova, E.K.; et al. Beringian standstill and spread of native American founders. *PLoS ONE* **2007**, *2*, e829. [[CrossRef](#)]
46. Perego, U.A.; Angerhofer, N.; Pala, M.; Olivieri, A.; Lancioni, H.; Kashani, B.H.; Carossa, V.; Ekins, J.E.; Gómez-Carballa, A.; Huber, G.; et al. The initial peopling of the Americas: A growing number of founding mitochondrial genomes from Beringia. *Genome Res.* **2010**, *20*, 1174–1179. [[CrossRef](#)]

47. Llamas, B.; Fehren-Schmitz, L.; Valverde, G.; Soubrier, J.; Mallick, S.; Rohland, N.; Nordenfelt, S.; Valdiosera, C.; Richards, S.M.; Rohrlach, A.; et al. Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Sci. Adv.* **2016**, *2*, e1501385. [[CrossRef](#)] [[PubMed](#)]
48. Morales-Arce, A.Y.; McCafferty, G.; Hand, J.; Schmill, N.; McGrath, K.; Speller, C. Ancient mitochondrial DNA and population dynamics in postclassic Central Mexico: Tlatelolco (AD 1325–1520) and Cholula (AD 900–1350). *Archaeol. Anthropol. Sci.* **2019**, *11*, 3459–3475. [[CrossRef](#)]
49. Verdugo, C.; Zhu, K.; Kassadjikova, K.; Berg, L.; Forst, J.; Galloway, A.; Brady, J.E.; Fehren-Schmitz, L. An investigation of ancient Maya intentional dental modification practices at Midnight Terror Cave using anthroposcopic and paleogenomic methods. *J. Archaeol. Sci.* **2020**, *115*, 105096. [[CrossRef](#)]
50. Salazar, A.H. *Estudio Genético Poblacional de Restos Óseos Prehispánicos de Una Subpoblación de Teotihuacan*; CINVESTAV: México City, Mexico, 2007.
51. Aguirre-Samudio, A.J.; González-Sobrino, B.Z.; Álvarez-Sandoval, B.A.; Montiel, R.; Serrano-Sánchez, C.; Meza-Peñaloza, A. Genetic history of classic period Teotihuacan burials in Central Mexico. *Rev. Argent. Antropol. Biol.* **2017**, *19*, 1–14.
52. Álvarez-Sandoval, B.A.; Manzanilla, L.R.; Montiel, R. Sex determination in highly fragmented human DNA by High-Resolution Melting (HRM) analysis. *PLoS ONE* **2014**, *9*. [[CrossRef](#)]
53. Álvarez-Sandoval, B.A.; Manzanilla, L.R.; González-Ruiz, M.; Malgosa, A.; Montiel, R. Genetic evidence supports the multiethnic character of teopanazgo, a neighborhood center of teotihuacan, Mexico (ad 200–600). *PLoS ONE* **2015**, *10*, e0132371. [[CrossRef](#)]
54. Mata-Míguez, J.; Overholtzer, L.; Rodríguez-Alegría, E.; Kemp, B.M.; Bolnick, D.A. The genetic impact of Aztec imperialism: Ancient mitochondrial DNA evidence from Xaltocan, Mexico. *Am. J. Phys. Anthropol.* **2012**, *149*, 504–516. [[CrossRef](#)]
55. Kemp, B.M.; Reséndez, A.; Román Berrelleza, J.A.; Malhi, R.; Smith, D.G. An analysis of ancient Aztec mtDNA from Tlatelolco: Pre-Columbian relations and the spread of Uto-Aztecan. In *Biomolecular Archaeology: Genetic Approaches to the Past*; Reed, D.M., Ed.; Southern Illinois University: Carbondale, IL, USA, 2005; pp. 22–46.
56. Navarro, E.S. *De La Mesoamérica Prehispánica a la Colonial: La Huella del DNA Antiguo*; Universitat Autònoma de Barcelona: Barcelona, Spain, 2006.
57. De La Cruz, I.; González-Oliver, A.; Kemp, B.M.; Román, J.A.; Smith, D.G.; Torre-Blanco, A. Sex identification of children sacrificed to the ancient Aztec rain gods in Tlatelolco. *Curr. Anthropol.* **2008**, *49*, 519–526. [[CrossRef](#)]
58. Martín, A.I.J. *Parentesco Biológico Entre los Pobladores Prehispánicos de Cholula, Mediante el Análisis Molecular de Sus Restos Óseos*; Escuela Nacional de Antropología e Historia: Ciudad de México, Mexico, 2002.
59. Chatters, J.C.; Kennett, D.J.; Asmerom, Y.; Kemp, B.M.; Polyak, V.; Lberto, N.B.; Beddows, P.A.; Reinhardt, E.; Arroyo-Cabrales, J.; Bolnick, D.A.; et al. Late pleistocene human skeleton and mtDNA link paleoamericans and modern native Americans. *Science* **2014**, *344*, 750–755. [[CrossRef](#)]
60. Ochoa-Lugo, M.I.; de Muñoz, M.L.; Pérez-Ramírez, G.; Beaty, K.G.; López-Armenta, M.; Cervini-Silva, J.; Moreno-Galeana, M.; Meza, A.M.; Ramos, E.; Crawford, M.H.; et al. Genetic affiliation of pre-hispanic and contemporary Mayas through maternal lineage. *Hum. Biol.* **2016**, *88*, 136–167. [[CrossRef](#)] [[PubMed](#)]
61. González-Oliver, A.; Márquez-Morfín, L.; Jiménez, J.C.; Torre-Blanco, A. Founding amerindian mitochondrial DNA lineages in ancient Maya from Xcaret, Quintana Roo. *Am. J. Phys. Anthropol.* **2001**, *116*, 230–235. [[CrossRef](#)]
62. Merriwether, D.A.; Reed, D.M.; Ferrell, R.E. Ancient and contemporary Mitochondrial DNA variation in the Maya. In *Bones of the Maya: Studies of Ancient Skeletons*; Smithsonian Institution Press: Washington, DC, USA, 1997; pp. 208–217. ISBN 0817353763.
63. Mizuno, F.; Kumagai, M.; Kurosaki, K.; Hayashi, M.; Sugiyama, S.; Ueda, S.; Wang, L. Imputation approach for deducing a complete mitogenome sequence from low-depth-coverage next-generation sequencing data: Application to ancient remains from the Moon Pyramid, Mexico. *J. Hum. Genet.* **2017**, *62*, 631–635. [[CrossRef](#)]
64. Manzanilla, L.R. *Multiethnicity and Migration at Teopanazgo*; University Press of Florida: Gainesville, FL, USA, 2017; ISBN 9780813054285.
65. Plunket, P.; Uruñuela, G. Preclassic household patterns preserved under volcanic ash at Tetimpa, Puebla, Mexico. *Lat. Am. Antiq.* **1998**, *9*, 287–309. [[CrossRef](#)]
66. Cowgill, G.L. State and society at Teotihuacan, Mexico. *Annu. Rev. Anthropol.* **1997**, *26*, 129–161. [[CrossRef](#)]

67. Froese, T.; Gershenson, C.; Manzanilla, L.R. Can government be self-organized? A mathematical model of the collective social organization of ancient Teotihuacan, Central Mexico. *PLoS ONE* **2014**, *9*. [[CrossRef](#)]
68. Siebe, C. Age and archaeological implications of Xitle volcano, southwestern Basin of Mexico-City. *J. Volcanol. Geotherm. Res.* **2000**, *104*, 45–64. [[CrossRef](#)]
69. Appel, G.I.M. Elementos traza aplicados al análisis de la paleodieta en Teopacazco. In *Estudios Arqueométricos del Centro de Barrio de Teopacazco en Teotihuacan*; Manzanilla, L.R., Ed.; Coordinación de Humanidades—Coordinación de la Investigación Científica, UNAM: Mexico City, Mexico, 2012; pp. 325–345.
70. Puente, P.M.; Alvarado, E.C.; Naim, L.R.M.; Trujano, F.J.O. Estudio de la paleodieta empleando isótopos estables de los elementos carbono, oxígeno y nitrógeno en restos humanos y de fauna encontrados en el barrio teotihuacano de Teopacazco. In *Estudios Arqueométricos del Centro de Barrio de Teopacazco en Teotihuacan*; Manzanilla, L.R., Ed.; Coordinación de Humanidades—Coordinación de la Investigación Científica, UNAM: Mexico City, Mexico, 2012; pp. 347–423.
71. Schaaf, P.; Solís, G.; Manzanilla, L.R.; Hernández, T.; Lailson, B.; Horn, P. Isótopos de estroncio aplicados a estudios de migración humana en el centro de barrio de Teopacazco, Teotihuacan. In *Estudios Arqueométricos del Centro de Barrio de Teopacazco en Teotihuacan*; Manzanilla, L.R., Ed.; Coordinación de Humanidades—Coordinación de la Investigación Científica, UNAM: Mexico City, Mexico, 2012; pp. 425–448.
72. Hassig, R. *War and Society in Ancient Mesoamerica*; University of California: Oakland, CA, USA, 1992.
73. Carballo, D.M. Effigy vessels, religious integration, and the origins of the central Mexican pantheon. *Anc. Mesoam.* **2007**, *18*, 53–67. [[CrossRef](#)]
74. Lachniet, M.S.; Bernal, J.P.; Asmerom, Y.; Polyak, V.; Piperno, D. A 2400 yr Mesoamerican rainfall reconstruction links climate and cultural change. *Geology* **2012**, *40*, 259–262. [[CrossRef](#)]
75. Acuna-Soto, R.; Stahle, D.W.; Therrell, M.D.; Chavez, S.G.; Cleaveland, M.K. Drought, epidemic disease, and the fall of classic period cultures in Mesoamerica (AD 750–950). Hemorrhagic fevers as a cause of massive population loss. *Med. Hypotheses* **2005**, *65*, 405–409. [[CrossRef](#)] [[PubMed](#)]
76. Acuna-Soto, R.; Stahle, D.W.; Cleaveland, M.K.; Therrell, M.D. Megadrought and megadeath in 16th century Mexico. *Emerg. Infect. Dis.* **2002**, *8*, 360–362. [[CrossRef](#)] [[PubMed](#)]
77. Kubler, G. La traza colonial de Cholula. *Estud. Hist. Novohisp.* **1968**, *2*, 1–30. [[CrossRef](#)]
78. Brumfiel, E.M. Aztec state making: Ecology, structure, and the origin of the state. *Am. Anthropol.* **1983**, *85*, 261–284. [[CrossRef](#)]
79. Gibson, C. *The Aztecs under Spanish Rule: A History of the Indians of the Valley of Mexico, 1519–1810*; Stanford University: Stanford, CA, USA, 1964.
80. Hicks, F. *Xaltocan under Mexican Domination, 1435–1520*; Caciques their people; Marcus, J., Zeitlin, J.F., Eds.; Museum Anthropology, University of Michigan: Ann Arbor, MI, USA, 1994; pp. 67–85.
81. Overholtzer, L. Otomies at Xaltocan: A case of ethnic change. In Proceedings of the 76th Annual Meeting of the Society for American Archaeology, Sacramento, CA, USA, 30 March–3 April 2011.
82. Kemp, B.M.; Lindo, J.; Bolnick, D.A.; Malhi, R.S.; Chatters, J.C. Response to comment on “late pleistocene human skeleton and mtDNA link paleoamericans and modern native Americans”. *Science* **2015**, *347*, 835. [[CrossRef](#)]
83. Prüfer, K.; Meyer, M. Comment on “late pleistocene human skeleton and mtDNA link paleoamericans and modern native Americans”. *Science* **2015**, *347*, 835. [[CrossRef](#)]
84. Rasmussen, M.; Anzick, S.L.; Waters, M.R.; Skoglund, P.; Degiorgio, M.; Stafford, T.W.; Rasmussen, S.; Moltke, I.; Albrechtsen, A.; Doyle, S.M.; et al. The genome of a late Pleistocene human from a Clovis burial site in western Montana. *Nature* **2014**, *506*, 225–229. [[CrossRef](#)]
85. Wade, L. The arrival of strangers. *Science* **2020**, *367*, 968–973. [[CrossRef](#)]
86. Laluz, C.; Pérez-Pérez, A.; Prats, E.; Cornudella, L.; Turbón, D. Lack of founding Amerindian mitochondrial DNA lineages in extinct Aborigines from Tierra del Fuego–Patagonia. *Hum. Mol. Genet.* **1997**, *6*, 41–46. [[CrossRef](#)]
87. Starikovskaya, Y.B.; Sukernik, R.I.; Schurr, T.G.; Kogelnik, A.M.; Wallace, D.C. mtDNA diversity in Chukchi and Siberian Eskimos: Implications for the genetic history of ancient Beringia and the peopling of the New World. *Am. J. Hum. Genet.* **1998**, *63*, 1473–1491. [[CrossRef](#)]
88. McCafferty, G.; Dennett, C. Ethnogenesis and hybridity in Proto-historic Nicaragua. *Archaeol. Rev. Camb.* **2013**, *28*, 191–215.

89. Umaña, A.C.; Rojas, E.I. Mapa de la distribución territorial aproximada de las lenguas indígenas habladas en Costa Rica y en sectores colindantes de Nicaragua y de Panamá en el siglo XVI. *Estud. Linguist. Chibcha* **2009**, *XXVIII*, 109–112.
90. Ioannidis, A.G.; Blanco-Portillo, J.; Sandoval, K.; Hagelberg, E.; Miquel-Poblete, J.F.; Moreno-Mayar, J.V.; Rodríguez-Rodríguez, J.E.; Quinto-Cortés, C.D.; Auckland, K.; Parks, T.; et al. Native American gene flow into Polynesia predating Easter Island settlement. *Nature* **2020**, *583*, 572–577. [[CrossRef](#)] [[PubMed](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Chapter IV: The Genetic History of Pre-Columbian Mesoamerica

Statement of Authorship

Title of Paper	The Genetic History of pre-Columbian Mesoamerica
Publication Status	<input type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input checked="" type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	

Principal Author

Name of Principal Author (Candidate)	Xavier Roca Rada
Contribution to the Paper	Conceptualisation, performed laboratory experiments, bioinformatic processing, data analyses, results interpretation and drafted the manuscript.
Overall percentage (%)	70%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 18/10/2022

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Prof. Dr. Lars Fehren-Schmitz
Contribution to the Paper	performed laboratory experiments
Signature	Date 10/20/2022

Name of Co-Author	Deguilloux Marie-France
Contribution to the Paper	DNA isolation, libraries construction and shotgun sequencing (screening) of Michoacan human remains (ancient DNA facilities of the PACEA lab, Bordeaux University)
Signature	Date 21/10/2022

Name of Co-Author	LINDA R. MANZANILLA
Contribution to the Paper	SAMPLES FROM MY EXCAVATIONS AT TEOPANCAZCO, TEOTIHUACAN AND CONTEXTUAL DATA
Signature	Date OCTOBER 22, 2022

Name of Co-Author	Dr Yassine Scuilmi		
Contribution to the Paper	Co supervised the Mr Rocca Rada, helped interpret the results, and provided input on the manuscript.		
Signature		Date	20/10/2022

Name of Co-Author	João Teixeira		
Contribution to the Paper	Supervision, technical and general guidance, edited the manuscript		
Signature		Date	18/10/2022

Name of Co-Author	Cristina Santos		
Contribution to the Paper	Provide samples		
Signature		Date	24/10/2022

Name of Co-Author	Assumpció Maigosa		
Contribution to the Paper	Provided samples		
Signature		Date	27/10/2022

Name of Co-Author	Grégory Pereira		
Contribution to the Paper	Provision of samples and data on archaeological contexts		
Signature		Date	28/10/2022

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Conceptualisation, technical and general guidance, edited the manuscript		
Signature		Date	18/10/2022

1. Introduction

The main aim of this chapter is to investigate the Pre-Columbian Mesoamerican genetic history. As reviewed in Chapter III¹, Mesoamericans probably carried an invaluable genetic diversity partly lost during the Spanish conquest and the subsequent colonial period. However, aDNA research in Mesoamerica is extremely limited, despite a rich archaeological record and Indigenous cultural diversity. To fill this gap, I analysed 42 archaeological samples from 5 different archaeological sites and time periods (320–1,400 CE): Teopancazco (TEO; n=10), Naachtun (NA; n=7), El Rincón (ER; n=13), Malpaís Prieto (MP; n=6), and Tlatelolco (TLA; n=6) (Figure 1).

It is worth noting that this project suffered several setbacks due to the COVID-19 pandemic, such as the inability to travel to Mexico for sample collection and community engagement. Despite sample size being one of the main limitations of this study, it offers the first Pre-Columbian Mesoamerican aDNA genetic screening that includes full mitogenomes and nuclear data. It is still work in progress, with further analyses proposed in the discussion of this chapter.

2. Methods

2.1. Archaeological sites

2.1.1. Teotihuacan–Teopancazco

As previously described in Chapter III¹, Teotihuacan was a powerful political and military city centre from the Late Preclassic and Classic periods (200 BCE–650 CE) located in the Basin of Mexico (Figure 1). Teotihuacan controlled trade routes and attracted large masses of people from distant regions, coordinating them into a seemingly effective multi-ethnic system. The metropolis grew into the largest city of Mesoamerica and one of the world's largest and most urbanised population centres of the first millennium CE, reaching 80,000–100,000 inhabitants

by approximately 150 CE²⁻⁵. The city's internal structure consisted of various neighbourhoods. While most neighbourhoods were segregated by the province of origin of its inhabitants, some neighbourhoods were multi-ethnic settlements, such as Teopancazco⁵ (TEO). In this chapter, I analysed 10 samples from TEO with different preservation status that date from four different stages of the Classic period (see Results, Table 1).

2.1.2. Naachtun

Naachtun (NAA) was a major city centre of the early days of the Mayan civilisation in the Yucatán Peninsula. It is currently located in northern Petén, Guatemala, approximately 2 km south of the Mexican border. Early on in its history (before 378 CE), NAA established an alliance with the Tikal dynasty, which hastened its first apogee⁶. Later, during the first half of the Late Classic (also known as Epiclassic in the Basin of Mexico), this centre was submitted to the Kanu'l kings of Calakmul⁷. After its emancipation from this rule around 730 CE, NAA enjoyed a second apogee that lasted until the end of the Late Classic (950 CE).

The 13 studied individuals from this archaeological site came from an archaeological complex that started by being a single-structure household and evolved into a fully developed residential compound during at least three hundred years during the Late Classic (500-800 CE) (see Results, Table 1). Archaeological analyses suggested that these individuals probably were part of an intermediate elite involved in artistic production for higher-ranking households⁸.

2.1.3. Michoacán–El Rincón and Malpaís Prieto

The Loma Alta culture of the Late Preclassic and Classic periods (100 BCE–400 CE) is the first culture that has been described in Michoacán (western Mexican state), although some previous agricultural activities have been reported⁹. The Loma Alta culture has been described for having among other characteristics: (i) elaborated pottery production; (ii) cremation as the

main funerary ritual; and (iii) a connection with Teotihuacan¹⁰. However, a new cultural dynamic developed in the Epiclassic period (600–900 CE) with the emergence of several settlements across Michoacán and relatively far away from each other, an intensive exploitation of the natural resources, and successive funeral burials for high-status people¹¹. El Rincón (ER) is a settlement that dates from this period, concretely from the 7th century CE. A successive funeral burial of 13 individuals was analysed (see Results, Table 1). This period ended with the eruption of the Malpaís Prieto volcano and the abandonment of most settlements¹¹.

The Early Postclassic period (900–1,200 CE) remains highly unknown and is characterised by the apogee of the El Palacio-La Crucita settlement in Zacapu, whose culture had a connection with the Toltec Empire from the Basin of Mexico¹¹. This period ended with the abandonment of the northern areas of Zacapu, the massive colonisation of the areas surrounding the Malpaís Prieto volcano and the arrival of the Uacúsechas and Chichimecas from northern Mexico in the region¹². Then, four highly populated settlements—Malpaís Prieto (MP), Las Milpillas, El Infernillo and El Palacio—emerged during the Middle Postclassic¹². These four settlements were a direct precedent of the following Purépecha Empire (also known as the Tarascan Empire) that emerged at the beginning of the 16th century CE in Pátzcuaro after the abandonment of the cities from Zacapu. The Purépecha Empire was contemporary with and an enemy of the Aztec Empire, against which it fought many wars before the arrival of the Spanish¹¹. Six individuals from the 14th century CE individually buried in the same funerary context coming from the Middle Postclassic MP settlement were screened using different bones and library preparations (see Results, Table 1).

2.1.4. Tlatelolco

Tlatelolco (TLA) was an Aztec city that was founded around 1,337 CE and is currently buried beneath modern-day Mexico City. Even though TLA was a sister city to Tenochtitlan (the

capital of the Aztec Empire), religious and cultural practices were distinct between the two cities. TLA was the biggest and most diversified marketplace in Mesoamerica during the Late Postclassic¹³. Kemp et al. (2005)¹⁴ and Solorzano Navarro (2006)¹⁵ previously analysed several individuals from a burial at the ceremonial centre of TLA, but only restriction fragment length polymorphism (RFLP) analyses were performed, sometimes followed by Sanger sequencing of the Hypervariable Region 1 (HVR1). In this study, I re-analysed with new techniques (high-throughput sequencing and targeted capture) six individuals previously reported by Solorzano Navarro (2006)¹⁵ (see Results, Table 1)

2.2. Previously published ancient Americans

In order to compare the newly generated data with existing surveys of ancient genetic variation in the region, I included 44 previously published ancient Americans from North America, Central America, and northern South America (Figure 1). Most of these individuals predate the settlement of Mesoamerica and are either part of the homogenous population associated with the Clovis culture¹⁶ that expanded southwards carrying the primary source of the ancestry of Central and South Americans¹⁷ (Anzick¹⁶, Belize^{17,18} and Las Locas¹⁹), or from Aridoamerica (Pre-Columbian mummies²⁰, Baja Mexico²¹ and Pericúes^{20,21}).

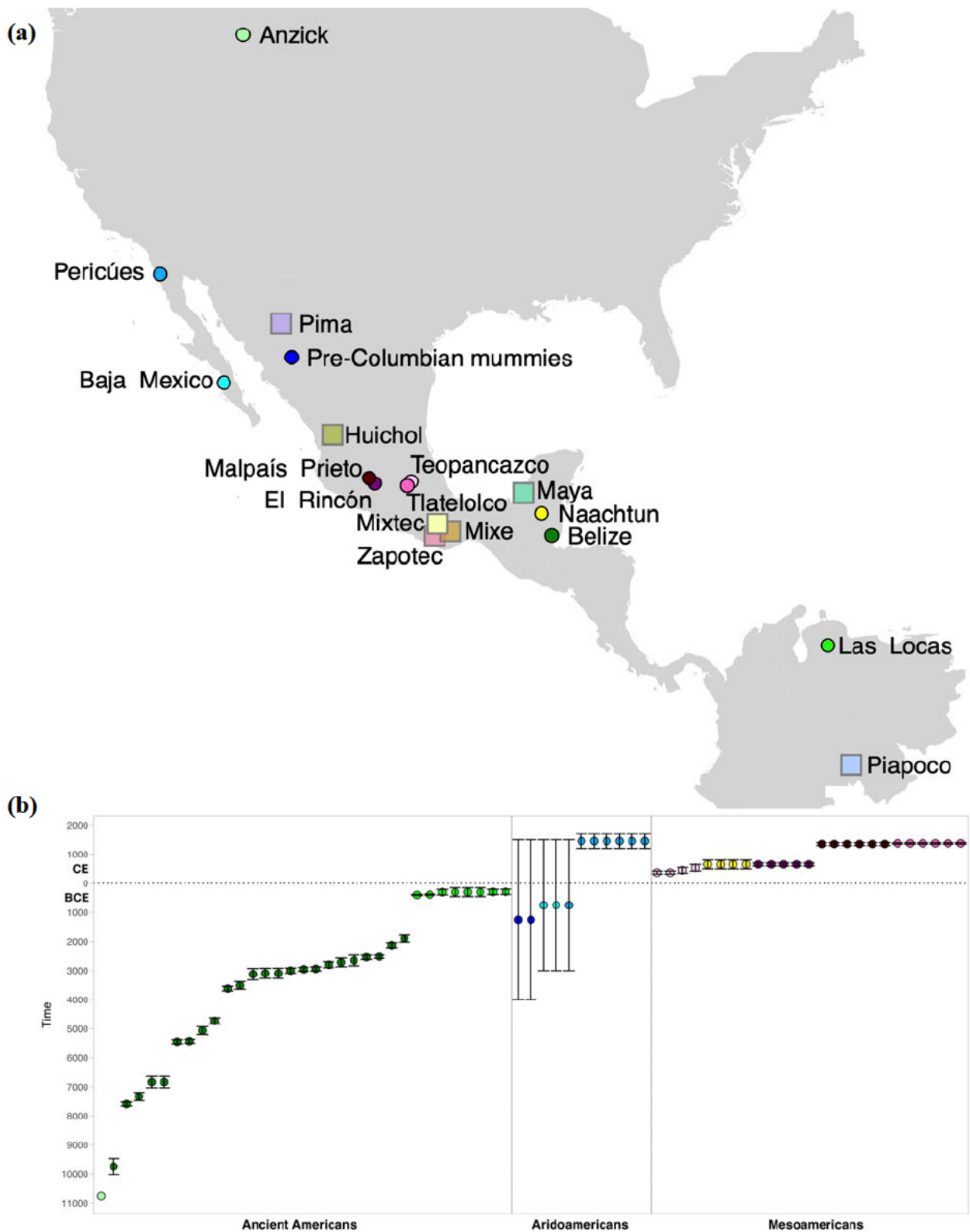


Figure 1. (a) Location and (b) age of the studied Mesoamericans and previously published ancient individuals used for the analyses of this chapter. Ancient individuals are represented in black circles with bright fill colours. Location of the modern-day Indigenous populations used for comparison purposes are represented in grey squares and light fill colours.

2.3. Ancient DNA laboratory work

Ancient DNA molecular processing was conducted in three different clean-room facilities—University of Adelaide’s Australian Centre for Ancient DNA (ACAD), Australia; University of Bordeaux, France; and the Paleogenomics Laboratory, University of California, Santa Cruz (UCSC), USA—following strict precautions to minimise contamination²². Samples were processed while wearing face mask, visors, hooded coverall, hair net, and gloves, in purpose-build aDNA facilities with positive air pressure. Standard precautions to avoid contaminations were employed²³.

A total of 42 well-preserved skeletal remains were included for this study. When teeth were used, powder was collected from the cementum, which is richer in DNA than dentin²⁴. DNA was extracted using a method optimised to retrieve degraded ancient DNA fragments²⁵ in all three ancient DNA facilities. While partially UDG-treated double-stranded DNA libraries were generated²⁶ at ACAD and University of Bordeaux, partially UDG-treated single-stranded DNA libraries were generated at UCSC²⁷. Paired-end shotgun sequencing using an Illumina two-colour chemistry (200 or 300 cycles) was performed by service providers at UCSC, South Australian Genomics Centre (SAGC; Adelaide, Australia), and the Kinghorn Centre for Clinical Genomics (Sydney, Australia).

All post-shotgun sequencing lab work was performed in ACAD’s post-PCR facilities. Libraries were reamplified and subjected to in-solution target hybridisation using an in-house protocol to capture mitochondrial fragments¹⁹. Subsequently, all libraries were captured again using the myBaits Expert Human Affinities Prime Plus Kit by DAICEL Arbor Biosciences (Ann Harbor, MI, USA), which targets 1.24 million nuclear single nucleotide polymorphisms (SNPs) (hereafter 1240k) used in population genetics²⁸, as well as the mitochondrial genome and 46 thousand SNPs on the Y chromosome. After each capture, libraries were paired-end sequenced

at the Kinghorn Centre for Clinical Genomics (Sydney, NSW, Australia) using an Illumina NovaSeq (200 cycles).

2.4. Sequencing data processing

Raw reads from the shotgun sequencing run were processed and mapped against the human reference genome (GRCh37d5) using the nf-core/eager²⁹ v2.4.0 pipeline. To determine data authenticity, I evaluated the endogenous DNA percentage, determined the fragment size distribution, and calculated the post-mortem damage rate at read termini using DamageProfiler³⁰.

Raw reads from all other sequencing runs were processed and mapped against the human reference genome (hg19) using the nf-core/eager²⁹ v2.4.0 pipeline. A subset of libraries passed authenticity quality thresholds to undergo downstream analyses (See Results, Table 1, and Table 2). Reads that were retained were trimmed 2 bp from each end using the trimBam function of bamUtil (<https://github.com/statgen/bamUtil>). I applied read q30 and Q30 base quality filters through samtools³¹ version 1.10 mpileup function and I then performed pseudohaploidisation on the ancient samples and subsetted the genome-wide data to match the number of SNPs in the 1240k capture panel using pileupCaller (<https://github.com/stschiff/sequenceTools>). Due to the highly degraded nature of aDNA, it is often not possible to call diploid genotypes and even if a site presents more than one allele, it could derive from actual heterozygosity, sequencing errors or deamination. Hence, pseudohaploidisation is a standard step in aDNA studies and it consists in randomly selection a single allele to represent each locus.

2.5. Sex determination

Using shotgun data, genetic sex determination was assigned using the method described by Gower et al. (2019). This method uses dosage of reads mapping to the X chromosome and autosomes to determine the genetic sex of an individual. Under the assumption that the number of reads obtained should reflect the chromosome copy numbers and chromosome lengths, two binomial models are constructed, one for males and one for females. A likelihood ratio test is then used to distinguish between the two models. A read dosage on the X chromosome (M_x) close to 0.5 or 1 means that the individual is assigned as male or female, respectively.

To compare genetic sex results, I also used SexDetERRmine (<https://github.com/nf-core/modules/tree/master/modules/nf-core/sexdeterminer>) with default quality cut-off values for -q30 and -Q30. SexDetERRmine calculates X- and Y- ratios, corresponding to the number of reads mapping to each of the sex chromosomes relative to the autosomes.

2.6. Uniparental markers

2.6.1. Y Chromosome

To obtain Y-chromosome haplotypes for the male individuals, I converted the Eigenstrat files into Packedped files and then into a VCF file selecting only the Y chromosome sites. The VCF file was inputted into yhaplo v1.1.2³² to automatically determine the haplotypes.

