Archaeogenetic perspectives on the hunter-gatherers and prehistoric farmers of the Mediterranean

Dissertation

To fulfill the
requirements for the degree of
„Doctor of Philosophy" (PhD)

Submitted to the Council of the Faculty
of Biological Sciences
of the Friedrich Schiller University Jena

by M.Sc. Ecology & Evolution Marieke Sophia van de Loosdrecht
born on 26.12.1988 in Almelo, the Netherlands

Gutachter:

1. Prof. Dr. Johannes Krause (Jena)

2. Prof. Dr. Christina Warinner (Jena)

3. Prof. Dr. Maanasa Raghavan (Chicago)

Beginn der Promotion: 12.10.2020
Dissertation eingereicht am: 25.09.2020
Tag der öffentlichen Verteidigung: 08.03.2021

Table of contents

Abbreviations

A	adenine
aDNA	ancient DNA
AMH	Anatomically Modern Human
AMS	Accelerator Mass Spectrometer
ANE	Ancient North Eurasians
BAL	*Balma Guilanyà* (in group labels)
BCE	Before the Common Era
bp	base pairs
BWA	Burrows-Wheeler Aligner
C	cytosine
°C	degree Celsius
calBP	calibrated years before present (before 1950)
CHA	*Cueva de Chaves* (in group labels)
CHG	Caucasus Hunter-Gatherers
CI	confidence interval
cM	centimorgan
CMS	*Moita do Sebastião* (in group labels)
CT	Computerized Tomography
DNA	deoxyribonucleic acid
ds	double-stranded
EHG	Eastern European Hunter-gatherers
ELT	*Cova de Els Trocs* (in group labels)
EM	Early Mesolithic (in group labels)
EN	Early Neolithic (in group labels)
FUC	*Fuente Celada* (in group labels)
G	guanine
HO	Human Origins
HGs	hunter-gatherers
IBD	identity-by-descent

LD	linkage disequilibrium
LGM	Last Glacial Maximum
LM	Late Mesolithic (in group labels)
LSA	Late Stone Age
MAPQ	mapping quality (Phred-scaled)
MDS	multidimensional scaling
MN	Middle Neolithic (in group labels)
MSA	Middle Stone Age
mtDNA	mitochondrial DNA
N	Neolithic (in group labels)
NE	Near Eastern (in *f*-statistics)
NGS	Next Generation Sequencing
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PMMR	pairwise mismatch rate
rCRS	revised Cambridge Reference Sequence
RSRS	Reconstructed Sapiens Reference Sequence
SE	standard error
SNP	Single Nucleotide Polymorphism
ss	single-stranded
T	thymine
TAF	*Grotte des Pigeons/Taforalt* (in group labels)
TMRCA	Time To the Most Recent Common Ancestor
UDG	uracil DNA glycosylase
UP	Upper Palaeolithic
UV	Ultraviolet
UZZ	*Grotta dell'Uzzo* (in group labels)
WHG	Western European Hunter-Gatherers
X	fold genomic coverage/read depth of SNP
yBP	years before present (before 1950)

1. Introduction

Prehistoric humans depended intimately on the rich biodiversity and relatively hospitable climate of the Mediterranean ecosystem. As such, the Mediterranean Basin, located at the crossroads of Africa, Europe and Asia, had a significant influence on prehistoric human migrations, demographic structure and behavioural development. Some of the oldest remains of anatomically modern humans (AMH) and various early aspects of behavioural modernity have been retrieved from northern Africa (Hublin et al., 2017; Richter et al., 2017; Scerri, 2017). During cold glacial epochs early humans and the game that they hunted may have contracted their habitats to refugial areas in Mediterranean coastal regions (Schmitt, 2007). But after the Last Glacial Maximum (LGM) 26,500-19,000 years before present (yBP), the gradually warming climate in the eastern Mediterranean allowed the resident hunter-gatherers to increasingly shift towards more versatile and sedentary subsistence strategies (Scarre, 2013). This would eventually give rise to the earliest fully sedentary farming communities in the Near East with plant and animal domesticates by at least ~10,500 yBP (Bellwood, 2004). Notably, agricultural practices similar to those in the Near East emerged earlier in the Mediterranean coastal regions than in many other European and African regions (Zilhão, 2001).

Scholars from various scientific disciplines have aimed to elucidate elements of the Mediterranean past. While linguists compare and reconstruct languages and archaeologists analyse lithic industries, symbolic art, pots or metallurgy, population geneticists can deduce past migrations, admixture events and other demographic processes by investigating patterns of genomic variation. Especially the analysis of ancient DNA (aDNA) is powerful in reconstructing demographic events in the human past (e.g. Slatkin & Racimo, 2016; Skoglund & Mathieson, 2018; Olalde & Posth, 2020). In archaeogenetic studies the principles of evolutionary and population genetics are applied to ancient remains retrieved from archaeological contexts. Under the condition that sufficient authentic aDNA can be recovered, this provides genomic reference points directly from the ancient individuals in the past. As a result, it is for example possible to distinguish whether changes in technology and subsistence strategy occurred because local people adopted new ideas (acculturation), were replaced by an incoming group with a different genomic ancestry (demic replacement), or a combination of these processes (genomic admixture).

This dissertation illustrates the power of aDNA in its current state-of-the-art to elucidate key events in human prehistory by reconstructing genome-wide data of hunter-gatherers and early farmers from the Mediterranean, an area that usually shows poor DNA preservation due to

warm and wet climatic conditions. The archaeogenetic investigations presented here advance our understanding of the population ancestry and demographic transitions in Europe and northern Africa after the LGM, and explore how these relate to transitions in material culture and subsistence strategy before and during the emergence of agricultural practises. In addition, the ancient genome-wide data presented here for peoples from both the European and African circum-Mediterranean coastal regions provide an unprecedented opportunity to investigate whether humans crossed the Mediterranean Sea, and admixed and/or exchanged material artefacts and ideas about subsistence strategies during prehistory.

This thesis is built upon decades of advancements in population genetics, archaeology in general, and bioarchaeology and archaeogenetics in particular. In the following part of this introduction I highlight the methodological and conceptual developments that are central to the ancient genomic data acquisition of this work. Subsequently, I provide a topical background on the human prehistory and population genomics relevant to the circum-Mediterranean region, emphasising previous archaeogenetic studies, to introduce the rationale of this dissertation.

1.1. The ancient DNA revolution: current state-of-the-art

Since the first successful attempts to recover aDNA fragments about 35 years ago (Higuchi et al., 1984; Pääbo, 1985), the methodologies and technologies for aDNA retrieval have been developing at a very fast pace. The developments have focused on mitigating the intrinsic properties of aDNA and contaminant DNA molecules. Three technical advances and conceptual shifts revolutionised aDNA research in recent years: next generation sequencing (NGS), DNA hybridisation enrichment and sampling of the inner ear bone. These developments have led to a multi-fold increase in the quantity of ancient genomic data that could be generated per individual, and opened up the possibility to successfully reconstruct genomic information from individuals with highly degraded aDNA (Slatkin & Racimo, 2016; Skoglund & Mathieson, 2018; Olalde & Posth, 2020).

Various post-mortem degradation processes contribute to the progressive decay of DNA, including hydrolysis (e.g. deamination of nucleotides), oxidation (e.g. depurination) and microbial digestion (Lindahl, 1993; Hofreiter et al., 2001; Bollingo et al., 2008; Kendall et al., 2018). Due to these degradation processes, abasic and atypic nucleotides, interstrand cross-links and strand breaks accumulate in DNA molecules over time. This eventually leads to aDNA fragments with many nucleotide modifications that are typically short (average length <60 base pairs), and

present only at very low endogenous concentrations within a vast background of environmental microbial and fungal DNA molecules (Pääbo, 1989; Green et al., 2008; Sawyer et al., 2012). The most commonly observed nucleotide modifications in aDNA fragments result from nucleotide deamination, in particular the deamination of cytosine (C) into uracil (U) in single-stranded overhanging ends (e.g. Lindahl, 1993; Hofreiter et al., 2001; Briggs et al., 2007). The rate of aDNA degradation is affected by thermal age (Sawyer et al., 2012) and various environmental factors, including temperature, pH, and exposure to oxygen and water (Lindahl, 1993; Smith et al., 2003; Bollingo et al., 2008; Kendall et al., 2018). As a result, ancient genomes for individuals that date to the Holocene (<11,700 yBP) and from cooler, higher latitudes dominate our current genomic record (Slatkin & Racimo, 2016).

In the first two decades of aDNA research, the go-to method of scholars who aimed to recover the low proportion of endogenous aDNA was to selectively target and amplify ancient molecules with known regions of interest, using a combination of polymerase chain reaction (PCR) and Sanger sequencing (Rizzi et al., 2012 and references therein). The results of these early studies, however, were frequently affected by contaminants. After some high-profile aDNA studies with extravagant claims were disputed for their credibility (e.g. Woodward et al., 1994), the scientific community agreed on the standardisation of more stringent criteria for proper authentication and reproducibility (e.g. Cooper & Poinar, 2000; Gilbert et al., 2005). These included measures to minimise contamination, such as performing pre-amplification steps in dedicated clean rooms, and processing blank controls.

The invention of next generation sequencing (NGS), a type of high-throughput sequencing, revolutionised the field of archaeogenetics about fifteen years ago (Green et al., 2006; Poinar et al., 2006). With NGS, massive amounts of sequencing data could be generated per ancient individual (Mardis, 2008). These methods have allowed to mitigate most of the limiting effects that the properties of aDNA can have on a reliable sequence reconstruction, such as low endogenous aDNA, high post-mortem damage and the unknown quantity of exogenous or contaminant DNA molecules. In contrast to PCR-based methods that only target one or a few genomic regions, NGS uses untargeted cluster amplification for massive parallel sequencing of DNA molecules in DNA libraries that can be re-amplified when necessary. During DNA library preparation, universal sequencing adapters with intrinsic primer annealing regions are attached to the terminal ends of aDNA fragments in an aDNA extract (Meyer & Kircher, 2010). This process is not species-specific; i.e. the composition of the DNA libraries approximates the total metagenomic pool of DNA molecules preserved in the ancient specimen, including exogenous sources such as from the environment. In addition, to mitigate the effects of post-mortem damage

on down-stream analyses, aDNA fragments can be treated with uracil-DNA-glycosylase (UDG) and endonuclease VIII ("USER" treatment) to partially or fully remove abasic and uracil sites prior to DNA library preparation (Briggs et al., 2010; Rohland et al., 2015).

The billions of DNA sequences generated by high-throughput sequencing also opened up new possibilities for detecting and mitigating contaminants, and validating aDNA authenticity. Each DNA library can be labelled with a unique oligo barcode (index pair) (Kircher et al., 2012). This barcode can be used to discern between the read-outs of different libraries after sequencing, and facilitates greatly in the detection and removal of various contamination sources, such cross-contamination, background contamination from the lab and carry-over from previous sequencing runs. Moreover, the sequence reads obtained with NGS retain the nucleotide misincorporation patterns characteristic for aDNA (Hofreiter et al., 2001; Briggs et al., 2007) For example, sites where cytosine deaminated into uracil appear as C->T misincorporations in the DNA reads in comparison to the human reference genome, whereas such changes are usually not found in modern contaminating DNA (Krause et al., 2010). Damage typically accumulates in the terminal read ends, providing a valuable authentication criterion (Lindahl, 1993; Briggs et al., 2007; Sawyer et al., 2012). Moreover, methods have been introduced that assess human contamination by quantifying polymorphisms in haploid genomic regions, such as on the mitogenome (Fu et al., 2014; Renaud et al., 2015) or the X-chromosome in males (Korneliussen et al., 2014).

Optimisation of DNA extraction techniques and the introduction of single-stranded library preparation protocols have further improved the recovery efficiency of aDNA (Meyer et al., 2012; Dabney et al., 2013; Gansauge & Meyer, 2013; Gansauge et al., 2017). Compared to double-stranded library preparation, single-stranded protocols have proved to be effective in converting fragments that are single-stranded or contain breaks, and showed a vastly improved recovery rate for DNA fragments shorter than 50 base pairs (Gansauge & Meyer, 2013). For example, Meyer and colleagues used these protocols to successfully reconstruct parts of the nuclear genome from a ~430,000 yBP Middle Pleistocene hominin from *Sima de los Huesos* in Spain (Meyer et al., 2016).

The effectiveness of reconstructing ancient human genomes was further boosted by capture methods that selectively target and enrich for genomic regions of interest, such as the mitogenome (e.g. Briggs et al., 2009; Maricic et al., 2010), specific exome (protein-coding) regions (e.g. Burbano et al., 2010), one chromosome (e.g. Fu et al., 2013), or the entire genome (e.g. Carpenter et al., 2013). Moreover, in 2015 a capture panel was developed to target ~1,2 million autosomal SNPs (the so-called "1240k panel") for ancestry analyses (Mathieson et al., 2015). In targeted capture approaches, magnetic-beaded probes designed to target genome regions of

interest are hybridised to complementary DNA fragments either in solid-phase on an array surface (e.g. Burbano et al., 2010) or in solution (e.g. Fu et al., 2013). This can substantially increase the relative fraction of endogenous DNA molecules in the total metagenomic pool in the DNA library. In-solution hybridisation enrichment has made it possible to recover sufficient genomic information for ancestry analysis from individuals with endogenous concentrations as low as 0.3%, and was shown to be up to ~50 times more effective than sole shotgun sequencing (Mathieson et al., 2015). The use of hybridisation enrichment is hence a very cost-efficient approach to generating genomic data from ancient humans for ancestry analyses.

Another impactful recent discovery in aDNA research has been the observation that the petrous part of the temporal bone (*pars petrosa* or petrous bone) in the cranium can provide a particularly rich source of endogenous human DNA (Gamba et al., 2014; Pinhasi et al., 2015; Hansen et al., 2017; Gaudio et al., 2019). The petrous bone contains the inner ear and is among the densest parts of the human skeleton. In addition to retaining relatively higher concentrations of total aDNA, the petrous bone also preserves relatively higher amounts of nuclear DNA (Hansen et al., 2017; Furtwängler et al., 2018). This has opened up possibilities to investigate the genomic material from individuals from environments in which DNA decays at an accelerated rate, such as deserts, the Mediterranean and (sub)tropical environments (Llorente et al., 2015; Pinhasi et al., 2015; Lazaridis et al., 2016; Skoglund et al., 2016). As a result, the petrous bone has become the most sought-after skeletal element for demographic studies based on ancient human genomes.

1.2. The ~10,000-year trajectory from foraging to farming in the Mediterranean region

At the height of the LGM ~20,500 yBP, ice sheets covered most of northern and central Europe, and appeared at lower elevations in southern Europe and northern Africa (Hughes et al. 2006). As a result, fauna and flora as well as early humans contracted their habitats to localised areas where environmental conditions remained comparably hospitable, such as the Mediterranean coastal regions of southern Iberia, southern Italy, the Balkans, Levant, Maghreb and Cyrenaica (Bar-Yosef, 1990; Gamble, 1993; Barton et al., 2005; Schmitt, 2007; Jones et al., 2016).

The hunter-gatherer groups that resided in these “glacial refugia” had a central role in the repopulation of Europe and northern Africa after the LGM, and frequently brought with them new technological innovations and subsistence strategies (Figure-1). The lithic technology at the heart

of the adaptive success of the glacial and post-glacial groups was the production of a large diversity of microliths (Clarkson et al., 2018). These small, standardised flakes made of flint or chert with cutting edges could be inserted into wooden or bone hafts to manufacture a large variety of composite tools for specialised use (Scarre, 2013). The exploitation of microliths intensified towards the end of the last Ice Age. Hunter-gatherers manufactured increasingly smaller types (miniaturisation) and made more frequent use of the microburin technique[1] to produce geometric microliths (Clarkson et al., 2018). Notably, the post-LGM microlith-producing hunter-gatherers of northern Africa and the Near East gradually changed to increasingly more sedentary lifestyles that included the management of selected plant species and storage of food (Bar-Yosef & Meadow, 1995; Barton & Bouzouggar, 2013). Eventually, these developments gave rise to the earliest fully sedentary farming communities with plant and animal domesticates in the Near East ~12,000-9,000 yBP (Bar-Yosef & Meadow, 1995; Ibáñez-Estévez et al., 2018) (Figure-1). From the Near East, agricultural practices and pottery (known as the "Neolithic package") would spread across Europe and Africa. The Neolithic transition in Europe, which started 8,800/8,400-7,500 yBP, marks the end of the Epipalaeolithic and Mesolithic in central and western Europe, the exact timing of which differs between regions (Ibáñez-Estévez et al., 2017). Notably, these agricultural practices spread much more rapidly in the central and western Mediterranean coastal areas than in many other areas in Europe, presumably due to maritime routes (Zilhão, 2001; Isern et al., 2017).

Archaeological scholars have postulated various direct connections between microlithic foragers and early farmers across the western and central Mediterranean, based on similarities observed in material technology or synchronous developments in subsistence strategies (e.g. reviewed in Straus, 2001; Zilhão, 2014; Biagi & Starnini, 2016). During glacial maximum aridity epochs, sea levels may have been ~120m lower than today (Lambeck et al., 2014). As a result, the distance between Africa and Europe was especially narrow at two sea corridors: the Gibraltar Strait in the western and the Siculo-Tunisian Strait in the central Mediterranean. Microlithic hunter-gatherers and Neolithic farmers may have crossed to the opposite seashore over the nearly closed land bridges that formed from surfacing island chains, or in prehistoric watercrafts made of logs, reed, papyrus or hides for sea-faring (reviewed in Ammerman, 2010; Howitt-Marshall & Runnels, 2016; Bailey et al., 2020). Most scholars agree that some Epipalaeolithic forager groups in the Mediterranean were already seafaring by ~13,500 yBP, but that the practice would become

[1] Procedure for cutting up lithic blades to manufacture (geometric) microliths that generates microburins as side products.

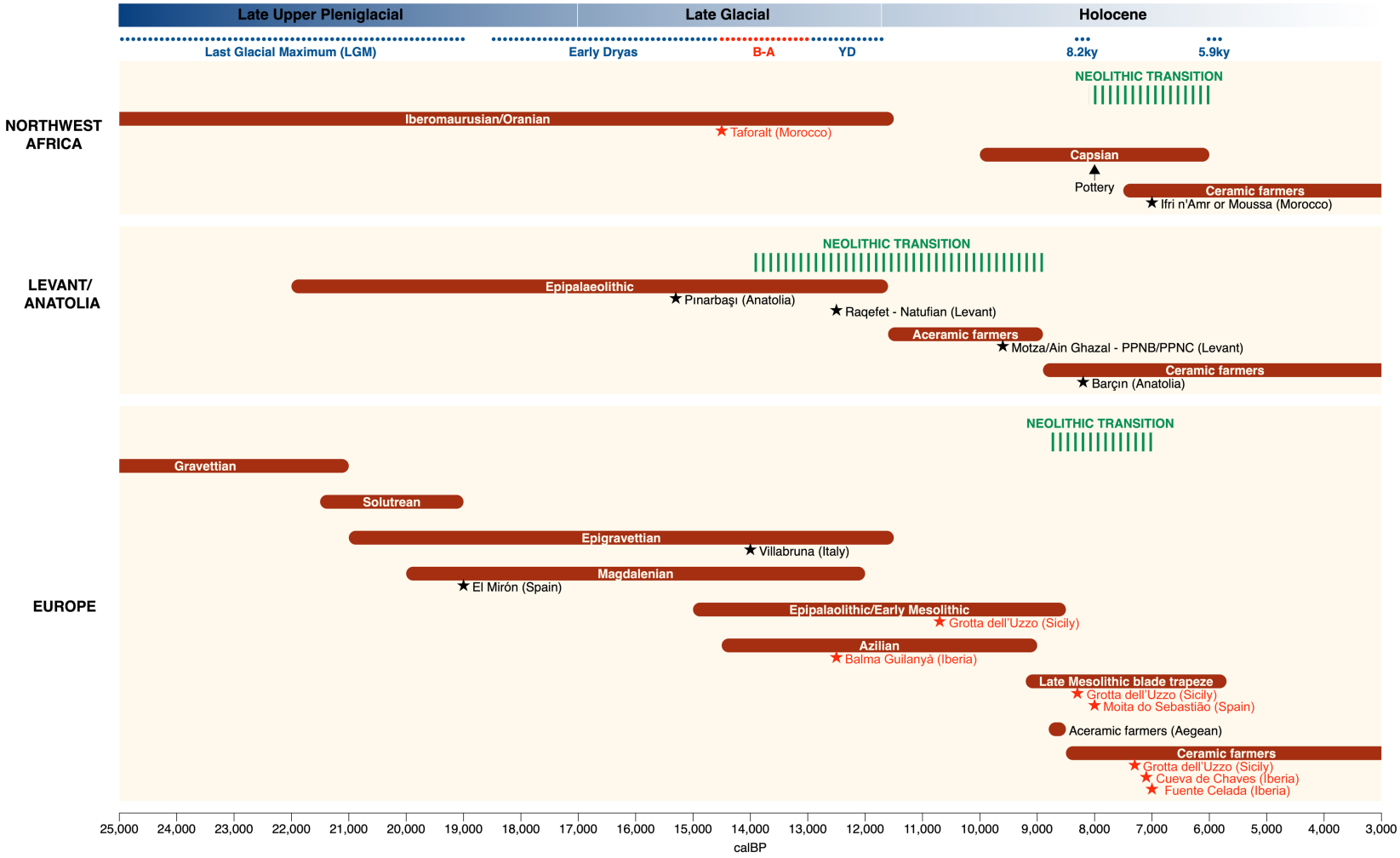


Figure-1 | Timeline showing a generalised chronology of relevant archeological industries and climate events in the Mediterranean. Stars indicate key ancient individuals or groups generated in this dissertation (red) or previously published (black). B-A: Bølling-Allerød interstadial, YD: Younger Dryas. Date ranges are based on information in Scarre, 2013; Hogue & Barton, 2016; Ibáñez-Estévez et al., 2017; Zilhão. 2014; Mulazzani et al., 2015; Petrullo & Delaplace, 2020

widely known and well-established among Neolithic farmers (Lo Presti et al., 2019; Ammerman, 2020; Galanidou et al., 2020).

In the following subsections I detail further relevant archaeological and genomic background on the population transitions during and after the last Ice Age in northwestern Africa (Section 1.2.1) and Europe (Section 1.2.2), and during the Neolithic transition in the Mediterranean (Section 1.2.3) to more specifically introduce the rationale and research questions of this dissertation.

1.2.1. Northwestern Africa during and after the last Ice Age

Late Stone Age[2] (LSA) microlithic industries can be found in the archaeological record of northern Africa from ~40,000 yBP onwards up until ~6,000 yBP (Garcea, 2010; Barton & Bouzouggar, 2013). The Iberomaurusian (or Oranian) microlithic bladelet complex, often made with the microburin technique, appeared ~26,000 yBP in the Maghreb in northwestern Africa and ~19,000-17,000 yBP in northeastern Africa (Cyrenaica, sometimes referred to as Eastern Oranian) (Douka et al., 2014; Hogue & Barton, 2016) (Figure-1). The Iberomaurusian foragers used grindstones to process edible plants and pigments, and buried their dead in cemeteries. Noteworthy anthropological characteristics include the practice of tooth evulsion and trepanation (Dastugue, 1962; Humphrey et al., 2012; De Grootte & Humphrey, 2016). Iberomaurusian sites seem to have been particularly dense and intensively occupied in Mediterranean coastal regions during episodes of regional hyper-aridity. This has led scholars to propose that Iberomaurusian hunter-gatherers actively sought out these regions as refugia areas during the LGM (Barton et al., 2005, but see: Poti & Weniger, 2019).

As the climate became generally wetter and warmer during the early Holocene, Capsian assemblages emerged in the Maghreb ~10,000-6,000 yBP (Lubell et al., 1984; Petrullo & Delaplace, 2020). The Capsian was a microlithic tool complex similar to the Late Iberomaurusian industries, albeit with a greater diversity of lithics (Rahmani, 2004; Petrullo & Delaplace, 2020). Notably, from ~8,000 yBP onwards the Capsian hunter-gatherers can be found alongside an

[2] Archaeologists have distinguished various regional and supra-regional microlithic varieties in northern Africa that they have referred to as either Late Stone Age, Upper Palaeolithic, Epipalaeolithic or Mesolithic technologies (e.g. Garcea, 2010). This difference in naming may partly stem from observed similarities with material cultures in (North) Africa, the Near East or southern Europe.

increasing number of elements associated with a food-producing economy, including pottery and pastoral activities (Lindstädter, 2008; Mulazzani et al., 2015) (Figure-1).
Agricultural food-producing economies similar to the European Mediterranean Early Neolithic, with polished stone tools, Cardial pottery, domesticated crops and animals, appeared in coastal northwestern African regions ~7,500-7,350 yBP (see for more detail Section 1.2.3) (e.g. Zilhão, 2014 and references therein) (Figure-1).

The increasing diversity of (geometric) microlithic tools, use of grindstones to process wild plants, baskets for storage and burial of dead in cemeteries indicates that the hunter-gatherers gradually transitioned into more localised societies with increased sedentism and broader subsistence strategies (Barton & Bouzouggar, 2013). A long outstanding question in North African archaeology is whether these transitions were local inventions (demic continuity) or introduced by incoming people from outside the Maghreb (demic diffusion or replacement) (reviewed in Lindstädter, 2008; Garcea, 2010; Barton & Bouzouggar, 2013). Whereas archaeological analyses have allowed to distinguish between the various observed technological industries in the (North) African record and date their respective transitions, genetic research has shed important insights on population connections and the demographic processes that underlie the changes in subsistence strategies and technology.

Present-day North Africans are genetically more similar to Near Easterners and Europeans than to sub-Saharan Africans (Henn et al., 2012). It remains unknown, however, when and in what archaeological contexts, this genomic structure within the African continent became established. Studies based on modern autosomal data suggested that extensive bidirectional migrations introduced genomic influences from neighbouring Mediterranean Europe and the Near East into northern Africa (Henn et al., 2012; Arauna et al., 2017; Serra-Vidal et al., 2019). These studies have discerned four major ancestry layers in present-day North Africans: i) an autochthonous Maghrebi component that derives from Stone Age hunter-gatherers (initially estimated ~12,000 yBP), ii) a Middle Eastern component from expanding Early Neolithic farmers that was reintroduced during the Arab conquest ~1,400 yBP, iii) a southern European component, likely the result of Iron Age Punic/Carthage explorations ~2,700-2,100 yBP and more recent historical events, and iv) a sub-Saharan African component, possibly reflecting trans-Saharan slave trade ~1,200-700 yBP. However, modern human genomes may not be ideal to investigate population genomic events deeper in prehistory, due to the possibility that more recent admixture events replaced or diluted the genomic signals of older ones. Therefore, to accurately dissect the different ancestry layers and their respective temporal contexts in North African population genomic history, aDNA is needed from individuals who lived in the past directly.

Unfortunately DNA rapidly decays in warm conditions, as in most African environments. As such, prior to the start of this dissertation research, ancient genome-wide data had only been successfully generated for 26 ancient African individuals, the oldest one being an 8,100 yBP nomadic pastoralist from Tanzania (Gallego Llorente et al., 2015; Schlebusch et al., 2017; Schuenemann et al., 2017; Skoglund et al., 2017). The oldest known genome-wide reference points for northern Africa came from three (Pre-) Ptolemaic Egyptian mummies dated to ~2,600-2,000 yBP, whose ancestry profiles indicated that a strong genetic connection between Near Easterners and northern Africans had already been established by that time (Schuenemann et al., 2017). It has remained unknown, however, if this genomic connection dates to the expansion of farming practises from the Near East or to earlier expansions, such as those of nomadic pastoralists or hunter-gatherer groups associated with microlithic industries.

Manuscript A of this dissertation provides the first insights into the pre-farming genomic history of northern Africa in reconstructing and analysing seven ancient genomes retrieved from ~15,000 yBP Iberomaurusian hunter-gatherers, who were retrieved from the oldest known cemetery in Africa near *Taforalt* in northeastern Morocco (Humphrey et al., 2012) (Figure-1).

1.2.2. Europe during and after the last Ice Age

In the last decade, the analysis of ancient genomes from Palaeolithic and Mesolithic European hunter-gatherers greatly contributed to our understanding of the repopulation of Europe (e.g. Skoglund et al., 2012/2014; Lazaridis et al., 2014/2016; Olalde et al., 2014/2019; Raghavan et al., 2014; Seguin-Orlando et al., 2014; Haak et al., 2015; Jones et al., 2015; Mathieson et al., 2015/2018; Broushaki et al., 2016; Fu et al., 2016; Hofmanová et al., 2016; Günther et al., 2018; Mittnik et al., 2018; Feldman et al., 2019). One of the most significant findings was that many of the major lithic industries in the archaeological record could be linked to distinct genomic profiles (e.g. Fu et al., 2016). This has provided valuable insights in the population genomic ancestry and demographic processes underlying archaeological and subsistence strategy transitions during European prehistory.

During the LGM, early humans may have deserted northwestern Europe completely, since there is no evidence for occupation (Steward & Stringer, 2012). However, towards the end of the LGM ~19,000 yBP Europe was repopulated by expanding groups associated with the Magdalenian technocomplex (Fernández-López de Pablo et al., 2019) (Figure-1). Both their material artefacts and distinct genomic ancestry can be found until ~12,000 yBP across

southwestern and central Europe (Bonilla et al., 2012; Maier, 2015; Fu et al., 2016). This group of Magdalenian hunter-gatherers was assigned as the El Mirón genetic cluster, named after the oldest individual. Notably, the Magdalenian ancestry was found to have been partly derived from the ancestry in the pre-LGM Aurignacian-associated hunter-gatherers (Fu et al., 2016). Moreover, Epigravettian assemblages appeared in Italy and southeastern Europe, often overlying the pre-LGM Gravettian industries (Djindjian, 2016).

The abrupt warm and moist climate period of the Bølling-Allerød interstadial ~14,500-12,900 yBP is thought to have triggered additional human expansions and a gradual repopulation of European areas further and further north (Fernández-López de Pablo et al., 2019). In western Europe north of the Alps, the Azilian developed from the Magdalenian, and technologies related to these, as well as to the Italian Epigravettian, emerged in Iberia below the Pyrenees (Bonilla et al., 2012; Naudinot et al., 2017) (Figure-1). During this time period, ancestry as found in a 14,000 yBP Epigravettian hunter-gatherer from *Villabruna* in northern Italy appeared throughout continental Europe (Villabruna cluster, Fu et al., 2016).

Throughout the Mesolithic, which coincided with the onset of the Holocene ~11,700 yBP (Walker et al., 2009), the Villabruna cluster ancestry spread further across Europe. This ancestry became the dominant component in the genomes of foragers from western Europe, coined Western European hunter-gatherers (WHGs) (Lazaridis et al., 2014; Fu et al., 2016). However, hunter-gatherers from eastern Europe were found to retain some Ancient North Eurasian (ANE) ancestry related to Upper Palaeolithic Siberians, and thus coined Eastern European hunter-gatherers (EHGs) (Haak et al., 2015; Mittnik et al., 2018). Genetic interactions of European hunter-gatherers over millennia spanning the Upper Palaeolithic to the Late Mesolithic resulted in an ancestry cline, in which WHG ancestry was highest in hunter-gatherers from western Europe, and those further towards the east having increasingly higher proportions of EHG and ANE ancestry (Lazaridis et al., 2014; Fu et al., 2016; Lipson et al., 2017; Mathieson et al., 2018).

All in all, these studies indicated that post-glacial hunter-gatherers in Europe had derived their ancestry from at least two genomically distinct hunter-gatherer groups; one group contributed WHG and the other EHG/ANE ancestry. However, the knowledge on the genomic diversity of post-glacial hunter-gatherers in the western and central Mediterranean has remained rather limited. Prior to the start of this dissertation research, ancient genome-wide data was available for only three Mesolithic Iberian foragers (Olalde et al., 2014; González-Fortes et al., 2017). Notably, the Magdalenian-associated ancestry as found in the ~19,000 yBP *El Mirón* hunter-gatherer from Iberia represented only a minor part of the ancestry in the Iberian Mesolithic *La Braña* forager (Fu et al., 2016). In the absence of more genomic reference points for Upper

Palaeolithic and Mesolithic Iberian foragers it has remained unknown if Magdalenian-ancestry was totally replaced during post-glacial expansions or in some areas locally preserved. Moreover, although a few ancient genomes are known for (Epi)Gravettian foragers from Italy, no ancient genomes were so far successfully recovered from Mesolithic foragers.

The limited understanding of the genomic diversity of the foragers in Iberia and southern Italy is unfortunate because the foragers in these regions are prime candidates to have retained ancestry from different glacial hunter-gatherer groups. A systematic investigation into the genomic diversity and structure of the foragers in Iberia and southern Italy can hence provide a crucial geographical and temporal context to the observed genetic clines among glacial and postglacial European hunter-gatherers. As such, I reconstructed and analysed a genomic transect of Upper Palaeolithic and Mesolithic HGs from Iberia in Manuscript B, and from Sicily in Manuscript C.

In addition, archaeologists proposed direct material connections between the European (Epi)Gravettian or Magdalenian and contemporaneous North African Iberomaurusian assemblages during the last Ice Age, as well as between European Late Mesolithic blade-and-trapeze industries and contemporaneous North African Capsian assemblages (e.g. Straus, 2001; Cortés Sánchez et al., 2012; Perrin & Binder, 2014). The generation of ancient genome-wide data for individuals in both the African (Manuscript A) and European Mediterranean coastal regions (Manuscripts B & C) allows for a first systematic investigation into the evidence for direct gene flow across the Mediterranean Sea in periods before the spread of farming (Figure-1).

1.2.3. The Neolithic Revolution: sedentary farming communities expand into the Mediterranean

One of the most impactful changes in human history was the transition in subsistence practices from comparably mobile hunting and gathering to sedentary farming. The primary zone of Neolithisation in western Eurasia spanned the Fertile Crescent, eastern and central Anatolia and to a lesser extent western Anatolia and the Aegean (Özdoğan, 1997). In the period between ~14,500-9,000 yBP, the hunter-gatherers from this region gradually invented and adopted the elements for agricultural practises associated with the Near Eastern “Neolithic Package”, including food storage, domestic plant and animal cultivation, sedentary housing, pottery, polished stone tools and other portable artefacts (Zvelebil, 1989; Özdoğan, 2011). By ~8,500 yBP, European early farmers associated with and related to the Starčevo-Körös-Criș cultural complex had established communities across the Aegean and Balkan Peninsula (Özdoğan, 2014).

Archaeologists proposed that the agricultural practises expanded along two major archaeologically-defined routes into Europe (reviewed in: Whittle, 1996; Price, 2000). Early farmers followed a continental route out of the Balkan Peninsula and spread along the Danube river into central Europe, associated with the Neolithic Linear Pottery culture (Linearbandkeramik, LBK), and from there further north and west (Figure-2, Danubian/Continental Route). In addition, early farmers expanded directly into the Mediterranean Basin associated with Impressa Ware ceramic culture in the eastern and central Mediterranean and with Cardial Ware in the western Mediterranean (Mediterranean Route (Guilaine & Manen, 2007). Notably, in early farming settlements in southern Italy, Sicily and Mediterranean coastal Iberia Impressa Ware pottery has been found that overlaps with or precedes the appearance of Cardial Ware up to a few hundreds of years (reviewed in e.g. Ibáñez -Estévez et al., 2017; Guilaine, 2018). Approximately contemporaneous with the appearance of the Impressa Ware in southern Iberia, early farming

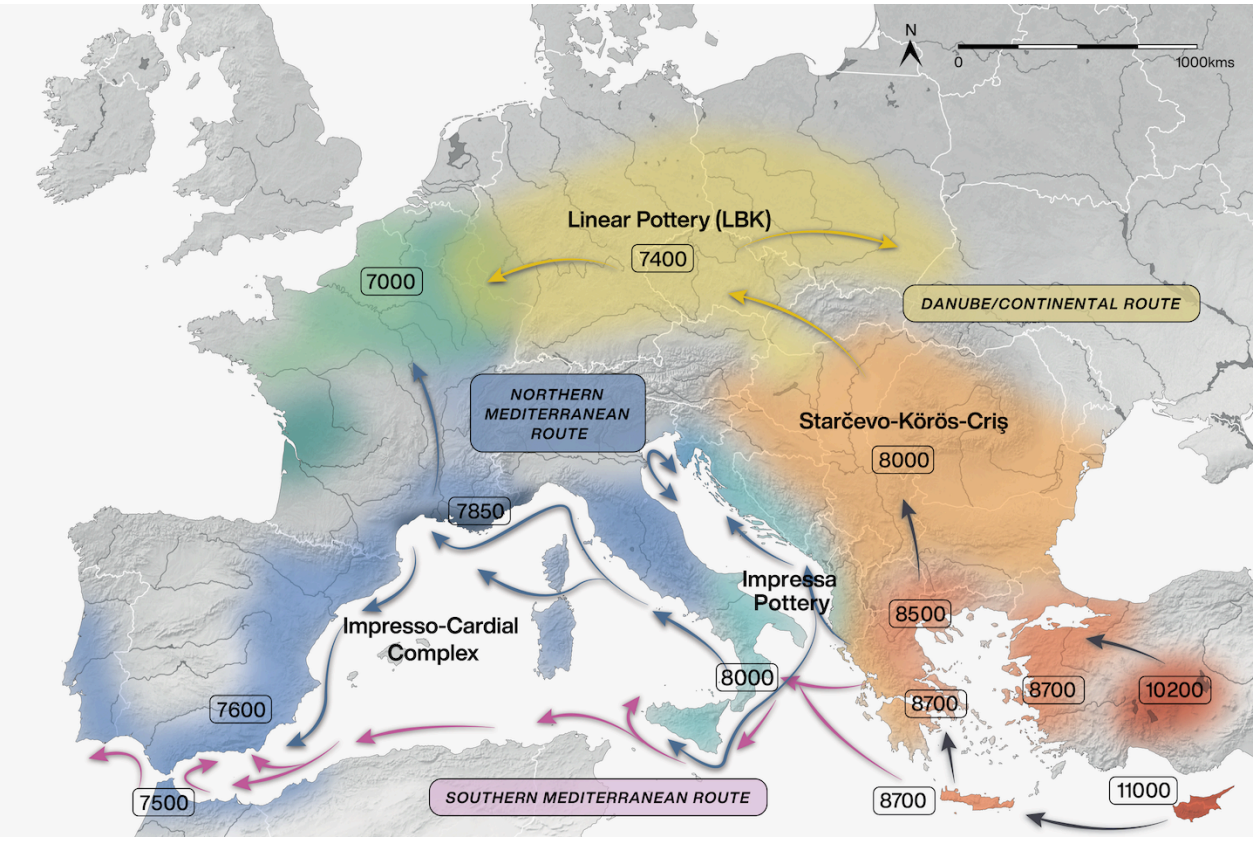


Figure-2 | Expansion of farming practises in Europe and northwestern Africa. Dates are in calBP. The map is based on Figure-S11 in Rivollat et al., 2020 (with permission). The arrows for the Danube/Continental and Northern Mediterranean Route are based on a map of Gronenborn et al., 2019 (RGZM/OREA), and those for the Southern Mediterranean Route on Manen et al., 2007. The date for the arrival of the ceramic Early Neolithic in Morocco is from Cortés Sánchez et al., 2012 and Martínez-Sánchez et al., 2018.

communities with similar material characteristics, such as *Almagra* type pottery, appeared in the coastal regions of the Maghreb ~7,500 yBP (reviewed in Cortés Sánchez et al., 2012; Martínez-Sánchez et al., 2018). Some archaeologists have proposed that these early farming practises associated with Impressa Ware have their origin in a separate pioneering event in southern Italy and Sicily that was followed by a subsequent expansion along the Northwest African Mediterranean coastline into the Maghreb and then via the Gibraltar Strait into Iberia (Southern Mediterranean Route, Figure-2) (Manen et al., 2007). Others proposed an expansion route of early farmers out of the Balkans into peninsular Italy to Sicily, further westwards to Mediterranean France and Iberia, and perhaps via the Gibraltar Strait into the Maghreb (Northern Mediterranean Route, Figure-2) (reviewed in Zilhão, 2014).

Ancient DNA studies have contributed substantially to the understanding of the Neolithisation of Europe. The studies focused on the routes along which early farmers expanded, and the extent of population replacement, mixing or acculturation. A plethora of aDNA studies confirmed that large-scale expansions of populations related to the Neolithic farmers from *Barçın* in northwestern Anatolia facilitated the agricultural transition in Europe (e.g. Skoglund et al., 2012/2014; Lazaridis et al., 2014/2016; Günther et al., 2015; Haak et al., 2015; Mathieson et al., 2015; Olalde et al., 2015; Hofmanová et al., 2016; Kılınç et al., 2016; Omrak et al., 2016). Notably, following the initial expansion hunter-gatherer ancestry was found to have increased substantially in the genomes of farmers from the Early to the Middle Neolithic, in particular in those from central Europe and Iberia (e.g. Günther et al., 2015; Haak et al., 2015; Mathieson et al., 2015/2018; Olalde et al., 2015; Hofmanová et al., 2016; Martiniano et al., 2017). Olalde et al. showed that this resurgence of hunter-gatherer ancestry was predominantly of a local origin, and could be used to track the expansion dynamics of farmers (Olalde et al., 2015). More specifically, the amount of forager ancestry in early farmers was found to follow roughly an east to west cline (Lipson et al., 2017; Rivollat et al., 2020).

Moreover, aDNA studies have also indicated that the underlying demographic processes for the Neolithisation likely differed among regions within Europe (e.g. Skoglund et al., 2014; González-Fortes et al., 2017; Jones et al., 2017; Lipson et al., 2017; Mathieson et al., 2018; Rivollat et al., 2020). It is important to note, however, that most of these studies focussed on the population dynamics of the farmers who spread along the continental (Danubian) route. The picture that has emerged from these studies is that in central, western and northern Europe the early farmer settlements along the Danubian expansion route showed all characteristics of the Near Eastern Neolithic package, including pots, animal and plant domesticates. The early farmers in these regions carried only a very minor genomic contribution from local European hunter-

gatherers (Skoglund et al., 2012; Haak et al., 2015; Mathieson et al., 2015; Hofmanová et al., 2016). In contrast, in some regions in the Baltic, Balkans and southern France acculturation seemingly exerted a stronger influence during the Neolithic transition (Jones et al., 2017; Mathieson et al., 2018; Rivollat et al., 2020). Similar to the Balkans and southern France, there are archaeological indications that acculturation had a significant role in the transition to farming in the Mediterranean coastal regions of southern Italy, Sicily, Iberia and the Maghreb (Cortés Sánchez et al., 2012; Mulazzani et al., 2016; Gronenborg et al., 2017; Isern et al., 2017; Dunne et al., 2019).

Prior to the start of this dissertation research ancient genome-wide data for Late Mesolithic hunter-gatherers and early farmers in southern Italy, Sicily and northwestern Africa was absent and scarce in Iberia. As a result, details on demographic processes underlying the Neolithisation processes in the western and central Mediterranean region have remained unclear. Also, it has remained an open question whether there is genomic support for the long-held hypothesis in archaeology for a distinct Mediterranean expansion route, and whether this route involved a direct connection between southern Europe and northwestern Africa via the Strait of Sicily or Gibraltar. Manuscripts B and C contribute ancient genomic reference points from individuals that cover the Mesolithic-Neolithic transition in Iberia and Sicily, respectively, to investigate these topics.

2. Aims and objectives

This dissertation aims to recover ancient genome-wide data of prehistoric peoples from the Mediterranean to examine i) transitions in hunter-gatherer genomic structure after the Last Glacial Maximum in light of proposed archaeological expansions, and ii) the demographic processes that underlie the transition to agricultural practises. Special emphasis is placed on examining the genomic evidence for direct population interactions across the Mediterranean Sea.

Using a combination of state-of-the-art aDNA techniques optimised for the retrieval of highly degraded DNA, each of the manuscripts adds genomic reference points from Mediterranean areas where human occupation locally persisted during the Last Glacial Maximum, or with some of the earliest indications for agricultural practises and pottery (Figure-3). Manuscript A presents genome-wide data for ~15,000 yBP Iberomaurusian foragers from the Mediterranean coast in Morocco. Manuscripts B and C report a genomic transect of various post-glacial hunter-gatherer and early farmer groups from Iberia and Sicily, respectively.

The research questions addressed in this dissertation are:

1) What is the genomic profile of the hunter-gatherers associated with the Iberomaurusian complex?
2) Does the genomic evidence support an origin for the Iberomaurusian microlithic industry in Northwest Africa or outside of the region?
3) What are the expansion dynamics and admixture events that resulted in the genetic clines observed among the glacial and post-glacial hunter-gatherers of Europe?
4) Is there genomic evidence that northern African or southern European microlithic hunter-gatherers crossed the Mediterranean Sea?
5) Was the transition to agriculture in the central and western Mediterranean driven by acculturation of local foragers or population replacement by incoming farmers?
6) Along which route did farming practices spread into the central and western Mediterranean?
7) Is there genomic evidence for a separate southern Mediterranean Neolithic expansion route that involved sea crossings at the Strait of Sicily or Gibraltar?

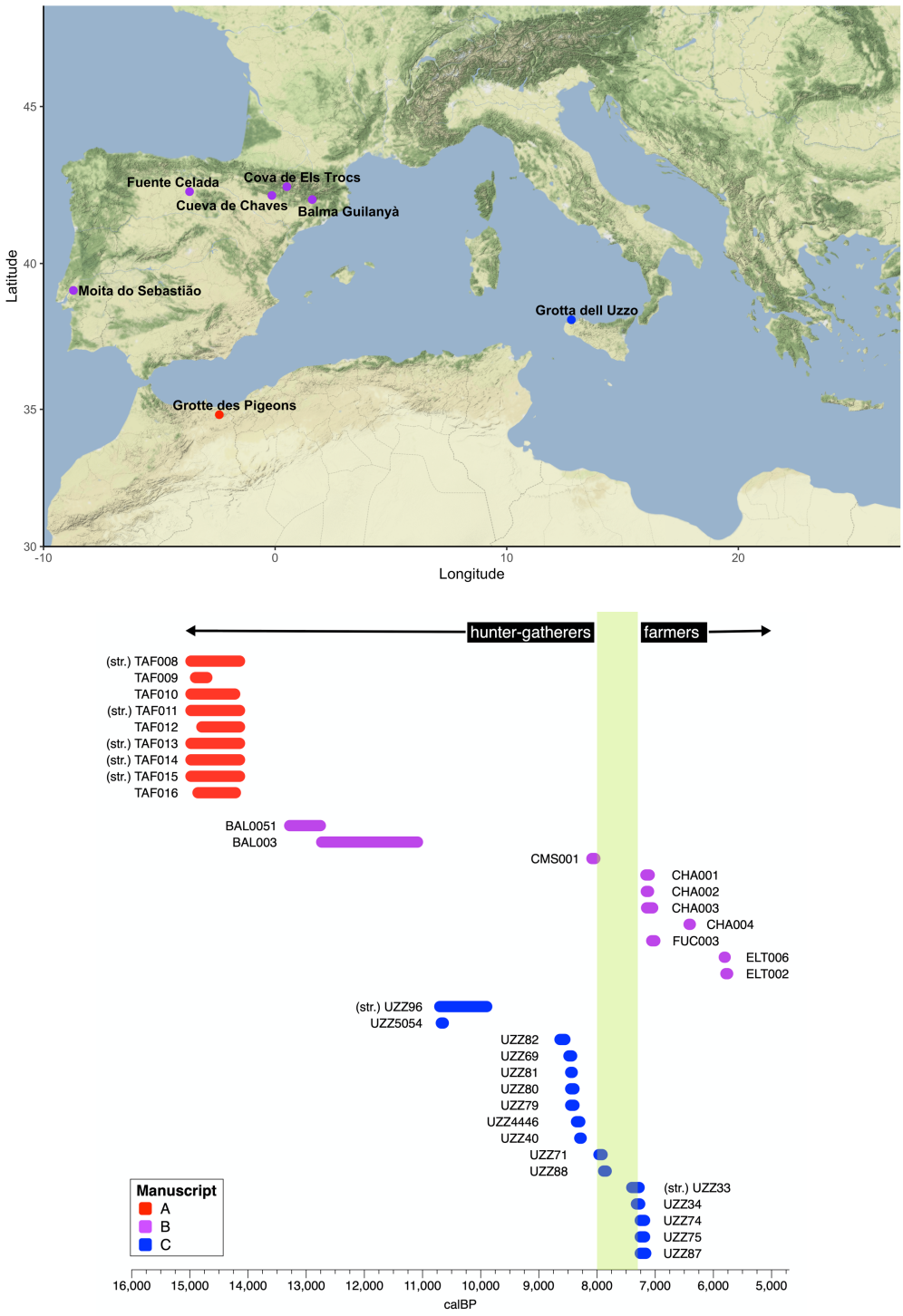


Figure-3 | Geographical, temporal and archaeological context of the generated ancient genome-wide data in the manuscripts of this dissertation. The onset of the Neolithic transition is marked in green. TAF: *Grotte des Pigeons* (Late Stone Age - Iberomaurusian), BAL: *Balma Guilanyà* (Late Upper Paleolithic - Azilian), CMS*: Moita do Sebastião* (Late Mesolithic - Geometric), FUC: *Fuente Celada* (Early Neolithic - Cardial), CHA: *Cueva de Chaves* (Early Neolithic - Epicardial), ELT: *Cova de Els Trocs* (Middle Neolithic), UZZ: *Grotta dell'Uzzo* (Mesolithic, Late Mesolithic - Castelnovian, and Early Neolithic - Stentinello).

3. Overview of manuscripts

This dissertation includes the following three publications:

3.1. Manuscript A

"Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations"

Marieke Sophia van de Loosdrecht, Abdeljalil Bouzouggar, Louise Humphrey, Cosimo Posth, Nick Barton, Ayinuer Aximu-Petri, Birgit Nickel, Sarah Nagel, El Hassan Talbi, Mohammed Abdeljalil El Hajraoui, Saaïd Amzazi, Jean-Jacques Hublin, Svante Pääbo, Stephan Schiffels, Matthias Meyer, Wolfgang Haak, Choongwon Jeong, Johannes Krause

Published in *Science* (May 2018)

Synopsis: In Manuscript A, I aimed to obtain the first insights into the population genetic history of northern Africa prior to the introduction of agriculture.

I reported genome-wide data for nine ~15,000 calBP Iberomaurusian hunter-gatherers retrieved from *Grotte des Pigeons*, a cave site near *Taforalt* in northeastern Morocco (Figure-3). These are the hitherto oldest genomic reference points for Africa. *Grotte des Pigeons* was intensively occupied during the Middle and Late Stone Age, making it a crucial site to investigate the population genetic history of northern Africa. I extracted DNA from petrous bones, generated single-stranded DNA libraries and used in-solution capture to selectively enrich the whole mitogenome and >1.2 million ancestry informative SNPs in the nuclear genome. The DNA fragments of seven individuals, six genetic males and one female, showed post-mortem degradation characteristics typical of authentic aDNA. Interestingly, the Iberomaurusian foragers carried mitogenome haplogroups U6a (six individuals) and M1b (one individual), and all males carried lineages of the Y-haplogroup E1b1b1a1 (M-78). U6 and M1 haplogroups typically occur at a high frequency among present-day North Africans, and have been associated with a back-migration to Africa from Eurasia. These characteristic North African haplogroups were hence already present in northwestern Africa by at least ~15,000 yBP.

First, I investigated how Iberomaurusian hunter-gatherers relate to ancient and present-day sub-Saharan Africans and Eurasians using a combination of several *f*-statistic-based methods. I found two deeply diverged ancestry components in the Iberomaurusian hunter-gatherers. The major component in their genomic profiles is best approximated by the Eurasian ancestry as preserved among ~15,000-9,000 calBP Natufian hunter-gatherers from the Levant. In addition, I found that about one-third of their ancestry is uniquely shared with sub-Saharan Africans, in particular present-day West and East Africans. These results suggest that the Iberomaurusian foragers were deeply related to both sub-Saharan Africans and Eurasians.

Subsequently, I aimed to characterise the sub-Saharan African component further by modeling it as one, two or three-way mixtures of several known ancient and present-day African groups. However, none of these admixture models could fit the ancestry profile of the Iberomaurusian foragers fully. The best fitting model approximated the sub-Saharan component as West African (Yoruba) ancestry but left unexplained residual genomic variance with affinities to deeply diverged sub-Saharan lineages in central, East and South Africa. Based on this I proposed that the Iberomaurusian hunter-gatherers may have derived up to one-third of their ancestry from an unknown ghost population that did not leave any unadmixed descendants.

Lastly, I explicitly examined whether the Iberomaurusian hunter-gatherers carried admixture signals that would indicate a direct genetic exchange with European hunter-gatherers across the Mediterranean Sea. I found no genomic evidence for direct population contact, despite the various long-standing archaeological hypotheses for material connections between Middle or Late Stone Age hunter-gatherers from northern Africa and Palaeolithic or Mesolithic hunter-gatherers from Europe.

3.2. Manuscript B

"Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula"

Vanessa Villalba-Mouco, Marieke Sophia van de Loosdrecht, Cosimo Posth, Rafael Mora, Jorge Martínez-Moreno, Manuel Rojo-Guerra, Domingo C. Salazar-García, José I. Royo-Guillén, Michael Kunst, Hélène Rougier, Isabelle Crevecoeur, Héctor Arcusa-Magallón, Cristina Tejedor-Rodríguez, Iñigo García-Martínez de Lagrán, Rafael Garrido-Pena, Kurt W. Alt, Choongwon Jeong, Stephan Schiffels, Pilar Utrilla, Johannes Krause, Wolfgang Haak

Published in *Current Biology* (April 2019)

Synopsis: In Manuscript B, I aimed to refine the hunter-gatherer genomic structure in the Iberian Peninsula during the Upper Palaeolithic and Mesolithic, and to characterise the hunter-gatherer ancestry sources that contributed to Neolithic farmer groups.

I reconstructed the genome-wide ancestry profiles of ten prehistoric Iberians (~13,000-6,000 calBP) that were retrieved from various sites and archaeological contexts: *Balma Guilanyà* (Late Upper Paleolithic), *Moita do Sebastião* (Late Mesolithic), *Cueva de Chaves* (Early and Middle Neolithic), *Fuente Celada* (Early Neolithic) and *Cova de Els Trocs* (Middle Neolithic) (Figure-3). I extracted DNA from various skeletal elements, created double- and single-stranded DNA libraries and used in-solution capture to selectively enrich for the full mitogenome and >1.2 million ancestry informative SNPs in the nuclear genome. In addition, I increased the SNP coverage for the previously published ~15,000 calBP *GoyetQ2* hunter-gatherer from central Europe to more accurately model Magdalenian-associated ancestry.

First, I searched for patterns of population genomic substructure among the Iberian hunter-gatherers using various *f*-statistic-based methods. I found that the Iberian hunter-gatherers fall along a previously unknown cline that is driven by asymmetric affinities to Villabruna- and Magdalenian-related ancestry. Mesolithic hunter-gatherers from the Cantabrian region in northern Iberia (*Canes1* and *La Braña)* carried more Villabruna-related ancestry, whereas the Upper Palaeolithic (*El Mirón*, *Balma*) and Mesolithic hunter-gatherers from northwestern Iberia (*Chan*) and the Atlantic coast (*Moita*) retained more Magdalenian-associated ancestry. Since the oldest hunter-gatherer from Iberia, ~19,000 calBP *El Mirón*, carried both ancestries, this suggests that the population substructure in the Iberian Peninsula formed from admixture following repopulation

expansions after the LGM. Moreover, Iberian hunter-gatherers seemingly retained higher admixture proportions of the Magdalenian-associated ancestry than contemporaneous hunter-gatherers from central Europe.

Subsequently, I investigated whether the Mesolithic hunter-gatherer genomic substructure was preserved in the subsequent Iberian early farmers. I found consistently higher proportions of Magdalenian-associated ancestry in early farmers from Iberia compared to those outside Iberia. Notably, the hunter-gatherer proportion in early farmers from southern Iberia was solely derived from Magdalenian-associated ancestry. This implies that the Mesolithic substructure had persisted among early farmers, and therefore that local Iberian hunter-gatherers had contributed directly to the gene pool of incoming early farmers.

Lastly, I examined the genomic evidence for a connection with northwestern Africa across the Strait of Gibraltar. Some archaeologists have proposed that the lithics of the microburin technique and geometrics as found among hunter-gatherers in southern, eastern and western Europe had their origins in the Capsian lithics produced by the post-glacial foragers in the Maghreb in northwestern Africa. However, I found no support for direct gene flow between North Africans with Iberomaurusian-related ancestry and the Geometric Mesolithic-associated individual from *Moita do Sebastião* in western Iberia.

Note: a supplementary data file and a video abstract can be accessed via the following DOI: https://doi.org/10.1016/j.cub.2019.02.006

3.3. Manuscript C

"Genomic and dietary transitions during the Mesolithic and Early Neolithic in Sicily"

Marieke Sophia van de Loosdrecht, Marcello Mannino, Sahra Talamo, Vanessa Villalba-Mouco, Cosimo Posth, Franziska Aron, Guido Brandt, Marta Burri, Cäcilia Freund, Rita Radzeviciute, Raphaela Stahl, Antje Wissgott, Lysann Klausnitzer, Sarah Nagel, Matthias Meyer, Antonio Tagliacozzo, Marcello Piperno, Sebastiano Tusa, Carmine Collina, Vittoria Schimmenti, Rosaria Di Salvo, Kay Prüfer, Jean-Jacques Hublin, Stephan Schiffels, Choongwon Jeong, Wolfgang Haak, Johannes Krause

Submitted to *Science Advances* (March 2020)

Synopsis: In Manuscript C, I aimed to characterise the Mesolithic genomic landscape and clarify the underlying demographic processes of the Neolithic transition in Sicily.

I jointly analysed genomic and stable isotopic data for a transect of 19 prehistoric individuals that were retrieved from *Grotta dell'Uzzo* on the island of Sicily, southern Italy (~11,700-7,200 calBP, Figure-3). I extracted DNA from various skeletal elements, created double- and single-stranded DNA libraries, and used in-solution capture to selectively enrich for the complete mitogenome and >1.2 million ancestry informative SNPs in the nuclear genome. The individuals clustered in three genetic groups: Early Mesolithic foragers, Late Mesolithic foragers, and Early Neolithic farmers.

First, I compared the ancestry of the Sicilian Early and Late Mesolithic hunter-gatherers, and quantified the degree of ancestry continuation between them with several *f*-statistic-based methods. I found that the Sicilian Late Mesolithic hunter-gatherers derived the majority of their ancestry from the preceding local Early Mesolithic foragers. However, about a quarter of their ancestry was derived from an EHG source related to hunter-gatherers in eastern/southeastern Europe and the Near East. This implies a significant introduction of non-local ancestry in Sicily during the Mesolithic. This genomic transition appears to have been linked to a transition in diet; the isotope values indicated that the Late Mesolithic hunter-gatherers consumed significantly more marine-based protein than the Early Mesolithic.

A further examination showed that not only the Late Mesolithic Sicilian foragers, but also contemporaneous hunter-gatherers from the Balkans, Baltic and Iberia carried the distinct EHG-

related ancestry, albeit in different proportions. Importantly, all these Late Mesolithic hunter-gatherers are associated with blade-and-trapeze industries. I therefore concluded that the EHG-related ancestry in European Late Mesolithic foragers probably reflects the spread of the blade-and-trapeze complex. Some archaeologists proposed that the blade-and-trapeze industries in Europe were introduced from Crimea via southeastern Europe or from the Maghreb in northwestern Africa. The substantial EHG-related ancestry in the genomes of Late Mesolithic European hunter-gatherers associated with the blade-and-trapeze complex provides support for an introduction from eastern/southeastern Europe. Notably, the isotopic values of the Late Mesolithic Sicilian foragers are similar to previously reported values for some contemporaneous Iron Gates hunter-gatherers, hence providing additional support for a connection to southeastern Europe.

Lastly, I characterised the ancestry of the Sicilian early farmers to investigate the Neolithisation process. Using various *f*-statistics I found that the Sicilian early farmers showed the highest genomic similarity to early farmers from the Balkans, followed by farmers associated with the Danubian expansion route. This provides support for a northern Mediterranean expansion route for these Sicilian early farmers. To examine the role of acculturation in the Neolithic transition, I quantified the amount of hunter-gatherer ancestry in the Sicilian early farmers. I estimated that the local Sicilian Mesolithic foragers contributed a maximum of only ~7% ancestry. This implies a large-scale demographic transition during the Early Neolithic. The isotopic values indicated that these early farmers had a predominantly terrestrial-based farming diet, suggesting that a genomic and dietary transition may have occurred in tandem during the Early Neolithic. Strikingly, the radiocarbon dates for two individuals with a typical Late Mesolithic ancestry are intermediate to those of the Late Mesolithic foragers and Early Neolithic farmers. These two individuals had distinctive diets and may have lived during a time when some of the earliest aspects of Impressa Ware appeared at the site. These individuals therefore may provide tentative initial evidence that, prior to the full-scale replacement by early farmers during the Early Neolithic, Sicilian foragers adopted some elements of early farming.

4. Author contributions

1. van de Loosdrecht, M.S., Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., Nickel, B., Nagel, S., Talbi, E.H., El Hajraoui, M.A., Amzazi, S., Hublin, J-J., Pääbo, S., Schiffels, S., Meyer, M., Haak, W.*, Jeong, C.*, Krause, J.* (2018**). Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations**. *Science*, 360(6388), 548-552. doi:10.1126/science.aar8380

Author contributions:
*These authors jointly supervised this work: Johannes Krause, Wolfgang Haak, Choongwon Jeong.

Johannes Krause, Abdeljalil Bouzouggar and Louise Humphrey conceived the study. Abdeljalil Bouzouggar, Louise Humphrey, Nick Barton and Jean-Jacques Hublin provided archaeological material and input for the archaeological interpretation. I sampled the petrous bones, extracted DNA, and generated double-stranded libraries to screen for DNA preservation at the Max Planck Institute for the Science of Human History (MPI-SHH) in Jena. Then I generated extracts for high production libraries that were processed into single-stranded libraries and captured by Birgit Nickel and Sarah Nagel, with the help of Ayinuer Aximu-Petri and Matthias Meyer, at the MPI for Evolutionary Anthropology (MPI-EVA) in Leipzig. All captured libraries were sequenced at the MPI-SHH by the laboratory staff. I performed the sequence quality analyses with help from Cosimo Posth. Under the supervision of Choongwon Jeong, I conducted the analyses for hypothesis formulation (PCA, ADMIXTURE), broad ancestry characterisation (f_3-, f_4-statistics, Neanderthal admixture), *qpAdm*-based ancestry modelling for the characterisation of the Eurasian ancestry, Y-chromosome haplotyping and phenotypic SNP analyses. Choongwon Jeong performed the *qpAdm*-based ancestry modelling for the sub-Saharan African component and the scaffold graph to test for European HG ancestry. Wolfgang Haak helped me with the mitogenome haplotyping and Cosimo Posth performed the BEAST analysis. The supplementary sections related to the archaeological context were written by Abdeljalil Bouzouggar, Louise Humphrey and Nick Barton (S1,2), and those addressing the genomic analyses by myself (S3-8, 10-12) and Choongwon Jeong (S8, 9, 13), with input from all co-authors. I wrote the majority of the main report with considerable contributions from Choongwon Jeong, minor contributions from Cosimo Posth, and input from all co-authors, under the supervision of Wolfgang Haak and Johannes Krause.

Sample procurement	0%
Laboratory work	70%
Bioinformatic data processing & quality control	90%
Population genomic analysis	60%
Manuscript writing	70%

2. Villalba-Mouco, V., **van de Loosdrecht, M**. **S.**, Posth, C., Mora, R., Martínez-Moreno, J., Rojo-Guerra, M., Rojo-Guerra, M., Salazar-García, D.C., Royo-Guillén, J.I., Kunst, M., Rougier, H., Crevecoeur, I., Arcusa-Magallón, H., Tejedor-Rodríguez, C., García-Martínez de Lagrán, I., Garrido-Pena, R., Alt, K.W., Jeong, C., Schiffels, S., Utrilla, P., Krause, J., Haak, W. (2019). **Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula**. *Current Biology*, 29(7), 1169-1177. doi:10.1016/j.cub.2019.02.006

Author contributions:

Vanessa Villalba-Mouco, Johannes Krause, and Wolfgang Haak conceived the study. Rafael Mora, Jorge Martínez-Moreno, Manuel Rojo-Guerra, Domingo Salazar-García, José Royo-Guillén, Michael Kunst, Hélène Rougier, Isabelle Crevecoeur, Héctor Arcusa-Magallón, Cristina Tejedor-Rodríguez, Iñigo García-Martínez de Lagrán, Rafael Garrido-Pena, Kurt Alt and Pilar Utrilla assembled the ancient human remains and provided the archaeological context. Most of the laboratory work was performed by Vanessa Villalba-Mouco, visiting PhD student from the University of Zaragoza, who I instructed and co-supervised. Vanessa Villalba-Mouco performed the quality control analyses under the co-supervision of, and partly together with, me and Cosimo Posth. The population genomic analyses were performed by Vanessa Villalba-Mouco under the co-supervision of Wolfgang Haak, Cosimo Posth and myself. In parallel to Vanessa, I analysed the Palaeolithic and Mesolithic foragers and contributed results (*f*-statistics to demonstrate the WHG-Magdalenian ancestry cline), and helped with the design and results interpretation of the *qpWave*- and *qpAdm*-based ancestry models. Vanessa Villalba-Mouco, with contributions from Cosimo Posth and supervised by Wolfgang Haak, wrote the majority of the main report and supplementary materials with input from all co-authors. I contributed in particular to the “Genetic Structure in Iberian Hunter-Gatherers” section in the main text, co-wrote the method sections related to the lab work, sequence read processing and quality control, and was actively involved in the writing of preliminary drafts.

Sample procurement	0%
Laboratory work	30%
Bioinformatic data processing & quality control	30%
Population genetic analysis	20%
Manuscript writing	15%

3. van de Loosdrecht, M. S., Mannino, M. A., Talamo, S., Villalba-Mouco, V., Posth, C., Aron, F., Brandt, G., Burri, M., Freund, C., Radzeviciute, R., Stahl, R., Wissgott, A., Klausnitzer, L., Nagel., Meyer, M., Tagliacozzo, A., Piperno, M., Tusa, S., Collina, C., Schimmenti, V., Di Salvo, R., Prüfer, K., Hublin, J-J., Schiffels, S., Jeong, C., Haak, W.*, Krause, J.* (2020). **Genomic and dietary transitions during the Mesolithic and Early Neolithic in Sicily**. *BiorXiv*, 2020.2003.2011.986158. doi:10.1101/2020.03.11.986158

Author contributions:

* These authors jointly supervised this work: Johannes Krause, Wolfgang Haak

Marcello Mannino, Wolfgang Haak and Johannes Krause conceived the study. Marcello Mannino, Antonio Tagliacozzo, Marcello Piperno, Sebastiano Tusa, Carmine Collina, Vittoria Schimmenti, and Rosaria Di Salvo provided the ancient human remains and input for the archaeological interpretation. I sampled 102 human skeletal elements, and generated extracts and double-stranded libraries for screening and capture at the MPI-SHH. For the skeletal remains of 19 individuals, I created additional extracts for single-stranded libraries that were generated by Sarah Nagel at the MPI-EVA. All the 1240k and mitogenome capture enrichments and sequencing were done at the MPI-SHH by the laboratory staff (Franziska Aron, Guido Brandt, Marta Burri, Rita Radzeviciute, Raphaela Stahl and Antje Wissgott). Lysann Klausnitzer and Sahra Talamo extracted collagen and performed the AMS radiocarbon dating. Sahra Talamo conducted the radiocarbon dating analysis, and Marcello Mannino the stable isotope analysis. I performed the data quality control and nearly all the population genomic analyses. Kay Prüfer provided a script to quantify nucleotide diversity, and helped with the interpretation of its results. Marcello Mannino wrote the supplementary sections related to the archaeological context (S1), whereas I wrote the population genomics ones (S2-7). I wrote the majority of the main report with input from all co-authors, under the supervision of Wolfgang Haak and Johannes Krause.

Sample procurement	0%
Laboratory work	70%
Bioinformatic data processing & quality control	90%
Population genetic analysis	90%
Manuscript writing	90%

Supervisors:

Johannes Krause

.

5. Manuscript A

Van de Loosdrecht et al. 2018. *Science*

PALEOGENOMICS

Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations

Marieke van de Loosdrecht,[1] Abdeljalil Bouzouggar,[2,3]*† Louise Humphrey,[4] Cosimo Posth,[1] Nick Barton,[5] Ayinuer Aximu-Petri,[6] Birgit Nickel,[6] Sarah Nagel,[6] El Hassan Talbi,[7] Mohammed Abdeljalil El Hajraoui,[2] Saaïd Amzazi,[8] Jean-Jacques Hublin,[3] Svante Pääbo,[6] Stephan Schiffels,[1] Matthias Meyer,[6] Wolfgang Haak,[1]† Choongwon Jeong,[1]*† Johannes Krause[1]*†

North Africa is a key region for understanding human history, but the genetic history of its people is largely unknown. We present genomic data from seven 15,000-year-old modern humans, attributed to the Iberomaurusian culture, from Morocco. We find a genetic affinity with early Holocene Near Easterners, best represented by Levantine Natufians, suggesting a pre-agricultural connection between Africa and the Near East. We do not find evidence for gene flow from Paleolithic Europeans to Late Pleistocene North Africans. The Taforalt individuals derive one-third of their ancestry from sub-Saharan Africans, best approximated by a mixture of genetic components preserved in present-day West and East Africans. Thus, we provide direct evidence for genetic interactions between modern humans across Africa and Eurasia in the Pleistocene.

Under typical conditions (i.e., aside from intermittent greening periods), the Sahara desert poses an ecogeographic barrier for human migration between North and sub-Saharan Africa (*1*). Sub-Saharan Africa is home to the most deeply divergent genetic lineages among present-day humans (*2*), and the general view is that all Eurasians mostly descend from a single group of humans that dispersed outside of sub-Saharan Africa around 50,000 to 100,000 years before the present (yr B.P.) (*3*). This group likely represented only a small fraction of the genetic diversity within Africa, most closely related to a Holocene East African group (*4*). Present-day North Africans share a majority of their ancestry with present-day Near Easterners but not with sub-Saharan Africans (*5*). Thus, from a genetic perspective, present-day North Africa is largely a part of Eurasia. However, the temporal depth of this genetic connection between the Near East and North Africa is poorly understood and has been estimated only indirectly from present-day mitochondrial DNA (mtDNA) variation (*6*, *7*).

Owing to challenging conditions for DNA preservation, relatively few ancient genomes have been recovered from Africa. Genome-wide data from 23 individuals have been reported from South and East Africa, with the oldest dating back to 8100 yr B.P. (*4*, *8*, *9*). In North Africa, a genomic study of Egyptian mummies from the first millennium BCE showed that the genetic connection between the Near East and North Africa was established by that time (*5*). However, the genetic affinity of North African populations at a greater time depth has remained unknown.

Here we present genome-wide data from seven individuals, directly dated between 15,100 and 13,900 calibrated years before present (cal. yr B.P.) (table S1), from Grotte des Pigeons near Taforalt in eastern Morocco (*10*). These genomic data provide a critical reference point to help explain the deep genetic history of North Africa and the broader Middle East (Fig. 1). The Taforalt individuals are associated with the Later Stone Age Iberomaurusian culture, whose origin is debated. These individuals may have descended either directly from the manufacturers of the preceding Middle Stone Age technologies (Aterian or local West African bladelet technologies) or from an exogenous population with ties to the Upper Paleolithic technocomplexes of the Near East or Southern Europe (*10*, *11*).

For nine Taforalt individuals (table S2), we created double-indexed single-stranded DNA libraries (*12*) for next-generation sequencing of DNA isolated from petrous bones. We then used in-solution capture probes (*13*) to enrich libraries for the whole mitochondrial genome and ~1,240,000 single-nucleotide polymorphisms (SNPs) in the nuclear genome (*14*). The DNA fragments obtained from seven individuals, six genetic males and one female, had postmortem degradation characteristics typical of ancient DNA (tables S3 to S5 and fig. S6). We reconstructed the mitochondrial genomes of all seven individuals (102× to 1701× coverage, unmerged libraries; table S4) while maintaining a low level of contamination from the DNA of modern humans (1 to 8%; table S4). For the nuclear data analysis, in which ancient DNA is more susceptible to contamination than in mitochondrial analyses, we analyzed five individuals (four males and one female) on the basis of coverage (table S3, merged libraries) and negligible modern human contamination for males (1.7 to 2.5%; table S5). For each individual, we randomly chose a single base per site as a haploid genotype. We intersected our new data with data from a panel of worldwide present-day populations, genotyped on the Affymetrix Human Origins array for ~600,000 markers, as well as ancient genomic data covering Europe, the Near East, and sub-Saharan Africa (*4*, *8*, *15*–*17*). The final data set includes 593,124 intersecting autosomal SNPs with 183,041 to 544,232 SNP positions covered for each of the five individuals (table S3). For group-based analyses involving other ancient individuals, we adopted the population labels from the original studies (*4*, *16*). We found an overall high genetic relatedness between the Taforalt individuals, suggesting a strong population bottleneck (fig. S26).

We analyzed the genetic affinities of the Taforalt individuals by performing principal components analysis and model-based clustering of worldwide data (Fig. 2). When projected onto the top principal components of African and west Eurasian populations, the Taforalt individuals form a distinct cluster in an intermediate position between present-day North Africans [e.g., Amazighes (Berbers), Mozabites, and Saharawis] and East Africans (e.g., Afars, Oromos, and Somalis) (Fig. 2A). Consistently, we find that all males with sufficient nuclear DNA preservation carry Y haplogroup E1b1b1a1 (M-78; table S16). This haplogroup occurs most frequently in present-day North and East African populations (*18*). The closely related E1b1b1b (M-123) haplogroup has been reported for Epipaleolithic Natufians and Pre-Pottery Neolithic Levantines (Levant_N) (*16*). Unsupervised genetic clustering also suggests a connection of Taforalt to the Near East. The three major components that make up the Taforalt genomes are maximized in early Holocene Levantines, East African hunter-gatherer Hadza from north-central Tanzania, and West Africans (number of genetic clusters K = 10; Fig. 2B). In contrast, present-day North Africans have smaller sub-Saharan African components with minimal Hadza-related contribution (Fig. 2B).

We calculated outgroup f_3 statistics of the form f_3(Taforalt, X; Mbuti) across worldwide ancient and present-day test populations. Consistent with previous analyses, we find that ancient Near Eastern populations, especially Epipaleolithic Natufians and early Neolithic Levantines, show the highest

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History (MPI-SHH), Jena, Kahlaische Strasse 10, D-07745, Germany. [2]Origin and Evolution of *Homo sapiens* in Morocco Research Group, Institut National des Sciences de l'Archéologie et du Patrimoine, Hay Riad, Madinat Al Irfane, Angle rues 5 et 7, Rabat-Instituts, 10 000 Rabat, Morocco. [3]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology (MPI-EVA), Leipzig, Deutscher Platz 6, D-04103, Germany. [4]Department of Earth Sciences, The Natural History Museum, London SW7 5BD, UK. [5]Institute of Archaeology, University of Oxford, 36 Beaumont Street, Oxford OX1 2PG, UK. [6]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology (MPI-EVA), Leipzig, Deutscher Platz 6, D-04103, Germany. [7]Faculté des Sciences, Campus d'Al Qods, Université Mohammed Premier, B.P. 717 Oujda, Morocco. [8]Mohammed V University, Avenue Ibn Batouta, Rabat, Morocco.
***Corresponding author. Email: krause@shh.mpg.de (J.K.); jeong@shh.mpg.de (C.J.); bouzouggar@eva.mpg.de (A.B.)**
†These authors contributed equally to this work.

outgroup f_3 values with Taforalt (Fig. 3A). This is confirmed by f_4 symmetry statistics of the form f_4(Chimpanzee, Taforalt; NE_1, NE_2) that measure a relative affinity of a pair of Near Eastern (NE) groups to Taforalt. A positive value indicates that NE_2 is closer than NE_1 to Taforalt. We consistently find positive f_4 values when the NE_2 group is Natufian or Levant_N and the NE_1 group is representative of other populations [z score = 2.2 to 11.0 standard error (SE); table S6]. Congruent to the outgtoup-f_3 results, the Natufian population shows higher affinity to Taforalt than does the Levant_N group (z score = 2.2 SE; table S6). This indicates that the early Holocene Levantine populations, overlapping with or postdating our Taforalt individuals by up to 6000 years (*16*), are most closely related to the Taforalt group, among Near Eastern populations. Next, we evaluated whether the Taforalt individuals have sub-Saharan African ancestry by calculating f_4(Chimpanzee, X; Natufian, Taforalt). We observe significant positive f_4 values for all sub-Saharan African groups and significant negative values for all Eurasian populations, supporting a substantial contribution from sub-Saharan Africa (Fig. 3B). West Africans, such as Mende and Yoruba, most strongly pull out the sub-Saharan African ancestry in Taforalt (Fig. 3B and figs. S15 and S16).

We investigated whether two first-hand proxies, Natufians and West Africans, are sufficient to explain the Taforalt gene pool or whether a more complex admixture model is required. We thus tested whether Natufians could be a sufficient proxy for the Eurasian ancestry in Taforalt without explicit modeling of its African ancestry (fig. S18). This line of investigation was inspired by proposed archaeological connections between the Iberomaurusian and Upper Paleolithic cultures in Southern Europe, either via the Strait of Gibraltar (*19*) or Sicily (*20*). If this connection is true, both the Upper Paleolithic European and Natufian ancestries will be required to explain the Taforalt gene pool. For our admixture modeling with the program qpAdm (*16*), we chose outgroups that can distinguish sub-Saharan African, Natufian, and Paleolithic European ancestries but are blind to differences between sub-Saharan African lineages (*11*). A two-way admixture model, comprising Natufian and sub-Saharan African populations, does not significantly deviate from our data (χ^2 $P \geq 0.128$), with 63.5% Natufian and 36.5% sub-Saharan African ancestry, on average (table S8). Adding Paleolithic European lineages as a third source only marginally increased the model fit (χ^2 $P = 0.019$ to 0.128; table S9). Consistently, by using the qpGraph package (*21*), we find that a mixture of Natufian and Yoruba reasonably fits the Taforalt gene pool ($|z| \leq 3.7$; fig. S19 and table S10). Adding gene flow from Paleolithic Europeans does not improve the model fit and provides an ancestry contribution estimate of 0% (fig. S19). We thus find no evidence of gene flow from

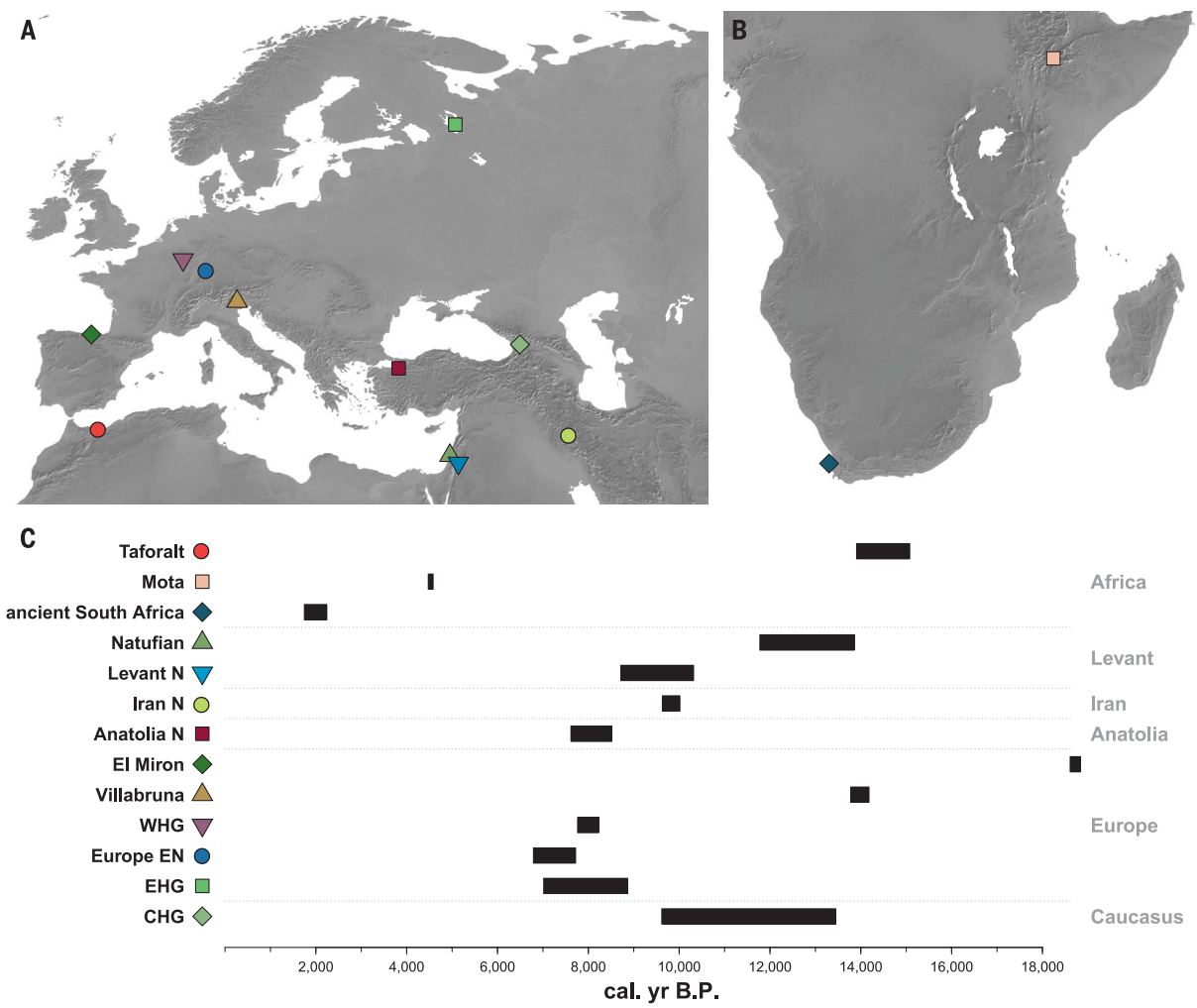


Fig. 1. Spatiotemporal locations of the Taforalt and other ancient genomes. (**A** and **B**) Geographic locations of representative ancient genomes from West Eurasia and Africa included in our analysis. The Pleistocene Taforalt site is denoted by a red circle. (**C**) The date range of each ancient group is marked by black bars, representing the range of 95% confidence intervals of radiocarbon dates across all dated individuals (cal. yr B.P. on the *x* axis). Group labels are taken from previous studies reporting each ancient genome (*4*, *16*, *27*). N, Neolithic; WHG, Western European hunter-gatherers; EHG, Eastern European hunter-gatherers; CHG, Caucasus hunter-gatherers.

Paleolithic Europeans into Taforalt within the resolution of our data.

We further characterized the sub-Saharan African–related ancestry in the Taforalt individuals by using f_4 statistics in the form f_4(Chimpanzee, African; Yoruba/Mende, Natufian). We find that Yoruba or Mende and Natufians are symmetrically related to two deeply divergent outgroups, an ancient South African group from 2000 yr B.P. (aSouthAfrica) and Mbuti Pygmy, respectively ($|z| \leq 1.564$ SE; table S11). Because f_4 statistics are linear under admixture, we expect the Taforalt population not to be any closer to these outgroups than Yoruba or Natufians if the two-way admixture model is correct. However, we find

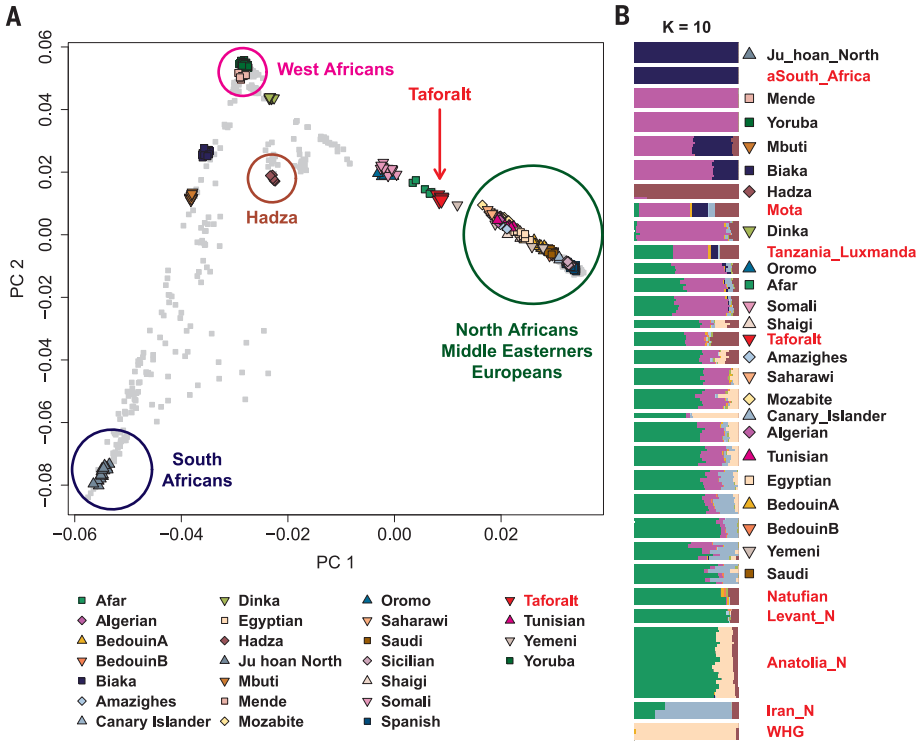


Fig. 2. Summary of the genetic profile of the Taforalt individuals. (**A**) The top two principal components (PCs) calculated from present-day African, Near Eastern, and Southern European individuals from 72 populations. The Taforalt individuals are projected thereon (red inverted triangles), and selected present-day populations are denoted by various colored symbols. Labels for other populations (denoted by small gray squares) are provided in fig. S8. (**B**) ADMIXTURE analysis results of chosen African and Middle Eastern populations (K = 10). Ancient individuals are labeled in red. Major ancestry components in Taforalt individuals are maximized in early Holocene Levantines (green), West Africans (purple), and East African Hadza (brown). The ancestry component prevalent in pre-Neolithic Europeans (beige) is absent in Taforalt.

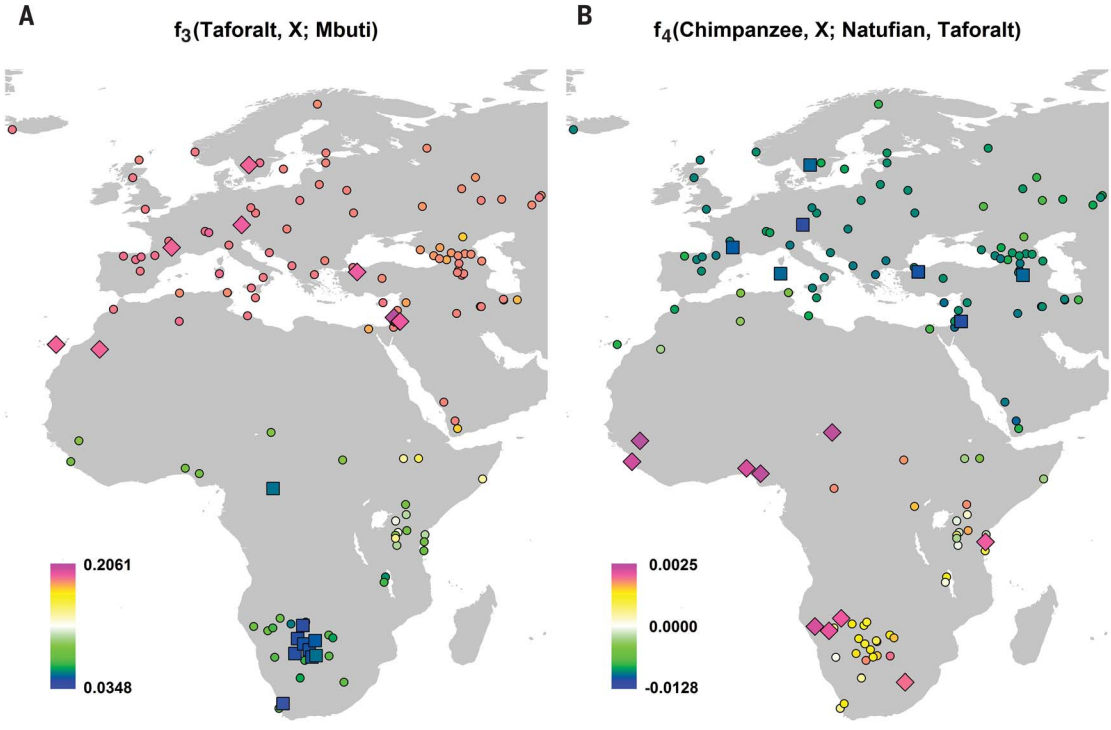


Fig. 3. Geographic distribution of the genetic affinity of the Taforalt group with worldwide populations. (**A**) Mean shared genetic drift with the Taforalt group, as measured by outgroup f_3 statistics in the form f_3 (Taforalt, X; Mbuti). Warm colors denote populations genetically close to Taforalt. Large diamonds and squares represent the 10 highest and lowest f_3 values, respectively. Early Holocene Levantine groups (Natufians and Neolithic Levantines) show the highest affinity with Taforalt. The statistics and their associated SEs for the top 30 signals are presented in fig. S14. (**B**) Extra genetic affinity with the Taforalt group in comparison to Natufians, as measured by f_4 statistics in the form f_4(Chimpanzee, X; Natufian, Taforalt). Large diamonds and squares represent the 10 most positive and negative f_4 values, respectively. Sub-Saharan Africans show high positive values, with West African Yoruba and Mende having the highest values, supporting the presence of sub-Saharan African ancestry in Taforalt individuals. In contrast, all Eurasian populations are genetically closer to Natufians than to the Taforalt group. The statistics and their associated SEs for the top 30 signals are presented in fig. S16.

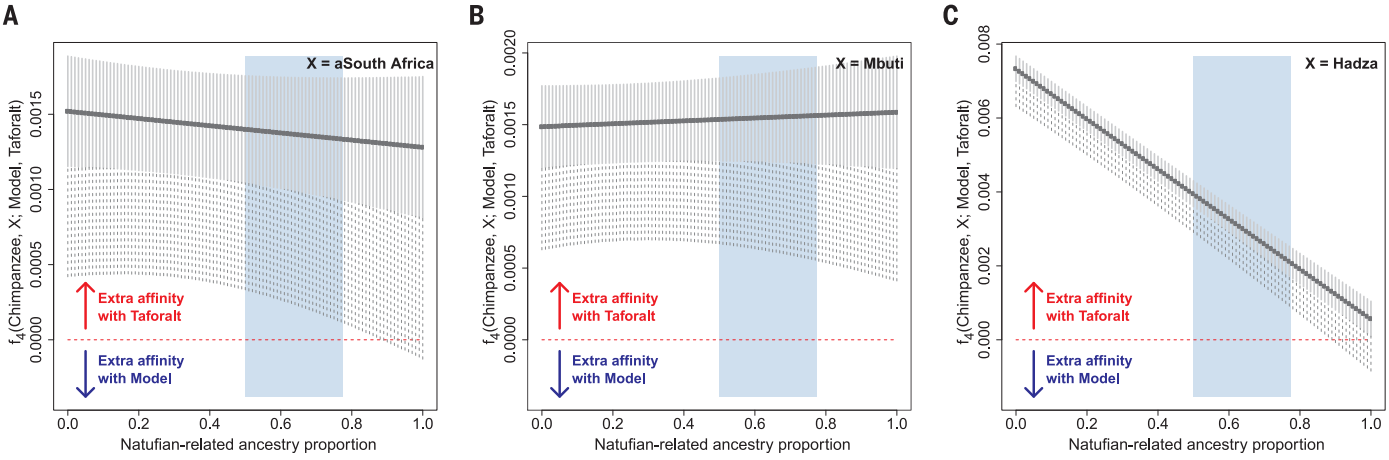


Fig. 4. Relative genetic affinity of representative sub-Saharan African groups to a mixture of Yoruba and Natufians in comparison to the Taforalt group. We measured f_4 statistics in the form f_4(Chimpanzee, African; Yoruba+Natufian, Taforalt) by using (**A**) aSouthAfrica, (**B**) Mbuti, and (**C**) Hadza as the African group. The f_4 statistics were calculated for the proportions of Natufian-related ancestry ranging from 0 to 100% in increments of 1%. The blue rectangle marks a plausible range of Natufian ancestry proportion, estimated by our qpAdm modeling [0.637 ± (2 × 0.069)]. Gray solid and dotted lines represent ±1 and −3 SE ranges, respectively. SEs were calculated by 5-centimorgan block jackknife method.

instead that the Taforalt group is significantly closer to both outgroups (aSouthAfrica and Mbuti) than any combination of Yoruba and Natufians ($z \geq 2.728$ SE; Fig. 4). A similar pattern is observed for the East African outgroups Dinka, Mota, and Hadza (table S11 and fig. S20). These results can only be explained by Taforalt harboring an ancestry that contains additional affinity with South, East, and Central African outgroups. None of the present-day or ancient Holocene African groups serve as a good proxy for this unknown ancestry, because adding them as the third source is still insufficient to match the model to the Taforalt gene pool (table S12 and fig. S21). However, we can exclude any branch in human genetic diversity more basal than the deepest known one represented by aSouthAfrica (*4*) as the source of this signal: it would result in a negative affinity to aSouthAfrica, not a positive one as we find (Fig. 4). Both an unknown archaic hominin and the recently proposed deep West African lineage (*4*) belong to this category and therefore cannot explain the Taforalt gene pool.

Mitochondrial consensus sequences of the Taforalt individuals belong to the U6a (six individuals) and M1b (one individual) haplogroups (*15*), which are mostly confined to present-day populations in North and East Africa (*7*). U6 and M1 have been proposed as markers for autochthonous Maghreb ancestry, which might have been originally introduced into this region by a back-to-Africa migration from West Asia (*6*, *7*). The occurrence of both haplogroups in the Taforalt individuals proves their pre-Holocene presence in the Maghreb. We used the BEAST v1.8.1 package (*24*) to analyze the seven ancient Taforalt individuals in combination with four Upper Paleolithic European mtDNA genomes (*22*, *23*) and present-day individuals belonging to U6 and M1 (*7*). By using a human mtDNA mutation rate inferred from tip calibration of ancient mtDNA genomes (*23*), we obtained divergence estimates for U6 at 37,000 yr B.P. (40,000 to 34,000 yr B.P. for 95% highest posterior density, HPD) and M1 at 24,000 yr B.P. (95% HPD: 29,000 to 20,000 yr B.P.) (table S15). Our estimated dates are considerably more recent than those of a study using present-day data only (45,000 ± 7000 yr B.P. for U6 and 37,000 ± 7000 yr B.P. for M1) (*7*) but are similar to those of Pennarun *et al.* (*25*). Moreover, we observed an asynchronous increase in the effective population size for U6 and M1 (fig. S24), which suggests that the demographic histories of these North and East African haplogroups do not coincide and might have been influenced by multiple expansions in the Late Pleistocene (*25*). Notably, the diversification of haplogroups U6a and M1 found for Taforalt is dated to ~24,000 yr B.P. (fig. S23), which is close in time to the earliest known appearance of the Iberomaurusian culture in Northwest Africa [25,845 to 25,270 cal. yr B.P. at Tamar Hat (*26*)].

The relationships of the Iberomaurusian culture with those of the preceding Middle Stone Age, including the local backed bladelet technologies in Northeast Africa, and the Epigravettian in Southern Europe have been questioned (*13*). The genetic profile of Taforalt suggests substantial Natufian-related and sub-Saharan African–related ancestries (63.5 and 36.5%, respectively) but not additional ancestry from Epigravettian or other Upper Paleolithic European populations. Therefore, we provide genomic evidence for a Late Pleistocene connection between North Africa and the Near East, predating the Neolithic transition by at least four millennia, while rejecting the hypothesis of a potential Epigravettian gene flow from Southern Europe into northern Africa, within the resolution of our data. Archaeogenetic studies on additional Iberomaurusian sites will be critical to evaluate the representativeness of Taforalt for the Iberomaurusian gene pool. We speculate that the Natufian-related ancestral population may have been widespread across North Africa and the Near East, associated with microlithic backed bladelet technologies that started to spread out in this area by at least 25,000 yr B.P. [(*10*) and references therein]. However, given the absence of ancient genomic data from a similar time frame for this broader area, the epicenter of expansion, if any, for this ancestral population remains unknown.

Although the oldest Iberomaurusian microlithic bladelet technologies are found earlier in the Maghreb than their equivalents in northeastern Africa (Cyrenaica) and the earliest Natufian in the Levant, the complex sub-Saharan ancestry in Taforalt makes our individuals an unlikely proxy for the ancestral population of later Natufians who do not harbor sub-Saharan ancestry. An epicenter in the Maghreb is plausible only if the sub-Saharan African admixture into Taforalt either postdated the expansion into the Levant or was a locally confined phenomenon. Alternatively, placing the epicenter in Cyrenaica or the Levant requires an additional explanation for the observed archaeological chronology.

REFERENCES AND NOTES

1. I. S. Castañeda *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20159–20163 (2009).
2. B. M. Henn *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5154–5162 (2011).
3. S. Mallick *et al.*, *Nature* **538**, 201–206 (2016).
4. P. Skoglund *et al.*, *Cell* **171**, 59–71.e21 (2017).
5. V. J. Schuenemann *et al.*, *Nat. Commun.* **8**, 15694–15704 (2017).
6. A. M. González *et al.*, *BMC Genomics* **8**, 223–234 (2007).
7. A. Olivieri *et al.*, *Science* **314**, 1767–1770 (2006).
8. M. Gallego Llorente *et al.*, *Science* **350**, 820–822 (2015).
9. C. M. Schlebusch *et al.*, *Science* **358**, 652–655 (2017).
10. R. N. Barton *et al.*, *J. Hum. Evol.* **65**, 266–281 (2013).
11. See supplementary materials.
12. M.-T. Gansauge *et al.*, *Nucleic Acids Res.* **45**, e79 (2017).
13. Q. Fu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2223–2227 (2013).
14. Q. Fu *et al.*, *Nature* **524**, 216–219 (2015).
15. E. R. Jones *et al.*, *Nat. Commun.* **6**, 8912–8919 (2015).
16. I. Lazaridis *et al.*, *Nature* **536**, 419–424 (2016).
17. M. Raghavan *et al.*, *Nature* **505**, 87–91 (2014).
18. F. Cruciani *et al.*, *Mol. Biol. Evol.* **24**, 1300–1311 (2007).
19. P. Pallary, in *Mémoires de la Société Historique Algérienne* (Jourdan, 1909), vol. 3.

20. G. Camps, *Les Civilisations Préhistoriques de l'Afrique du Nord et du Sahara* (Doin, 1974).
21. N. Patterson *et al.*, *Genetics* **192**, 1065–1093 (2012).
22. M. Hervella *et al.*, *Sci. Rep.* **6**, 25501–25505 (2016).
23. C. Posth *et al.*, *Curr. Biol.* **26**, 827–833 (2016).
24. A. J. Drummond, A. Rambaut, *BMC Evol. Biol.* **7**, 214 (2007).
25. E. Pennarun *et al.*, *BMC Evol. Biol.* **12**, 234–245 (2012).
26. J. T. Hogue, R. Barton, *Quat. Int.* **413**, 62–75 (2016).
27. Q. Fu *et al.*, *Nature* **534**, 200–205 (2016).

