LETTER – DISCOVERIES

# Mitogenome Diversity in Sardinians: a Genetic Window onto an Island's Past

Anna Olivieri,†,[1] Carlo Sidore,†,[2,3,4] Alessandro Achilli,†,[1] Andrea Angius,[2,4,5] Cosimo Posth,[6,7] Anja Furtwängler,[7] Stefania Brandini,[1] Marco Rosario Capodiferro,[1] Francesca Gandini,[1,8] Magdalena Zoledziewska,[2] Maristella Pitzalis,[2] Andrea Maschio,[2,3] Fabio Busonero,[2,3] Luca Lai,[9] Robin Skeates,[10] Maria Giuseppina Gradoli,[11] Jessica Beckett,[12] Michele Marongiu,[2] Vittorio Mazzarello,[4] Patrizia Marongiu[4], Salvatore Rubino[4], Teresa Rito,[13] Vincent Macaulay,[14] Ornella Semino,[1] Maria Pala,[8] Gonçalo R. Abecasis,[3] David Schlessinger,[15] Eduardo Conde-Sousa,[16] Pedro Soares,[16] Martin B. Richards,[8] Francesco Cucca,*,[2,4] and Antonio Torroni*,[1]

[1]Dipartimento di Biologia e Biotecnologie, Università di Pavia, 27100 Pavia, Italy.

[2]Istituto di Ricerca Genetica e Biomedica (IRGB), CNR, 09042 Monserrato, Italy.

[3]Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, Michigan, USA.

[4]Dipartimento di Scienze Biomediche, Università di Sassari, 07100 Sassari, Italy.

[5]Center for Advanced Studies, Research and Development in Sardinia (CRS4), AGCT Program, Parco Scientifico e Tecnologico della Sardegna, 09010 Pula, Italy.

[6]Max Planck Institute for the Science of Human History, 07745 Jena, Germany.

[7]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, 72070 Tübingen, Germany

[8]Department of Biological Sciences, School of Applied Sciences, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK.

[9]Department of Anthropology, University of South Florida, Tampa, FL 33620, USA.

[10]Department of Archaeology, Durham University, Durham DH1 3LE, U.K.

[11]School of Archaeology and Ancient History, University of Leicester, Leicester LE1 7RH, UK.

[12]Independent Contractor, Cagliari, Italy.

[13]Life and Health Sciences Research Institute (ICVS), School of Health Sciences & ICVS/3B's-PT Government Associate Laboratory, University of Minho, 4710-057 Braga, Portugal.

[14]School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8QQ, UK.
[15]Laboratory of Genetics, National Institute on Aging, US National Institutes of Health, Baltimore, MD 21224-6825, Maryland, USA.
[16]CBMA (Centre of Molecular and Environmental Biology), Department of Biology, University of Minho, Campus de Gualtar, 4710–057 Braga, Portugal.

†These authors contributed equally to this work.
***Corresponding authors**: E-mails: fcucca@uniss.it; antonio.torroni@unipv.it.

# Abstract

**Sardinians are "outliers" in the European genetic landscape and, according to paleogenomic nuclear data, the closest to early European Neolithic farmers. To learn more about their genetic ancestry, we analyzed 3,491 modern and 21 ancient mitogenomes from Sardinia. We observed that 78.4% of modern mitogenomes cluster into 89 haplogroups that most likely arose *in situ*. For each Sardinian-Specific Haplogroup (SSH), we also identified the upstream node in the phylogeny, from which non-Sardinian mitogenomes radiate. This provided minimum and maximum time estimates for the presence of each SSH on the island. In agreement with demographic evidence, almost all SSHs coalesce in the post-Nuragic, Nuragic and Neolithic-Copper Age periods. For some rare SSHs, however, we could not dismiss the possibility that they might have been on the island prior to the Neolithic, a scenario that would be in agreement with archeological evidence of a Mesolithic occupation of Sardinia.**

*Key words:* **mitochondrial genomes, mitochondrial DNA phylogeny, haplogroups, prehistory of Sardinia, origins of Europeans**