2.6.2. Mitochondrial DNA

Raw reads from all the sequencing runs were processed and mapped against the mitochondrial revised Cambridge Reference Sequence (rCRS)³³ using nf-core/eager²⁹ v2.4.0 pipeline. The sequencing read pileups were visualised in Geneious v2022.1.1 (Biomatters; <https://www.geneious.com>) and polymorphisms were called with a minimum coverage of 3

and a minimum variant frequency of 0.5. The assembly and the resulting list of SNPs were manually verified and compared to HaploGrep2 v2.4.0 (online implementation) from the Division of Genetic Epidemiology of the Medical University of Innsbruck (<https://haplogrep.uibk.ac.at/>)^{34,35}, based on Phylotree v17 (<http://www.phylotree.org/index.htm>)³⁶.

In addition, I calculated mitochondrial contamination using the mitoverse Haplocheck cloud service³⁷. This method checks for the contamination status by determining if multiple mitochondrial haplotypes are present within a single sample. Multiple haplotypes are detected through polymorphic variants and the resulting sequence is split into two components that then undergo haplotype assessment using haplogrep to determine contamination based on polymorphic site frequencies.

Finally, an unrooted Neighbour-Joining tree using the Tamura-Nei's genetic distance model with the studied ancient Mesoamericans was performed in Geneious v2022.1.1 (Biomatters; <https://www.geneious.com>).

2.7. Kinship analysis

I used the READ software to determine kinship between pairs of individuals³⁸. In the default settings, READ performs a normalisation step using the median of all pairwise differences as an estimate of the expected distance between two unrelated individuals.

2.8. Principal Component Analysis (PCA)

I performed a principal components analysis (PCA) using the smartpca program v10210 (EIGENSOFT) on the 1240k SNPs panel from 930 present-day worldwide populations of the HGDP³⁹, and projected the successful ancient samples from this study onto the modern-day

genetic variation. Then, I performed a PCA on the same SNPs from 55 present-day Indigenous populations from southern North America, Central America and northern South America^{16,39–41}, and projected the successful ancient samples from this study along with 43 ancient previously published Indigenous Americans^{16–21} ("Allen Ancient DNA Resource <https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>", v50.0) onto the modern-day genetic variation.

2.9. ADMIXTURE

ADMIXTURE is a model-based clustering approach for estimating ancestries³⁸ that was run with 1,121 modern-day and ancient individuals from Africa, Spain, East Asia, and the Americas^{16–21,39–59}. The model selected was unsupervised and 20 replicates were run. Common models among the different runs were identified, and clusters were aligned across different values of k using pong (<https://github.com/ramachandran-lab/pong>; v1.4.9). To determine the value of k that best represented the data, cross-validation errors were calculated, and the elbow method was applied.

2.10. f_3 -statistics

To explore the genetic affinities among the studied ancient individuals, I performed a series of outgroup f_3 -statistics in the form $f_3(\text{Mbuti}; \text{Test}, \text{X})$ using the qp3Pop function implemented in ADMIXTOOLS v7.0.2⁶⁰. Test refers to one archaeological site used for comparisons with each of the other sites, labelled X. Since the Mbuti population is an outgroup all the analysed samples, the resultant f_3 -statistic measures the shared drift between Test and X.

3. Results

3.1. Ancient genomic data screening

For each library, I evaluated the extent of endogenous aDNA via shotgun sequencing. I subsequently performed mtDNA and 1240k captures for all libraries except individual 24580 from ER, as it was the only one that had an endogenous content under 0.04%. As expected, hybridisation captures efficiently enriched for DNA targets, resulting in increased endogenous DNA proportions and a drop in the complexity of the libraries (Table 1).

After filtering individuals with: (i) more than 10,000 SNPs covered by at least one read in the 1240k capture panel (Table 1); (ii) a significant amount of reads mapping and covering the whole mitochondrial rCRS³³ (Table 1); (iii) a significant presence of the characteristic misincorporation patterns of aDNA (>3%) (Table 1); and (iv) absence or negligible levels of mitochondrial contamination (Table 2); I obtained informative dataset of 25 individuals: TEO (n=4), NA (n=4), ER (n=5), MP (n=6); and TLA (n=6). All the successful samples from TEO were males, all except one from MP were females, and all except one in NA and ER were females. In TL, I identified two females and four males (Table 2).

For replication purposes, individuals 24575_24581_ER and 25047_25048_TLA had more than one skeletal remain analysed, concretely, two different teeth per individual. While two single-stranded DNA libraries (L1605x and L1608) were performed on 24575_24581_ER at UCSC, two double-stranded DNA libraries (LP124_11 and LP121_13) were performed on 25047_25048_TLA at ACAD. Each pair of libraries analysed per individual presented highly similar results during quality assessment (Table 1). When comparing each library of the individual 25047_25048_TLA, the same mitochondrial polymorphisms were called, indicating the same haplotype and therefore the same individual (Figure 2). Even though the same mitochondrial haplogroup was called for both libraires of the individual 24575_24581_ER, not

all the same polymorphisms were called due to low coverage and missing positions (Figure 2). Kinship analyses suggested that both libraries came from the same individual.

In contrast, individual 24053_MP had two independent extractions and library preparations using the same petrous bone. A single-stranded DNA library was performed at UCSC (L1586) and a double-stranded DNA library at University of Bordeaux (A635). While the single-stranded DNA library (L1586) had higher endogenous DNA percentages at all different steps along the analytical pipeline and lower mean read lengths suggesting that shorter reads were retained in the process, the single-stranded DNA library also presented lower and asymmetrical rates of post-mortem damage at read termini (Table 1). When comparing each library, the same mitochondrial polymorphisms were called once again (Figure 2).

Individual Sep-33 (MP) had two different skeletal remains analysed (a metatarsal and a metacarpal) and two different library preparations were performed: single-stranded DNA library at UCSC (L1594) and a double-stranded DNA library at University of Bordeaux (A640), respectively. Interestingly, the metacarpal (double-stranded DNA library) qualitatively performed better (Table 1). When comparing each library, the same mitochondrial polymorphisms were called as well (Figure 2).

Individual Sep-41 (MP) had four different bones screened (two teeth, a metacarpal and an unidentified bone fragment) and four different libraries performed. All libraries were double-stranded and performed at University of Bordeaux (A637, A641, and A638, respectively), except the unidentified bone fragment that had a single-stranded DNA library performed at UCSC (L1588). Between double-stranded DNA libraires, once again the metacarpal qualitatively performed better. Even though it seems like the single-stranded DNA library (L1588) performed better than the other three—higher endogenous DNA percentages at all different steps along the analytical pipeline and a lower mean read length—, it has once again

lower rates of post-mortem damage at read termini. Quantitatively, the latter library has less reads mapping to the mitochondrial rCRS³³ than the other double-stranded libraries, being the library generated with the DNA extracted from the metacarpal the highest one. The same mitochondrial polymorphisms were called when comparing each library.

To maximise the genetic information per individual, I merged all the sequencing data generated from all the different types of libraries. I then could observe that individuals 24053, Sep-33, and Sep-41 presented low levels of mitochondrial contamination (Table 2). However, these contamination levels were negligible as the same mitochondrial polymorphisms were called when comparing each independently processed library. All these individuals include libraries initially screened at University of Bordeaux, suggesting a probable low-degree contamination during processing.

Finally, all the samples from TEO were compared according to their preservation status. In this case, only teeth were screened, and double-stranded DNA libraries were performed at ACAD. 20749B and 20750B skeletal remains were covered with cinnabar due to cultural funerary practices. They both presented poor results during quality assessment, but not as much as those individuals that presented mould stains, dryness, and a general observable bad preservation. All individuals that presented a macroscopic good preservation were successful, except one (20734B). In contrast, only one individual that presented a general observable bad preservation was successful (20745B).

Table 1. Summary of the studied archaeological samples and their paleogenetic results for quality assessment. Successful samples are highlighted in light grey.

Archaeological Site	Time period	Sample ID	Library ID	Bone	Preservation	Endogenous DNA (%)				5' C→T 1 st base (%)	3' G→A 1 st base (%)	Mean read length (bp)	#Reads mapped to rCRS ³³	#SNPs
						Post-Mapping (Shotgun)	Post-Dedup (Shotgun)	Post-Mapping (Shotgun and Captures)	Post-Dedup (Shotgun and Captures)					
TEO	Classic: Late Tlamimilolpa– Early Xolalpan 320–420 CE	20743B	LP109_7	Tooth	Bad Mould stains Dry	0.05	0.03	11.38	0.37	3.40	3.90	56.74	1,856	466
		20745B	LP109_8	Tooth	Bad Dry	0.68	0.5	18.38	14.91	9.70	9.60	63.52	75,735	185,555
		20747B	LP109_9	Tooth	Good	0.08	0.03	24.04	0.46	3.90	3.40	59.42	1,693	195
		20749B	LP109_10	Tooth	Bad Cinnabar	0.06	0.03	15.1	0.43	5.10	4.60	62.05	1,894	469
		20750B	LP109_11	Tooth	Bad Cinnabar	0.14	0.03	14.84	0.25	3.30	3.10	56.15	1,781	126
		20737B	LP109_3	Tooth	Good	0.8	0.63	73.84	65.44	9.40	9.50	67.3	32,505	279,163
	Classic: Xolalpan 350–550 CE	20741B	LP109_5	Tooth	Good	38.35	30.48	74.13	58.58	4.90	4.80	69.04	179,566	738,802
	Classic: Late Xolalpan 420–550 CE	20734B	LP109_1	Tooth	Good	0.1	0.06	12.21	3.14	9.60	10.30	59.30	15,911	18,745
	Classic: Late Xolalpan– Metepc 420– 650 CE	20736B	LP109_2	Tooth	Good	0.43	0.3	44.09	26.46	8.40	7.90	64.57	191,777	38,906
		20739B	LP109_4	Tooth	Average Dry	0.08	0.04	5.16	1	11.50	12.20	53.51	11,597	7,659
NAA	Late Classic 500-800 CE	24564	L1626	Tooth	-	4.87	3.88	89.31	73.07	0.70	0.80	68.12	1,482	10,363
		24565	L1628	Tooth	-	8.12	6.37	95.17	89.7	5.40	0.70	57.73	4,208	86,046
		24566	L1610	Tooth	-	0.04	0.02	66.87	17.67	3.20	1.10	65.90	1,692	9,341
		24567	L1566	Tooth	-	0.11	0.06	47.33	26.13	0.80	0.60	64.73	1,054	5,244
		24568	L1569	Tooth	-	15.48	12.88	96.86	92.53	8.20	0.50	68.51	15,410	632,088
		24569	L1625	Tooth	-	5.53	4.46	90.39	84.8	9.50	0.60	63.58	30,216	276,372
		24570	L1563x	Tooth	-	2.91	2.32	91.66	79.34	3.50	0.70	63.31	16,288	19,111

Chapter IV

ER	Epiclassic 7th Century CE	24573	L1570	Tooth	-	30.37	24.42	97.43	93.43	8.20	0.60	65.89	22,805	607,468	
		24572	L1612	Tooth	-	0.5	0.38	89.1	76.01	25.30	0.60	55.94	651	48,021	
		24579	L1607	Tooth	-	0.06	0.03	74.92	61.05	12.70	0.80	61.97	1,088	20,505	
		24571	L1585	Tooth	-	0.2	0.14	86.61	72.68	15.80	0.60	61.15	2,754	75,311	
		24577	L1596	Tooth	-	3.34	2.55	95.95	78.77	27.90	0.50	53.33	2,677	269,646	
		24580	L1617	Tooth	-	0	0	/	/	/	/	/	/	/	/
		24583	L1611	Tooth	-	0.29	0.2	85.65	56.47	17.40	0.60	55.84	1,433	49,523	
		24582	L1575	Tooth	-	4.9	3.78	96.64	89.93	10.40	0.60	64.80	25,469	423,751	
		24574	L1618	Tooth	-	0.42	0.3	88.34	79.24	17.00	0.60	60.51	1,135	86,342	
		24576	L1613x	Tooth	-	0.05	0.02	69.1	37.5	17.40	0.80	54.66	1,078	7,772	
		24584	L1592	Tooth	-	0.07	0.04	93.89	33.41	11.20	0.70	63.13	44,426	19,268	
		24578	L1564x	Tooth	-	0.68	0.52	85.12	68.55	2.90	0.80	65.78	1,734	7,664	
		24575_24581	L1605x	Tooth	-	0.2	0.13	88.08	50.41	19.10	0.60	54.70	2,372	115,474	
			L1608	Tooth	-	0.16	0.12	86.48	60.79	19.50	0.50	53.21	4,893		
MP	Middle Postclassic 14th Century CE	Sep-33	A640	Metacarpal	-	10.55	9.04	31.09	23.91	5.00	4.80	95.55	300,128	427,050	
			L1594	Metatarsal	-	0.47	0.37	90.19	78.24	3.00	0.40	68.53	34,381		
		24052	A636	Tooth	-	0.37	0.26	10.44	8.14	6.10	6.20	72.54	263,224	60,206	
		24056	A639	Metatarsal	-	32.79	28.68	61.73	44.78	4.80	4.70	83.75	530,582	536,087	
		24059	L1576	Bone fragment	-	20.56	17.33	94.1	89.21	2.90	0.50	67.39	52,545	587,231	
		Sep-41	A637	Tooth	-	1.09	0.85	25.02	16.18	7.40	7.40	80.06	267,281	506,875	
			A641	Tooth	-	3.19	2.66	24.43	18.1	5.20	5.10	85.35	143,587		
			A638	Metacarpal	-	5.54	4.66	30.77	26.47	5.20	5.20	91.16	292,030		
			L1588	Bone fragment	-	12.37	10.64	90.67	87.81	2.50	0.30	72.01	10,954		
		24053	L1586	Petrous	-	17.57	14.74	90.76	84.84	4.70	0.50	69.11	41,464	729,788	
A635	Petrous		-	1.07	0.84	18.75	13.59	8.00	8.00	65.48	20,440				

Chapter IV

TLA	Late Postclassic 1,350-1,400 CE	25042	LP121_7	Tooth	-	3.45	2.61	76.33	65.43	5.60	5.60	66.12	67,134	308,418
		25043	LP121_8	Tooth	-	15.15	11.73	77.95	71.43	3.60	3.50	68.23	36,113	387,069
		25044	LP121_9	Tooth	-	1.13	0.85	74.81	63.54	4.20	4.30	64.79	26,652	249,421
		25045	LP121_10	Tooth	-	6.17	4.62	83.27	71.93	4.10	4.00	64.94	81,619	431,431
		25046	LP124_10	Tooth	-	3.94	3.05	76	67.62	5.40	5.20	67.40	60,257	314,939
		25047_25048	LP124_11	Tooth	-	34.79	27.39	84.41	75.37	6.00	5.80	66.39	57,665	485,112
			LP121_13	Tooth	-	23.47	18.43	86.73	73.2	1.70	1.80	61.83	55,133	

Table 2. Summary of the studied archaeological samples, genetic sex determination, and uniparental markers haplogroup assignment. ND stands for “Not Detectable”.

Archaeological Site	Time period	Sample ID	Genetic Sex Determination	Mitochondrial Haplogroup	Mitochondrial Contamination Level	Y Chromosome Haplogroup
TEO	Classic: Late Tlamimilolpa–Early Xolalpan 320–420 CE	20745B	M	A2g	ND	Q-L53
		20737B	M	A2+(64)+16189	ND	Q-L54
	Classic: Xolalpan 350–550 CE	20741B	M	A2+(64)+@16111	ND	Q-M971
	Classic: Late Xolalpan–Metepec 420–650 CE	20736B	M	C1b	ND	<i>CT-M168</i>
NAA	Late Classic 500-800 CE	24565	M	A2+(64)+@16111	ND	Q-L53
		24568	F	C1c	ND	-
		24569	F	A2w1*2	ND	-
		24570	F	C1c+195	ND	-
ER	Epiclassic 7th Century CE	24573	F	C1c1b	ND	-
		24577	F	A2	ND	-
		24582	M	B2g2	ND	<i>BT-M42</i>
		24584	F	B2c1	ND	-
		24575_24581	F	B2c1	ND	-
MP	Middle Postclassic 14th Century CE	Sep-33	F	D1d2	0.024	-
		24052	F	A2+(64)	ND	-
		24056	F	C1c6	0.014	-
		24059	F	A2d	ND	-
		Sep-41	F	A2+(64)	0.018	-
		24053	M	A2+(64)	0.061	Q-M3
TLA	Late Postclassic 1,350-1,400 CE	25042	M	B2	ND	Q-M971
		25043	M	A2+(64)+@16111	ND	Q-M930
		25044	M	C1d+194	ND	Q-L54
		25045	F	A2+(64)+16189	ND	-
		25046	M	C1c6	ND	Q-L54
		25047_25048	F	D1i1	ND	-

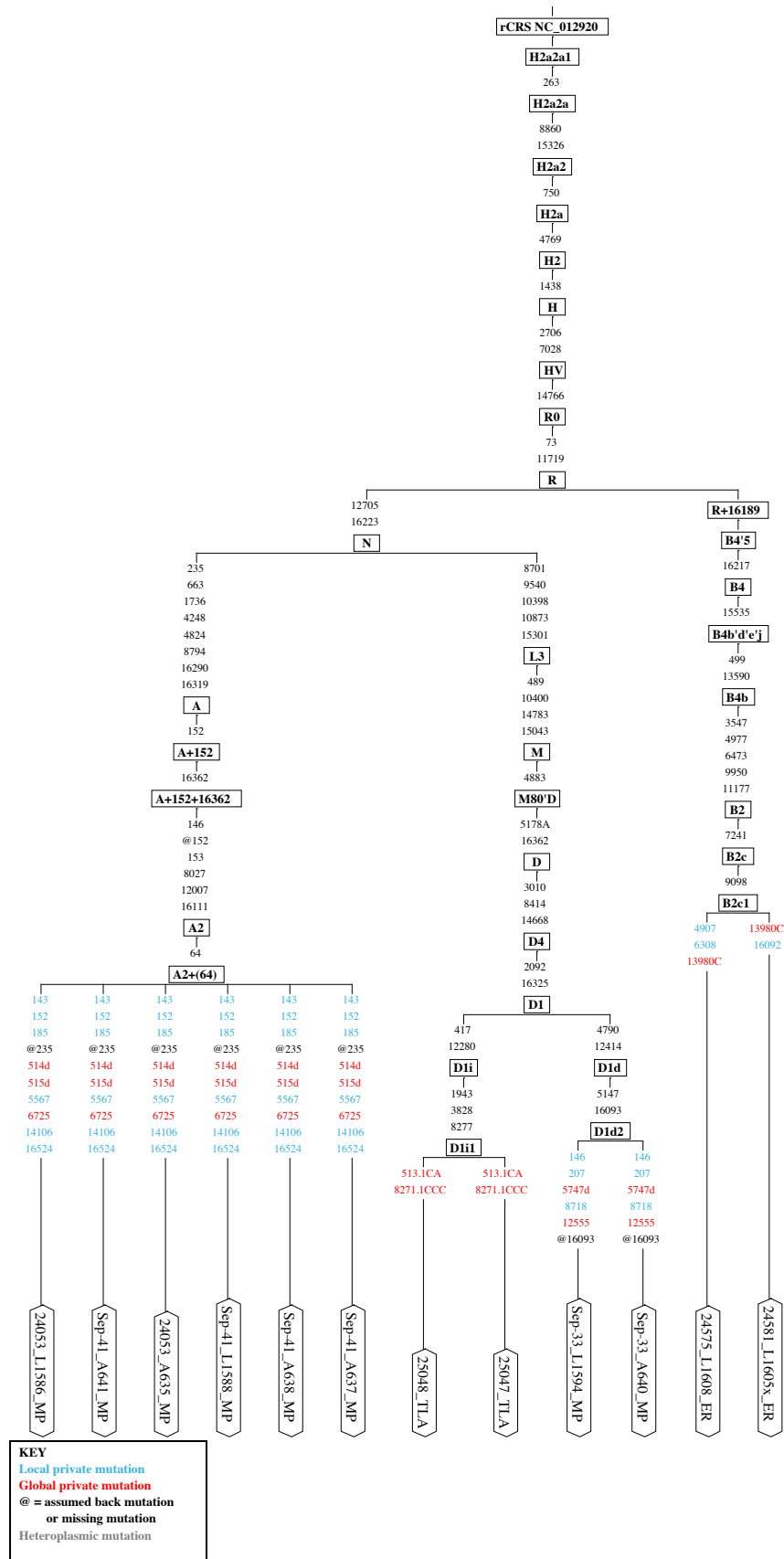


Figure 2. Phylogenetic tree of the mitochondrial sequences from each library.

3.2. Uniparental markers and kinship analysis

The most frequent mitochondrial haplogroup was A2 (48%) (Table 2 and Figure 3), followed by C1 (28%), B2 (16%), and D1 (8%). The other Indigenous American mitochondrial haplogroup, D4h3a, was not found¹.

Concretely, the archaeological site that had the highest frequency of haplogroup A2 was TEO (75%), which also had the only individual in this dataset (20736B) carrying subhaplogroup C1b (25%). Chronologically, the following archaeological site was NAA which presented two A2 subhaplogroups (50%) and two C1c subhaplogroups (50%).

In Michoacán, the highest frequency of haplogroup B2 was found in ER (60%), followed by one individual carrying A2 and another one C1c (20% each). However, in MP, the most frequent haplogroup was A2 (66.67%), followed by subhaplogroup C1c (16.66%) and the first individual in this dataset carrying haplogroup D1 (Sep-33; 16.66%). Interestingly, these frequencies suggest a turnover in mitochondrial frequencies in Michoacán between ER and MP, despite a small sample size.

Finally, TLA was the most diverse archaeological site, as only one haplogroup was found twice (A2; 33.33%). The other individuals carried different subhaplogroups: B2, C1c, C1d, and D1.

All the analysed mitochondrial haplotypes were different. Only the same A2+(64) haplotype was found in two individuals from MP. Notably, none of the samples were related to each other (Figure 4). In fact, the unrooted Neighbour-Joining phylogenetic tree shows an overall high mitochondrial diversity in pre-Columbian Mesoamerica (Figure 5).

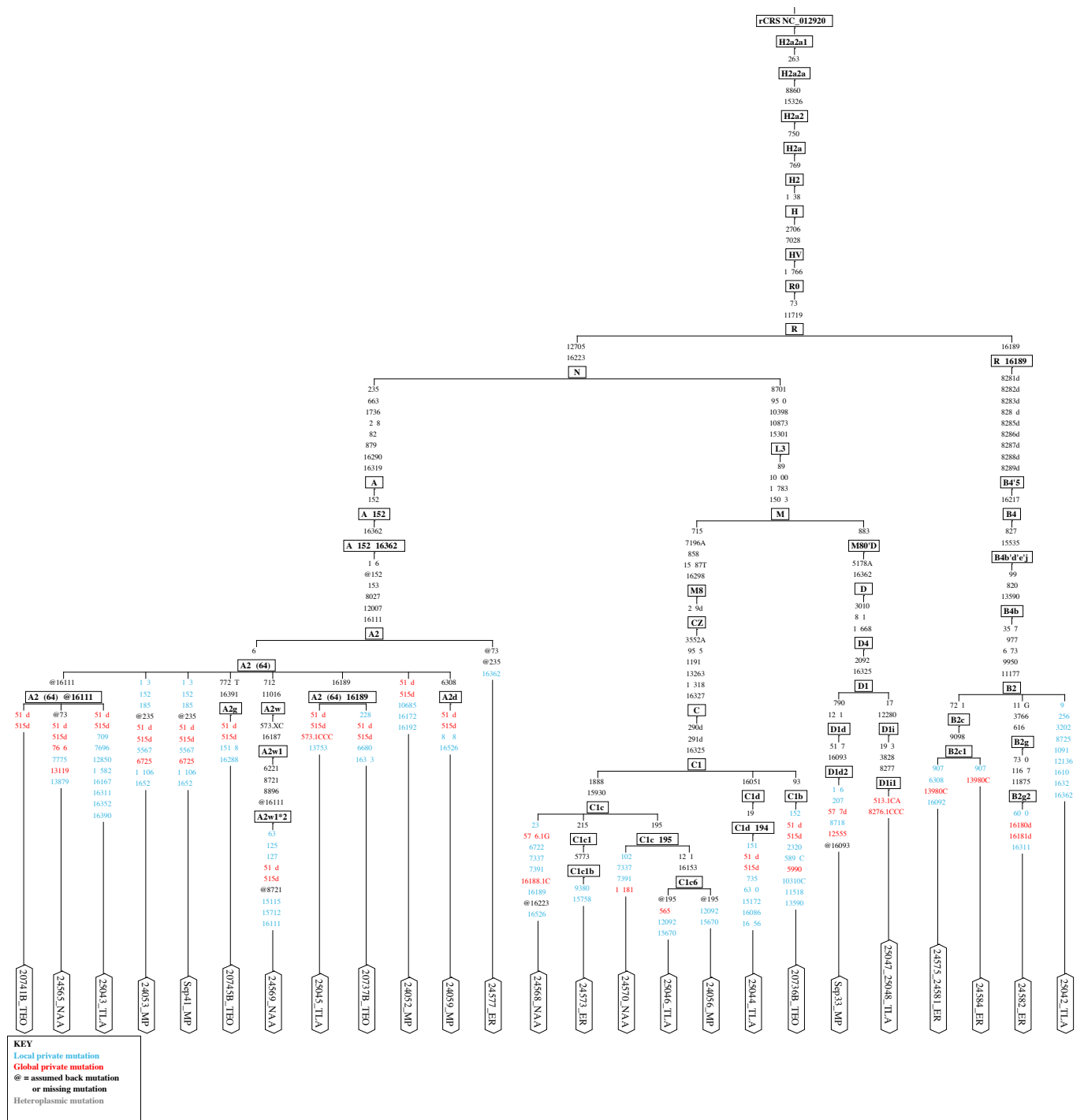


Figure 3. Phylogenetic tree of the mitochondrial sequences from each individual.

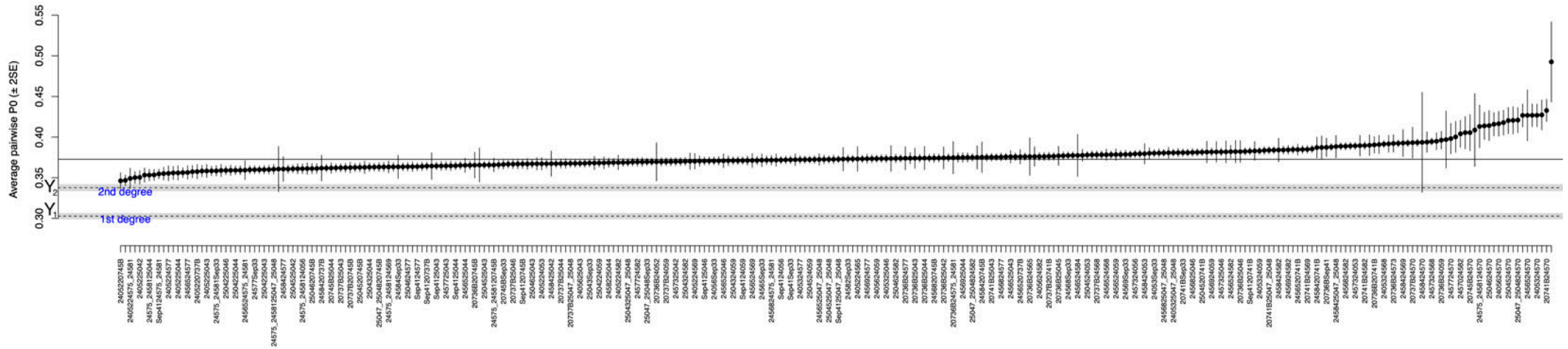


Figure 4. Kinship analysis using READ. Sorted non-normalised average P0 values for all pairwise comparisons between individuals. Error bars show two standard errors. The solid horizontal line indicates the median value used for normalisation. Dashed lines show the cut-offs used to classify the related individuals. The grey areas indicate 95% confidence intervals for the cut-offs, accounting for the uncertainty in estimating the average P0 in the pair used to set the baseline for unrelated individuals. $P0 < Y_1$ indicates two individuals are first degree relatives (parent-offspring or siblings); $Y_1 \leq P0 \leq Y_2$ indicates two individuals are second-degree relatives (i.e., grandchild-grandparent, half-siblings or uncle-aunt to nephew-niece); and if $P0 > Y_2$ the two individuals can be considered unrelated.

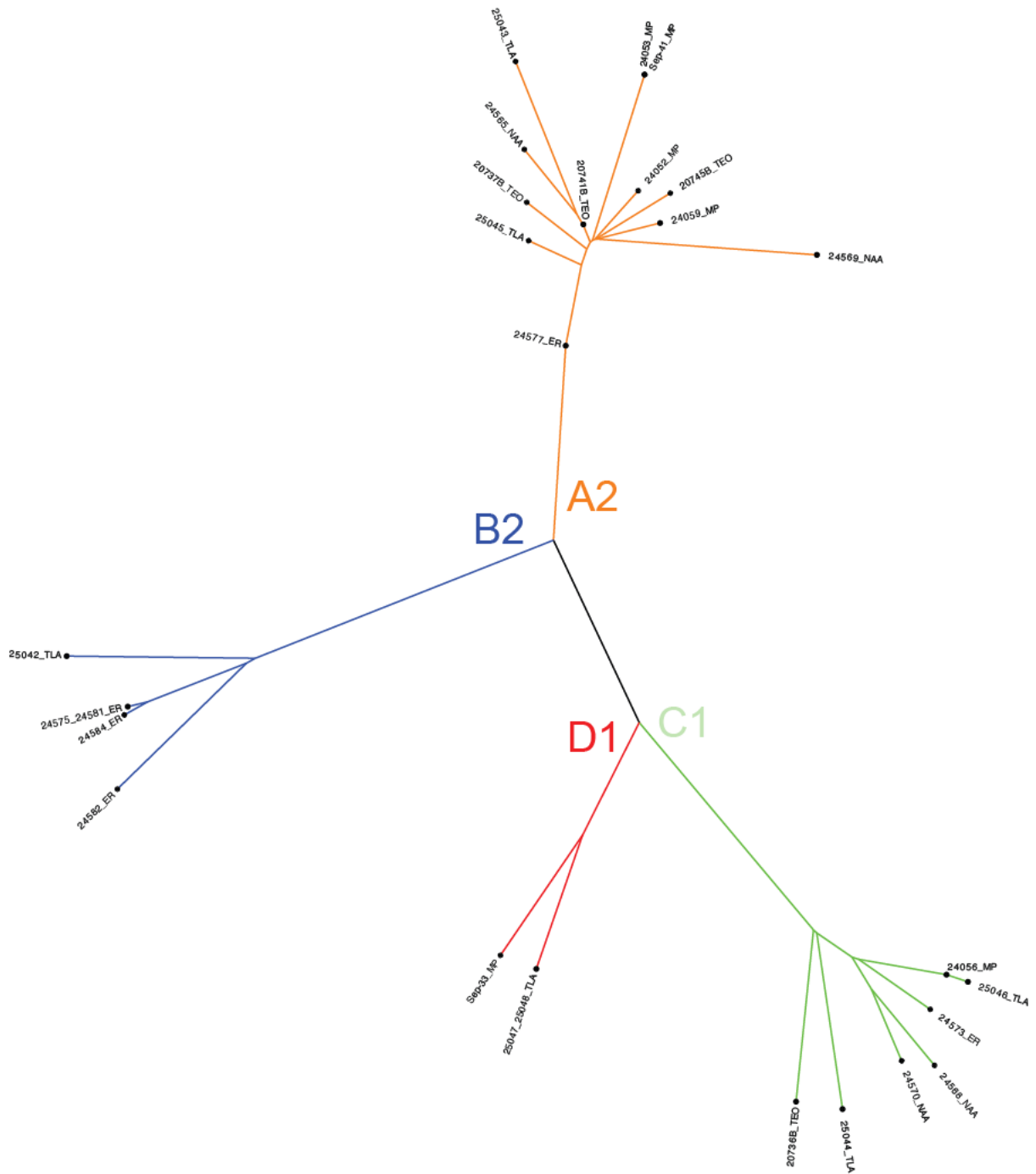


Figure 5. Unrooted Neighbour-Joining tree using Tamura-Nei's genetic distance model with the 25 studied ancient Mesoamericans.