ACKNOWLEDGMENTS

We thank H. Temming and A. Le Cabec (MPI-EVA) for CT scanning and G. Brandt, A. Wissgott, F. Aron, M. Burri, C. Freund, and R. Stahl (MPI-SHH) for DNA sequencing. **Funding:** This work was supported by the Max Planck Society, Institut National des Sciences de l'Archéologie et du Patrimoine (Protars grant P32/09-CNRST), the Natural Environment Research Council (grants EFCHED NER/T/S/2002/00700 and RESET NE/E015670/1), the Leverhulme Trust (grant F/08 735/F), the British Academy, Oxford University (Fell Fund, Boise and Meyerstein), the Natural History Museum (Human Origins Research Fund), and the Calleva Foundation. **Author contributions:** J.K., A.B., J.J.-H., and L.H. conceived of the study. A.B., L.H., N.B., and J.-J.H. provided archaeological material and input for the archaeological interpretation. M.v.d.L., B.N., and S.N. performed laboratory work with the help of A.A.-P. and M.M. M.v.d.L., C.J., C.P., and W.H. analyzed data. M.v.d.L., C.J., J.K., C.P., A.B., L.H., N.B., and M.M. wrote the manuscript with input from all coauthors. **Competing interests:** A.B. keeps an additional affiliation with the MPI-EVA; this institute supported his excavation efforts and worked with him on the site. This is also reflected in the coauthorship of J.J.-H., director of MPI-EVA. The authors declare no competing interests. **Data and materials availability:** Genomic data (BAM format) are available through the Sequence Read Archive (accession number SRP132033) and consensus mitogenome sequences (FASTA format) in GenBank (accession numbers MG936619 to MG936625).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6388/548/suppl/DC1
Supplementary Text
Figs. S1 to S26
Tables S1 to S16
References (*28–114*)

21 December 2017; accepted 28 February 2018
Published online 15 March 2018
10.1126/science.aar8380

6. Manuscript B

Villalba-Mouco et al. 2019. *Current Biology*

Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula

Vanessa Villalba-Mouco,[1,2] Marieke S. van de Loosdrecht,[1] Cosimo Posth,[1] Rafael Mora,[3] Jorge Martínez-Moreno,[3] Manuel Rojo-Guerra,[4] Domingo C. Salazar-García,[5] José I. Royo-Guillén,[6] Michael Kunst,[7] Hélène Rougier,[8] Isabelle Crevecoeur,[9] Héctor Arcusa-Magallón,[10] Cristina Tejedor-Rodríguez,[11] Iñigo García-Martínez de Lagrán,[12] Rafael Garrido-Pena,[13] Kurt W. Alt,[14,15] Choongwon Jeong,[1] Stephan Schiffels,[1] Pilar Utrilla,[2] Johannes Krause,[1] and Wolfgang Haak[1,16,*]

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Kahlaische Straße 10, 07745 Jena, Germany
[2]Departamento de Ciencias de la Antigüedad, Grupo Primeros Pobladores del Valle del Ebro (PPVE), Instituto de Investigación en Ciencias Ambientales (IUCA), Universidad de Zaragoza, Pedro Cerbuna, 50009 Zaragoza, Spain
[3]Centre d'Estudis del Patrimoni Arqueològic de la Prehistòria (CEPAP), Facultat de Lletres, Universitat Autònoma Barcelona, 08190 Bellaterra, Spain
[4]Department of Prehistory, University of Valladolid, Plaza del Campus, 47011 Valladolid, Spain
[5]Grupo de Investigación en Prehistoria IT-622-13 (UPV-EHU)/IKERBASQUE-Basque Foundation for Science, Euskal Herriko Unibertsitatea, Francisco Tomas y Valiente s/n., 01006 Vitoria, Spain
[6]Dirección General de Cultura y Patrimonio, Gobierno de Aragón, Avenida de Ranillas, 5 D., 50018 Zaragoza, Spain
[7]Instituto Arqueológico Alemán, Calle Serrano 159, E-28002 Madrid, Spain
[8]Department of Anthropology, California State University, Northridge, Northridge, California 91330, USA
[9]Université de Bordeaux, CNRS, UMR 5199-PACEA, 33615 Pessac Cedex, France
[10]Arcadia-FUNGE, Fundación General de la Universidad de Valladolid, 47002 Valladolid, Spain
[11]Juan de la Cierva-Formación Program, Institute of Heritage Sciences, Spanish National Research Council (Incipit, CSIC), Av. de Vigo, 15705 Santiago de Compostela, Spain
[12]Juan de la Cierva-Incorporación Program, Department of Prehistory, Valladolid University, Plaza del Campus, 47011 Valladolid, Spain
[13]Department of Prehistory, Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid, Spain
[14]Center of Natural and Cultural Human History, Danube Private University, Steiner Landstr. 124, A-3500 Krems, Austria
[15]Department of Biomedical Engineering, University of Basel, Gewerbestrasse 14-16, CH-4123 Allschwil, Switzerland
[16]Lead Contact
*Correspondence: haak@shh.mpg.de
https://doi.org/10.1016/j.cub.2019.02.006

SUMMARY

The Iberian Peninsula in southwestern Europe represents an important test case for the study of human population movements during prehistoric periods. During the Last Glacial Maximum (LGM), the peninsula formed a periglacial refugium [1] for hunter-gatherers (HGs) and thus served as a potential source for the re-peopling of northern latitudes [2]. The post-LGM genetic signature was previously described as a cline from Western HG (WHG) to Eastern HG (EHG), further shaped by later Holocene expansions from the Near East and the North Pontic steppes [3–9]. Western and central Europe were dominated by ancestry associated with the ~14,000-year-old individual from Villabruna, Italy, which had largely replaced earlier genetic ancestry, represented by 19,000–15,000-year-old individuals associated with the Magdalenian culture [2]. However, little is known about the genetic diversity in southern European refugia, the presence of distinct genetic clusters, and correspondence with geography. Here, we report new genome-wide data from 11 HGs and Neolithic individuals that highlight the late survival of Paleolithic ancestry in Iberia, reported previously in Magdalenian-associated individuals. We show that all Iberian HGs, including the oldest, a ~19,000-year-old individual from El Mirón in Spain, carry dual ancestry from both Villabruna and the Magdalenian-related individuals. Thus, our results suggest an early connection between two potential refugia, resulting in a genetic ancestry that survived in later Iberian HGs. Our new genomic data from Iberian Early and Middle Neolithic individuals show that the dual Iberian HG genomic legacy pertains in the peninsula, suggesting that expanding farmers mixed with local HGs.

RESULTS AND DISCUSSION

We successfully generated autosomal genome-wide data and mitochondrial genomes of 10 new individuals from key sites in the Iberian Peninsula ranging from ~13,000–6,000 calibrated years before present (years cal BP): Late Upper Paleolithic (n = 2), Mesolithic (n = 1), Early Neolithic (EN; n = 4), and Middle Neolithic (n = 3) (Figure 1, Data S1A, and STAR Methods). We furthermore improved the sequencing depth of one Upper Paleolithic individual from Troisième caverne of Goyet (Belgium) dated to ~15,000 years cal BP and associated with the Magdalenian culture [2]. For each individual, we generated multiple DNA libraries with unique index pairs [10, 11] for next-generation

Figure 1. Geo-chronological Location of Ancient Individuals from the Iberian Peninsula

(A) Map showing the geographical location of the new individuals and sites included in this study (black outlines) and relevant published data for HGs and EN/MN individuals from the Iberian Peninsula (no outlines).

(B) Radiocarbon dates of newly reported individuals in years cal BP (error bars indicate the 2-sigma range).

See also Data S1.

sequencing from ancient DNA (aDNA) extracted from teeth and bones [12] (Data S1B and STAR Methods). These were subsequently enriched using targeted in-solution capture for ~1240K informative nuclear SNPs [13], and independently for the complete mitogenome [14], and sequenced on Illumina platforms. All libraries contained short DNA fragments (46–65 bp length on average) with post mortem deamination patterns characteristic for aDNA (4%–16% for partial uracil-DNA glycosylase [UDG] and 19%–31% for non-UDG treated libraries at the first base pair position; Data S1C and S1D and STAR Methods). We estimated contamination rates from nuclear DNA in males to be 1.0%–3.4% for final merged libraries using ANGSD method 2 [15] and for mitogenomes in both sexes to be 0.14%–2.20% using ContamMix [16] (Data S1E and S1F and STAR Methods). After quality filtering, we obtained an endogenous DNA content of 1.3%–29.5% on the targeted 1240K SNPs (Data S1C and STAR Methods). We called pseudo-haploid genotypes for each individual (merged libraries) by randomly choosing a single base per site and intersected our data with a set of global modern populations genotyped for ~1240K nuclear SNP positions [17], including published ancient individuals from [2, 5, 7–9, 13, 14, 18–25]. The final dataset from the newly reported individuals contained 19,269–814,072 covered SNPs (Data S1G and STAR Methods). For principal-component analysis (PCA), we intersected our new data and published ancient individuals with a panel of worldwide modern populations genotyped on the Affymetrix Human Origins (HO) array [26].

Genetic Structure in Iberian Hunter-Gatherers

To characterize the genetic differentiation between HG individuals, we calculated genetic distances, defined as 1 – *f3,* where *f3* denotes the *f3*-outgroup statistics [26], for pairwise comparisons among all published and newly generated HG and visualized the results using multidimensional scaling (MDS) (STAR Methods, Figure 2A and STAR Methods). The hunter-gatherer (HG) individuals form distinguishable clusters on the MDS plot, supported by *f4*-statistics and clustering analysis (Figure S1 and STAR Methods), which we label in line with Fu et al. [2] as *Villabruna*, *Věstonice*, *Satsurblia*, and *Mal'ta* clusters, respectively. Henceforth, we present genetic clusters in italics and individuals in normal print. We introduce the *GoyetQ2* cluster (based on the highest genomic coverage) representing the Magdalenian-associated individuals Goyet Q-2, Hohle Fels 49, Rigney 1, El Mirón, and Burkhardtshöhle. With the newly generated data, we notice that Iberian HGs form a cline between the *GoyetQ2* and *Villabruna* clusters (Figures 2A and S2). This cline also includes El Mirón (the oldest individual from Iberia), which had previously been considered representing its own *El Mirón* cluster together with all individuals of the *GoyetQ2* cluster (yellow symbols in Figure 2A) [2]. Here, Canes 1 and La Braña 1 (Mesolithic individuals from the Cantabrian region in northern Iberia) are falling closer to the *Villabruna* cluster, while Chan (northwestern Iberia) and our newly reported individuals from Moita do Sebastião (Portuguese Atlantic coast) and Balma Guilanyà (Pre-Pyrenean region, northeastern Iberia) have more affinity with El Mirón, which is in turn closer to the Magdalenian *GoyetQ2* cluster.

This observation is confirmed by *f4*-statistics of the form *f4(GoyetQ2, Villabruna; test, Mbuti)*, which measures whether a test population shares more genetic drift with Goyet Q-2 than with the Villabruna individual. Three Iberian HGs (Chan, Moita do Sebastião, and El Mirón), as well as Hohle Fels 49 and the 35,000-year-old Goyet Q116-1, show significantly positive *f4*-values, indicating that these individuals shared more ancestry with Goyet Q-2 than with Villabruna (Figure 2B). This

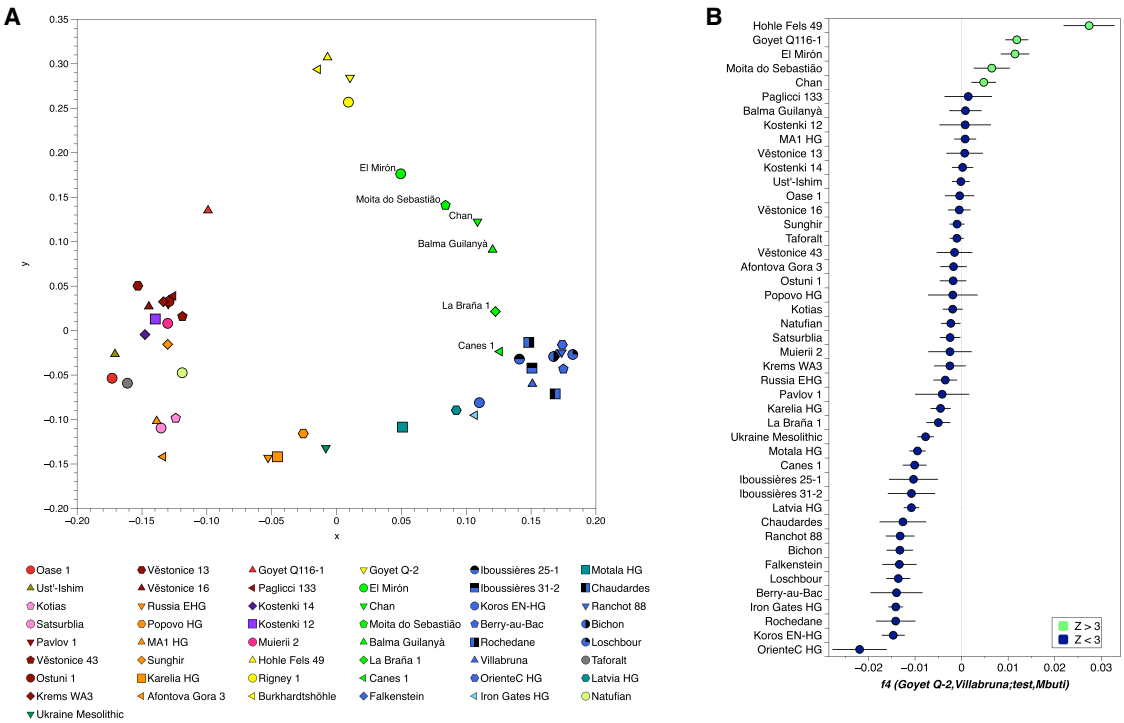


Figure 2. Genetic Distances between European HGs and Key *f4*-Statistics

(A) MDS plot of genetic distances between Eurasian HG individuals (>30,000 SNPs). The main genetic clusters defined previously [2] are: *Vĕstonice* (dark red), *Mal'ta* (orange: MA1 and Afontova Gora 3), *Satsurblia* (light pink: Kotias and Satsurblia), *Villabruna* (blue: Koros EN-HG, Berry-au-Bac 1, Rochedane, Villabruna, Chaudardes 1, Ranchot 88, La Braña 1, Loschbour), and *GoyetQ2* (yellow; newly defined). Iberian HGs are shown as green symbols.

(B) *f4*-statistics highlighting the excess affinity to Goyet Q-2 in Iberian and European HGs (>20,000 SNPs; error bars indicate ± 3 SE; $Z > 3$ [green]).

See also Figures S1 and S2.

heterogeneity in Iberian HGs cannot be explained by genetic drift alone (against which this type of F-statistics is robust) but only by admixture between two sources related to Goyet Q-2 and Villabruna, respectively. We visualize this admixture cline using contrasting *f3*-outgroup statistics of the form *f3(GoyetQ2; test, Mbuti)* and *f3(Villabruna; test, Mbuti)* (Figure 3A). The individuals from the *Villabruna* cluster deviate from the symmetry line x = y toward the y axis, expectedly, indicating excess genetic drift shared with Villabruna. In contrast, individuals of the *GoyetQ2* cluster deviate from the symmetry line x = y toward the x axis, indicating excess genetic drift with Goyet Q-2. Iberian HGs fall between the two clusters, which is inconsistent with them forming a clade with either group, but can only be explained by admixture. Here, Iberian HG Canes 1 and La Braña 1 share more Villabruna-like ancestry while El Mirón, Moita do Sebastião, and Chan share more Goyet Q-2-like ancestry.

To further confirm the potential admixture of El Mirón, we used *f4(GoyetQ2 cluster, El Mirón; Villabruna, Mbuti)* to test if Magdalenian-associated individuals were cladal with El Mirón. Here, we obtained significantly negative Z scores for Hohle Fels 49, Goyet Q-2, and Burkhardtshöhle. Among these, Goyet Q-2 has the highest data quality and most negative Z score and thus represents the best proxy for the non-*Villabruna*-like ancestry proportion in individuals such as El Mirón ($Z = -6.82$) (Data S1H). Based on this observation, we used the test *f4(Goyet Q-2, GoyetQ2 cluster; Villabruna, Mbuti)*, for which El Mirón is significantly negative (Figure 3B and Data S1H), confirming shared ancestry with Villabruna.

To show that the affinity of El Mirón with the Villabruna individual cannot be explained by El Mirón representing a basal split from Villabruna and the other *GoyetQ2* individuals, we used the test *f4(GoyetQ2 cluster, Villabruna; El Mirón, Mbuti)*. Here, all individuals of the *GoyetQ2* cluster are significantly positive, indicating that this cluster does not represent a sister branch of Villabruna and that El Mirón is not an outgroup to both the *Villabruna* and *GoyetQ2* clusters (Figure 3C and Data S1H). The mixed ancestry of El Mirón could also explain its reduced affinity to Goyet Q116-1 when compared to the younger *GoyetQ2* cluster in the test *f4(GoyetQ2 cluster, Goyet Q-2; Goyet Q116-1, Mbuti)* (Data S1H).

Having established two potential Paleolithic source populations surviving in Iberia from ~19,000 years cal BP onward, we used the admixture modeling programs *qpWave* and *qpAdm* (Figure 3D and STAR Methods) to explore the ancestry of all Iberian HGs. We used Villabruna and Goyet Q-2 as ultimate sources to model the dual ancestry in European HGs relative to outgroups that can distinguish these two sources from shared deeper ancestries (STAR Methods). Our two-source admixture model provides a good fit for the genetic profiles of most European HGs and is consistent with the cline between Villabruna- and Goyet Q-2-like ancestries described above (Figure 3D and STAR Methods). Here, Villabruna-like ancestry is the dominant component (69.8% ± 4.3% – 100%) in individuals of the *Villabruna* cluster,

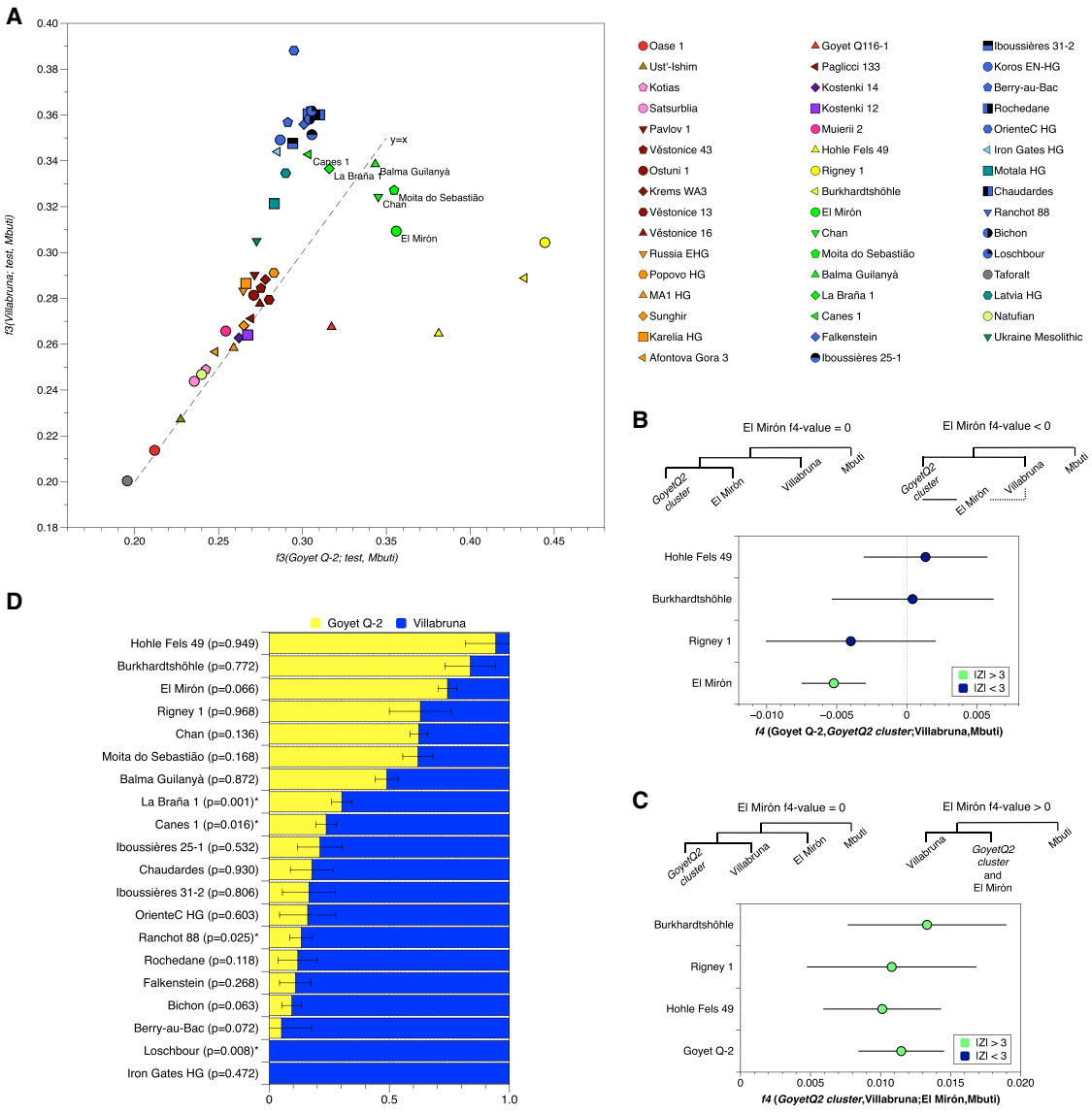


Figure 3. Key *f3*-Outgroup Tests, *f4*-Statistics, and *qpAdm* Results

(A) Biplot of *f3*-outgroup tests illustrating the Villabruna-like and Goyet Q-2-like ancestries in European HGs. The x = y axis marks full symmetry between Goyet Q-2- and Villabruna-like ancestries, and deviations mark excess ancestry shared with Goyet Q-2 (yellow) or Villabruna (blue).

(B) Results of *f4*-statistics highlighting the shared genetic drift between El Mirón and Villabruna individuals (≥20,000 SNPs; error bars indicated ±3 SE).

(C) Results of *f4*-statistics showing that El Mirón is not a sister clade of Villabruna (error bars indicated ±3 SE).

(D) Modeling European HGs as a two-way admixture of Villabruna- and Goyet Q-2-like ancestry (error bars indicated ±1 SE).

See also Figure 3B and Data S1.

and Goyet Q-2-like ancestry the dominant component (61.9% ± 6.3% – 94.3% ± 5.7%) in *GoyetQ2* cluster individuals (Figure 3D and Data S1I). These results underline the power of our outgroups and choice of proxies to differentiate Goyet Q-2- from Villabruna-like ancestry within our test individuals (STAR Methods; see Figure S3A for a replication with more proximal sources El Mirón and Loschbour). Congruent with the pattern observed in MDS (Figure 2A), clustering analysis (Figure S1), PCA (Figure S2), F-statistic-based tests (Figure 2B), and the biplot of *f3*-outgroup tests (Figure 3B), the two-source admixture model assigns a higher proportion of Goyet Q-2-like ancestry to Iberian HGs (ranging from 23.7% to 75.3%) than to contemporaneous Western HG (WHG) outside of Iberia. In fact, Balma Guilanyà, La Braña 1, and Canes 1 show elevated Villabruna admixture proportions but still higher Goyet Q-2 proportions than non-Iberian HGs (Figure 3D and Data S1I). We notice an additional contribution of Villabruna-like ancestry in the 12,000-year-old Balma Guilanyà individual from northeastern Iberia. Villabruna-like ancestry becomes even stronger during the Mesolithic in the Cantabrian region (La Braña 1 and Canes 1), suggesting extra HG flux into north/northeastern Iberia, which must have had a higher impact in this region. We were able to track a correlation between increasing Villabruna-like ancestry

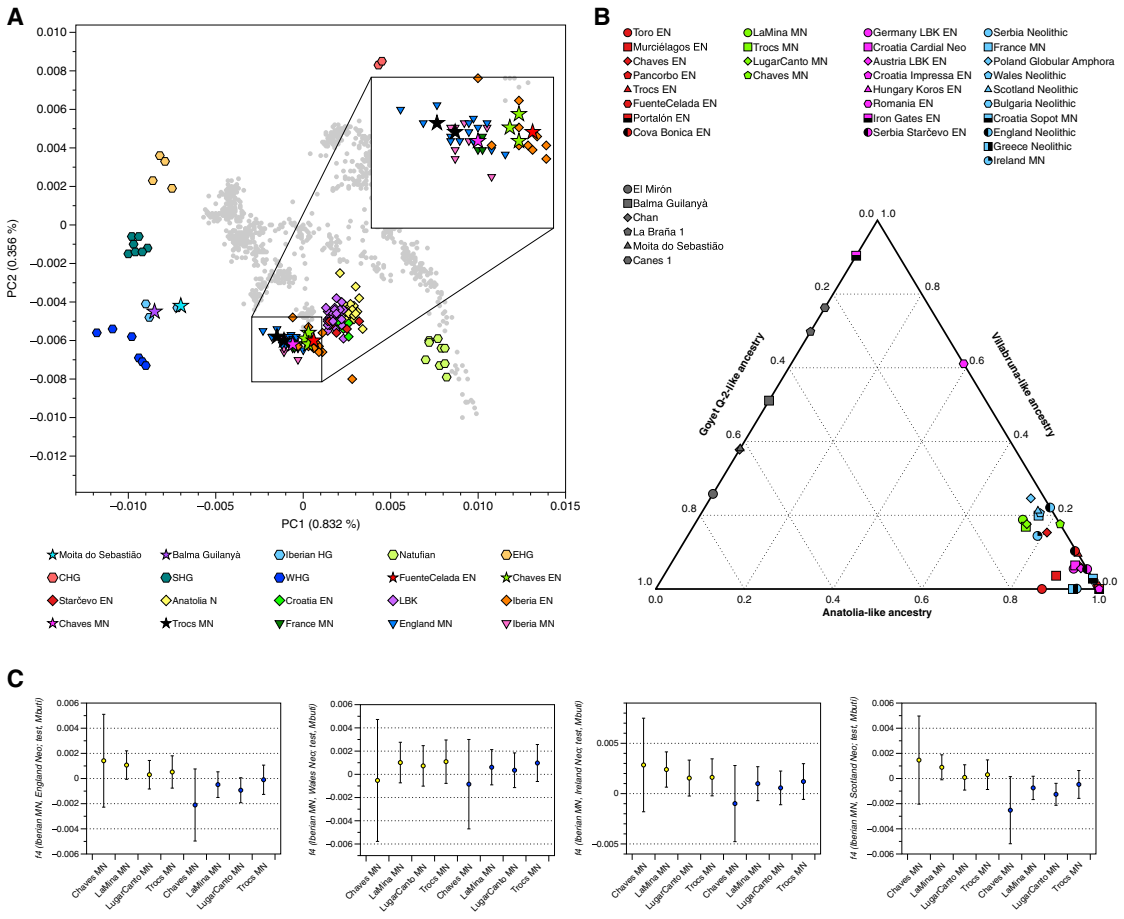


Figure 4. PCA Results and *qpAdm* Admixture Models

(A) Published ancient and newly reported individuals (stars) projected onto 777 present-day West Eurasians.

(B) Modeling EN and MN populations from Iberian and western Europe as admixture of three ancestral sources: Anatolian Neolithic, Goyet Q-2, and Villabruna (Data S1J).

(C) *f4(Iberian MN, Neolithic British Isles, test, Mbuti)*, where test is Goyet Q-2 (yellow) or Villabruna (blue), highlighting the excess of Goyet Q-2 ancestry in Iberian MN compared to Neolithic England and Scotland (error bars indicated ±3 SE). The affinity to Villabruna is shown for comparison to avoid potential bias created by unspecified HG attraction.

See also Data S1.

and time in this region (Figure S3A), while Mesolithic HGs outside this region (Chan and Moita do Sebāstiao) retain more *GoyetQ2* ancestry and do not fit this pattern (Figures S3A and S3C). Interestingly, we find no traces of African ancestry (Figure S3B and STAR Methods).

Dual Hunter-Gatherer Genetic Legacy in Iberian Neolithic Individuals

During the Neolithic transition ~7,600 years ago, human expansions reached the Iberian Peninsula relatively swiftly via expanding early farmers from western Anatolia [3–5]. The rapid expansion of EN individuals associated with farming practices across Europe resulted in a relatively low genetic variability in the reported Neolithic genomes, which makes it difficult to distinguish between the Mediterranean and Danubian routes of expansion of Neolithic lifeways [27, 28]. However, Olalde et al. [9] noted subtle regional differences between WHG individuals and used the proportion of HG ancestry from La Braña 1 in Neolithic Iberians to trace the expansion from southwestern Europe along the Atlantic coast to Britain. This movement corresponds well with the megalithic burial practices of these regions observed in the archaeological record [29, 30]. Under the assumption that these proportions reflect one, or potentially more, local admixture events along the routes of expansion, it is thus possible to distinguish Neolithic groups by their varying autochthonous HG signatures [18].

Given the presence of two ancestral lineages in Iberian HGs, we thus explored this potential genetic legacy in our newly generated EN and Middle Neolithic (MN) individuals. We first used PCA to assess the genetic affinities qualitatively (Figure 4A). Here, the new Neolithic Iberian individuals spread along a cline from Neolithic Anatolia to WHG, on which the new individuals cluster with contemporaneous Iberian Neolithic individuals [5, 7, 8, 13, 24, 25]. As shown before, MN individuals are shifted toward WHG individuals [3], including the newly reported MN individuals from Cova de Els Trocs and Cueva de Chaves.

Using *qpAdm* models consistent with those above (Figures 3D and S3), we aimed to trace and quantify the proportion of Goyet Q-2- and Villabruna-like HG ancestry in EN and MN groups from Iberia and western/central Europe as a mixture of three ancestral sources: Anatolian Neolithic, Goyet Q-2, and Villabruna, respectively. We show that EN Iberians shared a higher proportion of Goyet Q-2-like ancestry than EN individuals from outside Iberia (Figure 4B and Data S1J). Goyet Q-2-like ancestry is higher in EN from southern Iberia (Andalusia), suggesting additional admixture with local Iberian HGs, who carried mixed Upper Paleolithic ancestry.

Goyet Q-2 ancestry is continuously detectable in all Iberian MN individuals, including broadly contemporaneous individuals from Scotland, Wales, Ireland, and France, but not in Neolithic England, for which the *qpAdm* model with three sources fails ($p = 6.91e{-}05$) in favor of two sources, despite being poorly supported ($p = 0.001$) (Data S1J). Goyet Q-2-like ancestry is, however, highest in all Iberian MN (except Chaves MN) when compared with other MN populations that share a similar overall amount of HG ancestry (Figures 4B and 4C). We note the presence of Goyet Q-2 ancestry in MN Trocs, where this ancestry was not observed during the EN, but importantly also in MN individuals from France and Globular Amphora from Poland.

Olalde and colleagues [9] reported an elevated signal of La Braña 1 ancestry in Neolithic individuals from Wales and England (using KO1 HG from Hungary and Anatolian Neolithic as the other two sources) and thus argued for an Iberian contribution to the Neolithic in Britain [9]. We replicated these findings by using similar sources (El Mirón instead of Goyet Q-2; Data S1J), showing that these results are sensitive to the source populations used. However, our models with Goyet Q-2 as ultimate source highlight not only the admixed nature of La Braña 1 and El Mirón, but also that Goyet Q-2-like ancestry in MN individuals outside Iberia hints at Iberia as one possible source, but not the exclusive source, of the Neolithic in Britain. Further sampling from regions in today's France, the Netherlands, Belgium, Luxembourg, and Germany is needed to answer this question.

Conclusions

Our results highlight the unique genetic structure observed in Iberian HG individuals, which results from admixture of individuals related to the *GoyetQ2* and *Villabruna* clusters. This suggests a survival of two lineages of Late Pleistocene ancestry in Holocene western Europe, in particular the Iberian Peninsula, whereas HG ancestry in most other regions was largely replaced by Villabruna-like ancestry. With an age estimate of ~18,700 years cal BP for the El Mirón individual, the oldest representative of this mixed ancestry, the timing of this admixture suggests an early connection (*terminus ante quem*) between putative ancestries from different LGM refugia. It is possible that Goyet Q-2 ancestry could have existed in Iberia in unadmixed form, where it was complemented by Villabruna ancestry as early as ~18,700 years ago. Alternatively, both Magdalenian-associated Goyet Q-2 and Villabruna ancestries originated in regions outside Iberia and arrived in Iberia independently, where both lineages admixed, or had already existed in admixed form outside Iberia. Interestingly, the dual Upper Paleolithic ancestry was also found in EN individuals from the Iberian Peninsula, supporting the hypothesis of additional local admixture with resident HGs in Iberia during the time of the Mesolithic-Neolithic transition.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Archaeological sites and sample description
  - Ancient DNA processing and quality control
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Read processing and assessment of ancient DNA authenticity
  - Contamination tests
  - Genotyping and merging with dataset
  - Kinship relatedness and individual assessment
  - Phenotypic traits analysis
  - Mitochondrial and Y chromosome haplogroups
  - Population genetic analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found with this article online at https://doi.org/10.1016/j.cub.2019.02.006.

A video abstract is available at https://doi.org/10.1016/j.cub.2019.02.006#mmc4.

ACKNOWLEDGMENTS

We thank the members of the Archaeogenetics Department of the Max Planck Institute for the Science of Human History, especially Maïté Rivollat, Thiseas Lamnidis, Cody Parker, Rodrigo Barquera, Stephen Clayton, Aditya Kumar, and all technicians. We thank Iñigo Olalde for valuable comments on the manuscript. We are indebted to the Museo de Huesca, Patrick Semal, the Royal Belgian Institute of Natural Sciences, and all archaeologists involved in the excavations. The genetic research was funded by the Max Planck Society and the European Research Council ERC-CoG 771234 PALEoRIDER (W.H.). V.V.-M. was funded by a predoctoral scholarship of the Gobierno de Aragón and the Fondo Social Europeo (BOA20150701025) and a 3-month research stay grant (CH 76/16) by Programa CAI-Ibercaja de Estancias de Investigación. V.V.-M. and P.U. are members of the Spanish project HAR2014-59042-P (Transiciones climáticas y adaptaciones sociales en la prehistoria de la Cuenca del Ebro), and of the regional government of Aragon PPVE research group (H-07: Primeros Pobladores del Valle del Ebro). The Goyet project was funded by the Wenner-Gren Foundation (7837 to H.R.), the College of Social and Behavioral Sciences of CSUN, and the CSUN Competition for Research, Scholarship, and Creative Activity Awards.

AUTHOR CONTRIBUTIONS

V.V.-M., J.K., and W.H. conceived the study; R.M., J.M.-M., M.R.-G., D.C.S.-G., J.I.R.-G., M.K., H.R., I.C., H.A.-M., C.T.-R., I.G.-M.d.L., R.G.-P., K.W.A., and P.U. assembled samples and provided archaeological context; V.V.-M., M.S.v.d.L., and C.P. performed aDNA lab work and sequencing; V.V.-M., C.P., M.S.v.d.L., S.S., C.J., and W.H. analyzed data; and V.V.-M., C.P., M.S.v.d.L., and W.H. wrote the manuscript with input from all co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 16, 2018
Revised: January 4, 2019
Accepted: February 1, 2019
Published: March 14, 2019

REFERENCES

1. Stewart, J.R., and Stringer, C.B. (2012). Human evolution out of Africa: the role of refugia and climate change. Science *335*, 1317–1321.
2. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.
3. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.
4. Günther, T., and Jakobsson, M. (2016). Genes mirror migrations and cultures in prehistoric Europe-a population genomic perspective. Curr. Opin. Genet. Dev. *41*, 115–123.
5. Valdiosera, C., Günther, T., Vera-Rodríguez, J.C., Ureña, I., Iriarte, E., Rodríguez-Varela, R., Simões, L.G., Martínez-Sánchez, R.M., Svensson, E.M., Malmström, H., et al. (2018). Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. Proc. Natl. Acad. Sci. USA *115*, 3428–3433.
6. Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172.
7. Martiniano, R., Cassidy, L.M., Ó'Maoldúin, R., McLaughlin, R., Silva, N.M., Manco, L., Fidalgo, D., Pereira, T., Coelho, M.J., Serra, M., et al. (2017). The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. PLoS Genet. *13*, e1006852.
8. Olalde, I., Schroeder, H., Sandoval-Velasco, M., Vinner, L., Lobón, I., Ramirez, O., Civit, S., García Borja, P., Salazar-García, D.C., Talamo, S., et al. (2015). A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. Mol. Biol. Evol. *32*, 3132–3142.
9. Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. Nature *555*, 190–196.
10. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*.
11. Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. *40*, e3.
12. 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. USA *110*, 15758–15763.
13. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.
14. Mittnik, A., Wang, C.-C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R., Khartanovich, V., Moiseyev, V., Tõrv, M., et al. (2018). The genetic prehistory of the Baltic Sea region. Nat. Commun. *9*, 442.
15. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*, 356.
16. Fu, Q., Mittnik, 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.
17. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.
18. Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., et al. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature *551*, 368–372.
19. Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., et al. (2018). The genomic history of southeastern Europe. Nature *555*, 197–203.
20. Cassidy, L.M., Martiniano, R., Murphy, E.M., Teasdale, M.D., Mallory, J., Hartwell, B., and Bradley, D.G. (2016). Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. Proc. Natl. Acad. Sci. USA *113*, 368–373.
21. Sikora, M., Seguin-Orlando, A., Sousa, V.C., Albrechtsen, A., Korneliussen, T., Ko, A., Rasmussen, S., Dupanloup, I., Nigst, P.R., Bosch, M.D., et al. (2017). Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. Science *358*, 659–662.
22. van de Loosdrecht, M., Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., Nickel, B., Nagel, S., Talbi, E.H., El Hajraoui, M.A., et al. (2018). Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. Science *360*, 548–552.
23. Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat. Commun. *6*, 8912.
24. González-Fortes, G., Jones, E.R., Lightfoot, E., Bonsall, C., Lazar, C., Grandal-d'Anglade, A., Garralda, M.D., Drak, L., Siska, V., Simalcsik, A., et al. (2017). Paleogenomic Evidence for Multi-generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin. Curr. Biol. *27*, 1801–1810.
25. Fregel, R., Méndez, F.L., Bokbot, Y., Martín-Socas, D., Camalich-Massieu, M.D., Santana, J., Morales, J., Ávila-Arcos, M.C., Underhill, P.A., Shapiro, B., et al. (2018). Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. Proc. Natl. Acad. Sci. USA *115*, 6774–6779.
26. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.
27. Manning, K., Timpson, A., Colledge, S., Crema, E., Edinborough, K., Kerig, T., and Shennan, S. (2014). The chronology of culture: a comparative assessment of European Neolithic dating approaches. Antiquity *88*, 1065–1080.
28. Perrin, T., Manen, C., Valdeyron, N., and Guilaine, J. (2018). Beyond the sea… The Neolithic transition in the southwest of France. Quat. Int. *470*, 318–332.
29. Sherratt, A. (1995). Instruments of conversion? The role of megaliths in the mesolithic/neolithic transition in Northwest Europe. Oxf. J. Archaeol. *14*, 245–260.
30. Masset, C. (1993). Les dolmens: Sociétés néolithiques, practiques funéraires (Paris: Ed. Errance).
31. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

32. Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biol. *17*, 60.

33. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.

34. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. *44* (W1), W58-63.

35. Poznik, G.D. (2016). Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. bioRxiv. https://doi.org/10.1101/088716.

36. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

37. Monroy Kuhn, J.M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. PLoS ONE *13*, e0195491.

38. Terradas, X., Pallarés, M., Mora, R., and Moreno, J.M. (1993). Estudi preliminar de les ocupacions humanes de la balma de Guilanyà (Navès, Solsonès). Rev. d'Arqueologia Ponent, 231–248.

39. Martínez-Moreno, J., Mora, R., and Casanova, J. (2006). Balma Guilanyà y la ocupación de la vertiente sur del Prepirineo del Noreste de la Península Ibérica durante el Tardiglaciar. In La cuenca mediterránea durante el paleolítico superior: 38.000-10.000 años, pp. 444–457.

40. Martínez-Moreno, J., Mora, R., and Casanova, J. (2007). El contexto cronométrico y tecno-tipológico durante el Tardiglaciar y Postglaciar de la viertiente sur de los Pirineos orientales. Rev. d'Arqueologia Ponent, 7–44.

41. Martínez-Moreno, J., and Mora, R. (2009). Balma Guilanyà (Prepirineo de Lleida) y el Aziliense en el noreste de la Península Ibérica. Trabajos de Prehistoria *66*, 45–60.

42. Garcia-Guixé, E., Martínez-Moreno, J., Mora, R., Núñez, M., and Richards, M.P. (2009). Stable isotope analysis of human and animal remains from the Late Upper Palaeolithic site of Balma Guilanyà, southeastern Pre-Pyrenees, Spain. J. Archaeol. Sci. *36*, 1018–1026.

43. Ruiz, J., García-Sívoli, C., Martínez-Moreno, J., and Subirá, M.E. (2006). Los restos humanos del Tardiglaciar de Balma Guilanyà. In La cuenca mediterránea durante el paleolítico superior: 38.000-10.000 años, pp. 458–467.

44. Martzluff, M., Martínez-Moreno, J., Guilaine, J., Mora, R., and Casanova, J. (2012). Transformaciones culturales y cambios climáticos en los Pirineos catalanes entre el Tardiglaciar y Holoceno antiguo: Aziliense y Sauveterriense en Balma de la Margineda y Balma Guilanyà. Cuaternario y Geomorfología *26*, 61–78.

45. Straus, L.G. (2015). Chronostratigraphy of the Pleistocene/Holocene boundary: the Azilian problem in the Franco-Cantabrian region. Palaeohistoria *27*, 89–122.

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

47. Lubell, D., Jackes, M., Schwarcz, H., Knyf, M., and Meiklejohn, C. (1994). The Mesolithic-Neolithic transition in Portugal: isotopic and dental evidence of diet. J. Archaeol. Sci. *21*, 201–216.

48. Jackes, M., and Alvim, P. (1999). Reconstructing Moita do Sebastião, the first step. In Do Epipaleolítico ao Calcolítico na Península Iberica, Actas do IV Congresso de Arqueologia Peninsular, pp. 13–25.

49. Bicho, N.F. (1994). The End of the Paleolithic and the Mesolithic in Portugal. Curr. Anthropol. *35*, 664–674.

50. Gronenborn, D. (2017). Migrations before the Neolithic? The Late Mesolithic blade-and-trapeze horizon in Central Europe and beyond. Migration and Integration from Prehistory to the Middle Ages (Halle, Germany: LandesMuseum fur Vorgechichte), pp. 113–122.

51. Perrin, T., Marchand, G., Allard, P., Binder, D., Collina, C., Garcia Puchol, O., and Valdeyron, N. (2009). Le second Mesolithique d'Europe occidentale : origines et gradient chronologique. Annales de la Fondation Fyssen *24*, 160–176.

52. Utrilla Miranda, P., and Laborda Lorente, R. (2018). La cueva de Chaves (Bastaras, Huesca): 15 000 años de ocupación prehistórica. Trabajos de Prehistoria *75*, 248–269.

53. Castaños, P.M. (2004). Estudio arqueozoológico de los macromamíferos del Neolítico de la Cueva de Chaves (Huesca: Saldvie Estud. Prehist. y Arqueol), pp. 125–172.

54. Baldellou Martínez, V. (2011). La Cueva de Chaves (Bastarás-Casbas, Huesca). SAGVNTVM *12*, 141–144.

55. Utrilla, P., and Baldellou, V. (2001). Cantos pintados neolíticos de la Cueva de Chaves (Bastarás, Huesca: Saldvie Estud. Prehist. y Arqueol), pp. 45–126.

56. Utrilla, P., and Baldellou, V. (2007). Les galets peints de la Grotte de Chaves. Bull. la Société Préhistorique Ariège-Pyrénées *62*, 73–88.

57. Zilhão, J. (2001). Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. Proc. Natl. Acad. Sci. USA *98*, 14180–14185.

58. Isern, N., Zilhão, J., Fort, J., and Ammerman, A.J. (2017). Modeling the role of voyaging in the coastal spread of the Early Neolithic in the West Mediterranean. Proc. Natl. Acad. Sci. USA *114*, 897–902.

59. Bernabeu, J., Balaguer, L.M., Esquembre-Bebiá, M.A., Pérez, J.R.O., and Soler, J.d.D.B. (2009). La cerámica impresa mediterránea en el origen del Neolítico de la península Ibérica. In De Méditerranée et d'ailleurs...: mélanges offerts à Jean Guilaine, pp. 83–96.

60. Martins, H., Oms, F.X., Pereira, L., Pike, A.W.G., Rowsell, K., and Zilhão, J. (2015). Radiocarbon dating the beginning of the Neolithic in Iberia: new results, new problems. J. Mediterr. Archaeol. *28*, 105–131.

61. Villalba-Mouco, V., Utrilla, P., Laborda, R., Lorenzo, J.I., Martínez-Labarga, C., and Salazar-García, D.C. (2018). Reconstruction of human subsistence and husbandry strategies from the Iberian Early Neolithic: A stable isotope approach. Am. J. Phys. Anthropol. *167*, 257–271.

62. Cuenca-Romero, M.d.C.A., Carmona Ballestero, E., Pascual Blanco, S., Martínez, Díez, G., and Díez Pastor, C. (2011). El "campo de hoyos" calcolítico de Fuente Celada (Burgos): datos preliminares y perspectivas. Complutum *22*, 47–69.

63. Rojo-Guerra, M., Peña-Chocarro, L., Royo-Guillén, J.I., Tejedor, C., García-Martínez de Lagrán, I., Arcusa, H., Garrido Pena, R., Moreno, M., Mazzucco, N., Gibaja, J.F., et al. (2013). Pastores trashumantes del Neolítico Antiguo en un entorno de alta montaña: secuencia crono-cultural de la Cova de Els Trocs (San Feliú de Veri, Huesca: BSAA Arqueol), pp. 9–55.

64. Posth, C., Nägele, K., Colleran, H., Valentin, F., Bedford, S., Kami, K.W., Shing, R., Buckley, H., Kinaston, R., Walworth, M., et al. (2018). Language continuity despite population replacement in Remote Oceania. Nat. Ecol. Evol. *2*, 731–740.

65. 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. B Biol. Sci. *370*.

66. Maricic, T., Whitten, M., and Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS ONE *5*, e14004.

67. Fu, Q., Hajdinjak, M., Moldovan, O.T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. Nature *524*, 216–219.

68. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.

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

70. Gamba, C., Fernández, E., Tirado, M., Deguilloux, M.F., Pemonge, M.H., Utrilla, P., Edo, M., Molist, M., Rasteiro, R., Chikhi, L., and Arroyo-Pardo, E. (2012). Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. Mol. Ecol. *21*, 45–56.

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

72. Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-Del-Molino, D., van Dorp, L., López, S., Kousathanas, A., Link, V., et al. (2016). Early farmers from across Europe directly descended from Neolithic Aegeans. Proc. Natl. Acad. Sci. USA *113*, 6886–6891.

73. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.

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

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Ancient individual	This study/ Troisième caverne of Goyet archaeological site	Goyet Q-2
Ancient individual	This study/ Balma Guilanyà archaeological site	BAL001/ E1206 Shown to be identical with BAL005
Ancient individual	This study/ Balma Guilanyà archaeological site	BAL005/ BG E 3214 Shown to be identical with BAL001
Ancient individual	This study/ Balma Guilanyà archaeological site	BAL003/ E9605
Ancient individual	This study/ Moita do Sebastião archaeological site	CMS001/ 22
Ancient individual	This study/ Cueva de Chaves archaeological site	CHA001/ 84C
Ancient individual	This study/ Cueva de Chaves archaeological site	CHA002/ CH.NIG.11559
Ancient individual	This study/ Cueva de Chaves archaeological site	CHA003/ CH.NIG.11558
Ancient individual	This study/ Cueva de Chaves archaeological site	CHA004/ Ch.Banda13
Ancient individual	This study/ Fuente Celada archaeological site	FUC003/ H62 UE 622
Ancient individual	This study/ Cova de Els Trocs archaeological site	ELT002/ UE 69 C: 589 S:7 No Inv: 14227
Ancient individual	This study/ Cova de Els Trocs archaeological site	ELT006/ UE:1 C: 650 S:1 No Inv: 22404
Chemicals, Peptides, and Recombinant Proteins		
0.5 M EDTA pH 8.0	Life Technologies	AM9261
1x Tris-EDTA pH 8.0	AppliChem	A8569,0500
Proteinase K	Sigma-Aldrich	P2308-100MG
Guanidine hydrochloride	Sigma-Aldrich	G3272-500 g
3M Sodium Acetate pH 5,2	Sigma-Aldrich	S7899-500ML
Tween 20	Sigma-Aldrich	P9416-50ML
Water Chromasolv Plus	Sigma-Aldrich	34877-2.5L
Ethanol	Merck	1009832511
Isopropanol	Merck	1070222511
Buffer Tango	Life Technologies	BY5
T4 DNA Polymerase	New England Biosciences	M0203 L
T4 Polynucleotide Kinase	New England Biosciences	M0201 L
User Enzyme	New England Biosciences	M5505 L
Uracil Glycosylase inhibitor (UGI)	New England Biosciences	M0281 S
Bst 2.0 DNA Polymerase	New England Biosciences	M0537 S
BSA 20mg/mL	New England Biosciences	B9000 S
ATP	New England Biosciences	P0756 S
dNTPs 25 mM	Thermo Scientific	R1121
D1000 ScreenTapes	Agilent Technologies	5067-5582
D1000 Reagents	Agilent Technologies	5067-5583
Pfu Turbo Cx Hotstart DNA Polymerase	Agilent Technologies	600412

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Herculase II Fusion DNA Polymerase	Agilent Technologies	600679
1x TE-Puffer pH 8,0 low EDTA	AppliChem	A8569,0500
Sodiumhydroxide Pellets	Fisher Scientific	10306200
Sera-Mag Speed CM	GE Healthcare Lifescience	65152105050250
Dynabeads MyOne Streptavidin T1	Life Technologies	65601
GeneRuler Ultra Low Range DNA Ladder	Life Technologies	SM1211
10x GeneAmp PCR Gold Buffer and MgCl2	Life Technologies	4379874
Human Cot-1 DNA	Life Technologies	15279011
1M Tris-HCl pH 8.0	Life Technologies	15568025
20x SCC Buffer	Life Technologies	AM9770
UltraPure™ Salmon Sperm DNA Solution	Life Technologies	15632011
PEG 8000 Powder, Molecular Biology Grade	Promega	V3011
20% SDS Solution	Serva	39575.01
3M Sodium Acetate buffer solution pH 5,2	Sigma-Aldrich	S7899-500ML
5 M Sodium chloride solution	Sigma-Aldrich	S5150-1L
Denhardt's solution	Sigma-Aldrich	D9905-5Ml
Critical Commercial Assays		
High Pure Viral Nucleic Acid Large Volume Kit	Roche	5114403001
Quick Ligation Kit	New England Biosciences	M2200 L
MinElute PCR Purification Kit	QIAGEN	28006
DyNAmo Flash SYBR Green qPCR Kit	Life Technologies	F-415L
Oligo aCGH/Chip-on-Chip Hybridization Kit	Agilent Technologies	5188-5220
HighSeq 4000 SBS Kit	Ilumina	FC-410-1001/2
NextSeq 500/550 High Output Kit v2	Ilumina	FC-404-2002
Deposited Data		
Raw and analyzed data (European nucleotide archive)	This paper	ENA: PRJEB30985
Software and Algorithms		
Samtools	[31]	http://samtools.sourceforge.net/
EAGER	[32]	https://eager.readthedocs.io/en/latest/
ADMIXTOOLS	[25]	https://github.com/DReichLab/AdmixTools
smartpca	[33]	https://www.hsph.harvard.edu/alkes-price/software/
ANGSD	[15]	http://www.popgen.dk/angsd/index.php/Main_Page Contamination
Haplogrep 2	[34]	http://haplogrep.uibk.ac.at/
ContamMix	[16]	https://github.com/StanfordBioinformatics/DEFUNCT-env-modules/tree/master/contamMix
Yhaplo	[35]	https://github.com/23andMe/yhaplo
Geneious R8.1.974	[36]	https://www.geneious.com
READ	[37]	https://bitbucket.org/tguenther/read

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Wolfgang Haak (haak@shh.mpg.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The Iberian Peninsula in southwestern Europe is understood as a periglacial refugium for Pleistocene hunter-gatherers (HG) during the Last Glacial Maximum (LGM). The post-LGM genetic signature in western and central Europe was dominated by ancestry similar to the Villabruna individual, commonly described as WHG ancestry or '*Villabruna*' cluster [2]. This *Villabruna* cluster had largely

replaced the earlier *El Mirón* genetic cluster, comprised of 19,000-15,000-year-old individuals from central and western Europe associated with the Magdalenian culture [2].

By generating new genome-wide data from Belgian HG, Iberian HG and Neolithic individuals we aimed to further refine the HG genetic structure in the Iberian Peninsula during the Upper Paleolithic and Mesolithic, and to characterize the HG ancestry sources that contributed genetically to Neolithic groups. We hypothesize that (i) admixture events of different Upper Paleolithic HG ancestries resulted in genetic structure among various Iberian HG groups, which can be observed as asymmetric genetic affinities to the *Villabruna* and *El Mirón* cluster or another potential source, respectively, (ii) this structure during the Early Holocene is stronger in the Iberian Peninsula than in Central Europe (iii) the genetic structure was inherited by Neolithic individuals through mixture with local HG ancestry, especially in regions that initially were more affected by expanding farmers. All teeth and bone samples analyzed were obtained with relevant institutional permissions from the Gobierno de Aragón, Universitat Autònoma de Barcelona, Universidad de Valladolid, the German Archeological Institute Madrid.

Archaeological sites and sample description

Troisième caverne of Goyet (Upper Paleolithic)

The Troisième caverne of Goyet (Belgium) is a cave with an extensive Paleolithic record, from the Middle to the Upper Paleolithic periods (Aurignacian, Gravettian, and Magdalenian). The site was previously described and eight individuals were analyzed by Fu et al. [2]. For this study we have generated deeper sequencing data from individual Goyet Q-2 who is attributed to the Magdalenian period.

Goyet Q-2, juvenile individual (12,650 ± 50 BP [GrA-46168], 15,232–14,778 years cal BP [2-sigma value]) [2].

Balma Guilanyà (Late Upper Paleolithic)

Balma Guilanyà is a rock shelter located in Northeastern Iberia, at 1,150 m.a.s.l. (meters above sea level) in the Serra de Busa Pre-Pyrenean range (Navés, Lleida). After an initial test pit where Late Upper Paleolithic remains were recovered [38], the site was excavated between 2001 to 2008 [39, 40]. Two main chrono-cultural phases were defined. The oldest dates back to the Late Upper Paleolithic (15,000–11,000 years cal BP) and the youngest corresponds to the Early Mesolithic (11,000 – 9,500 years cal BP). The two chrono-cultural units are separated by a big fallen boulder which sealed the Late Upper Paleolithic levels [41]. A set of human skeletal remains were found under this big stone block without any evidence of funerary structures. Direct radiocarbon dates from two human remains (one human tooth and one human bone fragment) recovered from the same context dated to 13,380–12,660 years cal BP (Ua-34297) and 12,830–10,990 years cal BP (Ua- 34298) [42]. These dates fall inside the Bølling/Allerød interstadial and Younger Dryas stadial, which correspond to the Late Glacial. The Minimum Number of Individuals (MNI) was estimated to be three based on dental morphology: two adults and one immature individual [43]. The stable carbon and nitrogen isotope analyses performed on human bone collagen suggested a diet based on terrestrial herbivores, without any evidence of marine or freshwater resources [42]. The material cultural artifacts recovered from the same level as the human remains have been attributed to the Azilian techno-complex [41]. Balma Guilanyà shows clear technical parallels with the near Azilian site Balma Margineda [44]. However, in general the Azilian is considered to be more common in Vasco-Cantabrian northern Iberia and on the other side of the Pyrenees [45]. Here, we report the genome-wide data from two individuals from this site:

BAL0051, adult individual
BAL003, adult individual

Moita do Sebastião (Mesolithic)

This site was previously described in Szécsényi-Nagy et al. [46]. Moita do Sebastião is a Late Mesolithic shell midden site located in the Muge region (Salvaterra de Magos, Portugal) on the Atlantic coastline of Portugal. The Muge and Sado regions were very fertile estuaries and marshes during the Mesolithic, which were exploited by hunter-gatherers to obtain marine resources [47]. Although Mesolithic groups are not considered fully sedentary, Moita do Sebastião presents some cultural characteristics that suggest permanence at the site: the presence of post holes associated with hut building, and a big burial space [48]. These features have been interpreted as a systematic occupation of the estuarine areas, which is also reflected in the shell midden conformation. The Moita do Sebastião site was excavated by different archaeologists since the last century. The total minimum number of individuals (MNI) is unknown, but it could reach up to 100 individuals when summarizing the different campaigns [48]. The lithic assemblage of the Mesolithic phase is characterized by microburin technique and geometrics [49]. The Mesolithic geometric phenomenon is spread widely along eastern and western Europe and North Africa, but its origin is still debated [50]. Based on a chronological gradient an African origin was suggested, from where it spread into Europe through the South of Italy (Sicily) and from where it followed a Mediterranean expansion to Iberia [51]. In this study, we genetically analyze one Mesolithic individual from this site:

CMS001, adult individual, (7,240 ± 70 BP [To-131], 8,185–7,941 years cal BP [2-sigma value]) [46].

Cueva de Chaves (Early Neolithic)

Cueva de Chaves is located in Northeastern Iberia, at 663 m.a.s.l. in the Pre-Pyrenean mountain range of Sierra de Guara (Bastarás, Huesca). The site was excavated under the direction of Pilar Utrilla and Vicente Baldellou in between 1984 and 2007. Cueva de

Chaves was occupied during the Paleolithic, Neolithic, and sporadically during the Bronze Age and Late Roman periods. Neolithic deposits were divided in two archaeological levels and dated to the Early Neolithic period (Ia: 5,600–5,300 years cal BCE; Ib 5,300-5,000 years cal BCE) [52]. Both levels show a full Neolithic package consisting of domestic fauna [53], Cardial pottery [54] and schematic rock art painted on pebbles [55, 56]. The earliest Neolithic sites in the Iberian Peninsula are located in coastal areas [57, 58]. A long-standing hypothesis in archaeology to explain this is that the first arrival of the Neolithic in the Iberian Peninsula resulted from a Cardial expansion by a maritime route. In this context, Cueva de Chaves represents an interesting case study, because radiocarbon dates for the occupation of this cave overlap in time with other Cardial Early Neolithic sites in coastal Iberia [59, 60]. The pottery style, together with the radiocarbon dates, suggest an early expansion of the first farmers from coastal to the inland areas following the Ebro Basin [52]. An MNI of four individuals, directly radiocarbon dated, were recovered from this Early Neolithic context (although one radiocarbon date points back to the early Middle Neolithic). One of individuals was in a complete anatomical articulation. A human isotopic dietary study shows a high animal protein intake consumed by all individuals [61]. This was related to the existence of a specialized animal husbandry management community in which agriculture was not intensively developed. We included four Neolithic individuals for genetic analyses in this study:

CHA001, adult (6,230 ± 45 BP [GrA-26912], 7,257–7,006 years cal BP [2-sigma value]) [54].
CHA002, adult (6,227 ± 28 BP [MAMS 29127], 7,250–7,018 years cal BP [2-sigma value]) [61].
CHA003, infant (6,180 ± 54 BP [D-AMS 015821], 7,245–6,947 years cal BP [2-sigma value]) [61].
CHA004, adult (5,645 ± 31 BP [MAMS 28128], 6,494–6,321 years cal BP [2-sigma value]) [61].

Fuente Celada (Early Neolithic)

This site was described in Szécsényi-Nagy et al. [46]. Fuente Celada is an open-air settlement located in the northern Iberian Central Plateau (Quintanaduenas, Burgos). All the archaeological materials are from a rescue excavation carried out in 2008 by Alameda Cuenca-Romero et al. [62]. The site presents many negative structures, most of them from the Chalcolithic period, suggesting a habitat settlement. Some of these negative structures contain Chalcolithic humans remains. One of these burials gave an older date corresponding to the Early Neolithic. This burial contained an individual in a flexed position, with three bone rings close to the cervical vertebrae, which was interpreted as a necklace [62]. Here we include this Early Neolithic individual in our genetic analyses:

FUC003, adult (6,120 ± 30 BP [UGA-7565], 7,157–6,910 years cal BP [2-sigma value]) [62].

Cova de Els Trocs (Middle Neolithic)

This site was also described in Szécsényi-Nagy et al. [46]. Cova de Els Trocs is a cave located in Northeastern Iberia, at 1,564 m.a.s.l. in San Feliú de Veri (Bisaurri, Huesca) in the South of the Axial Pyrenees [63]. The excavation of the site is ongoing and lead by Manuel Rojo Guerra and José Ignacio Royo Guillén. Within the large stratigraphic sequence three different occupation phases are discerned that are supported by more than twenty radiocarbon dates [63]. The first phase corresponds to the Early Neolithic (ca. 5,300–4,800 years cal BCE.) for which genomic data has been published in Haak et al. [3] for seven individuals. The second phase dates back to the Middle Neolithic (ca. 4,500–4,300 years cal BCE.) when the cave was possibly used by animals and no human remains have been retrieved. During the third phase (ca. 4,000–3,700; 3,350–2,900 years cal BCE) the cave was again used as a burial place despite not being the only purpose. From the third phase, we have included two individuals in the present genomic study:

ELT002, adult (5,008 ± 23 BP [MAMS-16160], 5,882–5,658 years cal BP [2-sigma value]) [63].
ELT006, adult (5,035 ± 23 BP [MAMS-16165], 5,895–5,716 years cal BP [2-sigma value]) [63].

Ancient DNA processing and quality control

Sampling of ancient human remains

For the ancient individuals analyzed in this study, we sampled various bones (a humerus, phalanges, metacarpals, mandibles and a cranial fragment) and teeth (molars) in the clean room of the Max Planck Institute for the Science of Human History (MPI-SHH) in Jena, Germany, and at the Institute of Anthropology, Johannes Gutenberg University, Mainz (Data S1B). Prior to sampling, samples were irradiated with UV-light for 30 min at all sides. Different sampling methods were used for different bones types, including sandblasting, grinding with mortar and pestle, and cutting and drilling in the denser regions (Data S1B). Teeth surfaces were cleaned with a low concentration bleach solution (3%). For the teeth sampled at the MPI-SHH, the crown was separated from the root by cutting with a hand saw along the cementum/enamel junction followed by drilling inside the pulp chamber [64]. For the teeth sampled in Mainz the complete tooth was ground using a mixer mill [46].

DNA extraction

DNA extraction was done following a modified version of the Dabney protocol [12], with an initial amount of 50-100 mg of bone or tooth powder. Samples were digested with extraction buffer (EDTA, UV H_2O and Proteinase K) during 16-24h in a rotator at 37°C. The suspension was centrifuged and the supernatant transferred into binding buffer (GuHCl, UV H_2O and Isopropanol) and then into silica columns (High Pure Viral Nucleic Acid Kit; Roche). The columns were first washed with wash buffer (High Pure Viral Nucleic Acid Kit; Roche) and then eluted in 100 μL TET (TE-buffer with 0.05% Tween). We included one or two extraction blanks in each extraction series to check for cross-contamination between samples and background contamination from the lab.

Library preparation

A total of 27 double-stranded (ds) libraries were created from 25 μL DNA template extract at the MPI-SHH, following a protocol by Meyer & Kircher [10] with unique index pairs [11]. We used a partial Uracil DNA Glycosylase treatment (UDG-half) that repairs damaged nucleotides by removing deaminated cytosines except for the final nucleotides at the 5′ and 3′ read ends to retain a damage pattern characteristic for ancient DNA [65]. The libraries generated from Goyet Q-2 were ds-non-UDG and ss-UDG-half treated. We repaired the terminal ends of the DNA fragments using T4 DNA Polymerase (NBE) and joined the Illumina adaptors using the Quick Ligation Kit (NBE). We also added one or two library blanks per batch. One aliquot of each library was used to quantify the DNA copy number with IS7/IS8 primers [10] outside the clean room using DyNAmo SYBP Green qPCR Kit (Thermo Fisher Scientific) on the LightCycler 480 (Roche). Libraries were double indexed with unique index combinations [11] before doing PCR amplifications outside the cleanroom with PfuTurbo DNA Polymerase (Agilent). After amplification, the indexed products were purified with MinElute columns (QIAGEN) and eluted in 50 μL TET buffer and quantified with IS5/IS6 primers using the DyNAmo SYBP Green qPCR Kit (Thermo Fisher Scientific) on the LightCycler 480 (Roche) [10]. We used Herculase II Fusion DNA Polymerase (Agilent) with the same IS5/IS6 primers for the further amplification of the indexed products up to a copy number of 10 e-13 molecules/μL. After another purification round, we quantified the indexed libraries on a TapeStation (TapeStation Nucleic Acid System, Agilent 4200) and made a 10nM equimolar pool. Data S1B shows an overview of the extracts and libraries generated for each ancient individual.

Shotgun screening and in-solution enrichment of nuclear DNA (1240k capture) and mtDNA (mitocapture)

The pooled double indexed libraries were sequenced on an Illumina HiSeq2500 for a depth of ~5 million read cycles, using either a single (1x75bp reads) or double end (2x50bp reads) configuration. Reads were analyzed with EAGER 1.92.32 [32] to check the quality and quantity of endogenous human DNA in each library. We selected samples for targeted in-solution capture enrichment that showed a damage pattern characteristic for ancient DNA and with > 0.2% endogenous DNA. We further amplified these libraries with the IS5/IS6 primer set to a concentration of 200-400 ng/μL. After that, the libraries were hybridized in-solution to different oligonucleotide probe sets synthetized by Agilent Technologies to enrich for the complete mitogenome (mtDNA capture [66]) and for 1,196,358 informative nuclear SNP markers (1240K capture [67]).

QUANTIFICATION AND STATISTICAL ANALYSIS

Read processing and assessment of ancient DNA authenticity

We demultiplexed the sequenced libraries according to expected read indexes, allowing for one mismatch. We clipped adapters with AdapterRemoval v2.2.0 [68]. For paired end reads, we restricted to merged fragments with an overlap of at least 30 bp. Single end reads shorter than 30 bp were discarded. We mapped fragments to the Human Reference Genome Hs37d5 using the Burrows-Wheeler Aligner (BWA, v0.7.12-r1039) *aln* and *samse* commands (-l 16500, -n0.01, -q30) [69] and removed duplicate reads using DeDup v0.12.1. We excluded reads with a mapping quality phred score < 30. A summary of quality statistics is given for 1240K SNP captured libraries in Data S1C and for mtDNA captured libraries in Data S1D.

Contamination tests

Prior to genotype calling we assessed the level of contamination in the mitochondrial and nuclear genome using several methods.

DNA damage

We determined and plotted the deamination rate pattern in our UDG-half libraries using MapDamage v.2.0.6 from EAGER 1.92.32 [32]. Although damage rates at the terminal read ends vary (5.3%–14.8%) in libraries for individuals from different sites, all libraries show deamination patterns expected for ancient DNA (Data S1C). Then we trimmed the reads for 2 bp at both terminal ends of the UDG-half libraries to reduce the bias of deamination from our genotype calls. Non-UDG libraries generated from Goyet Q-2 were trimmed for 10 bp.

Contamination based on the match rate to the mtDNA dataset (ContamMix)

We used ContamMix 1.0.10 to estimate the mitochondrial contamination levels in our mito-captured libraries taking a worldwide mitochondrial dataset to compare as a potential contamination source [16] (Data S1F). We find contamination rates below 2.2% for all libraries (Data S1F). We visualized the mitochondrial read alignment with Geneious R8.1.974 [36] and manually checked for heterozygous calls to confirm the ContamMix estimates. We found a substantial mitogenome heterozygosity level for BAL003_MT in the manual check, contrasting its respective ContamMix estimate of 2.2%, and therefore excluded this library from further genome analyses.

Sex determination and X-contamination

We determined genetic sex by calculating the X-ratio (targeted X-Chromosome SNPs/ targeted autosomal SNPs) and Y-ratio (targeted Y-Chromosome SNPs/ targeted autosomal SNPs) (Data S1F). For uncontaminated libraries, we expect an X ratio ~1 and Y ratio ~0 for females and X and Y ratio of 0.5 in males [2]. Potential individuals that fall in an intermediate position could indicate the presence of DNA contamination.

Method 2 of the ANGSD package was used on merged and unmerged libraries from male individuals to test the heterozygosity of polymorphic sites on the X chromosome [15]. For low coverage libraries from the same individual with < 200 SNPs on the X chromosome we merged them into a single BAM file using samtools v0.1.19 [31] to facilitate contamination estimation of the merged libraries (Data S1F). Finally, merged libraries with less than 3.3% contamination were selected for population genetic analysis.

Genotyping and merging with dataset

After trimming of potentially damaged terminal ends bamfiles were genotyped with pileupCaller (https://github.com/stschiff/sequenceTools/tree/master/src-pileupCaller), which call one SNP per position considering the human genome as pseudo-haploid genome. Genotyped data were merged with Human Origins panel (~600K SNPs) [26] and 1240K panel [17]. For Goyet Q-2 the genotyping was applied to clipped and unclipped bamfiles, calling only transversions in the latter to avoid residual ancient DNA damage and merging these extra SNPs in the final genotype. The number of SNPs covered per individual is shown in Data S1G.

Kinship relatedness and individual assessment

We first calculated the pairwise mismatch rate between bam files to rule out a potential duplication of individuals. We found the same low mismatch rate comparing different bamfiles combination from libraries of BAL001 and BAL005, respectively, suggesting that both samples come from the same individual. We consequently merged them as BAL051.

We then used Relationship Estimation from Ancient DNA (READ) to estimate the degree of genetic kinship relatedness among the final set of individuals [37]. This method can determine first and second-degree relatedness among individuals and can also be used to test for potential cross-contamination among libraries from the same batch of sampling processing. We calculated the proportion of non-matching alleles and normalized the results separating Neolithic from HG individuals taking into account the potential genetic diversity within each group to calculate the proportions of non-matching alleles (*P0*). After the normalization of both groups, *P0* of HG ranged between 0.964-1.029 and between 0.974-1.074 for Neolithic individuals, which in both cases is higher than 0.90625, the top value for second-degree related individuals. In sum, there are no first- or second-degree relatives among our newly reported ancient Iberian individuals.

Phenotypic traits analysis

We extracted a list of 36 SNPs of functional importance or related to known phenotypic traits [22] (e.g., lactase persistence, pigmentation, eye colors) (Figure S4) and calculated the genotype likelihood based on the number of reads (using a quality filter q30) for each specific position to determinate the presence of the ancestral or derived alleles [22].

We interrogated different SNPs positions on the gene *OCA2* related to light eye color. We obtain the ancestral allele (rs12913832, 3 reads) in individuals CHA001 and ELT002, and heterozygous allele calls for ELT006, suggesting dark color eyes for all of them. The SNP coverage was not sufficient to reliable type the remaining individuals. Another allele from the same gene (rs1800404) related to eye pigmentation could support darker pigmentation in ELT006 than in ELT002. We also checked different SNPs positions in the gene *SLC45A2.* We obtain the ancestral alleles (rs1426654, 2 reads) in BAL0051 and (rs16891982, 4 reads) in CHA001 which suggest a darker skin color than ELT002 and ELT006, who are heterozygous or homozygous for the derived allele. The coverage in the other individuals is very low to allow comparisons. The allele rs3827760 of the *EDAR* gene, related to straight and thick hair, is ancestral in all individuals, albeit with variable coverage (CHA001, 4 reads; CHA003, 3 reads; ELT002, 14 reads and; ELT006, 8 reads). Also, as reported before for pre-farming and Neolithic individuals [13], none of our newly typed individuals show evidence for Lactase persistence. Individuals from the Neolithic times (ELT002, ELT006 and FUC003) show different combinations of derived and ancestral alleles of the gene *rs174546*, which is related to the capacity of regulation of the production of long-chain polyunsaturated fatty acids (FADS1/FADS2). Results are shown in Figure S4.

Mitochondrial and Y chromosome haplogroups

Using an in-house mtDNA capture assay, we could recover complete mitochondrial genomes from individuals CHA002, CHA003, CHA004, ELT002, ELT006, and FUC003. The coverage of the mtDNA genome for the rest or the samples ranges from 88.86%–99.99% (Data S1D). We used samtools v1.3.1 to extract reads from mitocapture data [31] and mapped them to the rCRS and called the consensus sequences using Geneious R8.1.974 [36]. We downloaded these consensus sequences in *fasta* format and they were used to determinate mitochondrial haplotypes using Haplogrep 2 [34] (Data S1F).

Iberian HG individuals from Balma Guilanyà and Moita do Sebastião belong to haplogroup U, together with the two MN individuals CHA004 and ELT006 (Data S1F). Individual BAL003 could be assigned to U2′3′4′7′8′9′, also found in the Paglicci 108 (~27,000 years cal BP, Italy), Rigney 1 (~15,500 years cal BP, France) [2], and Grotta d'Oriente C HG (~14,000 years cal BP, Italy) [19]. Individual BAL0051 belongs to U5b2a, also found in Neolithic Scotland [9]. Moita do Sebastião (CMS001) carries haplogroup U5b1, which was reported from MN, Bell Beaker and Middle Bronze Age individuals from Portugal and Spain [7, 9], and in the Ranchot 88 HG (~10,000 years cal BP, France) [2].

Early Neolithic individuals from Cueva de Chaves do not carry U haplogroups. Individual CHA001 could be assigned to haplogroup HV0+195, but was previously reported as K based on PCR-based results [70]. This haplogroup has been reported from MN Ireland [20], as well as MN Scotland and Bell Beaker individuals from England [9]. Individual CHA002 was assigned to K1a2a, which is common in Early Neolithic Iberia, e.g., Cova Bonica [8], Cova de Els Trocs [3] and Cueva del Toro [30], but also in Chalcolithic and Bell Beaker individuals from Iberia and Italy [9, 13]. CHA003 was assigned to K1a3a, so far reported from Neolithic Anatolia [13], Neolithic and Chalcolithic Scotland, and Bell Beaker individuals from Sicily [9]. Middle Neolithic individual CHA004 carried haplogroup U4a2f, found in HG from the Iron Gates, Romania, and Lithuania [14, 19]. The MN individual ELT002 carries haplogroup J1c1b, present in the Körös Early Neolithic [18], Neolithic from Scotland [9], Iberian Late Neolithic-Chalcolithic [7], and Bronze Age from Italy and Germany [6]. ELT006 was assigned to haplogroup U3a1, which has been reported from MN France and Germany [9, 19], and Chalcolithic Iberia [13]. Early Neolithic Fuente Celada carries haplogroup X2b+226, found in MN Hungary [71] and Middle Bronze Age Iberia [7]. X2b was

found in Iberian Late Neolithic [5], Chalcolithic [18] and Bell-Beaker individuals [9], Neolithic England [9] and Greece [72], and EN/LN Morocco [25].

For Y haplogroup determination, we first called the Y chromosome SNPs of the 1240K SNP panel from all male individuals using pileupCaller with MajorityCalling mode, (https://github.com/stschiff/sequenceTools), and mapping quality $\geq$30 and base quality $\geq$30 (Data S1E). Y chromosome haplogroups were called from the list of Y-SNPs included in the 1240K capture assay using the script *yhaplo* [35].

BAL0051 could be assigned to haplogroup I1, while BAL003 carries the C1a1a haplogroup. To the limits of our typing resolution, EN/MN individuals CHA001, CHA003, ELT002 and ELT006 share haplogroup I2a1b, which was also reported for Loschbour [73] and Motala HG [13], and other LN and Chalcolithic individuals from Iberia [7, 9], as well as Neolithic Scotland, France, England [9], and Lithuania [14]. Both C1 and I1/ I2 are considered typical European HG lineages prior to the arrival of farming. Interestingly, CHA002 was assigned to haplogroup R1b-M343, which together with an EN individual from Cova de Els Trocs (R1b1a) confirms the presence of R1b in Western Europe prior to the expansion of steppe pastoralists that established a related male lineage in Bronze Age Europe [3, 6, 9, 13, 19]. The geographical vicinity and contemporaneity of these two sites led us to run genomic kinship analysis in order to rule out any first or second degree of relatedness. Early Neolithic individual FUC003 carries the Y haplogroup G2a2a1, commonly found in other EN males from Neolithic Anatolia [13], Starçevo, LBK Hungary [18], *Impressa* from Croatia and Serbia Neolithic [19] and Czech Neolithic [9], but also in MN Croatia [19] and Chalcolithic Iberia [9].

Population genetic analysis

Labeling population groups

For the Paleolithic individuals, we adopted the labels from Fu et al. [2] and used the improved genotype calls from Mathieson et al. [19]. We included the HG with more than 15,000 SNPs covered in the PCA (Figure S1). If the HG individuals from the same site or chrono-cultural context clustered in the PCA analysis, we grouped them using the same label for the following population genetic analysis [14, 19, 21]. Neolithic individuals from Iberia were grouped by sites and by chrono-cultural context. For Neolithic individuals from outside of Iberia, we kept the label names from the respective initial publications [9, 13, 19, 20, 24, 74].

Principal Component Analysis

PCA analysis was run with the Human Origins dataset using *smartpca* v10210 (EIGENSOFT) with the option *SHRINKMODE* [33] using 777 modern populations to calculate eigenvectors on which aDNA samples were projected [74]. PC1 was multiplied by -1 (–PC1) in order to mirror geography.

F-statistics

D-statistics and F-statistics were calculated with qpDstat from ADMIXTOOLS (https://github.com/DReichLab). We used the 1240K panel to increase the number of SNPs covered by the ancient individuals and get more resolution in the statistic tests. Standard errors were calculated with the default block jacknife. We report and plot three standard errors in all F-statistics.

qpAdm and qpWave

We used *qpWave* and *qpAdm* from the ADMIXTOOLS package (https://github.com/DReichLab) to estimate admixture proportions. We used this framework to model and quantify the ancestry proportions of HG individuals in- and outside of Iberia (we only use HG or groups of HG with more than 30,000 SNPs). First, we tested whether Goyet Q-2 and Villabruna formed a clade with respect to the following set outgroups: Mota, Ust'-Ishim, Mal'ta 1 (MA1), Koros EN-HG, Goyet Q116-1, Mbuti, Papuan, Onge, Han, Karitiana and Natufian extending the set used by Olalde et al. [9]. The resulting *qpWave* model showed an extremely poor fit (p value = 3.07705115e-91), which means that our set of outgroups can be used to differentiate between Goyet Q-2 and Villabruna-related ancestry. We then modeled the ancestry in the HG groups as a mixture of Goyet Q-2 and Villabruna (Figure 3D and Data S1I). Alternatively, we also used El Mirón and Loschbour as potential source populations and the same ten outgroups, after checking that El Mirón and Loschbour are not equally related to the outgroups (p value = 5.41365181e-68) (Figure S3A and Data S1I).

We also used *qpWave* and *qpAdm* to explore the HG admixture in the Neolithic populations. In this case, the sources (left populations) were Anatolia Neolithic, Goyet Q-2 and Villabruna. We chose the same set of outgroups, and first checked that Anatolia Neolithic, Goyet Q-2 and Villabruna were not equally related to the outgroups (p value = 8.63552043e-92) (Figure 4B and Data S1J. We repeated the model with Anatolia Neolithic, El Mirón and Villabruna as sources using the same set of outgroups. However, the resulting *qpWave* model also resulted in a poor fit (p value = 9.5033482e-59) (Data S1J).

The results of all *qpWave* and *qpAdm* models are reported in Data S1I and S1J. The criteria to report these values were as follows: If the resulting p values were higher than 0.05 we report the three-sources model. If the p values were lower than 0.05 we show the best p value obtained from the three- or two-sources model (i.e., the nested model). In case of negative values for some of the sources, we report the two-sources model (nested model) and the respective p value of these models. We applied the same criteria to two-sources models.

Multi-Dimensional Scaling analysis (MDS)

We computed Multi-Dimensional Scaling (MDS) analysis using the R package *cmdscale* to measure the genetic dissimilarity among hunter-gatherers (HG), and then used the inverted [*1-f3(HG1; HG2, Mbuti)*] pairwise values among all the combinations [2].

DATA AND SOFTWARE AVAILABILITY

Data is available at ENA under study accession number PRJEB30985.

7. Manuscript C

Van de Loosdrecht et al. 2020. *Science Advances* (submitted March 2020)

Genomic and dietary transitions during the Mesolithic and Early Neolithic in Sicily

Short title

Genomic and dietary shifts in Sicilian prehistory

Authors

Marieke S. van de Loosdrecht,[1*] Marcello A. Mannino,[2,3*]† Sahra Talamo,[3,4] Vanessa Villalba-Mouco,[1] Cosimo Posth,[1] Franziska Aron,[1] Guido Brandt,[1] Marta Burri,[1] Cäcilia Freund,[1] Rita Radzeviciute,[1] Raphaela Stahl,[1] Antje Wissgott,[1] Lysann Klausnitzer,[3] Sarah Nagel,[5] Matthias Meyer,[5] Antonio Tagliacozzo,[6] Marcello Piperno,[7] Sebastiano Tusa,[8] Carmine Collina,[9] Vittoria Schimmenti,[10] Rosaria Di Salvo,[10] Kay Prüfer,[1,5] Jean-Jacques Hublin,[3,11] Stephan Schiffels,[1] Choongwon Jeong,[1,12] Wolfgang Haak,[1]† Johannes Krause[1*]†

Affiliations

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History (MPI-SHH), Jena, Kahlaische Strasse 10, D-07745, Germany.

[2]Department of Archeology and Heritage Studies, Aarhus University, Højbjerg, Moesgård Allé 20, 8270, Denmark.

[3]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology (MPI-EVA), Leipzig, Deutscher Platz 6, D-04103, Germany.

[4]Department of Chemistry G. Ciamician, Alma Mater Studiorum, Bologna University, Via Selmi, 2 , I-40126 Bologna, Italy.

[5]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology (MPI-EVA), Leipzig, Deutscher Platz 6, D-04103, Germany.

[6]Service of Bioarchaeology Museo delle Civiltà, museo preistorico etnografico “Luigi Pigorini”, P.le G. Marconi 14, Rome, Italy.

[7]Department of Ancient World Studies, Sapienza University of Rome, Rome, Via Palestro 63, 00185, Italy.

[8]Soprintendenza del Mare, Palermo, Via Lungarini 9, 90133, Italy.

[9]Museo Civico Biagio Greco, Mondragone (Caserta), Via Genova 2, 81034, Italy.

[10]Museo Archeologico Regionale "Antonino Salinas", Palermo, Via Bara all'Olivella 24, 90133, Italy.

[11]Collège de France, 11 place Marcellin Berthelot, 75005 Paris, France.

[12]School of Biological Sciences, Seoul National University, Seoul, 1 Gwanak-ro, 08826, Republic of Korea.

*Correspondence to: loosdrecht@shh.mpg.de (M.vdL), marcello.mannino@cas.au.dk (M.Ma), krause@shh.mpg.de (J.K)

† Co-supervised the study

Abstract

Southern Italy is a key region for understanding the agricultural transition in the Mediterranean due to its central position. We present a genomic transect for 19 prehistoric Sicilians that covers the Early Mesolithic to Early Neolithic period. We find that the Early Mesolithic hunter-gatherers (HGs) are a highly drifted sister lineage to Early Holocene western European HGs, whereas a quarter of the Late Mesolithic HGs ancestry is related to HGs from eastern Europe and the Near East. This indicates substantial gene flow from (south-)eastern Europe between the Early and Late Mesolithic. The Early Neolithic farmers are genetically most similar to those from the Balkan and Greece, and carry only a maximum of ~7% ancestry from Sicilian Mesolithic HGs. Ancestry changes match changes in dietary profile and material culture, except for two individuals who may provide tentative initial evidence that HGs adopted elements of farming in Sicily.

One-sentence summary

Genome-wide and isotopic data from prehistoric Sicilians reveal a pre-farming connection to (south-) eastern Europe, and tentative initial evidence that hunter-gatherers adopted some Neolithic aspects prior to near-total replacement by early farmers.

Key words

Ancient DNA, isotopes, Europe, Sicily, Ice Age, Mesolithic, Neolithic, Castelnovian, farming

MAIN TEXT

Introduction

Southern Italy and Sicily feature some of the earliest evidence for agricultural food production in the Central Mediterranean, starting as early as ~6,000 calBCE (*1*) or earlier (~6,200 calBCE (*2*)). In the Mediterranean area, two Early Neolithic meta-horizons had developed in parallel by ~5,500 calBCE (*3-5*). In the eastern and central Mediterranean, Early Neolithic farmers produced Impressa Wares with various decorative impressed designs made with a wide selection of tools. In contrast, in the western Mediterranean the decorative designs were preferentially made with *Cardium* seashell impressions, resulting in the typical Cardial Ware pottery (*6*). In Sicily and southern Italy two Impressa Ware horizons appeared rapidly in a timeframe of ~500 years. The very first aspect of Impressa Wares appeared 6,000-5,700 calBCE followed by the Impressed Ware of the Stentinello group (Stentinello/Kronio) around 5,800-5,500 calBCE (*1, 7, 8*). The Early Neolithic horizons in Sicily may have their origin in the early farming traditions in the Balkans (*9-11*).

Grotta dell'Uzzo, in northwestern Sicily, is a key site for understanding human prehistory in the Central Mediterranean, and has provided unique insights into the cultural, subsistence and dietary changes that took place in the transition from hunting and gathering to agro-pastoralism (*12-14*). The cave stratigraphy covers the late Upper Palaeolithic through the Mesolithic and up to the Middle Neolithic, with traces of later occupation. Quite uniquely in the Mediterranean region, deposits at Grotta dell'Uzzo show a continuous occupation during the Mesolithic (*12*). Zooarchaeological and isotopic investigations indicated shifts in the economy and diet of the cave occupants, who subsisted by hunting and gathering for most of the Mesolithic, started to exploit marine resources towards the end of the Mesolithic, and combined all the previous activities with the exploitation of domesticates and increased fishing during the Early Neolithic (*12-14*).

Genome-wide data has been published for six (Epi-)Gravettian HGs and one tentatively Sauveterrian Mesolithic HG from peninsular Italy, and one Late Epigravettian HG from *OrienteC* in Sicily (*15-17*). To date, no ancient genomes are available for Early Neolithic and Late Mesolithic individuals from Sicily or southern Italy. The question of whether the agricultural tradition was adopted by local HGs or brought to Sicily by incoming farmers, thus, remains open.

Results

Here, we investigated the biological processes underlying the transition from hunting and gathering to agropastoralism in Sicily. We reconstructed the genomes for 19 individuals from Grotta dell'Uzzo dating to a period from the Early Mesolithic ~8,810 calBCE to the Early Neolithic ~5,210 calBCE (Data file 1). We obtained a direct accelerator mass spectrometry (AMS) radiocarbon (^{14}C) date on the skeletal elements that were used for genetic analysis for 15 individuals, and determined carbon (δ^{13}C) and nitrogen (δ^{15}N) isotope values from the same bone collagen for dietary reconstruction (Data file 1).

We extracted DNA from bone and teeth in a dedicated clean room, built DNA libraries and enriched for ~1240k single nucleotide polymorphisms (SNPs) in the nuclear genome and independently for the complete mitogenome (*18*) using in-solution capture (*19*). We restricted our analyses to individuals with evidence of authentic DNA, and removed ~300k SNPs on CpG islands to minimize the effects of residual ancient DNA damage (*20*). The final data set includes 868,755 intersecting autosomal SNPs for which our newly reported individuals cover 53,352-796,174 SNP positions with an average read depth per target SNP of 0.09-9.39X (Data file 1). We compared our data to a global set of contemporary (*21*) and 377 ancient individuals from Europe, Asia and Africa (*15-17, 21-51*).

Genetic grouping of the ancient Sicilians

First, we aimed to group the individuals for genetic analysis. For this we co-analysed one Epigravettian HG (*OrienteC*) from the Grotta d'Oriente site on Favignana island in southwestern Sicily (12,250-11,850 calBCE, ^{14}C date on charcoal from the deposit (*15, 17*). The ancient Sicilians form three genetic groups that we distinguished based on the individuals' ^{14}C dates (Fig. 1B, Data file 1), position in Principal Component Analysis (PCA, Fig. 1C), mtDNA haplogroups (Supplementary Section S7), and degree of allele sharing in outgroup-f_3 statistics and qpWave-based ancestry models (Supplementary Section S2).

The first two groups consist of individuals that fall close to a cluster of western European Mesolithic hunter-gatherers ('WHG') that includes the 14,000-year-old individual from the Villabruna site in northern Italy ('*Villabruna*') (*16*) (Fig. 1C). The first and oldest genetic group, which we labelled Sicily Early Mesolithic (Sicily EM, n=3), contains the previously published Epigravettian *OrienteC* (12,250-11,850 calBCE (*15, 17*)) and the two oldest HGs from Grotta dell'Uzzo (~8,800-8,630 calBCE). These three individuals carried mitogenome lineages that fall within the U2'3'4'7'8'9 branch (Supplementary Section S7, and (*15, 17*) for *OrienteC*). From that haplogroup node they shared nine mutations specific to their

lineage and were differently related to each other with regard to three additional private mutations. U2'3'4'7'8'9 mitogenome lineages have already been reported for Upper Palaeolithic European HGs, such as *Paglicci108* associated with the Gravettian in Italy (26,400-25,000 calBCE (*52*)). The second genetic group, which we labelled as Sicily Late Mesolithic (Sicily LM, n=9), contains nine individuals dated to ~6,750-5,850 calBCE, The mitogenome haplogroups carried by the Sicily LM HGs are U4a2f (n=1), U5b2b (n=2), U5b2b1a (n=1), U5b3(d) (n=3), U5a1 (n=1), and U5a2+16294 (n=1), which are typical for European Late Mesolithic WHGs (*52, 53*).

The third and most recent genetic group, which we labelled as Sicily Early Neolithic (Sicily EN, n=7), contains seven individuals dated to ~5,460-5,220 calBCE. In PCA, these individuals show substantially Near-Eastern-related ancestry and fall close to early farmers from the Balkans (Croatia, Greece), Hungary, and Anatolia, but not Iberia (Fig. 1C) (*17, 38, 39, 54*). All the individuals in the Sicily EN group, with sufficient coverage for genome reconstruction, carried mitogenome haplogroups characteristic for European early farmers: U8b1b1 (n=2), K1a2 (n=1), N1a1a1 (n=1), J1c5 (n=1) and H (n=1) (*55*).

Mesolithic substructure and dynamics

Previous research has shown that the genetic diversity among European HGs after the Last Glacial Maximum (LGM) was shaped by various deeply diverged ancestries (*16, 17, 32, 38, 50, 56*). One such ancestry came from a group of pre-LGM individuals dating to ~30,000 calBCE and associated with the Gravettian industry (Věstonice cluster). Another one was from individuals associated with the Magdalenian industry (El Mirón cluster) that appeared in Europe by ~17,000 calBCE (*16*). Individuals of the Villabruna cluster, also referred to as western European hunter-gatherers (WHGs), appeared ~12,000 calBCE throughout continental Europe, and replaced most of the ancestry of the earlier clusters in European HGs. In Mesolithic HGs from eastern Europe (EHGs, ~6,000 calBCE) and the Iron Gates HGs from southeastern Europe (~9,500-5,800 calBCE), ancestry related to Upper Palaeolithic Siberians (Ancient North Eurasians, ANE) was found in addition to WHG ancestry.

To visualize the genetic differentiation among the Sicily EM and LM HGs, and their relation to other West Eurasian HGs, we plotted pairwise genetic distances calculated as f_3*(Mbuti; HG1, HG2)* in a Multidimensional Scaling (MDS) plot (Fig. 2A). The genetic variation among post-LGM European HGs is structured along two clines: 1) a WHG-EHG-ANE cline, confirming the genetic gradient found in Mesolithic HGs from western to eastern Europe and 2) a WHG-GoyetQ2 cline between WHG and Central European Magdalenian-associated individuals on which Iberian HGs take an intermediate position (*15-17,*

38, 50, 56). As previously reported for *OrienteC* (*15*), the Sicily EM HGs *UZZ5054* and *UZZ96* fall at the extreme WHG-end of both ancestry clines, slightly outside the genetic variation of the Villabruna cluster, named after the site name of its oldest representative individual (~12,230-11,830 calBCE) (Fig. 2A). The position of the Sicily EM HGs on the MDS plot either hints at a WHG ancestry component that is more basal than that found in Villabruna cluster individuals, and/or at substantial genetic drift. Compared to the Sicily EM HGs, the Sicily LM HGs fall closer to the Villabruna cluster in between Sicily EM HGs and Mesolithic Iron Gates HGs (Fig. 2A).

First, we investigated whether the Sicily EM HGs contain substantial lineage-specific genetic drift. We hence determined the nucleotide diversity (π) by calculating the average proportion of nucleotide mismatches for all possible combinations of individual pairs within the Sicily EM HG group, and compared that to the average of other HG groups. We indeed found significant lower nucleotide diversity for the Sicily EM HGs (95% confidence interval (95CI) $\pi = 0.161\text{-}0.170$), compared to HGs from Italy from the preceding Upper Palaeolithic (95CI $\pi = 0.227\text{-}0.239$), and the subsequent Sicily LM HGs (95CI $\pi = 0.217\text{-}0.223$), and later farmers (Fig. 3 and fig. S3.1). In addition, the nucleotide diversity for the Sicily EM HGs is ~20% lower compared to contemporaneous HG groups from Central Europe, Iberia and the Iron Gates (Fig. 3). The reduced genetic diversity of Sicily EN HG hence appears to be both geographical- and temporal-specific.

Secondly, we compared the ancestry component as found in Sicily EM and LM HGs, and their respective affinities to West Eurasian HGs, using an f_4-cladality test of the form *f_4(Chimp, X; Sicily EM HGs, Sicily LM HGs)*. We found a pattern that is linked with geography (Fig. 4A): Sicily EM HGs share significantly more alleles with HGs from western Europe, including Villabruna cluster HGs, and Iberian Upper Palaeolithic and Mesolithic HGs that carry Magdalenian-associated ancestry (*16, 50*). In contrast, the Sicily LM HGs share significantly more alleles with Upper Palaeolithic and Mesolithic HGs from (south-)eastern Europe and Russia, including *AfontovaGora3*, EHGs, *Mal'ta1* and Iron Gates HGs. Notably, comparing the ancestry in Sicily EM HGs with that of *Villabruna* with the cladality statistic *f_4(Chimp, X; Sicily EM HGs, Villabruna)* results in a similar geographical pattern (fig. S4.3). Also here, Sicily EM HGs share an excess of alleles with western European HGs, including the majority of Villabruna cluster individuals, whereas *Villabruna* does with (south-) eastern European HGs. Fu et al. (*16*) already showed an East Asian affinity for some individuals of the Villabruna cluster individuals compared to older individuals. However, using the Sicily EM HGs as a baseline for WHG ancestry pulls out the difference in genetic affinities to Magdalenian-associated and EHG/ANE-related ancestry more strongly between western and eastern West Eurasian HGs (Supplementary Section S4).

With more explicit modelling using qpGraph (*57*) we further examined the phylogenetic position of Sicily EM HGs. In the least complex scaffold tree Sicily EM HGs and *Villabruna* form a clade on an unadmixed branch, with *GoyetQ2,* a Magdalenian-associated HG, as an outgroup to both of them (fig. S5.3). However, trees that place either *Villabruna* or Sicily EM HGs on an admixed branch with an additional ancestry contribution from *AfontovoGora3* and *GoyetQ2*, respectively, fit the allele frequencies approximately equally well (Supplementary Section S5). Overall, the results suggest that the Sicily EM HGs represent a highly drifted branch closely related to the Villabruna cluster. We can however not rule out that Sicily EM HGs derived an ancestry contribution from Magdalenian-associated individuals or that Sicily EM HGs descended from a more basal lineage that admixed into Iberian HGs and Villabruna cluster individuals (fig. S5.5).

Subsequently, we characterized the ancestry profile of the Sicily LM HGs in more detail. Since the position of Sicily LM HGs on the MDS and PCA plots (Fig. 1D, 2A) is closer to *Villabruna* cluster individuals and EHGs, we investigated whether their gene pool is the result of admixture between the preceding Sicily EM HGs and a group high in EHG-ancestry, or is genetically drifted from Villabruna cluster HGs.

First, we used the outgroup f_3-statistic f_3*(Mbuti; Sicily LM HGs, X)* to investigate for various West-Eurasian HGs (*X)* which one is genetically closest to Sicily LM HGs (fig. S4.1). Sicily LM HGs shows the highest degree of allele sharing with Sicily EM HGs, followed by other individuals from the Villabruna cluster. Moreover, the statistic f_4 (Chimp, Sicily LM HGs; Sicily EM HGs, X) is strongly significantly negative for all tested West-Eurasian HGs, including Villabruna cluster individuals (fig. S4.2). This implies that the Sicily LM HGs form a clade with Sicily EM HGs to the exclusion of other West-Eurasian HGs. Taken together, these statistics indicate substantial continuity in ancestry between the Sicily EM and LM HGs. However, when Sicily EM HGs is used as baseline for the ancestry in Sicily LM HGs in the f_4-statistic f_4*(Chimp, X; Sicily EM HGs, Sicily LM HGs)*, additional admixture signals are found for various HGs from (south)-eastern Europe and Russia (Fig. 4A). Therefore, Sicily EM HGs do not represent the full gene pool of the Sicily LM HGs.

Subsequently, using qpAdm-based admixture models (*36*) we aimed to more explicitly model the gene pool of the Sicily LM HGs. We found that a three-way ancestry combination of 75.0±1.6% Sicily EM, 15.5±2.4% EHG and 9.5±2.8% Pınarbaşı, a ~13,300 calBCE HG from central Anatolia (*25*), results in a good fit (P = 0.123, Table 1). Notably, replacing the Sicily EM ancestry with that of *Loschbour* resulted in poorly fitting models (P_{Adm} = 4.27E-09, Table 1). The assigned ancestry components confirm both the substantial continuation of the local Sicily EM ancestry, and the influx of a non-local ancestry frsm (south-)eastern Europe, in Sicily during the Mesolithic. Moreover, the Upper Palaeolithic Pınarbaşı-related

ancestry in the Sicily LM HGs is striking, and underlines previous indications for a pre-Neolithic genetic connection between the Near East and European HGs by at least 12,000 calBCE (*16, 17, 25*).

In a last step, we characterised the ancestry profiles of the Sicily LM HGs on an individual level (Fig. 2B and Fig. 5A). The Sicily LM HGs form a heterogeneous group, with some individuals containing both the EHG and Near Eastern-related ancestry in addition to the preceding local Sicily EM ancestry, whereas others contain solely the additional EHG-related ancestry. The summed proportion for the non-local (EHG + Pınarbaşı) ancestry component in Sicily LM HGs ranges between 17±2% and 33±4% (Fig. 2B, Fig. 5A, Data file 1). Interestingly, whereas a Near Eastern-related ancestry component is discerned in many of the Mesolithic HGs from the Iron Gates and other areas in the Balkan, Baltic or Scandinavia, this is not the case for any of the Villabruna cluster individuals. In contrast to previous statements that the Villabruna cluster individuals form a genetically homogenous group (*16, 25*), here the individuals show rather diverse ancestry profiles with various combinations of Sicily EM, EHG and *GoyetQ2*-related ancestry. Since the Sicily LM HG gene pool contains the distinct Near Eastern-related ancestry but not the *GoyetQ2*-related ancestry, it is unlikely that the diversity of this group originated solely from genetic drift from these Villabruna cluster individuals. We speculate that a single hitherto unsampled population, with an ultimate origin perhaps in the Near East or Caucasus, might harbor the genetic diversity that fits the combined EHG- and Near Eastern-related ancestry in Sicily LM HGs and additional West-Eurasian HGs (e.g. related to the ~24.000 calBCE Caucasus HGs from Dzudzuana Cave (*56*).

Expanding early farmers replace local HGs in Sicily

Recent ancient human DNA studies have shown that the genetic variation in Early Neolithic groups from central and southwestern Europe is a subset from that found in early farmers from Barcin in northwestern Anatolia and Revenia in northern Greece (*30, 34-37, 39*). Interestingly, early farmers from Diros in Peloponnese Greece might harbour an ancestry component that places them outside of the genetic diversity represented by those from Anatolia Barcin (*17*).

In PCA (Fig. 1C, Data file 1) the Sicilian early farmers (Sicily EN) plot closest to Early Neolithic groups from the Balkan, Serbia, Hungary, Greece and Anatolia, but not from Iberia. With the f_3-outgroup statistic f_3*(Mbuti; Sicily EN, X)* we aimed to determine the best genetic proxies for the overall gene pool related to Sicily EN. Congruent to the PCA results, the Sicilian early farmers share most genetic drift with various Early Neolithic farmers from the Balkan and Central Europe (Fig. 4B). In addition, in the PCA the Sicilian early farmers do not fall on a cline towards the Sicilian Mesolithic HGs. This suggests that HG

ancestry is either absent or low, which would imply a large population replacement in Sicily with the appearance of the Early Neolithic horizons.

To investigate this more formally, we first tested whether Sicilian early farmers contain a distinct HG ancestry component that is not found in early farmers from Anatolia EN Barcin (*38*). For this we used the admixture f_4-statistic *f_4(Chimp, X; Anatolia EN Barcin, Sicily EN)* that tests for an excess of shared alleles between Sicilian early farmers and various West Eurasian HGs (*X*) when Anatolia EN Barcin is used as a baseline for the early farmer ancestry (fig. S6.1). Sicily EN shows strongly significant signals for admixture for the preceding local Sicily EM (z = 4.12) and Sicily LM HGs (z = 4.09).