## Main Text

Sardinia is an island that remained unconnected with the mainland even when the sea level was at its lowest during the LGM (Shackleton et al. 1984) and probably was the last of the large Mediterranean islands to be colonized by modern humans (Sondaar 1998). Modern Sardinians, a unique reservoir of distinct genetic signatures (Cavalli-Sforza et al. 1984; Pala et al. 2009; Francalacci et al. 2013; Sidore et al. 2015), on one hand apparently harbour the highest levels of nuclear genome similarity with European Neolithic farmers (Lazaridis et al. 2014) and an extensive similarity with the Late Neolithic/Chalcolithic Tyrolean Iceman (Keller et al. 2012; Sikora et al. 2014) but, on the other hand, they differ substantially from Near Eastern Neolithic farmers including those from Anatolia (Lazaridis et al. 2016). These findings have led to the view that, in modern Europe, Sardinians may have best preserved the gene pool of Neolithic farmers, possibly because their ancestors were less affected by subsequent Bronze Age dispersals across Europe (Haak et al. 2015).

Note that the above view does not necessarily imply that the first Sardinians were Neolithic farmers. On the contrary, there is archeological evidence indicating that humans were present on the island by at least 13 Kya (Hofmeijer et al. 1989; Dyson and Rowland 2007; Broodbank 2013). Moreover, a European pre-Neolithic origin for Y-chromosome haplogroup I2a1a1-M26, by far the most common in modern Sardinian males (38.9%) (Francalacci et al. 2013), has been postulated (Rootsi et al. 2004). Finally, a massive survey of whole-genome sequences from modern Sardinians has recently shown that the population of the mountainous Gennargentu region harbors higher levels of both hunter-gatherer and Neolithic farmer components relative to other Sardinian groups from less isolated areas. This has been interpreted as indicating that the hunter-gatherer component did not reach the island with more recent migrations from the continent, but it was either already present on the island when farmers arrived or due to previous admixture of the first incoming farmers with hunter-gatherers on the mainland (Chiang et al. 2016).

In this complex and partially contradictory scenario, the genetic perspective of the maternally transmitted mitochondrial genome is still almost completely unexplored. Therefore, to learn more about the ancestry of Sardinians and their genetic links with modern and ancient European (and other) populations, we analyzed a large dataset of 3,491 novel complete mitogenomes from modern islanders as well as 21 mitogenomes from ancient specimens. Among the modern samples, we removed 1,355 maternally related samples on the bases of pedigree data or kinship evaluation of nuclear genomes, and 44 samples with non-Sardinian maternal origins. We then assessed the phylogenetic relationships of the remaining mitogenomes (2,092 out of the initial 3,491), plus 124 previously published Sardinian mitogenomes, with all publicly available worldwide mitogenomes (more than 26,000, data not shown).

## Sardinian-Specific Haplogroups and Their Coalescence Ages

Our phylogenetic analyses revealed that 1,737 Sardinian mitogenomes (78.4%) clustered into 89 Sardinian-Specific Haplogroups (SSHs; see Methods for defining criteria) (Supplementary table S1; Supplementary fig. S1). For each SSH, using non-Sardinian mitogenomes, we also identified the upstream node in the phylogeny, from which non-Sardinian Closest External Mitogenomes (CEMs) radiate (Supplementary table S2, Supplementary fig. S2).

The finding that 78.4% of Sardinians harbor ethnic-specific haplogroups might appear surprising in the European context. However, a similar - though not so extreme - scenario is seen in the Basque-speaking regions of Spain, where a survey of haplogroup H mitogenomes (54.1% of the populations) identified six autochthonous sub-haplogroups encompassing 29.0% of all mitochondrial DNAs (Behar et al. 2012).

The 89 SSHs, 80 of which are defined here for the first time, include descendants from all major macro-haplogroups of the human mitochondrial DNA (mtDNA) tree (L, M, N and R) and are defined by a total of 104 mutations (Supplementary table S3). About 51% of modern Sardinian-specific mtDNAs fall into HV, 27% into JT, 17% into U and 5% into other lineages. These frequencies are close to those reported in typical western European populations. However, when assessed at a higher level of haplogroup resolution, they differ substantially from those in continental Europe. This is most marked for H1 and H3, with peak values on the island of 18.5% and 18.4%, respectively (Supplementary table S4).