Regarding the Y-chromosome, while I could not assign a more derived Y chromosome haplogroup branch within the Q-L53, Q-L54 and Q-M930 lineages (basal to present-day Eurasians and Americans) and the BT-M42 and CT-M168 lineages (basal to worldwide present-day populations) for most individuals due to low coverage of defining SNPs, the major American founder haplogroup (Q-M3) was found in the only male from MP (24053) and the minor American founder haplogroup (Q-M971) was found in a male from TEO (20741B) and another one from TLA (25042) (Table 2).

3.3. Genetic affinities with worldwide populations

I performed a PCA by projecting ancient samples onto PC axes computed using present-day populations from the HGDP³⁹ (Figure 6) and all samples fell in a cluster of Indigenous Americans, meaning that none of them had any possible admixture coming from Europe. Although this admixture is common in present-day populations, it is not expected in pre-Columbian populations.

3.4. Genetic affinities with American populations

I investigated the genetic ancestry of the 25 studied Mesoamericans performing an unsupervised ADMIXTURE analysis (Figure 7). For comparison purposes, African and Eurasian populations were used as well as 1,096 present-day Indigenous, cosmopolitan, and ancient individuals across the Americas^{16–21,39–59}. ADMIXTURE includes a cross-validation procedure that allows to identify the value of k for which the model has best predictive accuracy. A good value of k exhibits a low cross-validation error compared to other k values. Using the elbow method, I found that the data is best represented at $k=7$ (Figure 8). However, as it has overlapping error bars with $k=6$ and $k=8$, these two should not be completely rejected (Figure 9).

At $k=7$ (Figure 7), most of present-day cosmopolitan populations from Mexico, Puerto Rico, Colombia, and Peru present an admixed genetic ancestry composed by: (i) an Indigenous American ancestry (blue); (ii) a Spanish ancestry (red); and (iii) an African ancestry (orange). In contrast, most present-day Indigenous populations from North America, Mexico, Venezuela-Colombia, the Andes, and Brazil lack high degrees of Spanish and African genetic ancestries. Moreover, the East Asian ancestry (green) is barely found in present-day Indigenous Americans and decreases north-southwards, having its peak in Indigenous populations from North America and ancient individuals from North America, demonstrating the shared genetic history between Siberia and North America along the Bering Strait. Finally, Indigenous Brazilians (Karitiana and Surui) carry a distinct genetic ancestry (brown) that is found in extremely low levels across all present-day and ancient American populations.

Regarding the ancient individuals, two different genetic ancestries found in the Caribbeans are hardly found in other present-day and ancient individuals across the Americas. One being represented by ancient individuals from Cuba (purple) and the other one by ancient individuals from Puerto Rico, Hispaniola, and the Lesser Antilles (yellow). In fact, the “yellow” ancestry is slightly present in ancient Belizeans and Venezuelans and to a lesser degree in ancient South American populations.

Lastly, the studied ancient Mesoamerican individuals did not have any possible admixture coming from Europe. They mainly carried an Indigenous American ancestry (blue) and a slightly higher Caribbean-related genetic ancestry (yellow) compared to Aridoamericans, but lower than Belizeans.

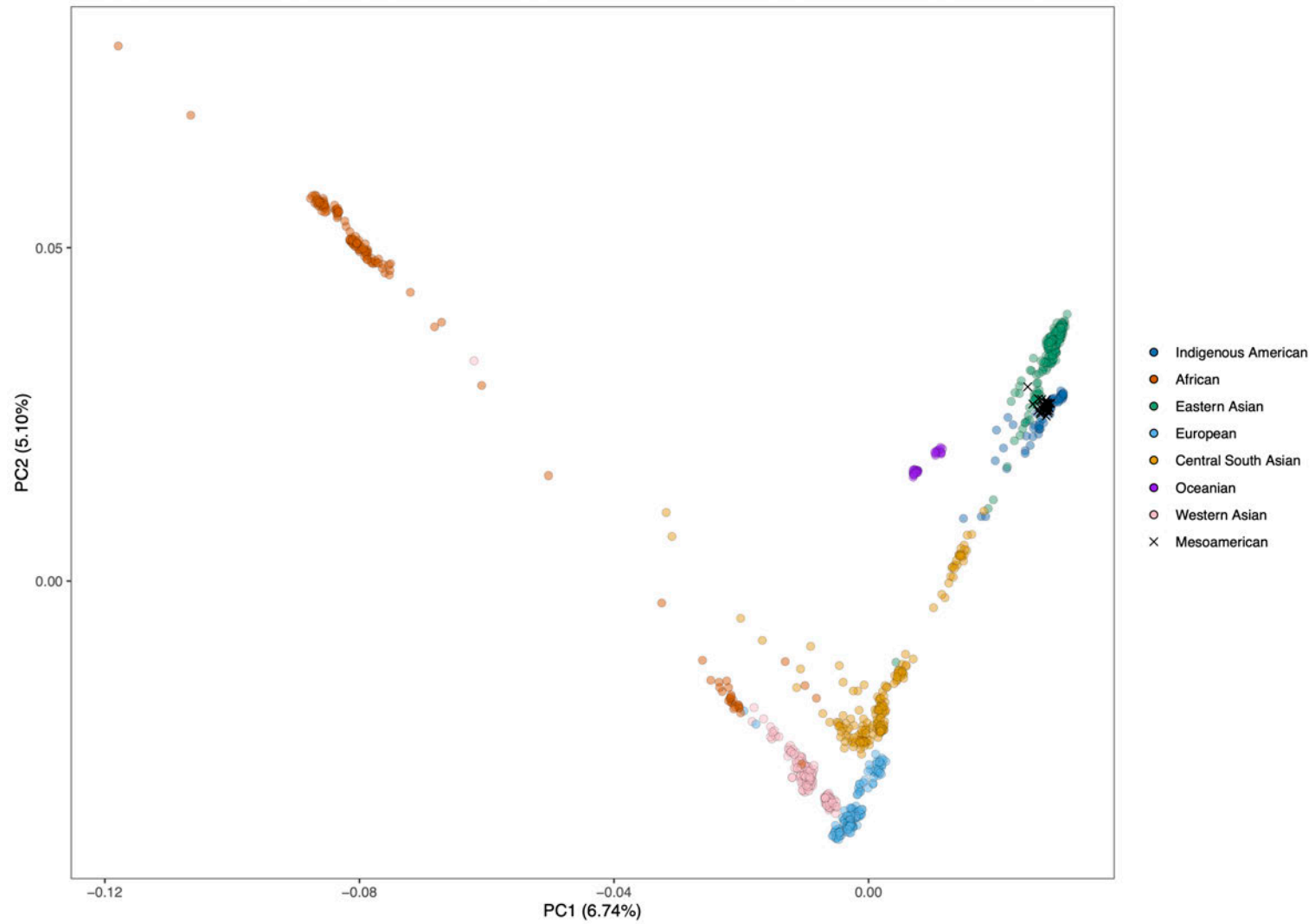


Figure 6. PCA computed using present-day worldwide populations of the HGDP with the studied ancient Mesoamericans (represented with crosses) projected onto PC axes.

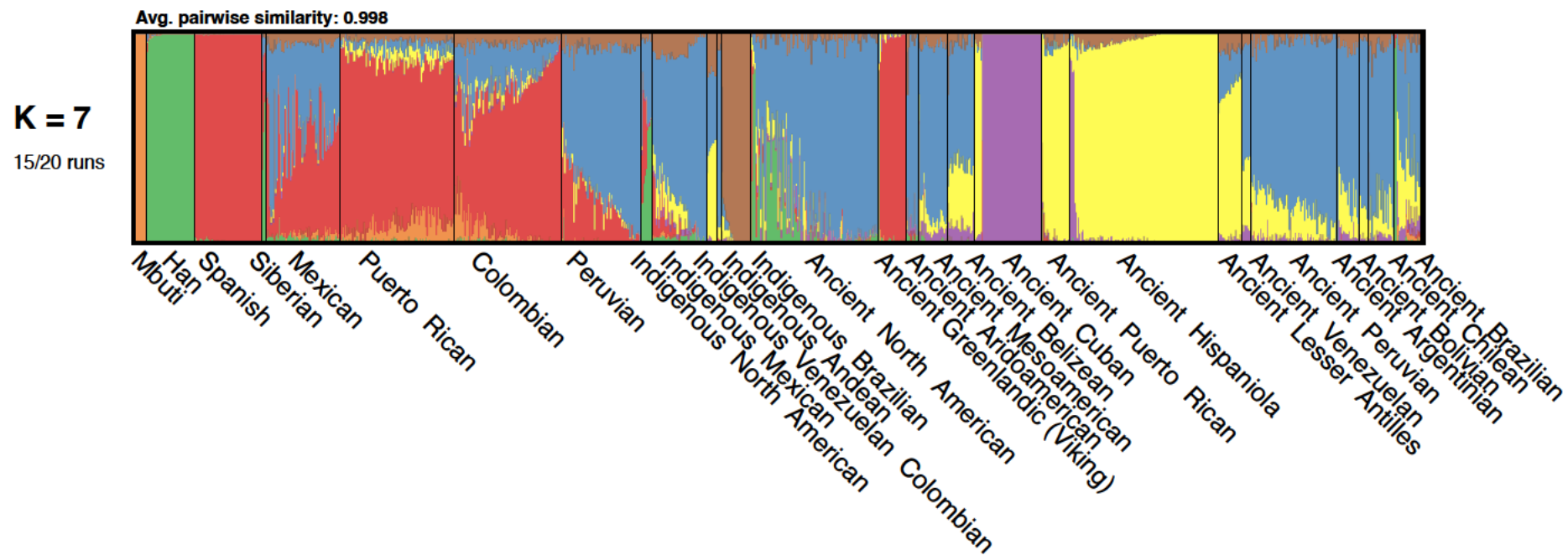


Figure 7. Ancestries estimated with ADMIXTURE at $k=7$, including Africans (Mbuti, orange), East Asians (Han, green), Europeans (Spanish, red) with cosmopolitan and Indigenous populations from Siberia and the Americas (Native Americans, blue; Indigenous Brazilians, brown) as well as ancient Americans (Cuban, purple; Caribbean, yellow).

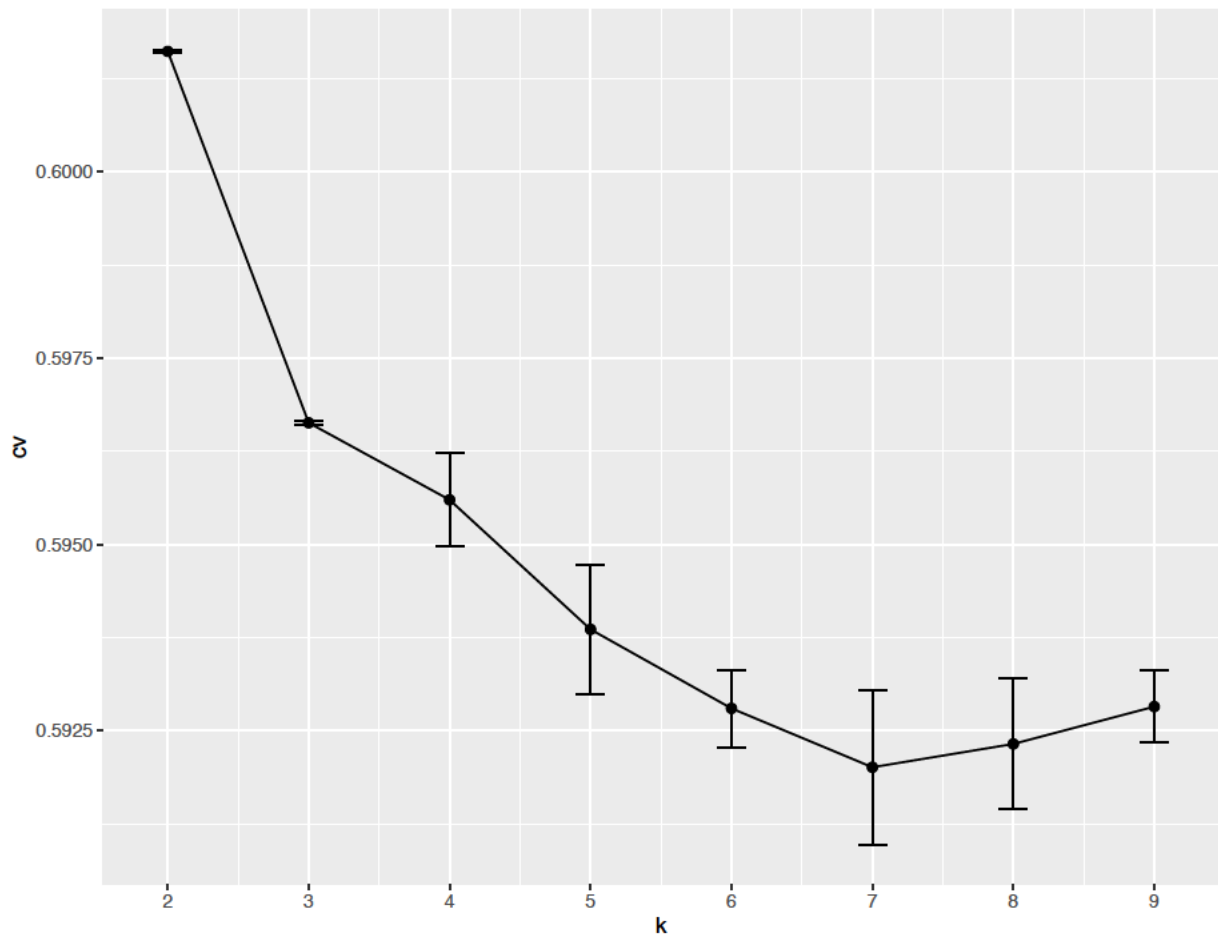


Figure 8. Cross-validation (cv) error plot. Data is best represented at $k=7$, despite having overlapping error bars with $k=6$ and $k=8$.

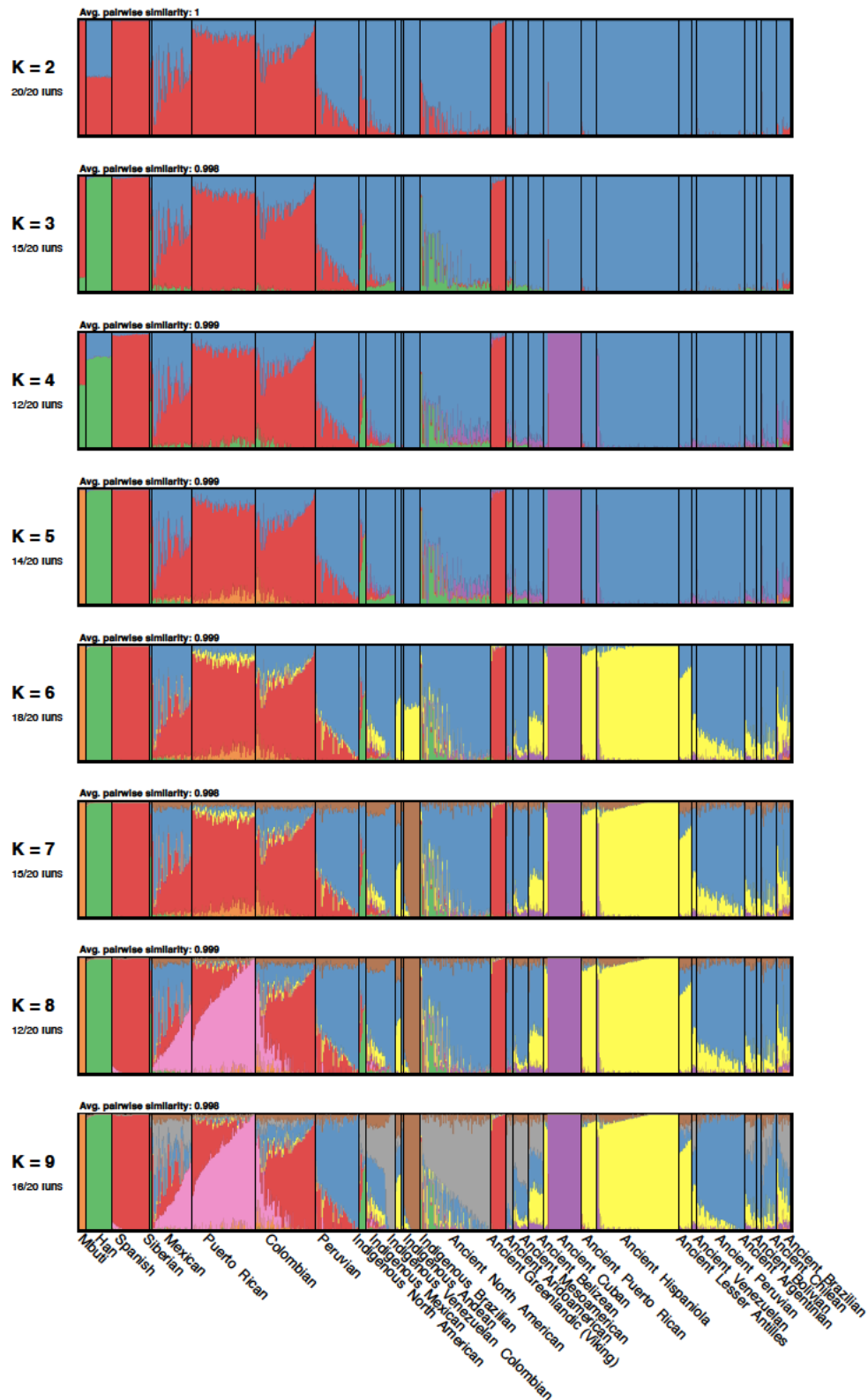


Figure 9. Ancestries estimated with ADMIXTURE at $k=1$ to $k=9$, including Africans, East Asians, Europeans with cosmopolitan and Indigenous populations from Siberia and the Americas as well as ancient Americans.

3.5. Genetic affinities with ancient and present-day Indigenous Central American and northern South American populations

Another PCA was performed (Figure 10) by projecting ancient samples onto PC axes computed using present-day Indigenous populations from northern Mexico (Pima), central Mexico (Huichol), southern Mexico (Zapotec, Mixe, and Mixtec), the Yucatán Peninsula in Mexico (Maya) and the Colombian Amazonian rainforest (Piapoco). PC1 (y-axis) represents a north-south geographical cline including Pima, Huichol, southern Mexicans (Zapotec, Mixe, Mixtec), and Piapoco-Maya. PC2 (x-axis) separates Piapoco and Maya. There are four different clusters of ancient samples: (i) an Aridoamerican-related cluster (Baja Mexico, Pericúes, and the pre-Columbian mummies; represented in shades of blue) falling between present-day Pima and Huichol; (ii) a Mesoamerican-related cluster (ER, MP, TEO, and TLA; represented in shades of pink) falling between Huichol and southern Mexicans (Zapotec, Mixe, and Mixtec) and, notably, the TEO individuals are quite diverse within the cluster; (iii) an Anzick-related cluster falling between Piapoco and southern Mexican-Mayan populations that includes the Anzick individual and the oldest known Central and northern South American samples from Belize and Venezuela (represented in shades of green). This cluster shifts towards present-day southern Mexicans and Mayans in the following order: Las Locas, Anzick and Belize. The younger the Belizean samples are, the closer they are to present-day populations from southern Mexico and the Yucatán Peninsula; and (iv) NAA cluster overlapping the latter cluster but with a higher affinity to present-day Mayans (represented in yellow).

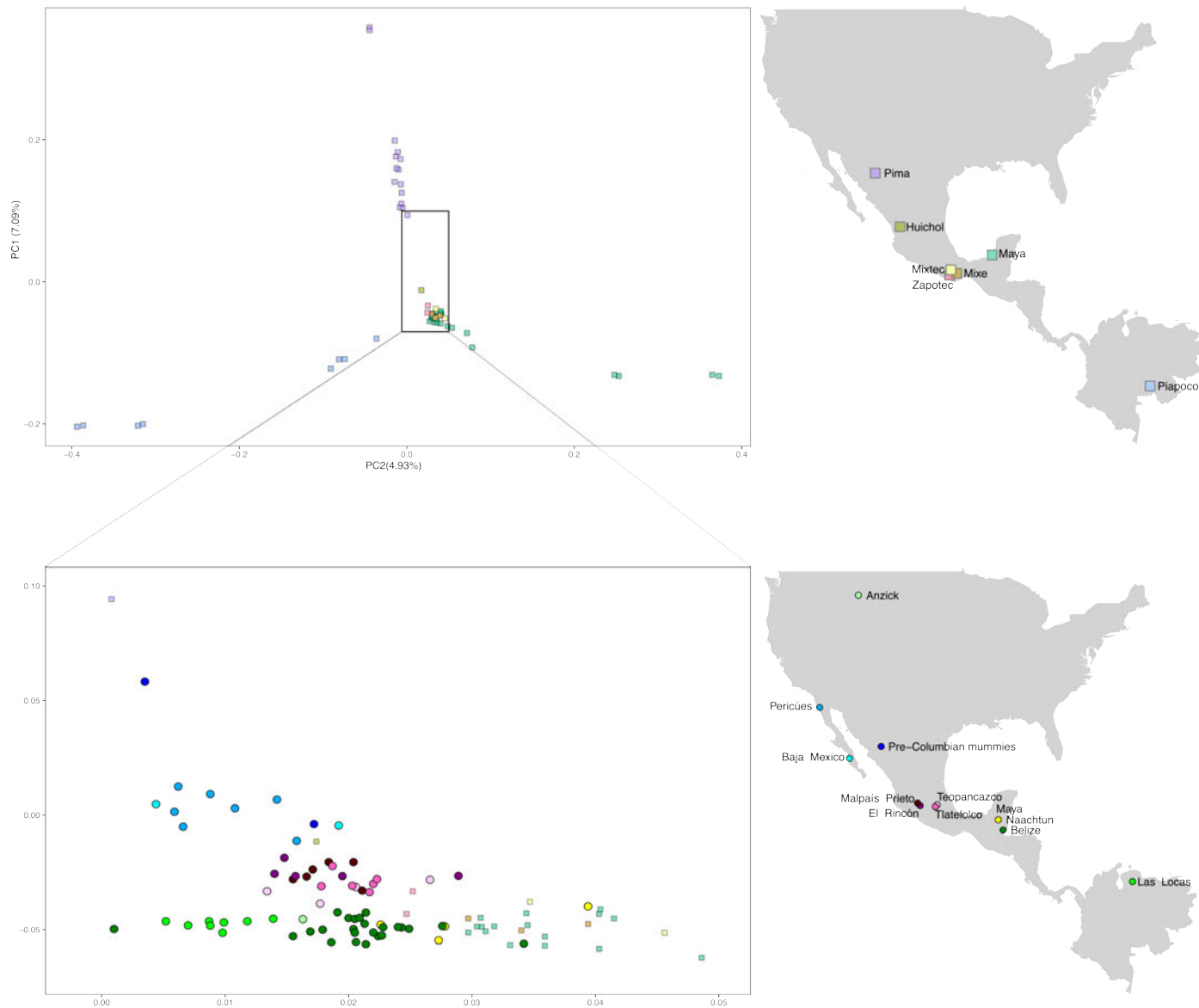


Figure 10. PCA computed using present-day Indigenous populations from Central America and northern South America (Pima, Huichol, Mixe, Mixtec, Zapotec, Maya, and Piapoco) with the ancient individuals projected onto PC axes in the zoomed-in plot. Previously published ancient individuals from northern Mexico are plotted in shades of blue forming an Aridoamerican-related cluster. The studied Mesoamerican samples form a unique cluster below the latter one and are plotted in shades of pink. Previously published individuals from Belize, Las Locas and the Anzick individual are plotted in shades of green forming a cluster containing the oldest samples of the area. The studied Mayan samples from NAA are plotted in yellow and are located close to the latter cluster and present-day Maya.

3.6. Genetic Affinities between ancient Mesoamericans

I explored genetic affinities between the studied ancient Mesoamericans by computing outgroup f_3 -statistics of the form $f_3(\text{Mbuti}; \text{Test}, X)$, where Test refers to one archaeological site used for comparisons with each of the other sites, labelled X (Figure 11). For each Test population, standard deviations strongly overlap means for many comparisons with X, meaning that each f_3 is relatively similar. Therefore, the interpretation solely based on means must be taken with reservations.

In the Basin of Mexico, the Classic TEO shared a higher genetic drift with the Late Postclassic TLA, indicating a possible genetic continuity (Figure 11a). In contrast, the Late Postclassic TLA shared a higher genetic drift with ER. However, it is worth noting that the second highest f_3 value has a higher Z-score and corresponds to the MP individuals from Michoacán, which were relatively contemporaneous to TLA (Figure 11e).

In the Yucatán Peninsula, the Late Classic NAA shared a higher genetic drift with the ER individuals from Michoacán, which were relatively contemporaneous to NAA. However, all the f_3 values computed with NAA in the Test position are slightly less supported (lower Z-scores) (Figure 11b).

Finally, both ER and MP shared a higher genetic drift with each other (Figure 11c and Figure 11d), suggesting a possible genetic continuity in Michoacán from the Epiclassic to the Middle Postclassic.

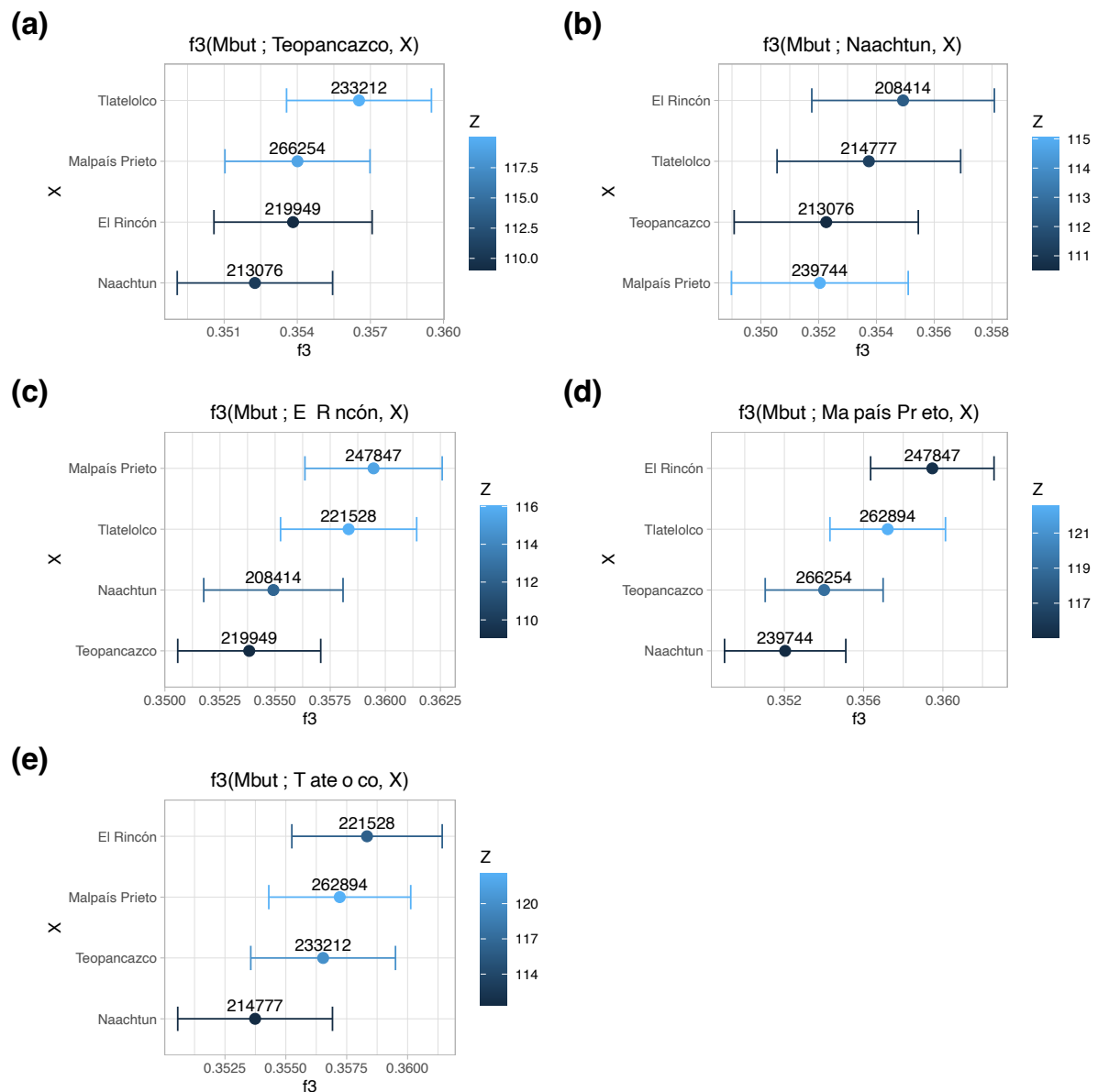


Figure 11. Outgroup f_3 -statistics of the form $f_3(\text{Mbuti}; \text{Test}, X)$ demonstrate the amount of shared drift between pairwise ancient Mesoamericans with Mbuti as the outgroup. The x-axis represents the f_3 -statistic values. Results are represented as mean \pm 1 SD and colours represent Z-scores. Numbers above each dot are the numbers of SNPs used to estimate the f_3 -statistics.