To test whether the HG component in Sicily EN is genetically closer to the ancestry of the local preceding Sicily LM HGs or to that of a specific non-local HG source, we performed the f_4-admixture statistic *f_4(Chimp, Sicily EN; Sicily LM HG, non-local HG)* (fig. S6.2). For HGs from central and southwestern Europe this statistic is negative, indicating that none of them is genetically closer to the HG ancestry in Sicily EN than Sicily LM HGs are. Contrastingly, Mesolithic HGs from southeastern Europe, including Iron Gates HGs, Croatia Mesolithic and Koros EN HGs, do show an excess of shared genetic drift with Sicily EN (fig. S6.2). However, southeastern European HGs and early farmers from Anatolia Barcin and the Balkan share a part of their ancestry (*17, 25*). The excess genetic attraction for southeastern European HGs might hence be driven by the farmer, rather than HG, ancestry component in Sicily EN.

To further investigate what combination of early farmer and HG ancestry fits the Sicily EN gene pool best, and their respective admixture proportions, we used qpWave (P_{Wave}) and qpAdm-based (P_{Adm}) ancestry models. We required a test result to be more extreme (larger) than P = 0.1 in order not to reject the null-hypothesis of a full ancestry fit (full rank). We found that the ancestry for approximately half of the Sicilian early farmers can be fitted as entirely early farmer ancestry as found in Anatolia EN Barcin, Greece EN Peloponnese, Croatia EN Cardial and Impressa, Hungary EN Koros or Germany LBK (Fig. 4, Data file 7). However, these early farmer sources were rejected as sole ancestry sources to the combined Sicily EN gene pool (max. $P_{Wave} \leq 0.011$ for Croatia EN Cardial, Data file 7). We could improve the model fit to the Sicily EN gene pool by adding ~4-9% ancestry from Sicily LM or Sicily EM HGs as a second source to early farmer ancestry from either Anatolia EN Barcin ($P_{Adm} \geq 0.016$), Greece EN Peloponnese ($P_{Adm} \geq 0.059$), Croatia EN Cardial ($P_{Adm} \geq 0.056$) or Hungary EN Koros ($P_{Adm} \geq 0.300$) (Table 1, Data file 7). Notably, three-way mixture models, in which the ancestry from Anatolia EN Barcin was combined with that from early farmers from Ganj Dareh as a proxy for the early farmer ancestry, improved the fit to the Sicily EN gene pool even further ($P_{Adm} \geq 0.335$, Table 1). We can therefore not exclude the possibility that the early farmer ancestry in Sicily EN falls partly outside of the broader genetic diversity of the early farmers from Europe and Anatolia Barcin (Supplementary Section S6).

Replacing the local Sicilian HG ancestry with that from Iron Gates HGs or many other Mesolithic European HGs often resulted in similar fits (Table 1, Data file 7). To the limits of the genetic resolution, we hence could not accurately discern whether the Sicilian early farmers derived their HG ancestry from a Mesolithic local Sicilian HG source or from a geographically more distant one. However, a local contribution appears to be the most plausible explanation, which is in line with the hypothesized continuity in occupation at Grotta dell'Uzzo (*14, 58-60*).

Discussion

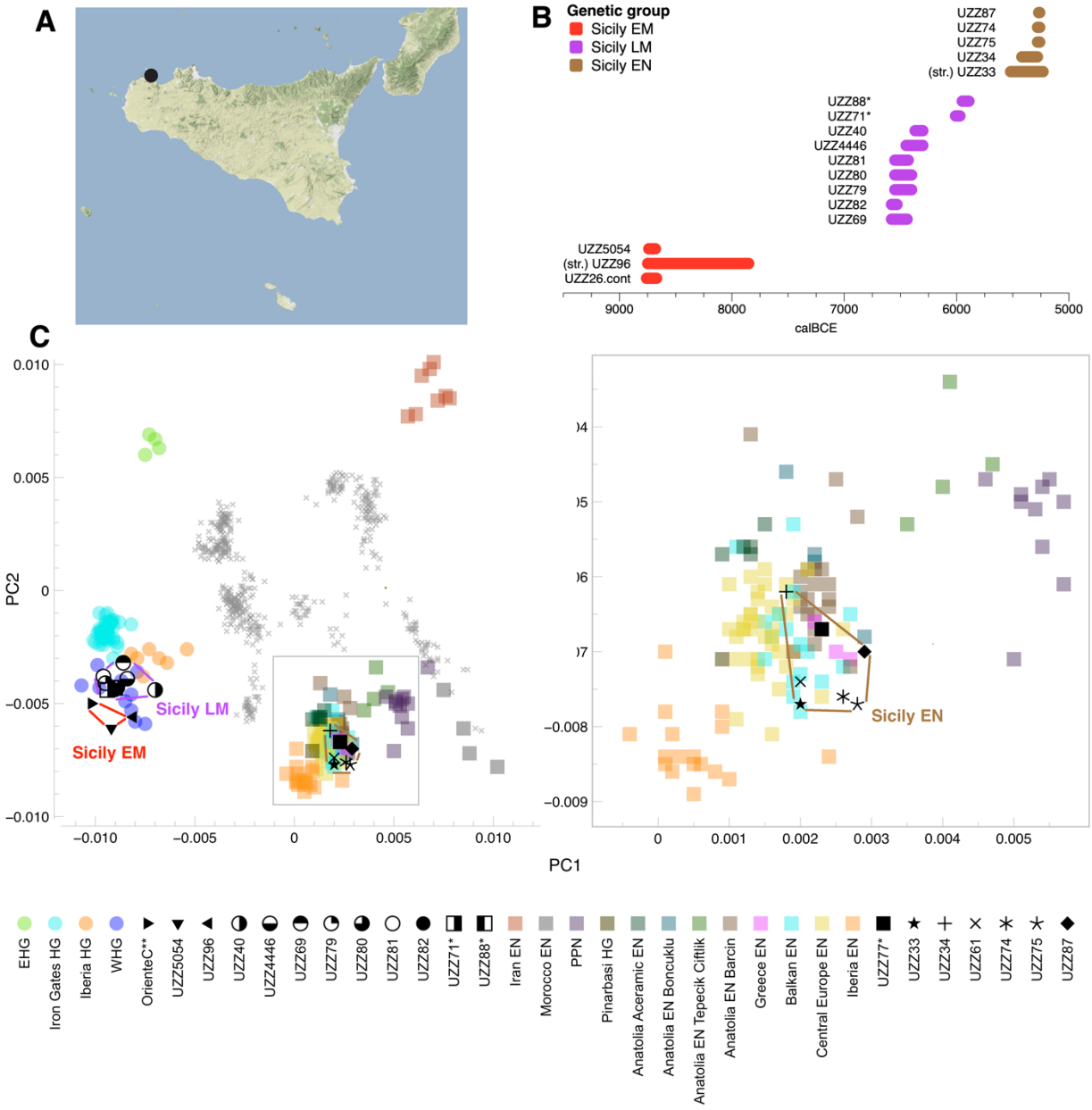
Southern Italy has long been viewed as a southern refugium (*61, 62*) during the LGM, ~25,000 years ago, from where Europe was repopulated (*16, 52, 63*). The earliest evidence for the presence of *Homo sapiens* in Sicily dates to ~17,000-16,000 calBCE, following the time when a land bridge connected the island to peninsular Italy (*64, 65*). Some Late Epigravettian sites in Sicily contain rock panels with engraved animal figures, which are indistinguishable from those of the Franco-Cantabrian style typical of the Magdalenian (*66*). Moreover, the presence of some decorated pebbles at sites in northwestern Sicily is also suggestive of cultural links with the Azilian in the French Pyrenees (*66*). Although there are many sites in peninsular Italy and Sicily with evidence of Late Upper Palaeolithic occupation, this decreases during the Mesolithic (*67*). The Early Mesolithic HGs from Grotta dell'Uzzo analysed here produced a lithic industry of Epigravettian tradition ((*68*), Supplementary Section S1) and, in continuity with their predecessors, subsisted mainly by hunting large terrestrial game with important contributions of plant foods and limited consumption of marine resources (*12*) (Fig. 5B). Here we showed that, compared to the ~12,000 calBCE *Villabruna* individual, the Sicily EM HGs have a higher genetic affinity to Magdalenian-associated Iberian HGs *El Miron* and *Balma Guilanyà*. Given the profoundly reduced genetic diversity in the Sicily EM HGs and their closely related non-identical mitogenomes in the U2'3'4'7'8'9 branch, we speculate that these individuals are unadmixed descendants from a peri-glacial refugium population.

The subsequent ~6,750-5,850 calBCE Late Mesolithic HGs derive between 15-37% of their ancestry from an EHG-related source with an affinity to the Near East. The substantial influx of non-local ancestry indicates a large genetic turnover in Sicily during the Mesolithic, and is matched by a change in diet that is characterized by a statistically significantly higher intake of marine-based protein (Fig. 5B). The seven oldest individuals in this group (dated ~6,750-6,250 calBCE) are tentatively assigned to the Castelnovian *sensu lato* facies (*12, 14, 69*) (Supplementary Section S1). The Castelnovian is part of the pan-European Late Mesolithic blade and trapeze lithic complex, and appeared throughout Italy ~6,800-6,500 calBCE ((*70*), D. Binder personal communication). These lithic industries have been argued to

originate from the Circum Pontic area (*71, 72*), with a possible ultimate origin from as far as eastern Asia (*73, 74*), or alternatively from the Capsian culture in northwestern Africa (*75, 76*). The ancestry profiles of the individuals associated with the Castelnovian *sensu lato* show a similarity to those of Mesolithic HGs from the Iron Gates, eastern Europe and the Baltic, hence providing support for a connection to the East.

The Sicilian early farmers carried almost exclusively ancestry characteristic for early European farmers. The preceding Late Mesolithic HGs may have contributed only a maximum of ~7% ancestry (Fig. 5A). Six individuals in this group are from layers that chronologically coincide with the presence of Early Neolithic Stentinello Wares, and one individual (*UZZ77*, undated) tentatively with older aspects of Impressa Ware (Supplementary Section S1). The isotope values of the Stentinello Ware associated early farmers are congruent with them having a terrestrial-based farming diet (Fig. 5B). Their ancestry composition points to a full-scale demographic transition during the Early Neolithic, similar to the Mesolithic-Neolithic transition in other regions in Europe (*30, 34, 36, 37, 53, 54, 77*). Intriguingly however, two individuals *UZZ71* and *UZZ88*, dated ~6,050-5,850 calBCE, chronologically coincide with layers at the site that may contain the very first aspects of Impressa Wares (*12*). These two individuals fall fully within the genetic diversity of the Late Mesolithic HGs associated with the Castelnovian *sensu lato*, despite postdating them by ~200 years (Fig. 1C, 5A, Supplementary Section S1, S2). Both these individuals show isotope values that are strikingly different from both the later Sicilian Early Neolithic farmers associated with Stentinello pottery and preceding Late Mesolithic HGs (Fig. 5B, Data file 1). The diet of the *UZZ71* individual included a large proportion of freshwater protein, similarly to what is recorded for Mesolithic hunter-gatherers from the Iron Gates on the Balkan peninsula (e.g. (*78, 79*)). On the other hand, *UZZ88* has an isotopic composition ($\delta^{13}C$ = -19.2‰, $\delta^{15}N$ = 7.1‰) that suggests a terrestrial-based farming diet with very low levels of animal protein consumption and that is significantly different from the hunter-gatherers it descended from (mean $\delta^{13}C$ = -16.6±1.8‰, mean $\delta^{15}N$ = 13.0±1.0‰), who relied for around half of their protein on seafood. Given the distinct diets and the intermediate ^{14}C dates, these two individuals might provide tentative initial evidence that HGs adopted elements of farming in Sicily, as was hypothesized by Tusa (*59, 60*). However, more extensive research on the stratigraphy of Grotta dell'Uzzo is necessary to determine whether the Impressa Ware aspects are indicative of a transitional period (*80*) or should be considered intrusive (*81*).

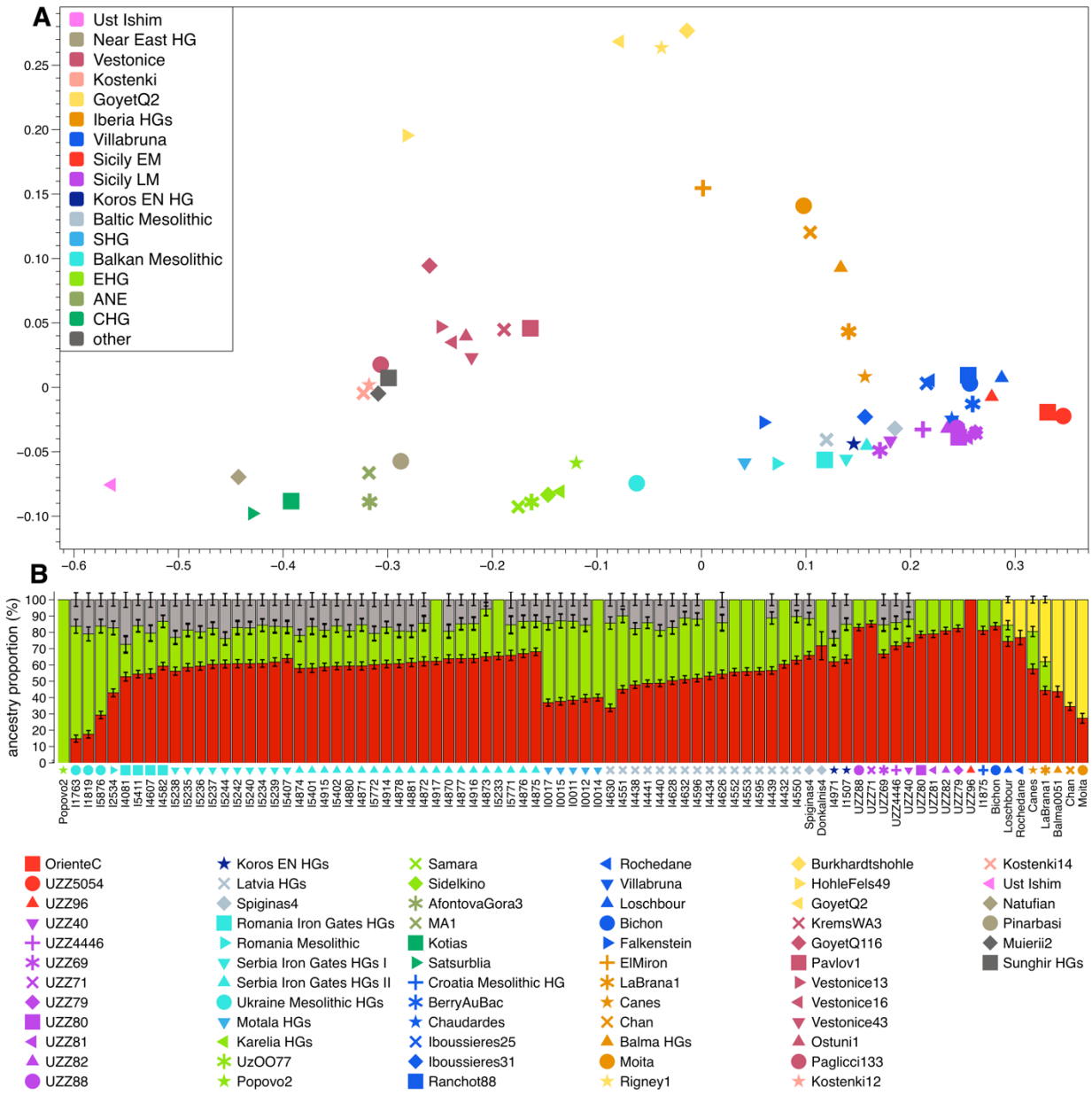
Although individuals that blur the Mesolithic and Neolithic dichotomy are rather rare, they have been reported before (*17, 29, 78*). Taken together, these individuals could indicate that hunter-gatherers may have met early farmers in different areas of the Mediterranean for which the frequency and exact geographical interaction sphere remains to be unravelled.



340

341 **Fig. 1. Genetic structure of ancient West Eurasians. (A)** The geographical location of Grotta dell'Uzzo
342 in Sicily. **(B)** Dating of the ancient Sicilians. Dates were determined from direct radiocarbon (^{14}C)
343 measurements, and for *UZZ96* and *UZZ33* from the stratigraphy (str.) (Supplementary Section S1). Dates
344 could not be inferred for individuals *UZZ61*, and -*77*. The individuals fall into three temporal groups, and
345 are coloured according to their assigned genetic group. *UZZ71, -88* and -*77* may be contemporaneous with
346 early Impressa Ware aspects at the site (marked by *). **(C)** PCA plots for the genetic distances between
347 West-Eurasians with principal components constructed from individuals from 43 modern Eurasian groups

(grey crosses). The ancient Sicilians are projected (black symbols) together with relevant previously published hunter-gatherers (coloured dots) and early farmers (coloured squares). We co-analyzed an Epigravettian HG from OrienteC (*15*) (marked by **). The genetic variation of the ancient Sicilians forms three genetic groups, that we labelled Sicily EM, LM and EN. The individuals in the Sicily EM and LM genetic group fall close to individuals from the Villabruna cluster that are characterized by high levels of WHG ancestry. In contrast, those from the Sicily EN genetic group contain substantially more Near Eastern ancestry, and fall among early farmers from the Balkan and Central Europe but not from Iberia.



355

356 **Fig. 2. Genetic profile characterization for the Sicilian HGs. (A)** MDS plot showing structure in the
357 genetic variation among West Eurasian HGs. Genetic distances are based on pairwise f_3-outgroup statistics
358 of the form f_3*(Mbuti; HG1, HG2)*. Colours reflect various ancestry clusters or geographical groups. The
359 genetic variation among West Eurasian HGs is structured along a WHG-EHG-ANE and WHG-GoyetQ2
360 ancestry cline. The Sicily EM HGs (red) fall at the extreme WHG end of both ancestry clines. Sicily LM
361 HGs (purple) fall among Villabruna cluster HGs (blue) in between Sicily EM HGs and Mesolithic HGs
362 from the Balkan (turquoise). **(B)** Individual ancestry profiles for Sicily LM HGs and other relevant West-
363 Eurasian HGs. Results are from qpWave- and qpAdm-based admixture models that inferred the ancestry of

each target HG as a one-, two- or three-way source mixture of ancestry approximated by Sicily EM (red: *UZZ5054, OrienteC), GoyetQ2* (yellow), EHG (lime green: Karelia HGs, *Uz0077, Samara, Sidelkino*), and *Pınarbaşı* (brown). Sicily EM, *GoyetQ2* and *Pınarbaşı* are taken as proxies for WHG-, Magdalenian- and Near Eastern HG-related ancestry, respectively. Error bars reflect 1 standard error (SE). Individuals with >150k SNPs covered are plotted. Sicily LM HGs, *Bichon* and the Croatia Mesolithic HG contain the highest proportions of Sicily EM ancestry. The Near Eastern-related ancestry is found in some Sicily LM HGs, and frequently among Mesolithic HGs from the Balkan, Baltic and Scandinavia. Congruent to their position in the MDS plot, the Sicily LM HGs ancestry profiles appear intermediate to those of *Bichon* and the Croatia Mesolithic HG (Villabruna cluster), and Mesolithic HGs from the Iron Gates in southeastern Europe.

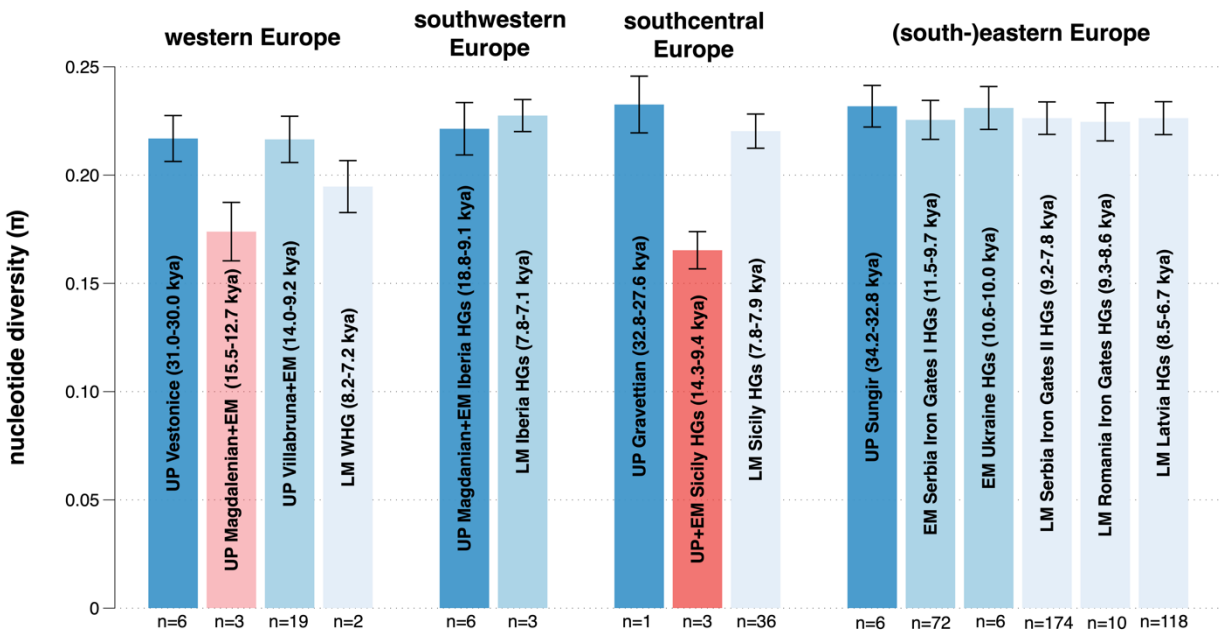


Fig. 3. The genetic diversity in Sicilian Early Mesolithic HGs is significantly reduced. Nucleotide diversity (π) was inferred from pseudo-haploid genotypes, calculated as the average proportion of nucleotide mismatches for autosomal SNPs covered in individual pairs within a given HG group. The averages for HG groups from different time periods in different regions in Europe are plotted. UP = Upper Palaeolithic, EM = Early Mesolithic, LM = Late Mesolithic. Individuals are grouped based on their assigned genetic cluster, geographical and temporal proximity (see Data file 1). The number of individual pairs (*n*) that is used to determine the average for each time period is given. Error bars reflect 95% confidence intervals from 5Mb jackknifing. The nucleotide diversity for UP + EM Sicily HGs (red) is significantly lower compared to that for HGs from other time periods or regions in Europe, except for Magdalenian-associated individuals from western Europe (UP Magdalenian + EM: *Rigney1*, *Burkhardshole*, *Hohlefels49, GoyetQ2*).

386

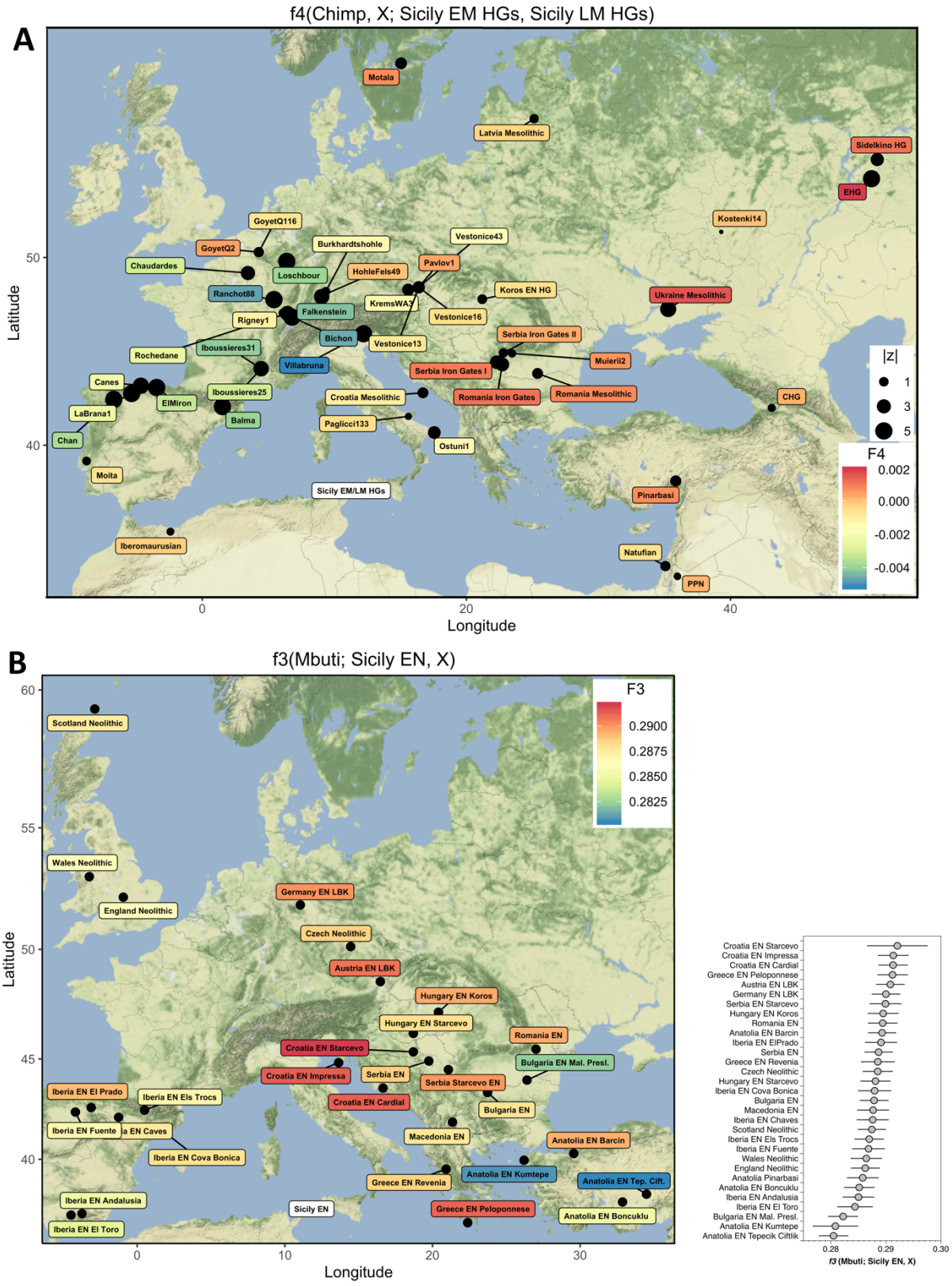
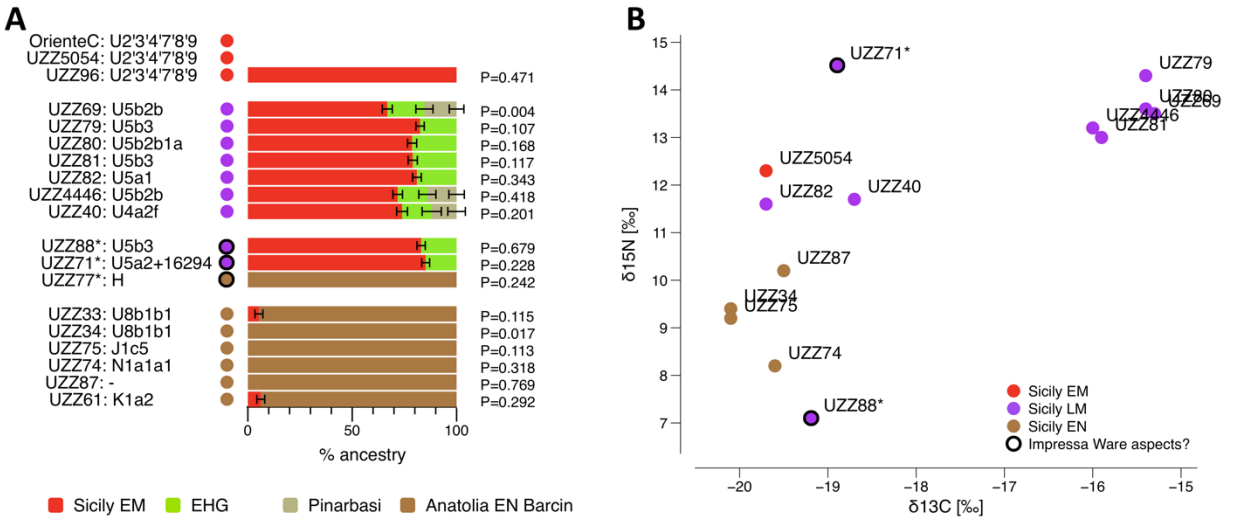

A
f4(Chimp, X; Sicily EM HGs, Sicily LM HGs)
Motala
Latvia Mesolithic
Sidelkino HG
EHG
Kostenki14
GoyetQ116
GoyetQ2
Burkhardtshohle
Vestonice43
Chaudardes
Loschbour
HohleFels49
Pavlov1
Ranchot88
Falkenstein
KremsWA3
Koros EN HG
Ukraine Mesolithic
Rigney1
Vestonice16
Bichon
Vestonice13
Serbia Iron Gates II
Iboussieres31
Rochedane
Villabruna
Serbia Iron Gates I
Mulerii2
Canes
Iboussieres25
Croatia Mesolithic
Romania Iron Gates
Romania Mesolithic
CHG
ElMiron
LaBrana1
Balma
Paglicci133
Chan
Ostuni1
Moita
Sicily EM/LM HGs
Pinarbasi
Iberomaurusian
Natufian
PPN
|z|
1
3
5
F4
0.002
0.000
-0.002
-0.004
Latitude
Longitude
50
40
0
20
40
B
f3(Mbuti; Sicily EN, X)
Scotland Neolithic
Wales Neolithic
England Neolithic
Germany EN LBK
Czech Neolithic
Austria EN LBK
Hungary EN Koros
Hungary EN Starcevo
Romania EN
Croatia EN Starcevo
Bulgaria EN Mal. Presl.
Croatia EN Impressa
Serbia EN
Serbia Starcevo EN
Iberia EN El Prado
Iberia EN Els Trocs
Croatia EN Cardial
Bulgaria EN
Iberia EN Fuente
Macedonia EN
Anatolia EN Barcin
Iberia EN Cova Bonica
Anatolia EN Kumtepe
Greece EN Revenia
Anatolia EN Tep. Cift.
Iberia EN Andalusia
Sicily EN
Greece EN Peloponnese
Anatolia EN Boncuklu
Iberia EN El Toro
F3
0.2900
0.2875
0.2850
0.2825
Latitude
Longitude
60
55
50
45
40
0
10
20
30
Croatia EN Starcevo
Croatia EN Impressa
Croatia EN Cardial
Greece EN Peloponnese
Austria EN LBK
Germany EN LBK
Serbia EN Starcevo
Hungary EN Koros
Romania EN
Anatolia EN Barcin
Iberia EN ElPrado
Serbia EN
Greece EN Revenia
Czech Neolithic
Hungary EN Starcevo
Iberia EN Cova Bonica
Bulgaria EN
Macedonia EN
Iberia EN Chaves
Scotland Neolithic
Iberia EN Els Trocs
Iberia EN Fuente
Wales Neolithic
England Neolithic
Anatolia Pinarbasi
Anatolia EN Boncuklu
Iberia EN Andalusia
Iberia EN El Toro
Bulgaria EN Mal. Presl.
Anatolia EN Kumtepe
Anatolia EN Tepecik Ciftlik
0.28
0.29
0.30
f3 (Mbuti; Sicily EN, X)

Fig. 4. Genomic affinity of the ancient Sicilians. (A) Comparing the ancestry in Sicily EM and LM HGs to various West Eurasian HGs (*X*), as measured by f_4*(Chimp, X; Sicily EM HGs, Sicily LM HGs*). Cooler colours indicate that *X* shares more genetic drift with Sicily EM HGs than with Sicily LM HGs, and warmer colours indicate the opposite. Dot sizes reflect |z|-scores. Not plotted: *AfontovaGora3* (f_4 = 0.0023, z = 3.73), *BerryAuBac* (f_4 = 0.0063, z = 5.10). Whereas Upper Palaeolithic and Mesolithic HGs from western Europe, including Villabruna cluster individuals, are genetically closer to Sicily EM HGs, those from eastern Europe are closer to Sicily LM HGs. **(B)** Early Neolithic Sicilian farmers show high genetic affinity to contemporaneous farmers from the Balkan (Croatia and Greece EN Peloponnese), as measured by f_3*(Mbuti; Sicily EN, X).* Warmer colours indicate higher levels of allele sharing. Error bars in the bar plot indicate 1 SE.



397

398 **Fig. 5. Genomic and dietary turnovers in Sicily during the Mesolithic and Early Neolithic.** The
399 coloured dots indicate the individuals' assigned genetic group with dark outlines those individuals that may
400 be contemporaneous to the earliest Impressa Ware aspects. **(A)** Individual ancestry profiles for the ancient
401 Sicilians determined from qpWave- and qpAdm-based ancestry models. The pre-Neolithic ancestry
402 proportion that is approximated by Sicily EM (*OrienteC/UZZ5054*) is in red, EHG in lime green, Pınarbaşı
403 in light brown, and the early farmer ancestry approximated by Anatolia EN Barcin in dark brown. The 5 cM
404 jackknifing standard errors are marked by horizontal bars. The mitogenomes haplogroups are given.
405 Compared to the preceding Sicily EM HGs, 15-37% of the ancestry in the Sicily LM HGs is from a non-
406 local source that is deeply related to both EHGs and Near Eastern HGs, such as Pınarbaşı. The Sicilian early
407 farmers contain almost entirely early farmer ancestry, indicating an almost complete ancestry replacement
408 during the Early Neolithic. **(B)** Isotope values for diet reconstruction of the ancient Sicilians. The values
409 for the Sicilian early farmers are indicative of a predominantly terrestrial-based farming diet. The Sicily
410 LM HGs associated with the Castelnovian *sensu lato* relied for around half of their protein on seafood.
411 *UZZ71* and *UZZ88*, two individuals that are tentatively contemporaneous to Impressa Ware aspects, show
412 dietary profiles that are strikingly different from the preceding Late Mesolithic Castelnovian HGs and later
413 Early Neolithic Stentinello farmers. *UZZ71* consumed a much higher proportion of freshwater protein,
414 similarly to what is recorded for Mesolithic HGs from the Iron Gates. *UZZ88* consumed more terrestrial
415 plants and less animal protein than the Sicilian early farmers.

Target	Source1	Source2	Source3	P_{Wave} (Sources)	P_{Adm} (Target + Sources)	Source1 (%)	Source2 (%)	Source3 (%)	SE1 (%)	SE2 (%)	SE3 (%)	Note
Sicily EM	Sicily LM	NA	NA	7.60E-34*	NA	NA	NA	NA	NA	NA	NA	A
Sicily LM	Sicily EM	EHG	Pınarbaşı	5.70E-113	0.12	75	15.5	9.5	1.6	2.4	2.8	B
Sicily LM	Sicily EM	EHG	Natufian	3.16E-172	0.17	75.2	17.3	7.5	1.6	1.9	2.1	
Sicily LM	Sicily EM	EHG	Iran EN Ganj Dareh	2.57E-135	0.09	78	15	7.1	1.4	2.7	2.2	
Sicily LM	Sicily EM	EHG	PPN	3.67E-218	0.15	76.1	16.5	7.4	1.5	2.1	2.1	
Sicily LM	Sicily EM	EHG	NA	2.19E-299	5.56E-03	78	22	NA	1.5	NA	NA	C
Sicily LM	Sicily EM	Pınarbaşı	NA	1.42E-224	4.72E-08	76.5	23.5	NA	1.7	NA	NA	
Sicily LM	Sicily EM	Natufian	NA	1.32E-310	2.03E-14	80.6	19.4	NA	1.9	NA	NA	
Sicily LM	Sicily EM	Iran EN Ganj Dareh	NA	0	4.68E-06	82.8	17.2	NA	1.3	NA	NA	
Sicily LM	Sicily EM	PPN	NA	0	9.44E-11	81.4	18.6	NA	1.5	NA	NA	
Sicily LM	Sicily EM	Serbia Iron Gates HG I	NA	7.17E-68	0.02	37.3	62.7	NA	2.9	NA	NA	D
Sicily LM	Sicily EM	Serbia Iron Gates HG II	NA	4.28E-68	0.05	35.4	64.6	NA	2.8	NA	NA	
Sicily LM	Sicily EM	Romania Iron Gates HG	NA	1.61E-69	4.56E-03	44.6	55.4	NA	2.9	NA	NA	
Sicily LM	Loschbour	EHG	NA	7.98E-255	2.90E-09	89.8	10.2	NA	1.8	NA	NA	E
Sicily LM	Loschbour	EHG	Pınarbaşı	3.07E-110	4.27E-09	88.1	7.4	4.6	2	2.6	3	
Sicily LM	Loschbour	EHG	Natufian	1.95E-194	1.60E-07	86.9	6.5	6.6	1.9	2.1	2	
Sicily LM	Loschbour	EHG	Iran EN Ganj Dareh	4.32E-125	1.59E-07	89.6	3.7	6.7	1.8	2.8	2.1	
Sicily LM	Loschbour	EHG	PPN	1.20E-213	4.15E-08	88	6.5	5.6	1.8	2.2	2	
Sicily EN	Anatolia EN Barcin	Sicily EM	NA	0	0.02	95.7	4.3	NA	1	NA	NA	F
Sicily EN	Anatolia EN Barcin	Sicily LM	NA	0	0.02	94.6	5.4	NA	1.2	NA	NA	
Sicily EN	Anatolia EN Barcin	Loschbour	NA	0	7.74E-03	95	5	NA	1.2	NA	NA	
Sicily EN	Anatolia EN Barcin	Serbia Iron Gates HG I	NA	0	0.01	93.8	6.2	NA	1.4	NA	NA	
Sicily EN	Anatolia EN Barcin	Serbia Iron Gates HG II	NA	0	0.01	93.9	6.1	NA	1.4	NA	NA	
Sicily EN	Anatolia EN Barcin	Romania Iron Gates HG	NA	0	0.03	92.6	7.4	NA	1.6	NA	NA	
Sicily EN	Anatolia EN Barcin	Iran EN Ganj Dareh	Sicily EM	1.02E-20	0.47	69.6	21.2	9.2	7.6	6.2	1.7	G
Sicily EN	Anatolia EN Barcin	Iran EN Ganj Dareh	Sicily LM	1.15E-24	0.34	71	18.6	10.5	7.3	5.7	1.9	
Sicily EN	Anatolia EN Barcin	Iran EN Ganj Dareh	Loschbour	7.39E-16	0.4	63	24.8	12.2	9.1	7	2.4	H
Sicily EN	Anatolia EN Barcin	Iran EN Ganj Dareh	Serbia Iron Gates HG I	7.47E-27	0.09	73.5	15.5	11	7.6	5.7	2.3	
Sicily EN	Anatolia EN Barcin	Iran EN Ganj Dareh	Serbia Iron Gates HG II	1.46E-26	0.15	72.8	16.2	11	7.6	5.7	2.2	
Sicily EN	Anatolia EN Barcin	Iran EN Ganj Dareh	Romania Iron Gates HG	4.01E-30	0.15	75	13.3	11.7	7.1	5.2	2.3	

416

Table 1. **Overview of key qpWave- and qpAdm-based ancestry models for Sicily EM HGs, Sicily LM HGs and Sicily EN farmers referred to in the main text.** For a more comprehensive overview, see Data file 7. For all qpAdm-based admixture models, the P_{Wave}-value for the *Sources* is small, indicating that that our used *Outgroups* can distinguish the ancestries between the *Sources*. Note A: The allele frequencies of the Sicily EM and LM HG gene pools do not fully overlap, and hence could not be fitted via one ancestry stream (P_{Wave}-value marked by * indicates a model for *Target* and *Source1*). B: A three-way mixture of Sicily EM, EHG and one of various Near Eastern sources results in a full ancestry fit to the Sicily LM HG gene pool. C: Modeling Sicily LM HGs either as a two-way mixture of Sicily EM and EHG or Near Eastern-related ancestry is rejected as a full model fit. D: Replacing the ancestry from EHG with that of Mesolithic Iron Gates HGs marginally improves the model fit. The proportion of assigned Sicily EM ancestry is lower compared to the model that uses EHG as a second ancestry source, due to the substantial amount of WHG ancestry in Iron Gates HGs ((*17*) and see Fig. 2B). E: Replacing the ancestry of Sicily EM with that of *Loschbour* in two or three-way mixtures are strongly rejected as full ancestry fits to the Sicily LM gene pool. F: A two-way mixture of the early farmer ancestry in Anatolia EN Barcin and a West-Eurasian HG source does not adequately fit the Sicily EN gene pool. G: Modelling the early farmer ancestry as a combination of Anatolia EN Barcin and an additional basal ancestry, approximated here by Iran EN Ganj Dareh, and the local preceding Sicilian Mesolithic HGs does result in a full fit to the Sicilian early farmer gene pool. H: The Sicily EN ancestry can also be adequately modelled using a non-local HG ancestry source in addition to early farmer ancestry as approximated by a combination Anatolia EN Barcin and Iran EN Ganj Dareh.

Materials and Methods

aDNA analysis. All pre-amplification laboratory work was performed in dedicated clean rooms (*82*) at the Max Planck Institute (MPI) for the Science of Human history (SHH) in Jena and MPI for Evolutionary Anthropology (EVA) in Leipzig, Germany. At the MPI-SHH the individuals were sampled for bone or tooth powder, originating from various skeletal elements (e.g. petrous, molars, teeth, humerus, phalange, tibia, see Data file 1). The outer layer of the skeletal elements was removed with high-pressured powdered aluminium oxide in a sandblasting instrument, and the element was irradiated with ultraviolet (UV) light for 15 minutes on all sides. The elements were then sampled using various strategies, including grinding with mortar and pestle or cutting and followed by drilling into denser regions (Data file 1). Subsequently, for each individual 1-8 extracts of 100uL were generated from ~50mg powder per extract, following a modified version of a silica-based DNA extraction method (*83*) described earlier (*50*) (Data file 1). At the MPI-SHH, 20uL undiluted extract aliquots were converted into double-indexed double stranded (ds-) libraries following established protocols (*40, 84*), some of them with a partial uracil-DNA glycosylase ('ds UDG-half') treatment (*85*) and others without ('ds non-UDG'). At the MPI-EVA, 30uL undiluted extract aliquot was converted into double-indexed single-stranded (ss-) libraries (*86*) with minor modifications detailed in (*87*), without UDG treatment ('ss non-UDG') (Data file 1). At the MPI-SHH, all the ds- and ss-libraries were shotgun sequenced to check for aDNA preservation, and subsequently enriched using in-solution capture probes following a modified version of (*19*) (described in (*25*)) for ~1240k single nucleotide polymorphisms (SNPs) in the nuclear genome (*17, 18, 54*) and independently for the complete mitogenome. Then the captured libraries were sequenced on an Illumina 224 HiSeq4000 platform using either a single end (1x75bp reads) or paired end configuration (2x50bp reads).

The sequenced reads were demultiplexed according to the expected index pair for each library, allowing one mismatch per 7 bp index, and subsequently processed using EAGER v1.92.21 (*88*). We used AdapterRemoval v2.2.0 (*89*) to clip adapters and Ns stretches of the reads. We merged paired end reads into a single sequence for regions with a minimal overlap of 30 bp, and single end reads smaller than 30 bp in length were discarded. The reads obtained from the nuclear capture were aligned against the human reference genome (hg19), and those from the mitogenome captured against the revised Cambridge Reference Sequence (rCRS). For mapping we used the Burrows-Wheeler Aligner (BWA v0.7.12) *aln* and *samse* programs (*90*) with a lenient stringency parameter of '-n 0.01' that allows more mismatches, and '-l 16500' to disable seeding. We excluded reads with Phred-scaled mapping quality (MAPQ) <25. Duplicate reads, identified by having identical strand orientation, start and end positions, were removed using DeDup v.0.12.1 (*88*).

aDNA authentication and quality control. We assessed the authenticity and contamination levels in our ancient DNA libraries (unmerged and merged per-individual) in several ways. First, we checked the cytosine deamination rates at the end of the reads (*91*) using DamageProfiler v0.3 (https://github.com/apeltzer/DamageProfiler). After merging the libraries for each individual, we observed 21-52% C>T mismatch rates at the first base in the terminal nucleotide at the 5'-end, an observation that is compatible with the presence of authentic ancient DNA molecules. Second, we tested for contamination of the nuclear genome in males based on the X-chromosomal polymorphism rate. We determined the genetic sex by calculating the X-rate (coverage of X-chromosomal SNPs/ coverage of autosomal SNPs) and Y-rate (coverage of Y-chromosomal SNPs/ coverage of autosomal SNPs) (*16*). Four individuals for which the libraries showed a Y-rate ≥ 0.49 we assigned the label 'male' and 14 individuals with Y-rates ≤ 0.07 as 'female'. The individual *UZZ26.cont* with an intermediate Y-rate of 0.17 we excluded from further genetic analyses. Then we tested for heterozygosity of the X-chromosome using ANGSD v0.910 (*92*) (≥ 200 X-chromosomal SNPs, covered at least twice (*16*). We found a nuclear contamination of 1.7-5.3% for the four male individuals (Data file 1), based on new Method1 (*93*). Third, we obtained two mtDNA contamination estimates for genetic males and females, using ContaMix v1.0.10 (*19*) and Schmutzi v1.0 (*94*) (Data file 1). Before running Schmutzi, we realigned the reads to the rCRS using CircularMapper v1.93.4 filtering with MAPQ < 30. After removing duplicate reads, we downsampled to ~30,000 reads per library. With Schmutzi we found low contamination estimates of 1-3% for all individuals with sufficient coverage (Data file 1). ContaMix returned estimates in the range of 0.0-5.6% for all individuals except for *UZZ69* (3.7-10.6%) and the lower coverage individual *UZZ096* (0.3-13.5%).

Dataset. For genotyping we extracted reads with high mapping quality (MAPQ ≥ 37) to the autosomes using samtools v1.3. The DNA damage plots indicated that misincorporations could extend up to 10 bp from the read termini in non-UDG treated and up to 3bp in UDG-half treated libraries. We hence clipped the reads accordingly, thereby removing G>A transitions from the terminal read ends in ds-libraries and C>T transitions in both ss- and ds-libraries. For each individual, we randomly chose a single base per SNP site as a pseudo-haploid genotype with our custom program 'pileupCaller' (https://github.com/stschiff/sequenceTools). We intersected our data with a global set of high-coverage genomes from the Simon Genome Diversity Project (SGDP) for ~1240k nuclear SNP positions (*21*), including previously reported ancient individuals from (*15-17, 21-51*). To minimize the effects of residual ancient DNA damage, we removed ~300k SNPs on CpG islands from the data set. CpG dinucleotides, where a cytosine is followed by a guanine nucleotide, are frequent targets of DNA methylation (*95*). Post-mortem cytosine deamination was shown to occur more frequently at methylated than unmethylated CpGs (*20*) resulting in excess of CpG → TpG conversions. The final data set includes 868,755 intersecting autosomal SNPs for which our newly reported individuals cover 53,352-796,174 SNP positions with an

497 average read depth per SNP of 0.09-9.39X (Data file 1). For principal component analyses (PCA) we
498 intersected our data and published ancient genomes with a panel of worldwide present-day populations,
499 genotyped on the Affymetrix Human Origins (HO) (*37, 57*). After filtering out CpG dinucleotides this data
500 set includes 441,774 SNPs.

501 **Kinship relatedness and individual assessment.** We determined pairwise mismatch rates (PMMRs) (*34,*
502 *96*) for pseudo-haploid genotypes to check for genetic duplicate individuals and first-degree relatives. If
503 two individuals show similar low PMMRs for inter- and intra-individual library comparisons, then this
504 indicates a genetic duplicate. Moreover, the expected PMMR for two first-degree related individuals falls
505 approximately in the middle of the baseline values for comparison between genetically unrelated and
506 identical individuals (*97*). We found a genetic triplicate (UZZ44, -45, -46) and quintuplicate (UZZ50-54),
507 and merged the respective libraries into *UZZ4446* and *UZZ5054*, respectively. In addition, *UZZ79* and
508 *UZZ81* showed an elevated PMMR indicative of a kinship relation (Data file 4).

509 **Mitogenome haplogroup determination.** We could reconstruct the mitochondrial genomes for 17
510 individuals (Data file 1). To obtain an automated mitochondrial haplogroup assignment we imported the
511 consensus sequences from Schmutzi into HaploGrep2 v2.1.1 ((*98*); available via:
512 https://haplogrep.uibk.ac.at/) based on phylotree (*99*) (mtDNA tree build 17, available via
513 http://www.phylotree.org/). In parallel, we manually haplotyped the reconstructed mitogenomes, based on
514 a procedure described in (*52*). We imported the bam.files for the merged libraries into Geneious v.9.0.5
515 (http://www.geneious.com) (*100*). After reassembling the reads against the revised Cambridge Reference
516 Sequence (rCRS) we called SNP variants with a minimum variant frequency of 0.7 and 2.0X coverage.
517 Using phylotree, we double-checked whether the called variants matched the expected diagnostic ones
518 based on the automated HaploGrep assignment. We did not consider known unstable nucleotide positions
519 309.1C(C), 315.1C, AC indels at 515-522, 16093C, 16182C, 16183C, 16193.1C(C) and 16519. We
520 extracted the consensus sequences based on a minimum of 75% base similarity. Using this approach, we
521 identified a total of twelve lineage-specific and private variants in the high coverage *UZZ5054* mitogenome.
522 Four of the lineage-specific variant positions were covered by only one or two reads in the low coverage
523 *UZZ96* and *OrienteC* genomes, and hence fell initially below our frequency threshold for variant detection.
524 However, since these variants were covered by a large number of reads in the closely related *UZZ5054*
525 mitogenome, for *UZZ96* and *OrienteC* we based the variant calls at these positions on the few reads
526 available and adjusted their consensus sequences accordingly (table S7.3).

527 **Y-chromosome haplogroup determination.** To determine the Y chromosome haplogroup for genetic
528 males we used the *yHaplo* program ((*101*), available via: https://github.com/23andMe/yhaplo). We based
529 our haplogroup assignment on 13,581 strand-unambiguous ancestry informative SNPs from the ISOGG

(International Society of Genetic Genealogy) data set. We called genotypes for these SNP sites by randomly choosing one allele with 'pileupCaller' (https://github.com/stschiff/sequenceTools). Using an in-house script (Choongwon Jeong) the genotypes were converted to an input format for *yHaplo*. In ancient genomes, missing data for key diagnostic sites may halt the automated search before the most derived haplogroup is reached. We therefore manually checked the coverage for informative SNPs further downstream for the haplogroup that was assigned to each individual. Also, our sequence reads from the non-UDG treated ss-libraries have an expected 3-5% C<->T mismatches due to residual ancient damage. C<->T mismatches at diagnostic SNP positions may result *yHaplo* to incorrectly assign a more derived haplogroup. We therefore checked whether the ancestral state variants matched those that are expected for the assigned haplogroup.

Principal component analysis (PCA). We computed principal components from individuals from 43 modern West Eurasian groups in the Human Origin panel (*37, 57*) using the *smartpca* program in the EIGENSOFT package v6.0.1 (*102*) with the parameter 'numeroutlieriter:0'. Ancient individuals were projected using 'lsqproject:YES' and 'shrinkmode:YES'.

***f*-statistics.** We performed *f*-statistics on the 1240k data set using ADMIXTOOLS (*57*). For f_3-outgroup statistics (*103*) we used *qp3Pop* and for f_4-statistics *qpDstat* with f4mode:YES. Standard errors (SEs) were determined using a weighted block jackknife over 5Mb blocks. F_3-outgroup statistics of the form $f_3(O;A,B)$ test the null hypothesis that O is a true outgroup to A and B. The strength of the f_3-statistic is a measure for the amount of genetic drift that A and B share after they branched off from a common ancestor with O, provided that A and B are not related by admixture. F_4-statistics of the form $f_4(X,Y; A,B)$ test the null hypothesis that the unrooted tree topology $((X,Y)(A,B))$, in which (X,Y) form a clade with regard to (A,B), reflects the true phylogeny. A positive value indicates that either X and A, or Y and B, share more drift than expected under the null hypothesis. A negative value indicates that the tree topology under the null-hypothesis is rejected into the other direction, due to more shared drift between Y and A, or X and B.

Multidimensional scaling (MDS). We performed MDS using the R package *cmdscale*. Euclidean distances were computed from the genetic distances among West-Eurasian HGs, as measured by $f_3(Mbuti; HG1, HG2)$ for all possible pairwise combinations (*16*). The first two principal components are plotted. We restricted the analyses to individuals with >30,000 autosomal SNPs covered. Relevant previously published West Eurasian HGs were pooled in groups according to their geographical or temporal context, following their initial publication labels (Data file 1). We deviated from the original labels with regard to Iron Gates HGs from Serbia by splitting them into an early and late subgroup, labelled here as 'Iron Gates HG Serbia I' (RC date: 10,000-7,500 calBCE) and 'Iron Gates HG Serbia II' (RC date: 7,500-5,700 calBCE) (*17*).

Nucleotide diversity. We selected a total of 120 West-Eurasian HGs with >100k SNPs covered, of which 103 were previously published (Data file 1), from four broad geographical regions "western" (n=18), "south-western" (n=7), "southern-central" (n=33), and "(south)-eastern" (n=62) Europe. We subgrouped the individuals further based on similar ^{14}C-dates and genetic cluster assignment (for an overview of the HG groups, see (Data file 1). E.g. we analysed the individuals associated with the Villabruna cluster and those high in Magdalenian-related ancestry in separate groups. We determined the nucleotide diversity (π) from pseudo-haploid genotypes by calculating the proportion of nucleotide mismatches for overlapping autosomal SNPs covered by at least one read in both individuals. We hence determined π from all possible combinations of individual pairs, rather than from all possible chromosome pairs, within a given group. We filtered out individual pairs that shared less than 30,000 SNPs covered. We calculated an average over all the individual pairs within a group and determined standard errors from block jackknifes over 5Mb windows and 95% confidence intervals (95CIs) from 1,000 bootstraps

Inference of mixture proportions. To characterize the ancestry of the ancient Sicilians we used the qpWave (*54*) and qpAdm (*36*) programs from the Admixtools v3.0 package, with the 'allsnps: YES' option. qpWave tests whether a set of *Left* populations is consistent with being related via as few as N streams of ancestry to a set of *Outgroup* populations. qpAdm tries to fit a *Target* as a linear combination of the *Left/Source* populations, and estimates the respective ancestry proportions that each of the *Left* populations contributed to the *Target*. Both qpWave and qpAdm are based on f_4-statistics of the form $f_4(X,O1;O2,O3)$, where *O1, O2, O3* are all the triplet combinations of the *Outgroup* populations, and *X* is a *Target* or *Left/Source* population. We used an *Outgroup* set from Mathieson et al. (*17*): *El Miron*, *Mota,* Mbuti, *Ust Ishim, Mal'ta, AfontovaGora3, GoyetQ116, Villabruna, Kostenki14, Vestonice16,* Karitiana, Papuan, Onge. We grouped individuals with a similar ancestry for Sicily EM, Sicily LM, EHG and CHG (Data file 1). Since missing data may inflate the P-values for this test, we required a test result to be smaller (less extreme) than P = 0.1 in order to reject the null-hypothesis of a full ancestry fit between the *Target* and the *Left/Source* population(s). Prior to running qpAdm we used qpWave to check whether the pairs of *Left/Source* populations are not equally related to the *Outgroups*.

Phylogeny modelling. We used the qpGraph program (*57*) to construct a phylogeny of ancestry lineages found among Palaeolithic and Mesolithic West-Eurasian HGs to clarify the genetic history of Sicily EM and LM HGs. For our modelling we used the parameters 'useallsnps: YES', 'forcezmode: YES', 'terse: NO'. We built the phylogeny models with increasing complexity by fitting representative West Eurasian HG ancestry lineages in the order: 1) Mbuti, 2) *Ust Ishim*, 3) *Kostenki14*, 4) *Mal'ta*, 5) *Vestonice16*, 6) *GoyetQ2,* 7) *AfontovaGora3*, 8) *Villabruna* or Sicily EM, 9) Sicily EM or *Villabruna* (Supplementary Section S5). We fitted the lineages one by one by adding them to all possible nodes as a branch without

admixture or as a binary admixture between two branches. We preferred the former over the latter if both models fitted the observed *f*-statistics equally well. We selected models that did not include trifurcations or 0% ancestry stream estimates, and for which the difference between the observed and fitted *f*-statistics were less extreme than 3.5 SEs.

Direct AMS ^{14}C bone dates. For 15 individuals we obtained a direct ^{14}C date from the skeletal element that was used for genetic analysis (Data file 1). All bone samples were pretreated at the Department of Human Evolution at the Max Planck Institute for Evolutionary Anthropology (MPI-EVA), Leipzig, Germany, using the method described in (*104*). For each skeletal element, 200-500mg of bone/tooth powder was decalcified in 0.5M HCl at room temperature ~4 hours until no CO_2 effervescence was observed. To remove humics, in a first step 0,1M NaOH was added for 30 minutes, followed by a final 0.5M HCl step for 15 minutes. The resulting solid was gelatinized following a protocol of (*105*) at pH 3 in a heater block at 75°C for 20h. The gelatin was then filtered in an Eeze-Filter™ (Elkay Laboratory Products (UK) Ltd.) to remove small (> 80mm) particles, and ultrafiltered (*106*) with Sartorius "VivaspinTurbo" 30 KDa ultrafilters. Prior to use, the filter was cleaned to remove carbon containing humectants (*107*). The samples were lyophilized for 48 hours. In order to monitor contamination introduced during the pre-treatment stage, a sample from a cave bear bone, kindly provided by D. Döppes (MAMS, Germany), was extracted along with the batch from La Ferrassie samples (*108*). In marine environments the radiocarbon is older than the true age, usually by ~400 years (marine reservoir effect). The specimens for which a correction was necessary are *UZZ4446* (40±10% marine), *UZZ81* (45±10% marine), *UZZ69*, *UZZ79* and *UZZ80* (50±10% marine). Corrections were made using the reservoir correction estimated for the Mediterranean Basin by (*109*), which is $\Delta R = 58 \pm 85$ ^{14}C yr.

Isotope analysis. For 14 individuals we determined the carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) isotope values for dietary inference (Data file 1). To assess the preservation of the collagen yield, C:N ratios, together with isotopic values are evaluated following the limits of (*110*).

80

H2: Supplementary Materials

Section S1. Grotta dell'Uzzo: archaeology and stratigraphic sequence

Section S2. Genetic grouping and substructure of the ancient Sicilians

Section S3. Elevated lineage-specific genetic drift in the Sicilian Early Mesolithic

Section S4. Characterizing the Sicilian Mesolithic HGs ancestry using F-statistics

Section S5. Investigating the phyologenetic position of the Early Mesolithic Sicilian HGs

Section S6. Characterizing the Sicilian early farmer ancestry using F-statistics

Section S7. Uniparental marker haplotyping

Data file S1. Summary table results ancient Sicilians, labels used in analyses, data in Supplementary Sections

Data file S2. Data underlying Fig. 1

Data file S3. Data underlying Fig. 2

Data file S4. Data underlying Fig. 3

Data file S5. Data underlying Fig. 4

Data file S6: Data underlying Fig. 5.

Data file S7. Admixture models for ancient Sicilians and West Eurasian HGs

635 **References and Notes**

636 1. D. Binder, P. Lanos, L. Angeli, L. Gomart, J. Guilaine, C. Manen, R. Maggi, I. M. Muntoni, C.
637 Panelli, G. Radi, C. Tozzi, D. Arobba, J. Battentier, M. Brandaglia, L. Bouby, F. Briois, A. Carré, C.
638 Delhon, L. Gourichon, P. Marinval, R. Nisbet, S. Rossi, P. Rowley-Conwy, S. Thiébault, Modelling the
639 earliest north-western dispersal of Mediterranean Impressed Wares: new dates and Bayesian chronological
640 model. *Documenta Praehistorica* 44, 54-77 (2017).

641 2. O. García-Puchol, A. A. Diez Castillo, S. Pardo-Gordo, *Timing the Western Mediterranean Last*
642 *Hunter-Gatherers and First Farmers*. O. García-Puchol, D. C. Salazar García, Eds., Times of Neolithic
643 Transition along the Western Mediterranean, Fundamental Issues in Archaeology (Springer, New York,
644 2017).

645 3. Í. García-Martínez de Lagrán, Recent Data and Approaches on the Neolithization of the Iberian
646 Peninsula. *European Journal of Archaeology* 18, 429-453 (2015).

647 4. H. Martins, F. X. Oms, L. Pereira, A. W. G. Pike, K. Rowsell, J. Zilhão, Radiocarbon Dating the
648 Beginning of the Neolithic in Iberia. *Journal of Mediterranean Archaeology* 28, 105-131 (2015).

649 5. J. Zilhão, Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west
650 Mediterranean Europe. *PNAS* 98, 14180-14185 (2001).

651 6. J. Guilaine, *Premiers bergers et paysans de l'Occident méditerranéen*. Études rurales (Mouton,
652 Paris, 1976), pp. 145.

653 7. J. Guilaine, A personal view of the neolithisation of the Western Mediterranean. *Quaternary*
654 *International* 470 211-225 (2018).

655 8. E. Natali, V. Forgia, The beginning of the Neolithic in Southern Italy and Sicily. *Quaternary*
656 *International* 470, 253-269 (2018).

657 9. S. Forenbaher, P. T. Miracle, The spread of farming in the Eastern Adriatic. *Documenta*
658 *Praehistorica* 33, 89-100 (2006).

659 10. M. A. Fugazzola Delpino, V. Tiné, Le statuine fittili femminili del Neolitico italiano. Iconografia
660 e contesto culturale. *Bullettino di Paletnologia Italiana (Roma)* 93-94, 19-51 (2002-2003).

661 11. C. A. T. Malone, The Italian Neolithic: A Synthesis of Research. *Journal of World Prehistory* 17,
662 235-312 (2003).

12. M. A. Mannino, S. Talamo, A. Tagliacozzo, I. Fiore, O. Nehlich, M. Piperno, S. Tusa, C. Collina, R. Di Salvo, V. Schimmenti, M. P. Richards, Climate-driven environmental changes around 8.200 years ago favoured increases in cetacean strandings and Mediterranean hunter-gatherers exploited them. *Sci. Rep.* 5, 1-12 (2015).

13. M. A. Mannino, K. D. Thomas, M. J. Leng, M. Piperno, S. Tusa, A. Tagliacozzo, Marine resources in the Mesolithic and Neolithic at the Grotta dell'Uzzo (Sicily): evidence from isotope analyses of marine shells. *Archaeometry* 49, 117-133 (2007).

14. A. Tagliacozzo, *Archeozoologia della Grotta dell'Uzzo, Sicilia. Da un'economia di caccia ad un'economia di pesca ed allevamento*. I. P. e. Z. d. Stato, Ed., Supplemento al Bullettino di Paletnologia Italiana (Istituto poligrafico e zecca dello Stato, Roma, 1993), vol. 84.

15. G. Catalano, P. F. Fabbri, D. Lo Vetro, S. Mallick, Late Upper Palaeolithic hunter-gatherers in the Central Mediterranean: new archaeological and genetic data from the Late Epigravettian burial Oriente C (Favignana, Sicily). *bioRxiv*, (2019).

16. Q. Fu, C. Posth, M. Hajdinjak, M. Petr, S. Mallick, D. Fernandes, A. Furtwängler, W. Haak, M. Meyer, A. Mittnik, B. Nickel, A. Peltzer, N. Rohland, V. Slon, S. Talamo, I. Lazaridis, M. Lipson, I. Mathieson, S. Schiffels, P. Skoglund, A. P. Derevianko, N. Drozdov, V. Slavinsky, A. Tsyvabkov, R. Grifoni Cremonesi, F. Mallegni, B. Gély, E. Vacca, M. R. González Morales, L. G. Straus, C. Neugebauer-Maresch, M. Teschler-Nicola, S. Constantin, O. T. Moldovan, S. Benazzi, M. Peresani, D. Coppola, M. Lari, S. Ricci, A. Ronchitelli, F. Valentin, C. Thevenet, K. Wehrberger, D. Grigorescu, H. Rougier, I. Crevecoeur, D. Flas, P. Semal, M. A. Mannino, C. Cupillard, H. Bocherens, N. J. Conard, K. Harvati, V. Moiseyev, D. G. Drucker, J. Svoboda, M. P. Richards, D. Caramelli, R. Pinhasi, J. Kelso, N. Patterson, J. Krause, S. Pääbo, D. Reich, The genetic history of Ice Age Europe. *Nature* 534, 200-205 (2016).

17. I. Mathieson, S. Alpaslan-Roodenberg, C. Posth, A. Szécsényi-Nagy, N. Rohland, S. Mallick, I. Olalde, N. Broomandkhoshbacht, F. Candilio, O. Cheronet, D. Fernandes, M. Ferry, B. Gamarra, G. Fortes, W. Haak, E. Harney, E. Jones, D. Keating, B. Krause-Kyora, I. Kucukkalipci, M. Michel, A. Mittnik, K. Nägele, M. Novak, J. Oppenheimer, N. Patterson, S. Pfrengle, K. Sirak, K. Stewardson, S. Vai, S. Alexandrov, K. Alt, R. Andreescu, D. Antonović, A. Ash, N. Atanassova, K. Bacvarov, M. Gusztáv, H. Bocherens, M. Bolus, A. Boroneanţ, Y. Boyadzhiev, A. Budnik, J. Burmaz, S. Chohadzhiev, N. Conard, R. Cottiaux, M. Čuka, C. Cupillard, D. Drucker, N. Elenski, M. Francken, B. Galabova, G. Ganetsovski, B. Gély, T. Hajdu, V. Handzhyiska, K. Harvati, T. Higham, S. Iliev, I. Janković, I. Karavanić, D. Kennett, D. Komšo, A. Kozak, D. Labuda, M. Lari, C. Lazar, M. Leppek, K. Leshtakov, D. Vetro, D. Los, I. Lozanov, M. Malina, F. Martini, K. McSweeney, H. Meller, M. Menđušić, P. Mirea, V. Moiseyev, V.

695 Petrova, T. Price, A. Simalcsik, L. Sineo, M. Šlaus, V. Slavchev, P. Stanev, A. Starović, T. Szeniczey, S.
696 Talamo, M. Teschler-Nicola, C. Thevenet, I. Valchev, F. Valentin, S. Vasilyev, F. Veljanovska, S.
697 Venelinova, E. Veselovskaya, B. Viola, C. Virag, J. Zaninović, S. Zäuner, P. Stockhammer, G. Catalano,
698 R. Krauß, D. Caramelli, G. Zariņa, B. Gaydarska, M. Lillie, A. Nikitin, I. Potekhina, A. Papathanasiou, D.
699 Borić, C. Bonsall, J. Krause, R. Pinhasi, D. Reich, The genomic history of southeastern Europe. *Nature*
700 555, 197-203 (2018).

701 18. Q. Fu, M. Hajdinjak, O. T. Moldovan, S. Constantin, S. Mallick, P. Skoglund, N. Patterson, N.
702 Rohland, I. Lazaridis, B. Nickel, T. B. Viola, K. Prüfer, M. Meyer, J. Kelso, D. Reich, S. Pääbo, An early
703 modern human from Romania with a recent Neanderthal ancestor. *Nature* 524, 216-219 (2015).

704 19. Q. Fu, M. Meyer, X. Gao, U. Stenzel, H. A. Burbano, J. Kelso, S. Pääbo, DNA analysis of an early
705 modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U.S.A* **110**, 2223-2227 (2013).

706 20. A. Seguin-Orlando, C. A. Hoover, S. K. Vasiliev, N. D. Ovodov, B. Shapiro, A. Cooper, E. M.
707 Rubin, E. Willerslev, L. Orlando, Amplification of TruSeq ancient DNA libraries with AccuPrime Pfx:
708 consequences on nucleotide misincorporation and methylation patterns. *STAR* 1, 1-9 (2015).

709 21. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S.
710 Nordenfelt, A. Tandon, P. Skoglund, Lazaridis I, Sankararaman S, Q. Fu, N. Rohland, G. Renaud, Y. Erlich,
711 T. F. Willems, C. Gallo, J. Spence, Y. Song, G. Poletti, F. Balloux, G. van Driem, P. de Knijff, I. Romero,
712 A. Jha, D. Behar, C. Bravi, C. Capelli, T. Hervig, A. Moreno-Estrada, O. Posukh, E. Balanovska, O.
713 Balanovsky, S. Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y.
714 Xue, M. Abdullah, A. Ruiz-Linares, C. Beall, A. Di Rienzo, C. Jeong, E. Starikovskaya, E. Metspalu, J.
715 Parik, R. Villems, B. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G. Stamatoyannopoulos, J. Wee, R.
716 Khusainova, E. Khusnutdinova, S. Litvinov, G. Ayodo, D. Comas, M. Hammer, T. Kivisild, W. Klitz, C.
717 Winkler, D. Labuda, M. Bamshad, L. Jorde, S. A. Tishkoff, W. Watkins, M. Metspalu, S. Dryomov, R.
718 Sukernik, L. Singh, K. Thangaraj, S. Pääbo, J. Kelso, N. Patterson, D. Reich, The Simons Genome Diversity
719 Project: 300 genomes from 142 diverse populations. *Nature* 538, 201-206 (2016).

720 22. M. Allentoft, Sikora M, S. KG, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder
721 H, Ahlström T, Vinner L, Malaspinas AS, Margaryan A, Higham T, Chivall D, Lynnerup N, Harvig L,
722 Baron J, Della Casa P, Dąbrowski P, Duffy PR, Ebel AV, Epimakhov A, Frei K, Furmanek, Gralak T,
723 Gromov A, Gronkiewicz S, Grupe G, Hajdu T, Jarysz R, Khartanovich V, Khokhlov A, Kiss V, Kolář J,
724 Kriiska A, Lasak I, Longhi C, McGlynn G, Merkevicius A, Merkyte I, Metspalu M, Mkrtchyan R,
725 Moiseyev V, Paja L, Pálfi G, Pokutta D, Pospieszny Ł, Price TD, Saag L, Sablin M, Shishlina N, Smrčka
726 V, Soenov VI, Szeverényi V, Tóth G, Trifanova SV, Varul L, Vicze M, Yepiskoposyan L, Zhitenev V,

727 Orlando L, Sicheritz-Pontén T, Brunak S, Nielsen R, K. K, W. E, Population genomics of Bronze Age
728 Eurasia. *Nature* 522, 167-172 (2015).

729 23. F. Broushaki, M. G. Thomas, V. Link, S. López, L. van Dorp, K. Kirsanow, J. Burger, Early
730 Neolithic genomes from the eastern Fertile Crescent. *Science* 353, 499-503 (2016).

731 24. P. de Barros Damgaard, R. Martiniano, J. Kamm, J. V. Moreno-Mayar, G. Kroonen, M. Peyrot, G.
732 Barjamovic, S. Rasmussen, C. Zacho, N. Baimukhanov, V. Zaibert, V. Merz, A. Biddanda, I. Merz, V.
733 Loman, V. Evdokimov, E. Usmanova, B. Hemphill, A. Seguin-Orlando, F. E. Yediay, I. Ullah, K. G.
734 Sjögren, K. H. Iversen, J. Choin, C. de la Fuente, M. Ilardo, H. Schroeder, V. Moiseyev, A. Gromov, A.
735 Polyakov, S. Omura, S. Y. Senyurt, H. Ahmad, C. McKenzie, A. Margaryan, A. Hameed, A. Samad, N.
736 Gul, M. H. Khokhar, O. I. Goriunova, V. I. Bazaliiskii, J. Novembre, A. W. Weber, L. Orlando, M. E.
737 Allentoft, R. Nielsen, K. Kristiansen, M. Sikora, A. K. Outram, R. Durbin, E. Willerslev, The first horse
738 herders and the impact of early Bronze Age steppe expansions into Asia. *Science* 360, eaar7711 (2018).

739 25. M. Feldman, E. Fernández-Domínguez, L. Reynolds, D. Baird, J. Pearson, I. Hershkovitz, H. May,
740 N. Goring-Morris, M. Benz, J. Gresky, R. Bianco, A. Fairbairn, G. Mustafaoğlu, P. W. Stockhammer, C.
741 Posth, W. Haak, C. Jeong, J. Krause, Late Pleistocene human genome suggests a local origin for the first
742 farmers of central Anatolia. *Nat. Commun.* 10, 1218 (2019).

743 26. R. Fregel, F. Méndez, Y. Bokbot, D. Martín-Socas, M. Camalich-Massieu, M. J. Santana J5, M.
744 Ávila-Arcos, P. Underhill, B. Shapiro, G. Wojcik, M. Rasmussen, A. Soares, J. Kapp, A. Sockell, F.
745 Rodríguez-Santos, A. Mikdad, A. Trujillo-Mederos, C. D. Bustamante, Ancient genomes from North Africa
746 evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proc Natl Acad Sci U S*
747 *A* 115, 6774-6779 (2018).

748 27. Q. Fu, H. Li, P. Moorjani, F. Jay, S. Slepchenko, A. Bondarev, P. Johnson, A. Aximu-Petri, K.
749 Prüfer, C. de Filippo, M. Meyer, N. Zwyns, D. Salazar-García, Y. Kuzmin, S. Keates, P. Kosintsev, D.
750 Razhev, M. Richards, N. Peristov, M. Lachmann, K. Douka, T. Higham, M. Slatkin, J. Hublin, D. Reich,
751 J. Kelso, T. Viola, S. Pääbo, Genome sequence of a 45,000-year-old modern human from western Siberia.
752 Nature. . *Nature* 514, 445-459 (2014).

753 28. M. Gallego Llorente, E. R. Jones, A. Eriksson, V. Siska, K. Arthur, M. Curtis, J. T. Stock, M.
754 Coltorti, P. Pieruccini, S. Stretton, F. Brock, T. F. G. Higham, Y. Park, M. Hoffreiter, D. G. Bradley, J.
755 Bhak, R. Pinhasi, A. Manica, Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern
756 Africa. *Science* 350, 820-822 (2015).

29. C. Gamba, E. R. Jones, M. D. Teasdale, R. McLaughlin, L, G. Gonzales-Fortes, V. Mattiangeli, L. Domboróczki, I. Kővári, I. Pap, A. Anders, A. Whittle, J. Dani, P. Raczky, T. F. G. Higham, M. Hofreiter, D. G. Bradley, R. Pinhasi, Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* 5, 5257-5265 (2014).

30. Z. Hofmanová, S. Kreutzer, G. Hellenthal, C. Sell, Y. Diekmann, D. Díez-del-Molino, L. van Dorp, S. López, A. Kousathanas, A. Link, K. Kirsanow, L. M. Cassidy, R. Martiniano, M. Strobel, A. Scheu, A. Kotsakis, P. Halstead, P. Triantaphyllou, N. Kyparissi-Apostolika, N. Urem-Kotsou, C. Ziota, F. Adaktylou, F. Gopalan, D. M. Bobo, L. Winkelbach, J. Blöcher, M. Unterländer, C. Leuenberger, Ç. Çilingiroğlu, B. Horejs, B. Gerritsen, S. J. Shennan, D. G. Bradley, M. Currat, K. R. Veeramah, D. Wegmann, M. G. Thomas, C. Papageorgopoulou, J. Burger, Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci. USA* 113, 6886-6891 (2016).

31. E. Jones, G. Zarina, V. Moiseyev, E. Lightfoot, P. Nigst, A. Manica, R. Pinhasi, D. Bradley, The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. *Curr Biol.* 27, 576-582 (2017).

32. E. R. Jones, G. Gonzalez-Fortes, S. Connell, V. Siska, A. Eriksson, R. Martiniano, R. McLaughlin, L, M. Gallego Llorente, L. Cassidy, C. Gamba, T. Meshveliani, O. Bar-Yosef, W. Müller, A. Belfer-Cohen, Z. Matskevich, N. Jakeli, T. F. G. Higham, M. Currat, D. Lordkipanidze, M. Hofreiter, A. Manica, R. Pinhasi, D. G. Bradley, Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* 6, 8912-8919 (2015).

33. A. Keller, A. Graefen, M. Ball, M. Matzas, V. Boisguerin, F. Maixner, P. Leidinger, C. Backes, R. Khairat, M. Forster, B. Stade, A. Franke, J. Mayer, J. Spangler, S. McLaughlin, M. Shah, C. Lee, T. T. Harkins, A. Sartori, A. Moreno-Estrada, B. Henn, M. Sikora, O. Semino, J. Chiaroni, S. Rootsi, N. M. Myres, V. M. Cabrera, P. A. Underhill, C. D. Bustamante, E. E. Vigl, M. Samadelli, G. Cipollini, J. Haas, H. Katus, B. D. O'Connor, M. R. Carlson, B. Meder, N. Blin, E. Meese, C. M. Pusch, A. Zink, New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* 3, 698 (2012).

34. G. Kılınç, A. Omrak, F. Özer, T. Günther, A. Büyükkarakaya, E. Bıçakçı, D. Baird, H. Dönertaş, A. Ghalichi, R. Yaka, D. Koptekin, S. Açan, P. Parvizi, M. Krzewińska, E. Daskalaki, E. Yüncü, N. Dağtaş, A. Fairbairn, J. Pearson, G. Mustafaoğlu, Y. Erdal, Y. Çakan, İ. Togan, M. Somel, J. Storå, M. Jakobsson, A. Götherström, The Demographic Development of the First Farmers in Anatolia. *Curr Biol.* 26, 2659-2666 (2016).

788 35. I. Lazaridis, A. Mittnik, N. Patterson, S. Mallick, N. Rohland, S. Pfrengle, A. Furtwängler, A.
789 Peltzer, C. Posth, A. Vasilakis, P. J. P. McGeorge, E. Konsolaki-Yannopoulou, G. Korres, H. Martlew, M.
790 Michalodimitrakis, M. Özsait, N. Özsait, A. Papathanasiou, M. Richards, S. A. Roodenberg, Y. Tzedakis,
791 R. Arnott, D. M. Fernandes, J. R. Hughey, D. M. Lotakis, P. A. Navas, Y. Maniatis, J. A.
792 Stamatoyannopoulos, K. Stewardson, P. Stockhammer, R. Pinhasi, D. Reich, J. Krause, G.
793 Stamatoyannopoulos, Genetic origins of the Minoans and Mycenaeans. *Nature* 548, 214-218 (2017).

794 36. I. Lazaridis, D. Nadel, G. Rollefson, D. Merrett, N. Rohland, S. Mallick, D. Fernandes, M. Novak,
795 B. Gamarra, K. Sirak, S. Connell, K. Stewardson, E. Harney, Q. Fu, G. Gonzales-Fortes, E. R. Jones, S.
796 Roodenberg, G. Lengyel, F. Bocquentin, B. Gasparian, J. Monge, M. Gregg, V. Eshed, A. Mizrahi, C.
797 Meiklejohn, F. Gerritsen, L. Bejenaru, M. Blüher, A. Campbell, G. Cavellari, D. Comas, P. Froguel, E.
798 Gilbert, S. Kerr, P. Kovacs, J. Krause, D. McGettigan, M. Merrigan, D. Merriwether, S. O'Reilly, M. B.
799 Richards, O. Semino, M. Shamoon-Pour, G. Stefanescu, M. Stumvoll, A. Tönjes, A. Torroni, J. Wilson, L.
800 Yengo, N. Hovhannisyan, N. Patterson, R. Pinhasi, D. Reich, Genomic insights into the origin of farming
801 in the ancient Near East. *Nature* 536, 419–424 (2016).

802 37. I. Lazaridis, N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. Sudmant, J.
803 Schraiber, S. Castellano, M. Lipson, B. Berger, C. Economou, R. Bollongino, Q. Fu, K. I. Bos, S.
804 Nordenfelt, H. Li, C. de Filippo, K. Prüfer, S. Sawyer, C. Posth, W. Haak, F. Hallgren, E. Fornander, N.
805 Rohland, D. Delsate, M. Francken, J. Guinet, J. Wahl, G. Ayodo, H. Babiker, G. Baillet, E. Balanovska, O.
806 Balanovsky, R. Barrantes, G. Bedoya, H. Ben-Ami, J. Bene, F. Berrada, C. Bravi, F. Brisighelli, G. Busby,
807 F. Cali, M. Churnosov, D. Cole, D. Corach, van Driem G, L. Damba, Dryomov S, Dugoujon JM, Fedorova
808 SA, Gallego Romero I, Gubina M, Hammer M, Henn BM, Hervig T, Hodoglugil U, Jha AR, Karachanak-
809 Yankova S, Khusainova R, Khusnutdinova E, Kittles R, Kivisild T, Klitz W, Kučinskas V, Kushniarevich
810 A, Laredj L, Litvinov S, Loukidis T, Mahley RW, Melegh B, Metspalu E, Molina J, Mountain J,
811 Näkkäläjärvi K, Nesheva D, Nyambo T, Osipova L, Parik J, Platonov F, Posukh O, Romano V,
812 Rothhammer F, Rudan I, Ruizbakiev R, Sahakyan H, Sajantila A, Salas A, Starikovskaya EB, Tarekegn A,
813 Toncheva D, Turdikulova S, Uktveryte I, Utevska O, Vasquez R, Villena M, Voevoda M, Winkler CA,
814 Yepiskoposyan L, Zalloua P, Zemunik T, Cooper A, Capelli C, Thomas MG, Ruiz-Linares A, Tishkoff SA,
815 Singh L, Thangaraj K, Villems R, Comas D, Sukernik R, Metspalu M, Meyer M, Eichler EE, Burger J, M.
816 Slatkin, Pääbo S, Kelso J, Reich D, K. J., Ancient human genomes suggest three ancestral populations for
817 present-day Europeans. *Nature* 513, 409–413 (2014).

818 38. M. Lipson, A. Szécsényi-Nagy, S. Mallick, A. Pósa, B. Stégmár, V. Keerl, N. Rohland, K.
819 Stewardson, M. Ferry, M. Michel, J. Oppenheimer, N. Broomandkhoshbacht, E. Harney, S. Nordenfelt, B.
820 Llamas, B. Gusztáv Mende, K. Köhler, K. Oross, M. Bondár, T. Marton, A. Osztás, J. Jakucs, T. Paluch,

F. Horváth, P. Csengeri, J. Koós, K. Sebők, A. Anders, P. Raczky, J. Regenye, J. Barna, S. Fábián, G. Serlegi, Z. Toldi, E. Gyöngyvér Nagy, J. Dani, E. Molnár, G. Pálfi, L. Márk, B. Melegh, Z. Bánfai, L. Domboróczki, J. Fernández-Eraso, J. Antonio Mujika-Alustiza, C. Alonso Fernández, J. Jiménez Echevarría, R. Bollongino, J. Orschiedt, K. Schierhold, H. Meller, A. Cooper, J. Burger, E. Bánffy, K. Alt, C. Lalueza-Fox, W. Haak, D. Reich, Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* 551, 368-372 (2017).

39. I. Mathieson, I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, H. Eadaoin, K. Stewardson, D. Fernandes, M. Novak, K. Sirak, C. Gamba, E. R. Jones, B. Llamas, S. Dryomov, J. Pickrell, J. L. Arsuaga, J. M. Bermúdez de Castro, E. Carbonell, F. Gerritsen, A. Khokhlov, P. Kuznetsov, M. Lozano, H. Meller, O. Mochalov, V. Moiseyev, M. A. R. Guerra, J. Roodenberg, J. M. Vergès, J. Krause, A. Cooper, K. W. Alt, D. Brown, D. Anthony, C. Lalueza-Fox, W. Haak, R. Pinhasi, D. Reich, Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528, 499–503 (2015).

40. M. Meyer, M. Kircher, M.-T. Gansauge, H. Li, F. Racimo, S. Mallick, J. Schraiber, F. Jay, K. Prüfer, C. de Filippo, P. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R. Green, K. Bryc, A. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M. Hammer, M. Shunkov, A. Derevianko, N. Patterson, A. Andrés, E. Eichler, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338, 222-226 (2012).

41. A. Mittnik, C. C. Wang, S. Pfrengle, M. Daubaras, G. Zariņa, F. Hallgren, R. Allmäe, V. Khartanovich, V. Moiseyev, M. Tõrv, A. Furtwängler, V. Andrades, A, M. Feldman, C. Economou, M. Oinonen, A. Vasks, E. Balanovska, D. Reich, R. Jankauskas, W. Haak, S. Schiffels, J. Krause, The genetic prehistory of the Baltic Sea region. *Nat Commun.* 9, (2018).

42. I. Olalde, M. E. Allentoft, F. Sánchez-Quinto, G. Santpere, C. W. Chiang, M. DeGiorgio, J. Prado-Martinez, J. A. Rodríguez, S. Rasmussen, J. Quilez, O. Ramírez, U. M. Marigorta, M. Fernández-Callejo, M. E. Prada, J. M. Encinas, R. Nielsen, M. G. Netea, J. Novembre, R. A. Sturm, P. Sabeti, T. Marquès-Bonet, A. Navarro, E. Willerslev, C. Lalueza-Fox, Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature.* 507, 225-228 (2014).

43. I. Olalde, S. Brace, M. E. Allentoft, I. Armit, K. Kristiansen, T. Booth, N. Rohland, Mallick, A. Szécsényi-Nagy, A. Mittnik, E. Altena, M. Lipson, I. Lazaridis, T. K. Harper, N. Patterson, N. Broomandkhoshbacht, Y. Diekmann, Z. Faltyskova, D. Fernandes, M. Ferry, E. Harney, P. de Knijff, M. Michel, J. Oppenheimer, K. Stewardson, A. Barclay, K. W. Alt, C. Liesau, P. Ríos, C. Blasco, J. V. Miguel, R. M. García, A. A. Fernández, E. Bánffy, M. Bernabò-Brea, D. Billoin, C. Bonsall, L. Bonsall, T. Allen, L. Büster, S. Carver, L. C. Navarro, O. E. Craig, G. T. Cook, B. Cunliffe, A. Denaire, K. E. Dinwiddy, N.

Dodwell, M. Ernée, C. Evans, M. Kuchařík, J. F. Farré, C. Fowler, M. Gazenbeek, R. G. Pena, M. Haber-Uriarte, E. Haduch, G. Hey, N. Jowett, T. Knowles, K. Massy, S. Pfrengle, P. Lefranc, O. Lemercier, A. Lefebvre, C. H. Martínez, V. G. Olmo, A. B. Ramírez, J. L. Maurandi, T. Majó, J. I. McKinley, K. McSweeney, B. G. Mende, A. Modi, G. Kulcsár, V. Kiss, A. Czene, R. Patay, A. Endrődi, K. Köhler, T. Hajdu, T. Szeniczey, J. Dani, Z. Bernert, M. Hoole, O. Cheronet, D. Keating, P. Velemínský, M. Dobeš, F. Candilio, F. Brown, R. F. Fernández, A. M. Herrero-Corral, S. Tusa, E. Carnieri, L. Lentini, A. Valenti, A. Zanini, C. Waddington, G. Delibes, E. Guerra-Doce, B. Neil, M. Brittain, M. Luke, R. Mortimer, J. Desideri, M. Besse, G. Brücken, M. Furmanek, A. Hałuszko, M. Mackiewicz, A. Rapiński, S. Leach, I. Soriano, K. T. Lillios, J. L. Cardoso, M. P. Pearson, P. Włodarczak, T. D. Price, P. Prieto, P. J. Rey, R. Risch, G. Rojo, M. A, A. Schmitt, J. Serralongue, A. M. Silva, V. Smrčka, L. Vergnaud, J. Zilhão, D. Caramelli, T. Higham, M. G. Thomas, D. J. Kennett, H. Fokkens, V. Heyd, A. Sheridan, K. G. Sjögren, P. W. Stockhammer, J. Krause, R. Pinhasi, W. Haak, I. Barnes, C. Lalueza-Fox, D. Reich, The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* 555, 190-119 (2018).

44. I. Olalde, H. Schroeder, M. Sandoval-Velasco, L. Vinner, I. Lobón, O. Ramirez, S. Civit, B. García, P, D. C. Salazar-García, S. Talamo, F. María, J, O. Xavier, F, M. Pedro, P. Martínez, M. Sanz, J. Daura, J. Zilhão, T. Marquès-Bonet, M. T. Gilbert, C. Lalueza-Fox, A common genetic origin for early farmers from Mediterranean Cardial and Central European LBK cultures. *Mol Biol Evol.* 32, 3132-3142 (2015).

45. A. Omrak, T. Günther, C. Valdiosera, E. M. Svensson, H. Malmström , H. Kiesewetter, W. Aylward, J. Storå, M. Jakobsson, A. Götherström, Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. *Curr Biol.* 26, 270-275 (2016).

46. M. Raghavan, P. Skoglund, K. E. Graf, M. Metspalu, A. Albrechtsen, I. Moltke, M. Rasmussen, T. J. Stafford, L. Orlando, E. Metspalu, M. Karmin, K. Tambets, S. Rootsi, R. Mägi, P. Campos, E. Balanovska, O. Balanovsky, E. Khusnutdinova, S. Litvinov, L. Osipova, S. Fedorova, M. Voevoda, M. DeGiorgio, T. Sicheritz-Ponten, S. Brunak, S. Demeshchenko, T. Kivisild, R. Villems, R. Nielsen, M. Jakobsson, E. Willerslev, Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505, 87–91 (2014).

47. M. Sikora, A. Seguin-Orlando, V. C. Sousa, A. Albrechtsen, T. S. Korneliussen, A. Ko, S. Rasmussen, I. Dupanloup, P. R. Nigst, M. D. Bosch, G. Renaud, M. E. Allentoft, A. Margaryan, S. V. Vasilyev, E. V. Veselovskaya, S. B. Borutskaya, T. Deviese, D. Comeskey, T. Higham, A. Manica, R. Foley, D. J. Meltzer, R. Nielsen, L. Excoffier, M. Mirazon-Lahr, L. Orlando, E. Willerslev, Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* 358, 659-662 (2017).

885 48. P. Skoglund, H. Malmström, A. Omrak, M. Raghavan, C. Valdiosera, T. Günther, P. Hall, K.
886 Tambets, J. Parik, K. Sjögren, J. Apel, E. Willerslev, J. Storå, A. Götherström, M. Jakobsson, Genomic
887 diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* 344, 747-750
888 (2014).

889 49. C. Valdiosera, T. Günther, J. C. Vera-Rodríguez, I. Ureña, E. Iriarte, R. Rodríguez-Varela, L. G.
890 Simões, R. M. Martínez-Sánchez, E. M. Svensson, H. Malmström, L. Rodríguez, d. Bermúdez, Castro, J.M,
891 E. Carbonell, A. Alday, V. Hernández, J.A, A. Götherström, J. M. Carretero, J. L. Arsuaga, C. I. Smith, M.
892 Jakobsson, Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations
893 at the far end of Eurasia. *Proc Natl Acad Sci U S A.* 115, 3428-3433 (2018).

894 50. V. Villalba-Mouco, M. S. van de Loosdrecht, C. Posth, R. Mora, J. Martínez-Moreno, M. Rojo-
895 Guerra, D. C. Salazar-García, J. I. Royo-Guillén, M. Kunst, H. Rougier, I. Crevecoeur, H. Arcusa-
896 Magallón, C. Tejedor-Rodríguez, I. García-Martínez de Lagrán, R. Garrido-Pena, K. W. Alt, C. Jeong, S.
897 Schiffels, P. Utrilla, J. Krause, W. Haak, Survival of Late Pleistocene Hunter-Gatherer Ancestry in the
898 Iberian Peninsula. *Current Biology* 29, 1169-1177 (2019).

899 51. L. Saag, L. Varul, C. L. Scheib, J. Stenderup, M. E. Allentoft, L. Saag, L. Pagani, M. Reidla, K.
900 Tambets, E. Metspalu, A. Kriiska, E. Willerslev, T. Kivisild, M. Metspalu, Extensive Farming in Estonia
901 Started through a Sex-Biased Migration from the Steppe. *Curr Biol.* 27, 2185-2193 (2017).

902 52. C. Posth, G. Renaud, A. Mittnik, D. G. Drucker, H. Rougier, C. Cupillard, F. Valentin, C. Thevenet,
903 A. Furtwängler, C. Wißing, M. Francken, M. Malina, M. Bolus, M. Lari, E. Gigli, G. Capecchi, I.
904 Crevecoeur, C. Beauval, D. Flas, M. Germonpré, J. Van der Plicht, R. Cottiaux, B. Gély, A. Ronchitelli, K.
905 Wehrberger, D. Grigorescu, J. Svoboda, P. Semal, Caramelli D, H. Bocherens, K. Harvati, N. Conard, W.
906 Haak, A. Powell, J. Krause, Pleistocene mitochondrial genomes suggest a single major dispersal of non-
907 Africans and a Late Glacial population turnover in Europe. *Current Biology* 26, 827-833 (2016).

908 53. B. Bramanti, M. Thomas, W. Haak, M. Unterlaender, P. Jores, K. Tambets, I. Antanaitis-Jacobs,
909 M. Haidle, R. Jankauskas, C. Kind, F. Lueth, T. Terberger, J. Hiller, S. Matsumura, P. Forster, B. Joachim,
910 Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326, 137-
911 140 (2009).

912 54. W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfelt,
913 E. Harney, K. Stewardson, Q. Fu, A. Mittnik, E. Bánffy, C. Economou, M. Francken, S. Friederich, R.
914 Pena, F. Hallgren, V. Khartanovich, A. Khokhlov, M. Kunst, P. Kuznetsov, H. Meller, O. Mochalov, V.
915 Moiseyev, N. Nicklisch, S. Pichler, R. Risch, M. Rojo Guerra, C. Roth, A. Szécsényi-Nagy, J. Wahl, M.

916 Meyer, J. Krause, D. Brown, D. Anthony, A. Cooper, K. W. Alt, D. Reich, Massive migration from the
917 steppe was a source for Indo-European languages in Europe. *Nature* 522, 207-2011 (2015).

918 55. G. Brandt, W. Haak, C. J. Adler, C. Roth, A. Szécsényi-Nagy, S. Karimnia, S. Möller-Rieker, H.
919 Meller, R. Ganslmeier, S. Friederich, V. Dresely, N. Nicklisch, J. K. Pickrell, F. Sirocko, D. Reich, A.
920 Cooper, K. W. Alt, The Genographic Consortium. Ancient DNA Reveals Key Stages in the Formation of
921 Central European Mitochondrial Genetic Diversity. *Science* 342, 257-261 (2013).

922 56. I. Lazaridis, A. Belfer-Cohen, S. Mallick, N. Patterson, O. Cheronet, N. Rohland, G. Bar-Oz, O.
923 Bar-Yosef, N. Jakeli, E. Kvavadze, D. Lordkipanidze, Z. Matzkevich, T. Meshveliani, B. J. Culleton, D. J.
924 Kennett, R. Pinhasi, D. Reich, Paleolithic DNA from the Caucasus reveals core of West Eurasian ancestry.
925 *BioRxiv*, (2018).

926 57. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster,
927 D. Reich, Ancient Admixture in Human History. *Genetics* 192, 1065-1093 (2012).

928 58. M. Piperno, *Papers in Italian Archaeology IV. Part II. Prehistory* C. Malone, S. Stoddart, Eds.,
929 International Series 244 (British Archaeological Reports, Oxford, 1985).

930 59. S. Tusa, *Papers in Italian Archaeology IV. Part II. Prehistory*. C. Malone, S. Stoddart, Eds.,
931 International Series 244 (British Archaeological Reports, Oxford, 1985).

932 60. S. Tusa, *Early Societies in Sicily: New developments in archaeological research* R. Leighton, Ed.,
933 (Accordia Specialist Studies on Italy 5, London, 1996).

934 61. G. N. Feliner, Southern European glacial refugia: A tale of tales. *Taxon* 60, 365-372 (2011).

935 62. T. Schmitt, Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers*
936 *in Zoology volume* 4, 11 (2007).

937 63. P. Soares, A. Achilli, O. Semino, W. Davies, V. Macaulay, B. H.J, A. Torroni, M. B. Richards, The
938 archaeogenetics of Europe. *Curr Biol.* 20, R174-183 (2010).

939 64. F. Antonioli, V. Lo Presti, M. Gasparo Morticelli, B. Laura, M. Mannino, M. R. Palombo, G.
940 Sannino, L. Ferranti, S. Furlani, K. Lambeck, S. Canese, R. Catalano, F. Chiocci, G. Mangano, G.
941 Scicchitano, R. Tonielli, Timing of the emergence of the Europe–Sicily bridge (40–17 cal ka BP) and its
942 implications for the spread of modern humans. *Geological Society London Special Publications* 411, 111-
943 144 (2014).

944 65. G. Di Maida, M. A. Mannino, B. Krause-Kyora, T. Z. T. Jensen, S. Talamo, Radiocarbon dating
945 and isotope analysis on the purported Aurignacian skeletal remains from Fontana Nuova (Ragusa, Italy).
946 *PLoS ONE* 14, e0213173 (2019).

947 66. M. Mussi, *Earliest Italy. An Overview of the Italian Palaeolithic and Mesolithic*. European Journal
948 of Archaeology (Springer US, New York, 2001), vol. 5, pp. 402.

949 67. P. Biagi, *The Widening Harvest. The Neolithic Transition in Europe: Looking Forward, Looking*
950 *Back*. A. J. Ammerman, P. Biagi, Eds., Aia Colloquia & Conference Papers (Book 6) (Archaeological
951 Institute of America, Boston, 2003), pp. 365.

952 68. D. Vetro, F. Martini, Mesolithic in Central-Southern Italy: Overview of lithic productions. *Quat.*
953 *Int* 423, 279-302 (2016).

954 69. C. Collina, *Le Neolithique ancien en Italie due sud: Evolution des industries lithiques entre VIIe et*
955 *VIe millenaire*. (Archaeopress Archaeology, Oxford, 2015), pp. 508.

956 70. D. Binder, C. Collina, R. Guilbert, T. Perrin, O. Garcia-Puchol, in *The emergence of pressure blade*
957 *making: From origin to modern experimentation* P. M. Desrosiers, Ed. (Springer Verlag, 2012), chap. 7,
958 pp. 199-218.

959 71. P. Biagi, D. Kiosak, The Mesolithic of the northwestern Pontic region New AMS dates for the
960 origin and spread of the blade and trapeze industries in southeastern Europe. *Eurasia Antiqua* 16, 21-41
961 (2010).

962 72. P. Biagi, E. Starnini, *The Origin and Spread of the Late Mesolithic Blade and Trapeze Industries*
963 *in Europe: Reconsidering J. G. D. Clark's Hypothesis Fifty Years After*. S. Țerna, B. Govedarica, Eds.,
964 Interactions, changes and meanings. Essays in honour of Igor Manzura on the occasion of his 60th birthday.
965 (Kishinev, 2016).

966 73. D. Gronenborn, *Migrations before the Neolithic? The Late Mesolithic blade-and-trapeze horizon*
967 *in central Europe and beyond*. Migration and Integration from Prehistory to the Middle Age (Tagungen des
968 Landesmuseums für Vorgeschichte Halle, Saale, 2017), vol. 17.

969 74. M.-L. Inizan, *The Emergence of Pressure Blade Making. From Origin to Modern Experimentation*.
970 P. M. Desrosiers, Ed., (Springer-Verlag, New York, 2012), pp. 536.

971 75. G. Marchand, T. Perrin, Why this revolution? Explaining the major technical shift inSouthwestern
972 Europe during the 7th millennium cal. BC. *Quat. Int* 428, 73-85 (2017).

76. T. Perrin, D. Binder, paper presented at the La transition néolithique en Méditerranée, Paris, 2014.

77. T. Günther, C. Valdiosera, H. Malmström, I. Ureña, R. Rodriguez-Varela, Ó. O. Sverrisdóttir, E. A. Daskalaki, P. Skoglund, T. Naidoo, E. M. Svensson, J. M. Bermúdez de Castro, E. Carbonell, M. Dunn, J. Storå, E. Iriarte, J. L. Arsuaga, J.-M. Carretero, A. Götherström, M. Jakobsson, Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc Natl Acad Sci U S A.* 112, 11917–11922 (2015).

78. C. Bonsall, R. Vasić, A. Boroneanţ, M. Roksandic, A. Soficaru, K. McSweeney, A. Evatt, Ü. Aguraiuja, C. Pickard, V. Dimitrijević, T. Higham, D. Hamilton, G. Cook, New AMS 14C Dates for Human Remains from Stone Age Sites in the Iron Gates Reach of the Danube, Southeast Europe. *Radiocarbon* 57, 33-46 (2015).

79. D. Borić, D. Price, Strontium isotopes document greater human mobility at the start of the Balkan Neolithic. *PNAS* 110, 3298-3303 (2013).

80. V. Tine, S. Tusa, paper presented at the Atti della XLI Riunione Scientifica, 16-19 novembre 2006, San Cipirello (PA), 2012.

81. C. Collina, *Sistemi tecnici e chaînes op_eratoires alla grotta dell'Uzzo (Trapani): analisi tecnologica delle industrie litiche dai livelli mesolitici e neolitici*. D. c. a. e. s. a. e. t. n. S. P. e, Protostorica, Eds., (Istituto Italiano di Preistoria e Protostoria, San Cipirello 2012), vol. Atti della XLI Riunione Scientifica, pp. 268.