It is most likely that the SSHs and their distinguishing mutational motifs arose *in situ* (fig. 1, panel A), even though the possibility that some of these motifs arose outside the island and, after their arrival in Sardinia, were lost in the ancestral sources, should not be overlooked. However, even in this case, the coalescence age of the SSH corresponds to the minimum time estimate for the presence of its founder mutational motif on the island (fig. 1, panel B). An overestimation of the SSH arrival/presence time in Sardinia would occur only if the founder haplotype as well as a derived haplotype both moved to Sardinia, but were also both lost in the ancestral source (fig. 1, panel D). This is not a very likely scenario, not only because it requires multiple events, but also given the great diversity of some of the deep-rooting founding lineages – for example, within the predominant haplogroups H1 and H3, where lineages are very sharply partitioned between the island and the mainland.

For all 89 SSHs, we assessed these minimum coalescence ages with both Maximum Likelihood (ML) and BEAST (Bayesian Evolutionary Analysis Sampling Trees) computations, employing two different mutation rates: one established on modern mitogenomes, which corrects for the effect of selection and is routinely applied in phylogeographic studies (Soares et al. 2009); the other using radiocarbon dated ancient mitogenomes as tip calibration points (Posth et al. 2016) (Supplementary table S3 and Supplementary fig. S2). In agreement with historical demographic evidence (Francalacci et al. 2013), all estimates indicate that more than 50% of SSHs coalesce in the post-Nuragic (<2 Kya) and Nuragic (~2-4 Kya) archaeological periods. However, not all of the remainder fall in the Neolithic-Copper Age period (~4-7.8 Kya) (Supplementary fig. S2). In particular, three rather rare SSHs, K1a2d, N1b1a9 and U5b1i1, corresponding to 3.1% of modern Sardinian-specific mitogenomes, showed with all or some of the approaches a mean coalescence age >7.8 Ky (table 1).

## A Pre-Neolithic Presence in the Island for Some Sardinian-Specific Haplogroups?

The postulated archeologically-based starting time of the Neolithic in Sardinia is 7.8 Ky (Berger and Guilaine 2009). Taking into account that coalescence ages correspond to the lower bound for their presence on the island, the observation that some SSHs might coalesce prior to that boundary, raises the possibility that their founding haplotypes were already on the island during the Mesolithic (fig. 2; Supplementary fig. S2).

Haplogroup K1a2d includes nine mitogenomes and, with both ML and BEAST and both mutation rates, shows coalescence ages (11-16 Ky) that, even when standard errors are included, predate the Neolithic (table 1). K1a2d is only one mutation away from K1a2, the previously defined node (Supplementary table S3; Supplementary fig. S1). From this node several other sub-branches depart (fig. 3), and the members of these branches are CEMs to the Sardinian-specific branch. Most of the K1a2 sub-branches (K1a2a-d) encompass only European mitogenomes (Costa et al. 2013), but several descending directly from the root of K1a2 have been identified also in the Near East, and include ancient samples. The oldest are two K1a2 mitogenomes from early Anatolian farmers, radiocarbon-dated to 8.3 and 8.0 Kya (Supplementary table S5). These observations indicate the Near East as the most likely ancestral source of K1a2. If so, our observations could be interpreted as indicating that K1a2 (female) carriers of Near Eastern ancestry arrived in Sardinia in the time frame between 18.7-14.5 Kya (ages of K1a2) and 16.0-11.0 Kya (ages of K1a2d) – that is, in Late Glacial times (Supplementary table S3).