4. Discussion

Recent advances in the recovery of authentic aDNA and the application of high-throughput DNA sequencing technologies are enabling the reconstruction of human history in tropical environments^{19,44,45,61}, opening the door to a deep exploration of ancient Mesoamerica.

4.1. aDNA preservation in Mesoamerica

I performed a general quality screening of the aDNA retrieved from Mesoamerican skeletal remains. Almost 60% of the Mesoamerican individuals passed good quality thresholds for population genetic analyses. The success rate in each archaeological site was: TEO (40%), NAA (57.1%), ER (35.5%), MP (100%) and TLA (100%). While the low success rate in TEO is due to our sample collection purposefully representing different preservation degrees at the archaeological site, in NAA and ER it may be due to general poor taphonomy.

For future studies, the following advices should be kept in mind: (i) skeletal remains covered with cinnabar or presenting mould stains, dryness or a general observable bad preservation should be avoided; (ii) while good quality aDNA was retrieved from teeth and petrous bones as expected^{62,63}, metacarpals performed surprisingly better than teeth which is in discordance with previous investigations⁶⁴; (iii) single-stranded DNA libraries had higher percentages of endogenous DNA and retained shorter DNA fragments, but presented lower and asymmetrical rates of post-mortem damage at read termini as previously shown⁶⁵; and (iv) replicates are important for reproducibility and authenticity.

4.2. Genetic history of pre-Columbian Mesoamerica

This study offers the first Mesoamerican aDNA genetic dataset that includes full mitogenomes and nuclear data. These preliminary results suggest that pre-Columbian Mesoamerica had a

high mitochondrial diversity. The most frequent mitochondrial haplogroup was A2 (48%) which is in accordance with previously published Mesoamerican aDNA data based on RFLP analyses and Sanger-sequenced HVR1¹. The frequencies of the other mitochondrial haplogroups were C1 (28%), B2 (16%), and D1 (8%); D4h3a was not found. All the analysed mitochondrial haplotypes were different—except one found in two individuals in MP—and none of the samples were related to each other.

The nuclear data evidenced that none of these individuals had any possible admixture coming from Europe, as expected in pre-Columbian populations. However, this admixture is common in present-day populations—especially in cosmopolitan populations and to a lesser degree in Indigenous populations—due to the Spanish conquest of Mexico, the subsequent colonial period, and the current globalisation of the world.

Nuclear genetic differences between present-day Indigenous populations as well as ancient Mesoamericans suggest a geographical correlation probably explained by isolation by distance. The studied Mesoamerican individuals from the Basin of Mexico (TEO and TLA) and Michoacán (ER and MP) had equidistant genetic affinities with present-day Huichol and southern Mexicans (Zapotec, Mixe, and Mixtec) as well as Mayans. In contrast, NAA presented a higher genetic affinity with present-day Mayans. Moreover, NAA also presented some genetic affinities with a group of ancient individuals that predates the settlement of Mesoamerica and includes the Anzick individual¹⁶ and the oldest known Central and northern South American samples from Belize^{17,18} and Venezuela¹⁹. This group of ancient individuals were part of a homogenous population that expanded southwards carrying the primary source of the ancestry of Central and South Americans, which was partially lost around 9,000 years ago due to a population replacement that carried the predominant ancestry currently found in South America¹⁷. However, genetically, little is known about the early dispersals in Central

America and even less during the first and a half millennium CE, when Mesoamerica was peopled by large intermingled cultural groups connected by trade and agriculture. The data presented in this chapter are a first attempt at filling this knowledge gap. Our paleogenetic analyses suggest that in the Basin of Mexico, TEO and TLA were genetically quite diverse and a likely product of genetic continuity from the Classic to the Postclassic. In contrast, in Michoacán, mitochondrial haplogroup frequencies suggest a possible turnover of lineages between the Epiclassic to the Middle Postclassic, which could possibly be due to the arrival of the Uacúsechas and Chichimecas from northern Mexico to the region¹². However, nuclear data suggest a possible genetic continuity. Differences between analyses using mitochondrial and nuclear data are due to stochastic factors associated with demography. That is why some genetic markers reflect population history, and some not. Even though there are limitations associated with the study of mtDNA as this genetic marker only represents the evolutionary history of the female population at a single locus, it does represent a fundamental element of a population's heritage and evolutionary history^{66,67}. Notably, interpretation solely based on these preliminary results must be taken with reservations due to the small sample size.

4.3. aDNA research in Mexico

As previously mentioned, this project suffered several setbacks due to the COVID-19 pandemic, such as the inability to travel to Mexico for sample collection and community engagement. However, consultation with present-day Indigenous populations for destructive analysis of human archaeological remains is not a standard procedure in Mexico. Assigning a present-day population as the single direct descendant of human archaeological remains is complex. This is because of socio-political and historical issues as well as the institutionalised imposition of the “mestizaje” narrative to minimise the voice of present-day Indigenous populations^{68,69}. Even though some recent ethical guidelines regarding worldwide DNA

research on human archaeological remains have been raised^{70,71}, an essential discussion that warrants aDNA research in Mexico is needed⁷². Consequently, the potential publication of the data presented in this chapter will depend on discussions I currently have with colleagues in Mexico, whom guidance will be critical.

4.4. Future perspectives

Further analyses could be performed to better explore the genetic history of pre-Columbian Mesoamerica. Firstly, the ancient mitochondrial genomes could be compared with present-day mitochondrial genomes to determine if the discovered genetic diversity is still part of the current genetic diversity or if it was lost during the Spanish conquest and the subsequent colonisation. Secondly, the Y chromosome haplogroups could be better assigned calculating and recording the number of derived and ancestral SNP calls as described in Rohrlach et al. (2021)⁷³ and/or running a more updated software such as HaploGrouper (https://gitlab.com/bio_anth_decode/haploGrouper). Thirdly, new paleogenomic methods have been developed to increase the resolution of kinship analyses^{74,75}, which is of relevance in this project to explore other kinship relationships beyond first- and second-degree relationships. Fourthly, the genetic affinities between ancient Mesoamericans could be further explored computing outgroup f_3 -statistics between them and observe the resulting pairwise matrix using a heatmap. Fifthly, admixture modelling could be explored with *qpAdm*⁷⁶, *qpWave* and *qpGraph*⁶⁰ to: (i) determine the magnitude, direction and number of gene flows required to model the relationships between the studied archaeological sites; (ii) identify the contribution of UpopA (a previously reported unsampled ghost population)⁵¹; and (iii) reconstruct the genetic relationships between ancient Americans through a phylogenetic tree allowing for the addition of admixture events. Sixthly, a re-exploration of some analyses could be performed including individuals from the Aridoamerican-Mesoamerican border region

recently reported in a pre-print by Villa-Islas et al. (2022)⁷⁷, when it successfully undergoes peer-review and is published. Lastly, further paleogenomic investigations across pre-Columbian Mesoamerica are needed to unravel the genetic history of this fascinating cultural region that is a key study area to gain a comprehensive understanding of the peopling and evolution of the Americas.

5. References

1. Roca-Rada X, Souilmi Y, Teixeira JC, Llamas B. Ancient DNA Studies in Pre-Columbian Mesoamerica. *Genes* (Basel) 2020;11(11):1346.
2. Smith ME. Mesoamerica's First World City. Teotihuacan in Comparative Perspective. In: Carballo DM, Hirth KG, Arroyo B, editors. *Teotihuacan: The World Beyond the City*. Dumbarton Oaks, Washington, DC.; 2020. p. 33–56.
3. Aguirre-Samudio AJ, González-Sobrino BZ, Álvarez-Sandoval BA, Montiel R, Serrano-Sánchez C, Meza-Peñaloza A. Genetic history of classic period Teotihuacan burials in Central Mexico. *Rev Argent Antropol Biol* 2017;19(1):1–14.
4. Álvarez-Sandoval BA, Manzanilla LR, González-Ruiz M, Malgosa A, Montiel R. Genetic evidence supports the multiethnic character of teopancazco, a neighborhood center of teotihuacan, Mexico (ad 200-600). *PLoS One* 2015;10(7):1–19.
5. Manzanilla LR. *Multiethnicity and Migration at Teopancazco*. University Press of Florida; 2017.
6. Nondédéo P, Lacadena García-Gallo A, Cases Martín JI. Teotihuacanos y Mayas En La 'Entrada' de 11 EB' (378 d.C.): Nuevos Datos de Naachtun, Peten, Guatemala. *Revista Española de Antropología Americana* 2019;49:53–75.
7. Nondédéo P. Naachtun: organisation, essor et histoire d'une capitale régionale maya. *Compte rendus de l'Académie des Inscriptions et Belles-Lettres* 2016 2017;juillet-octobre(fascicule 3):1211–28.
8. Goudiaby H, Nondédéo P. The funerary and architectural history of an ancient Maya residential group: Group 5N6, Naachtun, Guatemala. 2020.
9. Pétrequin P. Ocho mil años de la Cuenca de Zacapú: Evolución de los paisajes y primeros desmontes. Mexico: Centro de estudios mexicanos y centroamericanos; 1994.
10. Carot P. Arqueología de Michoacán: nuevas aportaciones de la historia Purhépecha. In: B. Braniff, editor. *Introducción a la arqueología de Occidente de México*. Colima, México: Universidad de Colima - INAH; 2004. p. 443–74.
11. Pereira G, Padilla Gutiérrez EF. *La ciudad perdida. Raíces de los soberanos tarascos*. CONACULTA-INAH. 2018.
12. Migeon G. *Residencias y Estructuras Civico-Ceremoniales Posclásicas Tarascas de la Región de Zacapu (Michoacán, México)*. Oxford: BAR; 2015.
13. López-Austin A, López-Luján L. *El pasado indígena*. Mexico, MX: El Colegio de México, Fondo de Cultura Económica.; 1996.
14. Kemp BM, Resendéz A, Román Berrelleza JA, Malhi R, Smith DG. An analysis of ancient Aztec mtDNA from Tlatelolco: Pre-Columbian relations and the spread of Uto-Aztecan. *Biomolecular Archaeology: Genetic Approaches to the Past* 2005;22–46.
15. Solorzano Navarro E. *De La Mesoamérica Prehispanica a la Colonial: La huella del DNA antiguo*. 2006.
16. Rasmussen M, Anzick SL, Waters MR, et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 2014;506(7487):225–9.
17. Posth C, Nakatsuka N, Lazaridis I, et al. Reconstructing the Deep Population History of Central and South America. *Cell* 2018;175(5):1185–97.

18. Kennett DJ, Lipson M, Prufer KM, et al. South-to-north migration preceded the advent of intensive farming in the Maya region. *Nat Commun* 2022;13(1).
19. Fernandes DM, Sirak KA, Ringbauer H, et al. A genetic history of the pre-contact Caribbean. *Nature* 2021;590(7844):103–10.
20. Raghavan M, Steinrücken M, Harris K, et al. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science (1979)* 2015;349(6250).
21. Scheib CL, Li H, Desai T, et al. Ancient human parallel lineages within North America contributed to a coastal expansion. *Science (1979)* 2018;360(6392):1024–7.
22. Llamas B, Willerslev E, Orlando L, Orlando L. Human evolution: a tale from ancient genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2017;372(1713).
23. Llamas B, Valverde G, Fehren-Schmitz L, Weyrich LS, Cooper A, Haak W. From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *Sci Technol Archaeol Res.* 2017;3(1):1–14.
24. Damgaard PB, Margaryan A, Schroeder H, Orlando L, Willerslev E, Allentoft ME. Improving access to endogenous DNA in ancient bones and teeth. *Sci Rep* 2015;5(11184):1–12.
25. Dabney J, Knapp M, Glocke I, et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences* 2013;110(39):15758–63.
26. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010;5(6).
27. Kapp JD, Green RE, Shapiro B. A fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. *Journal of Heredity* 2021;112(3):241–9.
28. Mathieson I, Lazaridis I, Rohland N, et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 2015;528(7583):499–503.
29. Fellows Yates JA, Lamnidis TC, Borry M, et al. Reproducible, portable, and efficient ancient genome reconstruction with nfcore/eager. *PeerJ* 2021;9:1–25.
30. Neukamm J, Peltzer A, Nieselt K. DamageProfiler: fast damage pattern calculation for ancient DNA. *Bioinformatics* 2021;37(April):3652–3.
31. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *Gigascience* 2021;10(2).
32. Poznik GD. Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. *bioRxiv* 2016;088716.
33. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 1999;23(2):147.
34. Weissensteiner H, Pacher D, Kloss-Brandstätter A, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res* 2016;44.

35. Kloss-Brandstätter A, Pacher D, Schönherr S, et al. HaploGrep: A fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat* 2011;32(1):25–32.
36. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 2009;30(2):386–94.
37. Weissensteiner H, Forer L, Fendt L, et al. Contamination detection in sequencing studies using the mitochondrial phylogeny. *Genome Res* 2021;31(2):309–16.
38. Kuhn JMM, Jakobsson M, Günther T. Estimating genetic kin relationships in prehistoric populations. *PLoS One* 2018;13(4).
39. Bergstrom A, McCarthy SA, Hui R, et al. Insights into human genetic variation and population history from 929 diverse genomes. 2019;1–23.
40. Skoglund P, Mallick S, Bortolini MC, et al. Genetic evidence for two founding populations of the Americas. *Nature* 2015;525(7567):104–8.
41. Mallick S, Li H, Lipson M, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 2016;538(7624):201–6.
42. Lazaridis I, Nadel D, Rollefson G, et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* 2016;536(7617):419–24.
43. Margaryan A, Lawson DJ, Sikora M, et al. Population genomics of the Viking world. *Nature* 2020;585(7825):390–6.
44. Nägele K, Posth C, Orbeagoza MI, et al. Genomic insights into the early peopling of the Caribbean. *Science* (1979) 2020;8697(June).
45. Nieves-Colón MA, Pestle WJ, Reynolds AW, et al. Ancient DNA Reconstructs the Genetic Legacies of Precontact Puerto Rico Communities. *Mol Biol Evol* 2020;37(3):611–26.
46. Bongers JL, Nakatsuka N, O’shea C, et al. Integration of ancient DNA with transdisciplinary dataset finds strong support for Inca resettlement in the south Peruvian coast.
47. de la Fuente C, Ávila-Arcos MC, Galimany J, et al. Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. *Proceedings of the National Academy of Sciences* 2018;115(17):E4006–12.
48. Nakatsuka N, Lazaridis I, Barbieri C, et al. A Paleogenomic Reconstruction of the Deep Population History of the Andes. *Cell* 2020;181(5):1131-1145.e21.
49. Lindo J, Achilli A, Perego UA, et al. Ancient individuals from the North American Northwest Coast reveal 10,000 years of regional genetic continuity. *Proc Natl Acad Sci U S A* 2017;114(16):4093–8.
50. Flegontov P, Altınışık NE, Changmai P, et al. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* 2019;570(7760):236–40.
51. Moreno-Mayar JV, Vinner L, de Barros Damgaard P, et al. Early human dispersals within the Americas. *Science* (1979) 2018;362(6419).
52. Lindo J, Haas R, Hofman C, et al. The genetic prehistory of the Andean highlands 7000 years BP though European contact. *Sci Adv* 2018;4.
53. Damgaard PDB, Marchi N, Rasmussen S, et al. 137 ancient human genomes from across the Eurasian steppes. *Nature* 2018;557:369–74.

54. Moreno-Mayar JV, Potter BA, Vinner L, et al. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* 2018;553(7687):203–7.
55. Kennett DJ, Plog S, George RJ, et al. Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat Commun* 2017;8.
56. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74.
57. Rasmussen M, Sikora M, Albrechtsen A, et al. The ancestry and affiliations of Kennewick Man. *Nature* 2015;523(7561):455–8.
58. Malaspinas A-S, Lao O, Schroeder H, et al. Two ancient human genomes reveal Polynesian ancestry among the indigenous Botocudos of Brazil. *Current Biology* 2014;24(21):R1035–7.
59. Rasmussen M, Li Y, Lindgreen S, et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 2010;463(7282):757–62.
60. Patterson N, Moorjani P, Luo Y, et al. Ancient admixture in human history. *Genetics* 2012;192(3):1065–93.
61. Morales-Arce AY, Snow MH, Kelley JH, Anne Katzenberg M. Ancient mitochondrial DNA and ancestry of Paquimé inhabitants, Casas Grandes (A.D. 1200–1450). *Am J Phys Anthropol* 2017;163(3):616–26.
62. Hansen HB, Damgaard PB, Margaryan A, et al. Comparing ancient DNA preservation in petrous bone and tooth cementum. *PLoS One* 2017;12(1).
63. Pinhasi R, Fernandes D, Sirak K, et al. Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS One* 2015;10(6).
64. Parker C, Rohrlach AB, Friederich S, et al. A systematic investigation of human DNA preservation in medieval skeletons. *Sci Rep* 2020;10(1).
65. Kapp JD, Green RE, Shapiro B. A fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. *Journal of Heredity* 2021;112(3):241–9.
66. Lalueza-Fox C. Agreements and misunderstandings among three scientific fields: Paleogenomics, archaeology, and human paleontology. *Curr Anthropol* 2013.
67. Rubinoff D, Holland BS. Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. *Syst Biol* 2005;54(6):952–61
68. Swanson P. *The Companion to Latin American Studies*. London: Routledge; 2003.
69. Bracho J. NARRATIVA E IDENTIDAD. EL MESTIZAJE Y SU REPRESENTACIÓN HISTORIOGRÁFICA. *Latinoamérica Revista de estudios Latinoamericanos* 2009;48:55–86.
70. Alpaslan-Roodenberg S, Anthony D, Babiker H, et al. Ethics of DNA research on human remains: five globally applicable guidelines. *Nature* 2021.
71. Wagner JK, Colwell C, Claw KG, et al. Fostering Responsible Research on Ancient DNA. *The American Journal of Human Genetics* 2020;107(2):183–95.
72. Ávila-Arcos MC, de la Fuente Castro C, Nieves-Colón MA, Raghavan M. Recommendations for Sustainable Ancient DNA Research in the Global South: Voices From a New Generation of Paleogenomicists. *Front Genet* 2022;13(April):1–8.

73. Rohrlach AB, Papac L, Childebayeva A, et al. Using Y-chromosome capture enrichment to resolve haplogroup H2 shows new evidence for a two-path Neolithic expansion to Western Europe. *Sci Rep* 2021;11(1):1–11.
74. Fowler C, Olalde I, Cummings V, et al. A high-resolution picture of kinship practices in an Early Neolithic tomb. *Nature* 2022;601(7894):584–7.
75. Hanghøj K, Moltke I, Andersen PA, Manica A, Korneliussen TS. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience* 2019;8(5).
76. Harney É, Patterson N, Reich D, Wakeley J. Assessing the performance of qpAdm: A statistical tool for studying population admixture. *Genetics* 2021;217(4).
77. Villa-Islas V, Izarraras-Gomez A, Larena M, et al. Demographic history and genetic structure in pre-Hispanic Central Mexico. *bioRxiv* 2022;2022.06.19.496730.

Part III: Southern Cone of South America

Chapter V: Ancient mitochondrial genomes from the Argentinian Pampas inform the early peopling of the Southern Cone of South America

Statement of Authorship

Title of Paper	Ancient mitochondrial genomes from the Argentinian Pampas inform the early peopling of the Southern Cone of South America
Publication Status	<input checked="" type="checkbox"/> Published <input type="checkbox"/> Accepted for Publication <input type="checkbox"/> Submitted for Publication <input type="checkbox"/> Unpublished and Unsubmitted work written in manuscript style
Publication Details	Roca-Rada X, Politis G, Messineo PG, et al. Ancient mitochondrial genomes from the Argentinian Pampas inform the early peopling of the Southern Cone of South America. iScience 2021;102553.

Principal Author

Name of Principal Author (Candidate)	Xavier Roca Rada			
Contribution to the Paper	Dry lab genetic analyses, interpretation, and drafted the manuscript			
Overall percentage (%)	50%			
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.			
Signature	<table border="1" style="width: 100%;"> <tr> <td style="width: 80%;"></td> <td style="width: 20%;">Date</td> <td>05/10/2022</td> </tr> </table>		Date	05/10/2022
	Date	05/10/2022		

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

- i. the candidate's stated contribution to the publication is accurate (as detailed above);
- ii. permission is granted for the candidate to include the publication in the thesis; and
- iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Gustavo Politis			
Contribution to the Paper	I provided archaeological samples, radiocarbon dates, contextual information and drafted the manuscript together with the main author and other co-authors			
Signature	<table border="1" style="width: 100%;"> <tr> <td style="width: 80%;"></td> <td style="width: 20%;">Date</td> <td>05/10/2022</td> </tr> </table>		Date	05/10/2022
	Date	05/10/2022		

Name of Co-Author	Pablo G. Messineo			
Contribution to the Paper	I provided archaeological samples, radiocarbon dates, contextual information, and interpreted the genetic results			
Signature	<table border="1" style="width: 100%;"> <tr> <td style="width: 80%;"></td> <td style="width: 20%;">Date</td> <td>05/10/2022</td> </tr> </table>		Date	05/10/2022
	Date	05/10/2022		

Name of Co-Author	Clara Scabuzzo			
Contribution to the Paper	provided archaeological samples, radiocarbon dates, and contextual information			
Signature	<table border="1" style="width: 100%;"> <tr> <td style="width: 80%;"></td> <td style="width: 20%;">Date</td> <td>26/10/2022</td> </tr> </table>		Date	26/10/2022
	Date	26/10/2022		

Name of Co-Author	Nahuel A. Scheifler		
Contribution to the Paper	I provided archaeological samples, radiocarbon dates, contextual information and interpreted the genetic results.		
Signature		Date	17/10/2022

Name of Co-Author	Mariela E. González		
Contribution to the Paper	I provided archaeological samples, radiocarbon dates, and contextual information		
Signature		Date	6/10/2022

Name of Co-Author	<i>Kelly M Harkins</i>		
Contribution to the Paper	<i>Wetlab work + some drylab data processing.</i>		
Signature		Date	<i>10/6/22</i>

Name of Co-Author	David Reich		
Contribution to the Paper	Provision of data		
Signature		Date	October 17, 2022

Name of Co-Author	Dr Yassine Souilmi		
Contribution to the Paper	Co supervised the Mr Roca Rada, helped interpret the results, and provided input on the manuscript.		
Signature		Date	5/10/2022

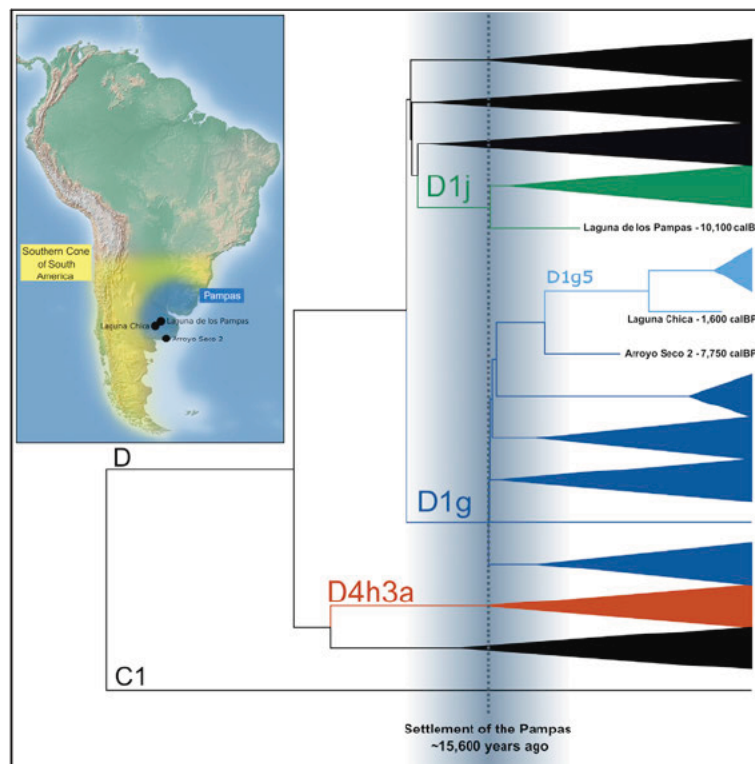
Name of Co-Author	João Teixeira		
Contribution to the Paper	Supervision, helped interpret the results, and provided input on the manuscript.		
Signature		Date	14/10/2022

Name of Co-Author	Bastien Llamas		
Contribution to the Paper	Wet and dry lab genetic analyses, drafted the manuscript together with the main author and other co-authors		
Signature		Date	06/10/2022

Name of Co-Author	Lars Fehren-Schmitz		
Contribution to the Paper	performed the wet lab and dry lab genetic analyses, drafted the manuscript together with the main author and other co-authors		
Signature		Date	05/10/2022

Article

Ancient mitochondrial genomes from the Argentinian Pampas inform the early peopling of the Southern Cone of South America



Xavier Roca Rada,
Gustavo Politis,
Pablo G.
Messineo, ..., João
C. Teixeira,
Bastien Llamas,
Lars Fehren
Schmitz

bastien.llamas@adelaide.edu
au (B.L.)
lfehren@ucsc.edu (L.F.S.)

Highlights

Analysis of 18 ancient human mitochondrial genomes from the Argentinian Pampas.

Genetic makeup of Early Mid Holocene Pampas distinct from later neighboring peoples.

Earliest individuals carrying region specific mitochondrial haplogroups D1j and D1g.

First Pampean settlers were part of a single and rapid dispersal ~15,600 years ago.

Roca Rada et al., *iScience* 24,
102553
June 25, 2021 © 2021 The
Author(s)
[https://doi.org/10.1016/
j.isci.2021.102553](https://doi.org/10.1016/j.isci.2021.102553)



Article

Ancient mitochondrial genomes from the Argentinian Pampas inform the early peopling of the Southern Cone of South America

Xavier Roca Rada,^{1,14,16} Gustavo Politis,^{2,3,14} Pablo G. Messineo,² Nahuel Scheifler,² Clara Scabuzzo,⁴ Mariela González,² Kelly M. Harkins,⁵ David Reich,^{6,7,8,9} Yassine Souilmi,^{1,10,11} João C. Teixeira,^{1,12} Bastien Llamas,^{1,10,11,12,15,*} and Lars Fehren Schmitz^{5,13,15,*}

SUMMARY

The Southern Cone of South America (SCSA) is a key region for investigations about the peopling of the Americas. However, little is known about the eastern sector, the Argentinian Pampas. We analyzed 18 mitochondrial genomes—7 of which are novel—from human skeletal remains from 3 Early to Late Holocene archaeological sites. The Pampas present a distinctive genetic makeup compared to other Middle to Late Holocene pre-Columbian SCSA populations. We also report the earliest individuals carrying SCSA-specific mitochondrial haplogroups D1j and D1g from Early and Middle Holocene, respectively. Using these deep calibration time points in Bayesian phylogenetic reconstructions, we suggest that the first settlers of the Pampas were part of a single and rapid dispersal ~15,600 years ago. Finally, we propose that present-day genetic differences between the Pampas and the rest of the SCSA are due to founder effects, genetic drift, and a partial population replacement ~9,000 years ago.

INTRODUCTION

The Southern Cone of South America (SCSA), formed by Argentina, Chile, Uruguay, and Southern Brazil, is a key study area to build a comprehensive picture of the peopling of the Americas. The region is geographically the most distant from Beringia, the entry point of the first settlers of the Americas, and the archaeological record shows that humans arrived in the SCSA roughly 14.3 thousand years ago (kya) (Dillehay et al., 2008; Politis and Prates, 2018), shortly after the initial colonization of the continent (~16 kya, based on ancient DNA (Llamas et al., 2016)), indicating a rapid spread southwards (Bodner et al., 2012; Prates et al., 2020). A growing body of genetic research addressing the population history of the SCSA has focused mostly on the southernmost archaeological and extant groups from Patagonia and Tierra del Fuego (Lalueza et al., 1997; García-Bour et al., 2004; de Saint Pierre et al., 2012b; de la Fuente et al., 2015, 2018; Crespo et al., 2017, 2018; Nakatsuka et al., 2020b). However, despite its potential importance for the peopling of this region, only a few recent studies included populations from the eastern sector of the SCSA, the Argentinian Pampas (Figure 1) (Perez et al., 2009; Llamas et al., 2016; Posth et al., 2018; Postillone et al., 2020b).