82. M. Gilbert, H. Bandelt, M. Hofreiter, I. Barnes, Assessing ancient DNA studies. *Trends Ecol. Evol.* 20, 541–544 (2005).

83. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. García, S. Pääbo, J.-L. Arsuaga, M. Meyer, 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 (2013).

84. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* 6, pdb.prot5448 (2010).

85. N. Rohland, E. Harney, S. Mallick, S. Nordenfelt, D. Reich, Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. B* 370, 20130624 (2015).

1001 86. M.-T. Gansauge, T. Gerber, I. Glocke, P. Korlevic, L. Lippik, S. Nagel, L. Riehl, A. Schmidt, M.
1002 Meyer, Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic*
1003 *Acids Res.* 45, e79 (2017).

1004 87. V. Slon, C. Hopfe, C. L. Weiß, F. Mafessoni, M. de la Rasilla, C. Lalueza-Fox, A. Rosas, M.
1005 Soressi, M. Knul, R. Miller, J. Stewart, A. P. Derevianko, Z. Jacobs, B. Li, R. Roberts, M. Shunkov, H. de
1006 Lumley, C. Perrenoud, I. Gušić, Ž. Kućan, P. Rudan, A. Aximu-Petri, E. Essel, S. Nagel, B. Nickel, A.
1007 Schmidt, K. Prüfer, J. Kelso, H. A. Burbano, S. Pääbo, M. Meyer, Neandertal and Denisovan DNA from
1008 Pleistocene sediments. *Science* 356, 605-608 (2017).

1009 88. A. Peltzer, G. Jäger, A. Herbig, A. Seitz, C. Kniep, J. Krause, K. Nieselt, EAGER: efficient ancient
1010 genome reconstruction. *Genome Biol.* 17, 60-73 (2016).

1011 89. M. Schubert, S. Lindgreen, L. Orland, AdapterRemoval v2: rapid adapter trimming, identification,
1012 and read merging. *BMC Res. Notes* 9, 88-93 (2016).

1013 90. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform.
1014 *Bioinformatics* 26, 589-595 (2009).

1015 91. A. W. Briggs, U. Stenzel, P. L. F. Johnson, R. E. Green, J. Kelso, K. Prüfer, M. Meyer, J. Krause,
1016 M. T. Ronan, M. Lachmann, S. Pääbo, Patterns of damage in genomic DNA sequences from a Neandertal.
1017 *Proc. Natl. Acad. Sci. U.S.A* 104, 14616–14621 (2007).

1018 92. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing
1019 Data. *BMC Bioinformatics* 15, 356-368 (2014).

1020 93. M. Rasmussen, X. Guo, Y. Wang, K. Lohmueller, S. Rasmussen, A. Albrechtsen, L. Skotte, S.
1021 Lindgreen, M. Metspalu, T. Jombart, T. Kivisild, W. Zhai, A. Eriksson, A. Manica, L. Orlando, F. De La
1022 Vega, S. Tridico, E. Metspalu, K. Nielsen, M. Ávila-Arcos, J. Moreno-Mayar, C. Muller, J. Dortch, M.
1023 Gilbert, O. Lund, A. Wesolowska, M. Karmin, L. Weinert, B. Wang, J. Li, S. Tai, F. Xiao, T. Hanihara, G.
1024 van Driem, A. Jha, F. Ricaut, P. de Knijff, A. Migliano, I. Gallego Romero, K. Kristiansen, D. Lambert, S.
1025 Brunak, P. Forster, B. Brinkmann, O. Nehlich, M. Bunce, M. Richards, R. Gupta, C. Bustamante, A. Krogh,
1026 R. Foley, M. Lahr, F. Balloux, T. Sicheritz-Pontén, R. Villems, R. Nielsen, J. Wang, E. Willerslev, An
1027 Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* 334, 94-98 (2011).

1028 94. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: estimation of contamination and
1029 endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* 16, 224-241 (2015).

95. E. Li, Y. Zhang, DNA methylation in mammals. *Cold Spring Harb Perspect Biol.* 6, a019133 (2014).

96. D. J. Kennett, S. Plog, R. J. George, B. J. Culleton, A. S. Watson, P. Skoglund, N. Rohland, S. Mallick, K. Stewardson, L. Kistler, S. A. LeBlanc, P. M. Whiteley, D. Reich, G. H. Perry, Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* 8, 14115 (2017).

97. M. van de Loosdrecht, A. Bouzouggar, L. Humphrey, C. Posth, N. Barton, A. Aximu-Petri, B. Nickel, S. Nagel, E. Talbi, M. El Hajraoui, S. Amzazi, J. Hublin, S. Pääbo, S. Schiffels, M. Meyer, W. Haak, C. Jeong, J. Krause, Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science* 360, 548-552 (2018).

98. H. Weissensteiner, D. Pacher, A. Kloss-Brandstätter, L. Forer, G. Specht, H. J. Bandelt, F. Kronenberg, A. Salas, S. Schönherr, *Nucleic Acids Res.* 44, W58-63 (2016).

99. M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 30, E386-394 (2009).

100. M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649 (2012).

101. G. D. Poznik, Y. Xue, F. L. Mendez, T. F. Willems, T. G. P. Consortium, C. D. Bustamante, C. Tyler-Smith, Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat. Genet.* 48, 593–599 (2016).

102. N. Patterson, A. L. Price, D. Reich, Population Structure and Eigenanalysis. *PLoS Genet* 2, e190 (2006).

103. D. Reich, K. Thangaraj, N. Patterson, A. L. Price, L. Singh, Reconstructing Indian population history. *Nature* 461, 489-494 (2009).

104. S. Talamo, M. Richards, A Comparison of Bone Pretreatment Methods for AMS Dating of Samples >30,000 BP. *Radiocarbon* 53, 443-449 (2011).

105. R. Longin, New Method of Collagen Extraction for Radiocarbon Dating. *Nature* 230, 241-242 (1971).

106. T. A. Brown, D. E. Nelson, J. S. Vogel, J. R. Southon, Improved Collagen Extraction by Modified Longin Method. *Radiocarbon* 30, 171-177 (1988).

107. F. Brock, C. Bronk Ramsey, T. Higham, Quality assurance of ultrafiltered bone dating. *Radiocarbon* 49, 187–192 (2007).

108. P. Korlević, S. Talamo, M. Meyer, A combined method for DNA analysis and radiocarbon dating from a single sample. *Scientific Reports* 8, (2018).

109. P. Reimer, G. McCormac, Marine Radiocarbon Reservoir Corrections for the Mediterranean and Aegean Seas. *Radiocarbon* 44, 159 (2002).

110. G. van Klinken, J, Bone Collagen Quality Indicators for Palaeodietary and Radiocarbon Measurements. *Journal of Archaeological Science* 26, 687-695 (1999).

111. G. Mannino, Il riparo dell'Uzzo. *Sicilia Archeologica (VI)* 23, 21-39 (1973).

112. L. Conte, S. Tusa, Approfondimento stratigrafico alla Grotta dell'Uzzo. *Atti della XLI Riunione Scientifica dell'Istituto Italiano di Preistoria e Protostoria*, 437-445 (2012).

113. M. Piperno, S. Scali, A. Tagliacozzo, Mesolitico e Neolitico alla Grotta dell'Uzzo (Trapani). Primi dati per un'interpretazione paleoeconomica. *Quaternaria* 22, 275-300 (1980).

114. M. Piperno, S. Tusa, Relazione preliminare sulla seconda campagna di scavi alla Grotta dell'Uzzo (Trapani). *Sicilia Archeologica (VI)* 21, 39-42 (1976).

115. M. Piperno, S. Tusa, I. Valente, Campagne di scavo 1977 e 1978 alla Grotta dell'Uzzo (Trapani). Relazione preliminare e datazione dei livelli mesolitici e neolitici. *Sicilia Archeologica (VI)* 42, 49-64 (1980).

116. S. M. Borgognini Tarli, A. Canci, M. Piperno, E. Repetto, Dati archeologici e antropologici sulle sepolture mesolitiche della Grotta dell'Uzzo (Trapani). *Bullettino di Paletnologia Italiana* 84, 85-179 (1993).

117. R. Di Salvo, M. A. Mannino, V. Schimmenti, L. Sineo, K. D. Thomas, paper presented at the Atti della XLI Riunione Scientifica dell'Istituto Italiano di Preistoria e Protostoria, 2012.

118. S. M. Borgognini Tarli, E. Repetto, Dietary patterns in the Mesolithic samples from Uzzo and Molara caves (Sicily): the evidence of teeth. *Journal of Human* 14, 241-254 (1985).

119. A. Guerreschi, F. Fontana, paper presented at the Atti della XLI Riunione Scientifica dell'Istituto Italiano di Preistoria e Protostoria, 2012.

120. D. Lo Vetro, F. Martini, paper presented at the Atti della XLI Riunione Scientifica dell'Istituto Italiano di Preistoria e Protostoria, 2012.

121. L. Costantini, M. Piperno, S. Tusa, *La néolithisation de la Sicile occidentale d'après les résultats des fouilles a la grotte de l'Uzzo (Trapani)*. J. Guilaine, J. Courtin, J.-L. Roudil, J.-L. Vernet, Eds., Premières Communautés Paysannes en Méditerranée Occidentale (CNRS Editions, Montpellier, 1987).

122. M. A. Mannino, K. D. Thomas, M. Piperno, S. Tusa, A. Tagliacozzo, Fine-tuning the radiocarbon chronology of the Grotta dell'Uzzo (Trapani). *Atti della Società per la Preistoria e Protostoria della Regione Friuli-Venezia Giulia* 15, 17-31 (2006).

123. M. Piperno, Scoperta di una sepoltura doppia epigravettiana nella Grotta dell'Uzzo (Trapani). *Kokalos* 22-23, 734-760 (1976-77).

124. C. Bronk Ramsey, Bayesian analysis of radiocarbon dates. *Radiocarbon* 51, 337-370 (2009).

125. P. J. Reimer, E. Bard, A. Bayliss, J. W. Beck, P. G. Blackwell, C. Bronk Ramsey, P. M. Grootes, T. P. Guilderson, H. Haflidason, I. Hajdas, C. Hatté, T. J. Heaton, D. L. Hoffmann, A. G. Hogg, K. A. Hughen, K. F. Kaiser, B. Kromer, S. W. Manning, M. Niu, R. W. Reimer, D. A. Richards, E. M. Scott, J. R. Southon, R. A. Staff, C. S. M. Turney, J. van der Plicht, IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* 55, 1869–1887 (2013).

126. K. Prüfer, snpAD: An ancient DNA genotype caller. *Bioinformatics* 34, 4165-4171 (2018).

127. A. Seguin-Orlando, T. S. Korneliussen, M. Sikora, A. S. Malaspinas, A. Manica, I. Moltke, A. Albrechtsen, A. Ko, A. Margaryan, V. Moiseyev, T. Goebel, M. Westaway, D. Lambert, V. Khartanovich, J. D. Wall, P. R. Nigst, R. A. Foley, M. M. Lahr, R. Nielsen, L. Orlando, E. Willerslev, Genomic structure in Europeans dating back at least 36,200 years. *Science* 346, 1113-1118 (2014).

128. Q. Fu, P. Rudan, S. Pääbo, J. Krause, Complete Mitochondrial Genomes Reveal Neolithic Expansion into Europe. *PLoS ONE* 7, e32473 (2012).

129. M. Pala, A. Achilli, A. Olivieri, B. Hooshiar Kashani, U. Perego, D. Sanna, E. Metspalu, K. Tambets, E. Tamm, M. Accetturo, V. Carossa, H. Lancioni, F. Panara, B. Zimmermann, G. Huber, N. Al-Zahery, F. Brisighelli, S. R. Woodward, P. Francalacci, W. Parson, A. Salas, D. Behar, R. Villems, O. Semino, H. J. Bandelt, A. Torroni, Mitochondrial haplogroup U5b3: a distant echo of the epipaleolithic in Italy and the legacy of the early Sardinians. *Am J Hum Genet.* 84, 814-821 (2009).

130. T. Günther, H. Malmström, E. M. Svensson, A. Omrak, F. Sánchez-Quinto, G. M. Kılınç, M. Krzewińska, G. Eriksson, M. Fraser, H. Edlund, A. R. Munters, A. Coutinho, L. G. Simões, M. Vicente, A.

1117 Sjölander, B. Jansen Sellevold, R. Jørgensen, P. Claes, M. D. Shriver, C. Valdiosera, M. G. Netea, J. Apel,
1118 K. Lidén, B. Skar, J. Storå, A. Götherström, M. Jakobsson, Population genomics of Mesolithic Scandinavia:
1119 Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol.* 16, e2003703
1120 (2018).

1121 131. C. Gamba, E. Fernández, M. Tirado, M. F. Deguilloux, M. H. Pemonge, P. Utrilla, M. Edo, M.
1122 Molist, R. Rasteiro, L. Chikhi, A.-P. E., Ancient DNA from an Early Neolithic Iberian population supports
1123 a pioneer colonization by first farmers. *Mol. Ecol.* 21, 45-56 (2012).

1124 132. A. Szécsényi-Nagy, C. Roth, G. Brandt, C. Rihuete-Herrada, C. Tejedor-Rodríguez, P. Held, Í.
1125 García-Martínez-de-Lagrán, M. H. Arcusa, S. Zesch, C. Knipper, E. Bánffy, S. Friederich, H. Meller, R. P.
1126 Bueno, B. R. Barroso, B. R. de Balbín, A. M. Herrero-Corral, F. R. Flores, F. C. Alonso, E. J. Jiménez, L.
1127 Rindlisbacher, C. Oliart, M. I. Fregeiro, I. Soriano, O. Vicente, R. Micó, V. Lull, D. J. Soler, P. J. A. López,
1128 M. C. Roca de Togores, P. M. S. Hernández, M. F. J. Jover, M. J. Lomba, F. A. Avilés, K. Lillios, A. Silva,
1129 R. M. Magalhães, L. M. Oosterbeek, C. Cunha, A. J. Waterman, J. Roig Buxó, A. Martínez, M. J. Ponce,
1130 O. M. Hunt, J. C. Mejías-García, E. J. C. Pecero, B. R. Cruz-Auñón, T. Tomé, B. E. Carmona, J. Cardoso,
1131 A. Araújo, C. Liesau von Lettow-Vorbeck, B. B. C, M. P. Ríos, A. Pujante, J. I. Royo-Guillén, M. A.
1132 Esquembre Beviá, V. M. Dos Santos Goncalves, R. Parreira, H. E. Morán, I. E. Méndez, M. J. Vega Y, G.
1133 R. Menduiña, C. V. Martínez, J. O. López, J. Krause, S. L. Pichler, R. Garrido-Pena, M. Kunst, R. Risch,
1134 M. A. Rojo-Guerra, W. Haak, K. W. Alt, The maternal genetic make-up of the Iberian Peninsula between
1135 the Neolithic and the Early Bronze Age. *Sci. Rep.* 7, 15644 (2017).

1136 133. W. Haak, O. Balanovsky, J. J. Sanchez, S. Koshel, V. Zaporozhchenko, C. J. Adler, C. S. I. Der
1137 Sarkissian, G. Brandt, C. Schwarz, N. Nicklisch, V. Dresely, B. Fritsch, E. Balanovska, R. Villems, H.
1138 Meller, K. W. Alt, A. Cooper, C. the Genographic, Ancient DNA from European Early Neolithic Farmers
1139 Reveals Their Near Eastern Affinities. *PLOS Biology* 8, e1000536 (2010).

1140 134. W. Haak, P. Forster, B. Bramanti, S. Matsumura, G. Brandt, M. Tänzer, R. Villems, C. Renfrew,
1141 D. Gronenborn, K. Alt, J. Burger, Ancient DNA from the first European farmers in 7500-year-old Neolithic
1142 site. *Science* 310, 1016-1018 (2005).

1143 135. T. Kivisild, The study of human Y chromosome variation through ancient DNA. *Hum. Genet.* 137,
1144 863 (2017).

1145 **Acknowledgements.** For helpful comments we thank Stephanie Eisenmann, Maite Rivollat, Thiseas
1146 Lamnidis and other members of the Department of Archaeogenetics, and Barbara Pavlek of the Minds &
1147 Traditions research group of the Max Planck Institute for the Science of Human History. We are grateful

for comments on the manuscript from Didier Binder from the French National Center for Scientific Research and Detlef Gronenborn from the Leibniz Research Institute for Archaeology. We thank David Reich, Shop Mallick and Ian Mathieson for access to unpublished data. We thank Sven Steinbrenner of the Department of Human Evolution from the Max Planck Institute for Evolutionary Anthropology for undertaking the stable isotope analysis. **Funding**. The Max Planck Society financed the genetic, isotopic and radiocarbon analyses. **Author contributions.** M.Ma, W.H, and J.K conceived the study. M.Ma, A.T, M.P, S.T, C.C, V.S and R.dS provided the ancient human remains and input for the archaeological interpretation. M.vdL, C.F, S.N and L.K performed laboratory work with the help of F.A, M.B, R.R, R.S, A.W, G.B and M.Me. M.vdL conducted the population genetic analyses with the help of K.P, C.J and W.H. S.T performed the AMS radiocarbon dating analysis, and M.Ma the isotope analysis. M.vdL, M.Ma, W.H, V.V-M, C.P, S.S, C.J, K.P and J.K. wrote the paper with input from all co-authors. **Competing interests.** The authors declare no competing interests. **Corresponding authors.** Correspondence to M.vdL, M.Ma or J.K. **Data and materials availability.** Genomic data (BAM format) are available through the Sequence Read Archive (accession number X) and consensus mitogenome sequences (FASTA format) in GenBank (accession numbers X to X).

8. Discussion

8.1. Archaeogenetic perspectives on the repopulation of northern Africa and Europe after the Last Glacial Maximum

The ancient genomes of hunter-gatherers from Mediterranean areas that harboured glacial refugia are key to furthering our understanding of the repopulation dynamics in northern Africa and Europe after the last Ice Age. Manuscript A reports the first genomic reference points for hunter-gatherers from the Mediterranean coast in northern Africa retrieved from the Maghreb, whereas Manuscripts B and C filled in gaps in the temporal and geographical record for foragers from southern Europe. In this section I highlight how the findings in this dissertation have advanced our understanding of the population transitions after the LGM associated with hunter-gatherers who manufactured microlithic industries in northwestern Africa (Section 8.1.1) and Europe (Section 8.1.2), and provide an extended discussion of parallel studies.

8.1.1. Northwestern Africa

The expansion of Iberomaurusian microlithic industries across northwestern Africa after the onset of the LGM ~25,000 yBP has remained a topic of debate among archaeologists. Some scholars have proposed an endemic origin in northwestern Africa, and others an introduction from Europe or the Near East via northeastern Africa (reviewed in Hogue, 2014). The findings in Manuscript A demonstrate that about two-thirds of the ancestry in the ~15,000 calBP Iberomaurusian foragers from Morocco is typical for Eurasians, and had a high similarity to that of Natufian microlithic hunter-gatherers who lived in the Near East ~15,000-9,000 yBP (Lazaridis et al., 2016). Notably, the substantial proportion of complex sub-Saharan African ancestry in the Iberomaurusian genomic profiles was absent in the Natufians. Based on this, I currently consider it more parsimonious that the Eurasian ancestry was introduced from the Near East into northern Africa rather than vice versa. However, it is important to note that the oldest Iberomaurusian sites are currently described for the Maghreb (~25,800 calBP) and not for northeastern Africa (Eastern Oranian, ~19,000 calBP) (Douka et al., 2014; Hogue & Barton, 2016). This leaves the possibility open that a population ancestral to the Late Iberomaurusian foragers analysed here had expanded from the Maghreb into the Near East prior to the admixture event that would introduce

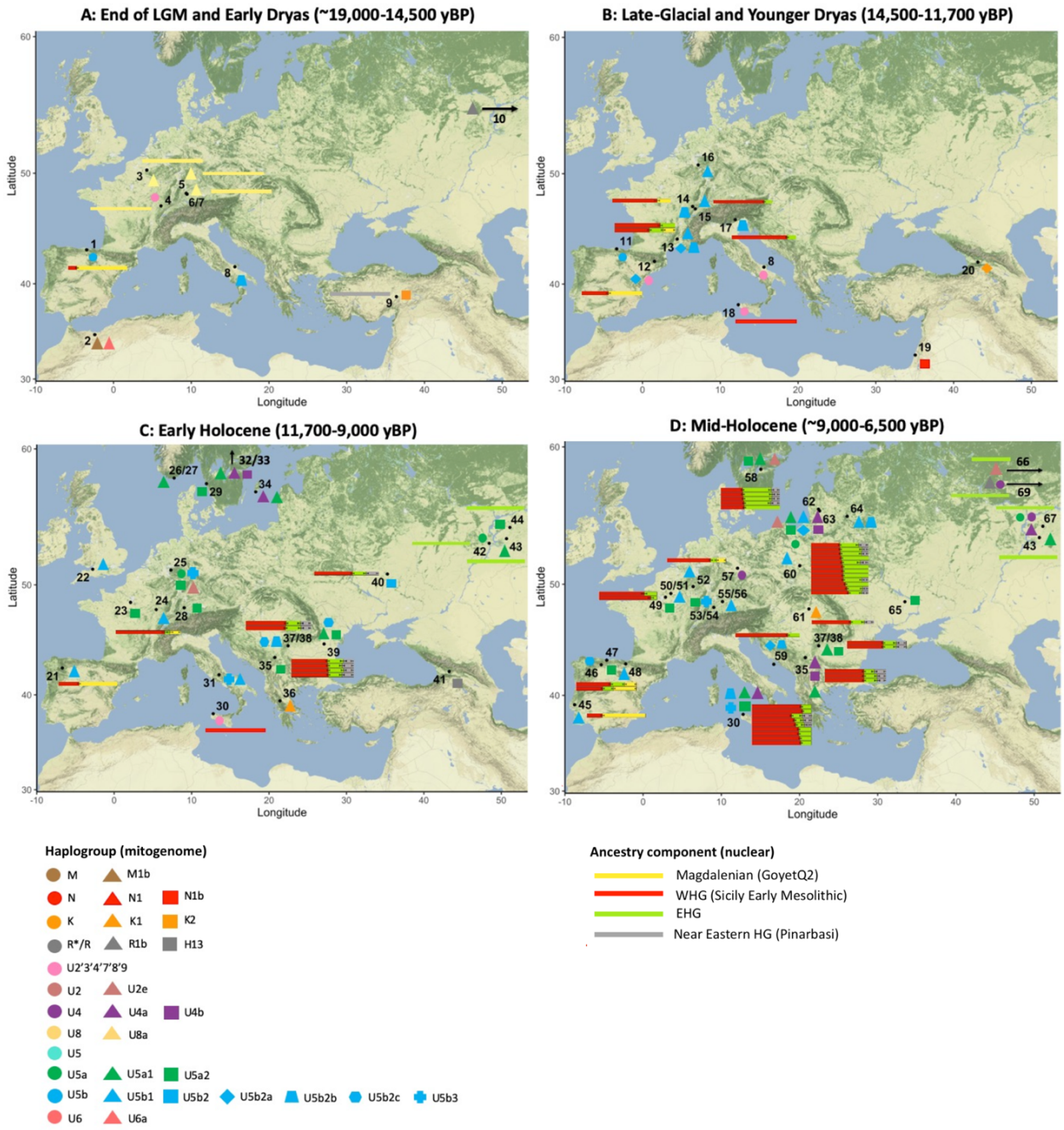


Figure-4 | Changes in post-LGM hunter-gatherer genomic structure in Europe and northern Africa over time. Symbols represent mitogenome haplogroups, and bars nuclear genomic ancestry profiles. The nuclear ancestry components are taken from the *qpWave*- and *qpAdm*-based models reported in Manuscript C, using *GoyetQ2* as a proxy for Magdalenian-related (yellow), Sicilian Early Mesolithic hunter-gatherers for WHG-related (red), EHGs for EHG-related (green) and *Pınarbaşı* for Near Eastern hunter-gatherer-related (grey) ancestries. Error bars reflect 1 standard error. **A)** End of the LGM and Early Dryas: Initial formation of a Magdalenian-WHG ancestry cline from central Europe to Iberia. **B)** Late Glacial and Younger Dryas: EHG ancestry, as part of the WHG/EHG mixture in some Villabruna cluster individuals, adds an additional dimension to the pre-existing Magdalenian-WHG ancestry cline.

the complex sub-Saharan African ancestry. Furthermore, some archaeologists have proposed that both the Iberomaurusian and the related Eastern Oranian industries may have their technological origins in the earlier LSA Dabban industry in Cyrenaica in northeastern Africa that had appeared ~40,000 yBP (Douka et al., 2014). However, the archaeological links between the Dabban and the contemporaneous Upper Palaeolithic technologies in the Near East, as well as the Middle Stone Age Aterian in northern Africa, remains highly ambiguous. Therefore, although the Iberomaurusian ancestry profiles establish an ultimate connection to the Near East, it is currently not possible to determine the exact temporal and archaeological context associated with this connection (for extensive archaeological discussions, see Garcea, 2010; Hogue, 2014; Poti & Weniger, 2019).

Another outstanding question is the temporal depth of the genomic connection between North Africa and the Near East. The genome-wide data of the Iberomaurusian hunter-gatherers doubled the time depth of reference points to understand the genomic prehistory of northern Africa. Their ancestry profiles indicated that significant Eurasian ancestry, tagged by the characteristic North African mitogenome lineages U6 and M1 (Maca-Meyer et al., 2003; Olivieri et al., 2006; Pereira et al., 2010), was already present in northwestern Africa by at least ~15,000 yBP (Figure-4A). This indicates that the genomic connection between northern Africa and the Near East long predates the arrival of nomadic pastoralism and sedentary farming practises. Notably, early farmers from Morocco retained a significant proportion of Iberomaurusian-related

---

Figure-4 (cont.) | **C)** Early Holocene: Ancestry with distal affinities to both EHG and *Pınarbaşı* (EHG/*Pınarbaşı*) appears in southeastern Europe. **D)** Mid-Holocene: The EHG/*Pınarbaşı* ancestry mixture can be found among Late Mesolithic foragers in Italy, the Baltic and Scandinavia, and is seemingly correlated with the spread of blade-and-trapeze industries in these regions.

Site 1: *El Mirón*, 2: *Grotte des Pigeons (Taforalt)*, 3: *Troisième*, 4: *Rigney*, 5: *Burkhardtshöhle*, 6: *Brillenhöhle*, 7: *Hohlefels*, 8: *Grotta Paglicci*, 9: *Pınarbaşı*, 10: *Afontova Gora 3*, 11: *Erralla*, 12: *Balma Guilanyà*, 13: *Aven des Iboussières*, 14: *Rochedane*, 15: *Grotte du Bichon*, 16: *Oberkassel*, 17: *Villabruna*, 18: *Grotta d'Oriente*, 19: *Raqefet*, 20: *Satsurblia*, 21: *Chan do Lindeiro*, 22: *Gough's Cave*, 23: *Les Closeaux*, 24: *Abri des Cabônes*, 25: *Blätterhöhle*, 26: *Hummervikholmen*, 27: *Kristiansand*, 28: *Falkensteiner Höhle*, 29: *Sandarna*, 30: *Grotta dell'Uzzo*, 31: *Continenza*, 32: *Stora Bjers*, 33: *Stora Förvar*, 34: *Hemse*, 35: *Vlasac*, 36: *Theopetra*, 37: *Hadučka Vodenica*, 38: *Schela Cladovei*, 39: *Padina*, 40: *Vasilyevka*, 41: *Kotias*, 42: *Chekalino*, 43: *Lebyazhinka*, 44: *Sidelkino*, 45: *Moita do Sebastião*, 46: *La Braña*, 47: *Canes*, 48: *Aizpea*, 49*: Les Vignolles*, 50: *BerryAuBac*, 51: *Les Fontinettes (Chaudardes)*, 52: *Loschbour*, 53: *Felsdach Inzigofen*, 54: *Hohlenstein-Stadel*, 55: *Bockstein Höhle*, 56: *Große Ofnet Höhle*, 57: *Bad Dürrenberg*, 58: *Motala*, 59: *Vela Spila*, 60: *Janisławice*, 61: *Koros*, 62: *Spiginas*, 63: *Zvejnieki*, 64: *Kretuonas*, 65: *Vovnigi*, 66: *Archangelsk*, 67: *Samara*, 68: *Yuzhnyy Oleni Ostrov*, 69: *Popova*.

ancestry, and minor proportions persist in the genomes of many present-day North African peoples (Fregel et al., 2018; Serra-Vidal et al., 2019). This underlines that Iberomaurusian-related ancestry remained endemic to northern Africa from at least ~15,000 yBP onwards (Henn et al., 2012).

8.1.2. Europe

On the opposite side of the Mediterranean Sea in Europe, the period that immediately followed the LGM is associated with several archaeological transitions. Manuscripts B and C reported a dozen new genomic reference points for Upper Palaeolithic and Mesolithic hunter-gatherers from Iberia and southern Italy, respectively. Together, these manuscripts have contributed to a more fine-scaled understanding of the population substructure and admixture dynamics of hunter-gatherers who inhabited Europe after the LGM. The most important findings include an admixture cline in Magdalenian-associated ancestry in Iberian hunter-gatherers (Manuscript B), and the spread of an EHG-related ancestry into Europe associated with the expansion of the Late Mesolithic blade-and-trapeze complex from eastern/southeastern Europe (Manuscript C). Below I detail some of these nuanced findings.

Late Pleniglacial expansions resulted in a Magdalenian-WHG ancestry cline

After the LGM, Magdalenian assemblages emerged and expanded in Iberia and western Europe, and Epigravettian assemblages in Italy and southeastern Europe (Maier, 2015; Djindjian, 2016; Fernández-López de Pablo et al., 2019). The European hunter-gatherers showed a large diversification of predominantly U2, U5, and U8 mitogenome haplogroups (Figure-4A) (Posth et al., 2016). Whereas U2 and U8 haplogroups were frequent among Magdalenian-associated individuals, haplogroup U5b was typical for individuals with high WHG ancestry, such as Epigravettian-associated individuals of the Villabruna genetic cluster.

The ~19,000 yBP *El Mirón* hunter-gatherer from Iberia is currently the oldest Magdalenian hunter-gatherer with ancient genome-wide data. Due to its geographical and temporal context, scholars have frequently assumed that this individual's ancestry most closely approximates that of the Magdalenian hunter-gatherers that expanded from southwestern Europe after the last Ice Age (e.g. Fu et al., 2016). However, the results in Manuscripts B and C indicate that *El Mirón* carried an ancestry from a WHG-related source that was not present in the slightly younger

Magdalenian-associated individuals from central Europe (Figure-4A). The admixed ancestry of *El Mirón* and the U5b haplogroup she carried suggests that the Magdalenian-WHG ancestry cline from Iberia to central Europe had already started to take form prior to the end of the LGM ~19,000 yBP. At Palaeolithic sites across the Iberian Peninsula and southwestern France, Solutrean technologies can be found until the end of the LGM, not seldomly directly underlying Early Magdalenian assemblages (Aura et al., 2012; Cascalheira et al., 2020 and references therein). Importantly, the Solutrean and Epigravettian complexes have been proposed to share a common technological origin in the Gravettian (e.g. Djindjian., 2016; Cascalheira et al., 2020). This raises the interesting possibility that the WHG ancestry and U5b haplogroup in *El Mirón* may have been carried over from preceding Solutrean-associated foragers in Iberia.

Europe: Late Glacial expansions resulted in substructure among Villabruna cluster individuals

The archaeological record in Europe shows major transitions in material culture during the Late Glacial ~14,500-11,700 yBP. The Magdalenian developed into the Azilian in western Europe, and technologies related to these and the Italian Epigravettian appeared in Iberia (Bonilla et al., 2012; Naudinot et al., 2017). During this period, WHG-related ancestry similar to that of the ~14,000 yBP *Villabruna* hunter-gatherer started to spread rapidly across Europe, seemingly tagged by a diversification of U5b mtDNA haplogroups (Figure-4B) (Fu et al., 2016; Posth et al., 2016). The results in Manuscripts B and C demonstrated that, rather than forming a genetically homogenous group, Villabruna cluster individuals had diverse genomic ancestries with influences from EHGs, Magdalenian-associated hunter-gatherers, or both (Figure-4B). A WHG/EHG mixture is detectable for the first time in *Villabruna* and some other hunter-gatherers in central Europe (e.g. *Bichon*, *Iboussieres*). This added EHG as an additional dimension to the Magdalenian-WHG ancestry cline in Iberia and central Europe. Notably, some hunter-gatherers in central Europe (*Rochedane*) and Iberia (*Balma*) did not show the affinity to EHG (Manuscript C). This observation is in line with my suggestion above that an initial Magdalenian-WHG ancestry cline had likely already formed in Iberia and central Europe prior to the European-wide expansion of Villabruna cluster individuals.

Furthermore, before publication of Manuscript B it was generally assumed that Villabruna cluster ancestry had replaced the Magdalenian ancestry in Iberia during the Mesolithic. This is because Mesolithic Iberian foragers, such as *La Braña*, *Canes1* and *Chan*, are genetically more similar to *Villabruna* and Mesolithic WHGs from central Europe than to ~19,000 yBP Magdalenian-associated *El Mirón* (Fu et al., 2016). However, the findings in Manuscript B indicated that the Villabruna cluster individuals from Iberia retained some Magdalenian-related ancestry throughout

the Mesolithic (Figure-4A-C). Moreover, in a recent study comparable proportions of Magdalenian-related ancestry were found in three Early Mesolithic foragers from France (Brunel et al., 2020). This suggests that Magdalenian-associated ancestry persisted among western European foragers after the last Ice Age.

Europe: Late Mesolithic blade-and-trapeze industry expansions introduced EHG-related ancestry

The onset of the Holocene ~11,700 yBP is characterised by a rapid increase in atmospheric temperatures (Rasmussen et al., 2006). As indicated in earlier studies, Manuscript C demonstrates that around this time an EHG-related ancestry with distal affinities to hunter-gatherers in Anatolia and the Near East (EHG/*Pınarbaşı* mixture) had appeared in southeastern Europe (Figure-4C) (Fu et al., 2016; Mathieson et al., 2018; Feldman et al., 2019). Further admixture of Early Mesolithic hunter-gatherers in central Europe and Iberia resulted in the introduction of *Villabruna*-like ancestry (WHG/EHG mixture) in Iberia (Figure-4C). U5a haplogroups appear predominantly as U5a2 lineages among Epipaleolithic and Early Mesolithic European hunter-gatherers in eastern, southeastern and central Europe, and additionally as U5a1 in northern Europe.

Moreover, the genomic transect analysis of the Sicilian hunter-gathers in Manuscript C provided one of the first systematic analyses to look into differences in ancestry between Early and Late Mesolithic hunter-gatherers. The results showed that the ancestry component related to both EHGs and Near Eastern hunter-gatherers (EHG/*Pınarbaşı*) spread further to Italy, the Baltic and Scandinavia as the Mesolithic progressed (Figure-4D). The spread of this ancestry seems to have coincided with the expansion of Late Mesolithic blade-and-trapeze industries in Europe, and was possibly tagged by U4a haplogroups.

8.2. Is there genomic evidence that prehistoric Mediterranean peoples were in direct contact across the Mediterranean Sea?

Archaeological scholars have postulated various prehistoric connections across the western Mediterranean via the Strait of Gibraltar and across the central Mediterranean via the Strait of Sicily based on observed similarities in material culture or synchronous developments in inferred subsistence strategies (reviewed in Straus, 2001; Zilhão, 2014; Gronenborn, 2017). The ancient genome-wide data for individuals from both the European and African Mediterranean coastal regions generated in this dissertation provided an unprecedented opportunity to

investigate systematically the genomic evidence for prehistoric sea crossings. However, to date the combined body of evidence based on the ancient genomes in this dissertation and other archaeogenetic studies does not support direct population connections between peoples inhabiting the European and African Mediterranean coastal areas prior to the Neolithic (all Manuscripts). In this section I detail relevant connections that were proposed by archaeologists for various time periods, and discuss their respective genomic support.

Middle and Upper Palaeolithic

In the past, scholars have frequently proposed direct material connections between Upper Palaeolithic[3] assemblages in Iberia or southern Italy and those in Mediterranean coastal northwestern Africa (Straus, 2001 and references therein; Castaño, 2007). One prominent supporter for dispersal across the Strait of Gibraltar was Pericot who proposed that the tanged pieces in Mediterranean Iberia and France from Upper Palaeolithic Solutrean contexts (25,000-19,000 yBP) had their origin in the Middle Paleolithic Aterian in northern Africa (Pericot, 1955). Moreover, based on a ~11,000 yBP harpoon from a Late Iberomaurusian context at *Taforalt*, Pallary proposed a connection between the Iberomaurusian in northern Africa and the Final Magdalenian in Iberia. The term “Iberomaurusian” is a fusion of the words ‘Ibero’ (Spanish) and ‘Maurusian’ (Mauritania, the Roman name for the geographical region) to draw attention to similarities between lithic assemblages in Morocco and Spain (Pallary, 1909). Some archaeologists have argued that the Taforalt harpoon, in combination with evidence for deep-sea fishing and probable navigation for microlithic cultures on both Mediterranean shores, forms a credible case for human contacts across the Strait of Gibraltar or Sicily during the terminal Pleistocene (reviewed in: Straus, 2001). However, other scholars have stressed that apart from this single harpoon, the Iberomaurusian assemblages have little in common specifically with the lithic industries of Mediterranean southern Iberia (Magdalenian, Solutrean) or with southern Italy (Gravettian, Epigravettian) (e.g. Zampetti, 1989; Aura Tortosa, 1995; Linstädter et al., 2012; Poti & Weniger, 2019).

All in all, despite the ample proposed indications from the archaeological record, I could not find genomic support for detectable direct population contacts between Upper Palaeolithic foragers in Iberia or Sicily and LSA Iberomaurusian foragers in northern Africa in either direction (All Manuscripts). Although many areas in northern Africa remain unsampled, the evidence so far

[3] Frequently referred to in Africa as Middle or Late Stone Age (e.g. Kleindienst, 2006)

suggests that the Mediterranean Sea posed a barrier to gene flow between peoples on the African and European shores around this time (Straus, 2001).

Late Mesolithic

The most remarkable change in material culture during the Mesolithic is the appearance of a horizon across western Eurasia with distinctive regular, parallel-sided blades and geometric trapezoidal microliths at the onset of the Late Mesolithic (Clark & Piggot, 1968). In northwestern Africa, similar trapezes and regular blades characterise the Upper Capsian (Rahmani, 2004). Scholars have debated on the geographical origins of the regular bladelets and the geometric trapezes, the expansion directions of these industries, and whether sea crossing were involved at the strait of Gibraltar or Sicily (recently reviewed in: Gronenborn, 2017). Some scholars have placed the origin of the blade-and trapeze industries in eastern Europe, where geometric trapezes had appeared in the Russian steppe zones and Ukraine ~9,500-9,000 yBP (Inizian, 2012; Biagi & Starnini, 2016). Yet another group of scholars proposed an origin in the Upper Capsian of northwestern Africa based on the early appearance of the bladelets and trapezes ~8,500 yBP, and their general similarity to the Late Mesolithic blade-and-trapeze industries of southern France, Iberia, and southern Italy.

In Manuscripts B and C, I generated ancient genome-wide data for Late Mesolithic hunter-gatherers from Iberia associated with the Geometric Mesolithic, the most abundant variety of blade-and-trapeze industry in Iberia, and from Sicily associated with the Castelnovian, a local variety of the blade-and-trapeze industry. Their genomic profiles contained a distinct ancestry that was deeply related to that of Mesolithic EHGs and Epipalaeolithic hunter-gatherers in Anatolia and the Near East. In contrast, I could not find indications for a distinct ancestry related to ~15,000 yBP Iberomaurusian hunter-gatherers, Neolithic farmers or present-day Amazighes (Berber) peoples from northwestern Africa. The ancestry profiles hence underline a connection to the related industries in eastern/southeastern Europe (Inizian, 2012; Biagi & Starnini, 2016), and not to the Capsian in northwestern Africa via the Strait of Sicily or Gibraltar (e.g: Cortés Sánchez et al., 2012; Perrin & Binder, 2014).

Early Neolithic

One of the outstanding questions in Mediterranean prehistoric archaeology is whether the introduction of early farming practises in Sicily and coastal areas in Iberia and the Maghreb involved the crossing of early farmers by boat at the Strait of Gibraltar and/or Sicily. In addition, two major expansion routes were proposed (reviewed e.g. in Zilhão, 2014) (see Figure-2). One

possible route followed the European Mediterranean coastlines out of the Balkans into peninsular Italy to Sicily, Mediterranean France and Iberia, and perhaps via the Gibraltar Strait into northwestern Africa (Northern Mediterranean Route). A second possible route involved an early pioneering event in peninsular southern Italy and Sicily, from where the early farmers crossed the Strait of Sicily and followed the northwest African Mediterranean coastline and the Gibraltar Strait into Iberia (Southern Mediterranean Route). The distinctive early occurrence of Impressa Ware pottery, such as of the *Almagra* type, was proposed as a signature cultural marker for this southern route expansion (Manen et al., 2007).

To address these outstanding questions with genomic evidence, I analysed ancient genomes from early farmers from Mediterranean regions with some of the oldest evidence for pottery and agriculture. The Early Neolithic Iberian farmers associated with Cardial and Epicardial pottery (Manuscript B), and from Sicily associated with Impressa (*Stentinello* type) pottery (Manuscript C), are genetically most similar to early farmers from the Balkans, and not to ~7,000 yBP early farmers from *Grotte d'Ifri n'Amr O'moussa* in Morocco who produced *Almagra* type pottery. The Iberian and Sicilian farmers appear to have derived their ancestry predominantly from northwestern Anatolian early farmers, similarly to previously published European early farmers, including those from other European Mediterranean coastal areas (e.g. Haak et al., 2015; Olalde et al., 2015). In contrast, the Early Neolithic Moroccan farmers showed a high genetic continuity with the preceding LSA ~15,000 yBP Iberomaurusian hunter-gatherers, as well as a high genetic similarity to early farmers from the Levant (Fregel et al., 2018).

Many areas in coastal northern Africa remain unsampled, as well as some of the oldest farmers communities in Sicily, peninsular southern Italy and southern Iberia. However, the genomic evidence put forth in this dissertation and other archaeogenetics studies indicates that the makers of both the Impressa and Cardial Ware potteries in Sicily and Iberia, including the *Almagra* type in southern Iberia, shared a common Balkan origin with central European early farmers who manufactured Linear Band Ceramics pottery (e.g. Lazaridis et al., 2014; Haak et al., 2015; Olalde et al., 2015; Valdiosera et al., 2018). This provides support for a northern Mediterranean expansion route (Zilhão, 2014). The evidence so far does not provide support for a separate introduction of early Impressa Ware in Sicily, Andalusia and northwestern Africa along a southern Mediterranean expansion route that involved early farmers crossing the Strait of Sicily and Gibraltar (Manen et al., 2007). Neither does the genomic evidence thus far support a Maghrebi origin for the Early Neolithic in Iberia (Manen et al., 2007; Marchand & Manen, 2009; Cortés Sánchez et al., 2012; Linstädter et al., 2012). Notably, ~5,000 yBP Late Neolithic farmers from *Kelif el Boroud* in Morocco were found to be genetically most closely related to farmers in

Andalusia in southern Iberia, and carried a similar genetic signature from Anatolian early farmer related ancestry associated with European early farmers (Fregel et al., 2018). This implies that farmers from Iberia had crossed the Strait of Gibraltar to northwestern Africa by at least 5,000 yBP (Zilhão., 2014).

8.3. Limitations of ancestry inferences using ancient DNA

Since the introduction of NGS approaches to archaeogenetic research, it has become increasingly clear that the intrinsic properties of aDNA not only affect its recovery in the lab (see Introduction section 1.1) but also its down-stream demographic and ancestry inferences (e.g. Günther & Nettleblad, 2019; Peyrégne & Prüfer, 2020). Similar to other ancient genome studies, I predominantly used *f*-statistic-based methods for the ancestry characterisation of the ancient individuals. *F*-statistics measure shared genetic drift based on allele frequencies, and provide sensitive statistical tests for admixture (Green et al., 2010; Patterson et al., 2012). However, *f*-statistic-based methods only provide sensible and reliable results when the underlying population allele-frequencies have been inferred from accurate genotype calls, and are based on polymorphic SNP loci that ascertain for sufficient genomic variance. In the remainder of this section I detail these limitations further, and address how these posed challenges for the research in this dissertation.

Reference bias may affect genotype calls

Genotype calls based on the reconstructed genomes from mapped aDNA reads may fail to accurately identify genetic variants because of the intrinsic properties of aDNA. The typical short fragment length and the larger fraction of mismatches from deamination may result in a higher proportion of aDNA reads to map aspecifically (Prüfer et al., 2010). Furthermore, reads that carry the reference variants will map more frequently and with higher specificity than reads with alternative variants (reference bias). This reference bias has been shown to disproportionally reduce the accuracy of the calling rate for the alternative allele at heterozygote site during genotyping, in particular at SNP loci that are covered by only a few reads (e.g. Günther & Nettleblad, 2019; Martiniano et al, 2020).

Pseudo-haploid genotype calling is a frequently used method to call SNP variants in human aDNA studies, including in this dissertation, and requires a minimum coverage of only one read at a given locus. As a result, pseudo-haploid variant calls for low coverage SNP loci can be

strongly affected by reference bias (for an analysis of the genome-wide date in Manuscript A, see Günther & Nettleblad, 2019). This underestimates the assessed genomic variability on the individual- and population-level, which may subsequently bias the measurements of shared genetic drift in *f*-statistics. However, this method does effectively call a larger number of overall SNPs from the low-coverage genomic data typical for ancient human remains. This often allows for more individuals to be included in *f*-statistic-based population-level genomic analyses, which results in a more accurate estimation of population allele frequencies.

SNP ascertainment may restrict ancestry characterisation

A second limitation is that the ~1.2 million SNP panel that I used in this dissertation for the capturing of autosomal genetic variation is based on pre-ascertained SNP loci that are known to be polymorphic in a subset of modern humans (Fu et al., 2015; Haak et al., 2015; Mathieson et al., 2018). The included SNPs are neutral common bi-allelic loci (minimal allele frequency >10%) with distinct frequency differences across populations from different geographical regions. Common variance has been shown to be effective in identifying continental-wide population structure and assigning ancestries (e.g. Li et al., 2008; Bergström et al., 2020). However, in order to infer more fine-scale population structure rare allele variance (minimum allele frequency <1%) is needed that is more geographically restricted (Slatkin et al., 1985; McVean et al., 2012; O'Conner et al., 2015). Without the information from rare variants it is not possible to detect fine-scale genetic substructure among closely related human groups. This made it for example challenging to differentiate between the various European early farmer groups and to investigate the more locally confined details of their expansion along the Mediterranean route (Manuscripts B and C).

Moreover, despite genetic diversity being higher in Africa than in any other continent, the SNP panel used here was ascertained for predominantly Eurasian populations and thus relatively poorly represents sub-Saharan African genetic diversity (ascertainment bias) (Lachance & Tishkoff, 2013). The limitation of this panel became clear during the ancestry characterisation of the ~15,000 yBP North African Iberomaurusian hunter-gatherers (Manuscript A). In this manuscript I discovered that the Iberomaurusian hunter-gatherers derived about one-third of their ancestry from a basal human lineage that did not leave any known unadmixed descendants. Such a "ghost ancestry component" is a typical result in aDNA studies when a proportion of the genomic variance of the ancient individual or group cannot be approximated by any known ancient or present-day human group (e.g. Reich et al. 2009; Lazaridis et al., 2016/2018; Durvasula & Sankararaman, 2020). In particular, hidden layers of archaic ancestry may confound the ancestry

decomposition of ancient genomic profiles that show affinity to sub-Saharan African and some of the most deeply diverged Eurasian lineages (Hammer et al., 2011; Lachance et al., 2012; Nielsen et al., 2017). Disentangling these ancestries may require the development of capture panels and analyses methods that do not rely on a reference genome for archaic hominins to detect archaic introgression (e.g. Durvasula & Sankararaman, 2020).

Despite the challenges associated with ancient genome analyses detailed above, the previous chapters have demonstrated ample examples of opportunities and advantages that aDNA analysis methods can offer to directly study demographic events in human prehistory.

8.4. Conclusion

This dissertation focused on the analysis of ancient genomes recovered from prehistoric peoples from the Mediterranean region to investigate transitions in population genomic structure after the LGM associated with archaeological expansions, including the spread of agricultural practises. A particular emphasis was placed on examining the genomic evidence for direct population contact across the Mediterranean Sea between peoples on the European and African shores.

Each of the manuscripts contributed crucial genomic reference points from Mediterranean areas where human occupation locally persisted during the LGM, or with some of the earliest indications for agricultural practises and pottery. Manuscript A provided the first available genomic reference points for hunter-gatherers from the Mediterranean coast in northern Africa. In addition, the two transects from Iberia and Sicily in Manuscripts B and C filled in important geographical and temporal gaps in the Mediterranean southern European genomic record. As a result, this dissertation research could provide archaeogenetic perspectives on five topics that have long been debated by scholars of Mediterranean prehistory.

1) The origin of the Iberomaurusian microlithic industry in northwestern Africa. The results in Manuscript A demonstrated that the Iberomaurusian foragers harboured a Eurasian ancestry that was similar to the ancestry found among microlithic Natufian foragers in the Near East. Whereas the Iberomaurusian genomic profiles contained additional complex sub-Saharan African ancestry, those of the Natufians did not. Additional ancient genomes from microlithic groups across northern Africa and the Near East are necessary to investigate whether the shared Eurasian ancestry in Natufians and Iberomaurusian foragers is characteristic for microlithic foragers in these regions. However, if this assumption is accurate, the initial genomic results

presented in this dissertation suggest that the hunter-gatherers that manufactured the Iberomaurusian industry had an ultimate origin in the Near East.

2) The repopulation of Europe after the LGM. Manuscript B demonstrated that, in addition to previously established contributions from hunter-gatherer groups that carried WHG and EHG ancestry, a third ancestry from Magdalenian-associated foragers persisted among post-LGM European foragers. In Manuscript C it was shown that many Late Mesolithic hunter-gatherers from Europe carried an ancestry component related not only to EHGs but also to Near Eastern hunter-gatherers. This component was probably linked to the expansion of blade-and-trapeze industries out of eastern/southeastern Europe. Multiple expansion and admixture events after the LGM hence resulted in complex population genomic substructure among the post-glacial hunter-gatherers of Europe.

3) Direct population interactions of microlithic hunter-gatherers across the Mediterranean Sea. Archaeologists have proposed direct material connections between the European Gravettian and Magdalenian with contemporaneous North African Iberomaurusian assemblages during the last Ice Age, and between European Late Mesolithic blade-and-trapeze industries and the North African Capsian after the Ice Age. The combined body of evidence in this dissertation, however, does not provide genomic evidence for admixture signals that would indicate that such direct crossings occurred.

4) The demographic processes underlying the transition to farming in the central and western Mediterranean. The results in this dissertation, combined with findings in other archaeogenetic studies, have shown that the ancestry profiles of early farmers from Iberia and Sicily retained only minor amounts of local hunter-gatherer ancestry. This indicates that, similar to central Europe, the transition to agriculture in the Mediterranean involved large-scale population replacements of local foragers by early farmers during the Early Neolithic. However, the results in Manuscript C indicated that Late Mesolithic foragers in Sicily may have adopted initial elements of agriculture, prior to the near-total population replacement of farmers. Similarly, studies demonstrated a substantial continuation of Iberomaurusian-related ancestry in early farmers from Morocco. These cases taken together may present initial genomic evidence that acculturation of local foragers had a significant role in the early phases of Neolithisation in the central and western Mediterranean.

5) The southern Mediterranean Neolithic expansion route. The genomic evidence provided in this dissertation and other archaeogenetic studies so far does not provide support for a separate introduction of early Impressa Ware in Sicily, Andalusia and northern Africa. Moreover, the genomic evidence thus far does not underline a Maghrebi origin for the Early Neolithic in

Iberia, or support for a southern Mediterranean expansion route that involved early farmers crossing the Strait of Sicily or Gibraltar. However, some of the oldest farmers communities in the Mediterranean remain to be sampled, including in Sicily, Calabria in southern Italy and Andalucía in southern Iberia.

8.5. Outlook

North African population genomic history remains greatly understudied, despite its importance for understanding the deeper prehistoric past of Africans, as well as non-Africans. Further research is necessary to clarify the geographical and archaeological context of the ancestral group that harboured the Eurasian ancestry associated with the expansion of the early microlithic industries, and the directionality of the expansion (i.e. from northern Africa into the Near East or vice versa, e.g. Lazaridis et al., 2018). Obtaining additional ancient genomes from glacial and post-glacial microlithic groups across northern Africa (e.g. Dabban, Eastern Oranian) and the Near East (e.g. Kebaran) will be crucial in addressing these questions. Also, these additional ancient genomes could clarify demographic interactions between various archaeological groups, and provide essential context as to how the subsistence strategy developed and intensified towards sedentary farming in these regions. For example, genomes of Capsian foragers may provide important insights regarding the degree of population continuity with the preceding Iberomaurusian foragers and during the subsequent Neolithisation of Northwest Africa.

Regarding the post-LGM repopulation of Europe, future research could focus on investigating in more detail the expansion dynamics of foragers who manufactured the Upper Palaeolithic Magdalenian and Epigravettian industries. Currently, there are no genomes available for Epigravettian foragers east of the Adriatic Sea, including from the Balkans. However, the Epigravettian foragers in this region are prime candidates for the source of the EHG ancestry as found in the Villabruna hunter-gatherer and related individuals in central and western Europe from ~14,000 yBP onwards. Furthermore, further studies could aim at obtaining a more detailed understanding of the Mesolithic pan-European blade-and-trapeze complex expansion. This would especially be beneficial to future studies on the Neolithic transition in Europe and northern Africa, which can take advantage of the more fine-scaled Late Mesolithic genomic landscape to trace further details in the expansion dynamics and routes of the early farmers (e.g. Manuscript B; Olalde et al., 2015; Lipson et al., 2017; Rivollat et al., 2020).

Lastly, future research could explore the possibility that in Mediterranean coastal regions acculturation of local foragers played a more significant role in the Neolithisation process than in central Europe. The genomic transect for Sicily in Manuscript C may represent a valuable case study that Late Mesolithic hunter-gatherers adopted elements of early farming, including Impressa Ware pottery. Although rather rare, such cases have been reported before in the Balkans, such as at the Iron Gates in Serbia and Romania and *Malak Preslavets* in Bulgaria, and very recently for the French Mediterranean coast (Bonsall et al., 2015; Gronenborn et al., 2017; Lipson et al. 2017; Mathieson et al., 2018; Rivollat et al., 2020). Also, the few available ancestry profiles for early farmers in the Maghreb showed a strong population genomic continuity with the local Iberomaurusian foragers (Fregel et al., 2018). These cases taken together may turn out to form the initial genomic evidence of a more general trend that hunter-gatherers interacted with early farmers in central and western Mediterranean regions (Cortés Sánchez et al., 2012; Mulazzani et al., 2016; Isern et al., 2017; Dunne et al., 2019; Rivollat et al., 2020). It is important to note that in many of these studies, including Manuscript C, the acculturation of foragers could be detected only because of the joint analyses of genomic ancestry, stable isotope data and precise AMS radiocarbon dates. Hence, similar multi-disciplinary approaches may form the most powerful research strategy for future research that aims to obtain a comprehensive understanding of the ~10,000-year transition process from nomadic foraging to fully sedentary farming in the Mediterranean.

References

Ammerman, A. J. (2010). The first Argonauts: Towards the study of the earliest seafaring in the Mediterranean. In A. Anderson, J. H. Barrett, & K. V. Boyle (Eds.), *The global origins and development of seafaring* (pp. 81-92). Cambridge: McDonald Institute for Archaeological Research.

Arauna, L. R., Mendoza-Revilla, J., Mas-Sandoval, A., Izaabel, H., Bekada, A., Benhamamouch, S., . . . Comas, D. (2017). Recent Historical Migrations Have Shaped the Gene Pool of Arabs and Berbers in North Africa. *Mol Biol Evol*, *34*(2), 318-329. doi:10.1093/molbev/msw218.

Aura Tortosa, J. E. (1995). *El Magdaleniense mediterráneo: la Cova del parpalló* (Gandia, Valencia) (Vol. 91). Valencia: Diputación de Valencia.

Aura Tortosa, J. E., Tiffagom, M., Jordá Pardo, J. F., Duarte, E., Fernández de la Vega, J., Santamaria, D., . . . Perez Ripoll, M. (2012). The Solutrean–Magdalenian transition: A view from Iberia. *Quatern Int, 272-273*, 75-87. doi:10.1016/j.quaint.2012.05.020

Bailey, G., Galanidou, N., Peeters, H., Jöns, H., & Mennenga, M. (2020). *The Archaeology of Europe's Drowned Landscapes* (G. Bailey, N. Galanidou, H. Peeters, H. Jöns, & M. Mennenga Eds.). Cham, Switzerland: Springer.

Bar-Yosef, O. (1990). The Last Glacial Maximum in the Mediterranean Levant. In C. Gamble & O. Soffer (Eds.), T*he World at 18,000 BP* (Vol. 2, pp. 58–77). London: Unwin Hyman.

Bar-Yosef, O., & Meadow, R. H. (1995). The origins of agriculture in the Near East. In T. D. Price & A. B. Gebauer (Eds.), *Last Hunters, First Farmers* (pp. 372): School of American Research Press.

Barton, N., Bouzouggar, A., Collcutt, S. N., Gale, R., Higham, T. F. G., Humphrey, L. T., . . . Malek, F. (2005). The Late Upper Palaeolithic occupation of the Moroccan Northwest Maghreb during the Last Glacial Maximum. *Afr Archaeol Rev, 22*, 77–100. doi:10.1007/s10437-005-4190-y

Barton, N., & Bouzouggar, A. (2013). Hunter-Gatherers of the Maghreb 25,000–6,000 Years Ago. In P. Mitchell & P. L. Lane (Eds.), *The Oxford Handbook of African Archaeology* (pp. 1052). Oxford: Oxford University Press.

Bellwood, P. S. (2004). *First Farmers: the Origins of Agricultural Societies* (M. Malden Ed. First ed.): Wiley-Blackwell.

Bergström, A., McCarthy, S. A., Hui, R., Almarri, M. A., Ayub, Q., Danecek, P., . . . Tyler-Smith, C. (2020). Insights into human genetic variation and population history from 929 diverse genomes. *Science, 367*(6484), eaay5012. doi:10.1126/science.aay5012

Biagi, P., & Starnini, E. (2016). T*he Origin and Spread of the Late Mesolithic Blade and Trapeze Industries in Europe: Reconsidering J. G. D. Clark's Hypothesis Fifty Years After.* Kishinev.

Bollongino, R., Tresset, A., & Vigne, J.-D. (2008). Environment and excavation: Pre-lab impacts on ancient DNA analyses. *C R Palevol, 7*(2), 91-98. doi:10.1016/j.crpv.2008.02.002

Bonsall, C., Vasić, R., Boroneanţ, A., Roksandic, M., Soficaru, A., McSweeney, K., . . . Cook, G. (2015). New AMS 14C Dates for Human Remains from Stone Age Sites in the Iron Gates Reach of the Danube, Southeast Europe. *Radiocarbon, 57*(1), 33-46. doi:10.2458/ azu_rc.57.18188

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prüfer, K., . . . Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A, 104*(37), 14616–14621. doi:10.1073/pnas.0704665104

Briggs, A. W., Good, J. M., Green, R. E. G., Krause, J., Maricic, T., Stenzel, U., . . . Pääbo, S. (2009). Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. *Science, 325*(5938), 318-321. doi:10.1126/science.1174462

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., & Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res, 38*(6), e87. doi:10.1093/nar/gkp1163

Broushaki, F., Thomas, M. G., Link, V., López, S., van Dorp, L., Kirsanow, K., . . . Burger, J. (2016). Early Neolithic genomes from the eastern Fertile Crescent. *Science, 353*(6298), 499. doi:10.1126/ science.aaf7943

Brunel, S., Bennett, E. A., Cardin, L., Garraud, D., Barrand Emam, H., Beylier, A., . . . Pruvost, M. (2020). Ancient genomes from present-day France unveil 7,000 years of its demographic history. *Proc Natl Acad Sci U S A, 117*(23), 12791. doi:10.1073/pnas.1918034117

Burbano, H. A., Hodges, E., Green, R. E., Briggs, A. W., Krause, J., Meyer, M., . . . Pääbo, S. (2010). Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. *Science, 328*(5979), 723-725. doi:10.1126/science.1188046

Carpenter, M., Buenrostro, J., Valdiosera, C., Schroeder, H., Allentoft, M., Sikora, M., . . . Bustamante, C. (2013). Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *Am J Hum Genet, 93*(5), 852-864. doi:10.1016/j.ajhg.2013.10.002

Cascalheira, J., Alcaraz-Castaño, M., Alcolea-González, J., de Andrés-Herrero, M., Arrizabalaga, A., Aura Tortosa, J. E., . . . Iriarte-Chiapusso, M.-J. (2020). Paleoenvironments and human adaptations during the Last Glacial Maximum in the Iberian Peninsula: A review. *Quatern Int.* doi:10.1016/j.quaint.2020.08.005

Castaño, M. A. (2007). El Ateriense del Norte de Africa y el Solutrense peninsular: ¿contactos transgibraltareños en el Pleistoceno Superior? *Munibe Antropologia - Arkeologia, 58*, 101-126.

Clark, G., & Piggott, S. (1968). *Prehistoric Societies* (3rd ed.). London: Hutchinson & Co.

Clarkson, C., Hiscock, P., Mackay, A., & Shipton, C. (2018). Small, sharp, and standardized: global convergence in backed-microlith technology. In M. O'Brian, B. Buchanan, & M. I. Eren (Eds.), *Convergent evolution in stone-tool technology* (pp. 304): The MIT Press.

Cooper, A., & Poinar, H. N. (2000). Ancient DNA: Do It Right or Not at All. *Science, 289*(5482), 1139. doi:10.1126/science.289.5482.1139b

Cortés Sánchez, M., Jiménez Espejo, F. J., Simón Vallejo, M. D., Gibaja Bao, J. F., Carvalho, A. F., Martinez-Ruiz, F., . . . Bicho, N. F. (2012). The Mesolithic–Neolithic transition in southern Iberia. *Quat Res, 77*(2), 221-234. doi:10.1016/j.yqres.2011.12.003

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., . . . 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*(39), 15758–15763. doi:10.1073/pnas. 1314445110

Dastugue, J. (1962). II. Pathologie osseuse. In D. Ferembach, J. Dastugue, & M.-J. Poitrat-Targowla (Eds.), *La nécropole épipaléolithique de Taforalt (Maroc oriental); étude des squelettes humains* (pp. 135-158). CNRS, Paris: Edita-Casablanca, Rabat.

De Groote, I., & Humphrey, L. T. (2016). Characterizing evulsion in the Later Stone Age Maghreb: Age, sex and effects on mastication. *Quatern Int, 413*(Part A), 50-61. doi:10.1016/ j.quaint.2015.08.082

Djindjian, F. (2016). Territories and economies of hunter–gatherer groups during the last glacial maximum in Europe. *Quatern Int, 412*, 37-43. doi:10.1016/j.quaint.2015.06.058

Douka, K., Jacobs, Z., Lane, C., Grün, R., Farr, L., Hunt, C., . . . Barker, G. (2014). The chronostratigraphy of the Haua Fteah cave (Cyrenaica, northeast Libya). *J Hum Evo, 66*, 39-63. doi:10.1016/j.jhevol. 2013.10.001

Dunne, J., Manning, K., Linstädter, J., Mikdad, A., Breeze, P., Hutterer, R., . . . Evershed, R. P. (2019). Pots, plants and animals: Broad-spectrum subsistence strategies in the Early Neolithic of the Moroccan Rif region. *Quatern Int.* doi:10.1016/j.quaint.2019.12.009

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Mol Biol Evol, 28*(8), 2239-2252. doi:10.1093/molbev/msr048

Durvasula, A., & Sankararaman, S. (2020). Recovering signals of ghost archaic introgression in African populations. *Sci Adv, 6(*7), eaax5097. doi:10.1126/sciadv.aax5097

Feldman, M., Fernández-Domínguez, E., Reynolds, L., Baird, D., Pearson, J., Hershkovitz, I., . . . Krause, J. (2019). Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. *Nat Commun, 10*(1), 1218. doi:10.1038/s41467-019-09209-7

Fernández-López de Pablo, J., Gutiérrez-Roig, M., Gómez-Puche, M., McLaughlin, R., Silva, F., & Lozano, S. (2019). Palaeodemographic modelling supports a population bottleneck during the Pleistocene-Holocene transition in Iberia. *Nat Commun, 10*(1), 1872. doi:10.1038/s41467-019-09833-3

Fregel, R., Méndez, F., Bokbot, Y., Martín-Socas, D., Camalich-Massieu, M., Santana J, M. J., . . . Bustamante, C. D. (2018). Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proc Natl Acad Sci U S A, 115*(26), 6774-6779. doi:10.1073/pnas.1800851115

Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H. A., Kelso, J., & Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci U S A, 110*(6), 2223-2227. doi:10.1073/pnas.1221359110

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S., Bondarev, A., . . . Pääbo, S. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature, 514*(7523), 445-459.

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., . . . Reich, D. (2016). The genetic history of Ice Age Europe. *Nature, 534*, 200-205. doi:10.1038/nature17993

Furtwängler, A., Reiter, E., Neumann, G. U., Siebke, I., Steuri, N., Hafner, A., . . . Krause, J. (2018). Ratio of mitochondrial to nuclear DNA affects contamination estimates in ancient DNA analysis. *Sci Rep, 8*, 14075. doi:10.1038/s41598-018-32083-0

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R., L, Gonzales-Fortes, G., Mattiangeli, V., . . . Pinhasi, R. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun, 5*, 5257-5265. doi:10.1038/ncomms6257

Gamble, C. (1993). In J. C. Chapman & P. M. Dolukhanov (Eds.), *Cultural Transformations and Interactions in Eastern Europe* (pp. 256). United Kingdom: Taylor & Francis Ltd

Gansauge, M.-T., & Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc, 8*(4), 737–748. doi:10.1038/nprot. 2013.038.

Gansauge, M.-T., Gerber, T., Glocke, I., Korlevic, P., Lippik, L., Nagel, S., . . . Meyer, M. (2017). Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res, 45*(10), e79. doi:10.1093/nar/gkx033

Garcea, E. A. A. (2010). The Lower and Upper Later Stone Age of North Africa. In E. A. A. Garcea (Ed.), *South-Eastern Mediterranean Peoples Between 130,000 and 10,000 Years Ago* (pp. 54-65): Oxbow Books.

Gaudio, D., Fernandes, D. M., Schmidt, R., Cheronet, O., Mazzarelli, D., Mattia, M., . . . Pinhasi, R. (2019). Genome-Wide DNA from Degraded Petrous Bones and the Assessment of Sex and Probable Geographic Origins of Forensic Cases. *Sci Rep, 9*(1), 8226. doi:10.1038/ s41598-019-44638-w

Gilbert, M., Bandelt, H., Hofreiter, M., & Barnes, I. (2005). Assessing ancient DNA studies. *Trends Ecol Evol, 20*(10), 541–544. doi:10.1016/j.tree.2005.07.005

González-Fortes, G., Jones, E., Lightfoot, E., Bonsall, C., Lazar, C., Grandal-d'Anglade, A., . . . Hofreiter, M. (2017). Paleogenomic Evidence for Multi-generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin. *Curr Biol, 27*(12), 1801-1810. doi: 10.1016/j.cub.2017.05.023

Green, R. E., Krause, J., Ptak, S., Briggs, A., Ronan, M., Simons, J., . . . Pääbo, S. (2006). The Neandertal genome and ancient DNA authenticity. *Nature, 444*(7117), 330—336. doi: 10.1038/nature05336.

Green, R. E., Malaspinas, A.-S., Krause, J., Briggs, A., Johnson, P., Uhler, C., . . . Pääbo, S. (2008). A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell, 134*(3), 416–426. doi:10.1016/j.cell.2008.06.021

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., . . . Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science, 328*(5979), 710. doi:10.1126/ science.1188021

Gronenborn, D. (2017). *Migrations before the Neolithic? The Late Mesolithic blade-and-trapeze horizon in central Europe and beyond* (H. Meller, F. Daim, J. Krause, & R. Risch Eds. Vol. 17). Saale: Tagungen des Landesmuseums für Vorgeschichte Halle (Saale).

Gronenborn, D., & Horejs, B. (Cartographer). (2019). *Map: Expansion of Farming in Western Eurasia, 9600 - 4000 Cal BC*. Retrieved from https://www.academia.edu/9424525/Map_ Expansion_of_farming_in_western_Eurasia_9600_4000_cal_BC_update_vers_2019_3_

Guilaine, J., & Manen, C. (2007). From Mesolithic to Early Neolithic in the western Mediterranean. In A. Whittle & V. Cummings (Eds.), *Going Over: The Mesolithic-Neolithic Transition in North-West Europe* (pp. 21–51). Oxford: Oxford University Press.

Guilaine, J. (2018). A personal view of the neolithisation of the Western Mediterranean. *Quatern Int, 470*, 211-225. doi:10.1016/j.quaint.2017.06.019

Günther, T., Valdiosera, C., Malmström, H., Ureña, I., Rodriguez-Varela, R., Sverrisdóttir, Ó. O., . . . Jakobsson, M. (2015). Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc Natl Acad Sci U S A, 112*(38), 11917–11922. doi:10.1073/ pnas.1509851112

Günther, T., Malmström, H., Svensson, E. M., Omrak, A., Sánchez-Quinto, F., Kılınç, G. M., . . . Jakobsson, M. (2018). Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol, 16*(1), e2003703. doi:10.1371/journal. pbio.2003703

Günther, T., & Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genet 15*(7), e1008302. doi:10.1371/journal. pgen.1008302

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., . . . Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature, 522*(7555), 207-2011. doi:10.1038/nature14317

Hammer, M. F., Woerner, A. E., Mendez, F. L., Watkins, J. C., & Wall, J. D. (2011). Genetic evidence for archaic admixture in Africa. *Proc Natl Acad Sci U S A, 108*(37), 15123. doi: 10.1073/pnas.11093 00108

Hansen, H. B., Damgaard, P. B., Margaryan, A., Stenderup, J., Lynnerup, N., Willerslev, E., & Allentoft, M. E. (2017). Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. *PLoS ONE, 12(*1), e0170940. doi:10.1371/journal.pone.0170940

Henn, B. M., Botigue, L. R., Gravel, S., Bustamante, C. D., & Comas, D. (2012). Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLoS Genet, 8*, e1002397. doi:10.1371/ journal.pgen.1002397

Hertler, C., Bruch, A., & Märker, M. (2014). The earliest stages of hominin dispersal in Africa and Eurasia. In P. Bellwood (Ed.), *The Global Prehistory of Human Migration* (pp. 458): Wiley-Blackwell.

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

Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-del-Molino, D., . . . Burger, J. (2016). Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc Natl Acad Sci U S A, 113*, 6886-6891. doi:10.1073/pnas.1523951113

Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, D., & Pääbo, S. (2001). DNA Sequences From Multiple Amplifications Reveal Artifacts Induced by Cytosine Deamination in Ancient DNA. *Nucleic Acids Res, 29*(23), 4793-4799. doi: 10.1093/nar/29.23.4793

Hogue, J. T. (2014). *The origin and development of the Pleistocene LSA in Northwest Africa: A case study from Grotte des Pigeons (Taforalt), Morocco.* (Doctoral). University of Oxford, Oxford. Retrieved from https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.677978

Hogue, J. T., & Barton, R. (2016). New radiocarbon dates for the earliest Later Stone Age microlithic technology in Northwest Africa. *Quatern Int, 413*(Part A), 62-75. doi:10.1016/ j.quaint.2015.11.144

Howitt-Marshall, D., & Runnels, C. (2016). Middle Pleistocene sea-crossings in the eastern Mediterranean? *J Anthropol Archaeol, 42*, 140-153. doi:10.1016/j.jaa.2016.04.005

Hublin, J.-J. (2017). New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. *Nature, 546*, 289-291.

Hughes, P. D., Woodward, J. C., & Gibbard, P. L. (2006). Quaternary glacial history of the Mediterranean mountains. Prog Phys Geogr: *Earth and Environment, 30*(3), 334-364. doi: 10.1191/0309133306 pp481ra

Humphrey, L., Bello, S. M., Turner, A., Bouzouggar, A., & Barton, R. (2012). Iberomaurusian funerary behaviour: Evidence from Grotte des Pigeons, Taforalt, Morocco. *J Hum Evol, 62*(2), 261-273. doi:10.1016/j.jhevol.2011.11.003.

Ibáñez-Estévez, J. J., Gibaja Bao, J. F., Gassin, B., & Mazzucco, N. (2017). Paths and Rhythms in the spread of agriculture in the Western Mediterranean: the contribution of the analysis of harvesting technology. In O. García-Puchol & D. C. Salazar García (Eds.), *Times of Neolithic Transition along the Western Mediterranean* (pp. 417). Cham, Switzerland: Springer International Publishing.

Ibáñez-Estévez, J. J., González-Urquijo, J., Teira-Mayolini, L. C., & Lazuén, T. (2018). The emergence of the Neolithic in the Near East: A protracted and multi-regional model. *Quatern Int, 470,* 226-252. doi:10.1016/j.quaint.2017.09.040

Inizan, M.-L. (2012). Pressure Débitage in the Old World: Forerunners, Researchers, Geopolitics – Handing on the Baton. In P. M. Desrosiers (Ed.), *The Emergence of Pressure Blade Making: From Origin to Modern Experimentation* (pp. 11-42). New York: Springer-Verlag.

Isern, N., Zilhão, J., Fort, J., & Ammerman, A. J. (2017). Modeling the role of voyaging in the coastal spread of the Early Neolithic in the West Mediterranean. *Proc Natl Acad Sci U S A, 114*(5), 897. doi:10.1073/pnas.1613413114

Jones, E. R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., . . . Bradley, D. G. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun, 6*, 8912-8919. doi:10.1038/ncomms9912

Jones, E. R., Zarina, G., Moiseyev, V., Lightfoot, E., Nigst, P., Manica, A., . . . Bradley, D. (2017). The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. *Curr Biol, 27*(4), 576-582. doi:10.1016/j.cub.2016.12.060.

Jones, S., Antoniadou, A., Barton, H., Drake, N., Farr, L., Hunt, C., . . . Barker, G. (2016). Patterns of Hominin Occupation and Cultural Diversity Across the Gebel Akhdar of Northern Libya Over the Last ~200 kyr. In S. C. Jones & B. A. Stewart (Eds.), *Africa from MIS 6-2: Population Dynamics and Paleoenvironments* (pp. 77-99). Dordrecht: Springer Netherlands.

Kendall, C., Eriksen, A. M. H., Kontopoulos, I., Collins, M. J., & Turner-Walker, G. (2018). Diagenesis of archaeological bone and tooth. P*alaeogeogr Palaeoclimatol Palaeoecol, 491,* 21-37. doi:10.1016/j.palaeo.2017.11.041

Kılınç, G., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, A., Bıçakçı, E., . . . Götherström, A. (2016). The Demographic Development of the First Farmers in Anatolia. *Curr Biol, 26*(19), 2659-2666. doi:10.1016/j.cub.2016.07.057

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res, 40*(1), e3. doi:10.1093/nar/gkr771

Kleindienst, M. R. (2006). On Naming Things. In E. Hovers & S. Kuhn (Eds.), *Transitions Before the Transition: Evolution and Stability in the Middle Paleolithic and Middle Stone Age* (pp. 13-28). New York: Springer-Verlag

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics, 15*, 356-368. doi:10.1186/s12859-014-0356-4

Krause, J., Briggs, A. W., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A. P., & Pääbo, S. (2010). A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. *Curr Biol, 20*, 231–236.

Lachance, J., Vernot, B., Elbers, Clara C., Ferwerda, B., Froment, A., Bodo, J.-M., . . . Tishkoff, Sarah A. (2012). Evolutionary History and Adaptation from High-Coverage Whole-Genome Sequences of Diverse African Hunter-Gatherers. *Cell, 150*(3), 457-469. doi:10.1016/j.cell.2012.07.009

Lachance, J., & Tishkoff, S. A. (2013). SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. *Bioessays, 35*(9), 780-786. doi:10.1002/bies.201300014

Lambeck, K., Rouby, H., Purcell, A., Sun, Y., & Sambridge, M. (2014). Sea level and global ice volumes from the Last Glacial Maximum to the Holocene. *Proc Natl Acad Sci U S A, 111*(43), 15296. doi:10.1073/pnas.1411762111

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., . . . J., K. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature, 513*(7518), 409–413. doi:10.1038/nature13673

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D., Rohland, N., Mallick, S., . . . Reich, D. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature, 536*(7617), 419–424. doi:10.1038/nature19310

Lazaridis, I., Belfer-Cohen, A., Mallick, S., Patterson, N., Cheronet, O., Rohland, N., . . . Reich, D. (2018). Paleolithic DNA from the Caucasus reveals core of West Eurasian ancestry. *BiorXiv*. doi:10.1101/423079

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., . . . Myers, R. M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science, 319*(5866), 1100-1104. doi:10.1126/science.1153717

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

Linstädter, J. (2008). The Epipalaeolithic-Neolithic-Transition in the Mediterranean region of Northwest Africa. *Quartär, 55*, 41-62.

Linstädter, J., Eiwanger, J., Mikdad, A., & Weniger, G.-C. (2012). Human occupation of Northwest Africa: A review of Middle Palaeolithic to Epipalaeolithic sites in Morocco. *Quatern Int, 274*, 158-174. doi:10.1016/j.quaint.2012.02.017

Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., . . . Reich, D. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature, 551*(7680), 368-372. doi:10.1038/nature24476

Llorente, M. G., Jones, E. R., Eriksson, A., Siska, V., Arthur, K. W., Arthur, J. W., . . . Manica, A. (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science, 350*(6262), 820. doi:10.1126/science.aad2879

Lo Presti, V., Antonioli, F., Palombo, M. R., Agnesi, V., Biolchi, S., Calcagnile, L., . . . Tusa, S. (2019). Palaeogeographical evolution of the Egadi Islands (western Sicily, Italy). Implications for late Pleistocene and early Holocene sea crossings by humans and other

mammals in the western Mediterranean. *Earth-Sci. Rev, 194*, 160-181. doi:10.1016/j.earscirev.2019.04.027

Lubell, D., Sheppard, P., & Jackes, M. (1984). Continuity in the Epipaleolithic of Northern Africa with Emphasis on the Maghreb. *Advances in World Archaeology, 3*, 143-191.

Maca-Meyer, N., González, A. M., Pestano, J., Flores, C., Larruga, J. M., & Cabrera, V. M. (2003). Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. *BMC Genet, 4*, 15-15. doi:10.1186/1471-2156-4-15

Maier, A. (2015). *The Central European Magdalenian: Regional Diversity and Internal Variability* (1st ed.). Dordrecht: Springer.

Manen, C., Marchand, G., & Calvalho, A. F. (2007). *Le Néolithique ancien de la Péninsule Ibérique: vers une nouvelle évaluation du mirage africain?* Paper presented at the Congrès du Centenaire: Un siècle de construction du discours scientifique en Préhistoire, Avignon.

Marchand, G., & Manen, C. (2009, 2009). *Mésolithique final et Néolithique ancien autour du détroit : une perspective septentrionale (Atlantique / Méditerranée).* Paper presented at the The last hunter-gatherers and the first farming communities in the South of the Iberian peninsula and North of Morocco, Workshop, Faro, novembre 2009. Faro, Universidade do Algarve, Faro, Portugal.

Mardis, E. R. (2008). Next-Generation DNA Sequencing Methods. *Annu Rev Genomics Hum Genet, 9*, 387-402. doi:10.1146/annurev.genom.9.081307.164359

Maricic, T., Whitten, M., & Pääbo, S. (2010). Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLoS ONE, 5*(11), e14004. doi:10.1371/journal.pone.0014004

Martínez-Sánchez, R. M., Vera-Rodríguez, J. C., Pérez-Jordà, G., Peña-Chocarro, L., & Bokbot, Y. (2018). The beginning of the Neolithic in northwestern Morocco. *Quatern Int, 470*, 485-496. doi:10.1016/j.quaint.2017.05.052

Martiniano, R., Cassidy, L., Ó'Maoldúin, R., McLaughlin, R., Silva, N., Manco, L., . . . Bradley, D. (2017). The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genet, 13*(7), e1006852. doi:10.1371/journal.pgen.1006852

Martiniano, R., Garrison, E., Jones, E. R., Manica, A., & Durbin, R. (2020). Removing reference bias and improving indel calling in ancient DNA data analysis by mapping to a sequence variation graph. *BiorXiv*, 782755. doi:10.1101/782755

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., . . . Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature, 528*(7583), 499–503. doi:10.1038/nature16152

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., . . . Reich, D. (2018). The genomic history of southeastern Europe. *Nature, 555*(7695), 197-203. doi:10.1038/nature25778

McVean, G. A., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., . . . The Genomes Project, C. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature, 491*(7422), 56-65. doi:10.1038/nature11632

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc, 6*, pdb.prot5448. doi:10.1101/pdb.prot5448.

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., . . . Pääbo, S. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science, 338(*6104), 222-226. doi:10.1126/science.1224344

Meyer, M., Arsuaga, J.-L., de Filippo, C., Nagel, S., Aximu-Petri, A., Nickel, B., . . . Pääbo, S. (2016). Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature, 531*, 504–507. doi:10.1038/nature17405

Mittnik, A., Wang, C. C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., . . . Krause, J. (2018). The genetic prehistory of the Baltic Sea region. *Nat Commun, 9*(442). doi:10.1038/s41467-018-028259

Mulazzani, S., Belhouchet, L., Salanova, L., Aouadi, N., Dridi, Y., Eddargach, W., . . . Zoughlami, J. (2016). The emergence of the Neolithic in North Africa: A new model for the Eastern Maghreb. *Quatern Int, 410*, 123-143. doi:10.1016/j.quaint.2015.11.089

Naudinot, N., Tomasso, A., Messager, E., Finsinger, W., Ruffaldi, P., & Langlais, M. (2017). Between Atlantic and Mediterranean: Changes in technology during the Late Glacial in Western Europe and the climate hypothesis. *Quatern Int, 428*, 33-49. doi:10.1016/j.quaint.2016.01.056

Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature, 541*(7637), 302-310. doi:10.1038/nature21347

Olalde, I., Allentoft, M. E., Sánchez-Quinto, F., Santpere, G., Chiang, C. W., DeGiorgio, M., . . . Lalueza-Fox, C. (2014). Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature, 507*(7491), 225-228. doi:10.1038/nature12960

Olalde, I., Schroeder, H., Sandoval-Velasco, M., Vinner, L., Lobón, I., Ramirez, O., . . . Lalueza-Fox, C. (2015). A common genetic origin for early farmers from Mediterranean Cardial and Central European LBK cultures. *Mol Biol Evol, 32*(12), 3132-3142. doi:10.1093/molbev/msv181

Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., . . . Reich, D. (2019). The genomic history of the Iberian Peninsula over the past 8000 years. *Science, 363*(6432), 1230. doi:10.1126/science.aav4040

Olalde, I., & Posth, C. (2020). Latest trends in archaeogenetic research of west Eurasians. *Curr Opin Genet Dev, 62*, 36-43. doi:10.1016/j.gde.2020.05.021

Olivieri, A., Achilli, A., Pala, M., Bandelt, H.-J., Fornarino, S., Al-Zahery, N., . . . Torroni, A. (2006). The mtDNA legacy of the Levantine Early Upper Palaeolithic in Africa. *Science, 314*(5806), 1767-1770. doi:10.1126/science.1135566

Omrak, A., Günther, T., Valdiosera, C., Svensson, E. M., Malmström , H., Kiesewetter, H., . . . Götherström, A. (2016). Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. *Curr Biol, 26*(2), 270-275. doi:10.1016/j.cub.2015.12.019

Özdoğan, M. (1997). The beginning of Neolithic economies in southeastern Europe: an Anatolian perspective. *Journal of European Archaeology Archive, 5*(2), 1-33. doi:10.1179/096576697800660267

Özdoğan, M. (2011). Archaeological Evidence on the Westward Expansion of Farming Communities from Eastern Anatolia to the Aegean and the Balkans. *Curr Anthropol, 52*, S415-S430. doi:10.1086/658895

Özdoğan, M. (2014). A new look at the introduction of the Neolithic way of life in Southeastern Europe. Changing paradigms of the expansion of the Neolithic way of life. *Doc Praehist, 41*, 33-49. doi:10.4312/dp.41.2

Pääbo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. *Nature, 314,* 644–645.

Pääbo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci U S A, 86*(6), 1939-1943. doi:10.1073/pnas.86.6.1939

Pallary, P. (1909). *Instructions pour la recherche préhistorique dans le Nord-Ouest de l'Afrique. In Mémoires de la Société Historique Algérienne* (Vol. No. 3). Algiers: Jourdan.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., . . . Reich, D. (2012). Ancient admixture in human history. *Genetics, 192*(3), 1065-1093. doi:10.1534/genetics.112.145037

Pereira, L., Silva, N. M., Franco-Duarte, R., Fernandes, V., Pereira, J. B., Costa, M. D., . . . Macaulay, V. (2010). Population expansion in the North African Late Pleistocene signalled

by mitochondrial DNA haplogroup U6. *BMC Evol Biol, 10*(1), 390. doi:10.1186/1471-2148-10-390

Pericot, L. (1955). *Sobre el problema de las relaciones preneolíticas entre España y Marruecos*. Paper presented at the Actas dell Congreso Arqueológico del Marruecos español, Tetuán, 1953.

Perrin, T., & Binder, D. (2014). *Le Mésolithique à trapèzes et la néolithisation de l'Europe sud-occidentale*. Paper presented at the La transition néolithique en Méditerranée, Paris.

Petrullo, G., & Delaplace, A. (2020). Common Cultural Markers in the Bone and Lithic Production of the Upper Capsian: A Comparative Approach. *Afr Archaeol Rev.* doi:10.1007/s10437-020-09382-x

Peyrégne, S., & Prüfer, K. (2020). Present-Day DNA Contamination in Ancient DNA Datasets. *Bioessays, 42*(9), 2000081. doi:10.1002/bies.202000081

Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., . . . Hofreiter, M. (2015). Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS ONE, 10*(6), e0129102. doi:10.1371/journal.pone.0129102

Poinar, H. N., Schwarz, C., Qi, J., Shapiro, B., Macphee, R. D., Buigues, B., . . . Schuster, S. C. (2006). Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science, 311*(5759), 392-394. doi:10.1126/science.1123360

Posth, C., Renaud, G., Mittnik, A., Drucker, D. G., Rougier, H., Cupillard, C., . . . Krause, J. (2016). Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a Late Glacial population turnover in Europe. *Curr Biol, 26*(6), 827-833. doi:10.1016/j.cub.2016.01.037

Potì, A., & Weniger, G.-C. (2019). Human Occupation of Northern Morocco at the Last Glacial Maximum. In A. Schmidt, J. o. Cascalheira, N. Bicho, & G.-C. Weniger (Eds.), *Human Adaptations to the Last Glacial Maximum: The Solutrean and its Neighbors* (pp. 44-64). Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK: Cambridge Scholars Publishing.

Price, T. D. (2000). *Europe's First Farmers* (T. D. Price Ed.). Cambridge, United Kingdom: Cambridge University Press.

Prüfer, K., Stenzel, U., Hofreiter, M., Pääbo, S., Kelso, J., & Green, R. E. (2010). Computational challenges in the analysis of ancient DNA. *Genome Biol, 11*(5), R47. doi:10.1186/gb-2010-11-5-r47

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., . . . Willerslev, E. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature, 505*(7481), 87-91. doi:10.1038/nature12736

Rahmani, N. (2004). Technological and Cultural Change Among the Last Hunter-Gatherers of the Maghreb: The Capsian (10,000–6000 B.P.). *J World Prehist, 18*, 57-105. doi: 10.1023/B: JOWO.0000038658.50738.eb

Rasmussen, S. O., Andersen, K. K., Svensson, A. M., Steffensen, J. P., Vinther, B. M., Clausen, H. B., . . . Ruth, U. (2006). A new Greenland ice core chronology for the last glacial termination. J *Geophys Res Atmos, 111*(D6). doi:10.1029/2005jd006079

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature, 461(7*263), 489-494. doi:10.1038/nature08365

Renaud, G., Slon, V., Duggan, A. T., & Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol, 16*, 224-241. doi:10.1186/s13059-015-0776-0

Richter, D., Grün, R., Joannes-Boyau, R., Steele, T. E., Amani, F., Rué, M., . . . McPherron, S. P. (2017). The age of the hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. *Nature, 546(7*657), 293-296. doi:10.1038/nature22335

Rivollat, M., Jeong, C., Schiffels, S., Küçükkalıpçı, İ., Pemonge, M.-H., Rohrlach, A. B., . . . Haak, W. (2020). Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. *Sci Adv, 6*(22), eaaz5344. doi:10.1126/ sciadv.aaz5344

Rizzi, E., Lari, M., Gigli, E., De Bellis, G., & Caramelli, D. (2012). Ancient DNA studies: new perspectives on old samples. *Genet Sel Evol, 44*. doi:10.1186/1297-9686-44-21

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., & Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Phil Trans R Soc B, 370*(1660), 20130624.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE, 7*(3), e34131. doi:10.1371/journal.pone.0034131

Scarre, C. (2013). *The human past: world prehistory & the development of human societies* (C. Scarre Ed. Third ed.). London: Thames & Hudson.

Scerri, E. M. L. (2017). The North African Middle Stone Age and its place in recent human evolution. *Evol Anthropol, 26*(3), 119-135. doi:10.1002/evan.21527

Schlebusch, C. M., Malmström, H., Günther, T., Sjödin, P., Coutinho, A., Edlund, H., . . . Jakobsson, M. (2017). Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science, 358*(6363), 652-655. doi:10.1126/science.aao6266

Schmitt, T. (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology* (Volume 4), 11. doi:10.1186/1742-9994-4-11

Schuenemann, V., Peltzer, A., Welte, B., van Pelt, W., Molak, M., Wang, C., . . . Krause, J. (2017). Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nat Commun, 8*, 15694-15704. doi:10.1038/ncomms15694

Seguin-Orlando, A., Korneliussen, T. S., Sikora, M., Malaspinas, A. S., Manica, A., Moltke, I., . . . Willerslev, E. (2014). Genomic structure in Europeans dating back at least 36,200 years. *Science, 346*(6213), 1113-1118. doi:10.1126/science.aaa0114

Serra-Vidal, G., Lucas-Sanchez, M., Fadhlaoui-Zid, K., Bekada, A., Zalloua, P., & Comas, D. (2019). Heterogeneity in Palaeolithic Population Continuity and Neolithic Expansion in North Africa. *Curr Biol, 29*(22), 3953-3959.e3954. doi:10.1016/j.cub.2019.09.050

Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., . . . Jakobsson, M. (2012). Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science, 336*(6080), 466. doi:10.1126/science.1216304

Skoglund, P., Malmström, H., Omrak, A., Raghavan, M., Valdiosera, C., Günther, T., . . . Jakobsson, M. (2014). Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science, 344*(6185), 747-750. doi:10.1126/science. 1253448

Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., . . . Reich, D. (2016). Genomic insights into the peopling of the Southwest Pacific. *Nature, 538*(7626), 510-513. doi: 10.1038/nature19844

Skoglund, P., Thompson, J. C., Prendergast, M. E., Mittnik, A., Sirak, K., Hajdinjak, M., . . . Reich, D. (2017). Reconstructing Prehistoric African Population Structure. *Cell, 171*(1), 59–71. e21. doi: 10.1016/j.cell.2017.08.049

Skoglund, P., & Mathieson, I. (2018). Ancient Genomics of Modern Humans: The First Decade. *Annu Rev Genomics Hum Genet, 19*, 381-404. doi:10.1146/annurev-genom-083117 021749

Slatkin, M. (1985). Rare Alleles as Indicators of Gene Flow. *Evolution, 39*(1), 53-65. doi:10.1111/j.1558-5646.1985.tb04079.x

Slatkin, M., & Racimo, F. (2016). Ancient DNA and human history. *Proc Natl Acad Sci U S A, 113*(23), 6380-6387.doi:10.1073/pnas.1524306113

Smith, C. I., Chamberlain, A. T., Riley, M. S., Stringer, C., & Collins, M. J. (2003). The Thermal History of Human Fossils and the Likelihood of Successful DNA Amplification. *J Hum Evol, 45*(3), 203-217. doi:10.1016/s0047-2484(03)00106-4

Stewart, J., & Stringer, C. (2012). Human Evolution Out of Africa: The Role of Refugia and Climate Change. *Science, 335*(6074), 1317-1321. doi:10.1126/science.1215627

Straus, L. G. (2001). Africa and Iberia in the Pleistocene. *Quatern Int, 75*(1), 91-102. doi:10.1016/S1040-6182(00)00081-1

Valdiosera, C., Günther, T., Vera-Rodríguez, J. C., Ureña, I., Iriarte, E., Rodríguez-Varela, R., . . . Jakobsson, M. (2018). Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. *Proc Natl Acad Sci U S A, 115*(13), 3428-3433. doi:10.1073/pnas.1717762115

van de Loosdrecht, M.S., Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., . . . Krause, J. (2018). Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science, 360*(6388), 548-552. doi:10.1126/science.aar8380

van de Loosdrecht, M. S., Mannino, M. A., Talamo, S., Villalba-Mouco, V., Posth, C., Aron, F., . . . Krause, J. (2020). Genomic and dietary transitions during the Mesolithic and Early Neolithic in Sicily. *BiorXiv*, 2020.2003.2011.986158. doi:10.1101/2020.03.11.986158

Villalba-Mouco, V., van de Loosdrecht, M. S., Posth, C., Mora, R., Martínez-Moreno, J., Rojo-Guerra, M., . . . Haak, W. (2019). Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula. *Curr Biol, 29*(7), 1169-1177. doi:10.1016/j.cub.2019.02.006

Villaverde Bonilla, V., Román, D., Ripoll, M. P., Bergadà, M. M., & Real, C. (2012). The end of the Upper Palaeolithic in the Mediterranean Basin of the Iberian Peninsula. Q*uatern Int, 272*(272-273), 272-273, 17-32. doi: 10.1016/j.quaint.2012.04.025

Walker, M., Johnsen, S., Rasmussen, S. O., Popp, T., Steffensen, J.-P., Gibbard, P., . . . Schwander, J. (2009). Formal definition and dating of the GSSP (Global Stratotype Section and Point) for the base of the Holocene using the Greenland NGRIP ice core, and selected auxiliary records. *J Quat Sci, 24*(1), 3-17. doi:10.1002/jqs.1227

Whittle, A. (1996). *Europe in the Neolithic.* The Creation of New Worlds. Cambridge: Cambridge University Press.

Woodward, S., Weyand, N., & Bunnell, M. (1994). DNA sequence from Cretaceous period bone fragments. *Science, 226*(5188), 1229-1232. doi:10.1126/science.7973705

Zampetti, D. (1989). La question des rapports entre la Sicile et l'Afrique du Nord pendant le Paleolithique superieur final: la contribution de l'archeologie. In I. Hershkovitz (Ed.), *People and Culture in Change.* Proceedings of the Second Symposium on Upper Palaeolithic, Mesolithic and Neolithic populations in Europe and the Mediterranean Basin. (Vol. 1, pp. 459-475). Oxford: BAR Int. Ser.

Zilhão, J. (2001). Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proc Natl Acad Sci U S A, 98*(24), 14180-14185. doi: 10.1073/pnas.2415 22898

Zilhão, J. (2014). Early prehistoric navigation in the Western Mediterranean: Implications for the Neolithic transition in Iberia and the Maghreb. *Eurasian Prehistory, 11*(1-2), 185-200. doi: http://hdl.handle.net/10451/31120

Zvelebil, M. (2001). The agricultural transition and the origins of Neolithic society in Europe. *Doc Praehist, 28*, 1-26. doi:10.4312/dp.28.1

Summary

Hunter-gatherers and prehistoric farmers depended intimately on the rich biodiversity and relatively hospitable climate of the Mediterranean environment. The inhabitants of the region had a central role in the repopulation of Europe and northern Africa after the Last Glacial Maximum (LGM) ~26,500-19,000 years before present (yBP), and in the subsequent development and spread of agricultural practises.

In this dissertation research, I applied state-of-the-art ancient DNA (aDNA) approaches to recover highly degraded DNA molecules from post-glacial hunter-gatherers and early farmers from the circum-Mediterranean to investigate the demographic processes that underlie the archaeological transitions prior to and during the expansion of agricultural practises. I placed a particular emphasis on investigating signals for direct admixture between peoples that inhabited the southern European and North African Mediterranean shores to examine the genomic evidence for sea crossings, as was proposed by archaeologists based on similarities in material culture.

In Manuscript A, I report the oldest ancient genome-wide data for Africa to date retrieved from nine ~15,000 calBP microlithic Iberomaurusian hunter-gatherers from the Mediterranean coast in Morocco (*Van de Loosdrecht et al. 2018. Science*). The Iberomaurusian genomic profiles contained substantial Eurasian ancestry that was similar to that of ~15,000-9,000 calBP microlithic Natufian foragers in the Near East. This indicates that the genomic connection observed between present-day northern African and Near Eastern peoples long predates the arrival of nomadic pastoralism and sedentary farming practises. In addition to the Natufian-like ancestry, the Iberomaurusian foragers additionally carried substantial complex sub-Saharan African ancestry from a “ghost lineage” related to West Africans that no longer exists in unadmixed form today.

In Manuscript B, I report a genomic transect of ten individuals from the Iberian Peninsula who were retrieved from various sites with Upper Palaeolithic to Middle Neolithic occupations (~13,000-6,000 calBP) (*Villalba-Mouco et al. 2019. Current Biology*). The results indicate that the genomic diversity of (post-)glacial Iberian hunter-gatherers formed a previously unknown genetic cline as a result of asymmetric affinities to Villabruna- and Magdalenian-related ancestry, associated with peoples from two different glacial refugia. Minor traces of this dual ancestry persisted among Iberian early farmers, indicating that local admixture had occurred with Iberian foragers during the expansion of agricultural practises.

In Manuscript C, I explore the multi-disciplinary potential of archaeogenetics, and jointly report genomic and isotopic data for dietary inference for a transect of 19 Mesolithic hunter-gatherers and Early Neolithic farmers from Sicily (~10,700-7,200 calBP) (*Van de Loosdrecht et al. 2020. BioRxiv*). The results provide evidence for major transitions during the Mesolithic and the Early Neolithic. Sicilian Late Mesolithic hunter-gatherers, as well as many contemporaneous European foragers, carried an ancestry component related to hunter-gatherers from Eastern Europe (EHGs) and the Near East. The spread of this component in Europe probably reflected the expansion of blade-and-trapeze industries from eastern/southeastern Europe. The earliest farmers from Sicily carried almost no hunter-gatherer ancestry, and were similar to early farmers from the Balkans. This indicates a full-scale replacement of the foragers by early farmers who had expanded along a northern Mediterranean route. However, the ancestry and dietary profiles of two individuals suggested that prior to the full-scale replacement by incoming farmers the foragers had adopted some aspects of the Neolithic package.

Each of the manuscripts contributes crucial genomic reference points for Mediterranean areas on both the European and African coasts where human occupation locally persisted during the LGM, or with some of the earliest indications for agricultural practises and pottery. Therefore, in the discussion of this dissertation I provide archaeogenetic perspectives on five topics that have long been debated by scholars of Mediterranean prehistory: 1) the origin of the Iberomaurusian industry in Northwest Africa; 2) the admixture dynamics of European hunter-gatherers after the LGM; 3) the genomic evidence for direct population interactions between South European and North African hunter-gatherers across the Mediterranean Sea; 4) the demographic processes underlying the transition from foraging to farming in the central and western Mediterranean; and 5) whether agricultural practises in Sicily, southern Iberia and northwestern Africa were introduced via a separate southern Mediterranean Neolithic expansion route involving the Strait of Sicily or Gibraltar. All in all, this dissertation highlights the importance and power of aDNA approaches in reconstructing past events in Mediterranean prehistory.

Zusammenfassung

Prähistorische Jäger-Sammler und Bauern waren in ihrer Lebensweise stark von der reichen Biodiversität und dem relativ milden Klima des mediterranen Raums abhängig. Die Bewohner dieser Region spielten eine zentrale Rolle in der Wiederbevölkerung Europas und Nordafrikas nach dem letzten glazialen Maximum (LGM) vor ca. 26,500-19,000 Jahren BP (Before Present) sowie in der darauffolgenden Entwicklung und Verbreitung landwirtschaftlicher Praktiken.