Haplogroup U5b1i1 includes 41 mitogenomes. Its mean coalescence ages are in the range of 9.7-13.0 Ky, although in most cases standard errors overlap with the arrival of the Neolithic in Sardinia (table 1). In modern populations, the CEMs to the Sardinian-specific haplogroup U5b1i1 are those departing from the newly identified U5b1i node (with an estimated age in the range of 10.7-13.5 Ky) and are all from western Europe (1 from Germany, 2 from the UK) (fig. 3), the same geographic origin as the U5b1 mitogenomes (upstream of U5b1i) from ancient samples (Supplementary table S5). Thus, in contrast with K1a2d, this matrilineal genetic component harbors deep ancestral roots in Western Europe.

Haplogroup N1b1a9 includes only four mitogenomes. Its mean coalescence ages are in the range of 7.3-9.4 Ky with standard errors always overlapping with the Neolithic (table 1). N1b1a9 is only one mutation away from the previously defined node N1b1a (Supplementary table S3; Supplementary fig. S1), from which numerous other branches depart (fig. 3). Most of these branches are shared between Europeans and Near Easterners, indicating the Near East as the likely homeland of N1b1a. This source is further supported by the geographical origin (Anatolia) of the only ancient N1b1a mitogenome (8.3 Ky) recovered so far (Supplementary table S5). Thus, the (female) N1b1a carriers of Near

Eastern ancestry might have arrived in Sardinia in the time frame between 17-11 Kya (age of N1b1a) and 9-7 Kya (age of N1b1a9) (Supplementary table S3).

## Founder Analysis and Coalescent Simulations

To further evaluate the arrival/presence times of U5b1i1, K1a2 and N1b1a in Sardinia, we performed a founder analysis (Richards et al. 2000; Macaulay and Richards 2013; Soares et al. 2016). This method assumes a strict division between potential source populations and sink population, and subtracts the diversity within the sink dataset that arose in the source region. In our case, the potential sources for the Sardinian mitogenomes were their modern CEMs (fig. 3). The migration scan, which plots the fraction of arriving lineages against time, showed single primary peaks at 17.0, 13.2, and 8.0 Kya for haplogroups K1a2, U5b1i1 and N1b1a, respectively (Supplementary fig. S3). Taking into account that founder analysis is conservative in that it provides only minimum estimates for the arrival time of each founder lineage, since the arrival necessarily predates the origin and expansion of the corresponding founder cluster, these peaks tend to support a pre-Neolithic presence of U5b1i1 and K1a2 in Sardinia. For N1b1a, the result does not completely rule out such a possibility, but makes it less plausible.

We then performed coalescent simulations under different demographic models (Hudson 2002) to test the two alternative scenarios: i) a single Neolithic occupation of the island at 8 Kya; ii) a first entry in the Late Paleolithic followed by another in the Neolithic. These simulations supported a dual migration scenario, with a first migration event in the Late Paleolithic at 12–15 Kya (effective population size between 500 and 1,500), followed by an Early Neolithic migration at about 8 Kya (effective population size of 35,000) (Supplementary table S6).

## Ancient Sardinian Mitogenomes

To further investigate the genetic ancestry of Sardinians, 21 prehistoric mitogenomes, from skeletal remains collected in a number of different rock-cut tombs, megalithic tombs, caves and rock shelters (Supplementary Materials), were also reconstructed and analyzed. Unfortunately, they were from the cultural phases of Sardinia between the Neolithic and the

Nuragic Final Bronze Age with radiocarbon datings in the range of 6.1 to 3.0 Kya, thus they could not shed further light on the issue of the potential pre-Neolithic presence of some SSHs on the island.