The Argentinian Pampas harbor some of the oldest known archaeological sites in the SCSA, with human presence in the region evident around 14,000 calibrated ¹⁴C years before present (calBP) (Politis et al., 2014; Politis et al., 2016) and a rich record of Late Pleistocene sites (~12.9–11 kya) (Politis, 2008; Mazzanti et al., 2012; Flegenheimer et al., 2013; Martínez et al., 2016). The faunal remains indicate that the subsistence was based on the hunting of camelids, extinct horse, and some extinct megamammals (e.g. giant ground sloths *Megatherium* and giant glyptodonts *Doedicurus*) (Martínez et al., 2016; Miotti et al., 2018). The fishtail projectile point (a typical early lithic tool) appeared both in the Pampas and Patagonia at ~13 kya, a period of time when population size increased in the SCSA (Bodner et al., 2012). However, during the beginning of the Early Holocene (~11–8 kya), there was a reduction of archaeological sites throughout the Pampas and Patagonia and a disappearance of megafaunal remains as well as fishtail projectile points. This change in the archaeological record precedes a proposed population replacement in parts of South America starting around 9 kya (Posth et al., 2018). While a different local retraction process

¹Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia

²INCUAPA CONICET, Facultad de Ciencias Sociales, Universidad Nacional del Centro de la Provincia de Buenos Aires, Olavarría, Buenos Aires, Argentina

³Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, Buenos Aires, Argentina

⁴CICYTTP CONICET, Provincia de Entre Ríos UADER División Arqueología, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata Dr. Materi y España (3105), Diamante, Entre Ríos Argentina

⁵UCSC Paleogenomics Department of Anthropology, University of California, Santa Cruz, Santa Cruz, CA 95064, USA

⁶Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

⁷Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

⁸Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

⁹Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA

¹⁰National Centre for Indigenous Genomics, Australian National University, Canberra, ACT 0200, Australia

¹¹Environment Institute, University of Adelaide, Adelaide, SA 5005, Australia

Continued



iScience 24, 102553, June 25, 2021 © 2021 The Author(s). This is an open access article under the CC BY NC ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1



Figure 1. Map of the Pampas region

The three archaeological sites from which samples for this study derive are highlighted (Laguna de los Pampas, Laguna Chica, and Arroyo Seco 2; black circles), as well as other Early (red circles) and Late Holocene (blue circles) sites mentioned in this study. Los Rieles has some samples that date from the Late Pleistocene. Monte Verde II has signs of human activity dating to the Late Pleistocene. The 19th century territories of the Kawésqar and Yámana populations are shadowed in green and yellow, respectively.

and a population replacement in the Middle Holocene (~8–4 kya), concretely in the latest period, have been proposed based on differences in skull morphology as well as a gap in the radiocarbon database (Barrantes and Perez, 2005), genetic evidence and more dates from archaeological sites indicate population continuity in the region since the Early Holocene (Politis, 2008; Mazzanti et al., 2015; Posth et al., 2018; Donadei, 2019). During the Late Holocene (<4 kya), the complexity of human populations in the Pampas increased due to the introduction of pottery most likely from the subtropical lowlands as well as the development of newly derived technologies from the bow and arrow adoption (Politis, 2008). During this period, macro-regional social link and exchange expanded, and, as a result, an individual with Central-Andes-associated genetic ancestry unexpectedly was found at the archaeological site of Laguna Chica, dating to 1700–1565 calBP (Nakatsuka et al., 2020a).

While phylogeographic studies of mitochondrial genomes (mitogenomes) from the past and present Native Americans have revealed that the major founding lineages (A2, B2, C1b, C1c, C1d, D1, and D4h3a) are widespread across the Americas with temporal and spatial variation (Perego et al., 2009; 2010; Bisso-Machado et al., 2012; Llamas et al., 2016), there are specific clades that show a highly restricted geographic distribution (Bodner et al., 2012; de Saint Pierre et al., 2012b; Gómez-Carballa et al., 2018). Thus, their evolutionary history can be highly informative to infer the population history of a particular region. Four clades are found nearly exclusively in people inhabiting the SCSA: B2i2, C1b13, D1g, and D1j (Bodner et al., 2012; de Saint Pierre et al., 2012b). The Pan-American minor founding lineage D4h3a is found at highest frequencies in Patagonia and Tierra del Fuego (Perego et al., 2009; de Saint Pierre et al., 2012a). A2 and B2, which are frequent all over South America, are virtually absent in populations of the extreme South (Lalueza et al., 1997; Garcia-Bour et al., 2004; Crespo et al., 2018).

In particular, the mitochondrial clades D1g and D1j play a prominent role in the discussions surrounding the peopling of the SCSA. The geographical distribution of both clades, identified by Bodner et al. (2012), differs significantly from one another; D1g is found most frequently in Argentinian and Chilean Patagonia, as well as in the Argentinian Pampas (Bodner et al., 2012; de Saint Pierre et al., 2012b; Crespo et al., 2018), while D1j is most frequent in northern and central Argentina (including the Pampas) and Southern Brazil but virtually absent in Chile (Bodner et al., 2012; Garcia et al., 2012; Crespo et al., 2018; de Saint Pierre, 2017). Based on their calculated divergence ages for both clades (D1g: 18.3 ± 2.4 kya; D1j: 13.9 ± 2.9

¹²Centre of Excellence for Australian Biodiversity and Heritage (CABA), University of Adelaide, Adelaide, SA 5005, Australia

¹³UCSC Genomics Institute, University of California, Santa Cruz, Santa Cruz, CA 95064, USA

¹⁴These authors contributed equally

¹⁵These authors contributed equally

¹⁶Lead contact

*Correspondence
 bastien.llamas@adelaide.edu.au (B.L.),
 lfehrens@ucsc.edu (L.F.S.)
<https://doi.org/10.1016/j.isci.2021.102553>



kya) and the prominence of Monte Verde II in southern Chile as the oldest known site in the SCSA (~14.3 kya) (Dillehay et al., 2008; Politis and Prates, 2018). Bodner et al. (2012) proposed that the most likely peopling scenario was with one initial founding population arriving via the Pacific coast and subsequently crossing the Andes into Argentina, followed by further diversification of both clades. However, other studies note that the practical non-overlapping distribution of both clades contradicts this hypothesis (de Saint Pierre et al., 2012b; Garcia et al., 2012; de Saint Pierre, 2017). Concretely, de Saint Pierre (2017) proposed much older divergence dates of the two clades (D1g: 22 ± 7 kya; D1j: 16.7 ± 9.4 kya), suggesting that the peopling of South America along the Pacific route might have happened much earlier than suggested by studies that include ancient mitogenomes (Llamas et al., 2016) or genomic data (Moreno-Mayar et al., 2018; Posth et al., 2018). Support for this earlier peopling hypothesis relies on the older dates reported for the Monte Verde locality (Dillehay et al., 2015). However, the association of these older dates with human occupation remains controversial (Politis and Prates, 2018; Prates et al., 2020).

We identify three potential shortcomings of studies addressing the peopling of the SCSA using mitogenomes: (1) Inferences are highly dependent on the prominence of one specific archaeological site; (2) Most of the mitochondrial data derive from either relatively recent archaeological samples or modern extant groups who might not have a local origin; (3) Divergence date estimates are highly dependent on the calibration method applied to the molecular clock when converting estimates of relative rates of molecular evolution into calendar years in the phylogenetic tree. Several studies have shown that calibration points close to the age of the events, such as radiocarbon-dated ancient DNA sequences (so-called tip calibrations), provide the most reliable date inferences (Rieux et al., 2014; Llamas et al., 2016, 2017; Posth et al., 2016).

In order to address these potential issues and to better understand the population history of the Argentinian Pampas and its implications in the peopling of the SCSA, we analyzed whole mitogenomes from well-dated human skeletal remains from three Early to Late Holocene archaeological sites in the Argentinian Pampas: Arroyo Seco 2 (AS2) (Politis et al., 2014; Politis et al., 2016), Laguna de los Pampas (LLP) (Messineo et al., 2018; Messineo et al., 2019), and Laguna Chica (LCH) (Scheifler et al., 2017; Messineo et al., 2019) (Figures 1 and S1–S3; supplemental information text). We further employed a tip-calibration approach to determine the divergence ages and to better understand the evolutionary history of the SCSA-specific mitochondrial clades D1g and D1j, as well as the Pan-American minor founding lineage D4h3a that has its highest frequencies in Patagonia and Tierra del Fuego.

RESULTS

Ancient DNA authenticity

We reconstructed 18 mitogenomes at an average coverage ranging from 22–419X (Tables 1 and S1). Twelve of these mitogenomes have been reported previously (Llamas et al., 2016; Posth et al., 2018; Nakatsuka et al., 2020a), whereas an additional seven are reported here for the first time. The DNA damage patterns observed at the terminal ends of the sequencing reads are indicative of ancient degraded DNA (Jónsson et al., 2013) and, combined with the low observed contamination estimates (0.5–3%), support the authenticity of the results (Table S1).

Radiocarbon dating

Radiocarbon dates obtained for some of the successfully sequenced samples show that one newly reported individual from Laguna de los Pampas dates from the initial Early Holocene (LLP.S2.E1; 10,223–9,764 calBP), while four individuals from Arroyo Seco 2 date from the terminal Early Holocene (8,960–8,188 calBP). The majority of samples from Arroyo Seco 2 ($n = 9$; 7,970–6,950 calBP) and Laguna Chica ($n = 3$; 7,724–6,650 calBP) date from the Middle Holocene with the exception of one individual from Laguna Chica that dates from the Late Holocene (1,627–1,565 calBP) (Table 1) (Politis et al., 2014; Messineo et al., 2018; Posth et al., 2018; Nakatsuka et al., 2020a). Importantly, a radiocarbon date for a burial in Laguna Chica (LCH.E1.3) is reported for the first time and corresponds to the early Middle Holocene (7,724–7,589 calBP).

Mitochondrial haplogroup diversity

The mitogenomes of the Early to Late Holocene individuals were assigned to 18 distinct haplotypes within known Native American mitochondrial haplogroups (Table 1). The diversity of haplotypes

Table 1. Calibrated dates and mitochondrial haplogroups of the 18 Early to Late Holocene individuals from the Argentinian Pampas.

Site	Sample	Calibrated date	Period	Mitochondrial haplogroup	GenBank ID	Reference
Laguna de los Pampas	LLP.S2.E1	10,223 9,764	Initial Early Holocene	D1j	MW291678	This study
Arroyo Seco 2	ASO_B27_S36	8,960 8,380	Terminal Early Holocene	C1b	MW291663	(Posth et al., 2018)
	ASO_B24_S31	8,545 8,188		C1b+16,311	MW291669	This study
	ASO_B13_S20	8,545 8,188 ^a		C1c	MW291667	This study
	ASO_S49	8,520 8,200	Middle Holocene	C1b	MW291664	(Posth et al., 2018)
	ASO_B3_S7	7,970 7,673		C1b	MW291671	This study
	ASO_B10_S15	7,920 7,660		D1g	MW291661	(Posth et al., 2018)
	ASO_B10_S17	7,920 7,660 ^b		C1b	MW291673	This study
	ASO_B9_S14a	7,832 7,573		D1	MW291672	(Llomas et al., 2016)
	ASO_B12_S19	7,570 7,300		A2	MW291666	(Llomas et al., 2016)
	ASO_B2_S6	7,570 7,290		A2	MW291670	(Posth et al., 2018)
	ASO_B2_S5	7,570 7,290 ^c		A2	MW291665	This study
	ASO_B1_S3	7,330 6,950		C1c	MW291668	(Posth et al., 2018)
	ASO_B1_S1	7,330 6,950 ^d		C1c	MW291662	This study
Laguna Chica	LCH.E1.3	7,724 7,589	Late Holocene	A2	MW291674	(Posth et al., 2018)
	LCH.E2.I2.1	6,960 6,790		B2b	MW291676	(Posth et al., 2018)
	LCH.E2.I1.2	6,780 6,650		C1b	MW291675	(Posth et al., 2018)
	LCH.E4.4	1,627 1,565		D1g5	MW291677	(Nakatsuka et al., 2020a)

^aIndirect date Burial 13 (skeleton 20; B13_S20) and burial 24 (skeleton 31; B24_S31) were buried at roughly the same depth and share some characteristics (primary burials with calcareous stones), which indicate they might have the same age (Politis et al., 2014)

^bIndirect date Burial 10 was formed by three fully articulated skeletons 1 (B10_S15), 2 (B10_S14), and 3 (B10_S17) buried at the same time. Date comes from B10_S15 and was extrapolated to B10_S17

^cIndirect date Burial 2 was formed by three fully articulated skeletons 4 (B2_S4), 5 (B2_S5), and 6 (B2_S6) buried at the same time. The ¹⁴C date comes from individual B2_S6 and was extrapolated to B2_S5

^dIndirect date Burial 1 was formed by three fully articulated skeletons 1 (B1_S1), 2 (B1_S2), and 3 (B1_S3) buried at the same time. Date comes from B1_S3 and was extrapolated to B1_S1

observed in the Early to Middle Holocene Pampas (excluding the Late Holocene individual from Laguna Chica) is characterized by the prevalence of C1 haplotypes (C1b: 35.3% and C1c: 17.6%), a higher frequency of A2 haplotypes (23.5%) compared to D1 haplotypes (17.6%), the detection of a B2b haplotype (6%), and the absence of D4h3a. Both B2b and C1c are absent in other pre-Columbian populations of the Eastern SCSA (de la Fuente et al., 2015; Crespo et al., 2017; Postillone et al., 2020b) but found in modern-day groups of central Argentina (including the Pampas) and northern Patagonia (Bailliet et al., 1994; Merriwether et al., 1995; Bobillo et al., 2010; Perego et al., 2010; Gómez-Carballa et al., 2016; Motti et al., 2020).

The oldest mitogenome obtained in this study from an individual found at the site Laguna de los Pampas (LLP.S2.E1; 10,223–9,764 calBP) exhibits a haplotype basal to D1j, according to the nomenclature suggested by [phylotree.org](https://www.phylotree.org) (mtDNA tree Build 17 [18 Feb 2016]) (van Oven, 2015). The haplotype shows the characteristic C16242T and T16311C control region mutation pattern but lacks the T152C substitution (Bodner et al., 2012).

Two D1 mitogenomes, one from the early Middle Holocene Arroyo Seco 2 (ASO_B10_S15; 7,920–7,660 calBP) and the other from the Late Holocene Laguna Chica (LCH.E4.4; 1,627–1,565 calBP), exhibit the mutational pattern A8116G-C16187T characteristic of haplogroup D1g (van Oven and Kayser, 2009; Bodner et al., 2012). The haplotype of the younger LCH.E4.4 shows an additional four substitutions (T55C-A56G-T10595C-T16209C) that are considered part of the mutational motif characterizing the D1g5 subclade, despite lacking three substitutions at nucleotide positions G499A-A3505G-A14693G (Bodner et al., 2012; García et al., 2012). The ASO_B10_S15 D1g mitogenome shares the characteristic substitution T55C, suggesting that the haplotype might be basal to the subclade D1g5.

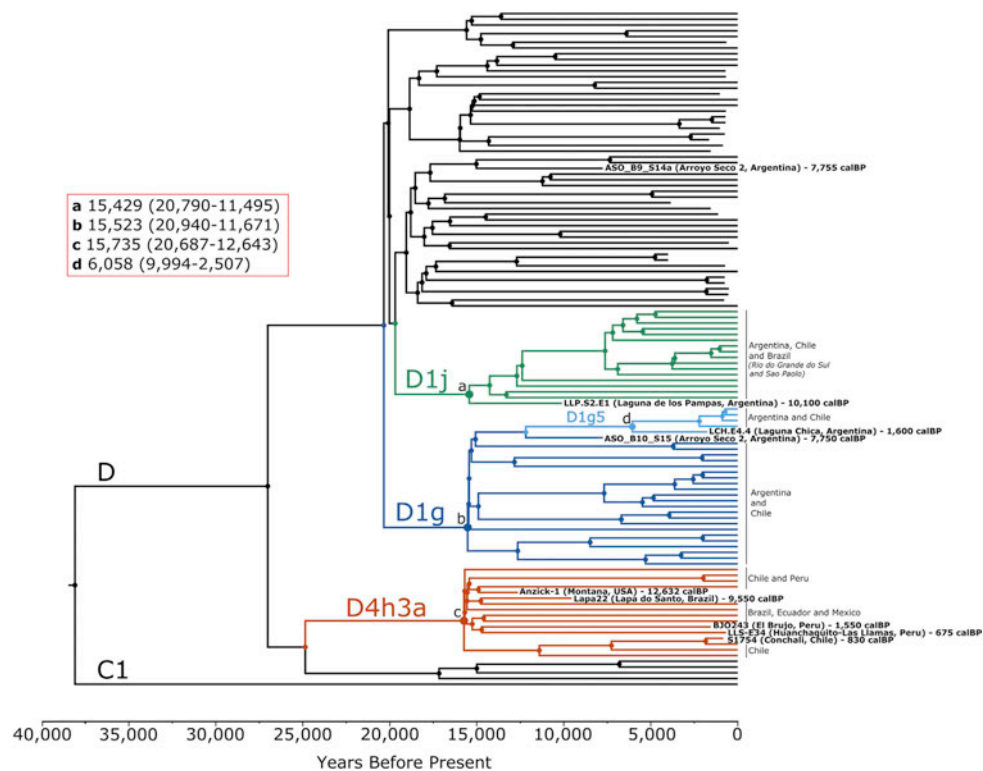


Figure 2. Maximum clade credibility tree for haplogroups D1 and D4h3a

The estimated mean coalescent times are shown for clades D1j (a), D1g (b), D4h3a (c), and subhaplogroup D1g5 (d) in years before present (credible intervals in parentheses)

Time-measured phylogenetic analyses

We further investigated the evolutionary relationship between the newly reported Early to Late Holocene D1j and D1g mitogenomes from the Argentinian Pampas and other mitogenomes from the SCSA and the Americas. To achieve this, we constructed a maximum clade credibility tree employing Bayesian phylogenetic inference using BEAST 1.8.0 (Drummond and Rambaut, 2007) (see STAR Methods). A total of 117 pre-Columbian and modern-day haplotypes belonging to haplogroups D1 and D4h3a (Bodner et al., 2012; de Saint Pierre et al., 2012b; Rasmussen et al., 2014; van Oven, 2015; Llamas et al., 2016; Posth et al., 2018) were included to estimate the coalescent times of these haplogroups. The resulting phylogenetic tree (Figures 2 and S4) confirms that the D1j mitogenome from Laguna de los Pampas (LLP.S2.E1) is basal to the entire D1j clade. The Middle Holocene D1g lineage from Arroyo Seco 2 (ASO_B10_S15) is positioned basal to a clade containing the Late Holocene Laguna Chica D1g lineage (LCH.E4.4), which again is ancestral to a number of mitogenomes falling into the subclade D1g5 containing contemporary Pampeans and Patagonians (Figure 2).

The radiocarbon dates of the ancient sequences from this and previously published studies were used as tip calibrations in the phylogenetic analysis (Table 1; Figure 2). We observe highly synchronous estimates of the times to the most recent common ancestor (TMRCA) for all three haplogroups D1g, D1j, and D4h3a (Figure S5), with an average divergence time of ~15.6 kya (Figure 2). The TMRCA estimates for D1g



(95% highest posterior density interval: 20.9–11.7 kya) and D1j (20.8–11.5 kya) fall well within the divergence age estimated for the founding lineage D1 in our study of 28.6–14.3 kya and previously published estimates of 23.7–12.5 kya (Llamas et al., 2016). Similarly, for D4h3a (21.6–12.6 kya), we were able to reproduce the previous TMRCA estimates of 20.9–12.6 kya (Llamas et al., 2016).

DISCUSSION

Few studies so far have addressed the past genetic diversity of Early to Late Holocene hunter-gatherer populations in the Argentinian Pampas and their genetic relationship with populations of the SCSA. Positioned at the Eastern sector edge of the SCSA and harboring some of the earliest known archaeological sites, the region might have played a crucial role during the peopling of the region.

Our analysis revealed a higher diversity of mitochondrial haplotypes and a different distribution of haplogroup frequencies in the Early to Middle Holocene Pampas compared to other published Middle to Late Holocene pre-Columbian Eastern Patagonian and SCSA populations (Cabana et al., 2006; Marrero et al., 2007; García and Demarchi, 2009; Bobillo et al., 2010; Nores and Demarchi, 2011; de Saint Pierre et al., 2012b; Motti, 2012; Cardoso et al., 2013; de la Fuente et al., 2015; Crespo et al., 2017; Arencibia et al., 2019; Motti et al., 2020; Postillone et al., 2020a, 2020b) (Table S2). These include Late Holocene groups from the geographically near site of Paso Alsina 1, at the eastern Pampa-Patagonia transition (Postillone et al., 2020b), and other sites along the Atlantic coast of Argentinian Patagonia (Motti et al., 2015; Crespo et al., 2017). The difference in haplogroup frequencies between our data set and those previously published is driven by the predominance of haplogroup C1 and the presence of B2b in the Early to Middle Holocene Pampas, while haplogroup D1 is predominant and B2 is mostly absent in the Middle to Late Holocene pre-Columbian Eastern Patagonian and SCSA populations. Even though haplogroup C1 was prevalent in a late 19th and early 20th century indigenous cemetery in Tierra del Fuego (La Candelaria), haplogroup D1 had a high frequency and B2 was not found (Motti et al., 2020). Interestingly, the Late Holocene study in the Córdoba Province (Argentina) presents similar frequencies to our study, as haplogroup C is also predominant and it is followed by B and A, with haplogroup D being the least frequent (Nores and Demarchi, 2011). Moreover, haplogroup B has also been reported at high frequencies in individuals from the Early Holocene site Baño Nuevo (Manríquez et al., 2011) and Late Holocene site Lago Salitroso (Arencibia et al., 2019), both in south-western Patagonia. Since these last studies are based on general identification of mitochondrial haplogroup variants or short, low-resolution hypervariable region 1 sequences, as opposed to complete mitogenomes, it is not possible to determine if those individuals carried the same haplotypes as observed in this study.

Lineages falling into the B2b clade are found throughout South America, most frequently in the Andes, and the clade most likely evolved in the North of Mesoamerica shortly after the initial entry into the Americas (Brandini et al., 2018). Notwithstanding, the most common B2b lineages in the Pampas and central Argentina today are the derived lineages B2b3 or B2b14 (Bobillo et al., 2010; Brandini et al., 2018), while the Middle Holocene Laguna Chica individual (LCH.E2-I2.1) exhibits a less derived B2b haplotype. Gómez-Carballa et al. (2018) performed a phylogeographic analysis of present-day B2 mitogenomes. Their study uses a bifurcation model to explain the presence of the two subclades in the Pampas and central Argentina, proposing that the less frequent B2b14 arrived from the Andes, and B2b3 arrived via an Amazonian/Atlantic route. The presence of the basal B2b haplotype in the Middle Holocene Pampas suggests that individuals carrying an ancestral B2b haplotype spread in the northern SCSA earlier than the B2b3-B2b14 divergence, with the derived clades either arriving in subsequent waves or differentiating later via random genetic drift. These results reinforce the need to include ancient DNA analysis to disentangle past demographic events. Like the B2b haplotype, most A2, C1b, and C1c haplotypes found in the Early and Middle Holocene individuals from the Pampas are relatively basal to the Native American founding haplotypes (Perego et al., 2010). This observation is in agreement with the proposed rapid dispersal of the initial settlers throughout the American continent (Bodner et al., 2012; Llamas et al., 2016; Prates et al., 2020). However, due to the lack of available information for these haplogroups in the SCSA, changes in haplotype frequency through the Holocene period cannot be inferred.

D1g and D1j are the two main mitochondrial haplogroups thought to have a major role in the peopling of the SCSA (Bodner et al., 2012; de Saint Pierre et al., 2012b; García et al., 2012; Crespo et al., 2017; Crespo et al., 2018; de Saint Pierre, 2017). The finding of a D1g haplotype in Arroyo Seco 2 (ASO_B10_S15) supports the idea that D1g was already in the Pampas during the early Middle Holocene. It has been previously

suggested that D1g arrived in the SCSA via the Pacific coast with trans-Andean migrations to the East based on its geographical distribution and frequency patterns in modern-day populations in southern Chile, Argentinian Patagonia, and the Pampas (Bodner et al., 2012; de Saint Pierre et al., 2012b; Crespo et al., 2017; de Saint Pierre, 2017). This early Middle Holocene D1g lineage from Arroyo Seco 2 is basal to a subclade containing the Late Holocene D1g5 mitogenome from Laguna Chica (LCH.E4.4), which is itself basal to the modern-day D1g5 lineages. The derived clade D1g5 has been proposed to have differentiated during the early peopling of the SCSA based on its geographically structured internal clades (Motti et al., 2019; Prieto et al., 2020). However, we find no evidence to corroborate this hypothesis as we estimated a divergence time of only ~6 kya for D1g5 (Figure 2), much later than the peopling of the SCSA, and indicative that D1g5 was present in the Pampas before migrating southwards.

While Bodner et al. (2012) proposed that both D1g and D1j arrived in the SCSA via the Pacific coast, García et al. (2012) and Postillone et al. (2020b) have proposed that D1j probably migrated from central Argentina. This hypothesis was based on the D1j geographic distribution, which extended along the Patagonian Atlantic coast in the Late Holocene groups (Motti et al., 2015; Crespo et al., 2017; Postillone et al., 2020b), and is presently more frequent in the Eastern Pampas, while absent in western Patagonia and the extreme South (Bodner et al., 2012; García et al., 2012). The Early Holocene D1j mitogenome in Laguna de los Pampas is basal in the D1j phylogeny and supports the hypothesis that D1j spread from the Pampas. Interestingly, a cranial morphometric study showed some affinities between the Early Holocene sites of Laguna de los Pampas and Lagoa Santa (Brazil) (Menéndez et al., 2015). These observations, among others, have been used to suggest a possible Atlantic migration route (Miotti, 2006). However, genetic evidence supporting the so-called Paleoamerican morphology has recently been disproven (Raghavan et al., 2015; Moreno-Mayar et al., 2018). Furthermore, the D1j haplotype from Laguna de los Pampas lacks the T152C substitution but has the characteristic C16242T and T16311C substitutions. García et al. (2012) argue that the mutations at T16311C and T152C co-occur in both D1j and other D1 haplotypes found in central Argentina and propose that the substitution at T152C preceded the one at C16242T. Again, the ancient mitogenome from Laguna de los Pampas does not support this hypothesis as our observations indicate that the substitution C16242T preceded T152C. Nevertheless, it should be noted that 152 and 16,311 are mutational hotspots as described by Soares et al. (2009), increasing the odds of a recurrent mutation event.

The divergence ages for D1g and D1j have been estimated several times in previous studies using a range of methods and calibrations of the mitochondrial mutation rate. Bodner et al. (2012) used the HKY85 model and the human-chimpanzee split time (Goodman et al., 1998)—as shown in Mishmar et al. (2003)—and reported 18.3 ± 2.4 kya for D1g and 13.9 ± 2.9 kya for D1j. They also used the corrected rho-based molecular clock proposed by Soares et al. (2009) (also based on the human-chimpanzee split time) and obtained similar time estimates (D1g: 19.7 ± 3 kya; D1j: 14.9 ± 4.7 kya). De Saint Pierre, Bravi, et al. (2012) employed a Bayesian phylogenetic method and multiple mutation rates to propose a divergence date of ~15 kya for D1g. In that study, they argued that the rate based on internal calibration points in the human mitochondrial tree (Endicott and Ho, 2008) was more concordant with the evolutionary process than when using any rate based on the deep human-chimpanzee split. However, de Saint Pierre (2017) recalculated both dates using a corrected rho-based molecular clock based on the human-chimpanzee split (Soares et al., 2009) and pushed the D1g and D1j divergence dates back to 22 ± 7 kya and 16.7 ± 9.4 kya, respectively. This result led to the proposal that D1g and D1j arrived in the SCSA via two temporally distinct migration pulses (de Saint Pierre, 2017). Here, we adopted a Bayesian phylogenetic method with tip dates calibration (Rieux and Balloux, 2016) and our coalescence time estimates for D1j, D1g, and even D4h3a were synchronous at ~15.6 kya. A main strength of the present study is the ability to directly calibrate the molecular clock for all 3 clades (D1g, D1j, and D4h3a) via radiocarbon-dated ancient mitochondrial sequences. While using deep nodes as opposed to tips to calibrate the molecular clock does not affect the tree topology in phylogenetic reconstructions, it does impact significantly the evolutionary substitution rate and ultimately the dating of demographic events (Rieux et al., 2014). In fact, tip calibration is a much more reliable method when time-stamped sequences are available (Rieux and Balloux, 2016), such as is the case for recent human mitogenome studies based on ancient DNA (Brotherton et al., 2013; Fu et al., 2013; Rieux et al., 2014; Llamas et al., 2016; Posth et al., 2016; Nieves-Colón et al., 2020).

Beyond mitochondrial DNA results, nuclear genome data from Posth et al. (2018) demonstrated that the ancient individuals from Arroyo Seco 2 have the highest affinity to present-day populations of the SCSA



and that Arroyo Seco 2 was an admixed population containing two main genetic ancestries, including the Early Holocene Lapa do Santo (Brazil)-associated ancestry and the Middle-Late Holocene ancestry predominant in present-day South Americans. Furthermore, they observed that ancient populations from central Chile (Los Rieles, 10,900 calBP) do not have a significant Lapa do Santo-associated ancestry, even though the most recent samples (Los Rieles, 5,100 calBP) also have the genetic ancestry related to all modern-day South Americans. These observations led [Posth et al. \(2018\)](#) to postulate a population replacement in South America ~ 9 kya. However, the Laguna de los Pampas individual analyzed in this study and carrying the D1j haplogroup predates such putative replacement event, which indicates that D1j was present before the replacement. Considering that Arroyo Seco 2 harbors both genetic ancestries, the proposed replacement event could not have been complete as some mitochondrial lineages persist since the Late Pleistocene until today. A partial population replacement might reasonably explain the high diversity of haplotypes and distinct distribution of haplogroup frequencies observed in the Early and Middle Holocene Pampas compared to the Middle and Late Holocene pre-Columbian SCSA populations.

Hence, the synchronous estimated divergence dates for D1g, D1j, and D4h3a reported in this study (~ 15.6 kya) suggest that the three clades might have emerged during the initial peopling of the Americas (~ 16 kya) ([Llamas et al., 2016](#)) and diversified en route to South America. At ~ 9 kya, a partial population replacement ([Posth et al., 2018](#)) as well as founder effects and genetic drift might have erased D1g and D1j outside the SCSA but not D4h3a. Alternatively, [Prates et al. \(2020\)](#) suggested that the earliest chronological threshold for the peopling of South America was ~ 15.5 kya (16.6–15 kya) using a quantitative analysis of screened radiocarbon databases. Therefore, it could also be possible that the settling of South America led to diversification and D1g and D1j arose in South America. However, this scenario would imply that although D4h3a emerged at the same time, it might have had a different genetic history given its distribution across the entire American Pacific coast both in pre-Columbian and contemporary populations and its high frequency in Patagonia and Tierra del Fuego ([Perego et al., 2009](#); [Moraga et al., 2010](#); [de Saint Pierre et al., 2012a](#)).