In dieser Dissertation habe ich auf Grundlage von Analysen alter DNA auf dem neuesten Stand der Technik, hoch degradierte DNA Moleküle von postglazialen Jäger-Sammlern und frühen Bauern aus dem zirkummediterranen Raum gewonnen, um demographische Prozesse vor und während der Expansion einer bäuerlich-produzierenden Subsistenzwirtschaft zu untersuchen. Ein besonderes Augenmerk lag auf der Untersuchung von Signalen der direkten genetischen Vermischung zwischen Menschen, die die südeuropäische und nordafrikanische Mittelmeerküste bewohnten, um genetische Belege für maritime Überquerungen zu erforschen, wie sie von Archäologen auf Grundlage von Ähnlichkeiten in der materiellen Kultur vorgeschlagen wurden.

In Manuskript A präsentiere ich die bislang ältesten genomweiten Daten aus Afrika von neun ~15,000 calBP datierenden Jäger-Sammlern des mikrolithischen Ibéromaurusien von der Mittelmeerküste Marokkos (*Van de Loosdrecht et al. 2018. Science*). Die genetischen Profile der Jäger-Sammler des Ibéromaurusien enthielten einen ausgeprägten Anteil eurasischer Abstammung, der Ähnlichkeiten mit Wildbeutern des Natufien um ca. 15,000-9,000 calBP im Nahen Osten aufwies. Dieses Ergebnis deutet darauf hin, dass die genetische Verbindung, die zwischen heutigen nordafrikanischen und Völkern aus dem Vorderen Orient besteht, in beiden Gebieten vor den Beginn von pastoraler und bäuerlicher Lebensweise datiert. Zusätzlich zu dieser Natufien-ähnlichen Abstammung trugen die Wildbeuter des Ibéromaurusien eine ausgeprägte, komplexe subsaharische Komponente von einer *ghost lineage*, die mit westafrikanischen Populationen verwandt ist, heute jedoch nicht mehr in unvermischter Form existiert.

In Manuskript B befasse ich mich mit einem genetischen Transekt bestehend aus zehn Individuen von verschiedenen Fundorten auf der iberischen Halbinsel, die vom Junpaläolithikum bis zum Mittelneolithikum (ca. 13,000-6,000 cal BP) datieren (*Villalba-Mouco et al. 2019. Current Biology*). Die Ergebnisse legen nahe, dass die genetische Diversität von (post-)glazialen iberischen Jäger-Sammlern eine bislang unbekannte genetische Kline bildeten, die das Resultat einer asymmetrischen Affinität zu Villabruna- und Magdalénien-ähnlichen Abstammungs-komponenten ist und mit den Populationen aus zwei unterschiedlichen glazialen Refugien

asoziiert werden kann. Geringere Spuren dieser dualen Abstammung bestanden in frühen iberischen Bauern fort und lassen vermuten, dass lokale Vermischungen mit iberischen Wildbeutern während der neolithischen Expansion stattfanden.

In Manuskript C untersuche ich das multidisziplinäre Potential von Archäogenetik und verwende DNA zusammen mit Isotopendaten, die Rückschlüsse auf die Ernährung zulassen, um einen genetischen Transekt bestehend aus 19 mesolithischen Jäger-Sammlern und frühen neolithischen Bauern von Sizilien (ca. 10,700-7,200 cal BP) zu untersuchen (*Van de Loosdrecht et al. 2020. BioRxiv*). Die Ergebnisse belegen den deutlichen Umbruch während des Mesolithikums und Frühneolithikums. Spätmesolithische Jäger-Sammler auf Sizilien trugen eine Abstammungskomponente, die auch in zahlreichen anderen zeitgleichen europäischen Wildbeutern nachgewiesen ist und die mit osteuropäischen Jäger-Sammlern und Jäger-Sammlern des Vorderen Orients verwandt ist. Die Ausbreitung dieser Komponente in Europa könnte die Expansion von Silexindustrien mit Klingen und Trapezen aus Ost- und Südosteuropa widerspiegeln. Die frühesten Bauern auf Sizilien trugen fast keine Jäger-Sammler Abstammung und waren genetisch ähnlich zu frühen Bauern auf dem Balkan. Dies lässt vermuten, dass Wildbeuter praktisch gänzlich von frühen Bauern abgelöst wurden, die sich entlang einer nördlichen mediterranen Route ausbreiteten. Die Abstammungs- und Ernährungsprofile von zwei Individuen lassen jedoch vermuten, dass vor der kompletten Ersetzung durch einwandernde Bauern die Wildbeuter einige Aspekte des neolithischen Pakets übernommen hatten.

Die Manuskripte tragen grundlegende genetische Referenzpunkte für den mediterranen Raum auf europäischer und afrikanischer Seite bei. Es handelt sich um Gegenden, in denen menschliche Besiedlung während des LGM fortbestand und in denen außerdem frühe Nachweise für bäuerliche Lebensweise und die Verwendung von Keramik vorliegen. In der Diskussion befasse ich mich daher aus einer archäogenetischen Perspektive mit fünf Themen, die seit langem Gegenstand archäologischer Debatten sind: 1) den Ursprung der Industrien des Ibéromaurusien in Nordwestafrika; 2) die Vermischungsdynamiken von europäischen Jäger-Sammlern nach dem LGM; 3) die genetischen Belege für direkte Interaktionen zwischen südeuropäischen und nordafrikanischen Jäger-Sammlern über das Mittelmeer hinweg; 4) die demographischen Prozesse, die dem Übergang von Wildbeutern zu Bauern im zentralen und westmediterranen Raum zugrunde lagen; und 5) der Frage, ob agrarische Praktiken in Sizilien, der südlichen iberischen Halbinsel und Nordwestafrika durch eine separate mediterrane Expansionsroute des Neolithikums eingeführt wurden, die die Straße von Sizilien und Gibraltar umfasste. Die Dissertation unterstreicht die Bedeutung und Wirksamkeit von Analysen alter DNA in der Rekonstruktion vergangener Ereignisse in der Vorgeschichte des Mittelmeerraums.

Eigenständigkeitserklärung

Entsprechend §5 Abs. 4 der Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität Jena, erkläre ich, dass mir die geltende Promotionsordnung der Fakultät bekannt ist. Ich bezeuge, dass ich die vorliegende Dissertation selbst angefertigt habe und keine Textabschnitte eines Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel, persönliche Mitteilungen sowie Quellen in meiner vorliegenden Arbeit angegeben habe. Zudem habe ich alle Personen, die mir bei der Auswahl und Auswertung sowie bei der Erstellung der Manuskripte unterstützt haben, in der Auflistung der Manuskripte und den entsprechenden Danksagungen namentlich erwähnt. Zudem versichere ich, dass ich die Hilfe eines Promotionsberaters nicht in Anspruch genommen haben und auch Dritten von mir keine unmittelbaren sowie mittelbaren geldwerte Leistungen für Arbeiten, die im Zusammenhang mit dieser Dissertation stehen, erhalten haben. Die vorliegende Promotion wurde zuvor weder für eine staatliche oder andere wissenschaftliche Prüfung eingereicht, also auch einer anderen Hochschule als Dissertation vorgelegt.

Jena, den 24.09.2020

Marieke Sophia van de Loosdrecht

Acknowledgments

I would like to thank the following people that helped me along this intense doctorate journey.

First and foremost, to my doctorate supervisors Prof. Dr. Johannes Krause and Dr. Wolfgang Haak for giving me the amazing opportunity to work with ancient human DNA in an excellent research institute, guiding me safely through the complex political environment of aDNA research, entrusting me with a large professional freedom to pursue my research ideas and to develop into an independent scientist.

To Dr. Choongwon Jeong for additional supervision and support throughout the Iberomaurusian project, and for teaching me a solid foundation of principles for analytical reasoning in ancient human population genomics and effective academic writing.

To the many contributing archaeologists for their collaboration, open-mindedness to aDNA research and sharing their expertise.

To the technical laboratory staff, Franziska Aron, Dr. Guido Brandt, Dr. Marta Burri, Cäcilia Freund, Rita Radzevičiūtė, Raphaela Stahl and Antje Wissgott for taking a large amount of work out of my hands by doing captures and sequencing.

To the various colleagues who helped me with the writing of this dissertation: Dr. Wolfgang Haak, Prof. Dr. Johannes Krause, Dr. Bettina Bock, Dr. Vanessa Villalba-Mouco, Stefanie Eisenmann; Dr. Didier Binder and Prof. Dr. Detlef Gronenborn for detailed comments; James Fellows Yates for proof-reading; Stefanie Eisenmann and Marcel Keller for translating the summary to German, and Michelle O´Reilly for helping with the figures.

To Aditya Kumar for simply being the best friend that I could have wished for.

To Aida Andrades Valtueña, Karen Giffin, Kerttu Majander, Cody Parker, Rodrigo Barquera, Ayshin Ghalichi, James Fellows Yates, Thiseas Christos Lamnidis, Barbara Pavlek, Elizabeth Nelson, Alexander Immel, Vanessa Villalba-Mouco, Maïté Rivollat, Ainash Childebayeva, Adam Rohrlach, Ke Wang and Nada Salem for your friendship, emotional support, inspiring conversations and making countless happy memories.

To Prof. Dr. Menno Gerkema for continuing to be a dear mentor up till today.

To my teachers and friends in the martial arts community, including Maximilian Beyer, Thomas Budich, Claudia Budich, Michael Richter, Jörg Funk, and Robert Pannicke, for providing much needed stability and safety when chaos hit hard in my private life in the middle of my doctorate study.

To my godparents, Angeline Paredes Santos-Martis and Ivani Paredes Santos, for their spiritual guidance.

To my grandmothers and -fathers. In particular Oma Fieke and her late husband Opa Hans who passed away in the summer of 2018.

To my brother for always making me laugh and reminding me how much I miss being at home.

To my father and mother for their love and care, supporting me in my educational and scientific pursuits, and always encouraging me to follow my heart and dreams.

Appendix

www.sciencemag.org/cgi/content/full/science.aar8380/DC1

Supplementary Materials for

Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations

Marieke van de Loosdrecht, Abdeljalil Bouzouggar,*† Louise Humphrey, Cosimo Posth, Nick Barton, Ayinuer Aximu-Petri, Birgit Nickel, Sarah Nagel, El Hassan Talbi, Mohammed Abdeljalil El Hajraoui, Saaïd Amzazi, Jean-Jacques Hublin, Svante Pääbo, Stephan Schiffels, Matthias Meyer, Wolfgang Haak,† Choongwon Jeong,*† Johannes Krause*†

*Corresponding author. Email: krause@shh.mpg.de (J.K.); jeong@shh.mpg.de (C.J.); bouzouggar@eva.mpg.de (A.B.)
†These authors contributed equally to this work.

Published 15 March 2018 on *Science* First Release
DOI: 10.1126/science.aar8380

This PDF file includes:

Overview Supplementary Text

S1. The LSA Iberomaurusian and its cultural origins

Abdeljalil Bouzouggar, Nick Barton

The Iberomaurusian is a microlithic bladelet technology found across the Maghreb (present-day Morocco, Algeria, Tunisia and parts of Libya, Fig. S1-S2). It forms part of the early Later Stone Age in North Africa (*10*) and marks a cultural break with the flake and blade technologies of the Middle Stone Age in this region (*28*). The term derives from the fusion of two words 'Ibero' (meaning Spanish) and 'Maurusian' (referring to 'Mauretania tingitana', the Roman name for this part of North Africa). It was first adopted by (*19*) to draw attention to similarities between lithic assemblages in Spain and Morocco that contained a profusion of tiny backed blades and acute points ('une profusion de très petites lames à dos retouché et à pointe très aiguë'). This definition was expanded by Camps who described it as a microlithic tradition dominated by backed bladelets, the latter reaching as high as between 40–80% of the retouched tool assemblages (*20*). The use of the microburin technique for producing microlithic backed points was a related feature of this technology.

Alternative names to the Iberomaurusian have sometimes been used for describing LSA backed bladelet assemblages in North Africa. These include the 'Oranian' (*29*) and 'Mouillian' (*30*) for sites in Algeria, while the Eastern Oranian was used in Cyrenaican Libya (*31*). In reality these microlithic industries are hard to distinguish from one another so we would see them generically as Later Stone Age but referring also to the common local name.

The chronology of the Iberomaurusian has been the subject of intensive study over the last ten years. New dating from Grotte des Pigeons, Taforalt, has established a long sequence of over 50 accelerator mass spectrometry (AMS) radiocarbon dates for Iberomaurusian cultural deposits covering a time-span from 23,459 to 12,548 cal. yBP (*32, 33*). Further dates for the Iberomaurusian have been published from Afalou Bou Rhummel (*34*), Ifri n'Ammar (*35*), Ifri el Baroud (*35*) and Kehf el Hammar (*36, 37*), none of which are older than Taforalt (Fig. S1). A slightly earlier age of 25,845-25,270 cal. yBP exists for the oldest Iberomaurusian deposits at Tamar Hat in Algeria (*26*). Critically, the start date for the Eastern Oranian at the Haua Fteah appears to be no earlier than ~17-19,000 yBP (*38*), which would suggest that the Iberomaurusian is older in the west than in Cyrenaica.

Origins with reference to the archaeological evidence

1. Indigenous development

The Iberomaurusian arose independently in North Africa with no presently known cultural antecedents. Its epicenter may have been in Algeria, from where it spread westwards into Morocco and east into Libya and Cyrenaica. The earliest dates for Tamar Hat and slightly younger ages from Grotte des Pigeons, Taforalt and Kehf el Hammar (*36*), and much younger dates from Libya and Cyrenaica are consistent with this scenario. They imply a cultural break around 25,000 cal. yBP.

2. Exogenous development

Theories have been advanced concerning the possible introduction of the Iberomaurusian from Europe across the Gibraltar or Sicily Straits (*39*). Currently there is little supporting evidence for this idea even though the similarities between the Iberomaurusian and the Italian Epigravettian were such that Camps referred to the Iberomaurusian as 'African Epigravettians' (*20*). Greatly reduced sea levels during the Last Glacial Maximum meant that during this period the actual distance between Tunisia and Sicily would have been in the order of nine kilometers (*40*), making the shorelines easily inter-visible. The weakness of this argument is that there are few

Epigravettian sites in either Sicily or southern Italy that can be dated to the LGM (*41, 42*). Similarly there are no known early Iberomaurusian sites in northern Tunisia but this might be due to a lack of fieldwork or that many of the sites lie submerged in now flooded areas of the continental shelf.

An alternative model is that populations with microlithic bladelet technologies followed a land route and spread westwards from the Near East to the Maghreb before 15,000 yBP. This would assume that the Iberomaurusian is related to the Natufian and earlier Epipaleolithic technologies of the Near East. However, the oldest Iberomaurusian microlithic bladelet technologies are mostly earlier than the equivalents in the intervening geographic areas of Egypt and Libya. Nevertheless it is possible to accommodate an intrusion of populations from the east that admixed with local populations and replaced MSA technologies with ones dominated by microlithic bladelets.

3. Continuity with the Middle Stone Age

This model would suggest that the Iberomaurusian emerged from the MSA or 'transitional technocomplexes' in Africa. At Taforalt, the layers underlying the Iberomaurusian are rich in adzes and adze flakes and a simple flake technology (*10*) that overlie typically MSA layers. The dating of the adze layers is still preliminary but suggests an age range of ~26,000 and 24,000 cal. yBP (*43*). Further west in Cyrenaica the transition to the Iberomaurusian/Oranian is marked by a gradual evolution from the Dabban (*31*). The Dabban industry contains typical North African MSA artifacts including small Levallois cores but also 'Upper Palaeolithic' tools such as backed blades and bladelets, chamfered tools, burins and end-scrapers. Continuity with the overlying layers is shown by increasing numbers of backed bladelets and a diminution of some 'UP' types (including chamfered pieces, awls). At Haua Fteah the dates for the Early Dabban are between ~39,000 and 32,000 yBP and the Late Dabban between ~22,000 and 19,000 yBP (*20*). Shared elements with the Dabban (e.g. chamfered blades) occur in assemblages as far east as the northern Levant (*44, 45*) but have so far not been widely observed further west than Cyrenaica. Dabban-like artifacts may occur sporadically in early Iberomaurusian assemblages at Tamar Hat (laterally retouched blades) and at Taforalt (laterally retouched blades and adzes) but other typical tools are absent. It is plausible that population bottlenecking in North Africa during MIS-3 led to intermittent isolation of human groups producing diversification in lithic traditions and the staged appearance of microlithic bladelet technologies from essentially MSA origins.

Another case for continuity exists in West Africa. The MSA persists for much longer than in North Africa (*46*). However, lithic assemblages from several sequences in West Africa show some similarities with the artifacts of the transitional phase at Taforalt with examples of adzes and bifacial pieces at Ounjougou, Mali (*47*), dated to 35,000 ± 1,000 years ago. Recently, sites in the Falémé Valley in Eastern Senegal have yielded bladelets and segments (*48*). For both Mali and Senegal, the sites are still poorly dated and there are no Iberomaurusian sites identified further north until the area of Agadir in southern Morocco.

S2. Taforalt site description and archaeological context of ancient individuals

Louise Humphrey, Abdeljalil Bouzouggar

In this section we provide the archeological context and direct radiocarbon dates for nine Taforalt individuals that were screened for DNA preservation and contamination in this study.

Taforalt

Grotte des Pigeons (also known as Taforalt) is located in the Beni-Snassen Mountains in eastern Morocco close to the border with Algeria at an elevation of 720m above sea level (34°48′38″ N, 2°24′30″ W). The site was first excavated by Armand Ruhlmann between 1944 and 1947, and subsequently by Jean Roche between 1951 and 1955, and 1969 and 1976. Archaeological deposits span the Aterian (Middle Stone Age, MSA) and Iberomaurusian (Later Stone Age, LSA) and the transition between these two periods (Fig. S3). The Roche excavations uncovered an extensive series of Iberomaurusian burials from the uppermost grey ashy deposits in two contiguous areas situated toward the back of the cave, designated Necropolis I and Necropolis II (*49, 50*).

New work on the Iberomaurusian sequence at Taforalt was conducted between 2003 and 2017 (*10, 33, 43, 51*). Excavation trenches were located on the south side of the cave (Sector 8), at the front of the cave (Sector 9) and a recess at the back of the cave in an area of restricted height (Sector 10). Taforalt provides unique information on the layout of Iberomaurusian technology and social organization. It is of great importance because it contains evidence of one of the earliest recognized LSA technologies. This technology yields various and typical artifacts of the Iberomaurusian including microliths, notches and denticulates and rare microburins (Fig. S4). A change in the technology was identified within the Yellow Series, which may be a response to fluctuating climatic and environmental conditions (*10*). The sedimentary transition between the Yellow and Grey Series coincides with a change in subsistence behavior including increased consumption of wild plant foods and land snails, and other behaviors associated with increased sedentism (*51, 52*). The presence of grindstones, used to process edible plants and pigments, provide further evidence for a change in diet at this time. Other non-lithic tools were identified within the archaeological deposits including an important assemblage of the bone tools. Evulsion of one or both upper central incisors (*53*) and the earliest documented example of trepanation are other distinctive characteristics of the Iberomaurusian in Taforalt (*49*).

No human burials or other human remains were found in Sector 8 or Sector 9. Excavations in Sector 10 revealed a series of closely spaced and inter-cutting primary burials within the uppermost grey ashy deposits. The distribution of articulated and disarticulated bones in Sector 10 indicates intensive use and reuse of the area, with earlier burials disturbed or truncated by subsequent burials (*52, 54*). Twelve partially articulated skeletons (7 adults and 5 infants) were recovered, together with another reasonably complete but disturbed infant skeleton, and an intrusive broken adult cranium and associated mandible (Individual 10). With the exception of Individual 10, there is no evidence to suggest that any of the bodies were incomplete or disarticulated at the time of deposition. Most of the individuals were buried in a seated or semi-reclined position, with the pelvis, feet and often the hands at the base of the grave and the head and knees uppermost. The adults and some of the infants were buried facing approximately towards the cave entrance. The partial loss of anatomical articulation of skeletal elements within many of the burials indicates that the bodies were surrounded by pockets of empty space during decomposition. This suggests that the bodies were loosely wrapped or covered by an organic material. The infant and adult burials were associated with a variety of funerary objects including horn cores, ochre stained stones and marine shells. Sector 10 and Necropolis I and II originally formed a contiguous and spatially demarcated collective burial area.

The sequence of burials in Sector 10 could be partly resolved based on their spatial relationships and the distribution of disturbed and undisturbed bones (Fig. S5). The burials comprise two main groups. The first group to be excavated was situated closer to the cave entrance and comprises Individuals 1-5 and Individuals 7 and 8. Individual 6 lay alongside the first group but its chronological relationship to those burials could not be established. The second group of burials

was situated close to the rear cave wall and comprises Individuals 9, 11, 12, 13 and 14. The intrusive skull of Individual 10 lay directly above the burials of individual 14, 13 and 11 but it may have been removed or displaced from an earlier burial.

The Iberomaurusian part of the sedimentary sequence in Sector 8 (Grey and upper Yellow Series) has been dated by radiocarbon accelerator mass spectrometry (AMS) using ultrafiltration, producing dates on bone and wood charcoal that span the period 20,500–12,600 yBP (*10*). Within Sector 8 there is a major stratigraphic division in the Iberomaurusian part of the sequence between the Grey Series (Roche's ashy deposits) and an underlying Yellow Series. The earliest calibrated age for the Grey Series is 15,204–14,261 cal. yBP. Seven human bone samples from Sector 10 have been directly dated by AMS using ultrafiltration (Table S1, (*52*)) These have yielded age estimations between 15,077 cal. yBP and 13,892 cal. yBP corresponding to the lower part the Grey Series deposits in Sector 8. All samples had good collagen preservation (>2% collagen bwt), carbon content (41-47%), and CN ratio (3.1-3.2). The sample from Individual 7 (OxA-16663) was small and the age estimate has a high standard error. The stratigraphic matrix, in combination with overlapping standard errors for the AMS radiocarbon dates for the directly dated individuals, suggests that all the Taforalt individuals reported here may have died within 200 years from one another.

S3. Isolation of ancient DNA and sequencing

Marieke van de Loosdrecht, Matthias Meyer, Cosimo Posth

CT-scanning

Prior to destructive sampling we CT (computed tomography)-scanned all petrous bones at the Max Planck Institute for Evolutionary Anthropology (MPI-EVA) in Leipzig, Germany. We anticipate that the CT-scanning of unique and valuable human specimens will become an integral part of archaeogenetics and bioarchaeology research to minimize the loss of valuable anatomical information. Although for X-ray radiation doses >2000 Gy both the ancient DNA quantity and frequency of C->T nucleotide misincorporations can be significantly reduced (the latter through newly induced strand breaks), no detectable effect was found for dose-levels <200 Gy (*55*). In this study the integrated X-ray dose delivered during the µCT acquisitions ranged from 0.5 to 5.5 Gy, which is hence highly unlikely to impair the ancient DNA preservation of our samples. We calculated the dose values based on the scanning parameters using the dose calculator provided in (*55*).

Bone sampling

We sampled the petrous bones in a dedicated clean room at the Max Planck Institute for the Science of Human History (MPI-SHH), Jena, Germany. Petrous were UV-irradiated for 30 min from multiple sides to reduce surface contamination. The bone samples were very fragile with soft and highly degraded, eroding surfaces. This made standard methods, such as cross-sectioning followed by drilling out bone powder from the areas around the cochlea, impossible. We thus isolated the inner-ear part, where DNA is expected to be better preserved (*56, 57*), by gradually shaving off the softer outer parts and ground the solid inner parts into bone powder with mortar and pestle.

DNA extraction

DNA extraction was done in a dedicated clean room at the MPI-SHH in Jena. We used ~50mg of bone powder to set up an extraction following (*58*) resulting in 100uL DNA extract. For each

bone sample we generated one or two DNA extracts (Table S2). A negative control was included for the whole extraction batch.

Library preparation and target enrichment of the nuclear and mitochondrial DNA
Library preparation was done in a dedicated clean room at the MPI-EVA in Leipzig. Single-stranded next-generation sequencing libraries were created from 30uL extract following an automated version of the protocol reported in (*12*) performed on an Agilent Technologies Bravo NGS Workstation (*59*). DNA libraries were not treated with uracil-DNA-glycosylase (UDG) in order to avoid loss of DNA molecules and library complexity associated with this treatment (*60*). Each library was barcoded with two unique 7 base pair (bp) indices (*61*). For each Taforalt individual we generated two to five indexed libraries (Table S2). In-solution hybridization capture enrichments were implemented on an Agilent Technologies Bravo NGS Workstation following a previously described protocol (*13*). Libraries were hybridized to multiple sets of oligonucleotide probes: one based on the revised Cambridge Reference Sequence (rCRS) ('mtDNA' capture) and the other one for 1,196,358 nuclear SNPs ('1240k' capture, SNP panels 1 and 2 in (*14*)).

DNA sequencing
For each probe set used we made a pool that combined the enriched libraries in equimolar concentrations. The capture pools were then sequenced on different lanes on an Illumina HiSeq4000 platform using a single end configuration (1x75bp reads) at the MPI-SHH in Jena.

S4. Authentication of ancient DNA, genetic sexing and contamination estimates

Marieke van de Loosdrecht, Choongwon Jeong, Cosimo Posth

Read processing
After sequencing, we demultiplexed the sequenced reads according to the expected index pair for each library, allowing for one mismatch per 7 bp index. To process FastQ files we used EAGER v1.92.44 (*62*), a computational pipeline specifically designed for the processing of ancient DNA data. Adapters were clipped with AdapterRemoval v2.2.0 (*63*) and fragments below 30 bp in length were discarded. We aligned reads from the mitochondrial and 1240k captures against the revised Cambridge Reference Sequence (rCRS) and the human reference genome (hg19), respectively, using the Burrows-Wheeler Aligner (BWA, v0.7.12) *aln* and *samse* programs (*64*). We used non-default parameters optimized for ancient DNA fragments (*65*) with '-n 0.01' to allow for more mismatches and '-l 16500' to turn seeding off. For duplicate reads, identified by having identical strand orientation, start and end positions, we included only the highest quality read. We excluded reads with Phred-scaled mapping quality (MAPQ) <30 for both the mitochondrial and nuclear data sets. After data processing, we find that the libraries enriched for nuclear DNA cover 0.3-73.5% of the SNPs in the 1240k panel at least once with an average read depth per SNP of 0.00-3.29X (Table S3). For libraries enriched for the mitochondrial genome, we find that reads cover 95-100% of the mitochondrial genome with an average coverage of 16-1,701X at each base position (Table S4). For each individual we merged the sequences from the different libraries with *samtools merge* (*64*). For an individual to be included in further population genetic analysis, we required the merged per-individual data to have reads covering > 200,000 SNPs on the target 1240k SNPs panel (Table S3).

Quality control and selection of libraries for analysis
We assessed the authenticity and level of contamination in our ancient DNA libraries (unmerged and merged per-individual) in three different ways. First, we checked if the DNA libraries show

the characteristic damage patterns of ancient DNA using DamageProfiler v0.3 (https://github.com/apeltzer/DamageProfiler) (Table S3-S4). Authentic ancient DNA molecules harbor a high rate of cytosine deamination at the terminal ends (*66*). Deaminated cytosines are misread as thymines, resulting in an apparent excess of thymines at read ends of ancient DNA fragments compared to the modern reference genome (*67*). We expect a high frequency of deaminated bases for the Taforalt individuals, considering their Upper Paleolithic age and their preservation in an environment in which DNA decays fast (*68, 69*). Indeed, we find above 50% C>T mismatch rate for our non-UDG libraries in the terminal nucleotide, gradually decreasing to a baseline rate of 2-3% up to around the 10^{th} base from each end (Fig. S6). We excluded all libraries from individuals TAF008 and TAF016 based on the much lower C>T mismatch rates (≤14.7%; Table S3) compared to the rest (≥47.8%; Table S3).

Second, we tested for mtDNA contamination using the program Schmutzi v1.0 (*70*), an iterative likelihood-based approach that jointly reconstructs the consensus mtDNA sequence for each mtDNA captured library and assesses the degree of present-day human contamination by comparison to a database containing 197 worldwide present-day human mtDNA sequences. In preparation for running Schmutzi, reads were realigned to the rCRS using CircularMapper v1.93.4 that takes into consideration the circularity of the mitochondrial genome. We removed reads with MAPQ <30 and duplicates and downsampled to 30,000 reads per library, which corresponds to an expected 80X base coverage, for ease of computation. We obtained low contamination levels and could reconstruct mitochondrial genomes for seven individuals (100% genome coverage, average base coverage per library 102-1,701X, Table S4). The consensus sequences for these individuals were imported into HaploGrep2 v2.1.1 (available via https://haplogrep.uibk.ac.at/) to obtain an automated mitochondrial haplogroup assignment (Table S4).

Third, we tested for contamination of the nuclear genome in males (all but one individual in Taforalt) based on the X-chromosomal polymorphism rate. By comparing the coverage on the X-chromosome relative to the autosomes (X-rate) and the Y-chromosome relative to the autosomes (Y-rate), we determined the genetic sex (*27*) (Table S5). Eight individuals for which the libraries showed a Y-rate >0.26 we assigned the label 'male' and one individual with library Y-rates <0.03 as 'female'. Then we tested for heterozygosity of the X-chromosome using ANGSD v0.910 (*71*). Since males have only one X-chromosome, once accounted for sequencing errors, the polymorphism rate provides a reliable and conservative contamination estimate given sufficient markers (>200 X-chromosomal SNPs, covered at least twice (*27*)). Our results suggest a nuclear contamination of <3% for four male individuals that we included in further analyses, and >10% for one male (TAF009) that we excluded from analyses (Table S5, based on new Method1 (*72*)).

Using the quality controls above we could identify four male and one female individuals for whom the unmerged libraries show no evidence of contamination. From these individuals we extracted reads with high mapping quality (MAPQ ≥37) to the autosomes using samtools v1.3 for genotyping. Since our damage plots indicate that C>T misincorporations can extend up to 10 bp from the read termini, we opted for a conservative masking when calling genotypes for C>T substitutions. We thus clipped 10 bp from the 5'- and 3'-end of the reads before randomly drawing a read to determine the haploid genotype for an allele at a C>T substitution site with our custom program 'pileupCaller' (https://github.com/stschiff/sequenceTools). Since for single-stranded libraries we do not expect an elevated G>A misincorporation rate, we used the unclipped reads to call genotypes for these and other substitutions. Then we merged our final genotype set for five Taforalt individuals with data presented in (*16*), that contains the Human Origins (HO) genotype set for worldwide present-day populations (*73*) and reference genotypes for ancient individuals including data from (*8, 15, 17*), and the data from (*4*) that includes sixteen ancient

sub-Saharan African individuals. We restricted our further population genetic analyses to 593,124 autosomal SNPs that overlap between these different genotype sets and adopt the population labels for ancient individuals from (*4, 16*).

S5. An overview of the genetic affinity of Taforalt with worldwide populations based on PCA and ADMIXTURE

Marieke van de Loosdrecht*, Choongwon Jeong*

To understand the genetic affinity of our Taforalt individuals with worldwide ancient and present-day populations, we first performed principal component (PC) and ADMIXTURE analyses. For PCA we constructed the PCs from present-day population sets from the HO panel to investigate the genetic diversity at three different geographical resolutions: global populations, populations restricted to Africa, the Near East and southern Europe, and West Eurasia. Then we projected the Taforalt individuals, and other relevant ancient individuals from the Near East (*16*) and sub-Saharan Africa (*4*) onto the top PCs. We used the program *smartpca* from the Eigensoft package v6.0.1 with option 'lsqproject: YES', and turned the outlier removal off.

The results for the PCA analyses that are based on African populations to construct the PCs show that the Taforalt individuals do not cluster with any present-day population. Instead they take an intermediate position between present-day North Africans and sub-Saharan Africans. Specifically, they are flanked by North African Mozabite and Saharawi, and by East African Afars on the Near Eastern and Sub-Saharan African directions, respectively (Fig. S7-S8). The East African groups, Afar, Oromo and Somali as well as a 3,000-year-old individual from Tanzania ('Tanzania_Luxmanda' in (*4*)), are the closest to Taforalt from the sub-Saharan African side. For these East African groups, which together with Taforalt take intermediate positions between sub-Saharan Africans and West Eurasians, genetic links with the Near East have been proposed. These links include an introduction of early Neolithic Levant-like ancestry (*16, 74*), associated with the spread of pastoralism, into East Africa ~4,000 yBP and South Africa ~2,000 yBP. A more recent link is the spread of Bronze Age Levant-like ancestry in sub-Saharan Africa together with agriculture during the Bantu expansion (*4*). The Bronze Age Levant gene pool has been proposed to be a mixture of the Neolithic Levant with individuals related to Caucasus hunter-gatherers and early Neolithic farmers from Iran, and from Anatolia (*16*).

When only the genetic variation of West Eurasians is taken into account to construct PCs, we find the Taforalt individuals to fall outside the genetic variation of present-day West Eurasians (Fig. S9). Consistent with the global and African PCAs, the Taforalt individuals form a tight cluster between Bedouins and East African nomadic groups. Interestingly, in this PCA that distinguishes West Eurasians well, Taforalt appears to be on the extreme end of a cline with early Neolithic Levantines and Natufians. This suggests a close connection with early Holocene Near Eastern groups among West Eurasians.

To obtain an alternative summary of the Taforalt gene pool, we also conducted an unsupervised genetic clustering of global populations using ADMIXTURE v1.3.0 (*75*). We removed duplicate individuals and individuals with a genotype missing rate >95% using PLINK v1.9 (*76*). We also performed a linkage disequilibrium (LD)-based SNP pruning using the '--indep-pairwise 200 20 0.2' option in PLINK, leaving 3,080 individuals and 201,361 SNPs to be analyzed. We reduced diploid to haploid calls by randomly sampling one allele to reduce the impact of artificial genetic drift in haploid ancient individuals on our clustering results. We ran five replicates with different random seeds for each K values ranging from 2 to 15 and took the replicate with the highest log

likelihood value. The most optimal levels of cluster breakdown, as indicated by lowest CV error, were K=9 to K=11 (Fig. S10). The two major components that comprise the Taforalt genomes (52.5% green and 31.3% purple; K=9; Fig. S11) are maximized in early Holocene Levantines (93.5% in Levant_N and 88.2% in Natufians) and West Africans (≥99.9% in Yoruba and Mende), respectively. Interestingly, when a component for East African hunter-gatherer Hadza (brown) is singled out at K=10, the model for Taforalt includes a substantial proportion of the Hadza-related component (19.3% West African and 25.7% Hadza-related; Fig. S11). In comparison, present-day North Africans have a much smaller sub-Saharan African component with no apparent link to Hadza, comprising 24.8% and 22.0% in Mozabite and Saharawi, respectively (Fig. S11).

Based on our results, we hypothesize that the ancient Taforalt individuals have a strong genetic affinity both with early Holocene Levantine groups and with sub-Saharan Africans. Also, the sub-Saharan African ancestry in the Taforalt individuals may have links to multiple sub-Saharan African lineages.

S6. Genetic affinity of Taforalt with early Holocene Levantines and present-day sub-Saharan Africans

Marieke van de Loosdrecht*, Choongwon Jeong*

To test our hypotheses on the genetic affinity of the Taforalt individuals with Near Eastern and sub-Saharan African ancestries, we first measured the amount of shared genetic drift with worldwide present-day and ancient populations using outgroup-f_3 statistics. Considering the potential presence of sub-Saharan African ancestry in Taforalt, we calculated the statistic for a few different outgroups; each outgroup provides a different weight for the sharing of sub-Saharan African or non-African ancestries, depending on their relationship with these ancestries. In short, an ancestry close to the outgroup is down-weighted, while one distant from it is up-weighted (Fig. S12).

First, we calculated f_3(Mbuti; Taforalt, X) using Mbuti (n=10) as an outgroup to find the Eurasian populations most closely related to Taforalt. Consistent with our PCA and ADMIXTURE results (Fig. 2), we find that ancient Levantine populations, especially Epipaleolithic Natufians and early Neolithic individuals from Pre-Pottery Neolithic B and C ('Levant_N'), show by far the highest outgroup-f_3 values with Taforalt (Fig. S13). Then, we formally tested if Natufian/Levant_N is more closely related to Taforalt than any other ancient Near Eastern populations, by calculating f_4-symmetry statistics in the form of f_4(Chimpanzee, Taforalt; NE_1, NE_2). If NE_2 shares more alleles with Taforalt than NE_1 does, i.e. if NE_2 is closer to Taforalt than NE_1 is, f_4 statistics will be significantly positive. Standard errors (SE) were estimated using 5 cM block jackknife method, as implemented in the qpDstat program (v711) in the admixtools v3.0 package (*21*). We find significant positive f_4 values when NE_2 are Natufian/Levant_N and NE_1 are others ($Z \geq 2.2$ SE; Table S6). Therefore, we conclude that the early Holocene Levantine populations, who postdate our Taforalt individuals by up to ~6000 years (*16*), are most closely related to Taforalt among Eurasian populations in our data set.

We also calculated f_3(Ust'-Ishim; Taforalt, X) to explore the affinity of Taforalt with African populations. Ust'-Ishim is an Upper Paleolithic Siberian individual, dated to 45,000 years before present (yBP), whose ancestry is symmetrically related to prehistoric European hunter-gatherers and present-day East Asians (*77*). In this case, North African populations, such as Saharawi and Mozabite, show the highest frequencies of allele sharing with Taforalt, suggesting similarity in their genetic profiles (Fig. S14). This matches well with their close positions in the PCA plot

(Fig. 2A). Following this, West African populations show high outgroup-f_3 values (Fig. S14). We found strong evidence visualizing the sub-Saharan African affinity of Taforalt by comparing this outgroup-f_3 with that of Natufian: f_3(Ust'-Ishim; Natufian, X) (Fig. S15). While Eurasian populations tightly fall on a line, all African populations clearly deviate from this line. This suggests that sub-Saharan Africans, most notably West Africans, share ancestry with Taforalt beyond what can be explained by their Natufian-like ancestry.

Next, we formally tested if the Taforalt individuals have sub-Saharan African ancestry by calculating f_4(Chimpanzee, X; Natufian, Taforalt). As expected, we observed significant positive f_4 values for all sub-Saharan Africans and significant negative values for all Eurasian populations (Fig. S16). A reduced level of Neanderthal ancestry cannot be the sole explanation for this, because we find that f_4(Chimpanzee, Altai Neanderthal; Natufian, Taforalt) is non-significant and positive (Z = -1.089 SE). Our results clearly support a dual ancestry of our Taforalt individuals, genetically related to both early Holocene Near Easterners and present-day sub-Saharan Africans.

Finally, we tried to detect additional signatures of admixture between the Eurasian and sub-Saharan African gene pools, using f_3 statistics with Taforalt as the target and a linkage disequilibrium (LD) decay-based method implemented in the ALDER v1.3 program (*78*). We could not find any Eurasian and sub-Saharan African population pairs with negative f_3(Taforalt; Eurasian, sub-Saharan African), suggesting a strong post-admixture genetic drift in our Taforalt individuals (Z > 44.320 SE). Neither could we detect a decay of admixture LD in our Taforalt individuals, suggesting that the admixture may not be a recent event (Fig. S17).

S7. Testing for Neanderthal admixture in Taforalt

Marieke van de Loosdrecht*, Choongwon Jeong*

An interesting feature of the Near Eastern gene pools is the genetic legacy of the so called 'Basal Eurasians', a hypothetical population that is basal to both Mesolithic European hunter-gatherers and East Eurasians (*16, 73*). This Basal Eurasian gene pool is also assumed to have no Neanderthal ancestry; i.e. it is an early branch of non-Africans that split off from the rest prior to the admixture with Neanderthals. If such characteristics are assumed, their contribution to Near Eastern, and accordingly to early Neolithic European, farmers can provide an explanation for interesting observations. These include the apparent closer relationship of East Eurasians to Ice Age Europeans than to Neolithic European farmers, and consequently a lower amount of Neanderthal ancestry in present-day West Eurasians (as a result of expansions of early farmers into Europe) than East Eurasians (*65, 79*). So far, the early Holocene populations from the Near East are known to have the highest proportion of Basal Eurasian ancestry, up to around 50%; e.g. the Mesolithic hunter-gatherers from the Caucasus (CHG, (*15*) and Iran (HotuIIIb, (*16*)), (pre-)/early Neolithic Levant (Natufian, Levant_N) and early Neolithic Iran (Iran_N, (*16*)). Expansion of these derived Basal Eurasian populations, and subsequent admixture with local populations, is considered to have substantially decreased the Neanderthal ancestry proportion in Near Eastern populations that changed from foraging to early food production (*16*). It is unknown where and when Basal Eurasians emerged and how the Holocene Near Easterners obtained this ancestry in large quantity. North Africa is a strong candidate for the place having kept Basal Eurasians because it is well connected to Eurasia. From this point of view, the Upper Paleolithic individuals from Taforalt are likely candidates as being direct descendants of, or closely derived from, the Basal Eurasian population.

For the Taforalt individuals to be considered as being Basal Eurasians, we expect that their genomes do not share significantly more alleles with the Neanderthal genome than that sub-Saharan Africans do. We therefore compared the amount of Neanderthal allele sharing between sub-Saharan African Yoruba and others by calculating f_4(Chimpanzee, Altai Neanderthal; Yoruba, *Test*). For Taforalt we obtain slightly positive f_4 values (f_4 = 0.000737, Z = 3.116 SE; Table S7), indicating a low amount of Neanderthal admixture. Typically, the allele sharing with Altai is lower in Taforalt than in early Holocene Near Eastern populations (e.g. Natufians: f_4 = 0.001132, Z = 2.922 and Levant_N: f_4 = 0.000880, Z = 3.082), and present-day North Africans and Near Easterners. When using other sub-Saharan African populations as a baseline (e.g. Mbuti, Hadza, aSouthAfrica, Ju|'hoan, Khoisan), we obtain qualitatively similar but statistically weaker results. This is likely due to a smaller sample size for these groups (n≤10) compared to Yoruba (n=70).

Considering the dual ancestry of the Taforalt individuals, we can explain the Altai affinity in Taforalt as a dilution of its Natufian-related ancestry with its significant proportion (~36.5%) of sub-Saharan African ancestry. Interestingly, the Neanderthal ancestry in Taforalt is higher than in early Neolithic Iran (Iran_N, f_4 = 0.000628, Z = 1.934). We can therefore deduce that the Taforalt individuals are not genetically closer to the hypothetical Basal Eurasian population than the early Holocene populations from Iran.

S8. Characterization of the Eurasian ancestry in Taforalt

Choongwon Jeong*, Marieke van de Loosdrecht*

In previous sections we showed that early Holocene Levantine populations, such as Epipaleolithic Natufians, are most closely related to the Eurasian ancestry in Taforalt (Fig. S13, Table S6). However, simple *F*-statistics are insufficient to show if the Natufian-like ancestry is enough to explain the Eurasian ancestry in Taforalt or if additional ancestries are required. This question is important in the context of archaeological studies on the Iberomaurusian culture, to which our Taforalt individuals are assigned. Some studies have hypothesized a prehistoric migration from Ice Age Europe, either via the strait of Gibraltar (*19*) or Sicily (*20, 39*), during the Last Glacial Maximum based on similarities in lithic technology. If the Taforalt individuals derive a substantial proportion of their ancestry from this hypothesized gene flow from Paleolithic Europeans, we expect that both Natufian- and Paleolithic European-related ancestries are necessary to model the Taforalt genomes.

To test this, we used admixture modeling of Taforalt to quantify and characterize their Eurasian ancestry (*16*). One complication is the presence of sub-Saharan African ancestry in Taforalt, for which the relationship with much later Holocene ancient and present-day African groups is not understood. To simplify our task, we performed admixture modeling without explicitly characterizing the relationship of the sub-Saharan ancestry in Taforalt with various sub-Saharan African ancestries in present-day populations.

For admixture modeling, we used the program qpAdm (v632) (*16*) of the admixtools v3.0 package. QpAdm can be viewed as a generalization of f_4 statistics jointly modeling multiple of them. It tests if the observed target population and the proposed admixture model for it are symmetrically related to a set of outgroups, and summarizes the results of multiple such comparisons into a single statistic (*16*). It also estimates ancestry proportion coefficients, and their 5 cM block jackknife SEs, by minimizing the difference between the target and the model. More specifically, qpAdm requires a target population (T), source/surrogate populations (S) and a

509 set of outgroups (O). Outgroups are differentially related to sources so that they can be
510 distinguished by f_4 statistics (Fig. S18). However, at the same time, outgroups must be related to
511 the target and the sources distantly enough so that a source and its related ancestry in the target
512 have a symmetrical genetic distance to all outgroups. An example of many scenarios to break this
513 prerequisite is a post-mixture gene flow from the target into an outgroup.
514
515 We made a heuristic choice of outgroups: Onge (n=11), Han (n=43), Papuan (n=18), Ust'-Ishim
516 (n=1), Kostenki14 (n=1) (*27*), MA-1 (n=1) (*17*), and Neolithic individuals from Iran ('Iran_N';
517 n=5) (*16*). We did not include any African group nor any archaic hominin in our outgroup set,
518 specifically to keep the program blind to the difference between various sub-Saharan African
519 ancestries. In doing so we could get a reliable estimate of sub-Saharan African ancestry
520 proportion in Taforalt. Iran_N was included among the outgroups because it shares the 'basal
521 Eurasian' ancestry with Natufians and therefore differentiates it from its sub-Saharan African
522 ancestry. Without Iran_N, the model becomes blind to the difference between sub-Saharan
523 African and basal Eurasian ancestries because both of them are basal to Eurasian ancestries to
524 which the rest of outgroups belong.
525
526 With the above choice of outgroups, we modeled Taforalt as a two-way mixture of Natufian and
527 one of nine sub-Saharan African populations: 2,000 yBP ancient South African hunter-gatherers
528 ('aSouthAfrica'; n=2) (*4*), Ju_hoan_North (n=22), Mende (n=8), Yoruba (n=70), Mbuti (n=10),
529 Biaka (n=20), Dinka (n=7), a 4,500 yBP ancient Ethiopian genome ('Mota'; n=1) (*8*) and Hadza
530 (n=5). Although these nine African groups are hugely divergent from each other and from
531 Natufians, we obtained nearly identical results for all nine two-way mixture models (Table S8).
532 All models show that a two-way mixture is a good fit to our data (χ^2 $p \geq 0.128$), with 36.5% sub-
533 Saharan African ancestry on average (35.2% to 38.2% with SE 6.9% to 8.0%; Table S8). Since
534 all the sub-Saharan African populations fit the data, this provides empirical support for our logic
535 to include only non-African outgroups in our qpAdm models.
536
537 We further explored if an additional gene flow from pre-Neolithic Europeans provides a better
538 model to fit the Taforalt gene pool than the previously explored two-way models, by adding
539 either Paleolithic 'Vestonice' (n=6), 'El Miron' (n=6) or 'Villabruna' (n=12) clusters (*27*) or
540 Mesolithic European hunter-gatherers ('WHG'; n=3) as the third source population. As expected,
541 the three-way admixture models, nesting simpler two-way ones, fit well to our data (χ^2 $p \geq 0.119$)
542 with similar estimates for the sub-Saharan African ancestry proportions (36.3% to 44.6% with SE
543 6.3% to 10.1%; Table S9). When contrasted with the simpler nested two-way models, we observe
544 that the three-way models perform only marginally, or not significantly better, than the two-way
545 ones (χ^2 $p = 0.019$ to 0.128; Table S9). We consider this marginal increase in the model fit,
546 especially with Villabruna cluster or with WHG, as not reliable enough to support the presence
547 for an additional contribution from pre-Neolithic Europeans for two reasons: i) the simpler two-
548 way models already provide a good fit and ii) the standard error estimates for both Natufian- and
549 European-related ancestries become large (10.8-17.3% and 6.8-10.4% for Natufian and European,
550 respectively; Table S9). We speculate that the low amount of genetic data for Natufians, together
551 with the absence of closely related outgroups, may cause a marginally better fit in some three-
552 way models.
553
554 We also tested the same question based on more explicit modeling of population relationship
555 using the qpGraph program (v6065) of the admixtools v3.0 package (*21*). We obtained a graph
556 fitting Mbuti, Yoruba, Ami, WHG, Mesolithic Eastern European hunter-gatherers from Samara
557 and Karelia ('EHG'; n=3) (*80, 81*), Iran_N and Natufian with no marked outlier by iteratively
558 adding one population at a time into the graph and choosing the best-fitting position on the graph
559 among all possible ones (max $|f_{4,\,expected} - f_{4,\,observed}|$ = 2.6 SE; Fig. S19A). Among all possible no-

admixture and two-way admixture models on this scaffold graph, the best model for Taforalt is a mixture of Natufian and Yoruba with a 30% contribution from a Yoruba-related branch, consistent with the qpAdm results (Fig. S19B). The best graph still deviates from the observed data in some *F*-statistics, but the outlier statistics do not suggest an unexplained extra affinity between Taforalt and WHG (Table S10). Likewise, adding additional gene flow from a WHG-related branch into Taforalt results in no increase in the model fit with 0% contribution estimate (Fig. S19C-E).

Based on both the qpAdm and qpGraph results, we conclude that gene flow from pre-Neolithic Europeans is not necessary to explain the Taforalt gene pool in our data's resolution. This is also consistent with our clustering results, which assign a distinct ancestry component to Natufian/Levant_N (dark green in Fig. S11) and to pre-Neolithic Europeans (salmon pink). Higher quality genome data from early Holocene Near Easterners will be critical to test the presence of a minor contribution from the pre-Neolithic Europeans beyond the resolution of our current data in the future.

S9. Characterization of sub-Saharan African ancestry in Taforalt

Choongwon Jeong, Marieke van de Loosdrecht

Previous studies of African population structure revealed several distinct ancestries across the continent (*82*). As the birthplace of *Homo sapiens*, sub-Saharan Africa is home to the most deeply diverged lineages of our species, including South African click-speaking populations ('Khoe-San'), East African click-speaking populations ('Hadza'), Central African rainforest hunter-gatherers ('Mbuti'), Niger-Congo speaking West Africans, and Nilo-Saharan speaking East Africans as extant representatives. Currently, it is mostly accepted that Khoe-San groups represent the deepest branch in modern humans, and Pygmies occupy intermediate positions between Khoe-San and other groups on the population tree (*82, 83*). Recently, Skoglund et al. suggested that Mota, a 4,500 yBP Ethiopian, and Hadza represent an ancient East African lineage, which is most closely related to non-Africans among all major African lineages (*4*). This finding also suggests that this East African lineage cannot be modeled as a sister group of West Africans, because it shares extra genetic affinity with Khoe-San (*4*). The exact relationship between major sub-Saharan African lineages is still poorly understood.

It is surprising that we observe a high proportion (36.5%) of sub-Saharan African ancestry in Taforalt. First, present-day North Africans do not have as high sub-Saharan African ancestry as the Taforalt individuals (Fig. 2B+S12). This may be attributed to more recent events, such as the historical Arab expansion. Also, the periodic expansion of the Saharan desert played a major role in limiting gene flow between North and sub-Saharan Africa throughout time. For example, a previous study of ancient Egyptian genomes shows that the genetic affinity with the Near East was even stronger in the first millennium BCE in Egypt (*5*). Importantly, our Taforalt individuals predate the most recent greening of the Sahara by several millennia (*84*). Thus, we may speculate that the sub-Saharan African ancestry in Taforalt derived from the gene pool of pre-LGM North Africans, who belong to the Middle Stone Age (MSA) cultures (*10*).

To characterize the sub-Saharan African ancestry in Taforalt, we first explored *F*-statistics between various African and ancient Near Eastern groups. As demonstrated above, both f_3(Ust'-Ishim; Taforalt, X) and f_4(Chimpanzee, X; Natufian, Taforalt) point out West Africans, such as Mende or Yoruba, to most strongly pull out the sub-Saharan African ancestry in Taforalt (Fig.

S14+S16). Therefore, it is likely that the Taforalt gene pool harbors a substantial component of West African-related ancestry.

However, such a simple two-way admixture model, with Natufian and Yoruba as sources, is insufficient to explain the Taforalt gene pool. F_4 statistics suggest that Yoruba and Natufian are symmetrically related to two deeply divergent outgroups, the 2,000 yBP ancient South African ('aSouthAfrica') and Mbuti (Z =0.495 and -0.248 SE, respectively; Table S11). Given that f_4 statistic is linear, we expect Taforalt not to be any closer to these outgroups than Yoruba or Natufian if the two-way admixture model is correct. Instead, we find that Taforalt is significantly closer to both outgroups than any combination of Yoruba and Natufian ($Z \geq 2.728$ SE; Fig. S20). A similar pattern is observed for the East African outgroups Dinka, Mota and Hadza; although they are closer to Natufian than Yoruba ($Z \geq 5.898$ SE; Table S11), they are more closely related to Taforalt than to either Natufian or Yoruba (and hence to any combination of them; Fig. S20). These results strongly suggest that Taforalt harbors an ancestry containing extra affinity with South, East and Central African outgroups.

We also find that adding one of these outgroups as the third source still does not provide a good fit in terms of the relationship with the rest of outgroups. First, f_4(Chimpanzee, X; model, Taforalt) for the three-way admixture models are still strongly positive across a range of ancestry proportion values around our estimates of 36.5% sub-Saharan African ancestry in Taforalt (Fig. S21). Second, qpAdm-based three-way admixture models fail to fit the data even when the minimum resolution for the sub-Saharan African ancestry was provided by adding the archaic hominin Denisovan as a sole non-Eurasian outgroup (Table S12). Therefore, we conclude that none of the South, Central and East African groups is a sister group of the sub-Saharan African ancestry in Taforalt. Given that our Taforalt individuals outdate even the most ancient Holocene African individual, the 4,500 yBP Mota, by over 10,000 years, this result is not surprising; a long-term gene flow between the various sub-Saharan African groups during the Holocene period is very likely to have generated a pattern that is not easily modeled as a rather simple admixture graph.

Finally, we exclude any ancestry more basal than the deepest known modern human ancestry represented by 'aSouthAfrica' and present-day Khoe-San speakers in South Africa as an additional source for the Taforalt gene pool. Such a basal ancestry will generate a negative affinity with the South African ancestry, not the positive one we observe in our f_4 statistics (Fig. 4). Because such a deep ancestry is an outgroup to both aSouthAfrica and any mixture of Yoruba and Natufian, f_4(Chimpanzee, aSouthAfrica, Natufian+Yoruba, deep ancestry) will have a negative value. Examples of such a deep ancestry includes any unknown archaic hominin and a hypothetical deep West African branch suggested by Skoglund et al. to explain the asymmetric relationship of Eurasian populations to Yoruba and Mende (*4*). Therefore, these hypothetical ancestries, even if they existed, cannot be used to fully characterize the Taforalt gene pool.

S10. Mitochondrial genome analyses in BEAST

Cosimo Posth, Wolfgang Haak, Marieke van de Loosdrecht

In a previous study on the mitochondrial Hypervariable Segment 1 for individuals associated with the Iberomaurusian culture from Afalou Bou Rhummel (Algeria) and Taforalt, the haplogroups U (U6), H, and R0 were reported (*85*).

Here we present complete mitochondrial genomes for seven Taforalt individuals. After an initial haplogroup assignment using HaploGrep2 v2.1.1 (available via https://haplogrep.uibk.ac.at/) (SOM S4)), we selected for each Taforalt individual the library for which the mitochondrial read assembly scored the highest on 'Overall rank'. We imported the bam.files into Geneious v.9.0.5 (http://www.geneious.com) (*86*), reassembled the reads against the Reconstructed Sapiens Reference Sequence (RSRS) (*87*) and called SNP variants following the procedure described in (supplementary material in (*23*)). We visually double-checked the assembly and compared the called SNP variants for each mitochondrial sequence to the diagnostic variants that are expected based on the initial haplogroup assignment, using the online mtDNA phylogeny tree (mtDNA tree Build 17, 18 Feb 2016, available via http://www.phylotree.org/). We did not consider known unstable nucleotide positions 309.1C(C), 315.1C, AC indels at 515-522, 16182C, 16183C, 16193.1C(C) and 16519. This allowed us to identify very few phylogenetically missing variants (likely due to back mutations at mutational hot spots such as nucleotide pair 195), as well as unambiguous variants private to the respective sequence haplotype. Using this approach, we defined sub-haplogroups U6a1b (n=1), U6a6b (n=1), U6a7(a) (n=2), U6a7b (n=2) and M1b (n=1). Of note, TAF011 and TAF012 do not carry five of the expected eight variants that define U6a7a, which means that this lineage falls basal to the currently reported lineages within branch U6a7a. The complete mitogenome consensus sequences (FastA) are available in Genbank under accession numbers MG936619-25.

Interestingly, the U6a and M1b mitochondrial haplogroups found in our Taforalt individuals (Table S13) are almost uniquely distributed within northern and eastern Africa (*7*). Haplogroups U6 and M1 have been proposed as signature haplogroups that represent an autochthonous Maghrebi ancestry signal and have putatively been linked to a back-to-Africa migration from West Asia (*7*). However, the timing and geographical direction for the spread of these haplotypes, and their association with a back-to-Africa migration, is poorly understood. Given the occurrence of both U6a and M1b haplogroups in the Taforalt individuals, here we can directly demonstrate a pre-Holocene presence of these autochthonous North African lineages in the Maghreb.

Interestingly, basal haplogroup U6 has been reported for ~35,000 yBP specimens found at Muierii cave in Romania (*22, 23*). We are therefore interested to know how the mtDNA genomes in our 15,000 cal. yBP North African individuals relate phylogenetically to the U6 and M mtDNA sequences found in Ice Age Europeans (*22, 23, 27*) and present-day humans (*7*). We prepared a multiple genome alignment with 93 sequences using MUSCLE (*88*) and reconstructed a Maximum Parsimony tree in MEGA6 (*89*) with 98% partial deletion and 1,000 iterations as bootstrap support (Fig. S22). As expected the ancient sequences show a shorter phylogenetic branch length. Six Taforalt individuals fall on the U6a branch and one on the M1b branch.

We also aimed to assess divergence times for specific haplogroups and infer changes in effective population size for populations carrying a specific haplogroup. Using MUSCLE (*88*) we aligned the seven visually corrected mtDNA sequences to 81 present-day mtDNA genomes (*7*), four Upper Paleolithic European mtDNA genomes (*22, 23*) belonging to haplogroups M and U6, and one African L3 mtDNA genome as an outgroup (Table S13). From the resulting multiple genome alignment of 93 mtDNA sequences we excluded all positions containing gaps or unassigned nucleotides using MEGA6 (*89*). This resulted in a total of 16,514 retained positions for each sequence in our data set. This dataset was run through ModelGenarator v.85 and we found that the best-supported substitution model is Tamura-Nei 93 with a fixed fraction of invariable sites but no gamma distributed rates. The complete deletion alignment was imported in *beauti,* a program part of the BEAST v1.8.1 package (*24*), where we selected the indicated model. We left the tip dates for present-day human mtDNA sequences as default zero, whereas the direct radiocarbon dates of the Upper Paleolithic haplogroup U6 (one Muierii and six Taforalt) and

haplogroup M individuals (two Goyet, one LaRochette and one Taforalt) served as time anchors (tip dates) for the ancient sequences (Table S13). For the eleven ancient individuals the 95,4% confidence interval of the radiocarbon date was selected as uniform prior distribution. A strict clock and an uncorrelated lognormal-distributed relaxed clock were tested to model rate variation among tree branches using a fixed mutation rate of $2.74*10^{-8}$ mutation/site/year, previously calculated with 66 dated ancient mtDNA sequences as tip calibration points (*23*). Furthermore, we investigated both the population constant size and the Bayesian skyline coalescent tree priors, the latter with group number set to 10 and a piecewise linear model. For each of the resulting four model combinations we performed a MCMC run of 50,000,000 iterations with individual sampling every 10,000 iterations and 10% burn-in. A marginal likelihood estimation (MLE) was implemented to assess the best-supported model using path sampling (PS) and stepping-stone sampling (*90*). The Bayesian skyline tree prior with a strict clock provided higher likelihoods than the other three models (Table S14). The resulting divergence times for the major haplogroups are reported in Table S15.

Our results provide the time to the most recent common ancestor (TMRCA) for macro-haplogroup M of ~50,000 yBP, similar to previous estimates with a much larger sample size (*23*). Here we estimate a divergence time for haplogroup U6 to 37,000 yBP (95% highest posterior density [HPD], 40-34,000 yBP), contrary to an older estimate of 45,000±7,000 yBP (*7*) but similar to more recent estimates of ~35,000 yBP (*91, 92*). Excluding the Upper Paleolithic European U6 sequence from Muierii cave results in a coalescence age for the African U6 haplogroup branch to 30,000 yBP (95% HPD, 34-25,000 yBP). The divergence time for the African-specific M1 we estimate to 24,000 yBP (95% HPD, 29-20,000 yBP), which is considerably younger and not overlapping with a previous estimate of 37,000±7,000 yBP by (*7*) but again similar to the date reported in (*92*) of ~26,000 yBP. The coalescence age for haplogroup M1b, which we find in one Taforalt individual, is 20,000 yBP (95% HPD, 24-16,000 yBP). The divergence time of haplogroup U6a found in six Taforalt individuals is calculated to 24,000 yBP (95% HPD, 28-20,000 yBP), and very closely approximates the divergence time of African haplogroup M1 (red and black distributions, respectively in Fig. S23).

In order to monitor the change in effective population size for haplogroups M1 and U6 we repeated BEAST analyses as reported above independently for both haplogroups. For haplogroup M1 this resulted in 51 present-day sequences (*7*), one Taforalt sequence and one L3 sequence as outgroup (16,549 positions considered). For haplogroup U6 this resulted in 30 present-day sequences (*7*), six Taforalt, Muierii1 and one L3 sequence as outgroup (16,552 positions considered). After running BEAST for both haplogroups with identical parameters as previously mentioned we used Tracer v1.6 (*24*) to reconstruct Bayesian skyline plots selecting linear change and the default 100 bins (Fig. S24). We find that the African-specific U6 and M1 lineages are characterized by a similar, almost exponential, increase in effective population size. However, the onset for their population expansion is not synchronous in time as previously reported in (*25*); whereas in our estimates the expansion for the M1 lineage begins at around 14,000 yBP, U6 starts expanding ~12.000 years earlier at ~26,000 yBP (Fig. S24). Interestingly, our estimates for divergence time of U6a and M1 (~24,000 yBP) and effective population size increase in U6 (26,000 yBP) approximates the earliest appearance of the oldest Iberomaurusian in Northwest Africa (25,845-25,270 cal. yBP at Tamar Hat (*26*)) and the emergence of a MSA-LSA Transitional Technology in Taforalt (24,769-23,940 cal. yBP (*32*)).

S11. Phenotypic informative SNPs analysis

Marieke van de Loosdrecht*, Choongwon Jeong*

762
763 We investigated individual genotypes of a dozen SNPs well known for their associations with
764 phenotypes of interest, such as lactose tolerance, skin pigmentation, hair straightness and
765 susceptibility for mycobacterial infections (*27, 80, 93*) (Fig. S25). We restricted this analysis to
766 the four Taforalt individuals with >450,000 SNPs mapping to the Human Origins panel. The
767 numbers of high quality bases matching with the ancestral (non-effect) and derived (effect) alleles
768 were retrieved from bam.files using the samtools *mpileup* function, after excluding reads with
769 Phred-scale mapping quality score (MAPQ) <30 and bases with base quality score (BASEQ) <30.
770 We also calculated genotype likelihoods using the UnifiedGenotyper module of the Genome
771 Analysis Toolkit (GATK) v.3.5.
772
773 Two derived allele variants in the *SLC24A5* gene associated with predicting light-skin color in
774 individuals with European and South Asian (Indian, Pakistani) ancestry are rs1426654 (derived
775 state A, ancestral state G (*94*)) and rs16891982 (derived state G, ancestral state C (*95*)).
776 Individuals with a homozygous derived state for both these SNPs have been found in early
777 Neolithic populations (Anatolia, Europe) (*16*)). Our results show that these derived alleles are
778 absent in the Taforalt individuals analyzed; all of them have a homozygous ancestral genotype for
779 both SNPs. The derived mutation for rs12913832 in the *OCA2* gene is associated with blue eye
780 color. A homozygous derived allele state at this position is the dominant determinant of light eye
781 color in present-day Europeans and occurs at 100% frequency in Mesolithic hunter-gatherers (*96,*
782 *97*). Individuals with the ancestral allele A (homozygous or heterozygous) for this SNP show
783 brown eye color 80% of the time (*97*). For all the Taforalt individuals we find a homozygous
784 ancestral genotype GG, predictive of brown eye color. In addition, all individuals show the
785 ancestral GG genotype for SNP rs12896399 located in the *SLC24A4* gene, providing further
786 support for dark eye pigmentation (*93*). The *TCHH1* gene codes for trichohyalin, a protein active
787 in hair follicle roots. For all Taforalt individuals we find the derived homozygous AA genotype
788 for SNP rs17646946 in this gene, which has been associated with straighter hair in Europeans
789 (allelic effect (β) = 0.4-0.5, explained variance = 6.11%) (*98*).
790
791 We also tested for SNPs located in *MCM6* gene for which derived allele variants are associated
792 with predicting lactose tolerance in Europeans (rs4988235) (*99*), Africans (rs41456145 (*100,*
793 *101*), rs4988236 (*101, 102*), rs145946881(*100, 101*)) and Middle Easterners (rs41380347) (*100,*
794 *101*). The Taforalt individuals have a homozygous ancestral genotype for all these SNP positions,
795 consistent with previous studies reporting the appearance of the European allele much later than
796 the Epipaleolithic (*80, 103*). Therefore, the Taforalt individuals were most likely lactose
797 intolerant and could not digest milk as adults.
798
799 Finally, we analyzed two SNP variants that have been studied in relation to susceptibility for
800 mycobacterial infections (*80*). A derived allele of rs4833103 in the *TLR1-6-10* gene, which
801 encodes an active functional component in the innate immune response, is associated with a
802 possibly increased resistance to leprosy, tuberculosis and other mycobacteria (*104, 105*). All
803 Taforalt individuals show the derived homozygote CC genotype. In present-day populations the
804 derived allele state has a high frequency in Eurasians but a low frequency, or absence, in sub-
805 Saharan Africans. Notably, the derived allele has been found in low frequencies in European, and
806 to a lesser degree in Anatolian, early Neolithic populations but not in any pre-Neolithic hunter-
807 gatherers so far (*80*). The high frequency of the derived state in the Taforalt individuals may
808 suggest that the presence of this variant predates the Neolithic transition in this region. In
809 addition, the derived allele of rs2269424 in the major histocompatibility complex on chromosome
810 6 may be associated with mycobacteria resistance (*80*). Three Taforalt individuals have the
811 heterozygous genotype and one individual is homozygous derived (AA).
812

S12. Y-chromosome haplogroup assignment

Choongwon Jeong*, Marieke van de Loosdrecht*

To determine the Y chromosome haplogroup of the Taforalt male individuals, we used the yHaplo program (*106*), which relies on the ancestry-informative markers in the ISOGG (International Society of Genetic Genealogy) data set. Specifically, we randomly chose one allele for each of 13,508 ISOGG SNPs, excluding strand-ambiguous (C/G and A/T) ones to represent individual genotypes and used it as an input for yHaplo. The program searches down from the root to the most derived branch of Y haplogroup tree based on derived markers supporting specific branches. Because missing data occurs frequently in ancient DNA data, this automated search often stops before the most derived position on a specific branch. This is not because the data positively support a deep-branching haplogroup but because data is missing for key informative SNPs that define particular nodes/branching points. To prevent this behavior, we manually checked haplogroup support further downstream from the last derived alleles that were detected automatically for each individual.

In all six males, we observe haplogroup E1b1b, more specifically E1b1b1a1 (M-78) in five of six (Table S16). This haplogroup is most frequent in present-day North and Northeast African populations, such as Oromo, Somali and Moroccan Arabs (*18*). A previous study reported that Natufians and Neolithic Levant individuals had E1b1b haplogroups, although they tended to belong to E1b1b1b (*16*).

S13. Genetic relatedness among Taforalt individuals

Choongwon Jeong, Marieke van de Loosdrecht

We estimated genetic relatedness among the ancient Taforalt individuals using pairwise mismatch rate of haploid genotypes, following an idea presented in (*107*). Specifically, we aim at estimating the coefficient of relationship (r) between a pair of individuals, defined as the probability of an allele from individual 1 having a copy of it (i.e. identity-by-descent, IBD) in individual 2. Therefore, assuming that an allele with population frequency p_i is sampled for the $i^{\text{-th}}$ SNP from individual 1, we have a probability $r/2$ and $1\text{-}r/2$ to sample an IBD and a non-IDB allele from individual 2, respectively. Then, the expected pairwise mismatch rate (PM) across all sites becomes:

$$PM = \Sigma\, Pr(\text{mismatch} \mid \text{IBD}) \times Pr(\text{IBD}) + Pr(\text{mismatch} \mid \text{non-IBD}) \times Pr(\text{non-IBD})$$
$$= \Sigma\, 0 \times (r/2) + (1 - p_i) \times (1 - r/2) = (1 - r/2) \times \Sigma\, (1 - p_i)$$

Therefore, we can expect that the pairwise mismatch rate between two unrelated individuals (r=0) is twice as big as that between two identical individuals (r=1), and it varies as a linear function of r in between. We calculated the pairwise mismatch rate between all pairs of sequencing libraries, which we have several for each individual, to obtain an accurate estimate of the baseline. We used pairs with at least 1,000 SNPs covered by both individuals to exclude extremely noisy estimates.

As expected, intra-individual library pairs show similarly low pairwise mismatch rates with a mean value of 0.126 (Fig. S26). We find that two individuals, TAF008 and TAF016, show a high level of mismatch rate with libraries from the other individuals, with the level expected for unrelated individuals: the mean value is 0.245, very close to being the double of 0.126 (Fig.

S26A). Given that these two individuals have much lower C>T misincorporation rate than the other individuals (Table S3), we suggest that they represent a contaminant and not endogenous genomes, and thus are unrelated to the rest of individuals. Removing these two individuals, the other seven individuals, including two with some nuclear contamination (≤ 11.9%; Table S5), show pairwise mismatch rates much lower than that for unrelated pairs but close to that for the first-degree pairs (mean value 0.199; Fig. S26B). Given the heterogeneous mitochondrial haplotypes and young age of all individuals, this overall close genetic relationship cannot be explained by their siblingship. Instead, we suggest that these individuals share extra genetic affinity with each other due to the strong population bottleneck the Taforalt group must have gone through. In addition, high coefficients of relatedness between Taforalt individuals in general, indicate substantial amount of genetic inbreeding.

Interestingly, library pairs from two individuals, TAF011 and TAF012, have an even further reduced pairwise mismatch rate, with a mean value of 0.165 across all library pairs (Fig. S26B). This value matches well with the mean value of the baseline (0.126 for intra-individual library pairs) and the inter-individual values (0.199), suggesting that they are first-degree relatives. Considering that they are infants and have identical mitochondrial haplotypes (Table S4), we conclude that they are siblings.

A recent study analyzed five Upper Paleolithic genomes from Sunghir and reported a limited kinship between them in spite of their shared material culture, overlapping radiocarbon dates and similar genetic compositions (*108*). Although the Taforalt individuals show a strong genetic similarity with each other, our results suggest that it is due to population bottleneck rather than recent inbreeding or kinship. This is because the within-individual comparisons do not show a reduction in heterozygosity in comparison to other populations. Together with the results from Sunghir, this may suggest that consanguineous mating between close relatives was not widespread in multiple Upper Paleolithic societies. However, it is too speculative to generalize such results based on limited per-site sample sizes, the big temporal and geographical distance between Taforalt and Sunghir, and the distinct age structure of the individuals analyzed.

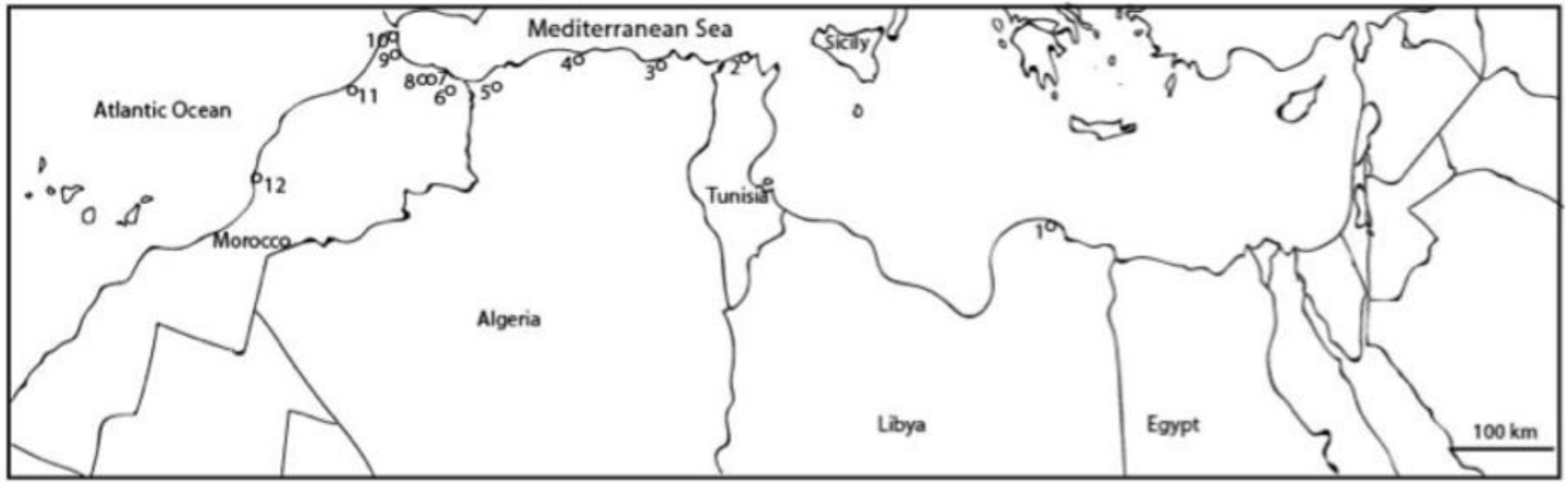


Fig. S1. Maps showing the location of the main Iberomaurusian sites in North Africa. 1.Haua Fteah, 2.Ouchtata, 3.Tamar Hat, 4.Rassel, 5.La Mouillah, 6.Taforalt, 7.Ifri n'Ammar, 8.Ifri el Baroud, 9.Kehf el Hammar, 10.Gar Cahal, 11.Contrebandiers, 12.Cap Rhir. See (*10*) for site descriptions.

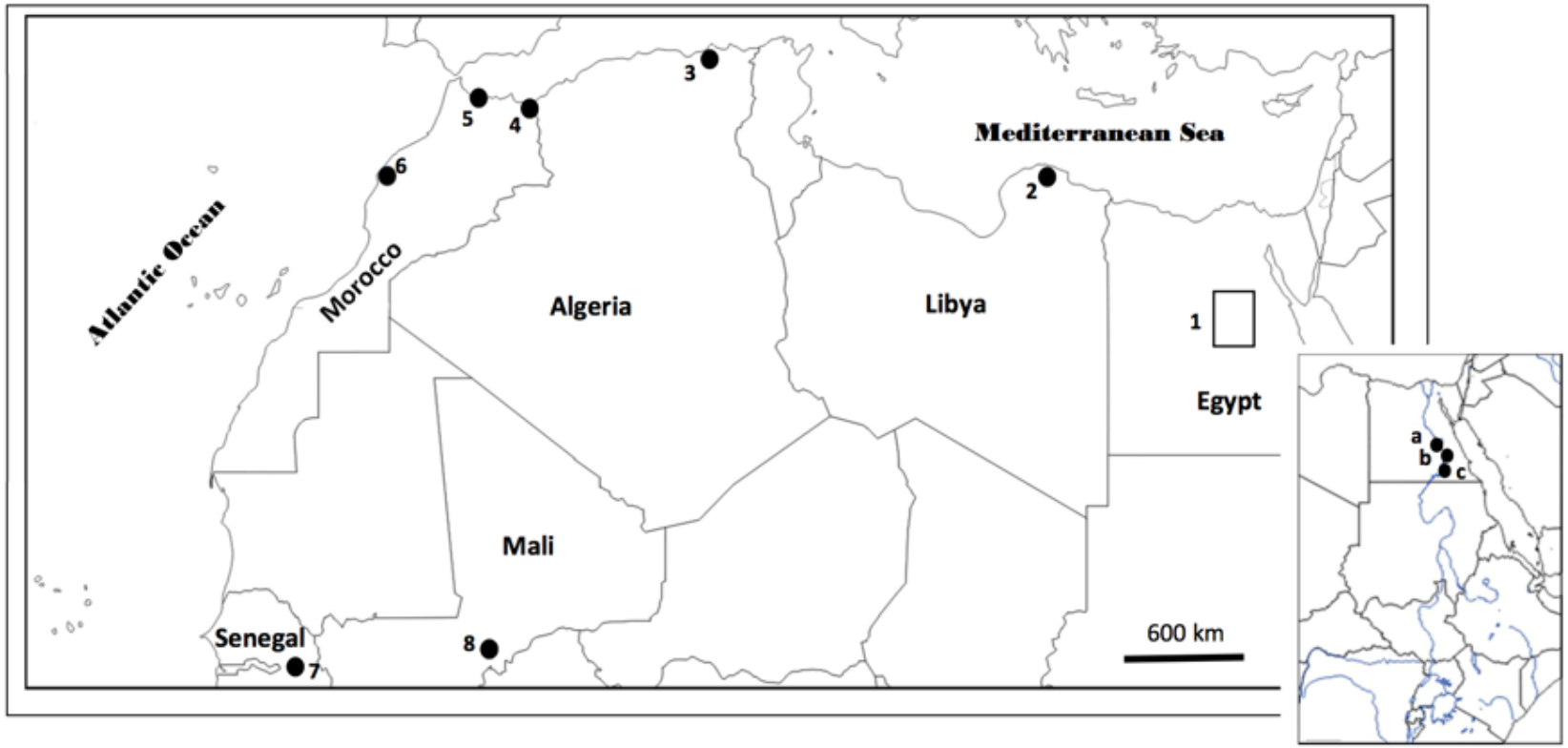


Fig. S2. Maps showing a few MSA and LSA sites in North and West Africa relevant to this study. a.Deir el Fakhouri (*109*), b.Gebel Silsila (*109*), c.Wadi Kubbaniya (*110, 111*), 2.Haua Fteah, 3.Tamar Hat, 4.Taforalt, 5.Kehf el Hammar, 6.Cap Rhir, 7.Falémé Valley (*48*), 8.Ounjougou (*47*). Individual site citations are given. For the other sites, see description in (*10*).

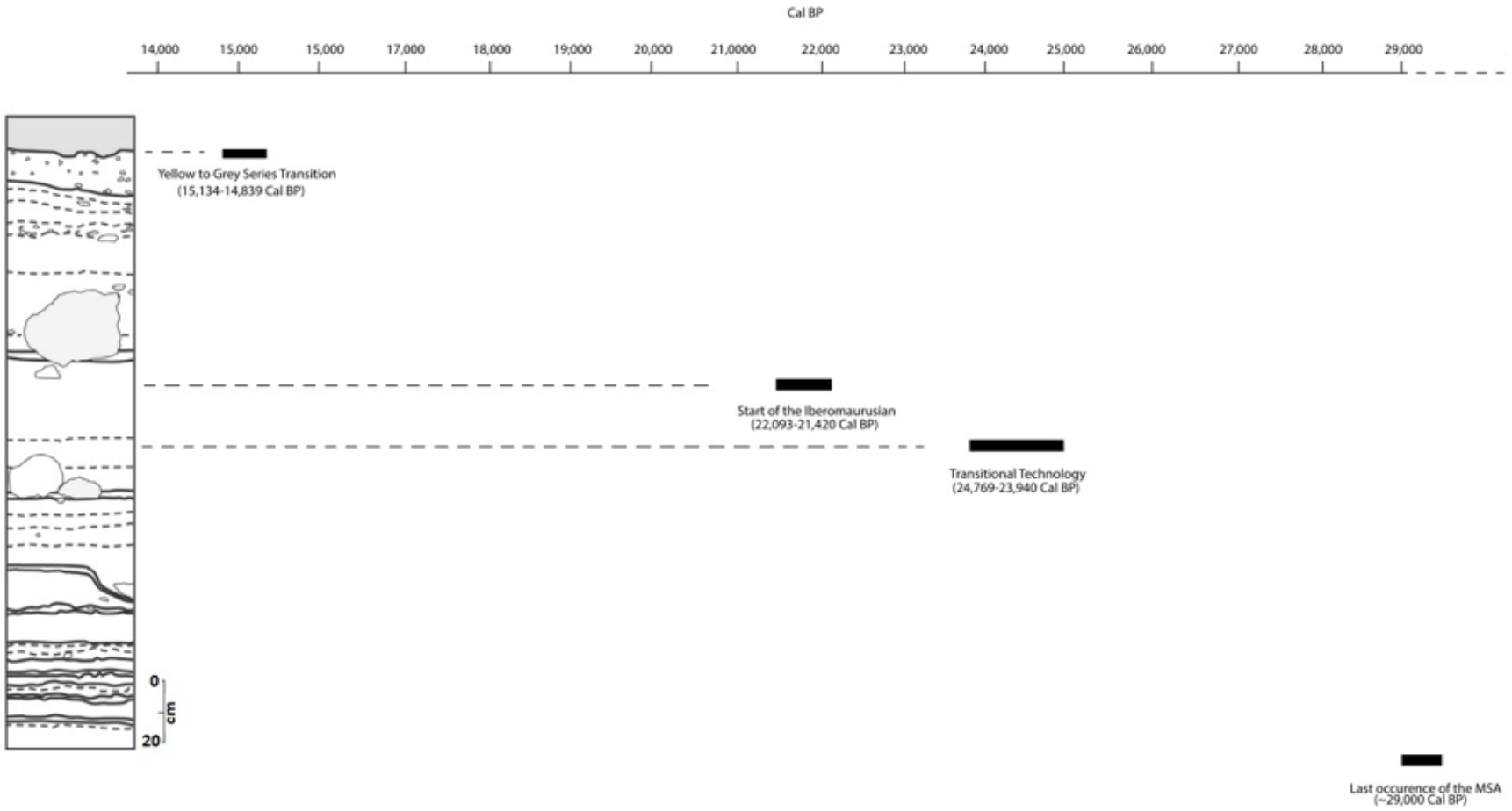
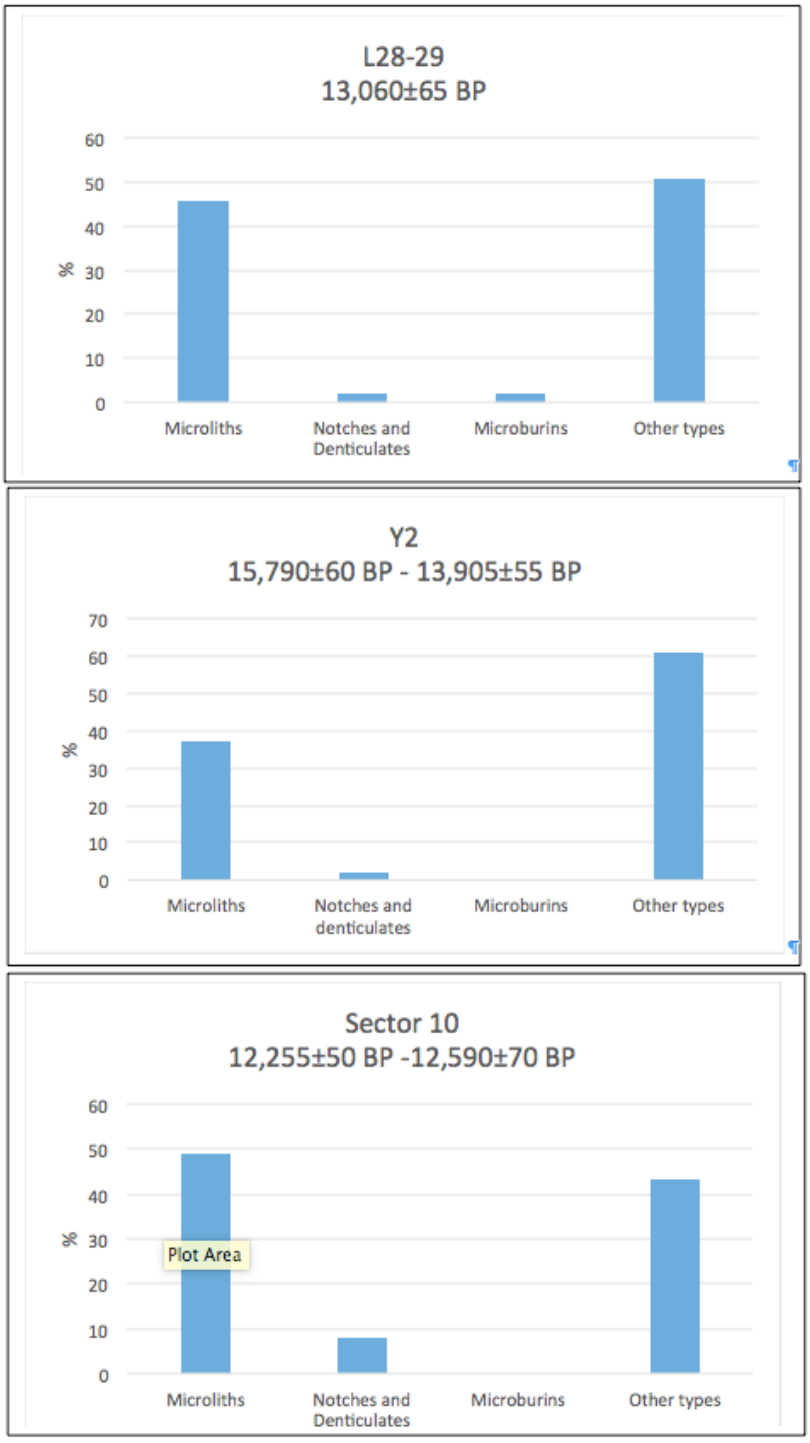


Fig. S3. Schematic section through the MSA and LSA Iberomaurusian deposits with their respective AMS radiocarbon dates. For more detailed information on the deposits and the dating, see (*10, 32*)

914



915
916
917
918 **Fig. S4. Frequencies of the main typological types in the Grey (L28-29) and Yellow (Y2)**
919 **Series & Sector 10.** Data from (*112*).
920

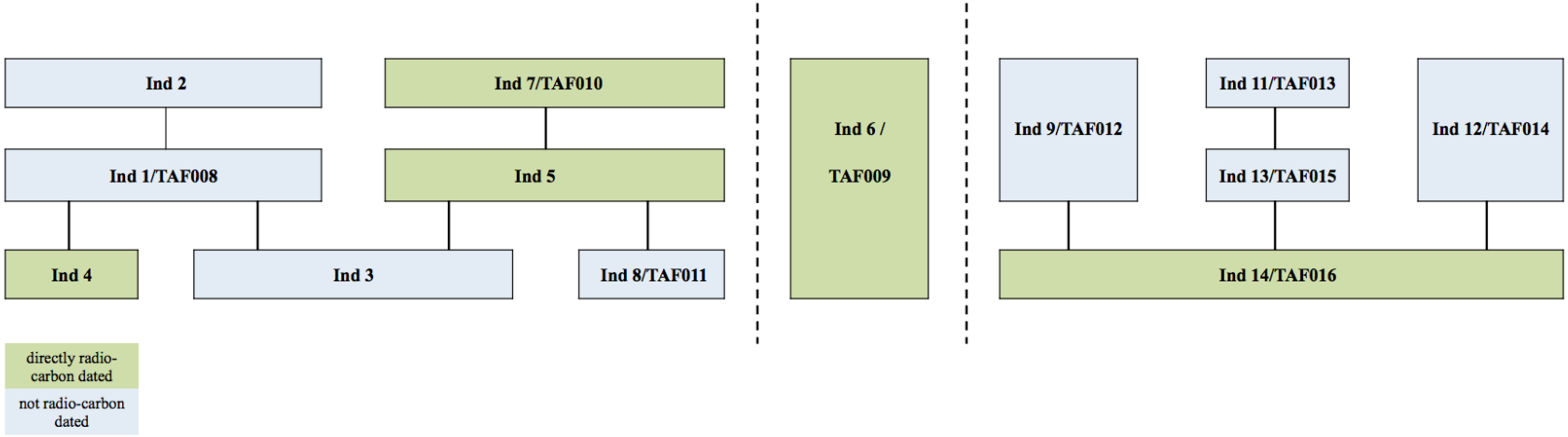


Fig. S5. A stratigraphic matrix that shows the burial context in Sector 10 for the Taforalt individuals included in this study. Ind# refers to the archaeological ID conforming to the numerical system from (*52*). TAF# is the individual ID used in this study for which we generated ancient genome data. Individuals 1-5 and 7-8 were situated closer to the cave entrance, and Individuals 9-14 towards the rear cave wall.

166

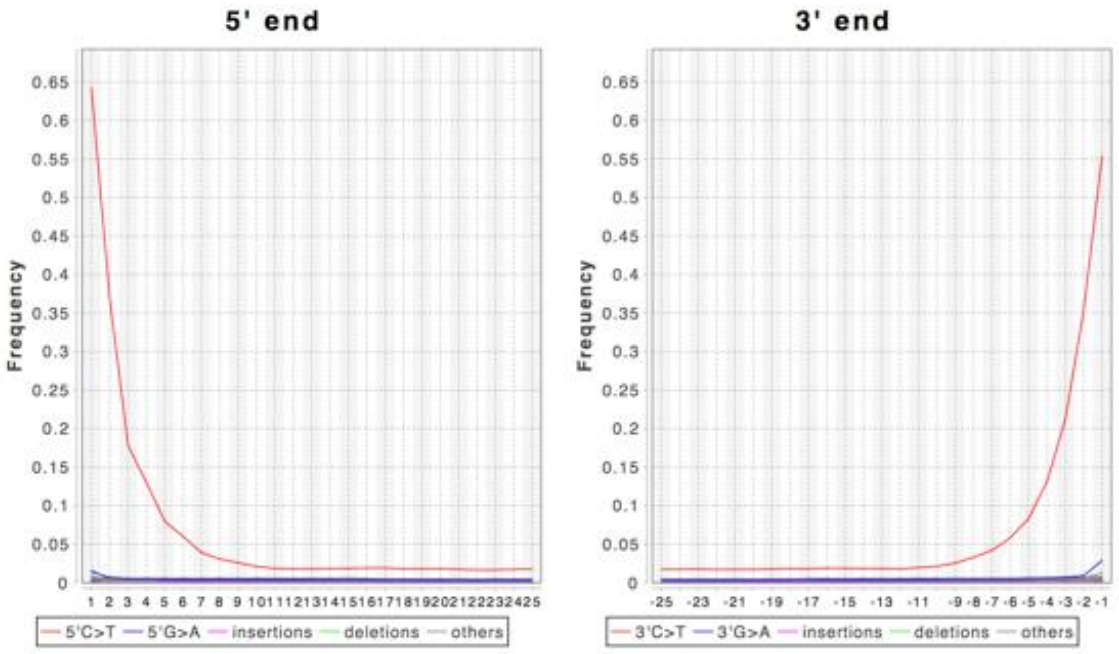


Fig. S6. A DNA damage plot of a representative sequencing library for Taforalt. High rates of C>T misincorporation are present at both ends of molecules, extending up to 10 base pairs inside. This plot is for the library TAF012.A0201.

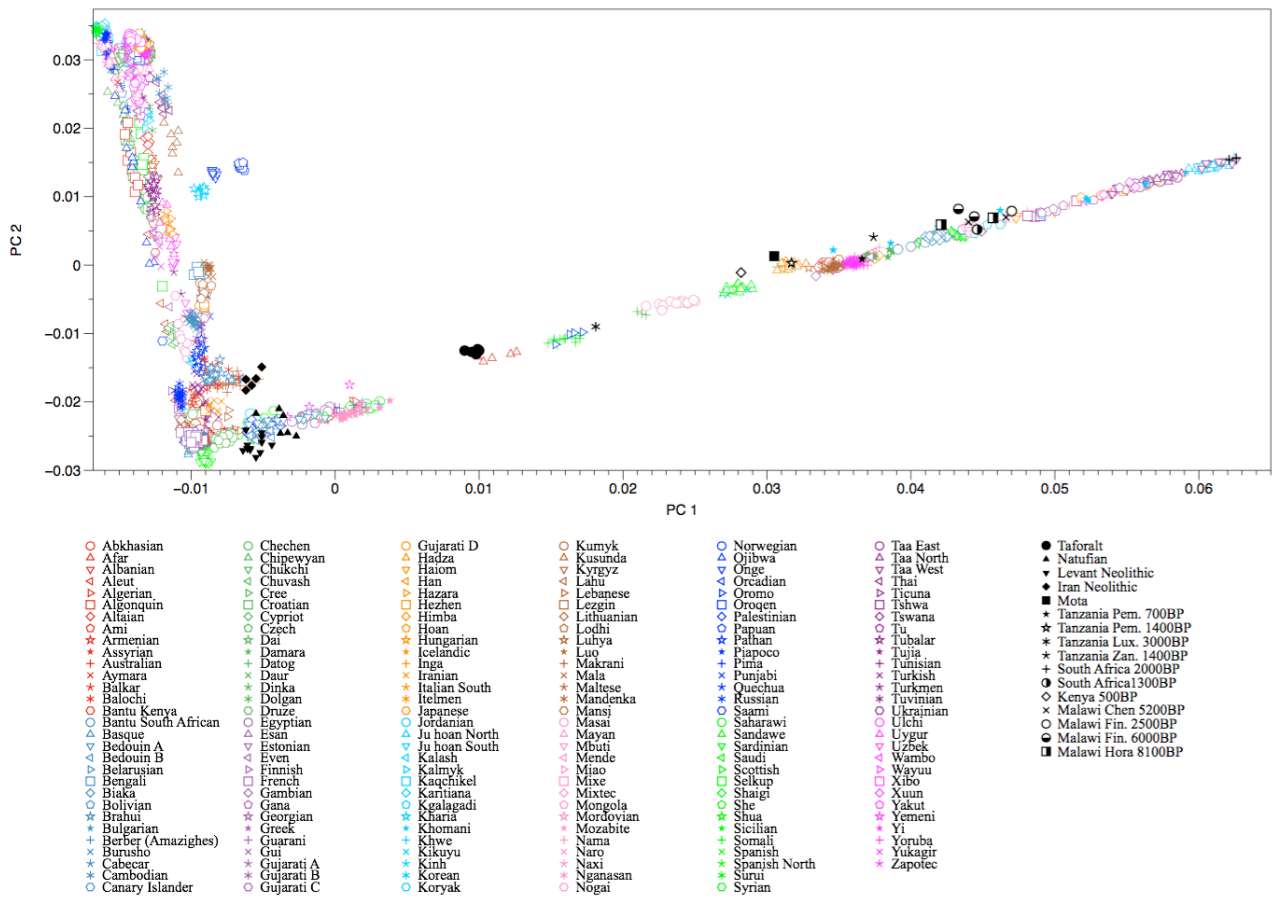


Fig. S7. **A global PCA plot showing the genetic relatedness of Taforalt to worldwide populations.** Taforalt (black filled circles), and other ancient individuals from Africa and the Middle East (black symbols), are projected on the PCs constructed from the genetic variation in 178 global present-day populations (colored symbols). Taforalt fall on an intermediate position between present-day North Africans and sub-Saharan (East) Africans.

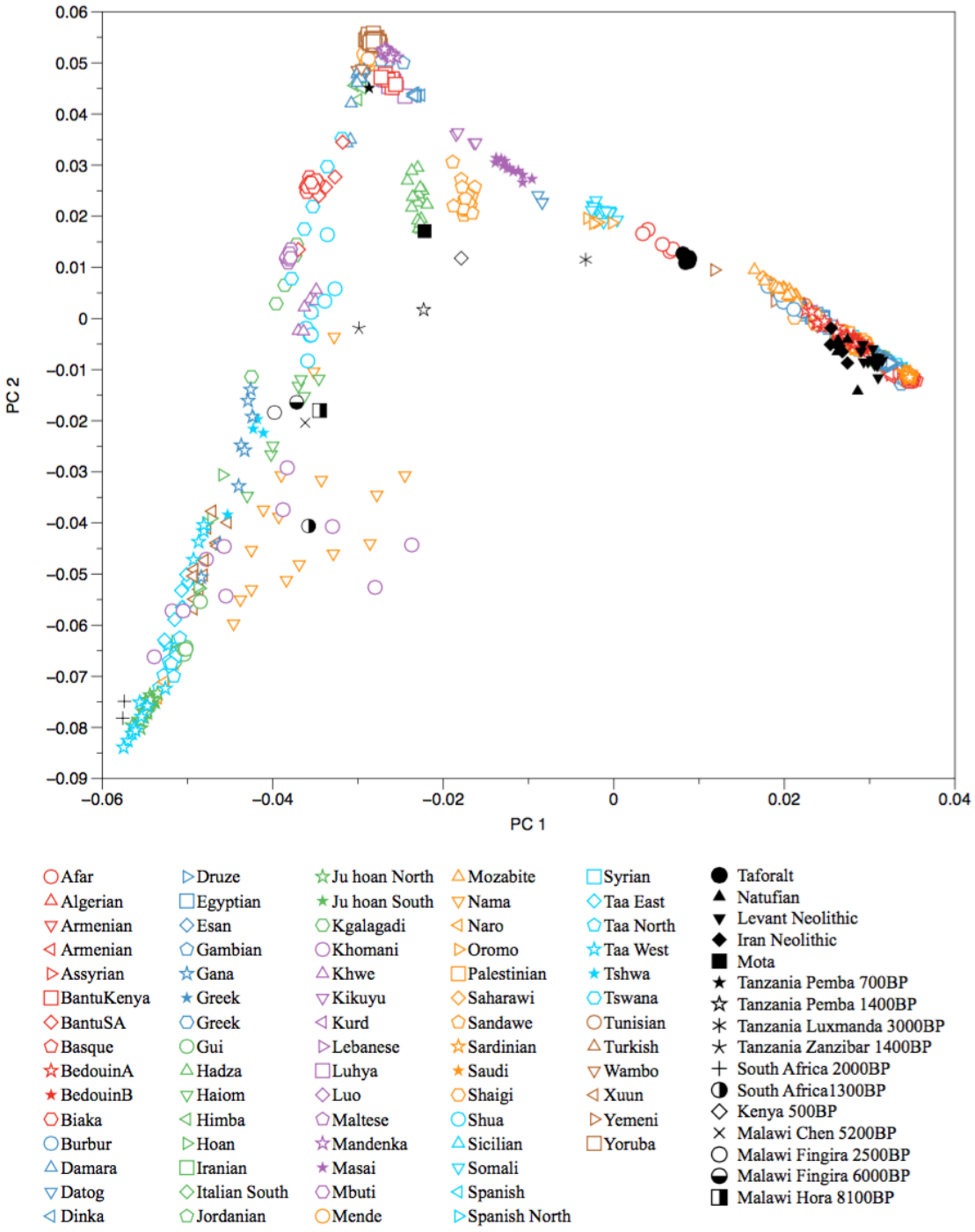


Fig. S8. Taforalt individuals on the top PCs of present-day African, Near Eastern and South European populations. Taforalt (black filled circles), and other ancient individuals from Africa and the Middle East (black symbols) were projected on PCs constructed from the genetic variation in 72 present-day populations (colored symbols). Taforalt individuals fall on closest to North African Mozabite and Saharawi, and to East African Afars on the Near Eastern and Sub-Saharan African directions, respectively.