They harbored 21 distinct haplotypes in 19 sub-haplogroups belonging to macro-haplogroups R0, JT and U (Supplementary table S7 and Supplementary fig. S4). These haplotypes were compared with those from modern Sardinians and with 417 ancient mitogenomes available in the literature (Supplementary table S5 and Supplementary fig. S5). The sub-haplogroups observed in ancient Sardinians are also present in modern Sardinians at the same (N = 15) or at a very close (N = 4) level of haplogroup resolution (Supplementary table S8). None of the ancient Sardinian mitogenomes clustered within a Sardinian-specific haplogroup, but four were closely related (Supplementary fig. S6). A Sardinian Bell Beaker mitogenome (MA108) of ~4.3 Kya turned out to be a member of a novel branch (HV0j1: 6.6±1.3 Ky), which was found in both a modern Sardinian and a continental Italian, and derives from a node (HV0j: 10.0±2.1 Ky) from which two other Sardinian mitogenomes diverge (Supplementary fig. S6). The phylogenetic age of HV0j1 is thus fully compatible with the radiocarbon dating of MA108. A similar conclusion emerges in all other cases in which an informative phylogenetic link between modern and ancient Sardinian samples was established: (i) MA78 (Early Bronze Age, ~4.0 Kya) is a direct molecular ancestor of the Sardinian-specific haplogroup H3u2 (~3.2 Kya); (ii) MA104 (Early Bronze Age, ~4.5 Kya) harbors one of the diagnostic mutations of the SSH K1a32 (~6.8 Kya); and (iii) MA88 (Early Bronze Age, ~4.2 Kya) shows the mutational motif of haplogroup U5b2b5 node (~12.6 Kya), from which the SSH U5b2b5a (~3.2 Kya) as well as other mtDNAs from Sardinia, Italy and the UK descend (Supplementary fig. S6). Ancient DNA links extend beyond Sardinia. A Copper Age (~5 Kya) mitogenome from Northern Spain (ATP16 in Günther et al. 2015) identifies Iberia as the likely homeland of the molecular ancestor of the SSH X2c2a (Supplementary fig. S6), and indicates that the founder mtDNA arrived in Sardinia between 8.9±1.7 Kya (age of X2c2) and 5.7±1.8 Kya (age of X2c2a) (Supplementary fig. S6).

We also compared modern and ancient Sardinian mitogenomes with the mitogenome (haplogroup K1f) of the Late Neolithic/Chalcolithic Tyrolean Iceman (radiocarbon-dated to

~5.3 Kya) (Ermini et al. 2008). One Sardinian-specific haplogroup (K1g1), present in ~2.1 % of Sardinians, is indeed related to Otzi's mitogenome as well as to other mitogenomes found in modern and ancient Europeans. However, the link is extremely distant in time, at the level of a very early node (K1 + 16362), which is only one mutation away from the root of K1 and is dated ~23.5 Ky (Supplementary fig. S7).

## Origins of the Most Ancient Sardinian-Specific Haplogroups

As described above, our analyses raise the possibility that several SSHs may have already been present on the island prior to the Neolithic. The most plausible candidates would include K1a2d and U5b1i1, which together comprise almost 3% of modern Sardinians, and possibly others that might have arrived at an early date but expanded with the Neolithic. This scenario remains uncertain for two reasons: (i) K1a2d is the only SSH for which the standard errors of the coalescence ages never overlap with the arrival of the Neolithic in Sardinia; (ii) the possibility illustrated in panel D of figure 1 (the founder haplotype as well as a derived haplotype both moved to Sardinia, but were also both lost in the ancestral source), which, even if rather unlikely, can not be ruled out. However, at the same time our analyses show that the scenario of a pre-Neolithic presence of one or more SSHs in Sardinia cannot be easily dismissed either.

Such a scenario would not only support archeological evidence of a Mesolithic occupation of Sardinia, and recent genome-wide studies, but could also suggest a dual ancestral origin of its first inhabitants. K1a2d is of Late Paleolithic Near Eastern ancestry, whereas U5b1i1 harbors deep ancestral roots in Paleolithic Western Europe, possibly paralleling the patrilineal source of the very frequent (38.9%) Y-chromosome haplogroup I2a1a1-M26 both in terms of geography and timing (Francalacci et al. 2013).

Recent genome-wide data from ancient specimens have shown that Palaeolithic Europeans from ~37 to ~14 Kya derive from a single ancestral population, but that this long-term genetic continuity was in part interrupted by the appearance of a novel genetic component related to modern Near Easterners ~14 Kya, during the first significant warming period (Bølling-Allerød interstadial) after the Last Glacial Maximum (LGM) (Fu et al. 2016)**.** The notion of a possible genetic input from the Near East into and across Europe in the late Pleistocene prior to the arrival of the Early Neolithic material culture in Greece ~8.5 Kya (Manning et al. 2014) is a novelty in human paleogenomics. Indeed, the prevailing conclusion of ancient DNA studies has been so far that Palaeolithic and Mesolithic hunter-gatherer European populations (characterized essentially only by haplogroups U8, U5 and U2 in terms of mitogenomes) differed genetically from early farmers, in turn implying that there was a wide-scale replacement across Europe from the Near East in the Early Neolithic, with limited assimilation of native Europeans (Pinhasi et al. 2012; Lazaridis et al. 2014; Omrak et al. 2016).