In conclusion, the most parsimonious explanation for our results is that at least D1g, D1j, and likely B2b arrived in the Pampas during the initial settlement of the region as inferred from archaeological evidence ($\sim 14,000$ calBP) ([Politis et al., 2014](#); [Politis et al., 2016](#)), molecular dating (~ 20 – 11 kya), and quantitative analysis of radiocarbon dates (16.6–15 kya) ([Prates et al., 2020](#)). Despite the caveat that mitochondrial data from the present study do not allow us to infer specific migration routes, it is unlikely that the peopling of the Pampas was through the Pacific coast followed by the crossing of the Southern Andes. The Patagonian Ice Sheet reached the 35th parallel south during the Late Glacial Maximum and it was still in the Le Glacial phase ~ 16 – 14 kya ([Dickinson, 2011](#); [Rabassa et al., 2011](#)). Even if trans-cordilleran passes were open during that time, it is unlikely that they were suitable for human transit. Thus, an Atlantic or inland route is more likely for the peopling of the Pampas. Additionally, a partial population replacement took place ~ 9 kya in South America and reached both the Pampas and the rest of the SCSA ([Posth et al., 2018](#)). Hence, some derived lineages that are currently specific to the Southern Cone could have been brought in by the second wave. For example, the basal B2b haplogroup found in Laguna Chica could have been present during the initial settlement of the Pampas, while present-day derived B2b3 and B2b14 could have come from the Andes and the Atlantic Route, respectively ([Gómez-Carballea et al., 2018](#)). Finally, Arroyo Seco 2 has the highest affinity to present-day SCSA indigenous populations ([Posth et al., 2018](#)), which suggests that the Pampas and the rest of the SCSA share some of their demographic history. Our results are consistent with the Pampas being one of the sources of the ancient Patagonian gene pool, and contemporary genetic differences observed between both regions could stem from founder effects, genetic drift, and the partial population replacement ~ 9 kya.

Limitations of the study

There are limitations associated with the study of mtDNA, as this genetic marker represents the evolutionary history of the female population at a single locus and sample sizes are often small. Moreover, mitochondrial data from the present study do not allow us to test specific migration routes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:



- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Archaeological context
 - Brief description of archaeological sites
- METHODS DETAILS
 - Radiocarbon dating and calibration
 - Ancient DNA laboratory work
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Contamination estimation
 - Mitochondrial DNA analyses
 - Phylogenetic analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.102553>.

ACKNOWLEDGMENTS

The reported investigation of the individuals from Arroyo Seco 2, Laguna Chica, and Laguna de los Pampas has been permitted by the government of the Province of Buenos Aires (No. de Registro 2014-3-A-125-1 and 2017-3-A-172-1). No indigenous communities in the counties have reclaimed Arroyo Seco 2 and Laguna de los Pampas remains. For Laguna Chica, in addition to obtaining permits from the provincial heritage institutions, the indigenous community living near the site (Comunidad Indígena Mapuche-Tehuelche Cacique Pincen) approved the study after consultation and participation in the rescue excavation (the skeletal remains will be re-buried). The authors want to acknowledge Beatriz Araujo Pincen, one of the leaders of the community as well as the local collaborators Ramon Coria, Juanjo Estevez, Ariel Grub, Marisa Martin, Sonia Finoccio, and Vicente Benito. The Municipality of Tres Arroyos was always very helpful during fieldwork. The authors want to also acknowledge Brendan Culleton for his assistance in the radiocarbon dating and for the summary of the methodology. X.R.-R. is supported by a Molecular Anthropology PhD scholarship from the University of Adelaide. L.F.-S. was supported by a U.S. National Science Foundation (NSF) grant (1515138) and the Wenner-Gren Foundation (SC-14-62). G.P. received support by PUE-CONICET 2015–2021 granted to INCUAPA and PICT 2015-2777 from ANPCYT, Argentina. P.G.M. received additional support by the National Geographic Society (#NGS-50543R-18) and CONICET (PIP No. 0414). Y.S. is supported by an Australian Research Council Discovery Project (ARC DP190103705). J.C.T. is supported by an ARC Indigenous Discovery Grant (IN180100017). B.L. is supported by an ARC Future Fellowship (FT170100448). K.M.H. was supported from 2015 to 2017 by the National Science Foundation (NSF) SBE postdoctoral fellowship (1513501).

AUTHOR CONTRIBUTIONS

X.R.-R., K.M.H., B.L., and L.F.-S. performed the wet lab and dry lab genetic analyses. G.P., P.G.M., N.S., C.S., and M.G. provided archaeological samples, radiocarbon dates, and contextual information. D.R. provided genetic data. All authors interpreted the genetic results. X.R.-R., G.P., B.L., and L.F.-S. wrote the manuscript with input from all co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 1, 2021

Revised: April 26, 2021

Accepted: May 14, 2021

Published: June 25, 2021



REFERENCES

- Arencibia, V., Crespo, C., García Guraieb, S., Russo, M G., Dejean, C B., and Gohli, R (2019) Análisis genético poblacional de grupos cazadores recolectores del Holoceno tardío del Lago Salitroso (Santa Cruz, Argentina) *RevArgAntropBiol* 21, 004
- Bailliet, G., Rothhammer, F., Carnese, F R., Bravi, C M., and Bianchi, N O (1994) Founder mitochondrial haplotypes in Amerindian populations *Am J Hum Genet* 55, 27–33
- Barrientos, G., and Perez, S I (2005) Was there a population replacement during the Late mid Holocene in the southeastern Pampas of Argentina? Archaeological evidence and Paleoeological basis *Quat Int* 132, 95–105
- Beverly, R K., Beaumont, W., Taz, D., Ormsby, K M., von Reden, K F., Santos, G M., and Southon, J R (2010) The keck carbon cycle ams laboratory, university of California, irvine status report *Radiocarbon* 52, 301–309
- Bisso Machado, R., Bortolini, M C., and Salzano, F M (2012) Uniparental genetic markers in South amerindians *Genet Mol Biol* 35, 365–387
- Bobillo, M C., Zimmermann, B., Sala, A., Huber, G., Röck, A., Bandelt, H J., Corach, D., and Parson, W (2010) Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina towards a first nationwide forensic mitochondrial DNA sequence database *Int J Leg Med* 124, 263–268
- Bodner, M., Perego, U A., Huber, G., Fendt, L., Röck, A W., Zimmermann, B., Olivieri, A., Gómez Carballa, A., Lancioni, H., Angerhofer, N., et al (2012) Rapid coastal spread of First Americans: novel insights from South America's Southern Cone mitochondrial genomes *Genome Res* 22, 811–820
- Brandini, S., Bergamaschi, P., Cerna, M F., Gandini, F., Bastaroli, F., Bertolini, E., Cereda, C., Ferretti, L., Gómez Carballa, A., Battaglia, V., et al (2018) The Paleo Indian entry into South America according to mitogenomes *Mol Biol Evol* 35, 299–311
- Bronk Ramsey, C. (2013) Recent and planned developments of the program OxCal *Radiocarbon* 55, 720–730
- Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Jane Adler, C., Richards, S M., Der Sarkissian, C., Ganslmeier, R., Friederich, S., et al (2013) Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans *Nat Commun* 4, 1764
- Cabana, G S., Merriwether, D A., Hunley, K., and Demarchi, D A (2006) Is the genetic structure of Gran Chaco populations unique? Interregional perspectives on native South American mitochondrial DNA variation *Am J Phys Anthropol* 131, 108–119
- Cardoso, S., Palencia Madrid, L., Valverde, L., Alfonso Sánchez, M A., Gómez Pérez, L., Alfaro, E., Bravi, C M., Dipiéri, J E., Peña, J A., and de Pancorbo, M M (2013) Mitochondrial DNA control region data reveal high prevalence of Native American lineages in Jujuy province, NW Argentina *Forensic Sci Int Genet* 7, e52–5
- Crespo, C., Favier Dubois, C., Russo, M., Lanata, J., and Dejean, C B. (2017) First analysis of ancient mtDNA genetic diversity in Northern coast of Argentinean Patagonia *J Archaeol Sci Rep* 12, 91–98
- Crespo, C M., Lanata, J L., Cardozo, D G., Avena, S A., and Dejean, C B. (2018) Ancient maternal lineages in hunter gatherer groups of Argentinean Patagonia: Settlement, population continuity and divergence *J Archaeol Sci Rep* 18, 689–695
- Dabney, J., Knapp, M., Glocke, I., Gansauge, M T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J L., and Meyer, M (2013) Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments *Proc Natl Acad Sci U S A* 110, 15758–15763
- Dickinson, W R (2011) Geological perspectives on the Monte Verde archeological site in Chile and pre Clovis coastal migration in the Americas *Quat Res* 76, 201–210
- Dillehay, T D., Ramirez, C., Pino, M., Collins, M B., Rossen, J., and Pino Navarro, J D (2008) Monte Verde seaweed, food, medicine, and the peopling of South America *Science* 320, 784–786
- Dillehay, T D., Ocampo, C., Saavedra, J., Sawakuchi, A O., Vega, R M., Pino, M., Collins, M B., Scott Cummings, L., Arregui, I., Villagran, X S., et al (2015) New archaeological evidence for an early human presence at Monte Verde, Chile *PLoS One* 10, e0141923–27
- Donadei, J P (2019) Local and nonlocal rocks technological strategies and raw material management: Hunter gatherer mobility for mid Holocene groups of eastern Tandilia range (Argentina) *J Archaeol Sci Rep* 24, 264–275
- Drummond, A J., and Rambaut, A (2007) BEAST Bayesian evolutionary analysis by sampling trees *BMC Evol Biol* 7, 214–218
- Edgar, R C (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput *Nucleic Acids Res* 32, 1792–1797
- Endicott, P., and Ho, S Y (2008) A bayesian evaluation of human mitochondrial substitution rates *Am J Hum Genet* 82, 895–902
- Flegenheimer, N., Miotti, L., and Mazza, N (2013) Rethinking early objects and landscape in the Southern Cone: fish tail point concentrations in the Pampas and Northern Patagonia. In *Paleoamerican Odyssey*, K Graf, C Ketron, and M Waters, eds (Center for the Study of First Americans), pp 359–376
- Fu, Q., Mitnik, A., Johnson, P L F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al (2013) A revised timescale for human evolution based on ancient mitochondrial genomes *Curr Biol* 23, 553–559
- de la Fuente, C., Galimany, J., Kemp, B M., Judd, K., Reyes, O., and Moraga, M (2015) Ancient marine hunter gatherers from Patagonia and Tierra Del Fuego: diversity and differentiation using uniparentally inherited genetic markers *Am J Phys Anthropol* 158, 719–729
- de la Fuente, C., Ávila Arcos, M C., Galimany, J., Carpenter, M L., Homburger, J R., Blanco, A., Contreras, P., Cruz Dávalos, D., Reyes, O., San Roman, M., et al (2018) Genomic insights into the origin and diversification of late maritime hunter gatherers from the Chilean Patagonia *Proc Natl Acad Sci U S A* 115, E4006–E4012
- García, A., and Demarchi, D A (2009) Incidence and distribution of native American mtDNA haplogroups in central Argentina *Hum Biol* 81, 59–69
- García, A., Pauro, M., Nores, R., Bravi, C M., and Demarchi, D A (2012) Phylogeography of mitochondrial haplogroup D1: an early spread of subhaplogroup D1j from Central Argentina *Am J Phys Anthropol* 149, 583–590
- García Bour, J., Pérez Pérez, A., Álvarez, S., Fernández, E., López Parra, A M., Arroyo Pardo, E., and Turbón, D (2004) Early Population Differentiation in Extinct Aborigines from Tierra del Fuego Patagonia: ancient mtDNA Sequences and Y Chromosome STR Characterization *Am J Phys Anthropol* 123, 361–370
- Gómez Carballa, A., Moreno, F., Álvarez Iglesias, V., Martín Torres, F., García Magariños, M., Pantoja Astudillo, J A., Aguirre Morales, E., Bustos, P., and Salas, A. (2016) Revealing latitudinal patterns of mitochondrial DNA diversity in Chileans *Forensic Sci Int Genet* 20, 81–88
- Gómez Carballa, A., Pardo Seco, J., Brandini, S., Achilli, A., Perego, U A., Coble, M D., Diegoli, T M., Álvarez Iglesias, V., Martín Torres, F., Olivieri, A., et al (2018) The peopling of South America and the trans Andean gene flow of the first settlers *Genome Res* 28, 767–779
- Goodman, M., Porter, C A., Czelusniak, J., Page, S L., Schneider, H., Shoshani, J., Gunnell, G., and Groves, C P (1998) Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence *Mol Phylogenet Evol* 9, 585–598
- Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P L F., and Orlando, L (2013) MapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters *Bioinformatics* 29, 1682–1684
- Lalueza, C., Pérez Pérez, A., Prats, E., Cornudella, L., and Turbón, D (1997) 'Lack of founding Amerindian mitochondrial DNA lineages in extinct Aborigines from Tierra del Fuego Patagonia' *Hum Mol Genet* 6, 41–46
- Lanfear, R., Calcott, B., Ho, S Y., and Guindon, S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses *Mol Biol Evol* 29, 1695–1701
- Li, H., and Durbin, R (2009) Fast and accurate short read alignment with Burrows Wheeler transform *Bioinformatics* 25, 1754–1760
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., and Abecasis, G R (2009) Durbin: 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools *Bioinformatics* 25, 2078–2079

- Llamas, B., Fehren Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., Nordenfelt, S., Valdiosera, C., Richards, S.M., Rohlfach, A., et al (2016) Ancient mitochondrial DNA provides high resolution time scale of the peopling of the Americas. *Sci Adv* 2, e1501385
- Llamas, B., Willerslev, E., and Orlando, L. (2017) Human evolution: a tale from ancient genomes. *Philos Trans R Soc Lond B Biol Sci* 372, 20150484
- Manríquez, G., Moraga, M., Santoro, C., Aspilla, E., Arriaza, B.T., and Rothhammer, F. (2011) Morphometric and mtDNA analyses of archaic skeletal remains from Southwestern South America. *Chungará (Arica)* 43, 283–292
- Marrero, A.R., Bravi, C., Stuart, S., Long, J.C., Pereira das Neves Leite, F., Komers, T., Carvalho, C.M., Pena, S.D., Ruiz Linares, A., Salzano, F.M., and Cátira Bortolini, M. (2007) Pre and post Columbian gene and cultural continuity: the case of the Gaucho from southern Brazil. *Hum Hered* 64, 160–171
- Martínez, G., Gutiérrez, M.A., Messineo, P.G., Kaufmann, C.A., and Rafuse, D.J. (2016) Subsistence strategies in Argentina during the late Pleistocene and early Holocene. *Quat Sci Rev* 144, 51–65
- Mazzanti, D.L., Martínez, G.A., and Quintana, C. (2012) Early settlements in eastern Tandilia, Buenos Aires Province, Argentina: archaeological contexts and site formation processes. In *Special Edition: Southbound Late Pleistocene Peopling of Latin America*, L. Miotti, M. Salemme, N. Flegenheimer, and T. Goebel, eds (Texas University), pp 115–119
- Mazzanti, D.L., Martínez, G.A., and Quintana, C. (2015) Asentamientos del Holoceno medio en Tandilia oriental: Aportes para el conocimiento de la dinámica poblacional de la región pampeana, Argentina. *Relaciones de la Sociedad Argentina de Antropología XL*, 209–231
- Menéndez, L.P., Perez, S.I., Pucciarelli, H.M., Bonomo, M., Messineo, P.G., Gonzalez, M.E., and Politis, G.G. (2015) Early holocene human remains from the Argentinean pampas: cranial variation in South America and the American peopling. *PaleoAmerica* 1, 251–265
- Merriwether, D.A., Rothhammer, F., and Ferrell, R.E. (1995) Distribution of the four founding lineage haplotypes in native Americans suggests a single wave of migration for the New World. *Am J Phys Anthropol* 98, 411–430
- Messineo, P.G., González, M.E., Álvarez, M.C., and Pal, N. (2018) Las ocupaciones humanas en la localidad Arqueológica Laguna de Los Pampas (Campo De Dunas Del Centro Pampeano, Argentina) Durante El Holoceno. *Latin Am Antiq* 29, 736–753
- Messineo, P.G., Barros, M.P., Pal, N., and Scheifler, N.A. (2019) Transporting rocks to an empty environment of lithic raw materials: The case of the Central Pampas Dunefields (Argentina). *J Archaeol Sci Rep* 25, 433–446
- Messineo, P.G., Scheifler, N.A., Álvarez, M.C., González, M., Pal, N., Barros, P., and Politis, G. (2019) A model of human occupation in the Central Pampas Dunefields of Argentina. *PaleoAmerica* A.J. Early Hum Migration Dispersal 5, 378–391
- Miotti, L.L. (2006) La fachada atlántica, como puerta de ingreso alternativa de la colonización humana de América del Sur durante la transición Pleistoceno/Holoceno. In 2° Simposio Internacional el Hombre Temprano en América, pp 155–188
- Miotti, L., Tonni, E., and Marchionni, L. (2018) What happened when the Pleistocene megafauna became extinct? *Quat Int* 473, 173–189
- Mishmar, D., Ruiz Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., et al. (2003) Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 100, 171–176
- Moraga, M., de Sanit Pierre, M., de Saint Torres, F., and Ríos, J. (2010) Kinship by maternal via between the last descendants of kawésqar ethnicity and burials in the patagonian channels: evidence from the study of mitochondrial lineages. *Magallania* 38, 103–114
- Moreno Mayar, J.V., Vinner, L., de Barros Damgaard, P., de la Fuente, C., Chan, J., Spence, J.P., Allentoft, M.E., Vimala, T., Racimo, F., Pinotti, T., et al. (2018) Early human dispersals within the Americas. *Science* 362, eaav2621
- Motti, J.M.B. (2012) Caracterización de linajes maternos en la población actual del Noroeste y Centro oeste argentinos (Universidad Nacional de La Plata)
- Motti, J.M.B., Hagelberg, E., Lindo, J., Malhi, R.S., Bravi, C.M., and Guichón, R.A. (2015) Primer genoma mitocondrial en restos humanos de la costa de Santa Cruz, Argentina. *Magallania* 43, 119–131
- Motti, J.M.B., Muñoz, A.S., Cruz, I., D'Angelo del Campo, M.D., Borrero, L.A., Bravi, C.M., and Guichón, R.A. (2019) Análisis de ADN mitocondrial en restos humanos del Holoceno tardío del sur de Santa Cruz (Arqueología de la Patagonia). *El pasado en las arenas*, pp 493–503
- Motti, J.M.B., Winingear, S., Valenzuela, L.O., Nieves Colón, M.A., Harkins, K.M., García Laborde, P., Bravi, C.M., Guichón, R.A., and Stone, A.C. (2020) Identification of the geographic origins of people buried in the cemetery of the Salesian Mission of Tierra del Fuego through the analyses of mtDNA and stable isotopes. *J Archaeol Sci Rep* 33, 102559
- Nakatsuka, N., Lazaridis, I., Barbieri, C., Skoglund, P., Rohland, N., Mallick, S., Posth, C., Harkins, Kinkaid, K., Ferry, M., Harney, E., et al. (2020a) A paleogenomic reconstruction of the deep population history of the Andes. *Cell* 181, 1131–e21
- Nakatsuka, N., Luisi, P., Motti, J.M.B., Salemme, M., Santiago, F., D'Angelo Del Campo, M.D., Vecchi, R.J., Espinosa Parrilla, Y., Prieto, A., Adamski, N., et al. (2020b) Ancient genomes in South Patagonia reveal population movements associated with technological shifts and geography. *Nat Commun* 11, 3868
- Nieves Colón, M.A., Pestle, W.J., Reynolds, A.W., Llamas, B., de la Fuente, C., Fowler, K., Skerry, K.M., Crespo Torres, E., Bustamante, C.D., and Stone, A.C. (2020) Ancient DNA reconstructs the genetic legacies of precontact Puerto Rico communities. *Mol Biol Evol* 37, 611–626
- Nores, R., and Demarchi, D.A. (2011) Análisis de haplogrupos mitocondriales en restos humanos de sitios arqueológicos de la provincia de Córdoba. *Revista Argentina de Antropología Biológica* 13, 43–54
- Tamura, K., Stecher, G., Peterson, D., and Filipiński, A.S. Kumar (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30, 2725–2729
- van Oven, M. (2015) PhyloTree Build 17: growing the human mitochondrial DNA tree. *Forensic Sci Int Genet Suppl Ser* 5, e392–e394
- van Oven, M., and Kayser, M. (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30, E386–E394
- Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., and Krause, J. (2016) EAGER: efficient ancient genome reconstruction. *Genome Biol* 17, 60
- Perego, U.A., Achilli, A., Anghofer, N., Accetturo, M., Pala, M., Olivieri, A., Hooshyar Kashani, B., Ritchie, K.H., Scozzari, R., Kong, Q.P., et al. (2009) Distinctive paleo Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol* 19, 1–8
- Perego, U.A., Anghofer, N., Pala, M., Olivieri, A., Lancioni, H., Hooshyar Kashani, B., Carossa, V., Ekins, J.E., Gómez Carballa, A., Huber, G., et al. (2010) The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Res* 20, 1174–1179
- Perez, S.I., Bernal, V., Gonzalez, P.N., Sardi, M., and Politis, G.G. (2009) Discrepancy between cranial and DNA data of early Americans: implications for American peopling. *PLoS One* 4, e5746
- Politis, G. (2008) The pampas and campos of South America. In *Handbook of South American Archaeology* (Springer), pp 235–261
- Politis, G.G., Gutiérrez, M.A., and Scabuzzo, C. (2014) In Estado actual de las investigaciones en el sitio arqueológico Arroyo Seco 2 (partido de tres arroyos, provincia de Buenos Aires, Argentina), G.G. Politis, M.A. Gutiérrez, and C. Scabuzzo, eds (INCUAPA CONICET, UNICEN), pp 329–370
- Politis, G.G., and Prates, L. (2018) Clocking the arrival of Homo sapiens in the southern cone of South America. In *New Perspectives on the Peopling of the Americas: Words, Bones, Genes, Tools*, K. Harvati, G. Jäger, and H. Reyes Centeno, eds (DFG Center for Advanced Studies Series), pp 79–106
- Politis, G.G., Gutiérrez, M.A., Rafuse, D.J., and Blasi, A. (2016) The arrival of Homo sapiens into the Southern Cone at 14,000 years ago. *PLoS One* 11, e0162870
- Posth, C., Renaud, G., Mittnik, A., Drucker, D.G., Rougier, H., Cupillard, C., Valentin, F., Thevenet, C., Furtwängler, A., Wißing, C., et al. (2016) Pleistocene mitochondrial genomes suggest a single major dispersal of non Africans and a late



- glacial population turnover in Europe. *Curr Biol* 26, 827–833
- Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., Lamnidis, T. C., Rohland, N., Nägele, K., Adamski, N., Bertolini, E., et al (2018) Reconstructing the deep population history of central and South America. *Cell* 175, 1185–1185.e22
- Postillone, M. B., Cobos, V. A., Urrutia, C., Dejean, C., Gonzalez, P. N., Perez, S. I., and Bernal, V. (2020a) Mitochondrial DNA diversity and evolutionary history of native human populations of Argentinean Northwest Patagonia (Argentina). *Hum Biol* 91, 57–79
- Postillone, M. B., Martinez, G., Flensburg, G., and Dejean, C. B. (2020b) First analysis of mitochondrial lineages from the eastern Pampa Patagonia transition during the final late Holocene. *Am J Phys Anthropol* 171, 659–670
- Prates, L., Politis, G. G., and Perez, S. I. (2020) Rapid radiation of humans in South America after the last glacial maximum: a radiocarbon based study. *PLoS one* 15, e0236023
- Prieto, A., Morano, S., Cárdenas, P., Sierpe, V., Calas, E., Christensen, M., Lefevre, C., Laroulandie, V., Espinosa Parrilla, Y., Ramirez, O., et al (2020) A Novel Child Burial from Tierra del Fuego: a Preliminary Report. *J Isl Coastal Archaeol* 15, 436–454
- Rabassa, J., Coronato, A., and Martínez, O. (2011) Late Cenozoic glaciations in Patagonia and Tierra del Fuego: an updated review. *Biol J Linn Soc* 103, 316–335
- Rafuse, D. J. (2017) Early to Middle holocene subsistence strategies in the pampas region: evidence from the Arroyo Seco 2 site. *J Archaeol Sci Rep* 12, 673–683
- Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila Arcos, M. C., Malaspina, A. S., et al (2015) Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349, aab3884
- Rasmussen, M., Anzick, S. L., Waters, M. R., Skoglund, P., DeGiorgio, M., Stafford, T. W., Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S. M., et al (2014) The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506, 225–229
- Rieux, A., and Balloux, F. (2016) Inferences from tip calibrated phylogenies: a review and a practical guide. *Mol Ecol* 25, 1911–1924
- Rieux, A., Eriksson, A., Li, M., Sobkowiak, B., Weinert, L. A., Warmuth, V., Ruiz Linares, A., Manica, A., and Balloux, F. (2014) Improved calibration of the human mitochondrial clock using ancient genomes. *Mol Biol Evol* 31, 2780–2792
- Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015) Partial uracil DNA glycosylase treatment for screening of ancient DNA. *Philos Trans R Soc Lond B Biol Sci* 370, 20130624
- de Saint Pierre, M. (2017) Antiquity of mtDNA lineage D1g from the southern cone of South America supports pre Clovis migration. *Quat Int* 444, 19–25
- de Saint Pierre, M., Bravi, C. M., Motti, J. M. B., Fuku, N., Tanaka, M., Llop, E., Bonatto, S. L., and Moraga, M. (2012a) An alternative model for the early peopling of Southern South America revealed by analyses of three mitochondrial DNA haplogroups. *PLoS One* 7, e43486
- de Saint Pierre, M., Gandini, F., Perego, U. A., Bodner, M., Gómez Carballa, A., Corach, D., Angerhofer, N., Woodward, S. R., Semino, O., Salas, A., et al (2012b) Arrival of paleo Indians to the southern cone of South America: new clues from mitogenomes. *PLoS One* 7, e51311–9
- Scheiffler, N. A., Messineo, P. G., and Antíñir, A. (2017) Cazadores recolectores en el sistema lagunar Hinojo Las Tunas (región pampeana, área Oeste) durante la transición Holoceno temprano medio y tardío. Primeros resultados de las investigaciones arqueológicas. *Comechingonia, Revista de Arqueología* 21, 287–314
- Schubert, M., and Lindgreen, S. (2016) Adapter removal v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* 9, 88
- Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., and Krause, J. M. J. (2014) Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc Natl Acad Sci USA* 111, 2229–2234
- Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V., and Richards, M. B. (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84, 740–759

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Pfu Turbo Cx Hotstart DNA Polymerase	Agilent Technologies	600412
Herculase II Fusion DNA Polymerase	Agilent Technologies	600679
2x HI RPM hybridization buffer	Agilent Technologies	5190 0403
0.5 M EDTA pH 8.0	BioExpress	E177
Sera Mag Magnetic Speed beads Carboxylate Modified (1 μm, 3EDAC/PAS)	GE LifeScience	6.51521E+13
USER enzyme	New England Biolabs	M5505
UGI	New England Biolabs	M0281
Bst DNA Polymerase 2.0, large frag.	New England Biolabs	M0537
PE buffer concentrate	QIAGEN	19065
Proteinase K	Sigma Aldrich	P6556
Guanidine hydrochloride	Sigma Aldrich	G3272
3M Sodium Acetate (pH 5.2)	Sigma Aldrich	S7899
Water	Sigma Aldrich	W4502
Tween 20	Sigma Aldrich	P9416
Isopropanol	Sigma Aldrich	650447
Ethanol	Sigma Aldrich	E7023
5M NaCl	Sigma Aldrich	S5150
1M NaOH	Sigma Aldrich	71463
20% SDS	Sigma Aldrich	5030
PEG 8000	Sigma Aldrich	89510
1 M Tris HCl pH 8.0	Sigma Aldrich	AM9856
dNTP Mix	Thermo Fisher Scientific	R1121
ATP	Thermo Fisher Scientific	R0441
10x Buffer Tango	Thermo Fisher Scientific	BY5
T4 Polynucleotide Kinase	Thermo Fisher Scientific	EK0032
T4 DNA Polymerase	Thermo Fisher Scientific	EP0062
T4 DNA Ligase	Thermo Fisher Scientific	EL0011
Maxima SYBR Green kit	Thermo Fisher Scientific	K0251
50x Denhardt's solution	Thermo Fisher Scientific	750018
SSC Buffer (20x)	Thermo Fisher Scientific	AM9770
GeneAmp 10x PCR Gold Buffer	Thermo Fisher Scientific	4379874
Dynabeads MyOne Streptavidin T1	Thermo Fisher Scientific	65602
Salmon sperm DNA	Thermo Fisher Scientific	15632 011
Human Cot 1 DNA	Thermo Fisher Scientific	15279011
DyNAmo HS SYBR Green qPCR Kit	Thermo Fisher Scientific	F410L
Methanol, certified ACS	WWR	EM MX0485 3
Acetone, certified ACS	WWR	BDH1101 4LP
Dichloromethane, certified ACS	WWR	EMD DX0835 3
Hydrochloric acid, 6N, 0.5N & 0.01N	WWR	EMD HX0603 3

(Continued on next page)

<i>Continued</i>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
High Pure Extender from Viral Nucleic Acid Large Volume Kit	Roche	5114403001
MinElute PCR Purification Kit	QIAGEN	28006
NextSeq® 500/550 High Output Kit v2 (150 cycles)	Illumina	FC 404 2002
HiSeq® 4000 SBS Kit (50/75 cycles)	Illumina	FC 410 1001/2
Deposited Data		
Sequencing Data	GenBank	MW291661 MW291678
Genotype Data	Reich Lab website	https://reich.hms.harvard.edu/datasets
Software and Algorithms		
Samtools v1.11	Li et al., 2009	http://samtools.sourceforge.net/
BWA v0.7.17 r1188	Li and Durbin, 2009	http://bio-bwa.sourceforge.net/
SeqPrep v2	https://github.com/jstjohn/SeqPrep	https://github.com/jstjohn/SeqPrep
AdapterRemoval v2	Schubert et al., 2016	https://github.com/MikkelSchubert/adapterremoval
Dedeup v0.12.07	Peltzer et al., 2016	https://eager.readthedocs.io/en/latest/
PMDtools v0.60	Skoglund et al., 2014	https://github.com/pontusk/PMDtools
Haplogrep v2.0		https://haplogrep.uibk.ac.at/index.html
ContamMix	Fu et al. (2013)	https://github.com/DReichLab/ADNA Tools
MEGA6	Tamura et al., 2013	https://www.megasoftware.net
mapDamage2.0	Jónsson et al. (2013)	https://gionhac.github.io/mapDamage/
MUSCLE v3.8.1551	Edgar, 2004	https://www.drive5.com/muscle/
Geneious v9.1.8		https://www.geneious.com/
PartitionFinder v2.1.1	Lanfear et al. (2012)	https://www.robertlanfear.com/partitionfinder/
BEAST 1.8.0	Drummond and Rambaut (2007)	https://beast.community
Tracer v1.7.1		https://github.com/beast-dev/tracer/releases/tag/v1.7.1
TreeAnnotator v2.6.0		https://beast2.blogs.auckland.ac.nz/treeannotator/
FigTree v1.4.4		http://tree.bio.ed.ac.uk/software/

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to and will be fulfilled by the lead contact Xavier Roca-Rada (xavier.rocarada@adelaide.edu.au).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The data underlying this article are available in GenBank under accession numbers MW291661–MW291678.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Archaeological context

We analyzed 18 whole mitochondrial genomes—7 of which are reported in this study for the first time—from human skeletal remains from three Early to Late Holocene archaeological sites in the Argentinian Pampas:

Arroyo Seco 2: 13

Laguna de los Pampas: 1

Laguna Chica: 4

Brief description of archaeological sites

The **Arroyo Seco 2 (AS2)** site (Figure S1) is located outside the city of Tres Arroyos. It is an open-air archaeological site situated on a low-lying knoll between a small temporary lake and a shallow creek (S 38°36' and W 60°24'). From 1979 to the most recent excavations in 2019, a total of 77 units (314 m²) were opened in the site, including shovel tests and 3 long trenches. AS2 is a multicomponent site with several occupation episodes and a chronological range from the Late Pleistocene to historical times (Politis et al., 2014; Politis et al., 2016; Rafuse, 2017).