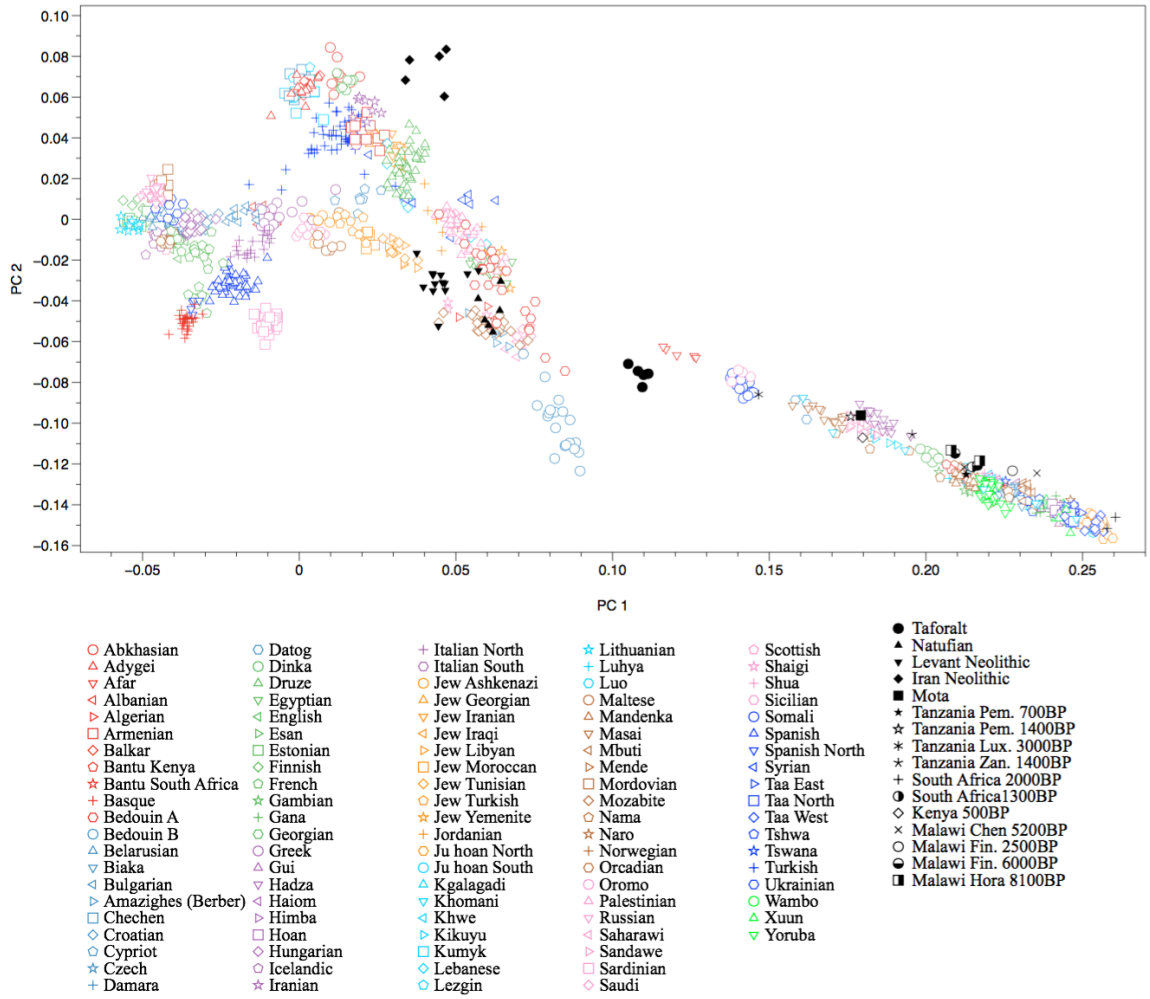


Fig. S9. Taforalt individuals on the top PCs of present-day West Eurasian populations. Taforalt individuals (black filled circles), other ancient African and Middle Eastern individuals (black symbols), and present-day Africans (colored symbols in the right side of Taforalt) are projected on PCs constructed from the genetic variation in 102 present-day populations (colored symbols in the left side of Taforalt).

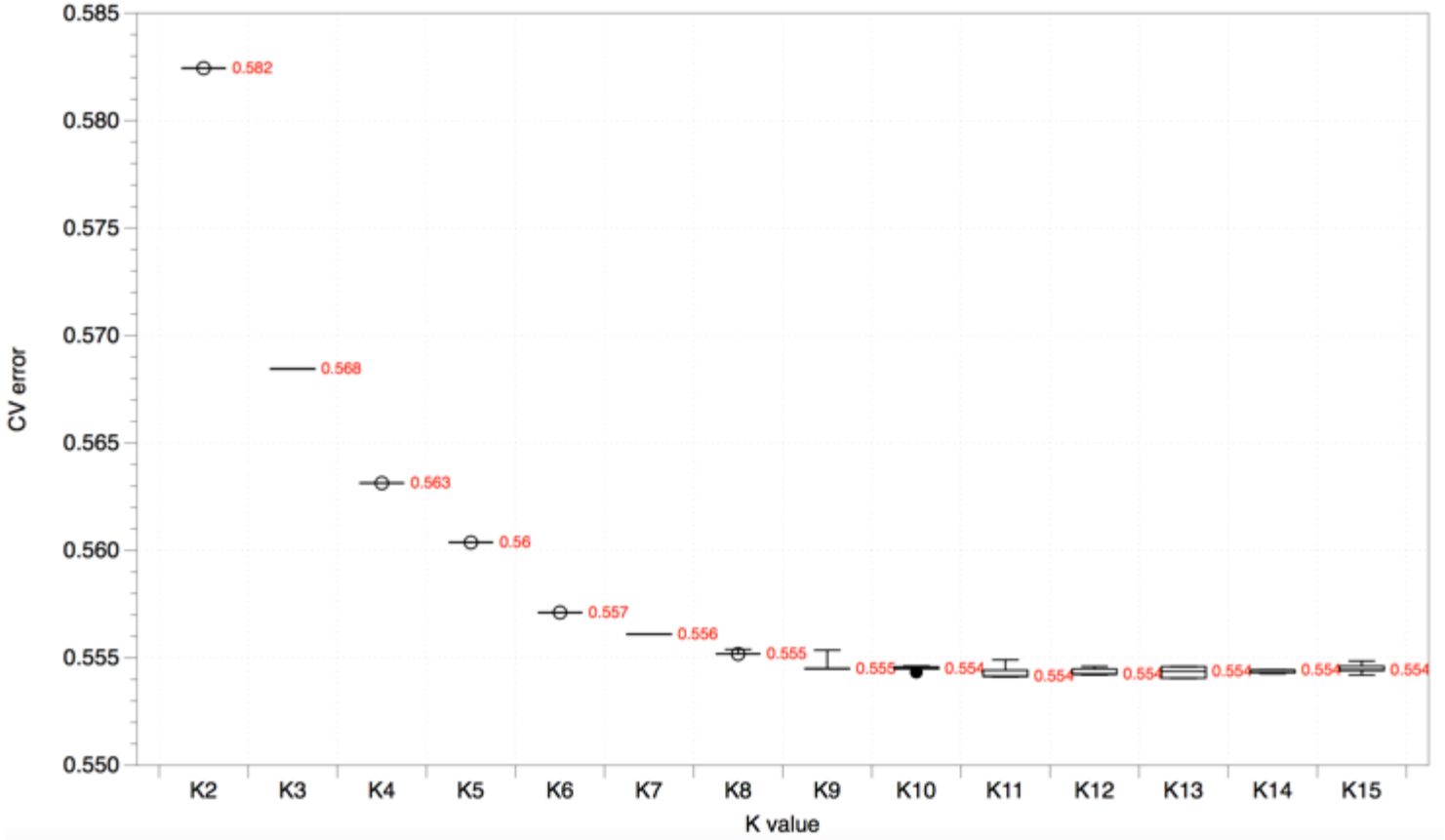


Fig. S10. Cross-validation (CV) error values as a function of the number of clusters (K) in ADMIXTURE analyses. For each K value, five replicates with random seeds were performed. The lowest CV errors are observed for K=9 to K=11.

966
967
968
969 **Fig. S11. ADMIXTURE results for a few informative K values.** In K=2, Taforalt individuals
970 are modeled as a mixture of sub-Saharan African and Eurasian ancestries. In K=9, major ancestry
971 components in Taforalt are maximized in early Holocene Levantines (green) and West Africans
972 (purple). In K=10, a new component is assigned to East African Hadza (brown) and it shows up
973 in Taforalt in large quantity. K=11 singles out Taforalt as a distinct genetic component. For none
974 of the K values Taforalt has a distinct ancestry uniquely maximized in Paleolithic Europeans.
975

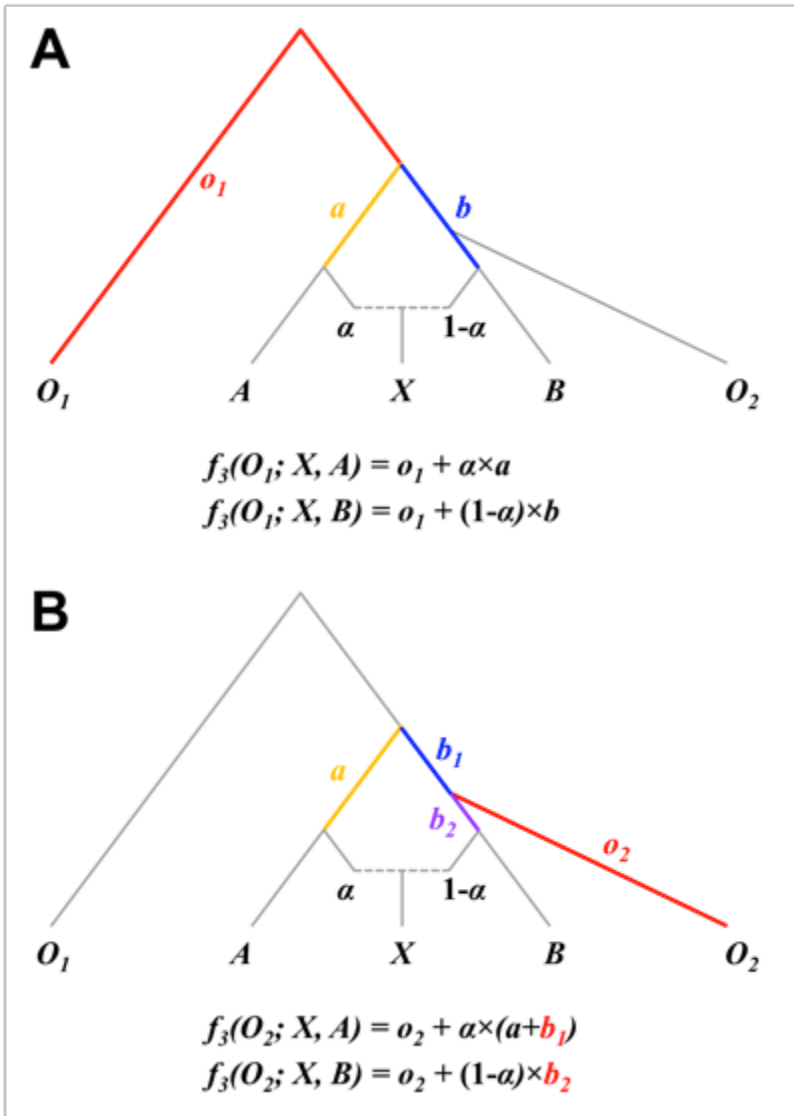
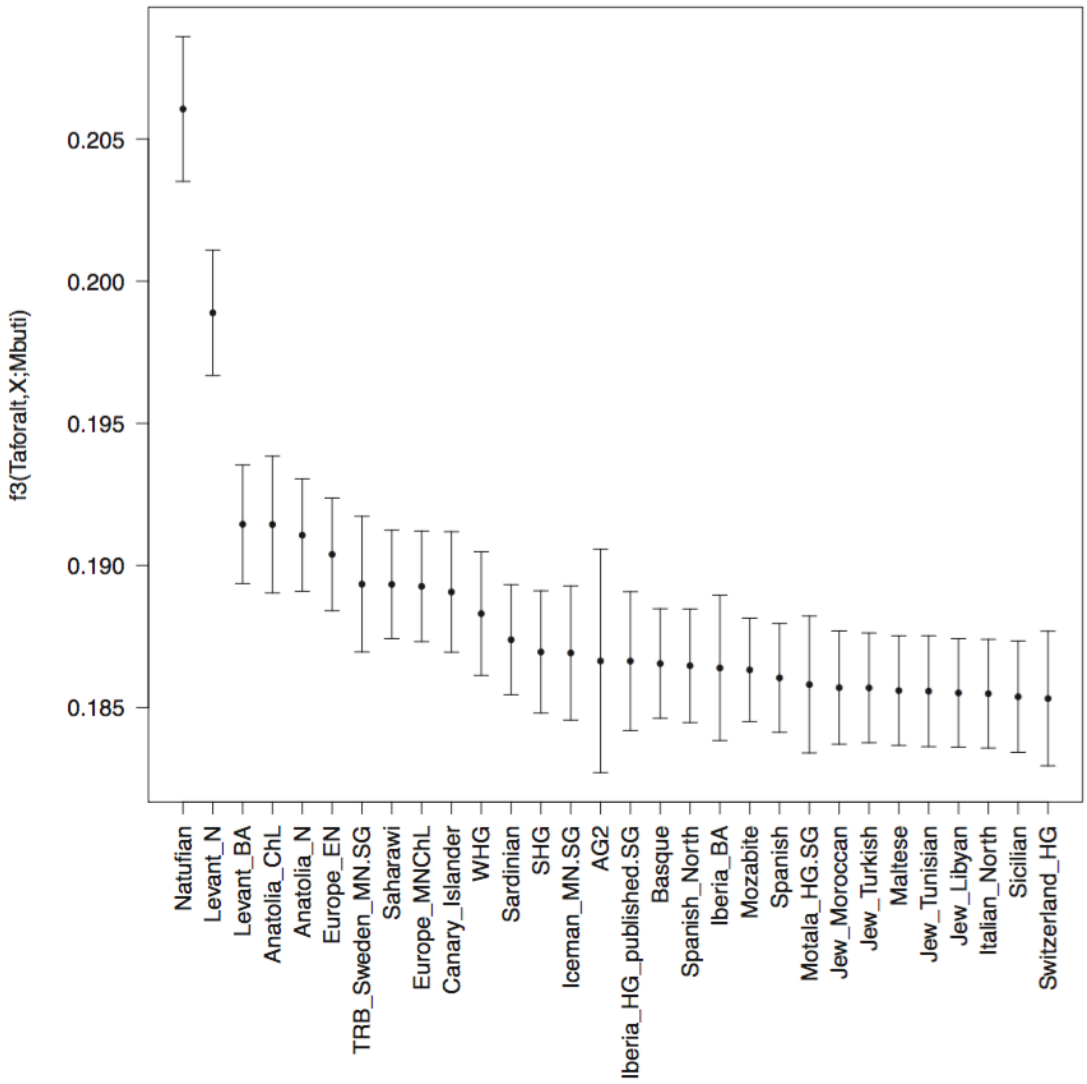
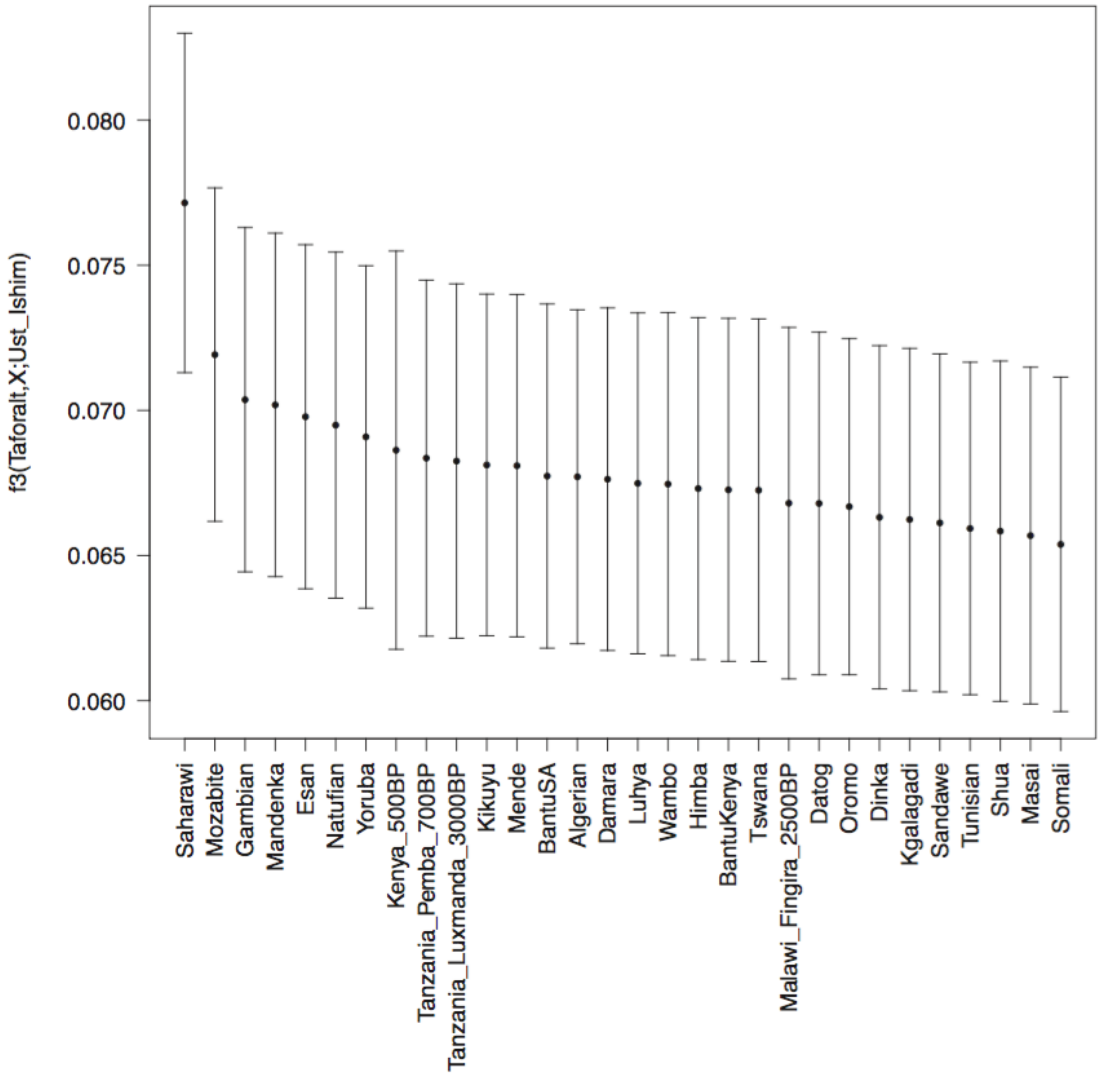


Fig. S12. Two-way admixture scenarios showing the effect of outgroup choice on the magnitude of outgroup-f_3 statistic. (A) When the true outgroup (O_1) is chosen, shared genetic drifts between (A, X) and (B, X) pairs (a and b, respectively), are proportional to the magnitude of outgroup-f_3. **(B)** When a group closer to B than to A is chosen as an operational outgroup, only a fraction of the shared genetic drift between X and B (b_2) contribute to the outgroup-f_3 between them, while the rest (b_1) contribute to outgroup-f_3 between X and A. Therefore, choosing an outgroup close to one of the sources down-weighs outgroup-f_3 of the admixed target with that source, while up-weighs that with the other sources.



989
990
991
992 **Fig. S13**. **The top 30 outgroup-f_3 signals between Taforalt and worldwide ancient and**
993 **present-day** populations, using Central African Mbuti as an outgroup. Early Holocene Levantine
994 populations, Natufian and Levant_N, show the biggest signal. The vertical bars represent ± 1
995 standard error (SE) estimated by 5 cM block jackknifing.
996



997
998
999
1000 **Fig. S14. The top 30 outgroup-f_3 signals between Taforalt and worldwide ancient and**
1001 **present-day populations, using a 45,000 yBP west Siberian genome Ust'-Ishim as an**
1002 **outgroup.** North African populations, such as Saharawi and Mozabite, show the biggest signals,
1003 followed by West Africans. The vertical bars represent ± 1 SE estimated by 5 cM block
1004 jackknifing.
1005
1006

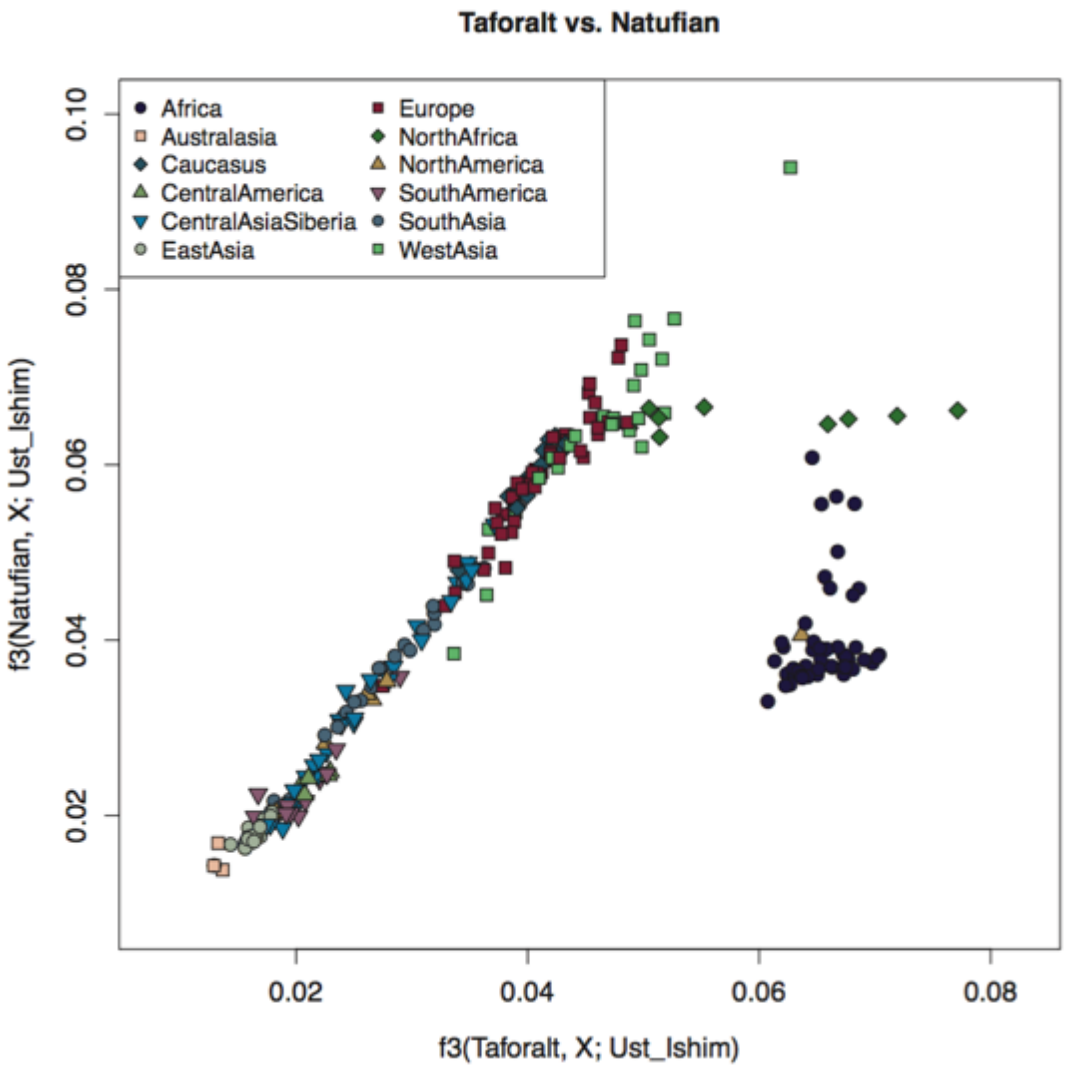


Fig. S15. A bi-plot of outgroup-f_3 statistics with the Upper Paleolithic Siberian Ust'-Ishim (x-axis) and Central African Mbuti (y-axis) as outgroups. All Eurasian populations fall on the line, suggesting that the relative affinity between Taforalt and Eurasian populations can be explained by the Natufian-like ancestry in Taforalt. In contrast, African populations deviate from the Eurasian line, showing the presence of shared genetic drift with Taforalt that cannot be explained by their shared Natufian-like ancestry.

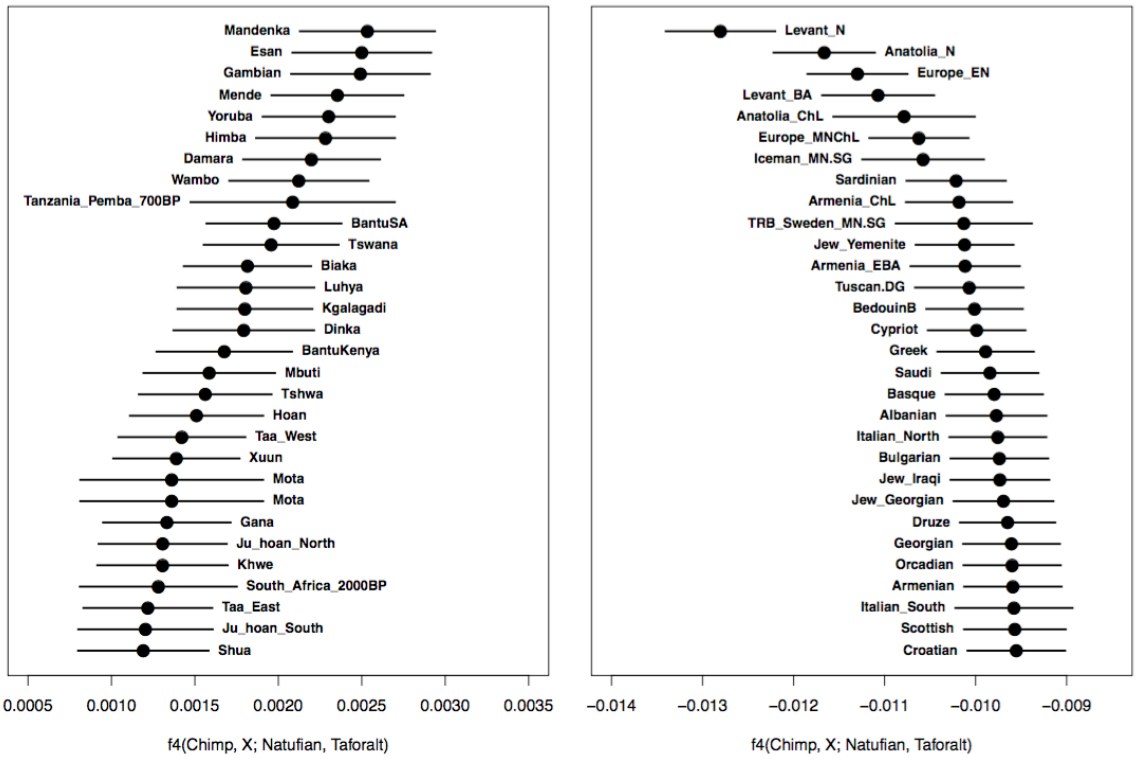


Fig. S16. The relative genetic affinity of Natufians and Taforalt against worldwide ancient and present-day populations, measured by f_4(Chimpanzee, X; Natufian, Taforalt). (A) The top 30 most positive f_4 values, suggesting extra genetic affinity with Taforalt, are found for Sub-Saharan Africans, most notably West Africans (Mandenka, Esan, Gambian, Mende, Yoruba) and Southeast Africans (Himba, Damara, Wambo). **(B)** In contrast, the top 30 most negative f_4 values are found for Eurasian populations, most notably Early Neolithic groups. Eurasians are hence genetically closer to Natufians than to Taforalt. The horizontal bars represent ± 1 SE estimated by 5 cM block jackknifing.

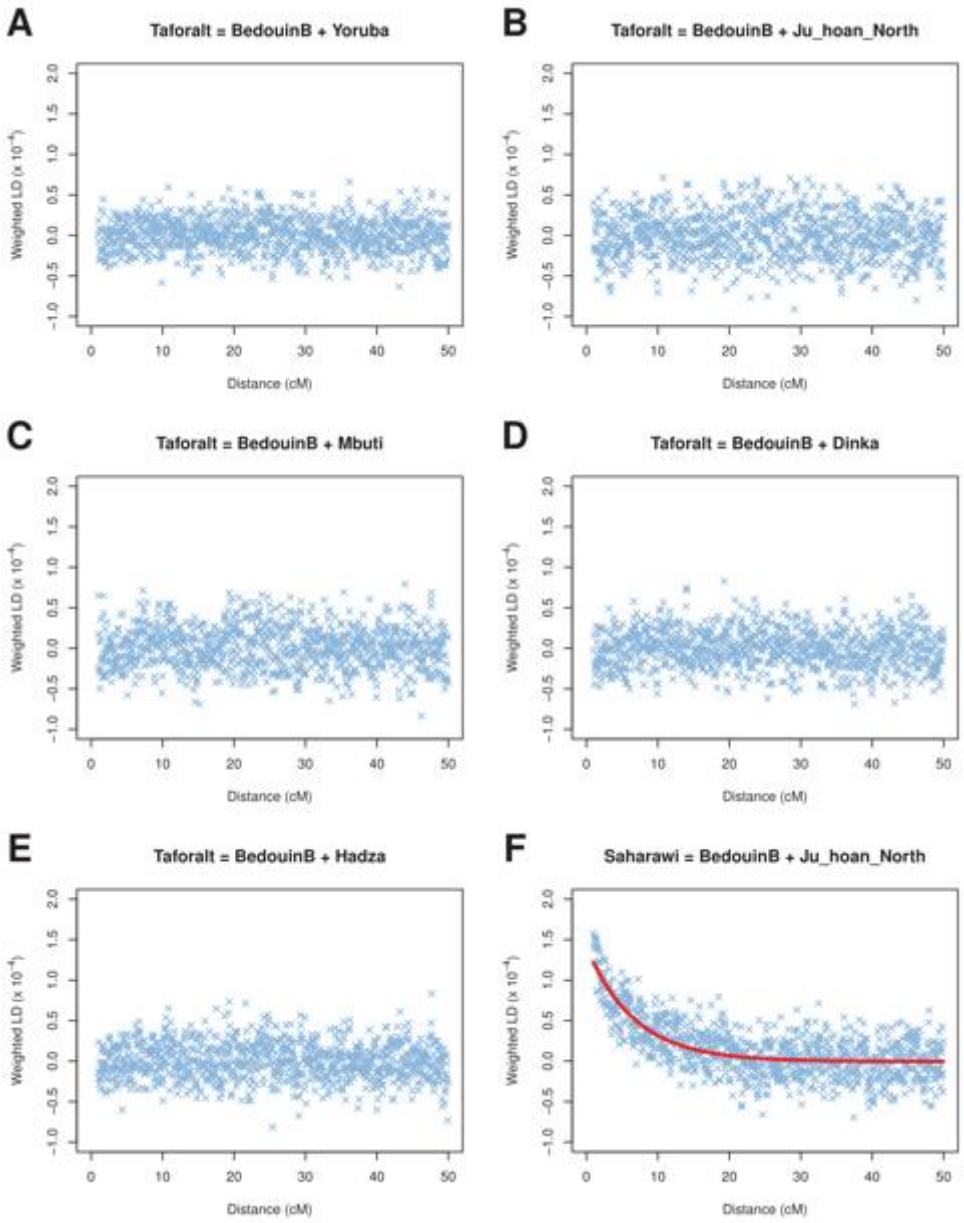
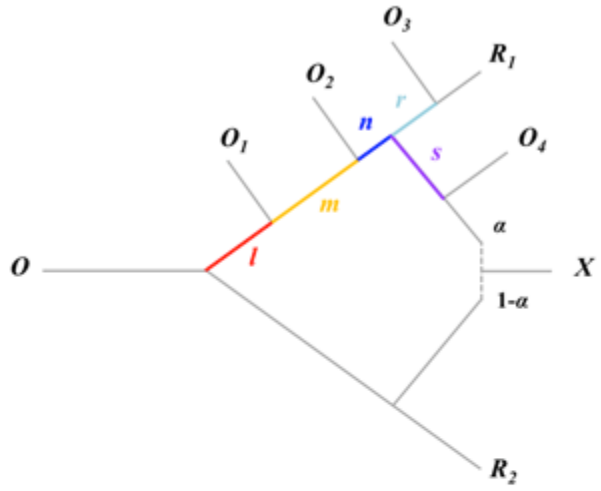


Fig. S17. Weighted LD decay in Taforalt. (A-E) We estimated the weighted LD decay pattern in Taforalt using BedouinB and one sub-Saharan African population as references. Regardless of our choice of reference, no exponential decay pattern against genetic distance was observed, suggesting that there is no signal of recent admixture in Taforalt. We obtained similar results when we tested different Near Eastern references, including Sardinian, Palestinian, Natufian and Levant_N (data not shown). **(F)** In contrast, present-day North African Saharawi shows a clear LD decay pattern with recent admixture time estimate (17.1 ± 3.1 generations ago; $p = 4.2\times10^{-8}$ for the significance of exponential decay).



A. Using O_1 as an outgroup

$Q_1 = f_4(O, O_1; R_1, X) = -(1-\alpha)\times l$

$Q_2 = f_4(O, O_1; R_2, X) = \alpha\times l$

$Q = \alpha\times Q_1 + (1-\alpha)\times Q_2 = 0$

B. Using O_2 as an outgroup

$Q_1 = f_4(O, O_2; R_1, X) = -(1-\alpha)\times(l+m)$

$Q_2 = f_4(O, O_2; R_2, X) = \alpha\times(l+m)$

$Q = \alpha\times Q_1 + (1-\alpha)\times Q_2 = 0$

C. Using O_3 as an outgroup

$Q_1 = f_4(O, O_3; R_1, X) = -(1-\alpha)\times(l+m+n)-r$

$Q_2 = f_4(O, O_3; R_2, X) = \alpha\times(l+m+n)$

$Q = \alpha\times Q_1 + (1-\alpha)\times Q_2 = -\alpha\times r \neq 0$

D. Using O_4 as an outgroup

$Q_1 = f_4(O, O_4; R_1, X) = -(1-\alpha)\times(l+m+n)+\alpha\times s$

$Q_2 = f_4(O, O_4; R_2, X) = \alpha\times(l+m+n+s)$

$Q = \alpha\times Q_1 + (1-\alpha)\times Q_2 = \alpha\times s \neq 0$

Fig. S18. A schematic of the impact of choosing outgroups for admixture modeling using qpAdm. In this simple two-way admixture scenario, the main goal is to distinguish ancestries in our target (X) associated with the sources A and B. For this purpose, we use one of the four outgroups (O_1 to O_4) which are more closely related to A than to B. **(A and B)** The outgroups (O_1 and O_2) are symmetrically related to R_1 and R_1-related ancestry in X. These outgroups therefore provide an unbiased estimate of α, the admixture proportion; i.e. the quantity Q becomes zero when we use the true value of α as the admixture proportion. Because O_2 is closer to R_1 than O_1 is, f_4 statistics have bigger magnitudes with O_2, and hence have a higher power of modeling. **(C and D)** We choose outgroups even more closely related to R_1, but they are not symmetrically related to R_1 and R_1-related ancestry in X. Therefore, modeling with these outgroups provide biased estimates of admixture proportion: i.e. the quantity Q becomes zero with the estimate of admixture proportion different from the true value α.

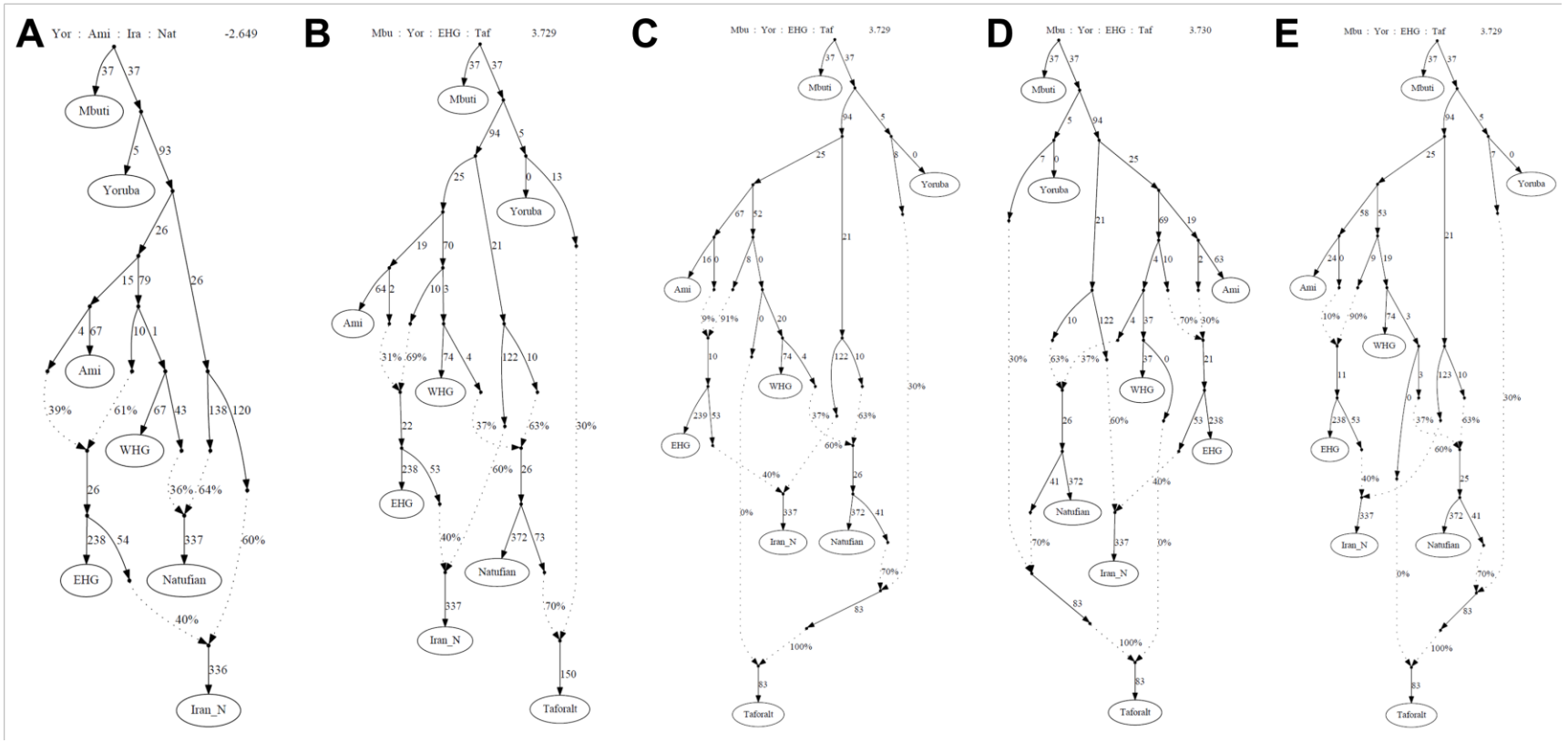


Figure S19. Placing Taforalt in the admixture graph based on the qpGraph program. (A) The best fitting scaffold graph including Mbuti, Yoruba, Ami, WHG, EHG, Iran_N and Natufian. **(B)** The best model of Taforalt on the scaffold graph is a mixture of Natufian (70%) and Yoruba (30%). **(C-E)** Adding an additional gene flow from a WHG-related branch into Taforalt uniformly estimates 0% contribution from this third ancestry and thus does not improve the model fit. On each graph, the largest f_4-statistic deviation from the expected value is provided in the unit of SE.

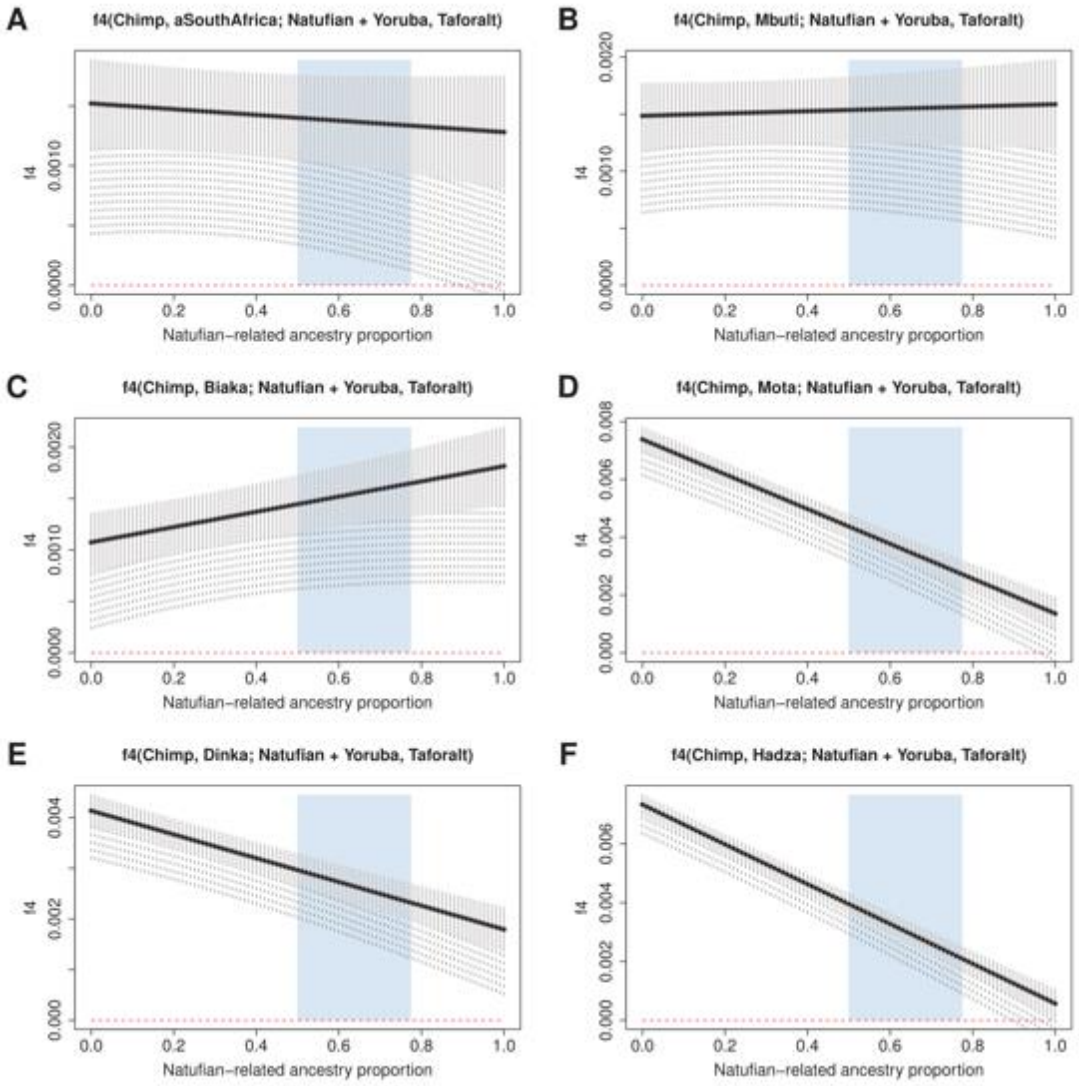


Fig. S20. Relative genetic affinity of representative sub-Saharan African groups to a mixture of Yoruba/Natufian and Taforalt, measured by f_4(Chimpanzee, African; Yoruba+Natufian, Taforalt). F_4 statistics were calculated for the proportions of Natufian-related ancestry ranging from 0% to 100% by an increment of 1%. The colored rectangle marks a plausible range of Natufian ancestry proportion, estimated by our qpAdm modeling (0.637 ± 2×0.069; Table S8). Thick grey solid lines and dotted grey lines represent ± 1 SE and -3 SE ranges, respectively. In all cases, South, Central and East African groups are more closely related to Taforalt than to any mixture of Yoruba and Natufians, strongly suggesting the presence of related ancestry in Taforalt. SEs were calculated by 5 cM block jackknifing.

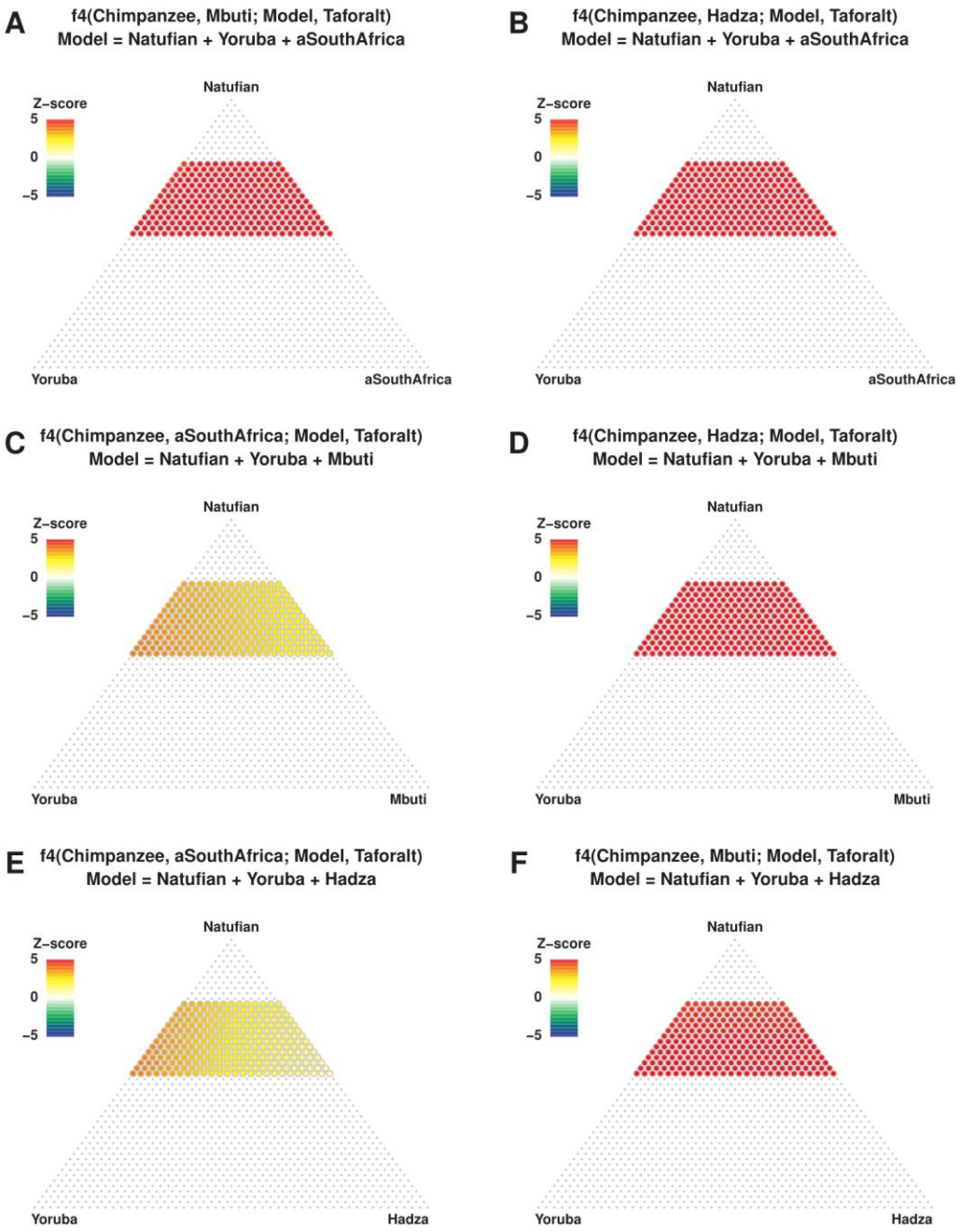


Fig. S21. F_4-based comparison of the Taforalt gene pool with three-way admixture models including Natufian, Yoruba and another sub-Saharan African group as sources. For the second sub-Saharan African source, we tested **(A and B)** aSouthAfrica, **(C and D)** Mbuti and **(E and F)** Hadza. We plot Z-scores within a plausible range of the Natufian-like ancestry proportion (0.499-0.775), covering ± 2 SE range from our qpAdm model (Table S12). Taforalt still shows extra affinity with South, Central or East Africans in comparison to the three-way admixture models including one of them as the third source.

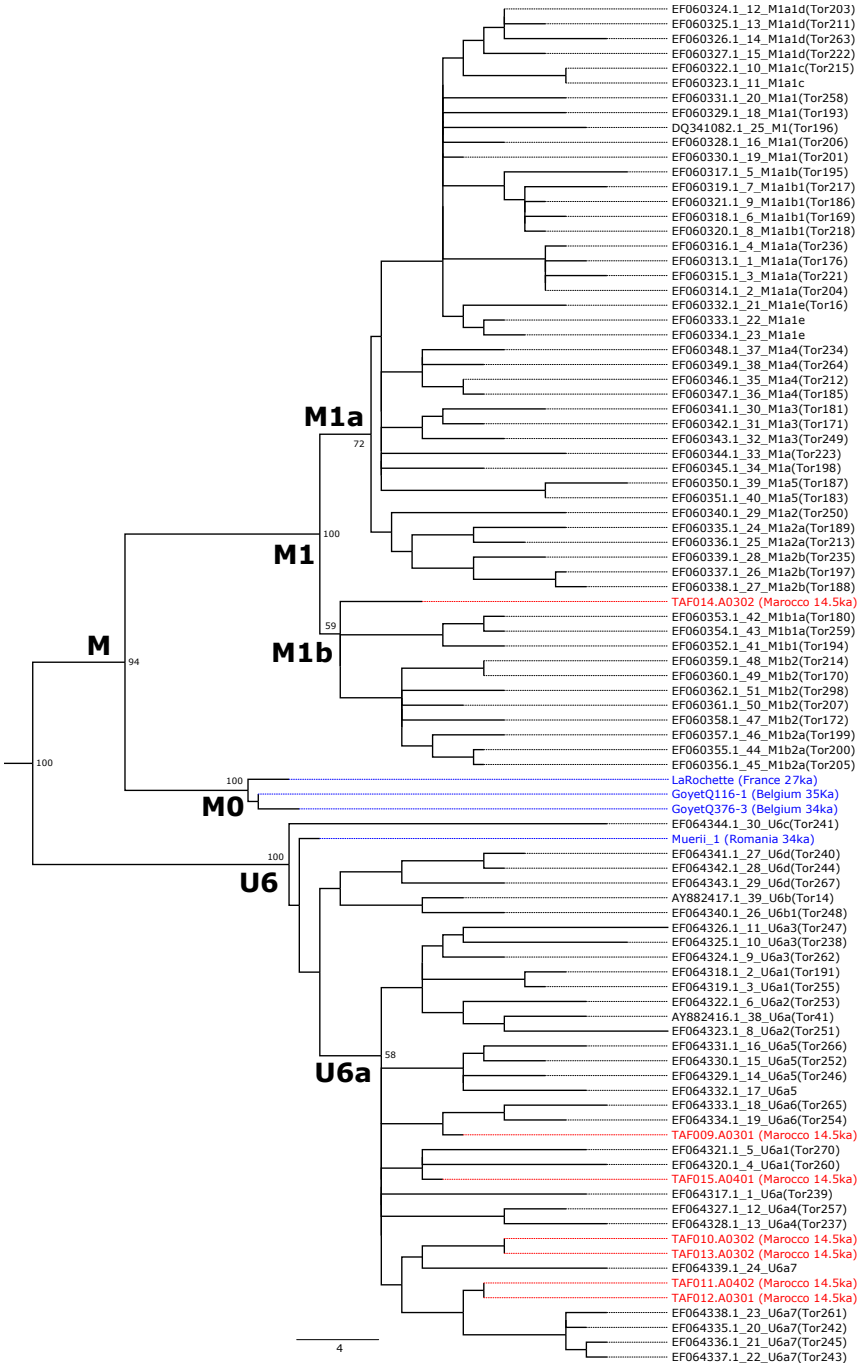


Fig. S22. Maximum Parsimony tree with 98% partial deletion and 1,000 iterations as bootstrap support for 81 modern human and 11 ancient U6 and M mtDNA genomes. Upper Paleolithic European mtDNAs are shown in blue and Taforalt individuals in red with averaged radiocarbon dates in thousands of years BP (ka) reported at each leaf. The major haplogroups are indicated at the branch nodes; M0 refers to the Upper Paleolithic European haplogroup M mtDNA sequences (*23*). The tree is rooted with a haplogroup L3 sequence (not shown). Six Taforalt mitogenomes fall on the U6a branch and one on the M1b branch.

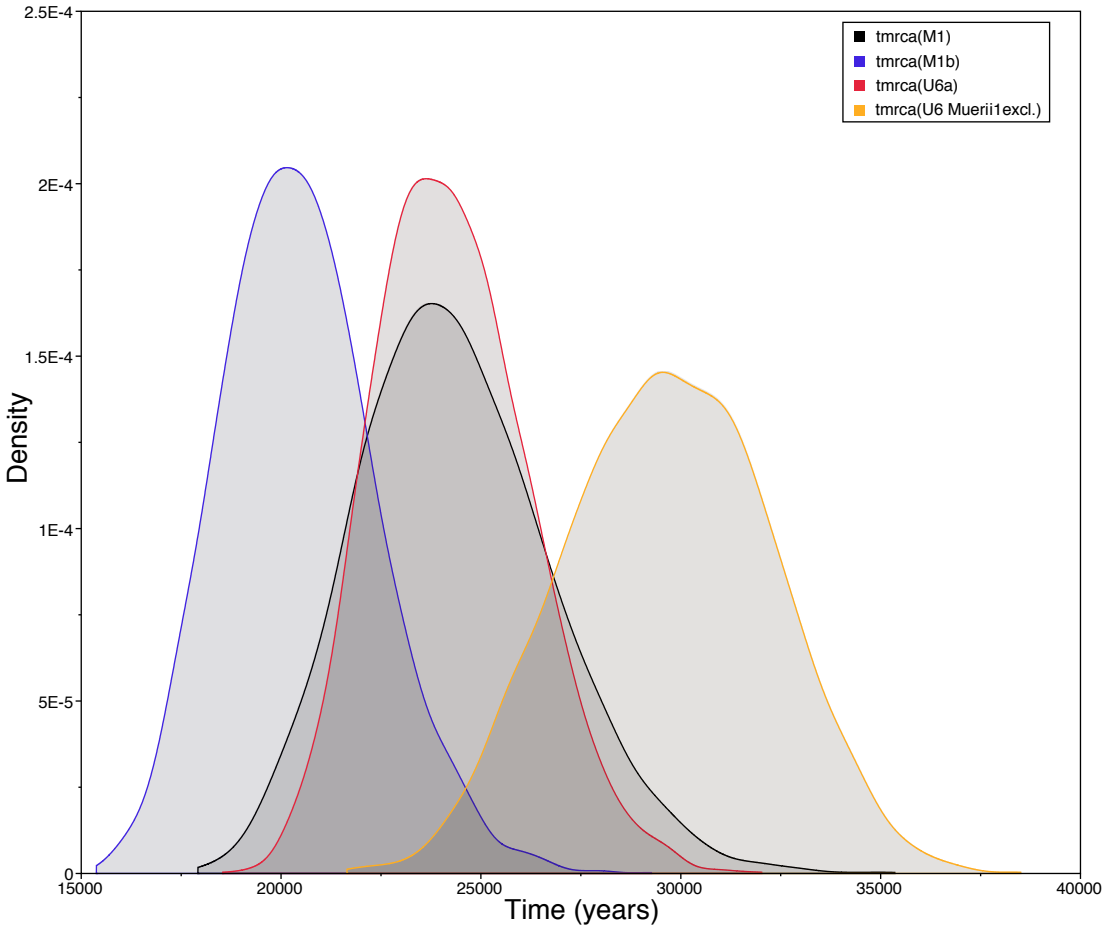
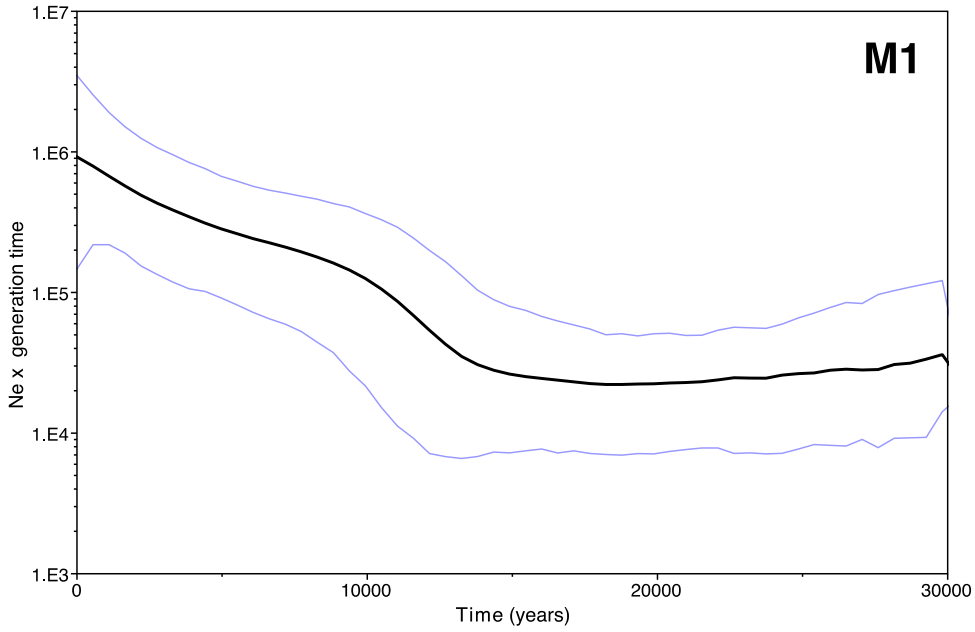


Fig. S23. Marginal probability distribution generated in BEAST for the time to the most recent common ancestor (TMRCA) of mtDNA haplogroups M1, M1b, U6 and U6a. The Upper Paleolithic European Muierii1 sequence is excluded from U6. The divergence time estimates of U6a (red distribution) and M1 (black distribution) largely overlap with a mean date of ~24,000 years ago.

184



B

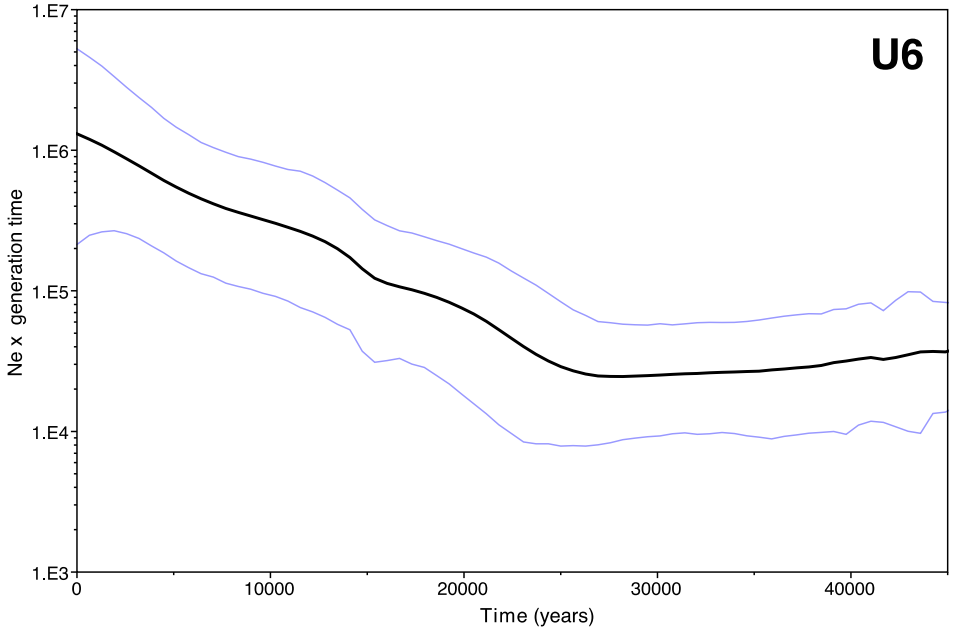


Fig. S24. Bayesian skyline plots for M1 and U6 mtDNA haplogroups to monitor changes in effective population size (N_e x generation time) through time (in years before present) The black line describes the mean value and the purple lines the 95% highest posterior density (HPD) interval. **(A)** Haplogroup M1, based on 52 sequences, shows an increase in effective population size ~14,000 years ago, whereas **(B)** haplogroup U6, based on 37 sequences, shows an increase ~26,000 years ago. In both cases we used a haplogroup L3 sequence as outgroup.

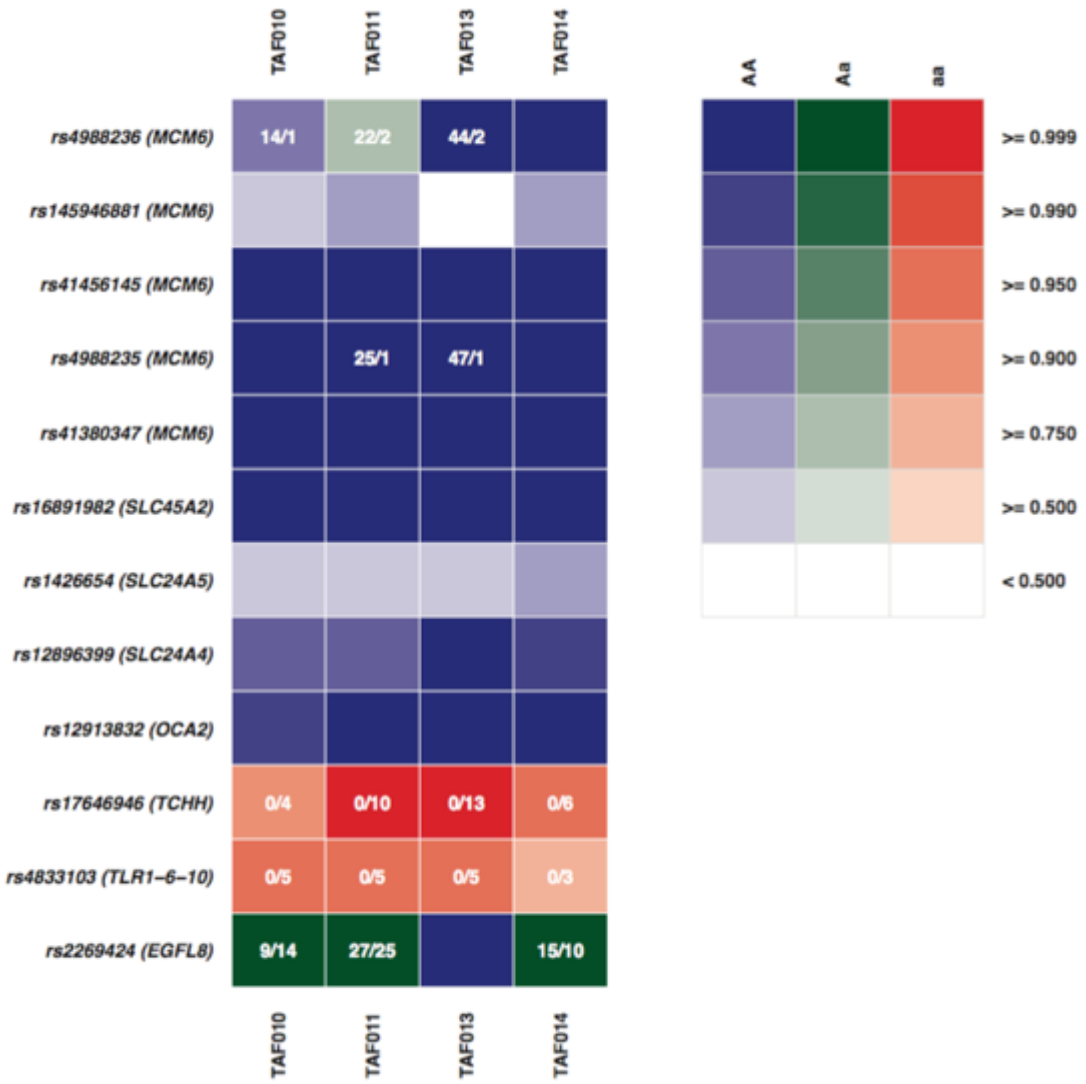


Fig. S25. Estimated phenotypic expression for thoroughly annotated SNPs predictive for eye color, skin pigmentation, lactose tolerance and mycobacteria susceptibility. Colors indicate the most likely genotype, shading indicates genotype likelihood and allele counts are given as an auxiliary text (# Ancestral / # Derived). AA = homozygous ancestral (non-effect), Aa = heterozygous, aa = homozygous derived (effect).

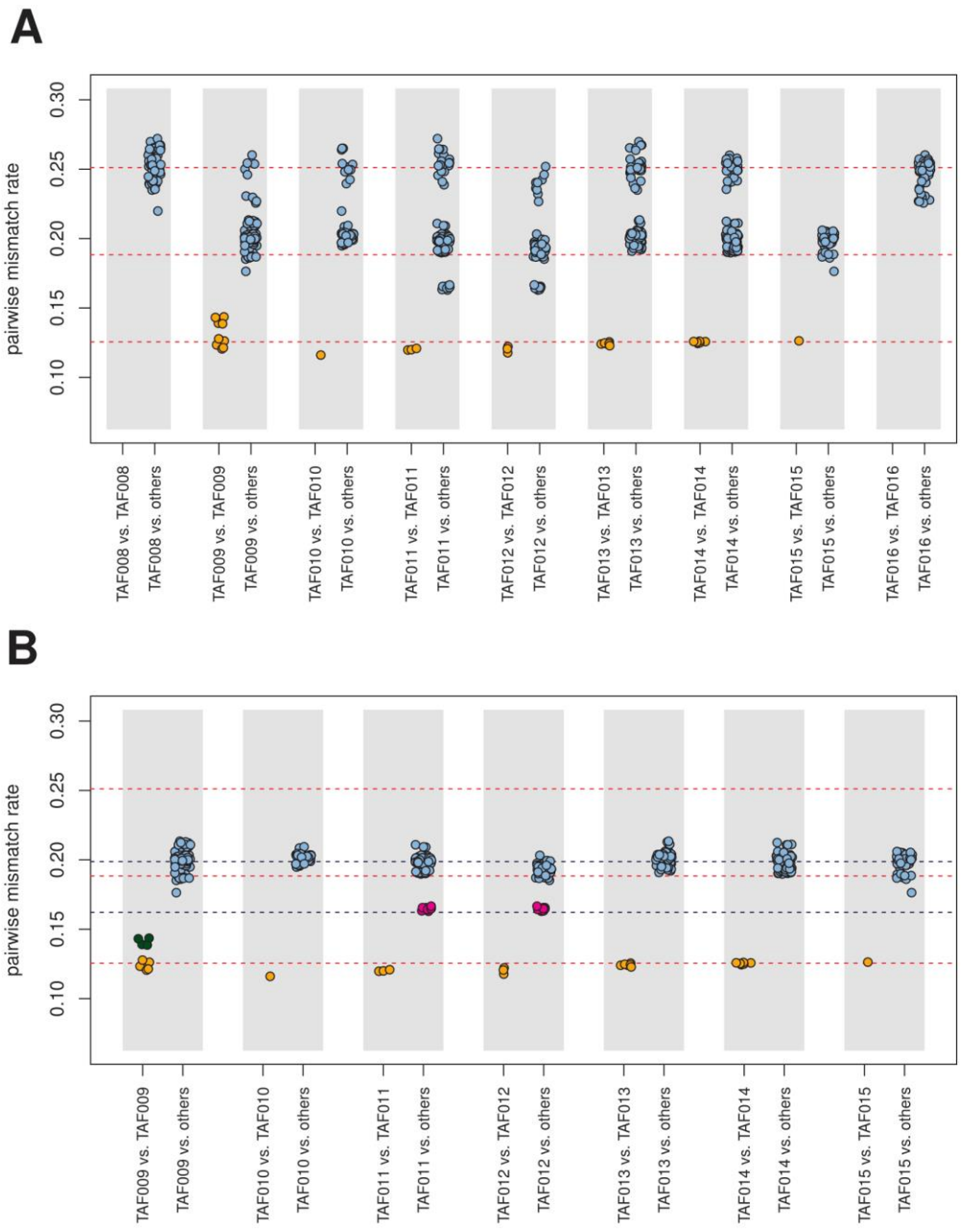


Fig. S26. Pairwise mismatch rate between all pairs of Taforalt sequencing libraries. The red dotted line on the bottom is the mean value of all intra-individual library pairs. **(A)** All nine individuals are present. The top and middle red lines mark 2 and 1.5 times of the bottom value, corresponding to the expected values for the unrelated (r=0) and the first degree relative (r=0.5) pairs, respectively. **(B)** Two individuals with low C>T misincorporation rates are removed. The upper blue dotted line shows the mean value of all inter-individual library pairs except for those between TAF011 and TAF012 (magenta dots). The lower blue line marks the average between the upper blue one and the red bottom line, corresponding to the first-degree relative controlling for the baseline inbreeding due to strong genetic drift. Dark green dots in TAF009 mark the four intra-individual comparisons including the most contaminated library TAF009.B0301.

Table S1. Archaeological information for Taforalt individuals used for dating and ancient genome analyses. All individuals are from Sector 10.

Skeleton ID[1]	RC dating?[2]	^{14}C laboratory ID	Dated specimen	^{14}C determination yBP[3]	Date ^{14}C cal. yBP 95.4%[4]	Genetic ID[5]	Ancient DNA specimen	Incl. mtDNA analyses?	Incl. autosomal DNA analyses?
Individual 1	no	-	-	-	-	TAF008	left petrous	no	no
Individual 6	yes	OxA-23779	humerus	12,255 ± 50	14,805-13,908	TAF009	left/right petrous	yes	no
Individual 7	no	OxA-16663	humerus	12,470 ± 100	15,077-14,132	TAF010	right petrous	yes	yes
Individual 8	no	-	-	-	-	TAF011	left petrous	yes	yes
Individual 9	yes	OxA-23780	clavicle	12,355 ± 50	14,890-14,049	TAF012	right petrous	yes	yes
Individual 11	no	-	-	-	-	TAF013	right petrous	yes	yes
Individual 12	no	-	-	-	-	TAF014	left petrous	yes	yes
Individual 13	no	-	-	-	-	TAF015	left petrous	yes	no
Individual 14	yes	OxA-23781	femur	12,410 ± 55	14,956-14,120	TAF016[6]	petrous	no	no
Individual 4	yes	OxA-23660	vertebra[7]	12,380 ± 55	14,930-14,076	-	-	-	-
Individual 5	yes	OxA-23778	metatarsal	12,265 ± 50	14,824-13,920	-	-	-	-

[1] Archaeological ID following the system of (*52*). [2] For radiocarbon dates, see (*52*). [3] Uncalibrated radiocarbon dates in yBP. [4] Calibrated (IntCal09) radiocarbon dates at 94.5% confidence interval. [5] Individual ID used for genomic analyses in this study. [6] A repeated dating for this individual (OxA-23782) yielded similar results (14C determination: 12,460 ± 55, calibrated date: 15,016-14,163 cal. yBP). [7] Displaced bone from Individual 4.

Table S2. Detailed information on prepared DNA extracts and libraries for the Taforalt individuals analyzed in this study. For each individual we made two to four DNA extracts. For each extract we made one or two single-stranded DNA libraries with unique index pairs using no UDG treatment. All extracts and libraries were created in a single batch for which we included one extraction and library blank, respectively.

Ind. ID	aDNA specimen	Extract ID (bone powder in mg)	Library ID (extract in uL)	P5 index sequence	P7 index sequence
TAF008	left petrous	TAF008.A02 (61)	TAF008.A0203 (30)	TAATGCG	TCATGGT
			TAF008.A0204 (30)	AGGTACC	AGAACCG
		TAF008.A03 (56)	TAF008.A0303 (30)	TGCGTCC	TGGAATA
			TAF008.A0304 (30)	GAATCTC	CAGGAGG
TAF009	left petrous	TAF009.A02 (51)	TAF009.A0201 (30)	GCATTGG	TTCGCAA
			TAF009.A0202 (30)	GATCTCG	AATTCAA
		TAF009.A03 (51)	TAF009.A0301 (30)	CAATATG	CGCGCAG
	right petrous	TAF009.B03 (52)	TAF009.B0301 (30)	CCGATTG	GGTCAAG
		TAF009.B04 (54)	TAF009.B0401 (30)	ATGCCGC	AATGATG
TAF010	right petrous	TAF010.A03 (50)	TAF010.A0301 (30)	AGAACCG	AACCAAG
			TAF010.A0302 (30)	TGGAATA	CGGCGTA
TAF011	left petrous	TAF011.A03 (41)	TAF011.A0301 (30)	AAGGTCT	AAGCTAA
		TAF011.A04 (52)	TAF011.A0401 (30)	ACTGGAC	GACGGCG
			TAF011.A0402 (30)	AGCAGGT	AGAAGAC
TAF012	right petrous	TAF012.A02 (54)	TAF012.A0201 (30)	AATGATG	AGAGCGC
			TAF012.A0202 (30)	AGTCAGA	GCCTACG
		TAF012.A03 (52)	TAF012.A0301 (30)	AACTAGA	TAATCAT
TAF013	right petrous	TAF013.A02 (50)	TAF013.A0201 (30)	AACCAAG	TAGGCCG
			TAF013.A0202 (30)	CGGCGTA	GGCATAG
		TAF013.A03 (51)	TAF013.A0301 (30)	GCAGTCC	TTCAACC
			TAF013.A0302 (30)	CTCGCGC	TTAACTC
TAF014	left petrous	TAF014.A02 (51)	TAF014.A0201 (30)	ATACTGA	AATAAGC
			TAF014.A0202 (30)	TACTTAG	AGCCTTG
		TAF014.A03 (53)	TAF014.A0301 (30)	AGAAGAC	AGAATTA
			TAF014.A0302 (30)	GTCCGGC	CAGCATC
TAF015	left petrous	TAF015.A03 (50)	TAF015.A0301 (30)	AACCTGC	AACGACC
		TAF015.A04 (49)	TAF015.A0401 (30)	GACGATT	CCAGCGG
TAF016	petrous	TAF016.A02 (50)	TAF016.A0201 (30)	TTCAACC	CGTATAT
			TAF016.A0202 (30)	TTAACTC	GCTAATC
		TAF016.A03 (50)	TAF016.A0301 (30)	TAGTCTA	GACTTCT
EXB	neg. extract control	EXB005.A10 (0)	EXB005.A1001 (30)	AGCCTTG	CGCAGCC
LIB	neg. library control		LIB001.A0125 (0)	CAGGAGG	GCAGTCC

Table S3. Details on library quality control for (unmerged and merged) libraries enriched for nuclear DNA, and library selection for autosomal analyses. We excluded the libraries for individuals TAF008, -9, 15, -16 given their much lower deamination rate and SNP coverage.

Library ID	Reads going into alignment	Reads mapped to genome	Reads on target after dup. rem. (MAPQ30)	SNPs on 1240k panel covered at least once	# SNPs hit on 1240k panel	Av. coverage base position (X) [st.err.]	Av. read length (bp)	% deamin. 5'- end	# SNPs hit on HO panel (after filter)	Incl. in autosomal analysis?
TAF008.A0203	1,884,246	1,857,762	5,567	0.3%	3,948	0.00 [0.07]	56	11.0%	2,386	
TAF008.A0204	1,843,325	1,832,817	4,794	0.3%	3,828	0.00 [0.07]	56	11.7%	2,261	
TAF008.A0303	2,956,034	2,942,791	7,390	0.3%	3,828	0.00 [0.07]	56	12.9%	2,194	
TAF008.A0304	2,828,559	2,810,621	8,446	0.4%	5,144	0.00 [0.08]	57	10.5%	3,087	
TAF008 (merged)					**15,998**				**8,856**	**No**
TAF009.A0201	5,575,137	5,328,308	66,662	4.7%	55,990	0.05 [0.25]	44	52.6%	36,272	
TAF009.A0202	4,746,484	4,537,899	64,628	5.1%	60,775	0.06 [0.27]	44	51.7%	39,387	
TAF009.A0301	4,518,613	4,274,225	63,840	5.1%	61,493	0.06 [0.27]	43	53.8%	39,177	
TAF009.B0301	5,245,065	5,029,883	100,024	10.5%	125,857	0.12 [0.40]	45	48.1%	77,960	
TAF009.B0401	4,948,072	4,697,900	66,605	4.9%	58,502	0.06 [0.26]	44	52.7%	38,737	
TAF009 (merged)					**287,074**				**155,794**	**No**
TAF010.A0301	7,035,890	6,723,985	546,296	57.0%	681,924	1.62 [2.42]	49	56.4%	393,136	
TAF010.A0302	7,577,053	7,256,767	540,407	55.0%	658,475	1.44 [2.14]	49	56.3%	390,643	
TAF010 (merged)					**829,806**				**456,526**	**Yes**
TAF011.A0301	6,595,596	6,232,545	454,837	68.6%	820,702	2.64 [3.57]	52	55.3%	467,724	
TAF011.A0401	4,646,720	4,396,667	512,270	68.2%	815,318	2.52 [3.51]	52	56.2%	456,986	
TAF011.A0402	6,569,010	6,244,885	530,960	72.6%	868,077	3.09 [4.14]	52	56.3%	480,052	
TAF011 (merged)					**1,013,869**				**544,232**	**Yes**
TAF012.A0201	5,292,269	5,037,288	81,450	11.6%	138,778	0.14 [0.43]	41	63.1%	83,780	
TAF012.A0202	4,710,604	4,431,869	78,897	12.4%	147,750	0.15 [0.46]	41	63.3%	93,094	
TAF012.A0301	6,353,427	6,047,326	83,324	12.5%	149,784	0.16 [0.46]	42	61.9%	96,304	
TAF012 (merged)					**343,764**				**183,041**	**Yes**
TAF013.A0201	4,846,011	4,483,054	459,613	56.5%	676,182	1.56 [2.30]	52	55.8%	385,874	
TAF013.A0202	4,981,008	4,671,356	446,644	58.5%	700,348	1.71 [2.47]	51	56.2%	412,538	
TAF013.A0301	5,026,349	4,687,304	502,554	60.3%	721,524	1.94 [2.87]	53	54.4%	425,201	
TAF013.A0302	4,709,059	4,412,490	497,887	59.2%	708,483	1.86 [2.80]	53	54.8%	420,176	
TAF013 (merged)					**974,776**				**531,021**	**Yes**
TAF014.A0201	3,458,419	3,221,811	330,137	55.1%	658,954	1.44 [2.12]	49	58.3%	389,766	
TAF014.A0202	4,354,092	4,061,545	336,823	56.8%	679,172	1.52 [2.19]	49	58.2%	399,173	
TAF014.A0301	5,599,630	5,204,847	299,457	53.1%	634,668	1.30 [1.92]	49	59.4%	378,969	
TAF014.A0302	3,767,866	3,513,783	281,545	57.0%	681,804	1.44 [2.01]	50	59.1%	392,970	
TAF014 (merged)					**961,954**				**521,631**	**Yes**
TAF015.A0301	5,061,712	4,749,671	43,162	3.5%	41,753	0.03 [0.22]	43	56.1%	27,820	
TAF015.A0401	5,348,793	5,031,066	43,406	3.9%	46,538	0.04 [0.23]	42	57.2%	28,288	
TAF015 (merged)					**84,128**				**44,820**	**No**
TAF016.A0201	4,243,140	4,213,380	19,691	1.7%	19,979	0.02 [0.17]	61	11.3%	11,581	
TAF016.A0202	4,365,126	4,326,241	20,850	1.7%	20,577	0.02 [0.17]	61	11.8%	12,312	
TAF016.A0301	4,068,627	4,036,241	15,501	1.3%	15,792	0.02 [0.15]	55	13.5%	9,251	
TAF016 (merged)					**54,405**				**28,699**	**No**

Table S4. Details on the library quality control and contamination estimates for libraries enriched in mitochondrial DNA. We excluded the libraries for individuals TAF008 and -16 for haplotyping and further analyses, given their much lower deamination rate and higher contamination estimate. We restricted the analyses for TAF009 to the three libraries with <3% contamination.

Library ID	Fragments going into alignment	Fragments mapped to genome	Fragments on target after dup. rem. (MAPQ30)	% rCRS covered	Av. coverage base position (X) [st.err.][1]	Av. fragment length	% 5'- end deamin.	mtDNA cont. est. [95% CI]	mtDNA haplotype[2]	Included in mtDNA analysis?
TAF008.A0203	6,631,155	3,513,783	5,567	97	19 [18]	55	6.7%	NA	NA	No
TAF008.A0204	6,335,054	1,843,325	4,794	95	16 [13]	54	5.3%	NA	NA	No
TAF008.A0303	8,455,030	2,956,034	7,390	99	24 [18]	53	4.7%	NA	NA	No
TAF008.A0304	7,933,698	2,828,559	8,446	99	27 [20]	54	4.8%	NA	NA	No
TAF009.A0201	9,359,897	5,575,137	66,662	100	161 [75]	40	52.1%	8% [7-9%]	U6a6b	No
TAF009.A0202	7,931,448	4,746,484	64,628	100	156 [70]	40	52.9%	7% [6-8%]	U6a6b	No
TAF009.A0301	7,645,154	4,518,613	63,840	100	150 [72]	39	54.2%	3% [2-4%]	U6a6b	Yes
TAF009.B0301	8,573,031	5,245,065	100,024	100	257 [93]	43	48.4%	3% [2-4%]	U6a6b	Yes
TAF009.B0401	8,548,858	4,948,072	66,605	100	158 [80]	39	54.3%	1% [0-2%]	U6a6b	Yes
TAF010.A0301	11,399,063	7,035,890	546,296	100	1,701 [215]	52	47.8%	1% [0-2%]	U6a7b	Yes
TAF010.A0302	12,258,343	7,577,053	540,407	100	1,681 [212]	52	47.9%	1% [0-2%]	U6a7b	Yes
TAF011.A0301	10,444,374	6,595,596	454,837	100	1,357 [188]	49	53.9%	1% [0-2%]	U6a7b	Yes
TAF011.A0401	7,604,829	4,646,720	512,270	100	1,576 [212]	51	51.7%	1% [0-2%]	U6a7b	Yes
TAF011.A0402	10,574,609	6,569,010	530,960	100	1,636 [201]	51	52.1%	1% [0-2%]	U6a7b	Yes
TAF012.A0201	8,449,174	5,292,269	81,450	100	190 [80]	39	61.2%	1% [0-2%]	U6a7	Yes
TAF012.A0202	7,596,800	4,710,604	78,897	100	184 [81]	39	60.4%	1% [0-2%]	U6a7	Yes
TAF012.A0301	10,185,106	6,353,427	83,324	100	195 [81]	39	59.8%	1% [0-2%]	U6a7	Yes
TAF013.A0201	8,241,256	4,846,011	459,613	100	1,377 [217]	50	54.1%	1% [0-2%]	U6a7b	Yes
TAF013.A0202	8,257,734	4,981,008	446,644	100	1,336 [213]	50	54.8%	1% [0-2%]	U6a7b	Yes
TAF013.A0301	8,324,946	5,026,349	502,554	100	1,530 [222]	50	53.5%	1% [0-2%]	U6a7b	Yes
TAF013.A0302	7,788,083	4,709,059	497,887	100	1,517 [226]	51	54.1%	1% [0-2%]	U6a7b	Yes
TAF014.A0201	5,722,005	3,458,419	330,137	100	944 [190]	47	55.4%	1% [0-2%]	M1b	Yes
TAF014.A0202	7,289,168	4,354,092	336,823	100	962 [190]	47	55.2%	1% [0-2%]	M1b	Yes
TAF014.A0301	9,191,765	5,599,630	299,457	100	845 [180]	47	55.9%	1% [0-2%]	M1b	Yes
TAF014.A0302	6,296,439	3,767,866	281,545	100	797 [176]	47	56.5%	1% [0-2%]	M1b	Yes
TAF015.A0301	8,517,181	5,061,712	43,162	100	102 [54]	39	54.5%	4% [3-5%]	U6a1b	Yes
TAF015.A0401	9,028,717	5,348,793	43,406	100	103 [56]	39	53.7%	3% [2-4%]	U6a1b	Yes
TAF016.A0201	7,053,485	4,243,140	19,691	100	68 [30]	57	8.4%	11% [10-12%]	NA	No
TAF016.A0202	7,342,323	4,365,126	20,850	100	72 [29]	57	8.7%	19% [18-20%]	NA	No
TAF016.A0301	6,799,788	4,068,627	15,501	100	49 [28]	52	14.7%	15% [14-16%]	NA	No

[1] Coverage reported here is calculated for unique reads after quality filtering and before down-sampling. [2] We obtained an automated mitochondrial haplogroup assignment from the consensus sequences generated by Schmutzi using HaploGrep and thereafter visually double-checked the mitochondrial sequences for the expected phylogenetically diagnostic SNP variants.

Table S5. Details on the genetic sex determination and X-chromosomal heterozygosity-based contamination estimates, before and after library merging. We excluded TAF008, TAF009, TAF015 and TAF016 individuals from autosomal population genetic analyses based on their unreliable or high contamination estimates.

Library ID	X rate	Y rate	Gen. sex	Cont. est. [SE][1]	# SNPs X-chrom.
TAF008.A0203	0.542	0.396	Male	NA	1
TAF008.A0204	0.424	0.445	Male	NA	2
TAF008.A0303	0.341	0.806	Male	NA	0
TAF008.A0304	0.694	0.262	Male	NA	3
TAF008 (merged)	**0.514**	**0.460**	**Male**	**NA**	**7**
TAF009.A0201	0.324	0.444	Male	5.2% [0%]	22
TAF009.A0202	0.330	0.456	Male	NA	19
TAF009.A0301	0.356	0.527	Male	6.8% [0%]	16
TAF009.B0301	0.439	0.444	Male	17.3% [0%]	94
TAF009.B0401	0.314	0.501	Male	11.3% [0%]	12
TAF009 (merged)	**0.369**	**0.469**	**Male**	**11.9% [0%]**	**295**
TAF010.A0301	0.285	0.457	Male	1.7% [0%]	2,048
TAF010.A0302	0.306	0.477	Male	1.6% [0%]	1,919
TAF010 (merged)	**0.295**	**0.466**	**Male**	**1.7% [0%]**	**4,307**
TAF011.A0301	0.304	0.501	Male	1.7% [0%]	3,833
TAF011.A0401	0.335	0.580	Male	1.9% [0%]	4,050
TAF011.A0402	0.346	0.588	Male	1.9% [0%]	5,172
TAF011 (merged)	**0.329**	**0.558**	**Male**	**1.8% [0%]**	**10,453**
TAF012.A0201	0.745	0.029	Female[2]	13.9% [0%]	200
TAF012.A0202	0.672	0.024	Female[2]	8.4% [0%]	205
TAF012.A0301	0.630	0.018	Female[2]	17.0% [0%]	222
TAF012 (merged)	**0.680**	**0.024**	**Female[2]**	**18.4% [0%]**	**1,102**
TAF013.A0201	0.294	0.469	Male	1.5% [0%]	1,970
TAF013.A0202	0.297	0.472	Male	2.2% [0%]	2,312
TAF013.A0301	0.285	0.465	Male	2.2% [0%]	2,592
TAF013.A0302	0.281	0.474	Male	2.1% [0%]	2,356
TAF013 (merged)	**0.289**	**0.470**	**Male**	**2.3% [0%]**	**8,403**
TAF014.A0201	0.312	0.490	Male	2.8% [0%]	1,932
TAF014.A0202	0.313	0.477	Male	2.3% [0%]	2,075
TAF014.A0301	0.302	0.472	Male	1.8% [0%]	1,626
TAF014.A0302	0.334	0.511	Male	2.5% [0%]	2,067
TAF014 (merged)	**0.315**	**0.488**	**Male**	**2.5% [0%]**	**7,987**
TAF015.A0301	0.296	0.440	Male	1.4% [NA]	10
TAF015.A0401	0.315	0.463	Male	NA	10
TAF015 (merged)	**0.306**	**0.452**	**Male**	**0.8% [0%]**	**27**
TAF016.A0201	0.437	0.430	Male	NA	16
TAF016.A0202	0.476	0.337	Male	NA	17
TAF016.A0301	0.698	0.273	Male	0.3% [0%]	25
TAF016 (merged)	**0.524**	**0.353**	**Male**	**6.3% [0%]**	**68**

[1] Values presented here are based on New method 1 (*72*). [2] The X-chromosomal heterozygosity rate can not be used to estimate nuclear contamination levels in females.

Table S6. Relative affinity of ancient Near Eastern populations with Taforalt measured by f_4(Chimpanzee, Taforalt; NE_1, NE_2). Positive values suggest that NE_2 shares more alleles with Taforalt than NE_1 does. Natufians share more alleles with Taforalt than any other Near Eastern groups ($Z \geq 2.2$ SE). Levant_N shares more alleles with Taforalt than all the others except Natufians ($Z \geq 2.2$ SE). Z-scores were estimated by dividing f_4 by 5 cM block jackknife SE.

NE_1	NE_2	f_4	Z	# SNPs
Levant_N	Natufian	0.001211	2.186	237,491
Iran_HotuIIIb		0.005924	4.454	46,024
Iran_N		0.006661	9.756	237,423
CHG		0.005492	8.982	266,013
Iran_ChL		0.004110	7.410	258,506
Armenia_ChL		0.003374	6.158	260,192
Anatolia_N		0.002664	5.537	266,035
Anatolia_ChL		0.002033	2.839	196,351
Europe_EN		0.002596	5.321	266,039
Natufian	Levant_N	-0.001211	-2.186	237,491
Iran_HotuIIIb		0.004182	4.239	70,082
Iran_N		0.005572	10.982	383,153
CHG		0.004180	9.268	440,356
Iran_ChL		0.003421	9.001	425,444
Armenia_ChL		0.002386	6.107	429,057
Anatolia_N		0.001583	5.076	440,387
Anatolia_ChL		0.001164	2.208	309,065
Europe_EN		0.001557	5.210	440,408

Table S7. Direct assessment of Neanderthal ancestry, using Yoruba as a baseline, in f_4-symmetry statistics in the form *f4*(Chimp, Altai; Yoruba, *Test*). Taforalt shares less alleles with Altai Neanderthals than early Holocene Levantines (Natufians, Levant_N) but more than early Neolithic individuals from Iran (Iran_N).

Test	f_4	Z	# SNPs
Ust_Ishim	0.001865	5.843	578,682
MA1	0.001857	5.197	416,496
WHG	0.001806	7.263	579,801
Europe_EN	0.001418	7.205	579,881
Iran_LN	0.001363	3.510	268,041
Anatolia_N	0.001308	6.513	579,325
Iranian	0.001307	7.157	579,888
Mozabite	0.001280	8.262	579,888
CHG	0.001158	4.347	579,743
Ju_hoan_North	0.001147	7.867	579,888
Egyptian	0.001137	7.344	579,888
Natufian	0.001132	2.922	267,482
Algerian	0.001087	6.707	579,888
aSouthAfrica	0.000999	3.294	515,991
Saharawi	0.000945	5.839	579,888
Levant_N	0.000880	3.082	443,349
Taforalt	0.000737	3.116	568,026
Iran_N	0.000628	1.934	453,548
Hadza	0.000363	2.955	579,888
Mbuti	0.000255	1.868	579,888
Mende	0.000113	1.434	579,888
Mota	0.000106	0.405	579,733

Table S8. Two-way admixture modeling of Taforalt using qpAdm, with Natufian and a sub-Saharan African group as sources. By not including any African group nor deeper-branching archaic hominin as an outgroup, we intend to keep sub-Saharan African lineages indistinguishable. As expected, all sub-Saharan African sources provide comparable results, which should be interpreted as a lumped sum sub-Saharan African ancestry proportion in Taforalt. SEs are estimated by 5 cM block jackknifing.

African source	P_{source}[1]	*P*-value[2]	$Coef_{AFR}$[3]	$Coef_{Nat}$[4]	SE
aSouthAfrica	1.77×10^{-20}	0.159	0.352	0.648	0.08
Ju_hoan_North	1.25×10^{-25}	0.207	0.381	0.619	0.075
Mende	4.44×10^{-31}	0.211	0.360	0.640	0.070
Yoruba	7.64×10^{-31}	0.217	0.363	0.637	0.069
Mbuti	8.25×10^{-28}	0.207	0.360	0.640	0.071
Biaka	4.23×10^{-29}	0.202	0.359	0.641	0.069
Dinka	5.93×10^{-28}	0.224	0.368	0.632	0.071
Mota	3.81×10^{-17}	0.128	0.362	0.638	0.080
Hadza	2.80×10^{-27}	0.286	0.379	0.621	0.072

[1] QpWave *P*-value for the distinctness of two sources (Natufian and sub-Saharan African). [2] QpAdm *P*-value for the sufficiency of the two-way admixture model for Taforalt. [3] The estimated contribution from the African source. [4] The estimated contribution from the Natufian-related gene pool.

Table S9. Three-way admixture modeling of Taforalt using qpAdm, with Natufian, a sub-Saharan African (AFR) and a pre-Neolithic European (EUR) group as sources. Three-way models are compared with the nested two-way models using a Chi-square test. SEs are estimated by 5 cM block jackknifing.

Source						
AFR	**EUR**	***P*[1]**	***P*$_{comp}$[2]**	**Coef$_{AFR}$[3]**	**Coef$_{Nat}$[4]**	**Coef$_{EUR}$[5]**
aSouthAfrica	Vestonice	0.119	0.110	0.394 ± 0.101	0.475 ± 0.173	0.132 ± 0.088
Ju_hoan_North		0.131	0.128	0.415 ± 0.090	0.469 ± 0.151	0.116 ± 0.079
Mende		0.308	0.056	0.411 ± 0.074	0.447 ± 0.127	0.142 ± 0.071
Yoruba		0.321	0.058	0.416 ± 0.074	0.445 ± 0.126	0.139 ± 0.070
Mbuti		0.182	0.089	0.402 ± 0.081	0.467 ± 0.140	0.131 ± 0.077
Biaka		0.213	0.088	0.399 ± 0.076	0.475 ± 0.130	0.126 ± 0.072
Dinka		0.506	0.030	0.443 ± 0.073	0.397 ± 0.123	0.160 ± 0.068
Mota		0.154	0.111	0.405 ± 0.094	0.466 ± 0.161	0.128 ± 0.085
Hadza		0.422	0.060	0.434 ± 0.077	0.430 ± 0.127	0.135 ± 0.069
aSouthAfrica	El Miron	0.462	0.073	0.363 ± 0.074	0.461 ± 0.132	0.176 ± 0.095
Ju_hoan_North		0.537	0.077	0.385 ± 0.069	0.455 ± 0.118	0.160 ± 0.087
Mende		0.738	0.054	0.375 ± 0.063	0.456 ± 0.108	0.170 ± 0.082
Yoruba		0.775	0.052	0.383 ± 0.064	0.446 ± 0.109	0.170 ± 0.081
Mbuti		0.609	0.054	0.369 ± 0.065	0.456 ± 0.114	0.175 ± 0.085
Biaka		0.629	0.063	0.372 ± 0.064	0.462 ± 0.112	0.166 ± 0.084
Dinka		0.919	0.034	0.394 ± 0.064	0.421 ± 0.108	0.185 ± 0.079
Mota		0.433	0.048	0.375 ± 0.072	0.433 ± 0.132	0.193 ± 0.095
Hadza		0.801	0.070	0.395 ± 0.067	0.446 ± 0.112	0.158 ± 0.081
aSouthAfrica	Villabruna	0.631	0.021	0.431 ± 0.071	0.330 ± 0.139	0.239 ± 0.096
Ju_hoan_North		0.596	0.036	0.446 ± 0.069	0.349 ± 0.134	0.205 ± 0.093
Mende		0.660	0.030	0.424 ± 0.063	0.367 ± 0.126	0.209 ± 0.090
Yoruba		0.636	0.034	0.427 ± 0.064	0.370 ± 0.127	0.203 ± 0.090
Mbuti		0.665	0.029	0.426 ± 0.064	0.360 ± 0.128	0.214 ± 0.091
Biaka		0.610	0.033	0.423 ± 0.064	0.371 ± 0.128	0.206 ± 0.091
Dinka		0.649	0.034	0.432 ± 0.065	0.366 ± 0.129	0.202 ± 0.090
Mota		0.486	0.024	0.444 ± 0.075	0.319 ± 0.153	0.237 ± 0.104
Hadza		0.594	0.064	0.431 ± 0.068	0.394 ± 0.132	0.175 ± 0.092
aSouthAfrica	WHG	0.665	0.019	0.427 ± 0.071	0.346 ± 0.133	0.227 ± 0.091
Ju_hoan_North		0.585	0.037	0.442 ± 0.069	0.366 ± 0.130	0.192 ± 0.089
Mende		0.597	0.037	0.418 ± 0.064	0.393 ± 0.124	0.189 ± 0.087
Yoruba		0.569	0.042	0.420 ± 0.065	0.398 ± 0.125	0.182 ± 0.087
Mbuti		0.639	0.031	0.421 ± 0.065	0.381 ± 0.125	0.197 ± 0.087
Biaka		0.573	0.037	0.418 ± 0.064	0.393 ± 0.125	0.189 ± 0.087
Dinka		0.555	0.047	0.423 ± 0.067	0.399 ± 0.127	0.178 ± 0.087
Mota		0.460	0.027	0.438 ± 0.075	0.345 ± 0.144	0.217 ± 0.097
Hadza		0.530	0.081	0.423 ± 0.068	0.421 ± 0.129	0.156 ± 0.088

[1] QpAdm *P*-value for the sufficiency of the three-way admixture model for Taforalt. [2] χ^2 *P*-value for the comparison of the nesting three-way model to the nested two-way one. [3] The estimated contribution from the African source ± SE. [4] The estimated contribution from the Natufian-related gene pool ± SE. [5] The estimated contribution from the pre-Neolithic European-related gene pool ± SE.

Table S10. A list of *F*-statistics with sizeable deviation from the expected values under the best qpGraph model of Taforalt. Here, we provide a list of eleven *F*-statistics with ≥ 2.5 SE deviant from the expected value. None of the statistics suggests an unmodeled extra affinity between Taforalt and WHG. All *F*-statistics are written in the form of f_4-statistic. SEs are estimated by 5 cM block jackknifing.

Pop1	Pop2	Pop3	Pop4	Expected f_4	Observed f_4	Difference f_4	SE	Z_{diff}
Yoruba	Ami	Iran_N	Natufian	-0.002982	-0.008546	-0.005564	0.001979	-2.812
Ami	WHG	Natufian	Taforalt	-0.008035	-0.013375	-0.005340	0.001938	-2.756
Mbuti	Ami	Yoruba	Taforalt	0.072459	0.076651	0.004192	0.001617	2.592
Yoruba	Ami	Natufian	Taforalt	-0.032664	-0.027802	0.004861	0.001755	2.770
Mbuti	Ami	Natufian	Taforalt	-0.031066	-0.025147	0.005919	0.002103	2.815
Mbuti	Ami	EHG	Taforalt	-0.052499	-0.046786	0.005713	0.002016	2.834
Mbuti	Yoruba	Iran_N	Taforalt	0.001597	0.004393	0.002796	0.000981	2.851
Ami	WHG	Iran_N	Natufian	0.010015	0.016774	0.006759	0.002230	3.030
Mbuti	Yoruba	Yoruba	Taforalt	-0.003726	-0.001155	0.002571	0.000805	3.193
Mbuti	Yoruba	Ami	Taforalt	0.001597	0.004197	0.002599	0.000797	3.261
Mbuti	Yoruba	EHG	Taforalt	0.001597	0.005105	0.003508	0.000941	3.729

Table S11. Relative genetic affinity of representative sub-Saharan African groups to West Africans (Yoruba or Mende) and Natufian, measured by f_4(Chimpanzee, African; Yoruba/Mende, Natufian). Mbuti and the ancient South African are symmetrically related to West Africans and Natufians ($|Z| \leq 1.564$), while ancient and present-day East Africans (Mota, Dinka and Hadza) are more closely related to Natufians than to West Africans ($Z \geq 5.898$). Z-score was calculated by dividing f_4 statistic by 5 cM block jackknife SE.

	Yoruba		Mende	
African	f_4	**Z**	f_4	**Z**
aSouthAfrica	0.000221	0.495	0.000504	1.110
Ju_hoan_North	0.001356	3.678	0.001784	4.701
Mende	-0.001538	-4.041		
Yoruba			-0.000154	-0.395
Mbuti	-0.000091	-0.248	0.000595	1.564
Biaka	-0.000735	-2.019	0.000149	0.393
Dinka	0.002349	5.898	0.003621	8.620
Mota	0.006009	11.227	0.006986	12.549
Hadza	0.00676	15.549	0.007762	17.328

Table S12. Three-way admixture modeling of Taforalt using qpAdm, with Natufian, Yoruba and another sub-Saharan African as the sources. We tried to achieve a minimum-level resolution in sub-Saharan African ancestry by adding an archaic hominin Denisovan in our outgroup set. Although Taforalt can numerically be modeled as a linear combination of the three sources ($P \geq 0.102$), the estimated ancestry proportions are out of range, so the models are not considered sensible. In the case of Mende, Mota and Hadza, the provided outgroups were not enough to distinguish all three sources (qpWave-$P \geq 0.154$). SEs are estimated by 5 cM block jackknifing.

2nd African	P_{source}[1]	*P*-value[2]	$Coef_{Yoruba}$[3]	$Coef_{AFR}$[4]	$Coef_{Nat}$[5]
aSouthAfrica	2.80×10^{-13}	0.165	-0.205 (0.121)	0.537 (0.191)	0.668 (0.071)
Ju_hoan_North	1.06×10^{-20}	0.186	-0.275 (0.137)	0.616 (0.200)	0.659 (0.065)
Mende	0.474	0.489	6.510 (8.644)	-5.990 (8.230)	0.481 (0.415)
Mbuti	6.77×10^{-14}	0.199	0.803 (0.300)	-0.440 (0.225)	0.637 (0.077)
Biaka	4.93×10^{-15}	0.210	0.914 (0.352)	-0.549 (0.277)	0.635 (0.077)
Dinka	4.80×10^{-5}	0.174	-0.849 (0.747)	1.219 (0.855)	0.630 (0.109)
Mota	0.154	0.102	1.450 (1.449)	-1.324 (1.552)	0.874 (0.108)
Hadza	0.405	0.504	-3.891 (4.183)	4.449 (4.523)	0.442 (0.344)

[1] QpWave *P*-value for the distinctness of three sources (Natufian, Yoruba and the second sub-Saharan African). [2] QpAdm *P*-value for the sufficiency of the three-way admixture model for Taforalt. [3] The estimated contribution from the Yoruba-related gene pool. [4] The estimated contribution from the second African-related source. [5] The estimated contribution from the Natufian-related gene pool.