The hunting/scavenging events of the early hunter-gatherers at the AS2 site likely reflect at least two episodes. Temporary campsites were established in the area for the butchering/processing of *Megatherium* and the extinct horses *Equus neogeus* (at ~12,200 ¹⁴C years BP) and *Hippidon* (at ca. 11,180 ¹⁴C years BP). During this period, other species of megafauna (*Toxodon*, *Hemiauchenia* and *Glossotherium*) were at the site, although the evidence of human agency is still inconclusive for these taxa. Lithic artefacts are mainly unifacially retouched quartzite flakes. Formal tools include side and end scrapers as well as bifacially retouched knives. All raw materials are non-local and have different proveniences (Tandilia Hill Range, Ventania Hill Range, Atlantic seashore, etc.). After the extinction of the megamammals, there is a gap in the human occupation of AS2. During the Early Holocene (~8,500 ¹⁴C years BP), the site was occupied again by guanaco (*Lama guanicoe*) hunters, who established several overlapping campsites. Medium and large triangular projectile points, as well as a variety of unifacial quartzite and chert tools, characterize the lithic technology during this period. Around this time, funerary activities are abundant of human skeletons (n = 50) of both sexes and all age categories, dating between 7,819 ± 61 ¹⁴C years BP and 4,487 ± 45 ¹⁴C years BP (n = 27 dates). Thirteen of these skeletons were analyzed for this study (Table 1). The burial modalities are varied, including simple and multiple primary burials and simple and multiple funeral packages. The earliest level of burial included five skeletons with projectile points (midsized triangular stemless) stuck between and within the bones. Grave goods consisted of marine shell beads and necklaces of canid canines were recorded in some skeletons, indicating an early and complex treatment of the dead.

The **Laguna de los Pampas (LLP)** (Figure S2) and **Laguna Chica (LC)** (Figure S3) archaeological localities are placed in the Central Pampas Dunefields, one of the aeolian unit system from the eastern center of Argentina. These localities are situated in the current margins of shallow lakes, with recurrent hunter-gatherer occupations through the Holocene (Messineo et al., 2019). In LLP, eight burials were recovered in two sectors of a lake's beach (sector 1: S 35°19'42" and W 61°31'50"; sector 2: S 35°19'56" and W 61°31'53") due to re-exposition by water erosion. Moreover, isolated human bone remains corresponding to other eight individuals were found. The inhumations dated from the Early and Middle Holocene (Messineo et al., 2019). Burial 1 of LLP includes a multiple primary burial composed of two individuals: one adult female dating from 8,971 ± 77 ¹⁴C years BP (LLP.S2.E1, AA-90127) and one sub-adult of 2–4 years old dating from 8,835 ± 83 ¹⁴C years BP (L.LLP.S2.43, AA-93221). Burials 2 and 3 are simple primary and they dated back to 5,688 ± 36 ¹⁴C years BP (L.LLP.S2.1062, AA-108848) and 5,819 ± 24 ¹⁴C years BP (L.LLP.S1.E3, MAMS-24770), respectively. Bone remains of burial 4 and 5 were found on the surface and the method of the burials were not determined. They dated back to 5,924 ± 40 ¹⁴C years BP (L.LLP.S2.E4, AA-106730) and 7,089 ± 37 ¹⁴C years BP (L.LLP.S1.2706, AA-110832), respectively (Messineo et al., 2019). Burials 6, 7 and 8 are simple primary but they still do not have radiocarbon dates. Only one of these skeletons was analyzed in this study (Table 1).

The **Laguna Chica (LCH)** archaeological locality is located in the current margins of a small, temporal, shallow lake (S 36°5'8.80" and W 62°20'31.44") in the southeast of the Hinojo-Las Tunas Shallow Lake System. Seven burials were identified: four in Sector A located in the southern part of the shallow lake and three in Sector B in the western area. Inhumations dated from the Middle to the Late Holocene (Scheifler et al., 2017; Messineo et al., 2019). These burials were partially exposed at the lake's beach due to water erosion. Only the cranium, the mandible and some remains of the thorax region were recovered from Burial N 1 (sample SC50-L763, LCH.E1.3). The morphological study of this individual's scarce remains determined a probable male adult (genetically confirmed) and dated back to 6,870 ± 30 ¹⁴C years BP (PSUAMS-6965). This



individual was associated with a decorated pendant made of a canine of a jaguar (yaguareté, *Panthera onca*). Burial N 2 contained multiple primary burials represented with two individuals who had a dorsal disposition of the bodies with the lower limbs flexed. Individual N 1 (sample SC50-L762, LCH.E2-11.2) was a female adult that dated back to $5,930 \pm 15^{14}\text{C}$ years BP (UCIAMS-185302). Individual N 2 (sample SC50-L761, LCH.E2-12.1) was a male adult that dated back to $6,080 \pm 15^{14}\text{C}$ years BP (UCIAMS-185303). Burial N 4 (sample SC50-L764, LCH.E4.4) was an infant of undetermined sex. This individual dated back to $1,750 \pm 15^{14}\text{C}$ years BP (UCIAMS-185301). Burials 3, 5 and 7 are simple primary represented by adult. The individual of the Burial 7 was associated with a bone tool made of a metapodial of guanaco (*Lama guanicoe*). Bone remains of burial 6 were found on the surface and the method of the burials were not determined. These last burials do not have radiocarbon dates yet.

Stratigraphic excavations have been performed in LLP and LCH locality, and abundant lithic materials and bones have been found. These lithic materials are characterized by a predominance of orthoquartzite and chert, which outcrops in the Tandilia Hill Range, 300 km far from the sites. Other lithic raw materials, such as granite, silicified dolomite, metaquartzite, rhyolite, basalt, silex, micaceous schist, and obsidian, were present in low frequencies. Some of these rocks come from diverse sectors of the Pampas (Ventania Hill Range, xerophytic woodland, and Tehuelche Mantle), and extra-regional areas. A high diversity of lithic tools was found, including side scrapers, end-scrapers, knives, multipurpose tools, triangular projectile points, etc. (Messineo et al., 2019). In addition, a standardized bone technology was also identified in LLP and centered in the production of beveled tools and blunted points made on guanaco tibiae (Messineo et al., 2019). The analysis of the material indicates that these sites were occupied during Holocene times. It might represent a succession of residential camps in the border of the shallow lakes by hunter-gatherers focused on the exploitation of guanaco (*Lama guanicoe*), although armadillo shell plates and *Rheidae* eggshells were well represented. In addition, complete eggs of greater rhea with a small perforation at the minor pole were found in both sites. They were possibly employed as flasks for storing water by hunter-gatherer groups in environments with shortage of sources of fresh water (Messineo et al., 2019).

METHODS DETAILS

Radiocarbon dating and calibration

The new AMS ^{14}C date reported in this article (LCH.E1.3; 7,724–7,589 calBP) has been generated at the Pennsylvania State University AMS Radiocarbon Laboratory, which is part of the Energy and Environmental Sustainability Laboratories (EESL). The Radiocarbon Laboratory is equipped with a National Electronics Corporation compact spectrometer with a 0.5MV accelerator (NEC 1.5SDH-1). The primary modifications impacting analytical measurement error are the use of a spherical ionizer ion source operating at high cathode voltage (7V) to generate intense C-beams, plus injection beam line changes for better ion-optical matching to the accelerator. The injector modifications include a second einzel lens plus an increased ion source voltage running at 47.5 kV combined with a redesigned large-gap injector magnet (DF01319) (Beverly et al., 2010). These alterations allow for analytical error in the 2–3‰ range for near modern samples under currents of up to 200 μA of ^{12}C and routinely generating 100–120 μA of ^{12}C from $\sim 0.7\text{mg}$ C samples. Radiocarbon ages are $\delta^{13}\text{C}$ -corrected for mass dependent fractionation with $\delta^{13}\text{C}$ values measured on the AMS and compared with OXII standards for normalization. Pre-treatment of sample (collagen extraction) was made in Laboratory of Pre-treatment for Isotopic Analysis (LAPREI) which is part of the Instituto de Investigaciones Arqueológicas y Paleontológicas del Cuaternario Pampeano (INCUAPA), Argentina, following standard protocols.

XAD Amino Acids: Sample was physically cleaned using hand tools and sectioned with disposable Dremel cut-off wheels and then demineralized in 0.5 N HCl for 2–3 days at 5°C. The demineralized collagen pseudomorph was gelatinized at 60°C in 1–2 mL 0.01 N HCl for eight to ten hours. Sample gelatine was pipetted into a pre-cleaned 10 mL disposable syringe with an attached 0.45 mm Millex Durapore PVDF filter (pre-cleaned with methanol and Nanopure H₂O) and driven into a thick-walled culture tube. The filtered solution was lyophilized, and percent gelatinization and yield determined by weight. The sample gelatine was then hydrolyzed in 2 mL 6 N HCl for 22 h at 110°C. Supelco ENVI-ChroSPE (Solid Phase Extraction; Sigma-Aldrich) columns were prepped with 2 washes of HCl (2 mL) and rinsed with 10 mL DI H₂O. With a 0.45 mm Millex Durapore filter attached, the SPE Column was equilibrated with 50 mL 6 N HCl and the washings discarded. 2 mL collagen hydrolyzate as HCl was pipetted onto the SPE column and driven with an additional 10 mL 6 N HCl dropwise with the syringe into a 20mml culture tube. The hydrolyzate was finally dried into a



viscous syrup by passing UHP N₂ gas over the sample heated at 50°C for 12 h. All calibrated ¹⁴C ages were calculated using OxCal version 4.3 (Bronk Ramsey, 2013).

Ancient DNA laboratory work

All samples were processed in the dedicated clean rooms at UCSC Paleogenomics Lab in Santa Cruz (UC-PL) or at David Reich's lab at Harvard Medical School (HMS), following strict precautions to minimize contamination (Llamas et al., 2017). DNA was extracted from tooth roots and bone samples (Table S1) using a commonly applied extraction method optimized to retain short DNA fragments (Dabney et al., 2013) with modifications described in Posth et al. (2018). Partially UDG treated internally barcoded double stranded DNA (dsDNA) libraries were generated following the protocol by Rohland et al. (2015). Sequencing libraries were enriched for mitochondrial DNA using in-solution hybridization capture at HMS or UC-PL following the protocols described in Llamas et al. (2016). Enriched libraries were sequenced (2x150) on Illumina NextSeq sequencers at HMS or UC-PL. Raw reads were processed and mapped against hg19 as described in Posth et al. (2018).

QUANTIFICATION AND STATISTICAL ANALYSIS

Contamination estimation

We evaluated evidence for ancient DNA authenticity by measuring the rate of damage in the first nucleotide, flagging individuals as potentially contaminated if they had a less than 3% cytosine-to-thymine substitution rate. Additionally, we estimated mtDNA contamination using contamMix version 1.0–12 Fu et al. (2013). The software was run with down-sampling to 50x for samples above that coverage, –trimBases 2, 8 threads, 4 chains, and 2 copies, taking the first one that finished.

Mitochondrial DNA analyses

The sequencing read pileups were visualized in Geneious v9.1.8 (Biomatters; available from <http://www.geneious.com/>). SNPs were called in Geneious for all polymorphisms with minimum coverage 2 and a minimum variant frequency 0.7. The assembly and the resulting list of SNPs were verified manually and compared to SNPs reported at phyloree.org (mtDNA tree Build 17 [18 Feb 2016] (van Oven, 2015)) as described in Llamas et al. (2016).

Phylogenetic analyses

A total of 117 pre-Columbian and modern-day haplotypes belonging to haplogroups D1 and D4h3a and one C1a outgroup were compiled (Bodner et al., 2012; de Saint Pierre et al., 2012a, 2012b; Rasmussen et al., 2014; van Oven, 2015; Llamas et al., 2016; Posth et al., 2018) and aligned using Geneious v2019.2.1. Common indels and mutational hotspots were excluded (nucleotide positions 309.1C(C), 315.1C, AC indels at 515–522, 16182C, 16183C, 16193.1C(C), and C16519T) (van Oven and Kayser, 2009). PartitionFinder v2.1.1 (Lanfear et al., 2012) was used to select best-fit partitioning schemes and models of molecular evolution. The best model was TRN+G (TN93) for the whole mitogenome. BEAST 1.8.0 (Drummond and Rambaut, 2007) was used with tip date calibrations to reconstruct the phylogeny and constrained monophyly for D, D1g, D1j and D4h3a.

Strict clock was rejected according to a preliminary analysis with uncorrelated lognormal relaxed clock, so a relaxed clock was used. Three Markov Chain Monte Carlo (MCMC) chains of 100 million steps were performed, with sampling of parameters every 10,000 steps. Parameter traces were monitored in Tracer v1.7.1 to ensure convergence of the MCMC chains and effective sample sizes higher than 200. The three chains performed with highly similar results. The initial 10 million steps were discarded as burn-in using TreeAnnotator v2.6.0 and the final tree was visualized using FigTree v1.4.4.

Supplemental information

- [Download .pdf \(11.4 MB\)](#)
-

Document S1. Figures S1–S5

- [Download .xlsx \(.01 MB\)](#)
-

Table S1. Complete information of the 18 Early to Late Holocene individuals from the Argentinian Pampas, related to table 1

- [Download .xlsx \(.01 MB\)](#)
-

Table S2. Haplogroup frequencies of other published Middle to Late Holocene pre-Columbian Eastern Patagonian and SCSA populations, related to figure 1

Thesis Discussion

The discovery that DNA survives in archaeological remains¹ revolutionised the fields of archaeology, anthropology, and evolutionary biology. Indeed, the last 40 years saw major technological and computational developments in the life sciences as well as new advances in the social sciences and humanities alongside paleogenetic research progresses, bringing us closer to a better understanding of the history of life on our planet.

Nowadays, molecular anthropology uses aDNA to answer questions about kinship structure of burial sites (e.g., kinship practices at chambered tombs in Early Neolithic Britain²), geographical origins of human populations (e.g., Bronze Age migrations from the Eurasian Steppe³), the migration patterns around the world [see Appendix or Skoglund and Mathieson, (2018)⁴], and molecular adaptations to novel environments⁵. Hence, aDNA can be used as a tool to tackle several questions from a narrow to a broad view, i.e., from a specific archaeological context to global demographic processes. Indeed, in this thesis I covered a large range of projects following this perspective, investigating: (i) a medical condition diagnosed from a single individual; (ii) the diachronic genetic changes at one archaeological site; (iii) the genetic history of a historical and cultural region; and (iv) the peopling of a subcontinental area. All these projects have been grouped according to their respective geographical area: Iberia, Mesoamerica, and the Southern Cone of South America.

Part I: Iberia

Paleogenomic research has thus far been most successful when reconstructing ancestry profiles of past individuals and their population affinities but aDNA analysis can also contribute to our understanding of the origin, incidence, and evolution of human health and disease^{6,7}. However, commonly applied medical methodologies may be highly inefficient due to the degraded nature of aDNA. Hence, our understanding of past human health and disease remains limited.

In Chapter I, I led an unprecedented integration of methods and knowledge in a transdisciplinary study—including morphology, osteology, and genetics—to describe a 1,000-year-old case of Klinefelter’s Syndrome (KS) from Portugal, highlighting the key role that paleogenetics can play in evolutionary medicine⁸. In fact, KS has probably been overlooked in the study of archaeological human remains due to its relatively low prevalence and difficulties associated with accurate diagnosis, which in living people relies mostly on karyotype analysis^{9–11}—an impossible fit for heavily fragmented aDNA. Moreover, the recent developments in human aDNA analyses do not usually include whole-genome sequencing but instead focus on the 1240k nuclear SNPs panel for population genetics analyses¹². Likewise, human skeletons found in archaeological contexts are usually not sufficiently well preserved to allow for complete and unbiased morphological and osteological analyses. Despite thousands of published ancient humans with genome-wide data¹³, only five putative KS cases have been reported from aDNA analysis before, none of which included a thorough morphological and osteological study of the skeletal remains nor an in-depth statistical assessment of the ploidy level of the sex chromosomes as the one presented here^{14–18}.

The reported individual was found in a mediaeval archaeological site of Torre Velha, north-eastern Portugal. This individual died after reaching adulthood, which allowed the analysis of a virtually complete adult skeleton and the evaluation of characteristic physical traits associated with KS, such as a robust male skeleton, a higher height compared to other individuals from Torre Velha, a bi-iliac width larger than expected for ancient Portuguese males, and a probable malocclusion and maxillary prognathism denounced by atypical dental wear. Bone densitometry analysis showed normal bone mineral density values, which is not consistent with a diagnostic of osteoporosis at the time of death, a trait previously associated with ~10-40% of adult KS cases¹⁹.

Genetic analysis confirmed KS through a variety of methods, including calculating the X- and Y-chromosomes read dosage (scaled coverage by sequence length; 0.902 and 0.298 respectively), detecting the presence of two X chromosomes²⁰, observing the expected X-chromosome heterozygosity²¹, and determining the Y-chromosome haplogroup. This work also introduces a novel statistical implementation that allows to probabilistically assign individuals to karyotypes based on the number of reads mapping to the X, Y, and autosomal chromosomes. This method showed an overwhelming support for an assignment of this individual to karyotype XXY and discarded any possibility that these results could come from contamination (posterior probability of approximately 0). The method has applications beyond aDNA and can be a potentially efficient statistical tool to analyse fragmented DNA from a variety of sources (e.g., cell-free DNA and DNA from forensic cases). In fact, cell-free DNA has recently been used as a marker for non-invasive prenatal diagnostics to detect aneuploidies²² and the application of our newly developed method would help the diagnosis of sex chromosome aneuploidies.

This study is a clear example of how transdisciplinary paleogenetic research has the potential to dramatically advance our understanding of past human health and push forward the medical field. This is best illustrated by the recent award of the 2022 Nobel prize in Physiology or Medicine to Svante Pääbo for his discoveries concerning the genomes of extinct hominins and human evolution.

Chapter II takes aDNA back to its original premise of tackling archaeological research questions with genetics and describes a study from the evaluation of possible relationships between individuals excavated at the same burial site to the exploration of population demographic processes, unifying various fields into a unique transdisciplinary research

framework. Specifically, I investigated the genetic context of the Mas d'en Boixos (MDB) archaeological site, located in the central Pre-littoral depression of Catalonia (Spain).

I have been studying MDB for a long time. During my Masters in Biological Anthropology at the Autonomous University of Barcelona back in 2017/18, I gained experience in aDNA laboratory techniques working with these samples. DNA was extracted in clean rooms applying commonly used methods optimised to retrieve short degraded aDNA fragments²³. At the time, the laboratory was still using RFLP analyses and Sanger-sequenced HVR1 amplicons, not genome-wide SNP capture or shotgun sequencing. During my Masters, I performed HVR1 amplicons sequencing using NGS in a service provider facility not specialised for aDNA work, which led to severe cross-contamination and misinterpretation of the results. During my PhD, I re-started the project and I generated partially UDG-treated double-stranded DNA libraries at ACAD, and I performed mitochondrial and nuclear genome-wide SNP enrichments that led to the preliminary results presented in Chapter II. It has been a long journey, but it offered me a unique opportunity to explore and learn from all the incremental methodological steps that the aDNA field has gone through.

I genetically analysed two individuals from the Middle Neolithic, nine from the Early Bronze Age and three (previously published) from the Iron Age²⁴. The Middle Neolithic burial was composed by either a mother and a son or two siblings. They shared the same mitochondrial haplotype (K1a4a1) and were most closely related to present-day Sardinians and Basques. While the genetic analysis classified these individuals as male and female, the anthropological study assigned both as females. Even though morphology is useful for several archaeological analyses, the variability of the human skeleton can lead to misdiagnosis that only genetics can discern.

All the individuals from the Early Bronze Age hypogeum carried different mitochondrial haplogroups and there was a second-degree relationship between one male and two other individuals. These results corroborated the previously proposed hypothesis that collective funerary practices during the Catalan Bronze Age were composed by an extended family structure unifying kinship and culture in small societies regardless of the different practices between coastal and mountain sites²⁵. All the individuals carried Steppe ancestry, and males carried R1b Y chromosome lineages, which is in accordance with the arrival of people from the Eurasian Steppe^{18,24,26}.

My preliminary population genetics analyses suggested that the Bronze Age transition in Iberia was not homogenous, as the Steppe ancestry expansion did not reach the whole peninsula in the same way. Even though it has been suggested this expansion was male driven²⁴, some mitochondrial lineages associated with Eastern populations have been detected in MDB, whereby females were probably migrating alongside males. In fact, Villalba-Mouco et al. 2021¹⁸ did not find any signal for male bias related to Steppe ancestry. Only more sampling in the Pyrenees, central Iberia and Portugal will help discerning more nuanced details of the Bronze Age transition both spatially and temporally.

Finally, the two males and the female from the Iron Age burials previously analysed by Olalde et al. (2019)²⁴ were relatively distinct from each other and closer to the Early Bronze Age individuals than the Middle Neolithic individuals. The genetic history of Iron Age Iberia is still highly unexplored, and it remains unclear if the only 24 previously published individuals are representative of the main Iron Age Iberian population whose most common funerary practice was cremation²⁷. Generally, all previously published Spanish Iron Age individuals have a higher Steppe ancestry than Bronze Age individuals, but lack of sampling in some areas, including Portugal, hamper the possibility to analyse genetic structure within the Peninsula.

Part II: Pre-Columbian Mesoamerica

Mesoamerica represents a critical tropical transition in the southward journey of the First Peoples of the Americas^{28,29}. Since its initial settling, Mesoamerica was highly populated by several cultures connected by commerce and farming. Hence, Mesoamericans probably carried a genetic diversity that was partly lost during the Spanish conquest and the subsequent colonial period³⁰.

In Chapter III, I reviewed previously published aDNA studies in pre-Columbian Mesoamerica. Despite a rich archaeological record and Indigenous cultural diversity, aDNA studies are limited and have thus far focused on the study of mitochondrial DNA in the Basin of Mexico as well as the Yucatán Peninsula and nearby territories, particularly during the Postclassic period (900 – 1,519 CE). Despite limitations associated with the poor preservation of samples in tropical areas, recent methodological improvements allow for a deeper analysis of genetic variation in pre-Columbian Mesoamerica populations to decipher their past genetic history and the mechanisms of adaptation to changing environments.

In Chapter IV, I investigated the past genetic history of Pre-Columbian Mesoamerica by analysing 25 newly sequenced ancient genomes from five different archaeological sites across Mesoamerica spanning the time periods between 320 and 1,400 CE. Despite the small sample size and setbacks related to the COVID-19 pandemic, this chapter offers the first (and ongoing) Mesoamerican aDNA genetic screening that includes full mitogenomes and nuclear data.

As the degraded nature of aDNA is further exacerbated in tropical environments (poor preservation in hot and wet environments^{31,32}), the first aim of this chapter was to evaluate sample preservation and success rate when attempting to retrieve endogenous human DNA from Mesoamerican skeletal remains. Also, in order to authenticate and reproduce results,

samples were processed in different laboratories, specifically at ACAD, UCSC and the University of Bordeaux. Skeletal remains covered with cinnabar or presenting mould stains, dryness or a general observable bad preservation did not yield DNA. As expected, good quality aDNA was retrieved from teeth and petrous bones^{33,34}. Surprisingly, metacarpals performed better than teeth, unlike previously reported by Parker et al. (2020)³⁵ who concluded that samples from the femur, metacarpal, ischial tuberosity, ribs, and clavícula, as well as any sample derived from cancellous (spongy) material were unlikely to yield high amounts of endogenous human DNA. Additionally, I prepared and tested different libraries and found that single-stranded DNA libraries presented a higher proportion of endogenous DNA and retained shorter DNA fragments. However, these libraries also presented lower and asymmetrical rates of post-mortem damage at read termini, as previously demonstrated³⁶.

Preliminary results suggest that pre-Columbian Mesoamerica had a high mitochondrial diversity, with A2 being the most frequent haplogroup (48%), which is in accordance with previously published Mesoamerican aDNA data based on RFLP analyses and Sanger-sequenced HVR1³⁷. All the analysed mitochondrial haplotypes were different—except for one that was found in two individuals from Malpaís Prieto—and none of the samples were related to each other according to kinship analyses performed using genome-wide SNP data. Nuclear genetic differences suggest a geographical correlation probably explained by isolation by distance. Concretely, individuals from the Basin of Mexico and Michoacán showed equidistant genetic affinities to present-day central and southern Mexicans (see Chapter IV, Figure 10).

In the Basin of Mexico, my paleogenetic analyses suggest that Teopancazco and Tlatelolco were genetically quite diverse and continuous from the Classic to the Postclassic (~300 – 1,400 CE). Further studies focusing on the Classic (150 – 650 CE) cosmopolitan neighbourhood of Teopancazco, will help resolving the demographic history of Teotihuacan and contextualize

the population history of Mesoamerica during this period. Likewise, studies focusing on the understudied Epiclassic (650 – 900 CE) as well as the Postclassic (900 – 1,519 CE) will enable testing for population continuity and determining the demographic processes associated with the fall of Teotihuacan and the subsequent empires.

In Michoacán, nuclear DNA data suggested a possible genetic continuity between the Epiclassic to the Middle Postclassic (~600 – 1,450 CE), but mitochondrial haplogroup frequencies instead suggest a possible genetic turnover, possibly related to the arrival of the Uacúsechas and Chichimecas from northern Mexico³⁸. Despite its advantages for the inference of population history, mitochondrial DNA is only a single locus that cannot capture the full complexity of such history^{39,40}. In the future, additional diachronic sampling in Michoacán and northern regions of Mexico will help disentangling these seemingly contrasting hypotheses.

The aDNA analysis of Naachtun revealed their high genetic affinity to present-day Mayans and a group of ancient individuals predating the settlement of Mesoamerica that include the Anzick individual⁴¹ and the oldest known Central and northern South American samples from Belize^{29,42} and Venezuela⁴³. Although the Mayan culture remained in the Yucatán peninsula throughout different cultural changes^{44,45}, its populations were never completely isolated and were also highly affected by the collapse of the Classic period. A genetic study across the whole Mayan region will help explaining their genetic affinity to other regions of Mesoamerica.

The data presented in this chapter are a first attempt to disentangle the genetic history of Mesoamerica during the first and a half millennium CE. However, interpretations solely based on these results must be taken with caution due to small sample sizes. Additional sampling across Mesoamerica is thus necessary to unravel population structure, adaptation, and detailed demographic inferences of its peoples. Mesoamerica is located at the crossroad between North and South America and remains genetically underexplored despite its key geographic position

and the extensive archaeological and anthropological records available. The inclusion of key areas of Mesoamerica is therefore essential for comprehensive genetic studies on: (i) the Gulf Coast, initially peopled by the Olmecs, the earliest known major Mesoamerican civilisation; (ii) Oaxaca, home to the important Preclassic–Classic city of Monte Albán; (iii) the West, including Michoacán, a geographically diverse area that led to the rise of highly diverse societies characterised by the shaft tomb tradition in the Preclassic and Classic periods; and (iv) the North, a transition area between Mesoamerica and Aridoamerica inhabited by both hunter-gatherers and farmers that was nearly abandoned during the Postclassic period and probably had a huge impact on Mesoamerican populations.

Part III: Southern Cone of South America

The Southern Cone of South America (SCSA) is a key study region to build a comprehensive picture of the peopling of the Americas as it is geographically the most distant location from Beringia. In Chapter V, I analysed eighteen ancient mitochondrial genomes—seven of which are novel—from three Early to Late Holocene (10,000–1,500 years ago) archaeological sites from the understudied Argentinian Pampas⁴⁶.

The Early to Middle Holocene Pampas presented a distinctive mitochondrial genetic makeup—high diversity, predominance of haplogroup C1 and presence of B2—compared to subsequent pre-Columbian populations from the SCSA—predominance of D1 and absence of B2.

Furthermore, this study reported the earliest individuals carrying mitochondrial subhaplogroups D1j and D1g dating from Early and Middle Holocene, respectively⁴⁶. In fact, D1g and D1j are nearly exclusively found in present-day populations from the SCSA and subhaplogroup D4h3a has its highest frequencies in the extreme south. As specific mitochondrial haplogroups with highly restricted geographic distribution can be used to infer

the local population history, I used D1g, D1j, and D4h3a to investigate the early peopling of the SCSA. Using these ancient mitochondrial genomes as deep calibration time points in Bayesian phylogenetic reconstructions, the estimated divergence dates for D1g, D1j, and D4h3a were synchronous (~15,600 years ago), suggesting that the three clades emerged during the initial peopling of the Americas (~16,000 years ago)²⁸ and diversified en route to South America. At ~9,000 years ago, a previously reported partial population replacement²⁹ as well as founder effects and genetic drift might have erased D1g and D1j outside the SCSA but not D4h3a. Alternatively, Prates et al. (2020)⁴⁷ suggested that the earliest chronological threshold for the peopling of South America was ~15,500 years ago and could have led to the diversification of D1g and D1j in South America. However, this scenario would imply that although D4h3a emerged at the same time, it might have had a different genetic history⁴⁸⁻⁵⁰.

This project is a clear example of how mitochondrial DNA is still highly relevant for phylogenetic and phylogeographic inferences to unravel the evolutionary path of human migrations. Even though there are limitations associated with the study of mtDNA as this genetic marker only represents the evolutionary history of the female population at a single locus, it does represent a fundamental element of a population's heritage and evolutionary history—certainly not the only, or definitive, source of genetic information, but nonetheless valid and accurate^{39,40}. One of the limitations from the present study is that the current mitochondrial data do not allow to infer specific migration routes. Only more aDNA research on early coastal and Amazonian sites of South America and availability of autosomal genetic markers will help to unravel if the peopling of the Pampas occurred via the Pacific coast followed by the crossing of the Southern Andes, or instead through an Atlantic/inland route.

aDNA ethics

Ethical issues regarding aDNA research encompass conceptualization, sampling and communication with appropriate local stakeholders and descendent communities⁵¹. Some recent ethical guidelines regarding worldwide DNA research on human archaeological remains have been raised^{51,52}: (i) researchers must ensure that all regulations were followed in the places where they work and from which the human remains derived; (ii) researchers must prepare a detailed plan prior to beginning any study; (iii) researchers must minimize damage to human remains; (iv) researchers must ensure that data are made available—publicly or upon request—following publication to allow critical re-examination of scientific findings; and (v) researchers must engage with stakeholders from the beginning of a study and ensure respect and sensitivity to other stakeholder perspectives.

Concretely in Mexico, consultation with present-day Indigenous populations for destructive analysis of human archaeological remains is not a standard procedure. Assigning a present-day population as the single direct descendant of human archaeological remains is complex. This is because of socio-political and historical issues as well as the institutionalised imposition of the “mestizaje” narrative to minimise the voice of present-day Indigenous populations^{53,54}. Hence, an essential discussion that warrants aDNA research in Mexico is needed.

References

1. Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. DNA sequences from the quagga, an extinct member of the horse family. *Nature* 1984;312(5991):282–4.
2. Fowler C, Olalde I, Cummings V, et al. A high-resolution picture of kinship practices in an Early Neolithic tomb. *Nature* 2022;601(7894):584–7.
3. Haak W, Lazaridis I, Patterson N, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 2015;522(7555):207–11.
4. Skoglund P, Mathieson I. Ancient genomics: a new view into human prehistory and evolution. *Annu Rev Genomics Hum Genet* 2018;19:381–404.
5. Mathieson I. Human adaptation over the past 40,000 years. *Curr Opin Genet Dev* 2020;62:97–104.
6. Benton ML, Abraham A, LaBella AL, Abbot P, Rokas A, Capra JA. The influence of evolutionary history on human health and disease. *Nat Rev Genet* 2021;22(5):269–83.
7. Llamas B, Roca-Rada X, Collen E. Ancient DNA helps trace the peopling of the world. *Biochem (Lond)* 2020;42(1):18–22.
8. Roca-Rada X, Tereso S, Rohrlach AB, et al. A 1000-year-old case of Klinefelter’s syndrome diagnosed by integrating morphology, osteology, and genetics. *The Lancet*. 2022;400(10353):691–2.
9. Lanfranco F, Kamischke A, Zitzmann M, Nieschlag E. Klinefelter’s syndrome. *Institute of Reproductive Medicine of the University of Münster* 2004;364(11):273–83.
10. Groth KA, Skakkebaek A, Høst C, Gravholt CH, Bojesen A. Klinefelter syndrome - A clinical update. *Journal of Clinical Endocrinology and Metabolism* 2013;98(1):20–30.
11. Herlihy AS, Halliday JL, Cock ML, McLachlan RI. The prevalence and diagnosis rates of Klinefelter syndrome: An Australian comparison. *Medical Journal of Australia* 2011;194(1):24–8.
12. Fu Q, Hajdinjak M, Moldovan OT, et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 2015;524(7564):216–9.
13. Racimo F, Sikora M, Vander Linden M, Schroeder H, Lalueza-Fox C. Beyond broad strokes: sociocultural insights from the study of ancient genomes. *Nat Rev Genet* 2020;21(6):355–66.
14. Sunna Ebenesersdóttir S, Sandoval-Velasco M, Gunnarsdóttir ED, et al. Ancient genomes from Iceland reveal the making of a human population. *Science (1979)* 2018;360(6392):1028–32.
15. Margaryan A, Lawson DJ, Sikora M, et al. Population genomics of the Viking world. *Nature* 2020;585(7825):390–6.
16. Rivollat M, Jeong C, Schiffels S, et al. Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. *Sci Adv* 2020;6(22):eaaz5344.
17. Moilanen U, Kirkinen T, Saari NJ, et al. A Woman with a Sword?-Weapon Grave at Suontaka Vesitorninmäki, Finland. *Eur J Archaeol* 2021;(June):1–19.

18. Villalba-Mouco V, Oliart C, Rihuete-Herrada C, et al. Genomic transformation and social organization during the Copper Age–Bronze Age transition in southern Iberia. *Sci Adv* 2021;7:21.
19. Breuil V, Euller-Ziegler L. Gonadal dysgenesis and bone metabolism. *Revue du Rhumatisme (Edition Francaise)* 2001;68(1):32–9.
20. Gower G, Fenderson LE, Salis AT, et al. Widespread male sex bias in mammal fossil and museum collections. *Proc Natl Acad Sci U S A* 2019;116(38):19019–24.
21. Rasmussen M, Guo X, Wang Y, et al. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science (1979)* 2011;334(6052):94–8.
22. Chiu RWK, Chan KCA, Gao Y, et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc Natl Acad Sci U S A* 2008;105(51):20458–63.
23. Dabney J, Knapp M, Glocke I, et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences* 2013;110(39):15758–63.
24. Olalde I, Mallick S, Patterson N, et al. The genomic history of the Iberian Peninsula over the past 8000 years. *Science (1979)* 2019;1234(March):1230–4.
25. Simón M, Jordana X, Armentano N, et al. The Presence of Nuclear Families in Prehistoric Collective Burials Revisited: The Bronze Age Burial of Montanissell Cave (Spain) in the Light of aDNA. *Am J Phys Anthropol* 2011;146(3):406–13.
26. Olalde I, Haak W, Barnes I, Lalueza-Fox C, Reich D. The Beaker Phenomenon and the Genomic Transformation of Northwest Europe. *Nature* 2018;555(7695):190–6.
27. López-Cachero FJ. Cremation Cemeteries in the Northeastern Iberian Peninsula: Funeral Diversity and Social Transformation during the Late Bronze and Early Iron Ages. *Eur J Archaeol* 2011;14(1–2):116–32.
28. Llamas B, Fehren-Schmitz L, Valverde G, et al. Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Sci Adv* 2016;2(4):e1501385.
29. Posth C, Nakatsuka N, Lazaridis I, et al. Reconstructing the Deep Population History of Central and South America. *Cell* 2018;175(5):1185–97.
30. Moreno-Estrada A, Gignoux CR, Fernández-López JC, et al. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science (1979)* 2014;344(6189):1280–5.
31. Orlando L, Gilbert MTP, Willerslev E. Reconstructing ancient genomes and epigenomes. *Nat Rev Genet* 2015;16(7):395–408.
32. Damgaard PB, Margaryan A, Schroeder H, Orlando L, Willerslev E, Allentoft ME. Improving access to endogenous DNA in ancient bones and teeth. *Sci Rep* 2015;5(11184):1–12.
33. Hansen HB, Damgaard PB, Margaryan A, et al. Comparing ancient DNA preservation in petrous bone and tooth cementum. *PLoS One* 2017;12(1).
34. Pinhasi R, Fernandes D, Sirak K, et al. Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS One* 2015;10(6).

35. Parker C, Rohrlach AB, Friederich S, et al. A systematic investigation of human DNA preservation in medieval skeletons. *Sci Rep* 2020;10(1).
36. Kapp JD, Green RE, Shapiro B. A fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. *Journal of Heredity* 2021;112(3):241–9.
37. Roca-Rada X, Souilmi Y, Teixeira JC, Llamas B. Ancient DNA Studies in Pre-Columbian Mesoamerica. *Genes (Basel)* 2020;11(11):1346.
38. Migeon G. *Residencias y Estructuras Civico-Ceremoniales Posclásicas Tarascas de la Región de Zacapu (Michoacán, México)*. Oxford: BAR; 2015.
39. Lalueza-Fox C. Agreements and misunderstandings among three scientific fields: Paleogenomics, archaeology, and human paleontology. *Curr Anthropol* 2013;54(SUPPL8.).
40. Rubinoff D, Holland BS. Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. *Syst Biol* 2005;54(6):952–61.
41. Rasmussen M, Anzick SL, Waters MR, et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 2014;506(7487):225–9.
42. Kennett DJ, Lipson M, Prufer KM, et al. South-to-north migration preceded the advent of intensive farming in the Maya region. *Nat Commun* 2022;13(1).
43. Fernandes DM, Sirak KA, Ringbauer H, et al. A genetic history of the pre-contact Caribbean. *Nature* 2021;590(7844):103–10.
44. Kirchhoff P. *Mesoamérica: sus límites geográficos, composición étnica y caracteres culturales*. *Acta Americana* 1943;
45. Weaver MP. *The Aztecs, Maya, and Their Predecessors: Archaeology of Mesoamerica*. San Diego: Academic Press; 1993.
46. Roca-Rada X, Politis G, Messineo PG, et al. Ancient mitochondrial genomes from the Argentinian Pampas inform the early peopling of the Southern Cone of South America. *iScience* 2021;102553.
47. Prates L, Politis GG, Perez SI. Rapid radiation of humans in South America after the last glacial maximum: A radiocarbon-based study. *PLoS One* 2020;15(7):e0236023.
48. Perego UA, Achilli A, Angerhofer N, et al. Distinctive Paleo-Indian Migration Routes from Beringia Marked by Two Rare mtDNA Haplogroups. *Current Biology* 2009;19(1):1–8.
49. Moraga M, Pierre M de saint, Torres F, Ríos J. Kinship by maternal via between the last descendants of kawéskar ethnicity and burials in the patagonian channels: Evidence from the study of mitochondrial lineages. *Magallania (Punta Arenas)* 2010;38(2):103–14.
50. de Saint Pierre M, Bravi CM, Motti JMB, et al. An alternative model for the early peopling of Southern South America revealed by analyses of three mitochondrial DNA haplogroups. *PLoS One* 2012;7(9).
51. Wagner JK, Colwell C, Claw KG, et al. Fostering Responsible Research on Ancient DNA. *The American Journal of Human Genetics* 2020;107(2):183–95.

52. Alpaslan-Roodenberg S, Anthony D, Babiker H, et al. Ethics of DNA research on human remains: five globally applicable guidelines. *Nature* 2021.
53. Swanson P. *The Companion to Latin American Studies*. London: Routledge; 2003.
54. Bracho J. NARRATIVA E IDENTIDAD. EL MESTIZAJE Y SU REPRESENTACIÓN HISTORIOGRÁFICA. *Latinoamérica Revista de estudios Latinoamericanos* 2009;48:55–86.

Appendix

Bastien Llamas, Xavier Roca Rada, Evelyn Collen; Ancient DNA helps trace the peopling of the world. *Biochem (Lond)* 31 January 2020; 42 (1): 18–22.

Ancient DNA

Ancient DNA helps trace the peopling of the world

Bastien Llamas,
Xavier Roca Rada and
Evelyn Collen (University
of Adelaide, Australia)

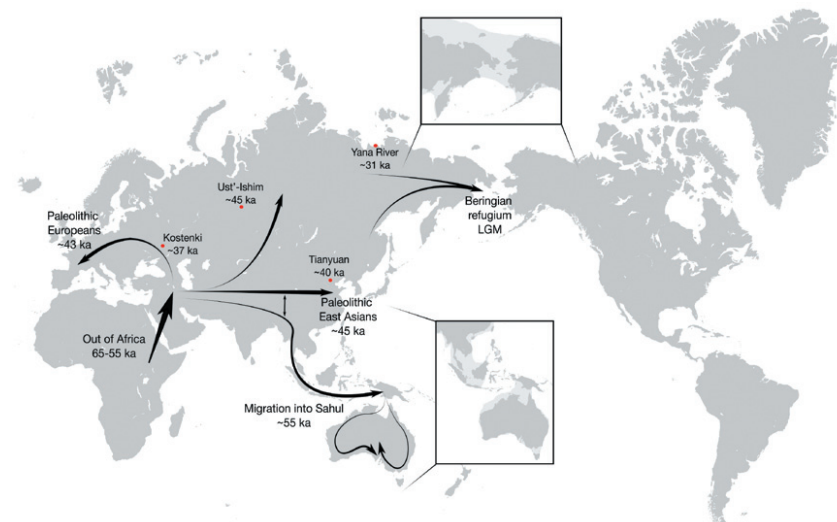
Many of us are fascinated by narratives regarding the origin and evolution of our species. Who are we? How did we people the world? Answers to these simple questions remain elusive even though researchers have been quite successful in describing past human morphology and culture using evidence from anthropology, archaeology, history, sociology and linguistics. However, when they address human migrations, archaeologists are somewhat restricted to surviving artifactual evidence and limited to descriptions of culture expansions, which may have occurred by the movement of either ideas or people. The advent of genomics, by which one can sequence whole or part of an individual's DNA, provided a powerful means to dig into past human demographic history. Notably, the coalescent theory posits that individuals in a population share genetic variants that originated from a common ancestor. This powerful theory is the basis for a number of bioanalytical innovations that utilize genetic data to reconstruct human movements around the world.

From classical archaeology to ancient DNA

During the last decade, the application of high-throughput DNA sequencing technologies to DNA extracted from archaeological remains has further revolutionized our understanding of human expansion and recent evolution. So-called ancient DNA (aDNA) provides accurate genetic snapshots of populations at a

given time and place, most of which are confounded in modern-day populations with subsequent demographic events such as admixture, population collapse or migration. Crucially, aDNA also enables accurately calibrating the molecular clock of human evolution. Not only can we identify splits between populations and infer migrations, we can also provide precise temporal context to these events.

Figure 1. Early movements of modern humans after they left Africa. Red dots are archaeological sites mentioned in the text, and arrows are only an approximation of the true migration process. The insets represent Beringia (**top**) and Sahul (**bottom**) with mainland extensions at the time of first human presence in these regions in light grey. LGM, Last Glacial Maximum.



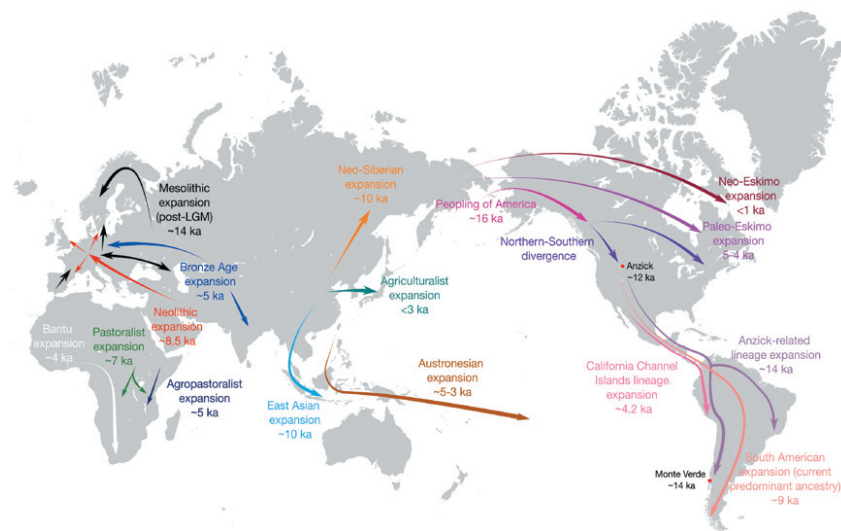


Figure 2. Modern human movements after the Last Glacial Maximum (LGM). Red dots are archaeological sites mentioned in the text, and arrows are only an approximation of the true migration process.

Equipped with the aDNA toolbox, researchers have now explored most regions of the world and tested hypotheses laid out by archaeologists, anthropologists and linguists. In all cases, aDNA findings were quite surprising (Figures 1 and 2).

Out of Africa and early non-African divergences

Africans have the highest levels of genetic diversity of any contemporary human population, and the oldest splits between human populations are found in sub-Saharan Africans. This evidence suggests that the origin of all anatomically modern humans (hereafter simply referred to as modern humans) is in Africa, although the exact location is still unknown.

Genetic studies of contemporary populations support an expansion of modern humans out of Africa around 65–55 thousand years ago (65–55 ka). Ancient genome sequences from extinct Neanderthals—archaic humans present in western and central Eurasia until 40 ka—revealed that all non-African populations have ~2% Neanderthal ancestry, supporting a single out-of-Africa event rapidly followed by admixture with Neanderthals. Furthermore, some populations in Asia and Oceania carry 3–6% Denisovan ancestry—Asian archaic humans identified through aDNA only. Admixture with archaic humans reveals the earliest split among non-African populations 60–45 ka: one lineage colonized mainland Eurasia, the other colonized New Guinea and Australia. The genomes of early Eurasian individuals such as

Ust’Ishim (45 ka), Kostenki (37 ka) and Tianyuan (40 ka) further constrain a split between western Eurasia and East Asia/Oceania to 55–45 ka.

Eurasia

Modern humans were widely distributed in Europe as early as 45 ka, but this early Palaeolithic European ancestry is virtually absent in contemporary populations. Three subsequent migration waves contributed to the modern-day European genetic makeup. First, Palaeolithic populations were replaced by Mesolithic hunter-gatherers coming from southern European and central Eurasian refugia ~14 ka, after the cold and arid conditions of the last Ice Age (also known as the Last Glacial Maximum (LGM); 28–18 ka). Second, the Neolithic farming populations of Anatolia, the Levant and northern Iran expanded throughout Europe ~8.5 ka and admixed with the Mesolithic hunter-gatherers. The Neolithic lifestyle (animal husbandry, agriculture and sedentarism) helped increase population size, but health became poorer. Third, Bronze Age herding populations migrated from the Eurasian steppe to central Europe ~5 ka. These Yamnaya and Afanasievo people already had a mixed ancestry related to various Russian and Caucasus hunter-gatherers. The Steppe migration revolutionized European technology with the notable introduction of the wheel and may have spread Indo-European languages.

aDNA from central and South Asia revealed complex patterns of migrations similar to Europe, where local Upper Palaeolithic hunter-gatherers admixed

Ancient DNA

with, or were replaced by, subsequent migrants from western Eurasia, including the Steppe expansion ~5 ka. In East and Southeast Asia, current Austroasiatic speakers exhibit a mixture of East Asian agriculturalist and eastern Eurasian hunter-gatherer ancestry. Present-day Japanese people result from an admixture between the indigenous Jomon population and later East Asian agriculturalists.

Island Southeast Asia and Oceania

Modern humans had probably crossed the Wallace line—the biogeographical zone that separates Asian and Oceanian ecozones—by 55 ka in an initial single wave to colonize Island Southeast Asia and Sahul (a prehistoric continent that consisted of New Guinea and Australia). These ancestors of Australo-Melanesian speakers (Papuaans, Aboriginal Australians, Torres Strait Islanders, Melanesians and Island Southeast Asia Negrito populations) underwent further genetic diversification and adaptation to new environments. Papuaans and Australians represent some of the oldest continuous indigenous cultures on earth due to isolation until European arrival. Island Southeast Asians west of the Wallace line further admixed with ancestors of Austroasiatic speakers from a mainland East Asian expansion ~10 ka. Then, an expansion associated with the Lapita culture spread the Austronesian languages and agriculture into Island Southeast Asia and Polynesian islands ~5–3 ka, further expanding as far as Hawaii and Easter Island. aDNA from Lapita individuals revealed they are highly related to present-day Taiwanese and have no Australo-Melanesian ancestry.

Siberia and Beringia

Ancestral North Siberians (represented by an individual from the Yana River, ~31 ka) diversified ~38 ka, soon after the basal split between western Eurasians and East Asians. The population did not contribute to present-day Siberians or Native Americans. During the LGM, Beringia was the land bridge exposed by lowered sea levels that connected northeastern Siberia and Alaska. Because northern Siberia was uninhabited due to adverse climatic conditions, humans who ventured into that region of the world were driven to a refugium in southeastern Beringia. However, the ice sheets that covered northern North America obstructed their dispersal further east, impeding access to the American continent. aDNA shows the isolated population had an Ancestral North Siberians ancestry and admixed with East Asians. This admixture resulted in two different lineages that diverged around 20 ka: the Ancient Palaeo-Siberians and the Ancestral Native Americans. The

genetic legacy of Ancient Palaeo-Siberians is currently restricted to populations in northeastern Siberia, whereas the direct ancestors of present-day Siberians were the Neo-Siberians, a population related to East Asians that expanded northwards ~10 ka.

America

The scenario of a southeastern Beringian refugium during the LGM is compatible with the Beringian Standstill Hypothesis, which proposes that Ancestral Native Americans were isolated for 2400–9000 years, giving rise to most of the non-Arctic Native American ancestry. Ancestral Native Americans entered North America ~25–15 ka, when ice sheets started to melt. This population split into two lineages either in Beringia or in North America ~17–15 ka: Northern Native Americans and Southern Native Americans (represented by the Anzick individual—~12 ka— associated with the Clovis culture). The peopling of America was swift and humans reached southern South America as early as ~14 ka (Monte Verde, Chile). Furthermore, a Siberian-related population described as Palaeo-Eskimos expanded into the Arctic ~5–4 ka. Palaeo-Eskimos were replaced by Neo-Eskimos, ancestors of the current Inuits, around 700 years ago.

Africa

Due to poor preservation conditions, aDNA studies have been challenging in Africa. Arguably, the main post-Out-of-Africa demographic event was the agricultural expansion of the Bantu-language population from the Highlands of Nigeria and Cameroon into sub-Saharan Africa around 4 ka. Bantu migrants generally admixed with local hunter-gatherers and in some places totally replaced the population. Additionally, two previous streams contributed to shaping the current African substructure: the migration of pastoralist populations from South Sudan to East and Central Africa ~7 ka, and the migration of agropastoralists from Ethiopia to Kenya and Tanzania ~5 ka.

aDNA and detection of adaptation

Genetic variants that confer better survival and reproduction tend to increase in frequency in a population and dominate over other variants. These adaptive variants leave distinctive genomic selection signatures in the form of unusually long genomic regions of low genetic diversity, where the variant and its surrounding genomic region increase in frequency in the population (Figure 3, left panel). Demographic processes such as admixture and bottlenecks may

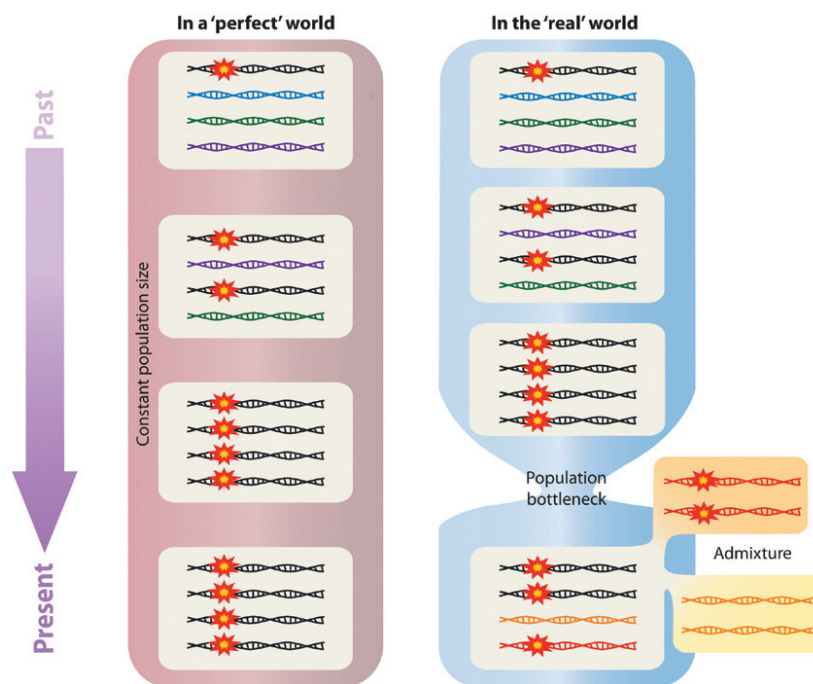


Figure 3. Left panel: detecting genetic signatures of selection in a 'perfect' world, i.e., in the absence of demographic processes that may break up the hallmarks of selection. A genetic variant (red and yellow star on a black genomic background) is selected for. The variant and its surrounding genomic sequence reach fixation in the contemporary population, so that all individuals have reduced diversity in this genomic region. Right panel: detecting signatures in the 'real' world is complicated by demographic processes. A population bottleneck reduces overall genome-wide diversity, making it more difficult to identify a region under selection (i.e., with low diversity). Admixture introduces new genomic backgrounds, also breaking up the signs of reduced diversity, even if the admixing population carries the same selected variant. aDNA allows observation of selection signatures occurring before demographic processes may disrupt the signal.

Downloaded from <http://portlandpress.com/bechemis/article-pdf/4/2/11/8967104/doi/10.1093/bechemis/article-pdf/4/2/11/8967104> by guest on 25 October 2022

potentially mask selection signatures (Figure 3, right panel), but aDNA is a powerful way to examine genomes for adaptive variants before known selection pressures, admixture and other demographic events occur.

As modern humans left Africa and peopled the world, they encountered new climatic conditions, food sources and pathogens. The Neolithic transition (notably the domestication of animals that carry zoonotic pathogens), the more recent Western Colonization and Industrial Revolution have also created drastic cultural and environmental changes. A major example of adaptation is the evolution of light skin pigmentation in western Europe. Originally, Palaeolithic Eurasians and Mesolithic Western Europeans carried variants linked to darker skin. Thanks to aDNA studies, we now know that some variants linked to lighter pigmentation were introduced by Neolithic farmers and Bronze Age herders. These only became widespread in Europe in the last 5000 years as a likely adaptation to counteract vitamin D deficiency due to lower levels of UV radiation at higher latitudes.

Another very notable example is the adaptation for lactase persistence in Europe, which enables milk digestion after weaning and into adulthood. This variant shows one of the strongest signatures of selection in

European populations. aDNA studies found this allele was rare (although present) in populations from Early Neolithic until the Bronze Age, and then very rapidly increased in frequency to ~70% in present-day northern Europe. The sudden rise in frequency comes at least 3000 years later than the archaeological record of dairy practices.

The evolution of height and the immune system are also the target of aDNA researchers, although these traits are challenging to study due to their highly polygenic nature.

Limitations and future of aDNA

aDNA research is limited by the availability of well-preserved human remains, and there are obvious ethical considerations regarding the destruction of samples for DNA studies. aDNA research also suffers from evolutionary genetics theoretical restrictions: divergence dates of genetic lineages can be inferred up to their most recent common ancestor, but we are blind to potentially extinct lineages that might have split earlier. For that reason, archaeological evidence of very early human presence in some regions of the world (Africa, Australia, Sumatra, China and the Arabian Peninsula), which seems

Ancient DNA

to predate diversification of currently known genetic lineages, might simply represent extinct populations that did not contribute ancestry to subsequent populations.

Ancient genetic data generated fascinating and often surprising discoveries about our past, but, as is the case for genetic results in general, great care must be taken when interpreting findings. While genetic studies involve grouping people into populations and observing the genetic differences between them, this cannot support racial or nationalistic ideologies. Nevertheless, aDNA is an invaluable tool to describe indigenous genetic diversity before the impact of European colonialism and more recent globalization. Beyond providing irrefutable genetic evidence of colonization-linked indigenous population collapse such as in South America, aDNA also provides useful evidence to repatriate indigenous remains locked in museums to where they belong. With the advent of personalized medicine, aDNA data may also help improve the health of indigenous people.

Exciting complementary disciplines that inform human evolution have recently emerged from the study of aDNA. Palaeomicrobiology, where genomes of ancient pathogens and microbes can be sequenced, provides invaluable insights about historical pandemics (such as the Black Death), but also the general health and lifestyle of past people. Similarly, palaeoepigenetics—which investigates DNA modifications that regulate gene expression—can contribute knowledge about morphology, age, lifestyle and eventually the environment of past individuals. There is definitely a future to aDNA research. ■



Bastien Llamas is an Australian Research Council Future Fellow and Associate Professor at the University of Adelaide, Australia. He completed a PhD in Molecular Biology in 2007 at the University of Montreal, Quebec, Canada.

He specializes in the analysis of aDNA extracted from past human populations and extinct animals to address outstanding questions in evolutionary biology.

Email: bastien.llamas@adelaide.edu.au



Xavier Roca Rada is a PhD student. He completed a bachelor's degree in biology at Universitat Autònoma de Barcelona and a master's in Biological Anthropology at Universitat de Barcelona. His current research focuses on the genetic and

evolutionary history of Pre-Columbian Mesoamerica.

Email: xavier.rocarada@adelaide.edu.au



Evelyn Collen completed a bachelor's degree in Molecular Biology at the University of Adelaide and is currently undertaking a PhD. She is especially interested in research at the interface of human anthropology, biology and

genetics to better inform human history. Her project focuses on the immunogenetic evolution of South American indigenous populations in the context of Western colonization.

Email: evelyn.phlox@adelaide.edu.au

Further reading

- Llamas, B., Willerslev, E. and Orlando, L. (2017) Human evolution: a tale from ancient genomes. *Philos. Trans. R. Soc. B: Biol. Sci.* **372**, 20150484
- Marciniak, S. and Perry, G.H. (2017) Harnessing ancient genomes to study the history of human adaptation. *Nat. Rev. Genet.* **18**, 659–674
- Nielsen, R., Akey, J.M., Jakobsson, M., Pritchard, J.K., Tishkoff, S. and Willerslev, E. (2017) Tracing the peopling of the world through genomics. *Nature* **541**, 302–310
- Pääbo, S. (2014) The human condition—a molecular approach. *Cell* **157**, 216–226
- Pääbo, S. (2015) The diverse origins of the human gene pool. *Nat. Rev. Genet.* **16**, 313–314
- Potter, B.A., Baichtal, J.F., Beaudoin, A.B. et al. (2018) Current evidence allows multiple models for the peopling of the Americas. *Sci. Adv.* **4**, eaat5473
- Skoglund, P. and Mathieson, I. (2018) Ancient genomics of modern humans: the first decade. *Annu. Rev. Genomics Hum. Genet.* **19**, 381–404