Table S13. Details on the ancient mitochondrial sequences and their dates that were used as time anchors in BEAST.

Individual ID	mtDNA haplogroup	Average date (yBP)	Date interval high (yBP)	Date interval low (yBP)	Dates from	Sequence from
TAF009	U6a6b	14,360	14,810	13,910	(*52*)[1]	this study
TAF010	U6a7b	14,605	15,080	14,130	(*52*)[1]	this study
TAF011	U6a7	14,480	14,930	14,030	this study[2]	this study
TAF012	U6a7	14,470	14,890	14,050	(*52*)[1]	this study
TAF013	U6a7b	14,480	14,930	14,030	this study[2]	this study
TAF014	M1b	14,480	14,930	14,030	this study[2]	this study
TAF015	U6a1b	14,480	14,930	14,030	this study[2]	this study
Muierii1	U6	34,035	34,360	33,710	(*113*)[1]	(*22*)
GoyetQ116-1	M	34,795	35,160	34,430	(*23*)[1]	(*23*)
GoyetQ376-3	M	33,540	33,940	33,140	(*23*)[1]	(*23*)
LaRochette	M	27,592	27,784	27,400	(*114*)[1]	(*23*)

[1] Date was obtained by direct dating, see (*23, 52, 113, 114*). [2] Date is based on an inferred average from direct radiocarbon dates for six Taforalt individuals (published in (*52*)).

Table S14. Likelihood results of four different BEAST models for coalescent tree priors (Constant size/Bayesian skyline) and rate variation among tree branches (strict clock/relaxed clock). A marginal likelihood estimation (MLE) was implemented to assess the best-supported model using path and stepping-stone sampling (*90*). The Bayesian skyline tree prior with strict clock provided higher likelihoods than the other three models.

Tree Prior	Clock	Log Marginal Likelihood	
		Stepping-Stone Sampling	Path Sampling
Constant	Strict	-26839.52	-26839.60
Constant	Relaxed	-26840.04	-26840.15
Skyline	Strict	-26801.28	-26800.98
Skyline	Relaxed	-26803.14	-26803.03

Table S15. Estimated divergence dates in BEAST for monophylethic M and U6 haplogroups and sub-haplogroups. HPD: highest posterior density. yBP: years before present.

	Estimated coalescence time (yBP)		
Haplogroup	**Mean**	**Median**	**95% HPD interval**
M	48,146	47,959	[41,040 - 55,711]
M1	24,328	24,157	[19,895 - 29,203]
M1b	20,446	20,350	[16,905 - 24,219]
U6	37,077	36,692	[34,233 - 40,597]
U6 (Muierii1 incl.)	35,719	35,454	[33,970 - 38,195]
U6 (Muierii1 excl.)	29,685	29,697	[25,019 - 34,667]
U6a	24,258	24,150	[20,850 - 28,028]
U6a1	18,659	18,548	[15,334 - 21,949]
U6a6	16,497	16,258	[14,306 - 19,120]
U6a7	21,426	21,312	[18,292 - 24,652]

Table S16. Y-haplogroup assignment for six Taforalt males. All individuals could be assigned to haplogroup E1b1b, and five of them more specifically to E1b1b1a1 (M-78).

Marker	Associated hg	SNP position hg19	Mutation (Anc.>Der.)	TAF009 (E1b1b1a1b1)	TAF010 (E1b1b1a1)	TAF011 (E1b1b1a1)	TAF013 (E1b1b1a1)	TAF014 (E1b1b1a1)	TAF015 (E1b1b)
P177	E1b	14159846	C>T				X	X	
P179	E1b1	14060308	A>C	X	X	X	X	X	X
P2	E1b1	21610831	G>A		X	X	X	X	
P178	E1b1	7401836	G>A	X	X	X	X	X	X
P180	E1b1	18601274	G>A		X	X	X		
P181	E1b1	17394111	C>G		X		X	X	
M215	E1b1b	15467824	A>G		X	X	X	X	
M243	E1b1b	15019092	T>C		X	X		X	
M5017	E1b1b	2827409	C>T				X		
M5038	E1b1b	17803995	T>C			X	X	X	
M5039	E1b1b	18637397	C>G			X	X	X	
M5048	E1b1b	22700429	G>A			X		X	
M5049	E1b1b	23021729	G>A		X	X	X	X	
M5082	E1b1b	7905833	C>T	X			X	X	
M5083	E1b1b	7906010	A>G		X	X	X	X	
M5084	E1b1b	8028896	C>T			X			
M5101	E1b1b	8692771	C>T		X	X	X	X	
M5183	E1b1b	16346117	C>T			X	X		
M5234	E1b1b	17612676	G>A	X	X	X	X	X	
M5251.1	E1b1b	18045601	C>T		X	X	X	X	
M5255	E1b1b	18093142	A>G			X	X		
M5282	E1b1b	19503700	T>C		X	X	X	X	X
M5305	E1b1b	21658631	G>C			X	X	X	
M5314	E1b1b	21909401	A>G		X	X	X		
M5316	E1b1b	21983698	G>A			X	X		
M5318	E1b1b	22001099	T>C				X		X
L336	E1b1b	21903853	G>A		X	X	X	X	
M5026	E1b1b1	14221285	G>T			X			
M5041	E1b1b1	21491115	A>G		X	X	X	X	
M5047	E1b1b1	21977569	C>T			X	X	X	
M5059	E1b1b1	7071937	G>T			X	X	X	
M5078	E1b1b1	7721674	G>A		X	X	X	X	
M5088	E1b1b1	8143135	A>C			X			
M5100	E1b1b1	8613281	G>A			X	X	X	
M5108	E1b1b1	8880108	G>A			X	X	X	
M5134	E1b1b1	14320934	C>A		X	X	X	X	

(Continue on the next page)

Table S16. *(Continued from the previous page)*

Marker	Associated hg	SNP position hg19	Mutation (Anc.>Der.)	TAF009 (E1b1b1a1b1)	TAF010 (E1b1b1a1)	TAF011 (E1b1b1a1)	TAF013 (E1b1b1a1)	TAF014 (E1b1b1a1)	TAF015 (E1b1b)
M5137	E1b1b1	14393170	A>C			X	X		
M5156	E1b1b1	15089380	A>G		X	X	X	X	
M5166	E1b1b1	15489314	T>C			X	X	X	
M5174	E1b1b1	15869722	G>A		X		X	X	
M5200	E1b1b1	16808859	A>G		X	X	X	X	
M5213	E1b1b1	17119601	G>T				X		
M5226	E1b1b1	17325559	G>T		X	X	X	X	
M5247	E1b1b1	17846754	G>A					X	
M5258	E1b1b1	18573064	A>G			X	X		
M5274	E1b1b1	19316389	A>T	X	X	X	X	X	
M5287	E1b1b1	21131760	T>C			X		X	
M5294	E1b1b1	21358197	T>C		X	X	X	X	
M5321	E1b1b1	22168598	G>A	X		X			
M5322	E1b1b1	22181731	G>A		X	X	X	X	
M5325	E1b1b1	22588254	T>G	X	X	X	X	X	
M5360	E1b1b1	23618826	C>T		X	X	X	X	
L117	E1b1b1	15026633	C>T			X	X	X	
L931	E1b1b1	17138251	A>G	X	X	X	X	X	
CTS9956	E1b1b1	19170454	C>T			X	X		
PF1575	E1b1b1	9389773	T>G	X	X	X	X	X	
PF1755	E1b1b1	18384838	G>A			X	X	X	
L546	E1b1b1a	17516070	C>T	X	X	X	X	X	
CTS2661	E1b1b1a	14410669	C>T		X		X	X	
CTS2270	E1b1b1a	14242950	G>A			X			
CTS4208	E1b1b1a	15479586	T>C		X	X	X	X	
CTS7924	E1b1b1a	17744738	A>G			X			
CTS10323	E1b1b1a	19396726	T>C		X	X	X	X	
PF2108	E1b1b1a	7804308	C>T		X	X	X	X	
PF2114	E1b1b1a	8232450	C>A			X	X	X	
PF2115	E1b1b1a	8361073	G>T		X		X	X	
PF2173	E1b1b1a	21036413	C>T		X	X			
PF2178	E1b1b1a	21583211	C>A			X		X	
PF2185	E1b1b1a	21860060	G>C		X	X			
PF2188	E1b1b1a	22080316	G>A		X	X	X		
M78	E1b1b1a1	21893303	C>T		X	X	X	X	

References and Notes

1. I. S. Castañeda, S. Mulitza, E. Schefuss, R. A. Lopes dos Santos, J. S. Sinninghe Damsté, S. Schouten, Wet phases in the Sahara/Sahel region and human migration patterns in North Africa. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20159–20163 (2009). doi:10.1073/pnas.0905771106 Medline

2. B. M. Henn, C. R. Gignoux, M. Jobin, J. M. Granka, J. M. Macpherson, J. M. Kidd, L. Rodríguez-Botigué, S. Ramachandran, L. Hon, A. Brisbin, A. A. Lin, P. A. Underhill, D. Comas, K. K. Kidd, P. J. Norman, P. Parham, C. D. Bustamante, J. L. Mountain, M. W. Feldman, Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5154–5162 (2011). doi:10.1073/pnas.1017511108 Medline

3. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y. S. Song, G. Poletti, F. Balloux, G. van Driem, P. de Knijff, I. G. Romero, A. R. Jha, D. M. Behar, C. M. Bravi, C. Capelli, T. Hervig, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O. Balanovsky, S. Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y. Xue, M. S. Abdullah, A. Ruiz-Linares, C. M. Beall, A. Di Rienzo, C. Jeong, E. B. Starikovskaya, E. Metspalu, J. Parik, R. Villems, B. M. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G. Stamatoyannopoulos, J. T. S. Wee, R. Khusainova, E. Khusnutdinova, S. Litvinov, G. Ayodo, D. Comas, M. F. Hammer, T. Kivisild, W. Klitz, C. A. Winkler, D. Labuda, M. Bamshad, L. B. Jorde, S. A. Tishkoff, W. S. Watkins, M. Metspalu, S. Dryomov, R. Sukernik, L. Singh, K. Thangaraj, S. Pääbo, J. Kelso, N. Patterson, D. Reich, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016). doi:10.1038/nature18964 Medline

4. P. Skoglund, J. C. Thompson, M. E. Prendergast, A. Mittnik, K. Sirak, M. Hajdinjak, T. Salie, N. Rohland, S. Mallick, A. Peltzer, A. Heinze, I. Olalde, M. Ferry, E. Harney, M. Michel, K. Stewardson, J. I. Cerezo-Román, C. Chiumia, A. Crowther, E. Gomani-Chindebvu, A. O. Gidna, K. M. Grillo, I. T. Helenius, G. Hellenthal, R. Helm, M. Horton, S. López, A. Z. P. Mabulla, J. Parkington, C. Shipton, M. G. Thomas, R. Tibesasa, M. Welling, V. M. Hayes, D. J. Kennett, R. Ramesar, M. Meyer, S. Pääbo, N. Patterson, A. G. Morris, N. Boivin, R. Pinhasi, J. Krause, D. Reich, Reconstructing prehistoric African population structure. *Cell* **171**, 59–71.e21 (2017). doi:10.1016/j.cell.2017.08.049 Medline

5. V. J. Schuenemann, A. Peltzer, B. Welte, W. P. van Pelt, M. Molak, C.-C. Wang, A. Furtwängler, C. Urban, E. Reiter, K. Nieselt, B. Teßmann, M. Francken, K. Harvati, W. Haak, S. Schiffels, J. Krause, Ancient Egyptian mummy genomes suggest an increase of sub-Saharan African ancestry in post-Roman periods. *Nat. Commun.* **8**, 15694–15704 (2017). doi:10.1038/ncomms15694 Medline

6. A. M. González, J. M. Larruga, K. K. Abu-Amero, Y. Shi, J. Pestano, V. M. Cabrera, Mitochondrial lineage M1 traces an early human backflow to Africa. *BMC Genomics* **8**, 223–234 (2007). doi:10.1186/1471-2164-8-223 Medline

7. A. Olivieri, A. Achilli, M. Pala, V. Battaglia, S. Fornarino, N. Al-Zahery, R. Scozzari, F. Cruciani, D. M. Behar, J.-M. Dugoujon, C. Coudray, A. S. Santachiara-Benerecetti, O. Semino, H.-J. Bandelt, A. Torroni, The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science* **314**, 1767–1770 (2006). doi:10.1126/science.1135566 Medline

8. M. Gallego Llorente, E. R. Jones, A. Eriksson, V. Siska, K. W. Arthur, J. W. Arthur, M. C. Curtis, J. T. Stock, M. Coltorti, P. Pieruccini, S. Stretton, F. Brock, T. Higham, Y. Park, M. Hofreiter, D. G. Bradley, J. Bhak, R. Pinhasi, A. Manica, Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* **350**, 820–822 (2015). doi:10.1126/science.aad2879 Medline

9. C. M. Schlebusch, H. Malmström, T. Günther, P. Sjödin, A. Coutinho, H. Edlund, A. R. Munters, M. Vicente, M. Steyn, H. Soodyall, M. Lombard, M. Jakobsson, Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science* **358**, 652–655 (2017). doi:10.1126/science.aao6266 Medline

10. R. N. Barton, A. Bouzouggar, J. T. Hogue, S. Lee, S. N. Collcutt, P. Ditchfield, Origins of the Iberomaurusian in NW Africa: New AMS radiocarbon dating of the Middle and Later Stone Age deposits at Taforalt Cave, Morocco. *J. Hum. Evol.* **65**, 266–281 (2013). doi:10.1016/j.jhevol.2013.06.003 Medline

11. See supplementary materials.

12. M.-T. Gansauge, T. Gerber, I. Glocke, P. Korlevic, L. Lippik, S. Nagel, L. M. Riehl, A. Schmidt, M. Meyer, Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, e79 (2017). Medline

13. Q. Fu, M. Meyer, X. Gao, U. Stenzel, H. A. Burbano, J. Kelso, S. Pääbo, DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2223–2227 (2013). doi:10.1073/pnas.1221359110 Medline

14. Q. Fu, M. Hajdinjak, O. T. Moldovan, S. Constantin, S. Mallick, P. Skoglund, N. Patterson, N. Rohland, I. Lazaridis, B. Nickel, B. Viola, K. Prüfer, M. Meyer, J. Kelso, D. Reich, S. Pääbo, An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015). doi:10.1038/nature14558 Medline

15. E. R. Jones, G. Gonzalez-Fortes, S. Connell, V. Siska, A. Eriksson, R. Martiniano, R. L. McLaughlin, M. Gallego Llorente, L. M. Cassidy, C. Gamba, T. Meshveliani, O. Bar-Yosef, W. Müller, A. Belfer-Cohen, Z. Matskevich, N. Jakeli, T. F. G. Higham, M. Currat, D. Lordkipanidze, M. Hofreiter, A. Manica, R. Pinhasi, D. G. Bradley, Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912–8919 (2015). doi:10.1038/ncomms9912 Medline

16. I. Lazaridis, D. Nadel, G. Rollefson, D. C. Merrett, N. Rohland, S. Mallick, D. Fernandes, M. Novak, B. Gamarra, K. Sirak, S. Connell, K. Stewardson, E. Harney, Q. Fu, G. Gonzalez-Fortes, E. R. Jones, S. A. Roodenberg, G. Lengyel, F. Bocquentin, B. Gasparian, J. M. Monge, M. Gregg, V. Eshed, A.-S. Mizrahi, C. Meiklejohn, F. Gerritsen, L. Bejenaru, M. Blüher, A. Campbell, G. Cavalleri, D. Comas, P. Froguel, E. Gilbert, S. M. Kerr, P. Kovacs, J. Krause, D. McGettigan, M. Merrigan, D. A. Merriwether, S. O'Reilly, M. B. Richards, O. Semino, M. Shamoon-Pour, G. Stefanescu, M. Stumvoll, A. Tönjes, A. Torroni, J. F. Wilson, L. Yengo, N. A. Hovhannisyan, N. Patterson, R. Pinhasi, D. Reich,

Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016). doi:10.1038/nature19310 Medline

17. M. Raghavan, P. Skoglund, K. E. Graf, M. Metspalu, A. Albrechtsen, I. Moltke, S. Rasmussen, T. W. Stafford Jr., L. Orlando, E. Metspalu, M. Karmin, K. Tambets, S. Rootsi, R. Mägi, P. F. Campos, E. Balanovska, O. Balanovsky, E. Khusnutdinova, S. Litvinov, L. P. Osipova, S. A. Fedorova, M. I. Voevoda, M. DeGiorgio, T. Sicheritz-Ponten, S. Brunak, S. Demeshchenko, T. Kivisild, R. Villems, R. Nielsen, M. Jakobsson, E. Willerslev, Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014). doi:10.1038/nature12736 Medline

18. F. Cruciani, R. La Fratta, B. Trombetta, P. Santolamazza, D. Sellitto, E. B. Colomb, J.-M. Dugoujon, F. Crivellaro, T. Benincasa, R. Pascone, P. Moral, E. Watson, B. Melegh, G. Barbujani, S. Fuselli, G. Vona, B. Zagradisnik, G. Assum, R. Brdicka, A. I. Kozlov, G. D. Efremov, A. Coppa, A. Novelletto, R. Scozzari, Tracing past human male movements in northern/eastern Africa and western Eurasia: New clues from Y-chromosomal haplogroups E-M78 and J-M12. *Mol. Biol. Evol.* **24**, 1300–1311 (2007). doi:10.1093/molbev/msm049 Medline

19. P. Pallary, in *Mémoires de la Société Historique Algérienne* (Jourdan, Algiers, 1909), vol. 3.

20. G. Camps, *Les Civilisations Préhistoriques de l'Afrique du Nord et du Sahara* (Doin, Paris, 1974), pp. 374.

21. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012). doi:10.1534/genetics.112.145037 Medline

22. M. Hervella, E. M. Svensson, A. Alberdi, T. Günther, N. Izagirre, A. R. Munters, S. Alonso, M. Ioana, F. Ridiche, A. Soficaru, M. Jakobsson, M. G. Netea, C. de-la-Rua, The mitogenome of a 35,000-year-old *Homo sapiens* from Europe supports a Palaeolithic back-migration to Africa. *Sci. Rep.* **6**, 25501–25505 (2016). doi:10.1038/srep25501 Medline

23. C. Posth, G. Renaud, A. Mittnik, D. G. Drucker, H. Rougier, C. Cupillard, F. Valentin, C. Thevenet, A. Furtwängler, C. Wißing, M. Francken, M. Malina, M. Bolus, M. Lari, E. Gigli, G. Capecchi, I. Crevecoeur, C. Beauval, D. Flas, M. Germonpré, J. van der Plicht, R. Cottiaux, B. Gély, A. Ronchitelli, K. Wehrberger, D. Grigorescu, J. Svoboda, P. Semal, D. Caramelli, H. Bocherens, K. Harvati, N. J. Conard, W. Haak, A. Powell, J. Krause, Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a Late Glacial population turnover in Europe. *Curr. Biol.* **26**, 827–833 (2016). doi:10.1016/j.cub.2016.01.037 Medline

24. A. J. Drummond, A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007). doi:10.1186/1471-2148-7-214 Medline

25. E. Pennarun, T. Kivisild, E. Metspalu, M. Metspalu, T. Reisberg, J.-P. Moisan, D. M. Behar, S. C. Jones, R. Villems, Divorcing the Late Upper Palaeolithic demographic histories of mtDNA haplogroups M1 and U6 in Africa. *BMC Evol. Biol.* **12**, 234–245 (2012). doi:10.1186/1471-2148-12-234 Medline

26. J. T. Hogue, R. Barton, New radiocarbon dates for the earliest Later Stone Age microlithic technology in Northwest Africa. *Quat. Int.* **413**, 62–75 (2016). doi:10.1016/j.quaint.2015.11.144

27. Q. Fu, C. Posth, M. Hajdinjak, M. Petr, S. Mallick, D. Fernandes, A. Furtwängler, W. Haak, M. Meyer, A. Mittnik, B. Nickel, A. Peltzer, N. Rohland, V. Slon, S. Talamo, I. Lazaridis, M. Lipson, I. Mathieson, S. Schiffels, P. Skoglund, A. P. Derevianko, N. Drozdov, V. Slavinsky, A. Tsybankov, R. G. Cremonesi, F. Mallegni, B. Gély, E. Vacca, M. R. G. Morales, L. G. Straus, C. Neugebauer-Maresch, M. Teschler-Nicola, S. Constantin, O. T. Moldovan, S. Benazzi, M. Peresani, D. Coppola, M. Lari, S. Ricci, A. Ronchitelli, F. Valentin, C. Thevenet, K. Wehrberger, D. Grigorescu, H. Rougier, I. Crevecoeur, D. Flas, P. Semal, M. A. Mannino, C. Cupillard, H. Bocherens, N. J. Conard, K. Harvati, V. Moiseyev, D. G. Drucker, J. Svoboda, M. P. Richards, D. Caramelli, R. Pinhasi, J. Kelso, N. Patterson, J. Krause, S. Pääbo, D. Reich, The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016). doi:10.1038/nature17993 Medline

28. A. Close, F. Wendorf, in *The World at 18,000 BP, Vol. 2. Low Latitudes* (Unwin Hyman, London, 1990), pp. 41–57.

29. E. Gobert, R. Vaufrey, Deux gisements extêmes d'Ibéromaurusien. *Anthropologie* **42**, 449–490 (1932).

30. C. Goetz, Note d'Archéologie préhistorique Nord-Africaine sur un foyer oranien de la sablière d'El-Kçar. *Bull. Soc. Préhist. Fr.* **38**, 262–265 (1941). doi:10.3406/bspf.1941.5550

31. C. McBurney, *The Haua Fteah (Cyrenaica) and the Stone Age of the South-East Mediterranean* (Cambridge University Press, Cambridge, 1967), pp. 404.

32. R. Barton, C. S. Lane, P. G. Albert, D. White, S. N. Collcutt, A. Bouzouggar, P. Ditchfield, L. Farr, A. Oh, L. Ottolini, V. C. Smith, P. Van Peer, K. Kindermann, The role of cryptotephra in refining the chronology of Late Pleistocene human evolution and cultural change in North Africa. *Quat. Sci. Rev.* **118**, 151–169 (2015). doi:10.1016/j.quascirev.2014.09.008

33. A. Bouzouggar, R. N. E. Barton, S. Blockley, C. Bronk-Ramsey, S. N. Collcutt, R. Gale, T. F. G. Higham, L. T. Humphrey, S. Parfitt, E. Turner, S. Ward, Reevaluating the age of the Iberomaurusian in Morocco. *Afr. Archaeol. Rev.* **25**, 3–19 (2008). doi:10.1007/s10437-008-9023-3

34. S. Hachi, L'Iberomaurusien, decouverte des fouilles d'Afalou (Bedjaıa, Algerie). *Anthropologie* **100**, 55–76 (1996).

35. J. Linstädter, J. Eiwanger, A. Mikdad, G. C. Weniger, Human occupation of Northwest Africa: A review of Middle Palaeolithic to Epipalaeolithic sites in Morocco. *Quat. Int.* **274**, 158–174 (2012). doi:10.1016/j.quaint.2012.02.017

36. R. Barton, A. Bouzouggar, S. N. Collcutt, R. Gale, T. F. G. Higham, L. T. Humphrey, S. Parfitt, E. Rhodes, C. B. Stringer, F. Malek, The Late Upper Palaeolithic occupation of the Moroccan Northwest Maghreb during the Last Glacial Maximum. *Afr. Archaeol. Rev.* **22**, 77–100 (2005). doi:10.1007/s10437-005-4190-y

37. A. Bouzouggar, in *I Seminario Hispano-Marroquí de Especialización en Arqueología*, R. B. Bernal D, Ramos J, Bouzouggar A., Eds. (Universidad de Cádiz, Cádiz, 2006), pp. 133–142.

38. K. Douka, Z. Jacobs, C. Lane, R. Grün, L. Farr, C. Hunt, R. H. Inglis, T. Reynolds, P. Albert, M. Aubert, V. Cullen, E. Hill, L. Kinsley, R. G. Roberts, E. L. Tomlinson, S. Wulf, G. Barker, The chronostratigraphy of the Haua Fteah cave (Cyrenaica, northeast Libya). *J. Hum. Evol.* **66**, 39–63 (2014). doi:10.1016/j.jhevol.2013.10.001 Medline

39. D. Ferembach, On the origin of the Iberomaurusians (Upper palaeolithic: North Africa). A new hypothesis. *J. Hum. Evol.* **14**, 393–397 (1985). doi:10.1016/S0047-2484(85)80047-6

40. A. Capart, D. Capart, *L'homme et des Déluges* (Hayez, Bruxelles, 1986), pp. 338.

41. M. A. Mannino, G. Catalano, S. Talamo, G. Mannino, R. Di Salvo, V. Schimmenti, C. Lalueza-Fox, A. Messina, D. Petruso, D. Caramelli, M. P. Richards, L. Sineo, Origin and diet of the prehistoric hunter-gatherers on the mediterranean island of Favignana (Ègadi Islands, Sicily). *PLOS ONE* **7**, e49802 (2012). doi:10.1371/journal.pone.0049802 Medline

42. M. A. Mannino, K. D. Thomas, M. J. Leng, R. Di Salvo, M. P. Richards, Stuck to the shore? Investigating prehistoric hunter-gatherer subsistence, mobility and territoriality in a Mediterranean coastal landscape through isotope analyses on marine mollusc shell carbonates and human bone collagen. *Quat. Int.* **244**, 88–104 (2011). doi:10.1016/j.quaint.2011.05.044

43. R. Barton, A. Bouzouggar, S. N. Collcutt, Y. Carrión Marco, L. Clark-Balzan, N. C. Debenham, J. Morales, Reconsidering the MSA to LSA transition at Taforalt Cave (Morocco) in the light of new multi-proxy dating evidence. *Quat. Int.* **413**, 36–49 (2016). doi:10.1016/j.quaint.2015.11.085

44. M. D. Bosch, M. A. Mannino, A. L. Prendergast, T. C. O'Connell, B. Demarchi, S. M. Taylor, L. Niven, J. van der Plicht, J.-J. Hublin, New chronology for Ksâr 'Akil (Lebanon) supports Levantine route of modern human dispersal into Europe. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7683–7688 (2015). doi:10.1073/pnas.1501529112 Medline

45. H. Watanabé, Les "éclats et lames à Chanfrein" et la technique de fracturation transversale dans un horizon paléolithique en Palestine. *Bull. Soc. Préhist. Fr.* **61**, 84–88 (1964). doi:10.3406/bspf.1964.8800

46. E. M. L. Scerri, The North African Middle Stone Age and its place in recent human evolution. *Evol. Anthropol.* **26**, 119–135 (2017). doi:10.1002/evan.21527 Medline

47. M. Rasse, S. Soriano, C. Tribolo, S. Stokes, E. Huysecom, La séquence pléistocène supérieur d'Ounjougou (Pays Dogon, Mali): Évolution géomorphologique, enregistrements sédimentaires et changements culturels. *Quaternaire* **15**, 329–341 (2004). doi:10.3406/quate.2004.1779

48. B. Chevrier, M. Rasse, L. Lespez, C. Tribolo, I. Hajdas, M. Guardiola Fígols, B. Lebrun, A. Leplongeon, A. Camara, É. Huysecom, West African Palaeolithic history: New archaeological and chronostratigraphic data from the Falémé valley, eastern Senegal. *Quat. Int.* **408**, 33–52 (2016). doi:10.1016/j.quaint.2015.11.060

49. J. Dastugue, in *La Nécropole Epipaléolithique de Taforalt (Maroc Oriental),* Ferembach, Ed. (Edita Casablanca, Rabat, 1962), pp. 133–158.

50. J. Roche, *L'Epipaléolithique Marocaine* (Fondation Calouste Gulbenkian, Lisbon, 1963).

51. V. Taylor, R. N. E. Barton, M. Bell, A. Bouzouggar, S. Collcutt, S. Black, J. T. Hogue, The Epipalaeolithic (Iberomaurusian) at Grotte des Pigeons (Taforalt), Morocco: A preliminary study of the land Mollusca. *Quat. Int.* **244**, 5–14 (2011). doi:10.1016/j.quaint.2011.04.041

52. L. T. Humphrey, I. De Groote, J. Morales, N. Barton, S. Collcutt, C. Bronk Ramsey, A. Bouzouggar, Earliest evidence for caries and exploitation of starchy plant foods in Pleistocene hunter-gatherers from Morocco. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 954–959 (2014). doi:10.1073/pnas.1318176111 Medline

53. I. De Groote, L. T. Humphrey, Characterizing evulsion in the Later Stone Age Maghreb: Age, sex and effects on mastication. *Quat. Int.* **413**, 50–61 (2016). doi:10.1016/j.quaint.2015.08.082

54. L. Humphrey, S. M. Bello, E. Turner, A. Bouzouggar, N. Barton, Iberomaurusian funerary behaviour: Evidence from Grotte des Pigeons, Taforalt, Morocco. *J. Hum. Evol.* **62**, 261–273 (2012). doi:10.1016/j.jhevol.2011.11.003 Medline

55. A. Immel, A. Le Cabec, M. Bonazzi, A. Herbig, H. Temming, V. J. Schuenemann, K. I. Bos, F. Langbein, K. Harvati, A. Bridault, G. Pion, M.-A. Julien, O. Krotova, N. J. Conard, S. C. Münzel, D. G. Drucker, B. Viola, J.-J. Hublin, P. Tafforeau, J. Krause, Effect of x-ray irradiation on ancient DNA in sub-fossil bones - Guidelines for safe x-ray imaging. *Sci. Rep.* **6**, 32969 (2016). doi:10.1038/srep32969 Medline

56. C. Gamba, E. R. Jones, M. D. Teasdale, R. L. McLaughlin, G. Gonzalez-Fortes, V. Mattiangeli, L. Domboróczki, I. Kővári, I. Pap, A. Anders, A. Whittle, J. Dani, P. Raczky, T. F. G. Higham, M. Hofreiter, D. G. Bradley, R. Pinhasi, Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257–5265 (2014). doi:10.1038/ncomms6257 Medline

57. R. Pinhasi, D. Fernandes, K. Sirak, M. Novak, S. Connell, S. Alpaslan-Roodenberg, F. Gerritsen, V. Moiseyev, A. Gromov, P. Raczky, A. Anders, M. Pietrusewsky, G. Rollefson, M. Jovanovic, H. Trinhhoang, G. Bar-Oz, M. Oxenham, H. Matsumura, M. Hofreiter, Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLOS ONE* **10**, e0129102 (2015). doi:10.1371/journal.pone.0129102 Medline

58. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. García, S. Pääbo, J.-L. Arsuaga, M. Meyer, 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 (2013). doi:10.1073/pnas.1314445110 Medline

59. V. Slon, C. Hopfe, C. L. Weiß, F. Mafessoni, M. de la Rasilla, C. Lalueza-Fox, A. Rosas, M. Soressi, M. V. Knul, R. Miller, J. R. Stewart, A. P. Derevianko, Z. Jacobs, B. Li, R. G. Roberts, M. V. Shunkov, H. de Lumley, C. Perrenoud, I. Gušić, Ž. Kućan, P. Rudan, A. Aximu-Petri, E. Essel, S. Nagel, B. Nickel, A. Schmidt, K. Prüfer, J. Kelso, H. A. Burbano, S. Pääbo, M. Meyer, Neandertal and Denisovan DNA from Pleistocene sediments. *Science* **356**, 605–608 (2017). doi:10.1126/science.aam9695 Medline

60. M.-T. Gansauge, M. Meyer, Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res.* **24**, 1543–1549 (2014). doi:10.1101/gr.174201.114 Medline

61. M. Kircher, S. Sawyer, M. Meyer, Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012). doi:10.1093/nar/gkr771 Medline

62. A. Peltzer, G. Jäger, A. Herbig, A. Seitz, C. Kniep, J. Krause, K. Nieselt, EAGER: Efficient ancient genome reconstruction. *Genome Biol.* **17**, 60–73 (2016). doi:10.1186/s13059-016-0918-z Medline

63. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88–93 (2016). doi:10.1186/s13104-016-1900-2 Medline

64. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi:10.1093/bioinformatics/btp698 Medline

65. M. Meyer, M. Kircher, M.-T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, F. Jay, K. Prüfer, C. de Filippo, P. H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R. E. Green, K. Bryc, A. W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M. F. Hammer, M. V. Shunkov, A. P. Derevianko, N. Patterson, A. M. Andrés, E. E. Eichler, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012). doi:10.1126/science.1224344 Medline

66. A. W. Briggs, U. Stenzel, P. L. F. Johnson, R. E. Green, J. Kelso, K. Prüfer, M. Meyer, J. Krause, M. T. Ronan, M. Lachmann, S. Pääbo, Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14616–14621 (2007). doi:10.1073/pnas.0704665104 Medline

67. M.-T. Gansauge, M. Meyer, Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013). doi:10.1038/nprot.2013.038 Medline

68. L. Kistler, R. Ware, O. Smith, M. Collins, R. G. Allaby, A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Res.* **45**, 6310–6320 (2017). doi:10.1093/nar/gkx361 Medline

69. S. Sawyer, J. Krause, K. Guschanski, V. Savolainen, S. Pääbo, Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLOS ONE* **7**, e34131 (2012). doi:10.1371/journal.pone.0034131 Medline

70. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224–241 (2015). doi:10.1186/s13059-015-0776-0 Medline

71. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356–368 (2014). doi:10.1186/s12859-014-0356-4 Medline

72. M. Rasmussen, X. Guo, Y. Wang, K. E. Lohmueller, S. Rasmussen, A. Albrechtsen, L. Skotte, S. Lindgreen, M. Metspalu, T. Jombart, T. Kivisild, W. Zhai, A. Eriksson, A. Manica, L. Orlando, F. M. De La Vega, S. Tridico, E. Metspalu, K. Nielsen, M. C. Ávila-Arcos, J. V. Moreno-Mayar, C. Muller, J. Dortch, M. T. P. Gilbert, O. Lund, A. Wesolowska, M. Karmin, L. A. Weinert, B. Wang, J. Li, S. Tai, F. Xiao, T. Hanihara, G. van Driem, A. R. Jha, F.-X. Ricaut, P. de Knijff, A. B. Migliano, I. Gallego Romero, K. Kristiansen, D. M. Lambert, S. Brunak, P. Forster, B. Brinkmann, O. Nehlich, M. Bunce, M. Richards, R. Gupta, C. D. Bustamante, A. Krogh, R. A. Foley, M. M. Lahr, F. Balloux, T. Sicheritz-Pontén, R. Villems, R. Nielsen, J. Wang, E. Willerslev, An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98 (2011). doi:10.1126/science.1211177 Medline

73. I. Lazaridis, N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, J. G. Schraiber, S. Castellano, M. Lipson, B. Berger, C. Economou, R. Bollongino, Q. Fu, K. I. Bos, S. Nordenfelt, H. Li, C. de Filippo, K. Prüfer, S. Sawyer, C. Posth, W. Haak, F. Hallgren, E. Fornander, N. Rohland, D. Delsate, M. Francken, J.-M. Guinet, J. Wahl, G. Ayodo, H. A. Babiker, G. Bailliet, E. Balanovska, O. Balanovsky, R. Barrantes, G. Bedoya, H. Ben-Ami, J. Bene, F. Berrada, C. M. Bravi, F. Brisighelli, G. B. J. Busby, F. Cali, M. Churnosov, D. E. C. Cole, D. Corach, L. Damba, G. van Driem, S. Dryomov, J.-M. Dugoujon, S. A. Fedorova, I. Gallego Romero, M. Gubina, M. Hammer, B. M. Henn, T. Hervig, U. Hodoglugil, A. R. Jha, S. Karachanak-Yankova, R. Khusainova, E. Khusnutdinova, R. Kittles, T. Kivisild, W. Klitz, V. Kučinskas, A. Kushniarevich, L. Laredj, S. Litvinov, T. Loukidis, R. W. Mahley, B. Melegh, E. Metspalu, J. Molina, J. Mountain, K. Näkkäläjärvi, D. Nesheva, T. Nyambo, L. Osipova, J. Parik, F. Platonov, O. Posukh, V. Romano, F. Rothhammer, I. Rudan, R. Ruizbakiev, H. Sahakyan, A. Sajantila, A. Salas, E. B. Starikovskaya, A. Tarekegn, D. Toncheva, S. Turdikulova, I. Uktveryte, O. Utevska, R. Vasquez, M. Villena, M. Voevoda, C. A. Winkler, L. Yepiskoposyan, P. Zalloua, T. Zemunik, A. Cooper, C. Capelli, M. G. Thomas, A. Ruiz-Linares, S. A. Tishkoff, L. Singh, K. Thangaraj, R. Villems, D. Comas, R. Sukernik, M. Metspalu, M. Meyer, E. E. Eichler, J. Burger, M. Slatkin, S. Pääbo, J. Kelso, D. Reich, J. Krause, Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014). doi:10.1038/nature13673 Medline

74. J. K. Pickrell, N. Patterson, P.-R. Loh, M. Lipson, B. Berger, M. Stoneking, B. Pakendorf, D. Reich, Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2632–2637 (2014). doi:10.1073/pnas.1313787111 Medline

75. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009). doi:10.1101/gr.094052.109 Medline

76. C. C. Chang, C. C. Chow, L. C. A. M. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7–22 (2015). doi:10.1186/s13742-015-0047-8 Medline

77. Q. Fu, H. Li, P. Moorjani, F. Jay, S. M. Slepchenko, A. A. Bondarev, P. L. F. Johnson, A. Aximu-Petri, K. Prüfer, C. de Filippo, M. Meyer, N. Zwyns, D. C. Salazar-García, Y. V. Kuzmin, S. G. Keates, P. A. Kosintsev, D. I. Razhev, M. P. Richards, N. V. Peristov, M. Lachmann, K. Douka, T. F. G. Higham, M. Slatkin, J.-J. Hublin, D. Reich, J. Kelso, T. B.

Viola, S. Pääbo, Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014). doi:10.1038/nature13810 Medline

78. P.-R. Loh, M. Lipson, N. Patterson, P. Moorjani, J. K. Pickrell, D. Reich, B. Berger, Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233–1254 (2013). doi:10.1534/genetics.112.147330 Medline

79. J. D. Wall, M. A. Yang, F. Jay, S. K. Kim, E. Y. Durand, L. S. Stevison, C. Gignoux, A. Woerner, M. F. Hammer, M. Slatkin, Higher levels of neanderthal ancestry in East Asians than in Europeans. *Genetics* **194**, 199–209 (2013). doi:10.1534/genetics.112.148213 Medline

80. I. Mathieson, I. Lazaridis, N. Rohland, S. Mallick, N. Patterson, S. A. Roodenberg, E. Harney, K. Stewardson, D. Fernandes, M. Novak, K. Sirak, C. Gamba, E. R. Jones, B. Llamas, S. Dryomov, J. Pickrell, J. L. Arsuaga, J. M. B. de Castro, E. Carbonell, F. Gerritsen, A. Khokhlov, P. Kuznetsov, M. Lozano, H. Meller, O. Mochalov, V. Moiseyev, M. A. R. Guerra, J. Roodenberg, J. M. Vergès, J. Krause, A. Cooper, K. W. Alt, D. Brown, D. Anthony, C. Lalueza-Fox, W. Haak, R. Pinhasi, D. Reich, Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015). doi:10.1038/nature16152 Medline

81. W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfelt, E. Harney, K. Stewardson, Q. Fu, A. Mittnik, E. Bánffy, C. Economou, M. Francken, S. Friederich, R. G. Pena, F. Hallgren, V. Khartanovich, A. Khokhlov, M. Kunst, P. Kuznetsov, H. Meller, O. Mochalov, V. Moiseyev, N. Nicklisch, S. L. Pichler, R. Risch, M. A. Rojo Guerra, C. Roth, A. Szécsényi-Nagy, J. Wahl, M. Meyer, J. Krause, D. Brown, D. Anthony, A. Cooper, K. W. Alt, D. Reich, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015). doi:10.1038/nature14317 Medline

82. C. M. Schlebusch, P. Skoglund, P. Sjödin, L. M. Gattepaille, D. Hernandez, F. Jay, S. Li, M. De Jongh, A. Singleton, M. G. B. Blum, H. Soodyall, M. Jakobsson, Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science* **338**, 374–379 (2012). doi:10.1126/science.1227721 Medline

83. J. Z. Li, D. M. Absher, H. Tang, A. M. Southwick, A. M. Casto, S. Ramachandran, H. M. Cann, G. S. Barsh, M. Feldman, L. L. Cavalli-Sforza, R. M. Myers, Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008). doi:10.1126/science.1153717 Medline

84. J. E. Tierney, F. S. R. Pausata, P. B. deMenocal, Rainfall regimes of the Green Sahara. *Sci. Adv.* **3**, e1601503 (2017). doi:10.1126/sciadv.1601503 Medline

85. R. Kefi, M. Hechmi, C. Naouali, H. Jmel, S. Hsouna, E. Bouzaid, S. Abdelhak, E. Beraud-Colomb, A. Stevanovitch, On the origin of Iberomaurusians: New data based on ancient mitochondrial DNA and phylogenetic analysis of Afalou and Taforalt populations. *Mitochondrial DNA Part A* **29**, 147–157 (2018). Medline

86. M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious Basic: An integrated and extendable desktop software platform for the

organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012). doi:10.1093/bioinformatics/bts199 Medline

87. D. M. Behar, M. van Oven, S. Rosset, M. Metspalu, E.-L. Loogväli, N. M. Silva, T. Kivisild, A. Torroni, R. Villems, A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012). doi:10.1016/j.ajhg.2012.03.002 Medline

88. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004). doi:10.1093/nar/gkh340 Medline

89. K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013). doi:10.1093/molbev/mst197 Medline

90. G. Baele, P. Lemey, Bayesian evolutionary model testing in the phylogenomics era: Matching model complexity with computational efficiency. *Bioinformatics* **29**, 1970–1979 (2013). doi:10.1093/bioinformatics/btt340 Medline

91. B. Secher, R. Fregel, J. M. Larruga, V. M. Cabrera, P. Endicott, J. J. Pestano, A. M. González, The history of the North African mitochondrial DNA haplogroup U6 gene flow into the African, Eurasian and American continents. *BMC Evol. Biol.* **14**, 109–125 (2014). doi:10.1186/1471-2148-14-109 Medline

92. C. L. Hernández, P. Soares, J. M. Dugoujon, A. Novelletto, J. N. Rodríguez, T. Rito, M. Oliveira, M. Melhaoui, A. Baali, L. Pereira, R. Calderón, Early Holocene and historic mtDNA African signatures in the Iberian Peninsula: The Andalusian region as a paradigm. *PLOS ONE* **10**, e0139784 (2015). doi:10.1371/journal.pone.0139784 Medline

93. K. L. Hart, S. L. Kimura, V. Mushailov, Z. M. Budimlija, M. Prinz, E. Wurmbach, Improved eye- and skin-color prediction based on 8 SNPs. *Croat. Med. J.* **54**, 248–256 (2013). doi:10.3325/cmj.2013.54.248 Medline

94. C. Basu Mallick, F. M. Iliescu, M. Möls, S. Hill, R. Tamang, G. Chaubey, R. Goto, S. Y. W. Ho, I. Gallego Romero, F. Crivellaro, G. Hudjashov, N. Rai, M. Metspalu, C. G. N. Mascie-Taylor, R. Pitchappan, L. Singh, M. Mirazon-Lahr, K. Thangaraj, R. Villems, T. Kivisild, The light skin allele of *SLC24A5* in South Asians and Europeans shares identity by descent. *PLOS Genet.* **9**, e1003912 (2013). doi:10.1371/journal.pgen.1003912 Medline

95. M. Mukherjee, S. Mukerjee, N. Sarkar-Roy, T. Ghosh, D. Kalpana, A. K. Sharma, Polymorphisms of four pigmentation genes (*SLC45A2*, *SLC24A5*, *MC1R* and *TYRP1*) among eleven endogamous populations of India. *J. Genet.* **92**, 135–139 (2013). doi:10.1007/s12041-013-0225-3 Medline

96. H. Eiberg, J. Troelsen, M. Nielsen, A. Mikkelsen, J. Mengel-From, K. W. Kjaer, L. Hansen, Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the *HERC2* gene inhibiting *OCA2* expression. *Hum. Genet.* **123**, 177–187 (2008). doi:10.1007/s00439-007-0460-x Medline

97. R. A. Sturm, D. L. Duffy, Z. Z. Zhao, F. P. N. Leite, M. S. Stark, N. K. Hayward, N. G. Martin, G. W. Montgomery, A single SNP in an evolutionary conserved region within

intron 86 of the *HERC2* gene determines human blue-brown eye color. *Am. J. Hum. Genet.* **82**, 424–431 (2008). doi:10.1016/j.ajhg.2007.11.005 Medline

98. S. E. Medland, D. R. Nyholt, J. N. Painter, B. P. McEvoy, A. F. McRae, G. Zhu, S. D. Gordon, M. A. R. Ferreira, M. J. Wright, A. K. Henders, M. J. Campbell, D. L. Duffy, N. K. Hansell, S. Macgregor, W. S. Slutske, A. C. Heath, G. W. Montgomery, N. G. Martin, Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am. J. Hum. Genet.* **85**, 750–755 (2009). doi:10.1016/j.ajhg.2009.10.009 Medline

99. N. S. Enattah, T. Sahi, E. Savilahti, J. D. Terwilliger, L. Peltonen, I. Järvelä, Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30**, 233–237 (2002). doi:10.1038/ng826 Medline

100. B. L. Jones, T. O. Raga, A. Liebert, P. Zmarz, E. Bekele, E. T. Danielsen, A. K. Olsen, N. Bradman, J. T. Troelsen, D. M. Swallow, Diversity of lactase persistence alleles in Ethiopia: Signature of a soft selective sweep. *Am. J. Hum. Genet.* **93**, 538–544 (2013). doi:10.1016/j.ajhg.2013.07.008 Medline

101. A. Ranciaro, M. C. Campbell, J. B. Hirbo, W.-Y. Ko, A. Froment, P. Anagnostou, M. J. Kotze, M. Ibrahim, T. Nyambo, S. A. Omar, S. A. Tishkoff, Genetic origins of lactase persistence and the spread of pastoralism in Africa. *Am. J. Hum. Genet.* **94**, 496–510 (2014). doi:10.1016/j.ajhg.2014.02.009 Medline

102. C. J. E. Ingram, M. F. Elamin, C. A. Mulcare, M. E. Weale, A. Tarekegn, T. O. Raga, E. Bekele, F. M. Elamin, M. G. Thomas, N. Bradman, D. M. Swallow, A novel polymorphism associated with lactose tolerance in Africa: Multiple causes for lactase persistence? *Hum. Genet.* **120**, 779–788 (2007). doi:10.1007/s00439-006-0291-1 Medline

103. M. E. Allentoft, M. Sikora, K.-G. Sjögren, S. Rasmussen, M. Rasmussen, J. Stenderup, P. B. Damgaard, H. Schroeder, T. Ahlström, L. Vinner, A.-S. Malaspinas, A. Margaryan, T. Higham, D. Chivall, N. Lynnerup, L. Harvig, J. Baron, P. Della Casa, P. Dąbrowski, P. R. Duffy, A. V. Ebel, A. Epimakhov, K. Frei, M. Furmanek, T. Gralak, A. Gromov, S. Gronkiewicz, G. Grupe, T. Hajdu, R. Jarysz, V. Khartanovich, A. Khokhlov, V. Kiss, J. Kolář, A. Kriiska, I. Lasak, C. Longhi, G. McGlynn, A. Merkevicius, I. Merkyte, M. Metspalu, R. Mkrtchyan, V. Moiseyev, L. Paja, G. Pálfi, D. Pokutta, Ł. Pospieszny, T. D. Price, L. Saag, M. Sablin, N. Shishlina, V. Smrčka, V. I. Soenov, V. Szeverényi, G. Tóth, S. V. Trifanova, L. Varul, M. Vicze, L. Yepiskoposyan, V. Zhitenev, L. Orlando, T. Sicheritz-Pontén, S. Brunak, R. Nielsen, K. Kristiansen, E. Willerslev, Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015). doi:10.1038/nature14507 Medline

104. L. B. Barreiro, M. Ben-Ali, H. Quach, G. Laval, E. Patin, J. K. Pickrell, C. Bouchier, M. Tichit, O. Neyrolles, B. Gicquel, J. R. Kidd, K. K. Kidd, A. Alcaïs, J. Ragimbeau, S. Pellegrini, L. Abel, J.-L. Casanova, L. Quintana-Murci, Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLOS Genet.* **5**, e1000562 (2009). doi:10.1371/journal.pgen.1000562 Medline

105. P. Uciechowski, H. Imhoff, C. Lange, C. G. Meyer, E. N. Browne, D. K. Kirsten, A. K. Schröder, B. Schaaf, A. Al-Lahham, R. R. Reinert, N. Reiling, H. Haase, A. Hatzmann, D. Fleischer, N. Heussen, M. Kleines, L. Rink, Susceptibility to tuberculosis is associated

with TLR1 polymorphisms resulting in a lack of TLR1 cell surface expression. *J. Leukoc. Biol.* **90**, 377–388 (2011). doi:10.1189/jlb.0409233 Medline

106. G. D. Poznik, Y. Xue, F. L. Mendez, T. F. Willems, A. Massaia, M. A. Wilson Sayres, Q. Ayub, S. A. McCarthy, A. Narechania, S. Kashin, Y. Chen, R. Banerjee, J. L. Rodriguez-Flores, M. Cerezo, H. Shao, M. Gymrek, A. Malhotra, S. Louzada, R. Desalle, G. R. S. Ritchie, E. Cerveira, T. W. Fitzgerald, E. Garrison, A. Marcketta, D. Mittelman, M. Romanovitch, C. Zhang, X. Zheng-Bradley, G. R. Abecasis, S. A. McCarroll, P. Flicek, P. A. Underhill, L. Coin, D. R. Zerbino, F. Yang, C. Lee, L. Clarke, A. Auton, Y. Erlich, R. E. Handsaker, C. D. Bustamante, C. Tyler-Smith; 1000 Genomes Project Consortium, Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat. Genet.* **48**, 593–599 (2016). doi:10.1038/ng.3559 Medline

107. D. J. Kennett, S. Plog, R. J. George, B. J. Culleton, A. S. Watson, P. Skoglund, N. Rohland, S. Mallick, K. Stewardson, L. Kistler, S. A. LeBlanc, P. M. Whiteley, D. Reich, G. H. Perry, Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017). doi:10.1038/ncomms14115 Medline

108. M. Sikora, A. Seguin-Orlando, V. C. Sousa, A. Albrechtsen, T. Korneliussen, A. Ko, S. Rasmussen, I. Dupanloup, P. R. Nigst, M. D. Bosch, G. Renaud, M. E. Allentoft, A. Margaryan, S. V. Vasilyev, E. V. Veselovskaya, S. B. Borutskaya, T. Deviese, D. Comeskey, T. Higham, A. Manica, R. Foley, D. J. Meltzer, R. Nielsen, L. Excoffier, M. Mirazon Lahr, L. Orlando, E. Willerslev, Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* **358**, 659–662 (2017). doi:10.1126/science.aao1807 Medline

109. P. M. Vermeersch, in *New Light on the Northeast African Past: Current Prehistoric Research*, F. Kees, R. Kuper, Eds. (Heinrich-Barth-Institut, Köln, 1992), pp. 100–153.

110. F. Wendorf, R. Schild, A. E. Close, Eds., *The Prehistory of Wadi Kubbaniya* (Southern Methodist University, Dallas, 1989).

111. P. M. Vermeersch, W. Van Neer, Nile behaviour and Late Palaeolithic humans in Upper Egypt during the Late Pleistocene. *Quat. Sci. Rev.* **130**, 155–167 (2015). doi:10.1016/j.quascirev.2015.03.025

112. J. T. Hogue, *The Origin and Development of the Pleistocene LSA in Northwest Africa: A Case Study from Grotte Des Pigeons (Taforalt), Morocco* (University of Oxford, Oxford, 2014), pp. 654.

113. A. Soficaru, A. Dobos, E. Trinkaus, Early modern humans from the Pestera Muierii, Baia de Fier, Romania. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17196–17201 (2006). doi:10.1073/pnas.0608443103 Medline

114. S. Benazzi, K. Douka, C. Fornai, C. C. Bauer, O. Kullmer, J. Svoboda, I. Pap, F. Mallegni, P. Bayle, M. Coquerelle, S. Condemi, A. Ronchitelli, K. Harvati, G. W. Weber, Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* **479**, 525–528 (2011). doi:10.1038/nature10617 Medline

Current Biology, Volume *29*

Supplemental Information

Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula

Vanessa Villalba-Mouco, Marieke S. van de Loosdrecht, Cosimo Posth, Rafael Mora, Jorge Martínez-Moreno, Manuel Rojo-Guerra, Domingo C. Salazar-García, José I. Royo-Guillén, Michael Kunst, Hélène Rougier, Isabelle Crevecoeur, Héctor Arcusa-Magallón, Cristina Tejedor-Rodríguez, Iñigo García-Martínez de Lagrán, Rafael Garrido-Pena, Kurt W. Alt, Choongwon Jeong, Stephan Schiffels, Pilar Utrilla, Johannes Krause, and Wolfgang Haak

Supplemental Figures

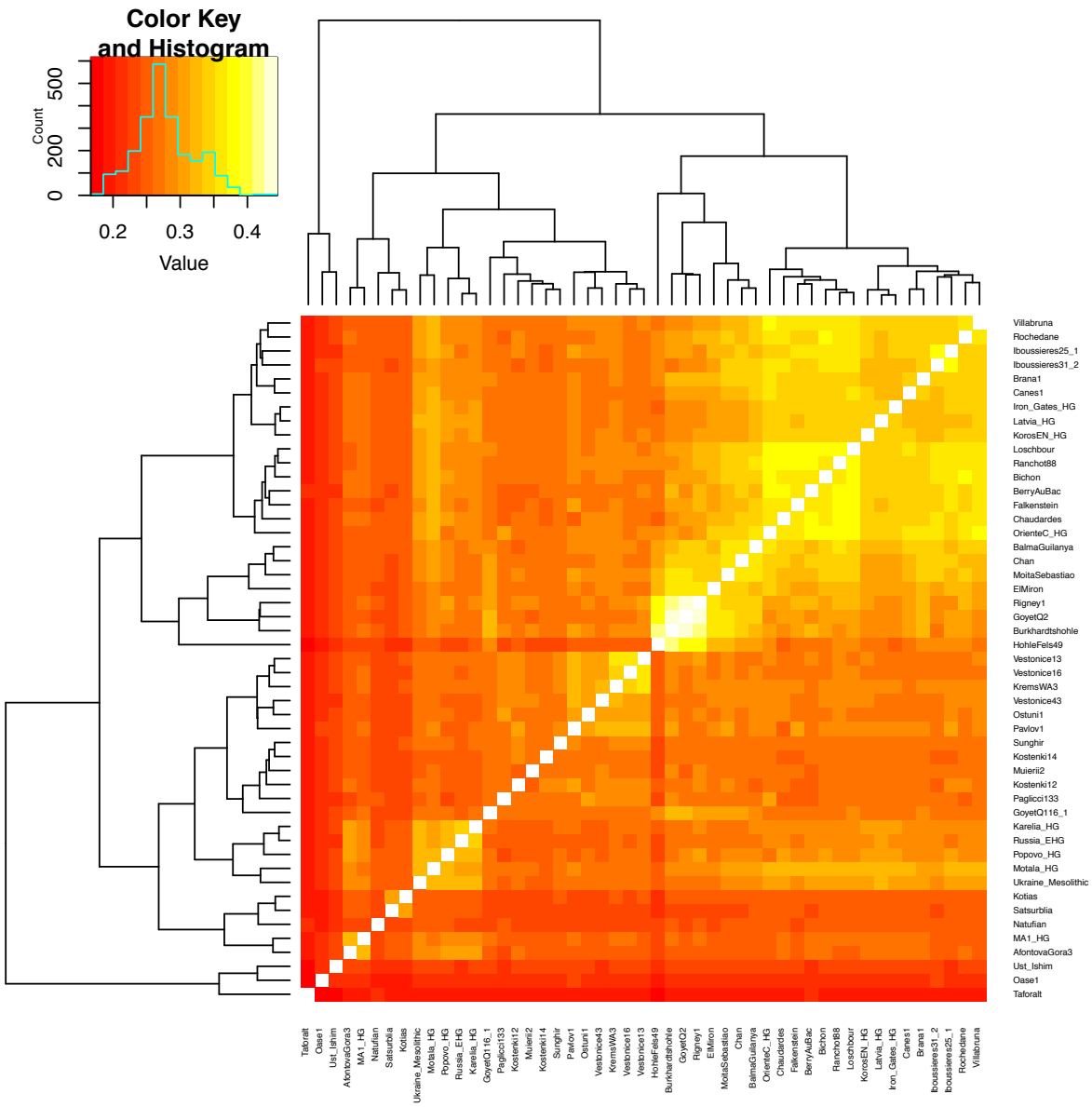


Figure S1. Heat plot showing the genetic distances between Eurasian HG, Related to Figure 2A. Genetic distances were calculated using *f3-outgroup* statistics of the form *f3(X;Y, Mbuti)*, with X and Y being Eurasian hunter-gatherers in all possible pairwise comparisons. The analysis has been restricted to samples with more than 30,000 SNPs, following Fu et al. [S1]. The clustering pattern is similar to the MDS plot (STAR methods, Figure 2A): the newly reported Moita do Sebastião [~8ky cal BP], Balma Guilanyà [~12ky cal BP], and Chan [~9ky cal BP] cluster with El Mirón and Goyet Q-2, whereas La Braña 1 and Canes 1 cluster with Villabruna.

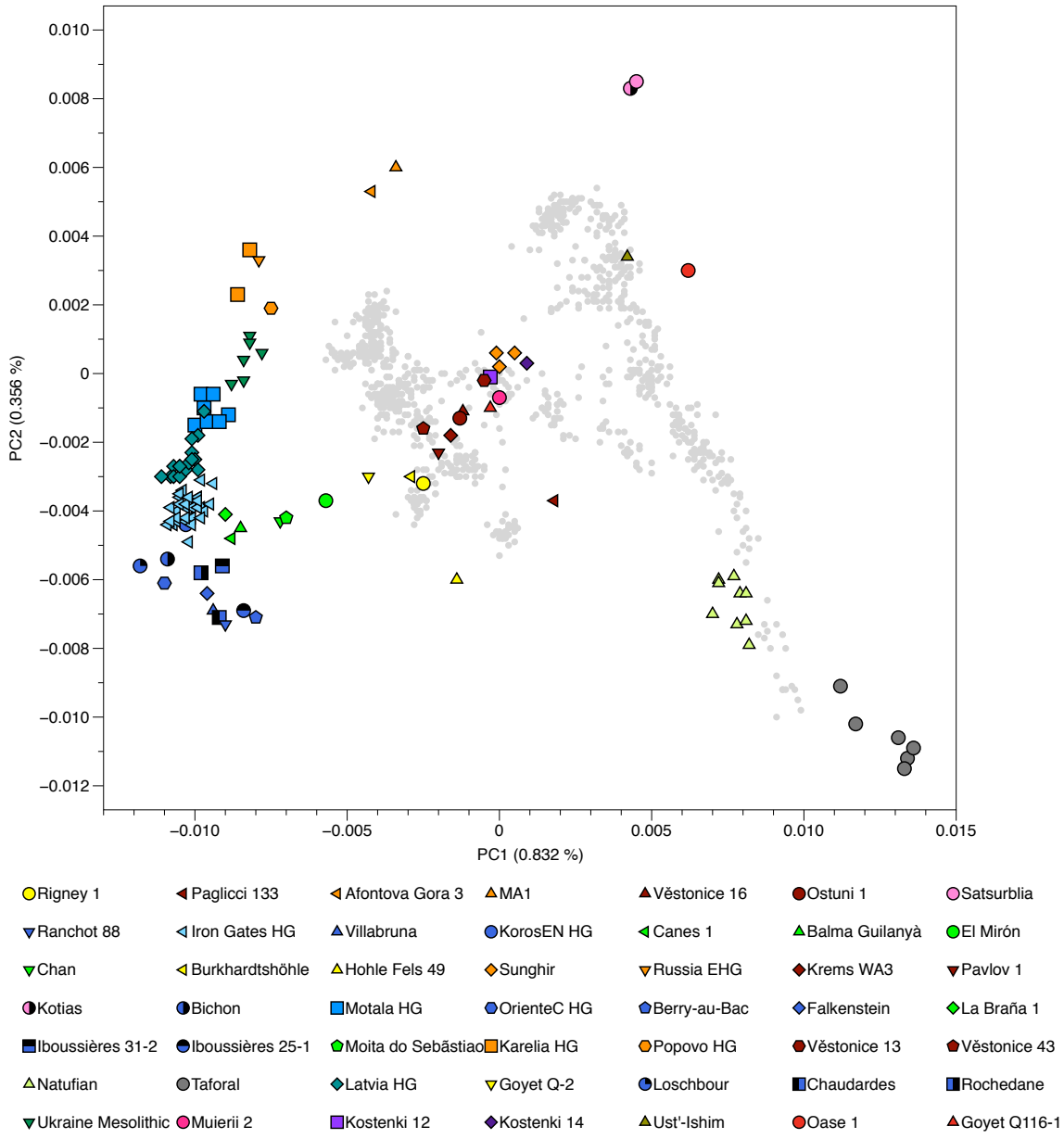


Figure S2. Principal Component Analysis of Hunter-gatherer individuals, Related to Figure 2A. PCA analysis calculated with 777 present day West Eurasians [S2] with option shrinkmode:YES on which HG individuals were projected.

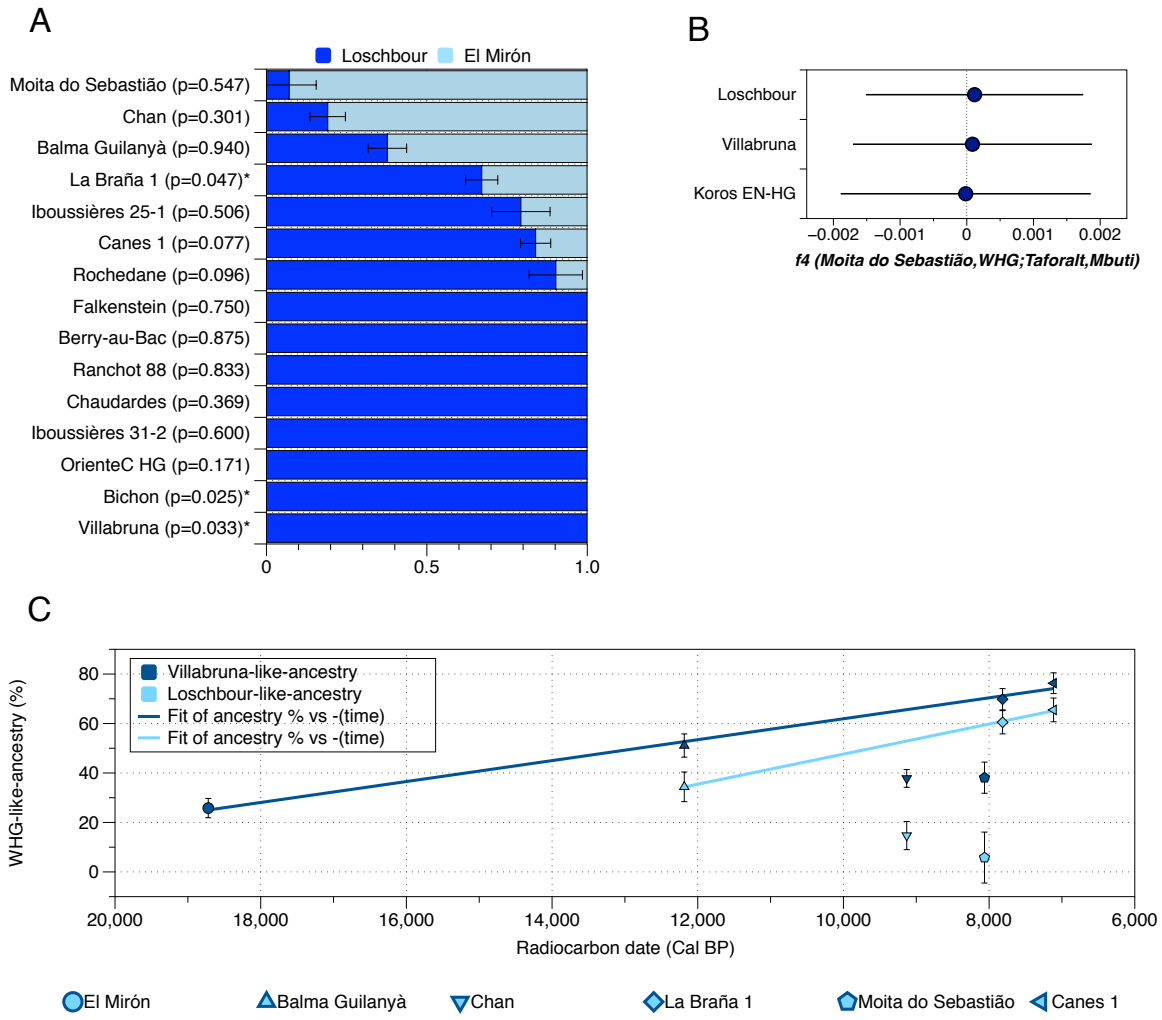


Figure S3. Hunter-gatherer ancestry and geographical correlation, **Related to Figure 3D**
A) Modelling European HG as admixture of Villabruna-like and El Mirón-like ancestry using *Loschbour* and *El Mirón* as proximal sources, respectively (error bars indicate ±1 standard error). **B)** f4-statistics showing **no** affinity between Geometric Mesolithic Moita do Sebastião from Portugal and Iberomaurusian HG from Taforalt, Morocco, North Africa. Taforalt individuals are a good proxy to test the African-Iberian connections due to the genetic continuity attested in North Africa from the Late Pleistocene to the Holocene (Early Neolithic) despite their chronologically older age [S3]; errors bars indicate ±3 standard errors **C)** Correlation between Villabruna-like ancestry (dark blue; Figure 3D) and Loschbour-like ancestry (light blue; Figure S3A) and time (error bars indicate the radiocarbon 2-sigma range). Both models result in a fit of R = 0.99 for individuals from north and northeast of Iberia Peninsula (to the exclusion of Chan and Moita do Sebastião), where we observe an increase of WHG-like ancestry similar to other parts of Europe.

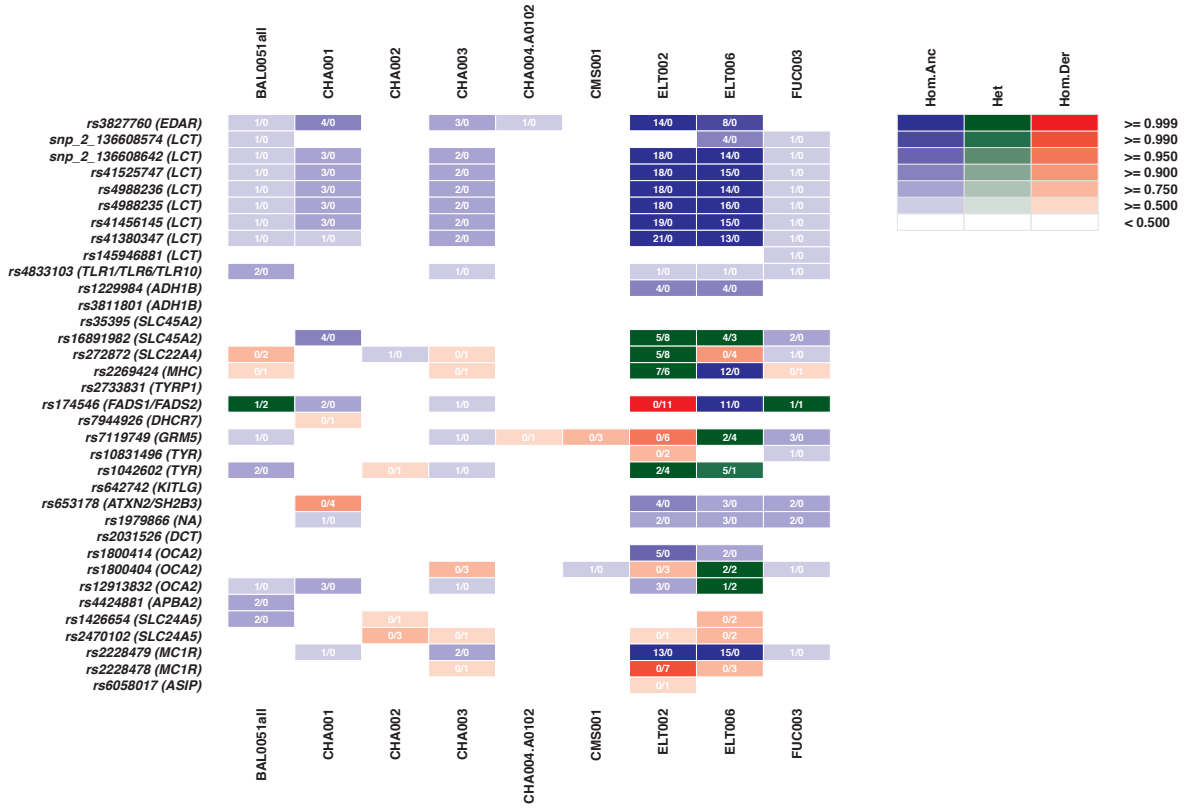


Figure S4. Summary of genotypes of phenotypic and functional SNPs, related to STAR methods. Colours indicate the homozygous ancestral/derived or heterozygous state of the SNPs reported in the left-hand column. Numbers in cells indicate the number of reads matching the ancestral/derived allele.

Supplemental References

S1. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., *et al.* (2016). The genetic history of Ice Age Europe. Nature *534*, 200-205.

S2. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., *et al.* (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419-424.

S3. Fregel, R., Méndez, F.L., Bokbot, Y., Martín-Socas, D., Camalich-Massieu, M.D., Santana, J., Morales, J., Ávila-Arcos, M.C., Underhill, P.A., Shapiro, B., *et al.* (2018). Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.1800851115

S4. Posth, C., Renaud, G., Mittnik, A., Drucker, D.G., Rougier, H., Cupillard, C., Valentin, F., Thevenet, C., Furtwängler, A., Wißing, C., et al. (2016). Current Biology *26*, 827-833.

Supplementary Text

Genomic and dietary transitions during the Mesolithic and Early Neolithic in Sicily

Marieke S. van de Loosdrecht,[1*] Marcello A. Mannino,[2,3*]† Sahra Talamo,[3,4] Vanessa Villalba-Mouco,[1] Cosimo Posth,[1] Franziska Aron,[1] Guido Brandt,[1] Marta Burri,[1] Cäcilia Freund,[1] Rita Radzeviciute,[1] Raphaela Stahl,[1] Antje Wissgott,[1] Lysann Klausnitzer,[3] Sarah Nagel,[5] Matthias Meyer,[5] Antonio Tagliacozzo,[6] Marcello Piperno,[7] Sebastiano Tusa,[8] Carmine Collina,[9] Vittoria Schimmenti,[10] Rosaria Di Salvo,[10] Kay Prüfer,[1,5] Jean-Jacques Hublin,[3,11] Stephan Schiffels,[1] Choongwon Jeong,[1,12] Wolfgang Haak,[1]† Johannes Krause[1*]†

SECTIONS

S1: Grotta dell'Uzzo: archaeology and stratigraphic sequence

1. The site, its burial ground and human remains

Grotta dell'Uzzo is a large shelter-like cave located in northwestern Sicily, along the eastern cliffs of the San Vito lo Capo peninsula (fig. S1.1). The site was visited hastily in 1927 by the French archaeologist Raymond Vaufrey, who did not realize its importance. The discovery of the deposit and its stratification was made in the early 1970s by Giovanni Mannino (*111*), who excavated a small test trench in the cave, exposing a sequence of *in situ* Mesolithic deposit (identified as *epipaleolitico*). Prehistoric deposits have been excavated during the 1970s, 1980s and in 2004 within a number of trenches both inside and outside the overhang of the cave (fig. S1.2). This revealed that the site was occupied from at least the late Upper Palaeolithic through the Mesolithic and into the Neolithic (*14, 112-115*). The cave was also occupied during the Bronze Age and throughout history, and until recently used by shepherds as a stable for sheep.

The main reason why Grotta dell'Uzzo is a key site for Mediterranean prehistory is that its long stratigraphic sequence covers the transition from hunter-gatherer to agro-pastoral economies (*12-14, 59, 60, 115*). It is one of few such sites, given that the number of sites in the Mediterranean with sequences from the late Mesolithic to the early Neolithic is rare, probably as a consequence of a decrease in hunter-gatherer populations at the end of the Mesolithic (*67*).

fig. S1.1. View of Grotta dell'Uzzo from the sea (photo by Marcello A. Mannino)

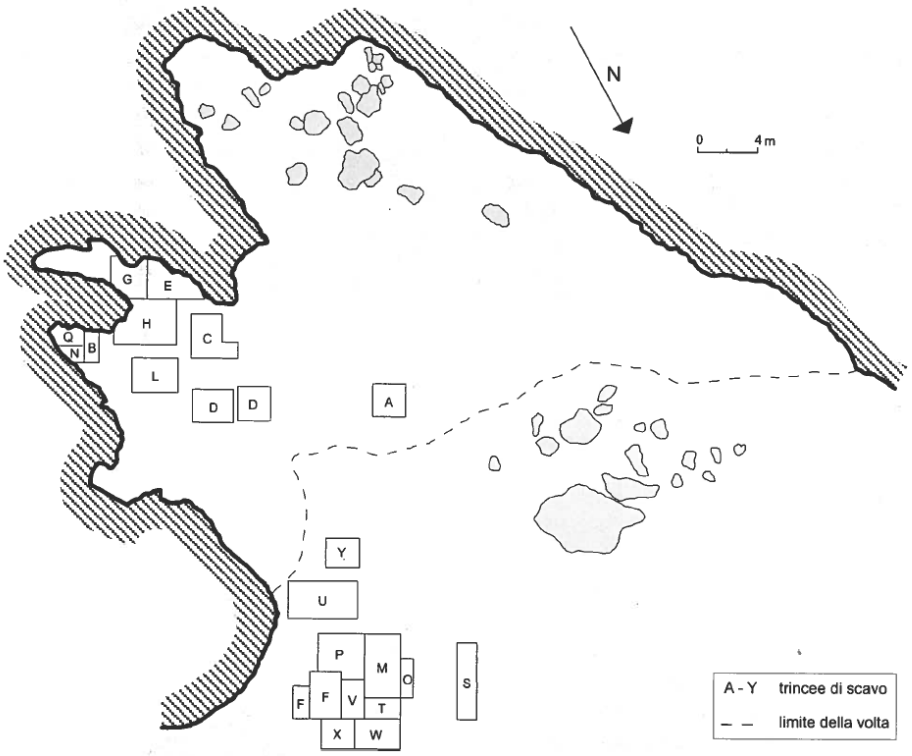


fig. S1.2. Plan of Grotta dell'Uzzo with the trenches excavated in the 1970s and 1980s (from: (*14*))

Another important feature of this cave site is that during the Mesolithic it was used as a burial ground. A total of 11 burials and 13 inhumated individuals (six males, four females and three infants) have been recovered at Grotta dell'Uzzo in the course of excavations in the 1970s, 1980s and 2004 close to the walls of the 'inner part' of the cave (*112, 116, 117*). Studies on the pathologies of the inhumated humans established that plant foods were an important component of the diet of the Mesolithic hunter-gatherers (*118*). On the other hand, isotopic and zooarchaeological investigations show that the occupants of Grotta dell'Uzzo relied heavily on animal protein, which through time originated increasingly from marine ecosystems (*12, 14*).

Human remains at the cave were, however, also found scattered through the deposits of Grotta dell'Uzzo. Radiocarbon dating ascertained that these remains were not only Mesolithic but dated to all the main phases of cave occupation, including the so-called Mesolithic-Neolithic transition phase and Neolithic phases (*12*). As part of that same study 70 human bones were sampled, of which 57 recovered from the burials and 13 commingled within the deposits. In total only 33 bones yielded collagen extracts and 10 of these were from the bones recovered outside of the burials. Only 40% of the bones from the burials yielded collagen extracts and not all of these met the quality criteria established by van Klinken (*110*), which is indicative of the poor state of preservation of the human skeletal remains from the burials (*12*). On the other hand, 77% of the commingled bones yielded collagen extracts, all of which are well preserved. For this reason, and because our aim was to obtain genetic and further isotopic information on the main periods of cave occupation, we decided to target the loose human remains, most of which have been directly dated within the remit of this project.

2. Cultural succession at Grotta dell'Uzzo from the Mesolithic to Early Neolithic

Mesolithic

The lithic industries from the two oldest two phases of the Mesolithic at Grotta dell'Uzzo have not been studied in detail, but they are contemporary to the occurrence in Sicily of facies of Epigravettian tradition across the island and the Undifferentiated Epipalaeolithic in the east, followed by Sauveterrian-like facies (*68*). In north-western Sicily microlithic industries of Epigravettian tradition (labelled as *Epigravettiano indifferenziato*) have been identified at Grotta dell'Uzzo in the Mesolithic phases I and II (*119*), as well as at Grotta dell'Isolidda (*68*) and Grotta di Cala Mancina (*120*) on the western coast of the San Vito lo Capo peninsula. These industries demonstrate strong techno-cultural affinities between the Late Epigravettian and early Holocene hunter-gatherers of Sicily. On the other hand, Sauveterrian industries have not been clearly identified at Grotta dell'Uzzo, but Sauveterrian-like facies have been retrieved from the nearby site of Grotta dell'Isolidda (*68*) and at the site at the westernmost end of Sicily on the island of Favignana at Grotta d'Oriente (*120*).

Levels 14 to 11 in both Trench F and Trench M have previously been defined as the so-called 'Mesolithic-Neolithic transition phase' (e.g. (*12-14, 59, 113, 114, 121, 122*), because this was an essentially Mesolithic phase with some Neolithic elements in its upper spits. A recent study of the lithic industry from these layers attributes this phase of cave occupation to the blade-and-trapeze techno-complex of the western Mediterranean Castelnovian tradition (*69*). The oldest date available for the lowermost spits of this phase obtained on charcoal from spits 14 and 13 of Trench F attributes this part to 7,000-6,590 calBCE ((*58*); P-2734, 7,910 ± 70 BP), which is one of the oldest chronological attributions for a blade-and-trapeze industry (Castelnovian *sensu lato*). The most recent reassessment of the radiocarbon chronology for Grotta dell'Uzzo, based on Bayesian modelling of the sequence of dates available for Trench F, suggests that the phase associated with the Castelnovian facies may have spanned from around 6,770 to 5,850 yrs calBCE (~8,770-7,850 yrs calBP (*12*)). The following phase in chronological continuity is the Neolithic phase I, which according to the above-mentioned Bayesian model may have spanned from around 6,050 to 5,400 yrs calBCE (~8,050-7,400 yrs cal BP; (*12*).

The blade-and-trapeze Castelnovian (*sensu lato*) complex of the VII millennium BCE is in techno-economic continuity with the Neolithic complexes of the archaic Impressed Ware and Stentinello culture of the VI millennium BCE. The production of trapezes constitutes the defining element of the lithic techno-complexes between the VII and VI millennia BCE. This was achieved through a notable standardization of the production processes, particularly through the application of pressure by different modalities (*69*). Nevertheless, the variability in some technical behaviours (e.g. bladelet fracturing techniques, presence/absence of the microburin technique, façonage processes of the trapeze truncations) is linked with a break and discontinuity in the Mesolithic-Neolithic technical traditions (*69*).

Early Neolithic

The early Neolithic in Sicily has been defined based on sites in the western part of the island (i.e. Grotta dell'Uzzo, Grotta del Kronio) and is characterized by three main cultural horizons, which in chronological order are: 'Archaic Impressed Ware' (*ceramiche impresse arcaiche*), 'Advanced Impressed Ware' (*ceramiche impresse evolute*) of facies Stentinello I and 'Advanced Impressed Ware' (*ceramiche impresse evolute*) of facies Stentinello I (*8*). The chronology of these horizons is largely based on the dating at Grotta dell'Uzzo, which for this part of the sequence does not see full consensus between the different scholars who worked on the site, depending on whether the beginning of the Neolithic is taken to

coincide with Spit 12 of Trench F, as proposed by Tiné and Tusa (*80*), or with Spit 10 of Trench F, as proposed by Tagliacozzo (*14*) and Collina (*69*).

3. Human remains sampled: cultural attribution and chronology

The interior of the cave

The first trench excavated at the site was Trench A, which was located within the overhang of the cave. This part of the deposit includes the oldest levels of occupation, which can be attributed to the Late Epigravettian. These are covered by Mesolithic layers, which have been divided into two horizons. The more precise chrono-typological attribution has not been refined due to the partial study of the lithic industries. Horizon 1, the most recent Mesolithic horizon, is characterized by the presence of a specific type of scraper with a reduced front adjacent to a deep laterally-retouched indentation (*123*). This tool is associated to a laminar industry with geometrics, represented by triangles, circular segments and rare trapezes. These characters are not present in Horizon 2, the oldest of the two Mesolithic horizons, in which the geometrics are rarer and replaced by a lithic industry with less differentiated (or more undifferentiated) characters. Below Horizon 2, in a different sedimentary layer, terminal Upper Palaeolithic finds have been recovered. Most trenches within the overhang of the cave (including Trench H) contained deposits of Mesolithic age, given that the Neolithic layers had been removed within it.

We sampled the following individuals from within the overhang of Grotta dell'Uzzo:

- *UZZ26.cont*: cranial fragment retrieved from Spit 8 in Trench A, can be attributed to the phase characterized by microlithic industries of Epigravettian tradition (Mesolithic I phase I). DNA and collagen were extracted from this specimen, but DNA preservation was insufficient to include this individual in our genetic analyses. The direct ^{14}C date on this individual is 9,436 ± 36 BP (8,810-8,620 calBCE).
- *UZZ61*: phalanx retrieved from the topsoil layer of Trench H, for which we analysed the ancient DNA and collagen. Direct ^{14}C date: TBA.
  Given the risk of post-depositional disturbance in top soil layers the stratigraphical position for this individual could not be used for a reliable archaeological assignment. However, since this individual has both a genetic profile and isotopic compositon which we consider this individual compatible with the Stentinello culture.

The deposits outside the cave: Trench F and M

Most of the human skeletal remains selected for our genetic and isotopic investigation (table S1.1) originate from trenches beyond the overhang of the cave (fig. S1.2). The outside of the cave, and in particular trenches F and M (*69*), contained thick stratigraphic sequences spanning through the Mesolithic and up to the pre-Stentinello (Impressed Ware) and Stentinello Neolithic phases. Trench F contained a Mesolithic deposit of 1,50m in thickness (*113*), overlain by a similarly thick deposit that in chronological order included the so-called 'Mesolithic-Neolithic transition phase' and two Neolithic phases. A recent study of the lithic industries from the transitional layers shows that this phase was associated with the blade-and-trapeze complex attributable to the Castelnovian *sensu lato* (*69*).

The sequence from Trench F is the reference stratigraphy for Grotta dell'Uzzo. The original stratigraphic scheme was proposed by Tagliacozzo (*14*). Based on the findings of Collina (*69*) and on-going investigations, we here classified the different phases as follows, using the age ranges for each phase as generated by Mannino et al. (*12*):

- Basal stratum / Late Upper Palaeolithic (Spits 48-33): Late Epigravettian
- Mesolithic I, phase I (Spits 32-23): Industries of Epigravettian tradition
- Mesolitihic I, phase II (Spits 22-15): (~11,100-8,500 yrs calBP)
- Mesolithic II (Spits 14-11): Castelnovian facies *sensu lato* (~8,770-7,850 yrs calBP)
- Neolithic phase I (Spits 10-6): Impressed Ware horizon (~8,050-7,400 yrs calBP)
- Neolithic phase II (Spits 5-1): Stentinello horizon (~7,520-7,130 yrs calBP)

Here, we present genomic and isotope data for the following individuals from the F-trench:

- *UZZ33:* tooth retrieved from Spit 4. Although we we could not obtain a direct ^{14}C date, this individual was found in the same layer as *UZZ34* that was directly dated.
- *UZZ34*: tooth retrieved from Spit 4, which corresponds to Neolithic I phase II Stentinello. Direct ^{14}C date: 6,351±24 BP, 5,470-5,230 calBCE.
  For both *UZZ33* and *UZZ34* the archaeological contextual attribution and ancestry profile are consistent with them being early farmers, most likely from a Stentinello horizon.
- *UZZ40*: tooth of *infans* retrieved from Spit 13, which corresponds to Mesolithic II, Castelnovian *sensu lato*. Direct ^{14}C date: 7,471±26 BP, 6,420-6,250 calBCE. This confirms the attribution based on stratigraphic and archaeological observations.
- *UZZ4446*: retrieved from Spit 15, which is stratigraphically assigned to the Mesolithic I phase I. For this individual DNA was extracted from two teeth (skeletal elements UZZ44 and –45) and a mandible fragment (element UZZ46). Collagen was extracted from element UZZ45. Direct ^{14}C date: 7,713±26 BP, 6,500-6,250 calBCE. The calibrated age corrected for a marine dietary contribution of 40±10%, attributes this specimen to the Mesolithic II and not to the Mesolithic I phase II. In relation to some cetacean bones, it is possible that materials moved post-depositionally down the sequence from the layer immediately above (Spits 14-11) into Spits 15 and 16 (*12*).
- *UZZ5054*: retrieved from Spits 19 and 20, which correspond to Mesolithic I phase II. DNA was extracted from five different teeth (skeletal elements UZZ50-54). We obtained a direct ^{14}C date on element UZZ51: 9,436±29 BP, 8,790-8,630 calBCE. This confirms the attribution based on stratigraphic and archaeological observations.

Trench M had a very similar stratigraphic sequence to Trench F, albeit including only four phases of cave occupation:

- Mesolithic I, phase II (Spits 18-15):
- Mesolithic II (Spits 14-11): Castelnovian facies *sensu lato* (blade-and-trapeze complex)
- Neolithic phase I (Spits 10-7): Impressed Ware culture
- Neolithic phase II (Spits 6-1): Stentinello culture

From the M trench, we present genomic and isotope data for the following individuals:

- *UZZ69*: for who we sampled a mandible retrieved from Spit 3, which corresponds to Neolithic I Stentinello. Direct ^{14}C date: 7,848±26 BP, 6,630-6,390 calBCE.
- *UZZ71,* for who we sampled a tooth retrieved from Spit 10, which corresponds to the Neolithic I Impressa. This individual shows an ancestry profile chacteristic for the individuals associated with the Mesolithic II Castelnovian lithic industry. Direct ^{14}C date: 7,127±25 BP, 6,060-5,920 calBCE.

The lack of an adequate freshwater isotopic baseline for Grotta dell'Uzzo, and the possibility that this individual may not be local, complicate issues linked to accurate reservoir correction. We have, thus, only calibrated the radiocarbon date for *UZZ71*, but not corrected the calibrated age for possible freshwater effects.

Trenches S, T, U, W and burial 8

As discussed above, only few trenches from Grotta dell'Uzzo have been studied and dated more in detail (i.e. trenches A, F and M). For this reason, and because radiocarbon dating has not been applied much on materials from other trenches, the stratigraphically and archaeologically based attribution to cultural phase hinges on observations recorded in the excavation notebooks. However, we have radiocarbon dated almost all the specimens from these poorly-studied trenches, so that their calibrated ages can be related to the stratigraphical and chronological 'master sequence' published for Trench F (*12*).

From Trench S, we present genomic and isotope data for two individuals:

- *UZZ74*: femur retrieved from Spit 5. Direct ^{14}C date: 6,310±23 BP, 5,330-5,210 calBCE. The direct ^{14}C date, isotopic and ancestry profile are consistent with this individual being an early farmer, most likely from a Stentinello horizon.
- *UZZ75*: petrous bone retrieved from Spit 15. Direct ^{14}C date: 6,310±23 BP, 5,330-5,210 calBCE. The direct ^{14}C date, isotopic and ancestry profile are consistent with this individual being an early farmer, most likely from a Stentinello horizon.

Although *UZZ74* and *UZZ75* have identical ^{14}C dates, we can exclude that these individuals are genetic identicals or have a kinship relation.

From Trench T, we present genomic data for one individual:

- *UZZ77*: tooth, which based on the excavation notebooks and finds recovered from its spit of origin (Spit 13) can be attributed the Neolithic I Impressed Ware horizon. Direct ^{14}C date: TBA.

From Trench W, we present genomic and isotope data for two individuals:

- *UZZ87*: humerus retrieved from Spit 2. Direct ^{14}C date: 6,286±24 BP, 5,320-5,210 calBCE. The archaeological contextual attribution, direct radiocarbon date, isotopic and ancestry profile are consistent with this individual being an early farmer, most likely from a Stentinello horizon.
- *UZZ88*: phalanx retrieved from Spit 14. Direct ^{14}C date: 7,036±25 BP, 6,000-5,840 calBCE. The radiocarbon date is in line with the contextual attribution, both indicating that this individual dates to the Neolithic I Impressed Ware horizon.

From the Mesolithic burial VIII, we present genomic and isotope data for one individual:

- *UZZ96*: molar (M2). This specimen originates from one of the inhumations, which all date to the Mesolithic I (*12*).
  The genetic profile and stratigraphy date are consistent with this individual being a hunter-gatherer from the Early Mesolithic, most likely from the facies Mesolithic I, phase II.

In addition, four individuals were retrieved from the top soil layer in Trench U for which a reliable archaeological assignment could not be ascertained on stratigraphic grounds. However, their direct ^{14}C dates, genetic ancestry and isotopic profiles are consistent with them being from the Mesolithic II

Castelnovian facies *sensu lato*. We found genetic evidence for a second-degree kinship relation between *UZZ79* (genetic female) and *UZZ81* (genetic male) (Extended Data Table-4).

- *UZZ79*: petrous bone. Direct ^{14}C date: 7,809±26 BP, 6,600-6,350 calBCE.
- *UZZ80*: petrous bone. Direct ^{14}C date: 7,809±26 BP, 6,600-6,350 calBCE.
- *UZZ81*: temporal bone fragment. Direct ^{14}C date: 7,807±26 BP, 6,600-6,380 calBCE.
- *UZZ82*: petrous bone. Direct ^{14}C date: 7,729±26 BP, 6,630-6,480 calBCE

Stratigraphic position	Genetic ID	Cultural phase	Skeletal element
A-8	UZZ26.cont	Mesolithic I phase I	cranial
F-4	UZZ33	Early Neolithic Stentinello	tooth
F-4	UZZ34	Early Neolithic Stentinello	tooth
F-13	UZZ40	Mesolithic II Castelnovian	tooth (infans)
F-15	UZZ4446	Mesolithic II Castelnovian	tooth(2x)/mandible
F-19/20	UZZ5054	Mesolithic I phase II	tooth (5x)
H-rim	UZZ61	Early Neolithic Stentinello	phalanx
M-3	UZZ69	Mesolithic II Castelnovian	mandible
M-7	UZZ71	Early Neolithic Impressa?	tooth
S-rim	UZZ74	Early Neolithic Stentinello	femur
S-5	UZZ75	Early Neolithic Stentinello	temporal/petrous
T-13	UZZ77	Early Neolithic Impressa?	tooth
U-rim	UZZ79	Mesolithic II Castelnovian	temporal/petrous
U-rim	UZZ80	Mesolithic II Castelnovian	temporal/petrous
U-rim	UZZ81	Mesolithic II Castelnovian	temporal fr
U-rim	UZZ82	Mesolithic II Castelnovian	temporal/petrous
W-2	UZZ87	Early Neolithic Stentinello	humerus
W-14	UZZ88	Early Neolithic Impressa?	phalanx
burial VIII	UZZ96	Mesolithic I phase I	M2 upper right

table S1.1. Cultural affiliation of the human remains investigated in this study. This table lists the attribution to cultural phase based on stratigraphic and archaeological grounds. The attribution of samples that have not been clearly assigned to a cultural phase on archaeological grounds is briefly treated in the specimen by specimen list above. In the case of Trench U, although all specimens come from the 'topsoil layer', we propose an attribution to the Mesolithic II phase (Castelnovian *sensu lato*) based on a radiocarbon date a delphinid specimen from this part of the deposit ((*12*); 8,083±26 BP: 6,780-6,350 calBCE).

Individual ID	R-EVA	MAMS	Trench-Spit	Phase	^{14}C date BP	Calibrated age BP 2σ	Calibrated age BC 2σ
UZZ26.cont	1918	40708	A-8	MESO1/1	9436±36	10760-10570	8810-8620
UZZ5054	1935	40710	F-19	MESO1/2	9436±29	10740-10580	8790-8630
UZZ82	1960	40722	U-rim	MESO2	7809±26	8650-8540	6690-6590
UZZ69	1948	40711	M-3	MESO2	7848±26	8580-8340	6630-6390
UZZ79	1957	40719	U-rim	MESO2	7809±26	8550-8300	6600-6350
UZZ80	1958	40720	U-rim	MESO2	7809±26	8550-8300	6600-6350
UZZ81	1959	40721	U-rim	MESO2	7807±26	8550-8330	6600-6380
UZZ82	1960	40722	U-rim	MESO2	7,729±26	8650-8540	6630-6480
UZZ4446	1930	40709	F-15	MESO2	7713±26	8450-8200	6500-6250
UZZ40	2880	40726	F-13	MESO2	7471±26	8370-8200	6420-6250
UZZ71	1950	43967	M-10	NEO1/1	7127±25	8010-7870	6060-5920
UZZ88	1965	40712	W-14	NEO1/1	7036±25	7940-7790	6000-5840
UZZ34	2879	40725	F-4	NEO1/2	6351±24	7420-7170	5470-5230
UZZ74	1953	40716	S-rim	NEO1/2	6310±23	7280-7160	5330-5210
UZZ75	1954	40717	S-5	NEO1/2	6310±23	7280-7160	5330-5210
UZZ87	1964	40723	W-2	NEO1/2	6286±24	7260-7160	5320-5210

table S1.2. Radiocarbon dates, calibrated and corrected ages of humans from Grotta dell'Uzzo. The AMS radiocarbon dates reported in this table were performed at the Klaus Tschira Laboratory of the Curt-Engelhorn-Zentrum Archaeometrie in Mannheim (MAMS). Dates were calibrated with the OxCal 4.2 software (*124*) using the IntCal13 curve and, in addition, the Marine13 curve for individuals that had consumed marine protein (*125*). The estimation of the amount of marine protein consumed is based on calculations made for specimen S-EVA 8010 (40±10% marine) by (*12*). The individuals for which a correction was necessary are *UZZ4446* (40±10% marine), *UZZ81* (45±10% marine), *UZZ69*, *UZZ79* and *UZZ80* (50±10% marine). Corrections were made using the reservoir correction estimated for the Mediterranean Basin by Reimer and McCormac (*109*), which is $\Delta R = 58 \pm 85$ ^{14}C yr. This table is useful to relate the dates from (*12*), which are calibrated BP, with those produced for the present study that are calibrated BCE.

S2. Genetic grouping and substructure of the ancient Sicilians

Here, we aimed to investigate genetic substructure among the ancient Sicilians and whether individuals could be grouped for genetic analyses. We co-analyzed an Epigravettian HG from OrienteC (*I2158* (*15*)).

1. Three genetic groups

First, we used f_3-outgroup statistics of the form f_3*(Mbuti; X1, X2)* for all individual pairs to quantify their levels of shared genetic drift for SNPs ascertained in Mbuti. We found that the ancient Sicilians in our transect form three major genetic groups that are characterized by elevated levels of shared genetic drift for within-group compared to between-group individual pairs (fig. S2.1). Alternatively, taking an approach presented by Lazaridis et al. (*56*), we used *qpWave (54)* to test for all possible pairs of individuals whether their gene pools are consistent with being derived from one ancestry stream (N = 1) with regard to set of *Outgroups* (table S2.1). If one ancestry stream suffices the model has full rank and the null-hypothesis can not be rejected, hence f_4*(Left1, OG1; OG2, OG3)* = f_4*(Left2, OG1; OG2, OG3)* for all triplets. If the latter is true, we assumed that the individuals in the *Left* pair are symmetrically related to the specific combination of *Outgroups* used, and therefore can be grouped for analyses. We found that the qpWave-based cladality models distinguished the same three genetic groups as with the pairwise f_3-outgroup statistics.

One group contains the two oldest Mesolithic individuals from Uzzo, *UZZ5054* and *UZZ96* (~8,800-8,630 calBCE) and the Epigravettian *OrienteC* HG (12,250-11,850 calBCE (*15, 17*)). These three individuals carried mitogenome lineages that fall within the U2'3'4'7'8'9 branch (Supplementary Section S7, and (*15, 17*) for *OrienteC*). We labelled this group as **Sicily Early Mesolithic (Sicily EM)**.

A second group contains nine individuals dated to ~6,750-5,850 calBCE. The seven oldest individuals in this group (dated ~6,750-6,250 calBCE) are tentatively assigned to the Mesolithic II Castelnovian archaeological context (Supplementary Section S1). Notably, the two youngest individuals in this group (*UZZ71* and *UZZ88*, dated ~6,050-5,850 calBCE) chronologically coincide with layers at the site that may contain the very first aspects of Impressa Wares (Supplementary Section S1). These two individuals fall fully within the genetic diversity of the other individuals in the Sicily LM group, despite postdating them by ~200 years. The mitogenome haplogroups carried by the all these individuals are typical for European Late Mesolithic WHGs (Supplementary Section S7). We labelled this group as **Sicily Late Mesolithic (Sicily LM)**. Notably, for some individual pairs in this group the qpWave cladality model for one ancestry stream is rejected ($P<0.1$) (table S3.1). This implies that the Sicily LM HGs form a heterogenous group with possibly additional underlying substructure.

The third and most recent genetic group contains seven individuals dated to ~5,460-5,220 calBCE. Six individuals in this group are from layers that chronologically coincide with the presence of Early Neolithic Stentinello Ware, and one individual (*UZZ77*, undated) tentatively with older aspects of Impressa Ware (Supplementary Section S1). All the individuals in the Sicily EN group carried mitogenome haplogroups characteristic for European early farmers (Supplementary Section S7). We labelled this group as **Sicily Early Neolithic (Sicily EN)**.

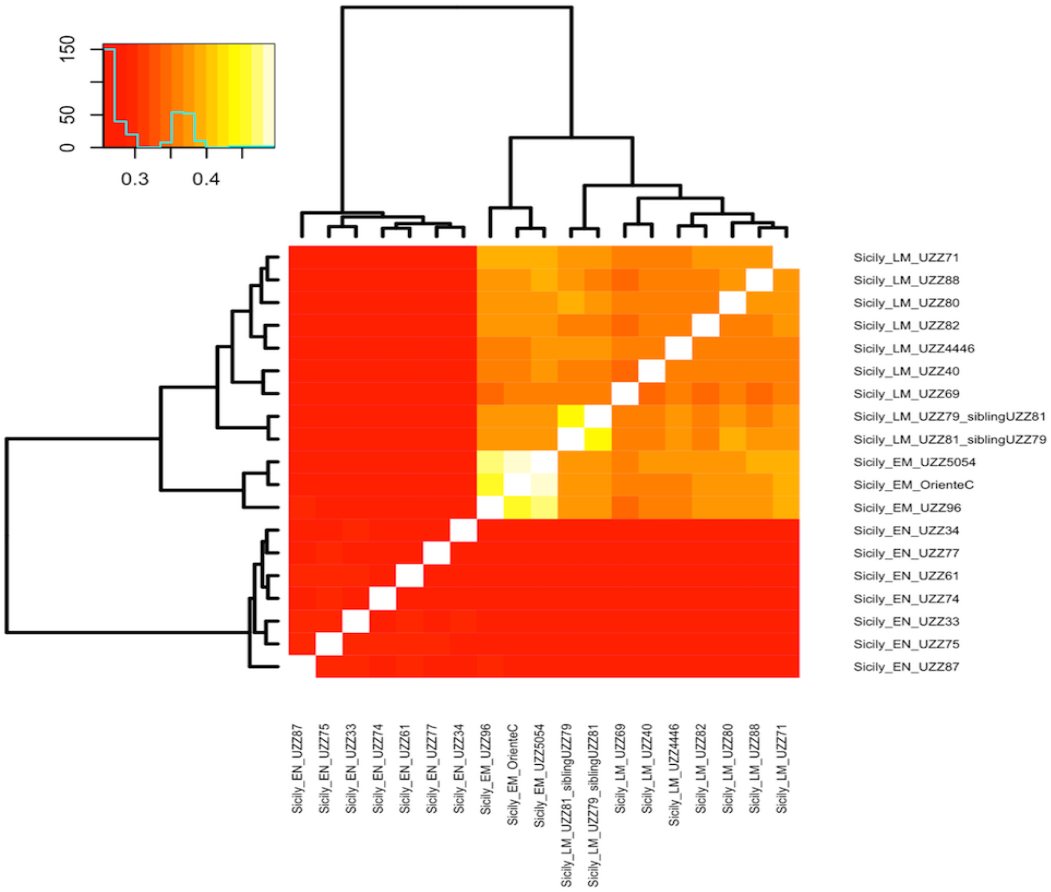


fig. S2.1. Heat map showing results for *f_3(Mbuti; X1, X2)* for all pairwise comparisons between the ancient individuals from Uzzo and one HG from OrienteC. Larger positive f_3-values indicate higher similarity in shared genetic covariance, hence stronger degrees of genetic relatedness between individuals. Three genetic groups appear. Notably, *UZZ5054* and *UZZ96* post-date *OrienteC* by ~3,250-3,450 years yet share very high levels of genetic drift (f_3*(Mbuti; OrienteC, UZZ5054)* = 0.495, f_3*(Mbuti; OrienteC, UZZ96)* = 0.461, and f_3*(Mbuti; UZZ5054, UZZ96)* = 0.465). The higher levels of shared genetic drift between the Sicily LM HGs UZZ79 and UZZ81 are the result of a direct kinship relation (see Extended Data Table-4).

	OrienteC	UZZ5054	UZZ096	UZZ4446	UZZ40	UZZ69	UZZ79	UZZ80	UZZ81	UZZ82	UZZ71	UZZ88	UZZ77	UZZ33	UZZ34	UZZ61	UZZ74	UZZ75	UZZ87
Genetic group	Sicily EM	Sicily EM		Sicily LM	Sicily LM	Sicily LM	Sicily LM	Sicily LM	Sicily LM	Sicily LM	Sicily LM	Sicily LM	Sicily EN	Sicily EN	Sicily EN	Sicily EN	Sicily EN	Sicily EN	Sicily EN
C14 date																			
# SNPs																			
OrienteC	NR	0.5660	0.9386	1.18E-15	1.37E-08	3.72E-18	1.52E-08	2.50E-12	9.03E-06	6.24E-10	2.46E-05	6.53E-09	2.36E-72	2.56E-128	3.04E-92	3.24E-164	7.82E-89	4.75E-178	6.72E-41
UZZ5054	0.5660	NR	0.4122	1.76E-18	2.81E-13	4.18E-31	3.15E-11	5.41E-16	1.48E-12	4.37E-11	9.03E-08	1.12E-08	7.68E-122	5.04E-197	2.26E-171	2.09E-214	8.32E-152	6.52E-228	3.48E-74
UZZ096	0.9386	0.4122	NR	0.0201	0.5583	1.37E-04	0.6865	0.0281	0.3366	3.48E-03	0.8156	0.7681	2.13E-22	6.99E-44	2.03E-26	1.09E-51	2.72E-32	1.74E-54	5.92E-10
UZZ4446	1.18E-15	1.76E-18	0.0201	NR	0.8770	0.0333	0.1123	0.1407	0.6736	0.0803	0.0812	0.0555	6.40E-56	1.19E-112	1.21E-95	2.46E-122	1.18E-71	8.30E-124	3.27E-33
UZZ40	1.37E-08	2.81E-13	0.5583	0.8770	NR	8.67E-03	0.0589	0.4519	0.8869	0.1467	0.0110	0.4252	9.03E-40	6.03E-84	1.62E-71	1.50E-104	9.15E-38	8.24E-99	1.75E-17
UZZ69	3.72E-18	4.18E-31	1.37E-04	0.0333	8.67E-03	NR	1.48E-04	2.41E-03	0.1819	3.41E-04	8.38E-09	1.01E-07	1.16E-67	1.53E-109	7.02E-81	4.33E-126	6.50E-85	1.81E-126	5.80E-35
UZZ79	1.52E-08	3.15E-11	0.6865	0.1123	0.0589	1.48E-04	NR	0.4076	0.9433	0.6787	0.8609	0.8865	8.34E-78	1.24E-137	9.21E-133	1.36E-162	5.14E-103	3.77E-162	5.80E-50
UZZ80	2.50E-12	5.41E-16	0.0281	0.1407	0.4519	2.41E-03	0.4076	NR	0.7294	0.9650	0.0705	0.4134	9.98E-90	1.17E-148	1.18E-123	1.13E-163	2.11E-94	4.68E-170	3.38E-50
UZZ81	9.03E-06	1.48E-12	0.3366	0.6736	0.8869	0.1819	0.9433	0.7294	NR	0.8256	0.6475	0.7192	7.99E-51	3.96E-93	9.51E-76	2.18E-123	1.05E-70	6.61E-118	3.40E-28
UZZ82	6.24E-10	4.37E-11	3.48E-03	0.0803	0.1467	3.41E-04	0.6787	0.9650	0.8256	NR	0.0958	0.6946	7.74E-82	1.85E-157	1.71E-126	2.81E-161	4.50E-106	8.55E-165	7.03E-44
UZZ71	2.46E-05	9.03E-08	0.8156	0.0812	0.0110	8.38E-09	0.8609	0.0705	0.6475	0.0958	NR	0.9360	1.11E-94	4.68E-146	2.09E-115	1.54E-157	3.37E-108	3.20E-186	2.41E-46
UZZ88	6.53E-09	1.12E-08	0.7681	0.0555	0.4252	1.01E-07	0.8865	0.4134	0.7192	0.6946	0.9360	NR	9.92E-76	8.46E-143	7.05E-125	2.57E-160	5.66E-96	1.24E-165	3.01E-41
UZZ77	2.36E-72	7.68E-122	2.13E-22	6.40E-56	9.03E-40	1.16E-67	8.34E-78	9.98E-90	7.99E-51	7.74E-82	1.11E-94	9.92E-76	NR	0.6138	0.0096	0.7563	0.9473	0.3566	0.7585
UZZ33	2.56E-128	5.04E-197	6.99E-44	1.19E-112	6.03E-84	1.53E-109	1.24E-137	1.17E-148	3.96E-93	1.85E-157	4.68E-146	8.46E-143	0.6138	NR	0.2600	0.1701	0.6806	0.5095	0.4760
UZZ34	3.04E-92	2.26E-171	2.03E-26	1.21E-95	1.62E-71	7.02E-81	9.21E-133	1.18E-123	9.51E-76	1.71E-126	2.09E-115	7.05E-125	0.0096	0.2600	NR	0.6946	0.2384	0.5425	0.4452
UZZ61	3.24E-164	2.09E-214	1.09E-51	2.46E-122	1.50E-104	4.33E-126	1.36E-162	1.13E-163	2.18E-123	2.81E-161	1.54E-157	2.57E-160	0.7563	0.1701	0.6946	NR	0.6973	0.2914	0.9762
UZZ74	7.82E-89	8.32E-152	2.72E-32	1.18E-71	9.15E-38	6.50E-85	5.14E-103	2.11E-94	1.05E-70	4.50E-106	3.37E-108	5.66E-96	0.9473	0.6806	0.2384	0.6973	NR	0.8842	0.0837
UZZ75	4.75E-178	6.52E-228	1.74E-54	8.30E-124	8.24E-99	1.81E-126	3.77E-162	4.68E-170	6.61E-118	8.55E-165	3.20E-186	1.24E-165	0.3566	0.5095	0.5425	0.2914	0.8842	NR	0.0263
UZZ87	6.72E-41	3.48E-74	5.92E-10	3.27E-33	1.75E-17	5.80E-35	5.80E-50	3.38E-50	3.40E-28	7.03E-44	2.41E-46	3.01E-41	0.7585	0.4760	0.4452	0.9762	0.0837	0.0263	NR

table S2.1. Ancestry similarity matrix for all individual pairs. Results are from qpWave-based cladality models. We used an *Outgroup* set from Mathieson et al. (*17*): *El Miron*, *Mota*, Mbuti, *Ust Ishim*, *Mal'ta*, *AfontovaGora3*, *GoyetQ116*, *Villabruna*, *Kostenki14*, *Vestonice16*, Karitiana, Papuan, Onge. Individuals in red have < 150k SNPs covered. Since missing data may inflate the P-values for this test, we required a test result to be smaller (less extreme) than P = 0.1 in order to reject the null-hypothesis of cladality. Models that provide a full ancestry fit (significance threshold: $P \geq 0.1$) are highlighted in green, those that approach the boundaries of the model ($0.01 < P < 0.1$) are in orange, and those for which a full ancestry fit can be rejected ($P < 0.01$) are in red. We found three genetic groups (boxes) that we labeled as Sicily EM (*OrienteC*, *UZZ5054*), Sicily LM (*UZZ4446*, *UZZ40*, *UZZ69*, *UZZ79*, *UZZ80*, *UZZ81*, *UZZ82*, *UZZ71*, *UZZ88*) and Sicily EN (*UZZ77*, *UZZ33*, *UZZ34*, *UZZ61*, *UZZ74*, *UZZ75*, *UZZ87*). The Sicily LM HGs form a heterogenous group with possible additional underlying substructure. *UZZ96* shows a high smilarity to individuals in both the Sicily EM and Sicily LM genetic group, congruent to its position in the MDS plot (Fig. 2A).

S3. Elevated lineage-specific genetic drift in the Sicilian Early Mesolithic HGs

Nucleotide diversity (π)

We selected a total of 120 West-Eurasian HGs with >150k SNPs covered on the 1240k panel, of which 103 were previously published (Extended Data Table 1), from four broad geographical regions that we labeled as "western" (n=18), "south-western" (n=7), "southern-central" (n=33), and "(south)-eastern" (n=62) Europe. We subgrouped the individuals further based on similar ^{14}C-dating and genetic cluster assignment (*16, 17, 36, 38, 50*) (for an overview of the HG groups, see Extended Data Table 1. E.g. we made seperate groups for individuals associated with the Villabruna cluster, and those high in Magdalenian-related ancestry. We determined the nucleotide diversity (π) from pseudo-haploid genotypes by calculating the average proportion of nucleotide mismatches for overlapping autosomal SNPs covered by at least one read by both individuals in a pair within a given group. Since individual pairs and not chromosome pairs are considered, this measure of nucleotide diversity does not include the global heterozygosity levels of individuals (e.g. (*32, 34, 48*)). We restricted to the set of ~870k CpG-filtered autosomal SNPs and removed individual pairs that shared less than 35,000 SNPs covered. Standard errors were determined from block jackknifes over 5Mb windows and 95% confidence intervals (95CIs) from 1,000 bootstraps. Then we calculated an average over all the individual pairs within a HG group.

We find a significantly lower nucleotide diversity (π) for individuals from the Early Mesolithic time period (π = 0.165, 95CI = 0.161-0.170), compared to those from the preceding Upper Paleolithic (π = 0.233, 95CI = 0.227-0.239), and subsequent Late Mesolithic (π = 0.220, 95CI = 0.217-0.223), Early Neolithic (π = 0.252, 95CI = 0.248-0.256), and later time periods (fig. S3.1).

In addition, the nucleotide diversity for the Sicily EM HGs is ~20% lower compared to contemporaneous Villabruna-cluster related individuals from Central Europe (π = 0.217, 95CI = 0.211-0.222), Magdalenian individuals from Iberia (π = 0.221, 95CI = 0.216-0.226) and the earliest Iron Gates HGs in Serbia (π = 0.226, 95CI = 0.222-0.229) (Fig. 3). Notably, we also find a reduction in genetic diversity for Upper Paleolithic HGs from Central Europe with Magdalenian-associated ancestry and in related Early Mesolithic HGs that are part of the ElMiron genetic cluster (*16*) ("western Europe UP Magdalenian + EM (15.5-12.7 kya)": π = 0.174, 95CI = 0.167-0.181). Intriguingly, this reduction is not found for the closely related Magdalenian- associated ElMiron genetic cluster HGs from Iberia, nor in contemporaneous individuals from Central Europe that are part of the Villabruna genetic cluster (*16*). These results underline previous suggestions for a possible genetic bottleneck in Central European Magdalenian individuals (*16, 52*).

To check whether our results are driven by differences in ascertainment bias, false positives due to sequencing errors or ancient DNA damage, we repeated the analysis for 94,469 autosomal SNP ascertained in Yoruba, an African outgroup, vis-à-vis (*17, 48*). We find that while the absolute values for the nucleotide diversity changes, the overall trend mirrors that of the full data set (fig. S3.2).

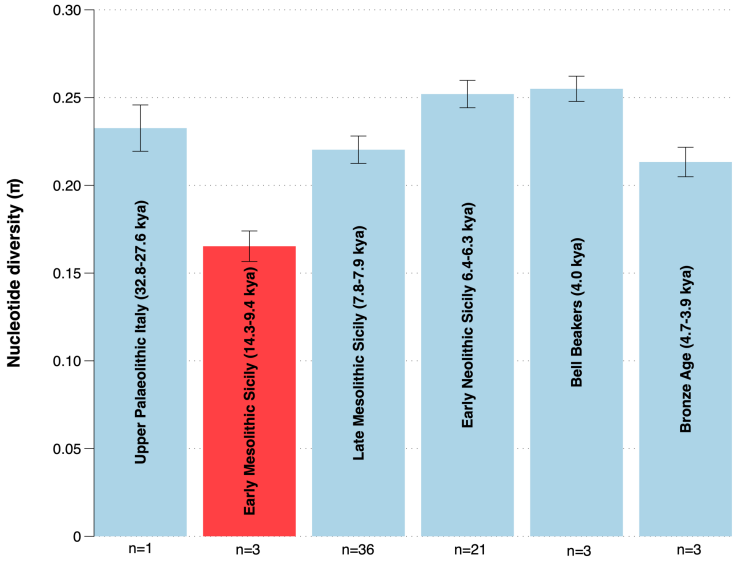


fig. S3.1. Changes in the nucleotide diversity over time for individuals from peninsular Italy and Sicily. The nucleotide diversity (π) is plotted for various transect groups in archaeo-chronological order. Upper Palaeolithic (Paglicci13, Ostuni1), Early Mesolithic (Sicily EM HGs), Late Mesolithic (Sicily LM HGs), Early Neolithic (Sicily EN farmers), Bell Beakers (Italy Bell Beakers), and Bronze Age (Italy Remedello), see Extended Data Table 1 for details on the grouping. The number of tests (n) that is used to determine the average for each time period is given. Error bars reflect 3 SEs.

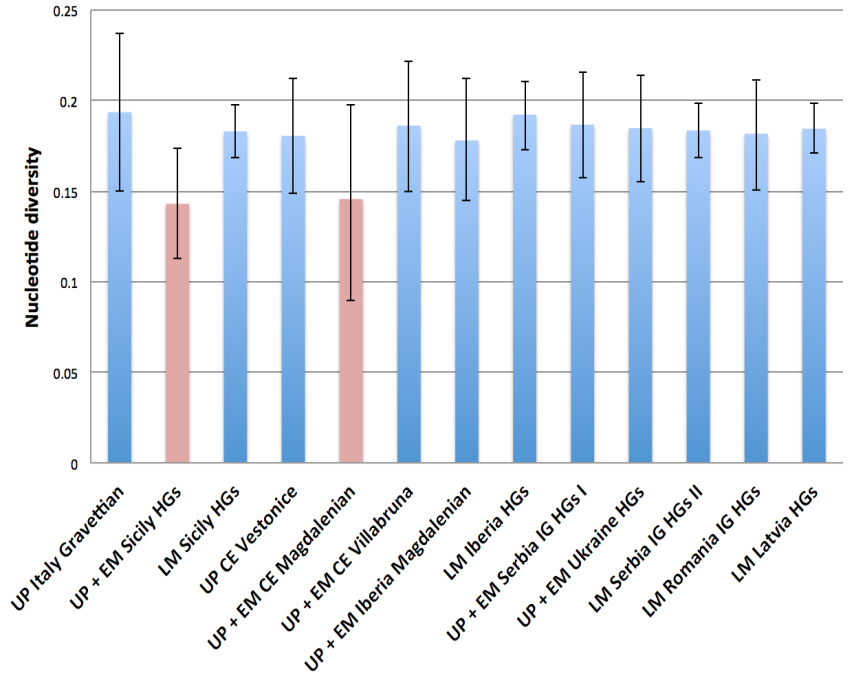


fig. S3.2. Nucleotide diversity (π) for the various Eurasian HG groups for autosomal SNP sites ascertained in Yoruba. Error bars reflect 3 SEs.

Individual Heterozygosity (H)

Secondly, we investigated whether the genetic diversity within the individual Sicily EM HGs is reduced compared to the later Sicily HGs and early farmers. We hence quantified the **individual heterozygosity (H)**, the proportion of sites considered heterozygous among all sites analysed. We calculated heterozygosity as the sum of heterozygous genotypes estimated using *SnpAD* (*126*) (version 0.3.3 with parameters --max_gtfreq=0.2) (table S3.1). Error profiles were calculated separately for single-stranded and double-stranded libraries, when both types of data were available. Confidence intervals (95%) (95CIs) were calculated using *snpADci*, which determines multiple testing corrected confidence intervals around heterozygous genotype frequencies. These confidence intervals were summed to arrive at confidence intervals for the heterozygosity.

As a quality check we investigated whether differences in calling rate for the alternative allele influenced the calculated heterozygous genotype frequencies. The calling rate may be biased when a heterozygous SNP site is covered by only a few reads. When the SNP depth is low the alternative allele may not be observed. Indeed, when the heterozygosity level (H) is plotted as a function of the read depth (X), individuals with an average SNP depth of 0.08-1.54X have a considerably lower calling rate for the alternative allele at heterozygous sites (fig. S3.3). This bias plateaus in our dataset in individuals with > 1.94X read coverage, and we hence used this as a cutoff for our analysis.

Subsequently, with the *boxplot.stats()* function in R we found that the individual heterozygosity for *UZZ88* is reduced compared to other Sicily LM HGs and does not fall within the variance of this group. We hence did not include this individual in the Sicily LM group average. We found that the average individual heterozygosity (H) for Sicily EM HGs is 30% lower compared to Sicily LM HGs (non-overlapping 95CIs), and 40% compared to Sicilian EN farmers (non-overlapping 95CIs) (fig. S3.4).

Individual label	# autosomal SNPs covered on 1240k	Average read depth for SNP (X coverage)	Individual Heterozygosity (H)	95% CI low. bound	95% CI up. bound
Sicily EM UZZ05054	502,957	3.5990	0.1422	0.1391	0.1451
Sicily EM OrienteC	155,489	NA	0.1628	0.1519	0.1722
Sicily EM UZZ096	48,824	0.1038	0.1042	0.0895	0.1218
Sicily LM UZZ069	449,167	2.7736	0.2086	0.2046	0.2119
Sicily LM UZZ079	534,685	7.1628	0.2165	0.2135	0.2197
Sicily LM UZZ080	561,466	9.1715	0.2139	0.2107	0.2166
Sicily LM UZZ082	517,375	4.7542	0.2132	0.2109	0.2179
Sicily LM UZZ040	175,573	0.5139	0.1206	0.1139	0.1267
Sicily LM UZZ04446	356,411	1.1895	0.2242	0.2193	0.2294
Sicily LM UZZ071	373,717	1.5397	0.1739	0.1703	0.1792
Sicily LM UZZ081	213,244	0.6539	0.1628	0.1567	0.1702
Sicily LM UZZ088	406,054	1.9261	0.1999	0.1963	0.2045
Sicily EN UZZ061	472,518	3.5955	0.2481	0.2438	0.2525
Sicily EN UZZ075	539,032	5.8031	0.2484	0.2458	0.2521
Sicily EN UZZ033	248,381	0.7065	0.2126	0.2061	0.2192
Sicily EN UZZ034	182,339	0.5185	0.1879	0.1802	0.1954
Sicily EN UZZ074	132,318	0.3468	0.0744	0.0667	0.0840
Sicily EN UZZ077	116,020	0.2542	0.2129	0.2022	0.2273
Sicily EN UZZ087	39,706	0.0806	0.1421	0.1209	0.1657

table S3.1. Individual heterozygosity (H) levels for the ancient Sicilians in our transect. 95% confidence intervals (95CI) were determined using block jackknifes over 5Mb windows and corrected for multiple testing. The total number of autosomal SNPs covered on the 1240k panel and average read depth for the SNPs are given. Individuals that were excluded from analysis are in red.

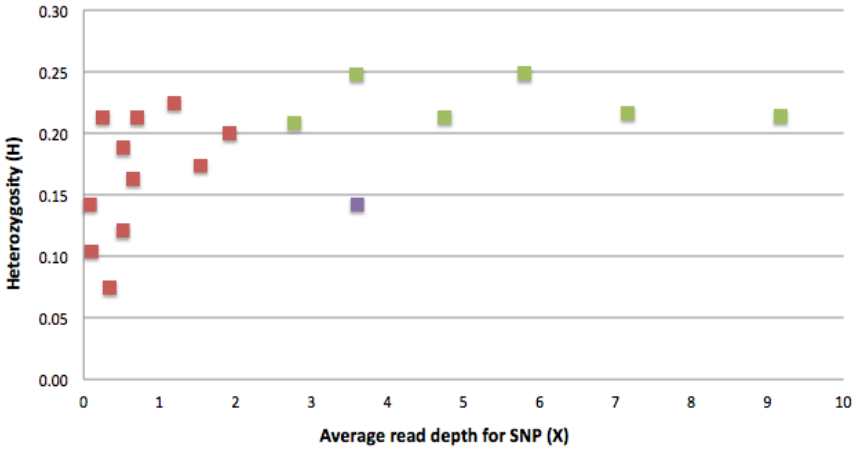


fig. S3.3. Plot for the individual Heterozygosity (H) levels in our ancient Sicilians as a function of their average SNP depth. Red: low coverage individuals that were excluded from analysis. For these individuals we found a systematic bias in the calling rate for the alternative allele at heterozygous sites. Green: individuals > 1.94X coverage that passed our quality threshold filters. Purple: Sicily EM HG *UZZ5054*. The observed lower heterozygosity resulted from its population genetic history. *OrienteC* is not plotted (H = 0.163, average SNP depth unknown).

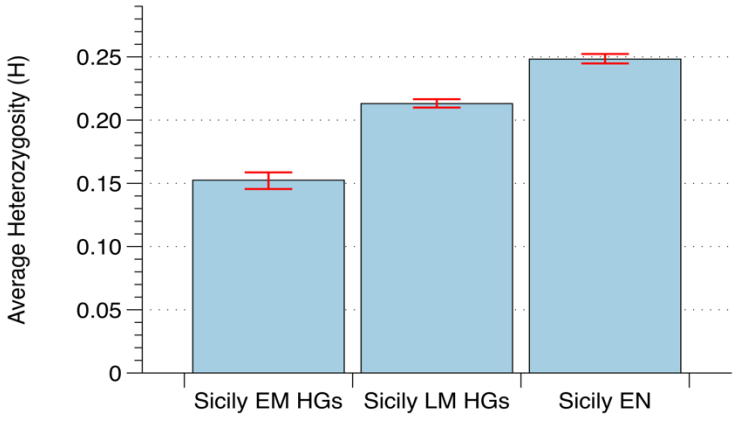
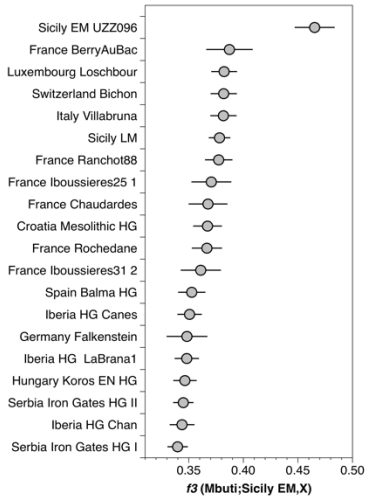


fig. S3.4. Barplot showing the average individual heterozygosity (H) for Sicily EM HGs, LM HGs and early farmers (EN). The 95% confidence intervals are given in red. The average individual heterozygosity (H) for Sicily EM HGs is 0.153 (95CI: 0.146-0.157), for Sicily LM HGs 0.213 (95CI: 0.210-0.217) and for Sicily early farmers (EN) 0.248 (95CI: 0.245-0.252). The confidence interval for Sicily EM HGs does not overlap with that for either Sicily LM HGs or Sicily EN.

S4. Characterizing the Sicilian Mesolithic HGs ancestry using F-statistics

First, we used outgroup f_3-statistics to investigate for various West-Eurasian HGs (*X*) which one is genetically closest to Sicily EM HGs and Sicily LM HGs, using *f_3(Mbuti; Sicily EM HGs, X)* and *f_3(Mbuti; Sicily LM HGs, X),* respectively (fig. S4.1). The highest amount of shared genetic drift for Sicily EM HGs is with Sicily EM HG *UZZ96*, followed by Villabruna cluster individuals and Sicily LM HGs. Sicily LM HGs show the highest degree of allele sharing with Sicily EM HGs, followed by other individuals from the Villabruna cluster.

Secondly, we performed f_4-cladality statistics of the form *f_4(Chimp, Sicily LM HGs; Sicily EM HGs, X)* and *f_4(Chimp, Sicily EM HGs; Sicily LM HGs, X)* (fig. S4.2). For almost all tested HGs *X* the statistic is ≤ 0, implying that Sicily LM and EM HGs form a clade to the exclusion of other West-Eurasian HGs. This suggests that the shared genetic drift level measured in the above f_3-outgroup statistics most likely reflects a direct ancestry connection between Sicily EM HGs and Sicily LM HGs. However, Sicily EM HGs do not represent all the ancestry in the Sicily LM HGs, since in the f_4- statistic *f_4(Chimp, X; Sicily EM HGs, Sicily LM HGs)* additional admixture signals are found for various HGs from (south)-eastern Europe and Russia (Fig. 4A).



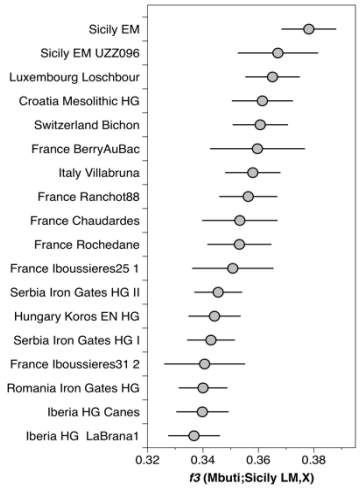
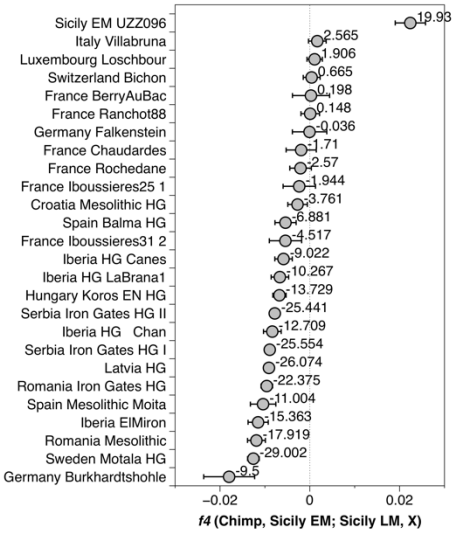


fig. S4.1. *F_3*-outgroup statistics for the Mesolithic Sicilian HGs. Error bars reflect 3 SEs.

.



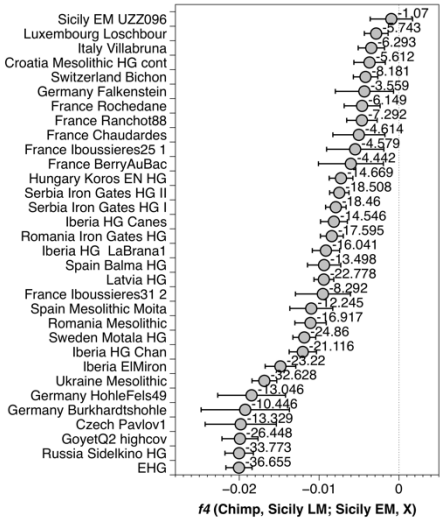


fig. S4.2. ***F_4*-cladality statistics for the Mesolithic Sicilian HGs.** Error bars reflect 3 SEs, z-values are given.

Subsequently, we aimed to compare the ancestry as found in *Villabruna* with that in the Mesolithic Sicilian HGs and their respective affinities to various West Eurasian HGs (*X*). Accordingly, we perfomed f_4-cladality tests of the form *f_4(Chimp, X; Sicily EM HGs, Villabruna),* and *f_4(Chimp, X; Sicily LM HGs, Villabruna)*. Notably, comparing the ancestry of Sicily EM HGs and *Villabruna* results in a similar geographical separation in their genetic affinitties to western and eastern European HGs as found for *f_4(Chimp, X; Sicily EM HGs, Sicily LM HGs)* (Fig. 4). Also here, Sicily EM HGs share an excess of alleles with western European HGs, including the majority of Villabruna cluster individuals, whereas *Villabruna* does with (south-)eastern European HGs (fig. S4.3 - Left). This suggests that *Villabruna* and Sicily LM HGs behave genetically similar in relation to Sicily EM HGs.

If the gene pool of the Sicily LM HGs is very similar to that of *Villabruna*, we do not expect any HGs from Eurasia to significantly share more alleles with either of them. In an f_4-cladality test, we hence expect *f_4(Chimp, X; Villabruna, Sicily LM HGs)* ≈ 0. However, EHG is marginally closer to Sicily LM HGs, whereas, Sicily EM HGs and some other West-Eurasian HGs from the Villabruna and Gravettian cluster are closer to *Villabruna* (fig. S4.3 - Right).

Taken together, these *F*-statistics suggest that *Villabruna*, Sicily EM HGs and Sicily LM HGs share significant ancestry but differ in their affinities towards eachother and to HGs from south(-eastern) Europe and HGs with Magdalenian-related ancestry from southwestern Europe.

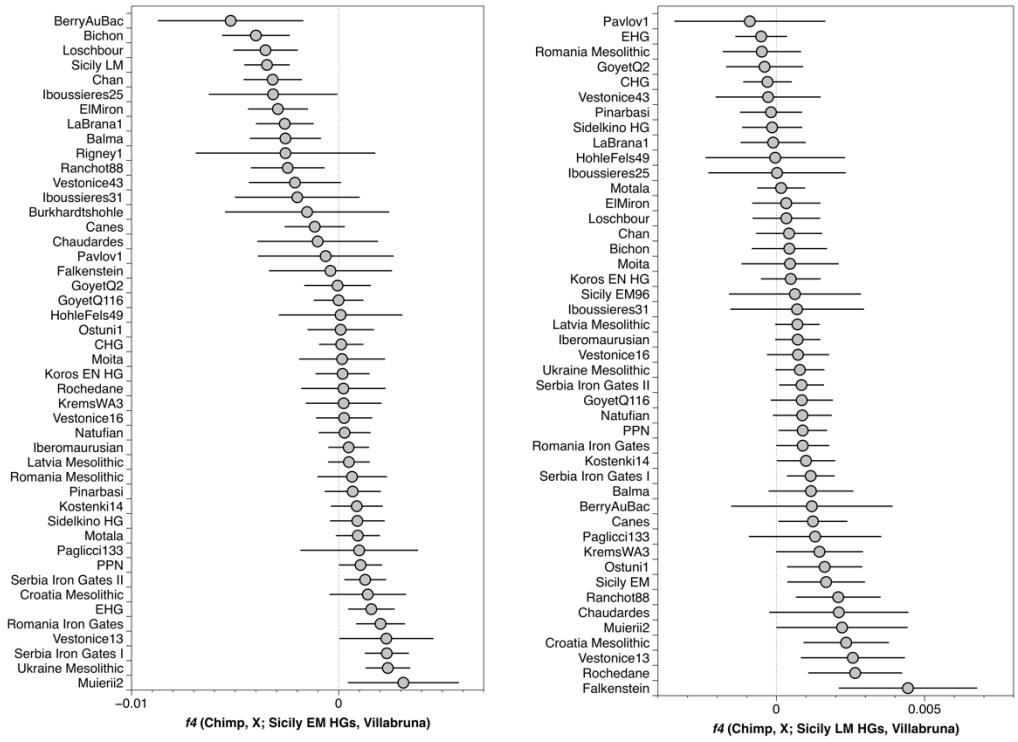


fig. S4.3. Cladality f_4-statistics to compare the ancestry in Villabruna to that of the Sicilian Mesolthic HGs with *f_4(Chimp, X; Sicily EM/LM HGs, Villabruna).* Error bars reflect 2 SEs.

S5. Investigating the phyologenetic position of the Early Mesolithic Sicilian HGs

Admixture graph models fit allele frequency correlations and allow us to hierarchically build an increasingly complex framework of ancestry streams that fit the genetic diversity observed. Here, we used the *qpGraph* program (*57*) to construct a phylogeny of ancestry lineages found among Palaeolithic and Mesolithic West-Eurasian HGs to further clarify the phylogenetic position of Sicily EM HGs in relation to *Villabruna*, other Early Mesolithic HGs from continental Europe (EM WHGs), and Magdalenian-associated HGs (e.g. *El Miron* and *GoyetQ2*).

We built the phylogeny models with increasing complexity by fitting representative West Eurasian HG ancestry lineages one by one to the phylogeny roughly in order of their respective ^{14}C dates. We added each of them to all possible nodes as a branch without admixture or as a binary admixture between two branches. We selected models that did not include trifurcations or 0% ancestry stream estimates, and for which the difference between the observed and fitted *f*-statistics were the lowest (the maximum deviation falls within 3.5 SEs for our preferred models). We preferred a model that fits a HG lineage as a branch without admixture over one with additional admixture if both of them fit the observed *f*-statistics equally well.

For this analysis, we grouped individuals that have similar ancestry:

Sicily EM	*I2158/OrienteC* (~14 kyBP), *UZZ05054* (~10.5 kyBP)
EM WHGs	Early Mesolithic WHGs: *Bichon* (~13.5 kyBP), *Rochedane* (~13 kyBP), *Iboussieres25* (~12 kyBP), *Iboussieres31* (~11.5 kyBP)

We started with a core model (CM-5) phylogeny fitting five populations that separates African (Mbuti) from non-African ancestry (*UstIshim* ~45 kyBP (*27*)), followed by a major split between basal West Eurasian (*Kostenki14*, ~36 kyBP (*127*)) and Ancient North Eurasian (ANE) ancestry (*Mal'ta*, ~24 kyBP (*36, 46*)) Subsequently, we added ~30 kyBP *Vestonice16* (*16*) associated with the Gravettian.

CM-5: 1) Mbuti, 2) *Ust Ishim*, 3) *Kostenki14*, 4) *Mal'ta*, 5) *Vestonice16*

We found two models that fit the data (fig. S5.1). In the least complex model *Vestonice16* is fitted as branch without admixture as a sister lineage to *Kostenki14*, and *Mal'ta* as an outgroup to both of them. In the alternative model, *Vestonice16* is on an admixed branch with a source related to *Mal'ta* contributing 92% and *Kostenki14* contributing 8%.

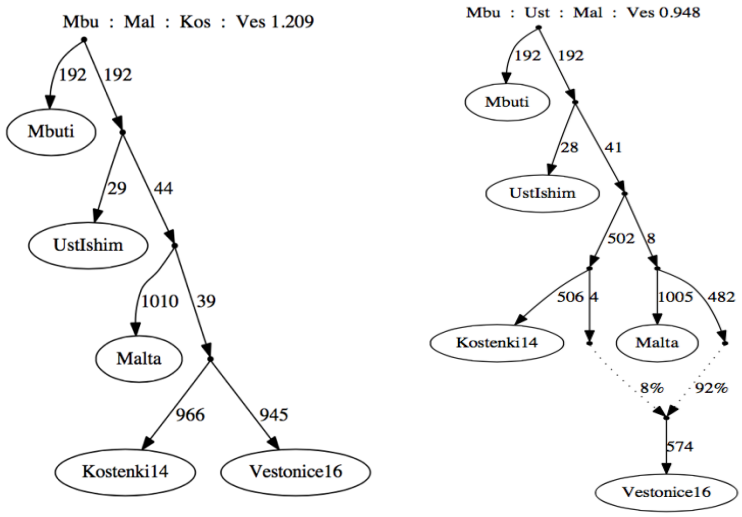


fig. S5.1. CM-5 models fitting 5 populations. Left: *Vestonice16* is fitted as unadmixed on a branch with *Kostenki14* with *Mal'ta* as an outgroup to both of them (1 outlier, max |f_4, expected – f_4, observed| = 1.209). Right: *Vestonice16* is admixed (1 outlier, max |f_4, expected – f_4, observed| = 0.948).

To the least complex model we then added either *GoyetQ116* (~35 kyBP Aurignacian (*16*)), *El Miron* (~19 kyBP Magdalenian (*16*)) or *GoyetQ2* (~15 kyBP Magdalenian (*16, 50*) as a representative of an ancestry lineage characteristic for Magdalenian-associated HGs (CM-6). Subsequently, we added ~18 kyBP *AfontovaGora3* (*16*) (CM-7), a more recent representative of the ANE ancestry lineage.

CM-6 = CM-5 + *GoyetQ116* or *El Miron* or *GoyetQ2*
CM-7 = CM-6 + *AfontovaGora3*

GoyetQ116 and *El Miron* can be fitted as branches without admixture in both core models (fig. S5.2 - Left & Center, only the least complex models are shown). In contrast *GoyetQ2* is best modeled as a mixture between 47% *Vestonice16* ancestry and 53% ancestry from a basal lineage that branches of basal to HGs with ANE ancestry. In all three models, *AfontovaGora3* is best fitted on a branch with *Mal'ta* without additional admixture (fig. S5.2 - Right).

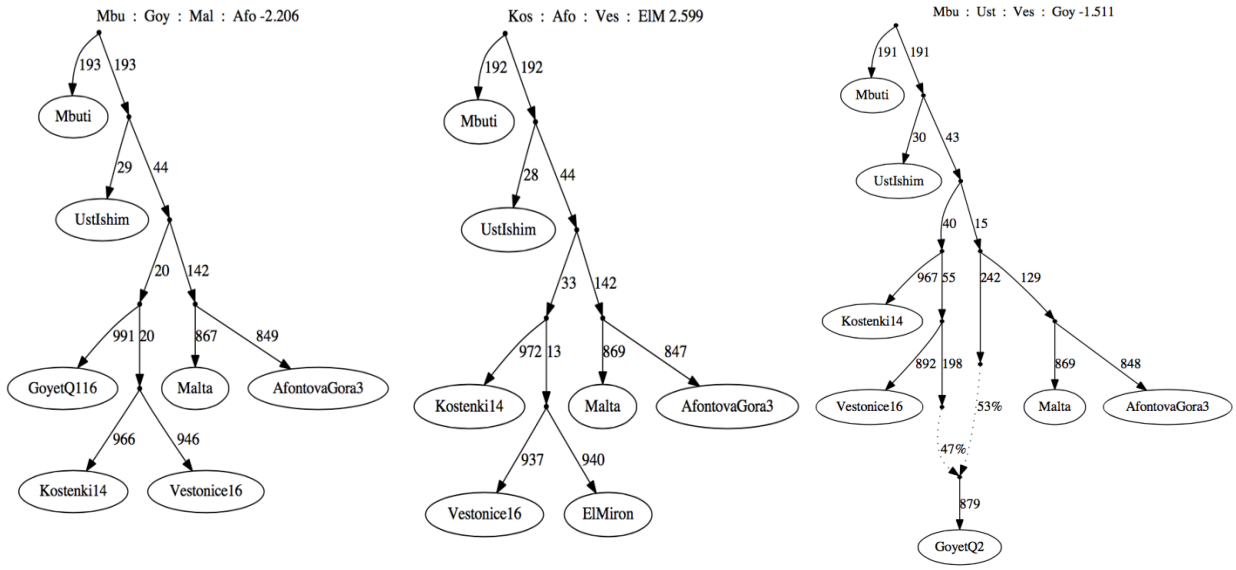


fig. S5.2. CM-7 models fitting different HGs as representatives of an ancestry found in Magdalenian-associated individuals. Left: *GoyetQ116* as a branch without admixture (1 outlier, $|f_4$, expected – f_4, observed$|$ = 2.206). Center: *El Miron* as a branch without admixture (1 outlier, $|f_4$, expected – f_4, observed$|$ = 2.599). Right: *GoyetQ2* as a mixture between *Vestonice16* and a lineage basal to ANE (1 outlier: $|f_4$, expected – f_4, observed$|$ = 1.511).

We then made an experimental model series to investigate how *Villabruna* and Sicily EM HGs relate to each other, and whether either of them requires an additional ancestry contribution from a Magdalenian-associated or ANE-related source. We added ~14 kyBP *Villabruna* (*16*) followed by the Sicily EM HGs, and *visa versa*, either as a branch without or with admixture to the seven population core models (EM-9a+b).

EM-9a: CM-7 + *Villabruna* + Sicily EM HGs
EM-9b: CM-7 + Sicily EM HGs + *Villabruna*

The least complex model that fits the gene pools of *Villabruna* and Sicily EM HGs is the same regardless of the order in which we added them (fig. S5.3 - Left): Sicily EM HGs and *Villabruna* form a clade, with *GoyetQ2* as an outgroup to both of them. Notably, models that include one admixture event fit the gene pools of both Sicily EM HGs and *Villabruna* approximately equally well. However, the branches that contributed ancestry are different for the two. When *Villabruna* is added to the graph first (EM-9a), Sicily EM HGs is fitted on an admixed branch that derives 37% ancestry from a *Villabruna*- and 63% from a *GoyetQ2*-related lineage (fig. S5.3 - Center). In contrast, when Sicily EM HGs is added first (EM-9b), *Villabruna* is fitted on an admixed branch that derives 97% ancestry from a lineage close to Sicily EM HGs and 3% ancestry from a lineage related to *AfontovaGora3* (fig. S5.3 - Right).

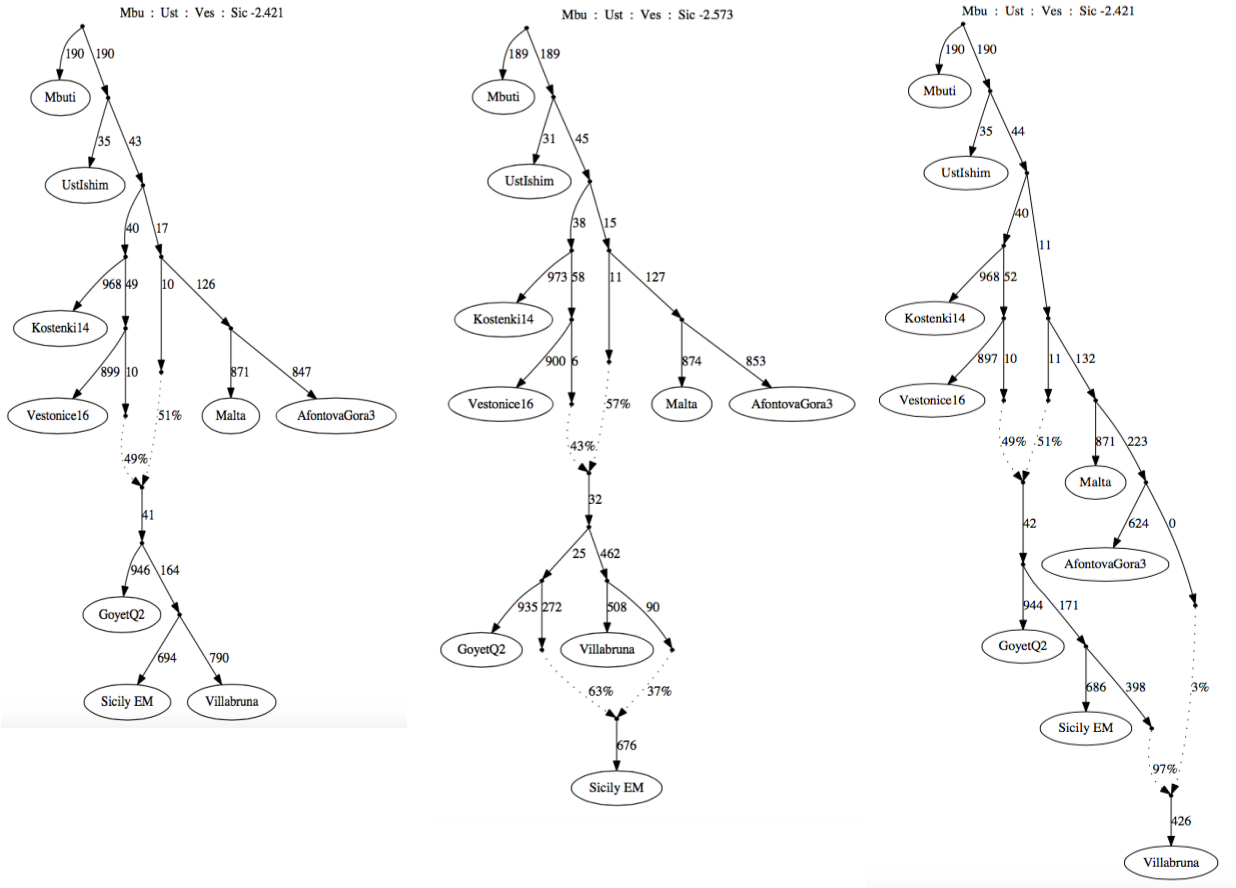


fig. S5.3. **EM-9 models fitting Sicily EM HGs and *Villabruna*.** Magdalenian-associated ancestry is represented by *GoyetQ2*, and ANE ancestry by *Mal'ta* and *AfontovaGora3*. Left: Sicily EM HGs and *Villabruna* form a clade with *GoyetQ2* as an outgroup to both of them (1 outlier, $|f_4$, expected – f_4, observed$|$ = 2.421). Center: Sicily EM HGs as a mixture between branches related to *Villabruna* and *GoyetQ2* (1 outlier, max $|f_4$, expected – f_4, observed$|$ = 2.573). Right: *Villabruna* as a mixture between Sicily EM HGs and a lineage related to *AfontovaGora3* (1 outlier, max $|f_4$, expected – f_4, observed$|$ = 2.421).

Changing the proxy for Magdalenian-related ancestry to *GoyetQ116* or *El Miron* does neither result in a consistent tree topology. Using *GoyetQ116* results in Sicily EM HGs and *Villabruna* being fitted as being cladal on an admixed branch that derives 96% ancestry from a lineage that related to *GoyetQ116* and 4% from *Vestonice16* (fig. S5.4 - Left). An alternative model fits Sicily EM HGs on a branch without admixture, and *Villabruna* on an admixed branch between 97% Sicily EM HG-related and 3% *AfontovaGora3*-related ancestry (fig. S5.4 - Center). Modeling the Magdalenian-associated ancestry with *El Miron* fits the ancestry in *Villabruna* as a mixture between 91% Sicily EM HG-related and 9% *El Miron*-related ancestry (fig. S5.4 - Right).

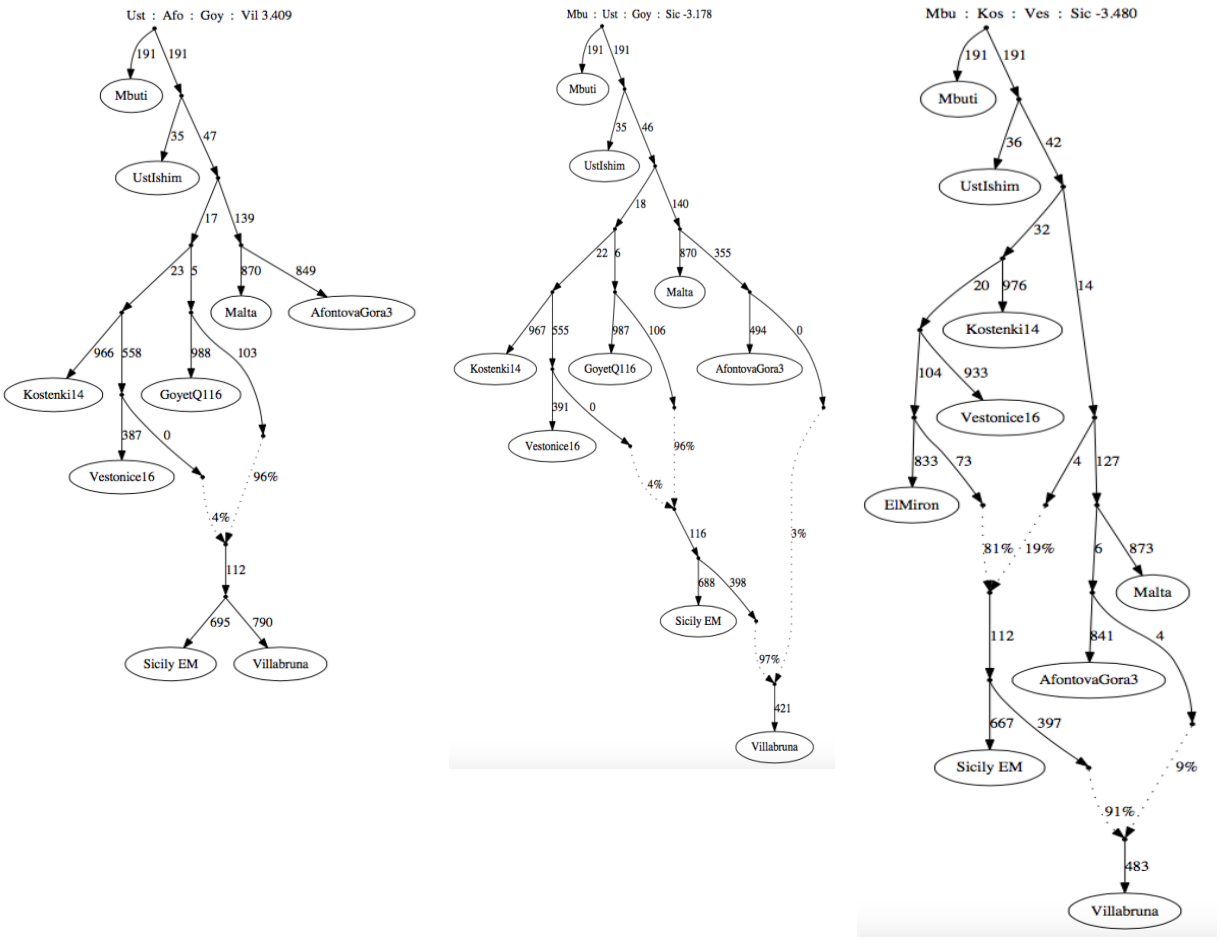


fig. S5.4. EM-9 models fitting nine populations, including Sicily EM HGs and *Villabruna*, with Magdalenian-associated ancestry represented by *GoyetQ116* (Left, Center) or *El Miron* (Right). Left: Sicily EM HGs and *Villabruna* form a clade on branch that derives 96% ancestry from a source close to *GoyetQ116* and 4% from *Vestonice16* (2 outliers, max $|f_4$, expected – f_4, observed$|$ = 3.409). Center: *Villabruna* as a mixture between branches related to Sicily EM HGs and *GoyetQ116* (1 outlier, max $|f_4$, expected – f_4, observed$|$ = 3.178). Right: *Villabruna* as a mixture between branches related to Sicily EM HGs and *El Miron*, respectively (4 outliers, max $|f_4$, expected – f_4, observed$|$ = 3.480).

In a second experimental model series we additionally included a group of early Mesolithic continental HGs (EM WHGs) dating to ~13.5-11.5 kyBP, which are approximately contemporaneous to *Villabruna* and Sicily EM HGs and part of the Villabruna genetic cluster (*17, 32*). To the seven populations core model we hence added *Villabruna*, followed by Early Mesolithic continental HGs (EM WHGs) (EM-9c), and then Sicily EM HGs (EM-10), either as a branch without or with admixture. Since *GoyetQ2* as a proxy for Magdalenian- associated ancestry has sofar resulted in the smallest discrepancy between the observed and fitted allele frequencies, we proceeded with this model (fig. S5.3 - Left).

EM-9c: CM-7[*GoyetQ2*] + *Villabruna* + EM WHGs
EM-10: CM-7[*GoyetQ2*] + *Villabruna* + EM WHGs + Sicily EM HGs

When EM WHGs is added to the graph with *Villabruna (*EM-9c), the EM WHGs are fitted on a branch to which *Villabruna* contributes 95% ancestry and *GoyetQ2* contributed 5% (fig. S5.5 - Left). A model that fits *Villabruna* and Sicily EM HGs as a clade on a branch without admixture (similar to fig. S5.3 - Left) results in more significant outlier statistics (3 outliers, max $|f_4$, expected – f_4, observed$|$ = 4.388), and hence is less likely to reflect the true tree topology.

Subsequently, when Sicily EM HGs is added (EM-10) it is placed on a branch without admixture that falls basal to both *Villabruna* and EM WHGs, with *GoyetQ2* as the immediate outgroup (fig. S5.5 – Right). Notably, the ancestry contribution from *GoyetQ2* to EM WHGs increases to 14%, and the ancestry contribution from *AfontovaGora3* to *Villabruna* increases to 9%. We can therefore not rule out that Sicily EM HGs descend from a more basal lineage that admixed into Iberian HGs and Villabruna cluster individuals.

All in all, in our models there is a complex interaction between distal affinities to ANE- and Magdalenian-related ancestries in *Villabruna* and other Early Mesolithic HGs from continental Europe, and Sicily EM HGs. Depending on the populations included in the scaffold graph we obtained different ancestry contributions and different tree topologies. We hence could not accurately resolve the phylogeny of the Sicily EM HGs. However, eventhough *Villabruna*, Sicily EM HGs and EM WHGs share a large proportion of their ancestry, our results hint at population substructure among these HGs. *Villabruna* shows a stronger affinity to ANE ancestry, whereas Sicily EM HGs and EM WHGs show a stronger affinity to Magdalenian-related ancestry.

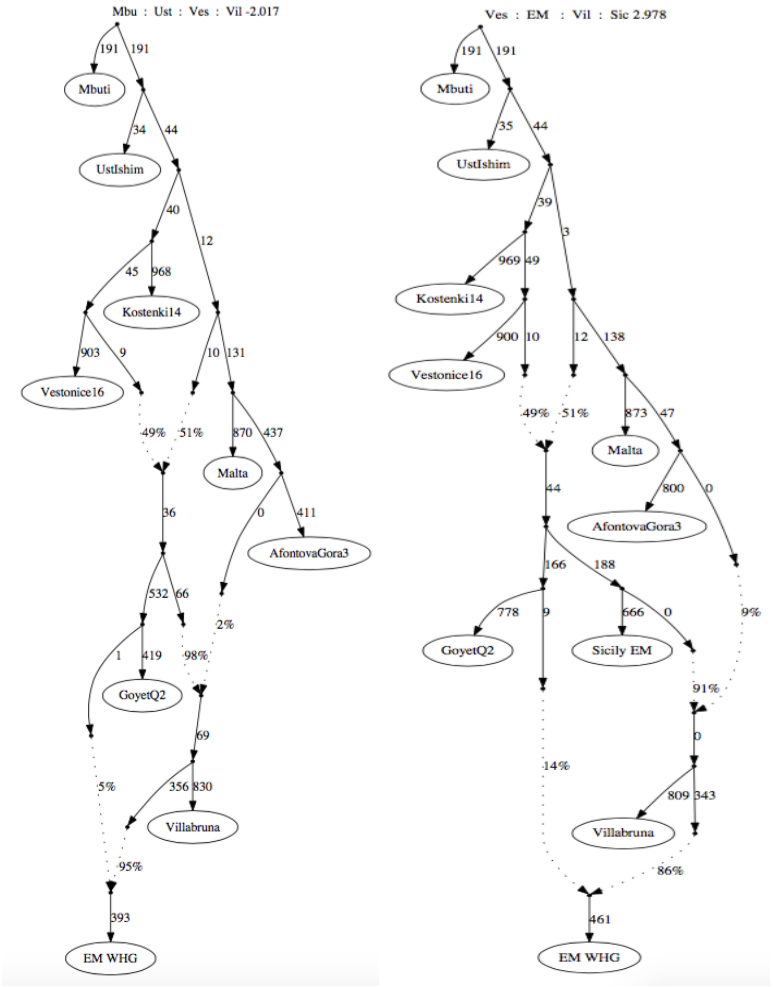


fig. S5.5. Scaffold graphs to investigate the genetic relation of *Villabruna*, EM WHGs and Sicily EM HGs. Magdalenian-associated ancestry is represented by *GoyetQ2*. Left: EM-9c scaffold graph fitting 9 populations, including *Villabruna* and EM WHGs. EM WHGs is fitted on an admixed branch that derives 95% ancestry from a source close to *Villabruna* and 5% from *GoyetQ2* (1 outlier, max $|f_4$, expected – f_4, observed$|$ = 2.017). Right: EM-10 scaffold graph fitting 10 populations to investigate the phylogenetic position of Sicily EM HGs relative to *Villabruna* and EM WHGs. Sicily EM HGs is fitted on a branch that falls basal to both *Villabruna* and EM WHGs. Sicily EM HGs contributed ancestry to *Villabruna*, and *Villabruna* contributed ancestry to EM WHGs (1 outlier, max $|f_4$, expected – f_4, observed$|$ = 2.978).

S6. Characterizing the Sicilian early farmer ancestry using *F*-statistics

First, we tested with an admixture f_4-statistic whether the Early Neolithic Sicilians contain a HG ancestry component in addition to their shared ancestry with early farmers from Anatolia Barcin of Greece Peloponnese. Accordingly, we performed *f_4(Chimp, X; Greece EN Peloponnese/Anatolia EN Barcin, Sicily EN),* where *X* are various West Eurasian HG lineages (fig. S6.1). By using Greece EN Peloponnese or Anatolia EN Barcin as a baseline for the early farmer ancestry in Sicily EN (*38*), we downweighted any shared ancestry related to this that is abundant in many Mesolithic HG lineages from southern Europe (e.g. Iron Gates HGs (*17*), and see fig. S6.2). We found that Sicily EN shows significant admixture signals for various HGs from Europe, including preceding local Sicily EM (z_{max} = 4.12) and Sicily LM HGs (z_{max} = 4.09) (fig. S6.1). Notably, alongside the Sicilian HGs, Mesolithic HGs from France (e.g. *Chaudardes*: z_{max} = 2.89, *Ranchot88:* z_{max} = 3.64), Croatia (z_{max} = 2.61), and Iberia (*LaBrana*: z_{max} = 2.96) are among the strongest signals for HG admixture in Sicily EN (fig. S6.1).

To test whether the Sicilian early farmers are closer to the Mesolithic HGs from France than to the preceding Sicily LM HGs, we performed f_4-cladality statistics of the form *f_4(Chimp, Sicily EN; Sicily LM HGs, X).* For this test we found no significantly positive f_4-values for the Mesolithic HGs from France (fig. S6.2). However, given the low number of ABBA and BABA trees, the similar ancestries in Sicily LM HGs and the Villabruna-cluster individuals from France may be driving the non-significance for this test. Notably, Mesolithic HGs from southeastern Europe, including *Croatia Mesolithic HG*, Koros EN HG and various Iron Gates HG groups, do share significantly more alleles with Sicily EN than the local Sicily LM HGs do. However, due to the pre-Neolithic gene flow between southeastern Europe and the Near East, the positive signals in this test for the shared genetic drift between Sicily EN and southeastern European HGs might either reflect an ongoing direct gene flow between these regions, e.g via maritime contact (*5*), or an indirect signal from the farmer ancestry that was brought in.

Subsequently, we aimed to find the closest proxy for the non-HG ancestry component in Sicily using *f_4(Chimp, X; Sicily LM HGs, Sicily EN),* where *X* are various Early Neolithic groups from West Eurasia (fig. S6.3). We here took the Sicily LM HG ancestry as a baseline for any HG ancestry in Sicily EN, and any HG ancestry broadly similar to it in *X*, if present. Hence, any Sicily LM-like HG ancestry is downweighted in this test and does not contribute to the f_4-statistic. We expected all Neolithic groups to result in a positive value for this test. However, the test group with the highest relative f_4-value was assumed to be the best proxy for the farmer ancestry component in Sicily EN. Again, we found the highest levels of shared genetic drift for various Early Neolithic farmers from the Balkan, followed by Hungary Koros and Anatolia Barcin (fig. S6.3).

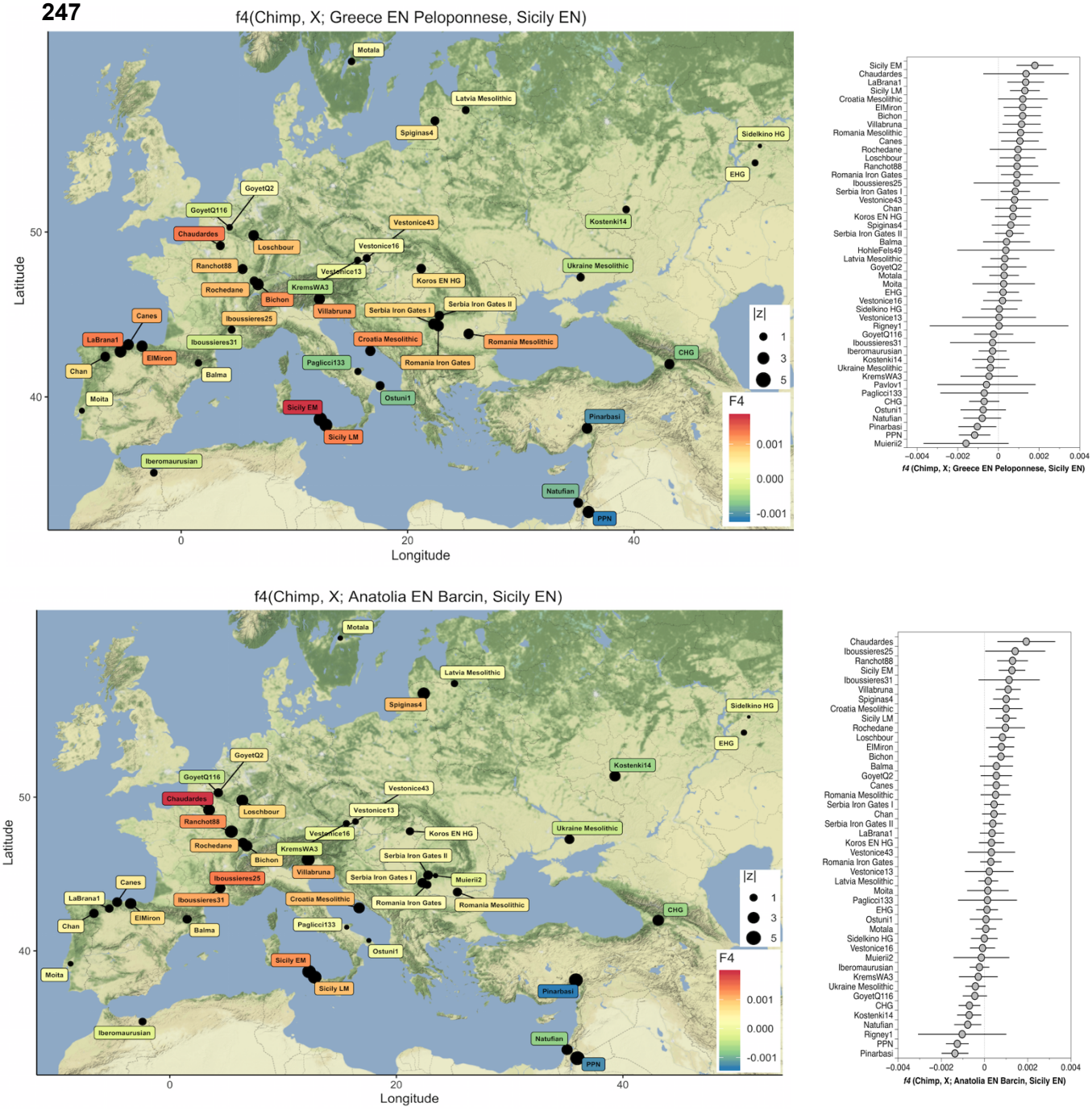


fig. S6.1. ***F*₄-admixture statistics for various HG ancestry sources in Sicily EN, relative to Greece EN Peloponnese (top) and Anatolia EN (bottom), of the form *f₄(Chimp, X; Greece EN Peloponnese/Anatolia EN Barcin, Sicily EN).*** Warmer colours reflect stronger signals of admixture. Dot sizes reflect |z|-scores and error bars 2 SEs.

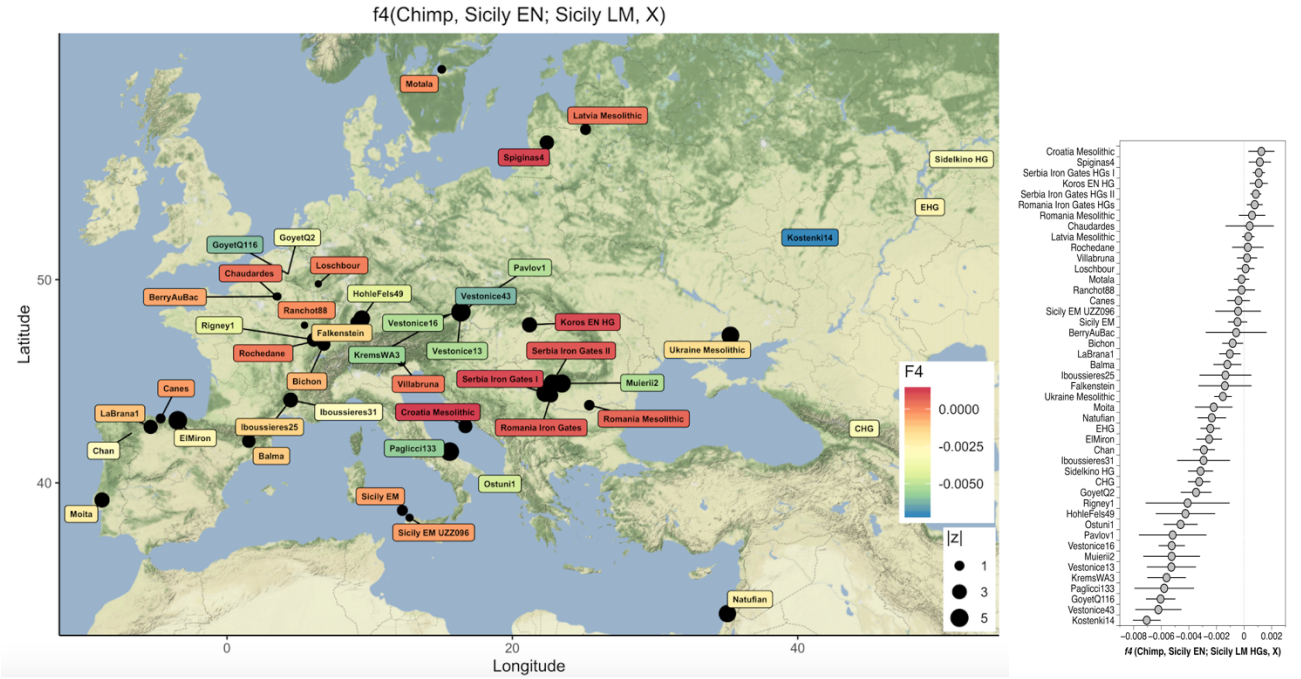


fig. S6.2. *F*$_4$-symmetry statistics of the form *f*$_4$*(Chimp, Sicily EN; Sicily LM HGs, X)* to compare the ancestry in Sicily EN farmers to that in Sicily LM HGs and various West-Eurasian HGs (*X*). There is no indication that Mesolithic HGs from France are genetically closer to the Sicilian early farmers than Sicily LM HGs are. Dot sizes reflect |z|-scores and error bars 2 SEs.

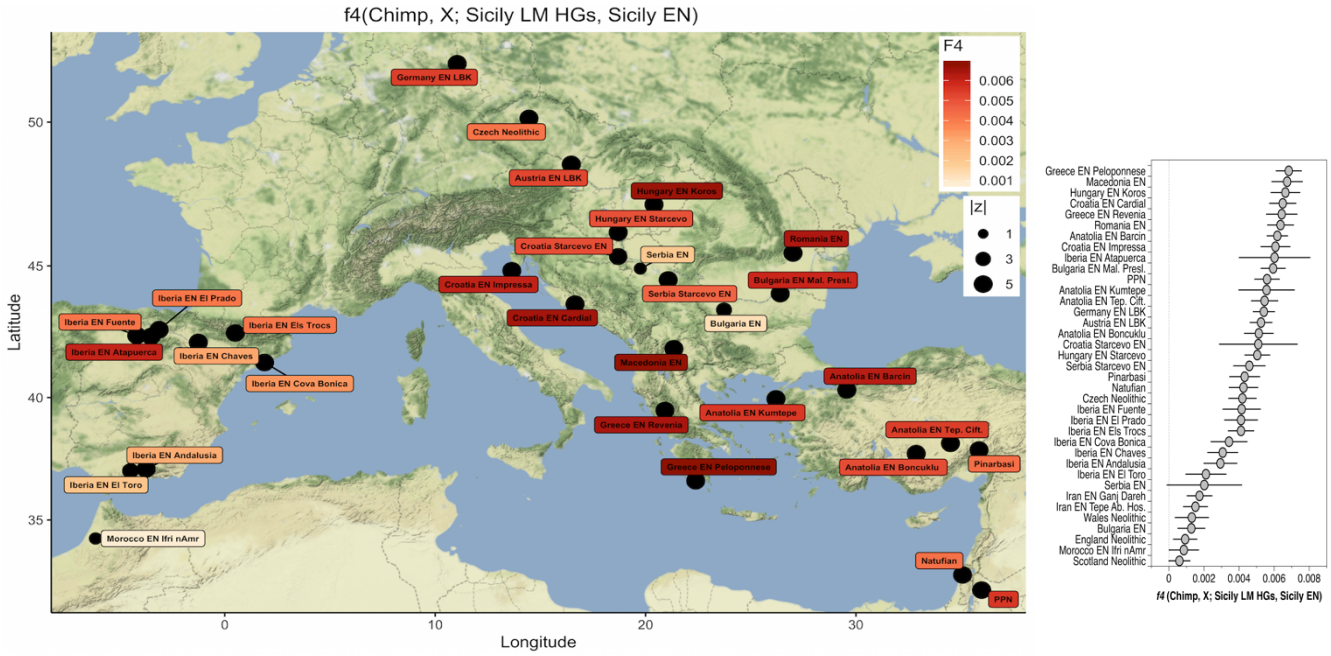


Fig. S6.3. *F*$_4$-admixture statistic of the form *f*$_4$*(Chimp, X; Sicily LM HGs, Sicily EN)* to find the closest proxy to the early farmer ancestry in Sicily, using Sicily LM HG as a baseline for the HG ancestry. Redder colours reflect higher levels of shared genetic drift between Sicily EN and the tested early farmer group X. Dot sizes reflect |z|-scores and error bars 2 SEs. Early farmers from the Balkan (Greece, Macedonia, Croatia) and Hungary Koros share the highest excess of alleles with Sicily EN.

Alternatively, we aimed to find the closest proxy for the non-HG ancestry in Sicily EN by measuring levels of shared genetic drift to various Early Neolithic European groups *X*, using Greece EN Peloponnese as a baseline, with f_4*(Chimp, Sicily EN; Greece EN Peloponnese, X)* (fig. S6.4). This test is similar to the f_3-outgroup statistic f_3*(Mbuti, Sicily EN, X)* (Fig. 4B), except that it downweights the ancestry that is shared between Sicily EN and Greece EN Peloponnese, and test group *X* and Greece EN Peloponnese. This has the advantage that any HG ancestry that is part of the gene pool shared between Sicily EN, Greece EN Peloponnese and *X* does not contribute to the *f*-statistic. In this way we can separate the population genetic affinities from distant HG admixture (e.g. from southeastern European HGs or Near Eastern HG groups), which may have resulted in population genetic substructure within the European Early Neolithic founder groups, from admixture from local HG groups *en route* as the Early Neolithic farmers expanded. Similar to the f_3-outgroup statistic results, we find various Early Neolithic farmers from the Balkan and Central Europe to be genetically most similar to Sicily EN (fig. S6.4).

Lastly, we performed the cladality test f_4*(Chimp, X; Greece EN Peloponnese, Sicily EN),* where *X* are various Early Neolithic groups from West Eurasia (fig. S6.5). By taking an immediate genetic outgroup to the Early Neolithic groups in Europe, Greece EN Peloponnese, any *en route* and locally admixed HG ancestry that is part of gene pool of Sicilian EN and is shared with test group *X* will contribute to a positive f_4-statistic. The test group with the most positive f_4-statistic hence will have a gene pool that is most similar in allele frequencies and variances compared to that of Sicily EN. We find that none of the Early Neolithic group from Europe shares significantly more alleles with Sicily EN than with Greece EN Peloponnese (fig. S6.5). Sicily EN and Greece EN Peloponnese either form a clade to the exclusion of other European Early Neolithic farmers, or the latter are genetically closer to Greece EN Peloponnese. This result could imply that the Early Neolithic Sicilians derive from an ancestral lineage that falls outside the genetic variation of other Early Neolithic groups in Europe. However, for statistics of the form f_4*(Chimp, Sicily EN; Greece EN Peloponnese, X)*, Sicily EN shares more alleles with various EN groups from the Balkan (Macedonia, Serbia, Croatia, Romania) and Central Europe than with Greece EN Peloponnese (fig. S6.4). Since the proportion of HG ancestry in Sicily EN exceeds that in Greece EN Peloponnese, the HG ancestry may cause outgroup attraction in the statistic f_4*(Chimp, X; Greece EN Peloponnese, Sicily EN*), forcing the outcomes more negative. We can, however, not exclude the possibility that the early farmer ancestry in Sicily EN falls partly outside of the broader genetic diversity of the early farmers from Europe, Greece EN Peloponnese or Anatolia Barcin. Assuming that the Late Mesolithic HGs from Sicily and the Balkan substantially overlapped both in lithic industry and genetic composition, parallel admixture events from local HGs in these regions with incoming early farmers would result in similar ancestry profiles for the early farmer groups from the Balkan and Sicily, respectively.

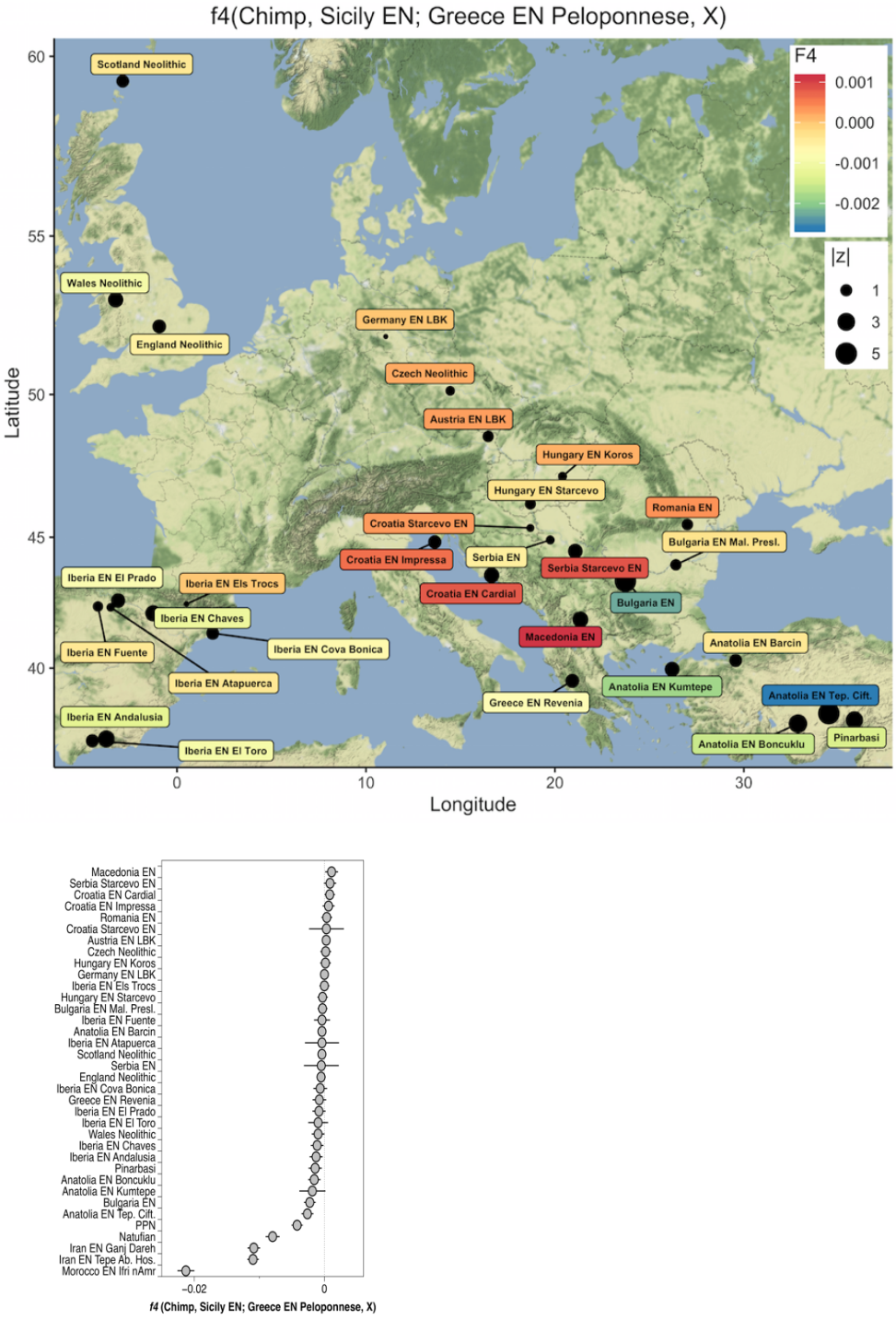


fig. S6.4. Cladality tests for various Early Neolithic groups from West Eurasia and Greece EN Peloponnese with respect to Sicily EN, of the form: *f_4(Chimp, Sicily EN; Greece EN Peloponnese, X).* Warmer colours correspond to positive f_4-values and indicate that Sicily EN gene pool is closer to that of the Early Neolithic test group *X*, whereas cooler colours correspond to negative f_4-values and indicate Sicily EN is closer to Greece EN Peloponnese. Compared to Greece EN Peloponnese, Early farmers from the Balkan and Central Europe share a moderate excess of alleles with Sicily EN. Dot sizes reflect |z|-scores and error bars 2 SEs.

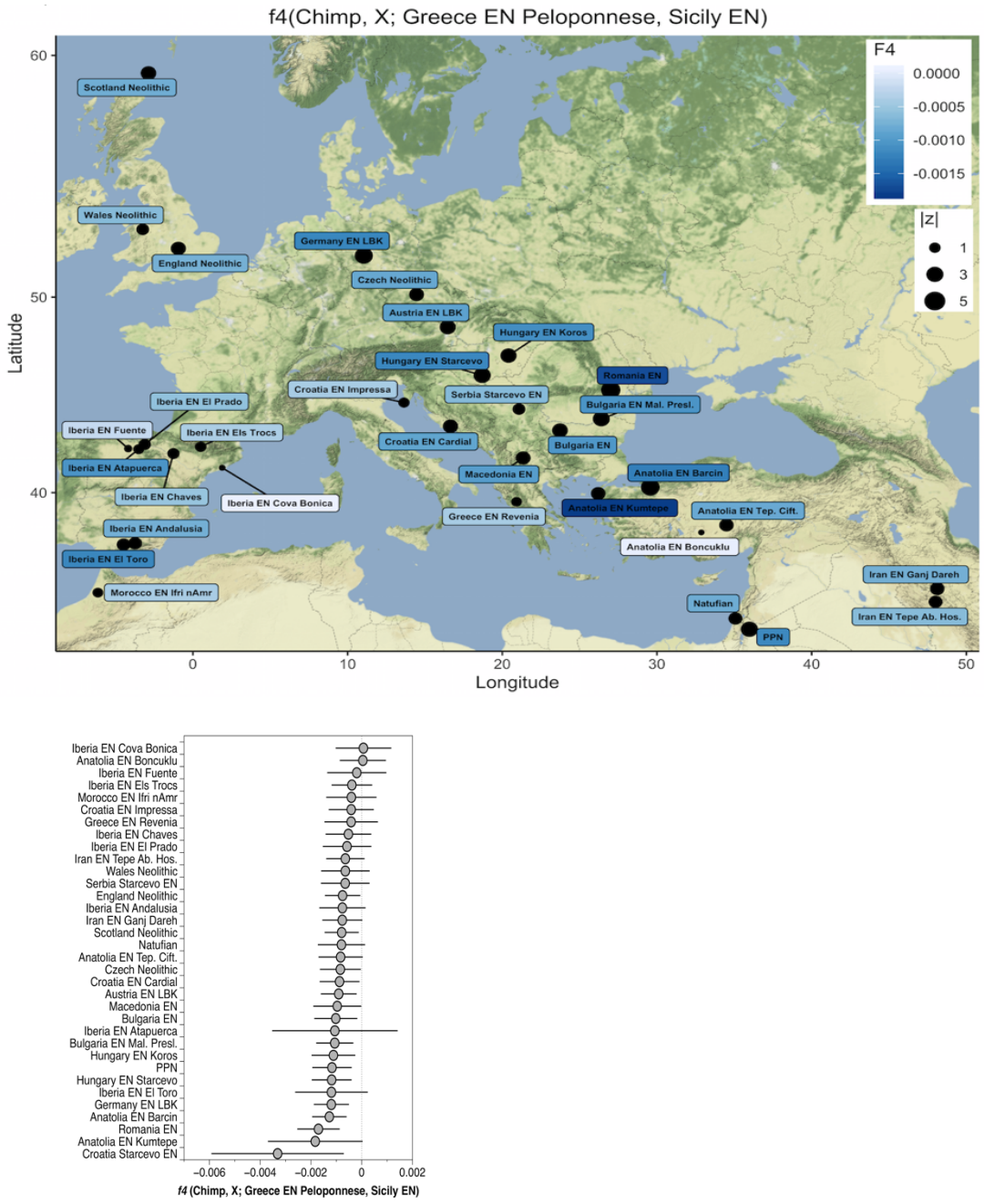


fig. S6.5. Comparing the ancestry in Greece EN Peloponnese and Sicily EN to various early farmer groups (*X*), using an f_4-cladality statistic of the form *f_4(Chimp, X; Greece EN Peloponnese, Sicily EN).* All the early farmer groups are either symmetrically related to Sicily EN and Greece EN Peloponnese (white) or genetically closer to the latter (dark blue). Dot sizes reflect |z|-scores and error bars 2 SEs.

S7: Uniparental marker haplotyping

A. Mitogenome haplotypes

We could reconstruct the mitochondrial genomes for 17 individuals (table S7.1, 98-100% genome coverage, mean base coverage 7 - 1,034X).

Sicily EM HGs

The two oldest HGs in our dataset, *UZZ5054*, *UZZ96* carried mitogenome lineages that fall within the U2'3'4'7'8'9 branch, and show a high similarity to the U2'3'4'7'8'9 haplotype that was previously reported for an Epigravettian HG from OrienteC (*I2158 - OrienteC*, (*15, 17*) (table S7.2). The three HGs have nine lineage-specific mutations in common and differently relate to each other with regard to three additional private mutations (table S7.2). U2'3'4'7'8'9 mitogenome lineages have been reported for Upper Palaeolithic European HGs associated with the Gravetian in Italy (*Paglicci108)*, Magdalenian in France (*Rigney*) and Azilian in Spain (*Balma Guilanyà*) (*50, 52*).

Individual ID	Site	^{14}C age (calBCE)	Study	Genetic group label	Genetic sex	Mitogenome haplogroup	% reference covered	Mean base coverage (X)	Sd. base coverage (X)
I2158	OrienteC	12,836-7,923 (strat.)	(*15, 17*)	Sicily EM	F	U2'3'4'7'8'9	55.6	2.5	2.0
UZZ5054	Uzzo	8,790-8,635	this study	Sicily EM	F	U2'3'4'7'8'9	100	401.4	122.1
UZZ96	Uzzo	9,150-6,550 (strat.)	this study	Sicily EM	F	U2'3'4'7'8'9	98.4	6.6	3.4
UZZ69	Uzzo	6,753-6,609	this study	Sicily LM	F	U5b2b	100	17.8	6.0
UZZ4446	Uzzo	6,599-6,477	this study	Sicily LM	F	U5b2b	100	326	95.1
UZZ71	Uzzo	n.a	this study	Sicily LM	F	U5a2+16294	100	259	75.9
UZZ79	Uzzo	6,684-6,596	this study	Sicily LM	F	U5b3d	100	96.4	24.7
UZZ88	Uzzo	5,989-5,850	this study	Sicily LM	F	U5b3d	100	106.6	31.2
UZZ81	Uzzo	6,682-6,595	this study	Sicily LM	M	U5b3d	100	39.4	12.3
UZZ80	Uzzo	6,683-6,596	this study	Sicily LM	F	U5b2b1a	100	344.2	86.8
UZZ82	Uzzo	6,628-6,481	this study	Sicily LM	F	U5a1	100	889.5	228.5
UZZ40	Uzzo	6,416-6,251	this study	Sicily LM	M	U4a2f	100	1034.3	248.8
UZZ61	Uzzo	n.a	this study	Sicily EN	M	K1a2	100	27.1	8.2
UZZ77	Uzzo	n.a	this study	Sicily EN	F	H	100	493.8	134.4
UZZ33	Uzzo	5,570-5,180 (strat.)	this study	Sicily EN	M	U8b1b1	100	521.5	139.7
UZZ34	Uzzo	5,461-5,231	this study	Sicily EN	F	U8b1b1	100	247.2	87.9
UZZ74	Uzzo	5,326-5,220	this study	Sicily EN	F	N1a1a1	99.6	32.3	12.6
UZZ75	Uzzo	5,327-5,220	this study	Sicily EN	F	J1c5	100	221.6	65.2

table S7.1. Details on the reconstructed mitogenomes and the assigned haplogroups.

Individual ID	MT haplogroup	Variants for called MT haplogroup (against rCRS)	Private mutations
UZZ5054	U2'3'4'7'8'9	73G 263G 750G 1438G A1811G 2706G 4769G 7028T 8860G 11467G 11719A 12308G 12372A 14766T 15326G	1406C 5999C 6152C 6498A 7403G 9991G 10020C 14152G 15466A 16274A 16297C
UZZ96	U2'3'4'7'8'9	73G 263G 750G 1438G A1811G 2706G 4769G 7028T 8860G 11467G 11719A 12308G 12372A 14766T 15326G	895T 5999C 6152C 6498A 7403G 10020C 14152G 15466A 16274A 16297C
I2158/ OrienteC	U2'3'4'7'8'9	750G 11719A 12308G 12372A 14766T 15326G	14152G 15466A 16274A 16297C

table S7.2. **Details on the mitogenome haplotypes for the Sicilian EM HGs.**

Private mutations	*OrienteC (I2158)*	*UZZ5054*	*UZZ96*
895T	no coverage	absent! 517X	9X
1406C	absent! 1X*	321X	absent! 6X
5999C	no coverage	381X	12X
6152C	no coverage	321X	7X
6498A	no coverage	435X	9X
7403G	no coverage	351X	1X*
9991G	no coverage	208X	absent! 3X
10020C	no coverage	150X	1X*
14152G	2X	357X	2X
15466A	2X	241X	1X*
16274A	2X	156X	3X
16297C	1X*	109X	2X

table S7.3. Private mutations with their coverage for the U2'3'4'7'8'9 mitogenome sequences in the Sicily EM HGs. When a private mutation is absent, the coverage for the reference allele is given.

Sicily LM HGs

We found that all the individuals in the Sicily LM genetic group carried U4a, U5a, and U5b mitogenome haplogroup lineages. All of these are characteristic for West Eurasian Mesolithic HGs (*52, 128*).

Two Castelnovian-associated HGs carried haplogroup U5b2b and one a more derived variant U5b2b1a (table S7.4). The individuals who harboured U5b2b (*UZZ69* and *UZZ4446*) shared five private mutations (5585A, 9833C, 12477C, 16311C, 16355T). None of these mutations are typically found on a more derived branch, including U5b2b1a. U5b2b haplotypes were frequently observed among Villabruna cluster individuals high in WHG ancestry (*52*). The oldest individuals found so far to have carried U5b2b are two Italian Epigravettian individuals from *Grotta Paglicci* and Villabruna, and two Epipalaeolithic HGs from *Rochedane* and *Aven des Iboussières* in France (*52*). The haplogroup was also found in low

frequency among Mesolithic HGs from southeastern Europe such as Croatia and Iron Gates fishermen from Serbia (~7,300-6,000 calBCE (*17*).

We also found haplogroup U5b3/U5b3d in two Castelnovian-associated HGs and in one individual tentatively contemporanous to early Impressa Ware (table S7.4). Notably, these individuals carried only one of the three expected variants that define U5b3d, and had three additional mutations in common (11836G, 16278T, 16385G). The two Castelnovian HGs, a genetic male (*UZZ79*) and female (*UZZ81*) also show a pairwise mismatch rate (PMMR) for autosomal SNP sites that is half of that found for unrelated individuals from this time period (see Extended Data Table-4). This underlines a first-degree genetic relatedness for these two individuals via at least the maternal side. Interestingly, the U5b3/U5b3d haplogroup has not been reported in European Mesolithic HGs thus far. However, Pala et al. (*129*) suggested an origin for U5b3 in the Italian Peninsula based on their analysis on the mitochondrial DNA variation observed among modern individuals. Notably, U5b3 has been found in an early Cardial farmer from the *El Portalon* cave at Sierra de Atapuerca in Spain, with a high amount of local HG ancestry (*77*). Additional sampling of Sicily Mesolithic HGs should indicate whether this haplogroup can be viewed as a general maternal lineage for the Mesolithic population in Sicily, or whether the individuals sampled here are genetic isolates.

In addition, we found U5a haplogroups in one Castelnovian-associated HG (*UZZ82*) and one individual tentatively contemporaneous to Impressa Ware (*UZZ71*) (table S7.4). *UZZ82* carried U5a1 with three additional private mutations (1007C, 3865G, 9380A). The U5a1 haplogroup has been reported for Mesolithic HGs from Russia and northern Europe (*39, 130*). *UZZ71* harboured U5a2+16294, a basal lineage to U5a2a. The more basal U5a2 haplogroup has been found in two Mesolithic hunter-gatherers from *Los Closeaux* and *Les Vignolles* in France (*52, 130*). The more derived haplogroup U5a2a is found in relatively higher frequency among Mesolithic HGs in general, more specifically in those from Ukraine, Serbia and Romania (*17*).

Lastly, for one Castelnovian-associated HG, *UZZ40* we found the rare haplogroup U4a2f without one of the four expected variants (G15172A is missing, table S7.3). Intriguingly, haplogroup U4a2f has been found also in a Cardial Ware individual from *Cueva de Chaves*, Iberia (*131*). U4a haplogroups are mostly found among Mesolithic HGs from northern Europe, the Baltic and Russia (*17, 41, 130*).

Individual ID	MT haplogroup	Variants for called MT haplogroup (against rCRS)	Private mutations	Missing mutations
UZZ79	U5b3d	73G 150T 263G 750G 1438G 2706G 3197C 4769G 7028T 7226A 7768G 8860G 9477A 11467G 11719A 12308G 12372A 13617C 14182C 14766T 15326G 16192T 16270T 16304C 16311C!	11836G 16278T 16385G	13830C 16067T
UZZ81	U5b3d	73G 150T 263G 750G 1438G 2706G 3197C 4769G 7028T 7226A 7768G 8860G 9477A 11467G 11719A 12308G 12372A 13617C 14182C 14766T 15326G 16192T 16270T 16304C 16311C!	11836G 16278T 16385G	13830C 16067T
UZZ88	U5b3d	73G 150T 263G 750G 1438G 2706G 3197C 4769G 7028T 7226A 7768G 8860G 9477A 11467G 11719A 12308G 12372A 13617C 14182C 14766T 15326G 16192T 16270T 16304C 16311C!	11836G 16278T 16385G	13830C 16067T
UZZ80	U5b2b1a	73G 150T 263G 750G 1438G 1721T 2706G 3197C A3861G 4769G 7028T 7768G 8860G 9477A 11467G 11653G 11719A 12308G 12372A 13617C 12634G 13630G 13637G 14182C 14766T 15326G 15497A 16192C! 16270T 16362C		
UZZ4446	U5b2b	73G 150T 263G 750G 1438G 1721T 2706G 3197C 4769G 7028T 7768G 8860G 9477A 11467G 11653G 11719A 12308G 12372A 13617C 12634G 13630G 13637G 14182C 14766T 15326G 16192C! 16270T	5585A 9833C 12477C 16311C 16355T	
UZZ69	U5b2b	73G 150T 263G 750G 1438G 1721T 2706G 3197C 4769G 7028T 7768G 8860G 9477A 11467G 11653G 11719A 12308G 12372A 13617C 12634G 13630G 13637G 14182C 14766T 15326G 16192C! 16270T	5585A 9833C 12477C 16311C 16355T	
UZZ82	U5a1	73G 263G 750G 1438G 2706G 3197C 4769G 7028T 8860G 9477A 11467G 11719A 12308G 12372A 13617C 14766T 14793G 15218G 15326G 16192T 16256T 16270T A16399G	1007C 3865G 9380A	
UZZ71	U5a2 + 16294	73G 263G 750G 1438G 2706G 3197C 4769G 7028T 8860G 9477A 11467G 11719A 12308G 12372A 13617C 14766T 14793G 15326G 16256T 16270T 16294T 16526A	3523G 5460A 15297C 16192C!	
UZZ40	U4a2f	73G 195C! 263G 310C 499A 750G 1189C 1438G 1811G 1978G 2706G 4646C 4769G 5999C 6047G 7028T 8818T 8860G 11332T 11467G 11719A 12308G 12372A 12397G 14620T 14766T 15326G 15693C 16356C		15172A

table S7.4. Details on the mitogenome haplotypes for the Sicily LM HGs.

Sicily EN

The early Sicilian farmers in our transect harboured mitogenome haplogroups characteristic for early farmers: U8b1b1 (n=2), K1a2 (n=1), N1a1a1 (n=1), H (n=1), and J1c5 (n=1) (table S7.5). All these haplogroups have previously been reported in early farmers from the Balkan, and in aceramic and ceramic Neolithic individuals from *Barcin* in north-western Anatolia (*17, 39*). Subsets of these were found among early farmers from all over Europe, albeit in different combinations and frequencies in the Balkan, Central Europe and Iberia (*132*).

U8b1b1, found in two of the early Sicilian farmers, has been reported for Starcevo early farmers from Croatia (17). Haplogroup K1a2 has been reported for early farmers from Romania, Germany LBK and northern Greece (*17, 30, 38*). In addition, K1a2 and the derived K1a2a haplogroup appear frequently among early farmers from Iberia. This includes a ~5,400 calBCE Cardial individual from *Cova Bonica* and a ~5,100 calBCE Epicardial individual from *Cova de Els Trocs* in northeastern Spain, and a ~5,000 calBCE individual from *Cueva del Toro* in southern Spain associated with ‘boquique’ and ‘almagra’ technique pottery (*26, 44, 54*).

The rare haplogroup N1a occurs at a relatively high frequency in LBK early farmers from Central Europe, but is much lower in Iberia (*132-134*). The N1a1a1 haplotype that we found in one Sicilian farmer was reported in Germany EN LBK and Hungary EN Starcevo farmers, and for one individual from *Cova de Els Trocs* (*17, 38, 54*). Interestingly, the more basal haplogroup N* was found in three Early Neolithic Cardial farmers from the *Can Sadurní* Cave in Catalonia, northern Spain.

Individual ID	MT haplogroup	Variants for called MT haplogroup (against rCRS)	Private mutations
UZZ33	U8b1b1	73G 195C! 263G 750G 1438G 1811A! 2706G 3480G 4769G 5165T 7028T 8860G 9055A 9698C 11467G 11719A 12308G 12372A 14053G 14167T 14766T 15326G 16189C! 16234T 16324C	
UZZ34	U8b1b1	73G 195C! 263G 750G 1438G 1811A! 2706G 3480G 4769G 5165T 7028T 8860G 9055A 9698C 11467G 11719A 12308G 12372A 14053G 14167T 14766T 15326G 16189C! 16234T 16324C	
UZZ61	K1a2	73G 263G 497T 750G 1189C 1438G 1811A! 2706G 3480G 4769G 7028T 8860G 9055A 9698C 10398G! 10550G 11025C 16224C 16311C! 11299C 11467G 11719A 12308G 12372A 14167T 14766T 14798C 15326G	152C! 9604G
UZZ77	H	263G 750G 1438G 2706A 4769G 7028C 8860G 15326G	
UZZ75	J1c5	73G 185A 228A 263G 295T 462T 489C 750G 1438G 2706G 3010A 4216C 4769G 5198G 7028T 8860G 10398G! 11251G 11719A 12612G 13708A 14766T 14798C 15326G 15452a 16069T 16126C	
UZZ74	N1a1a1	199C 204C 669C 750G 1438G 1719A 2702A 2706G 3336C 4769G 5315G 7028T 8860G 8901G 10238C 10398G! 11719A 12501A 12705T 13780G 14766T 15043A 15326G 16147A 16172C 16223T 16248T 16248T	5460A 11884G

table S7.5. **Details on the mitogenome haplotypes for the Sicilian early farmers.**

B. Y-chromosome haplotypes

We could determine the Y-haplogroup for four males (table S7.6). Two Sicilian LM HGs associated with the Castelnovian carried haplogroups I and I2a2, which both are characteristic for Upper Palaeolithic and Mesolithic HGs from West-Eurasia. Haplogroup I is commonly found among individuals associated with the Gravettian and part of the Vestonice genetic cluster, such as *Paglicci133* from Italy and *KremsWA3* from lower Austria, and in Magdalenian-associated individuals, such as *Hohlefels49* and *Burkhardtshohle (16).* In addition, I and the more derived I2, and I2a haplogroups are the most frequent haplotypes found among European Mesolithic HGs related to the Villabruna cluster. This includes from France the Epipaleolithic *Rochedane* (haplogroup I), Mesolithic *Chaudardes* (haplogroup I) and Mesolithic *BerryAuBac* (haplogroup I), from Hungary the *Koros* individual from a Neolithic context but an ancestry profile characteristic for WHG (haplogroup I2a), and Mesolithic *Loschbour* (haplogroup I2a1b) from Luxembourg (*16, 37, 39*). Moreover, haplogroup I2 was reported for *Bichon* from Switzerland associated with the Azilian (*32*). The haplogroup I2a2 that we find in one Sicily LM HG occurred relatively frequently among Mesolithic HGs from the Iron Gates and Latvia (*17*).

The two Sicilian early farmers carried haplogroups C1a2 and H. Interestingly, although C1 haplotypes are found among early farmers, these are in general considered to be more typical for pre-

Neolithic West-Eurasians. One of the oldest individuals to have carried C1a are *GoyetQ116* and *Vestonice16* associated with the Gravettian *(16).*The more derived haplogroup C1a2 that we find in one of the Sicilian early farmers, was found in the Mesolithic *LaBrana* HG from Iberia (*42*). However, C1a2 haplotypes have also been reported for early farmers from *Barcin* and *Tepe Ciflik*, as well as in the ~13,300 calBCE Pınarbaşı HG in Anatolia (*25, 34, 39*). C1a2 haplotypes were also found in early Cardial farmers from Croatia and early LBK farmers from Austria (*17*).

The Y-haplogroup H that we find in one of the Sicilian early farmers has been proposed to be among the genetic markers of the early farmer populations of Middle East that were introduced to Europe during the Neolithic transition (*135*).

Individual ID	Genetic group label	Mitogenome haplogroup	Y-chromosome haplogroup	Y derived SNPs supporting haplogroup determination
UZZ40	Sicily LM	U4a2f	I	I: L578 L755 L758 CTS48 CTS646 CTS7502 CTS8742 CTS9860 PF3640 PF3660 PF3665 PF3668 PF3796 PF3797 PF3809 PF3817 PF3822 PF3837 FGC2412 FGC2414
UZZ81	Sicily LM	U5b3/U5b3d	I2a2	I: V218.2 L578 L751 L755 L758 CTS4848 CTS6231 CTS6265 CTS7329 CTS7831 CTS8876 CTS9618 CTS11540 PF3640 PF3661 PF3668 PF3814 PF3837 FGC2413 I2a2: M436 P217 P218 L35 L37
UZZ33	Sicily EN	U8b1b1	H	HIJK: M578 H: M2713 M2896 M2936 M2945 M2992 M3035 M3058 M3062 M3070 Z4309
UZZ61	Sicily EN	K1a2	C1a2	C: P255 P260 V77 V183 V199 V232 C1a: CTS11043 C1a2: Z28922

table S7.6. Details on the Y-chromosome haplogroup assignments for genetic males.