A pre-Neolithic genetic input from the Near East is, however, not a novelty in the context of phylogeographic studies of modern mtDNA variation. These have proposed that many modern European mtDNA lineages within haplogroups J, T, I, W and R0a, and possibly others, entered Europe in the Late Glacial and postglacial periods from the Near East before the migration waves associated with the onset of farming (Pala et al. 2012; Olivieri et al. 2013; Gandini et al. 2016; Richards et al. 2016); and that these haplogroups typical of modern Europeans, often assumed to have dispersed from Anatolia only with the advent of the Neolithic, were instead already present in Mesolithic Mediterranean Europe, particularly in Italy (Pereira et al. 2016). This scenario is in line with our findings concerning K1a2d in Sardinia as well as with the recent detection of two K1c mitogenomes in Mesolithic Greece (Hofmanová et al. 2016).

The potential genetic legacy of Mesolithic Sardinians could be even higher than the ~3% represented by K1a2d and U5b1i1. We should stress that the assumptions of our analysis are highly conservative in this regard, because every migration of a lineage away from Sardinia to the continent would here be recorded instead as the signal of a Sardinian founder event, thus reducing the age of the SSH. For example, a large fraction of the SSHs within H1 and H3 (Supplementary fig. S2), the two most common haplogroups in modern Sardinians (18.5% and 18.4%, respectively; Supplementary table S4), are only one (sometimes fast-evolving) mutation away from the H1 and H3 founding nodes and/or have CEMs departing from the H1 and H3 nodes. Therefore, their estimated coalescence ages represent upper bounds for the presence of H1 and H3 mitogenomes in the island.

As shown in fig. 2, the ages of H1 and H3 leave open the possibility that both were also present in Sardinia prior to the Neolithic. Notably, the frequency of H3 in Sardinia (18.4%) is the highest reported till now, and haplogroup H3 harbors a very peculiar geographical distribution. The highest frequencies are in western Mediterranean (Sardinians, Basques and other Iberians), with a sharp decrease towards Central and Eastern Europe and only very few occurrences in the Near East (fig. 4 and Supplementary table S9), which founder analyses explain as recent incursions. Given that the population size trends for the Sardinian H3 mtDNAs indicate an expansion beginning between 9.0 and 10.5 Kya (fig. 4), it is tempting to link such an expansion to a pre-Neolithic arrival and diffusion of H3 on the island, most likely from a Western Mediterranean source, as previously suggested (Achilli et al. 2004; Torroni et al. 2006; Soares et al. 2010), possibly the same ancestral source of the ancestors of U5b1i1 and Y-chromosome haplogroup I2a1a1-M26 (Francalacci et al. 2015).

Our detection of potential pre-Neolithic signals in the mitogenome pool of contemporary Sardinians remains to be tested with studies of ancient DNA. If confirmed, a pre-Neolithic presence of H3 (and possibly also of H1 and other lineages, for example within JT) on the island alongside K1a2d and U5b1i1 (and most likely other lineages that we have not detected) would indicate a more substantial genetic legacy of Mesolithic Sardinians to the modern people of Sardinia. However, it is also important to realise that even if H3 (and H1) arrived in Sardinia only with the Neolithic, they most likely came from either Spain or elsewhere in the western Mediterranean, and not from the Near East. This would imply that they are likely the result of autochthonous west Mediterranean Mesolithic acculturation, in the wider European context.

In conclusion, contemporary Sardinians harbor a unique genetic heritage as a result of their distinct history and relative isolation from the demographic upheavals of continental Europe. Whilst the major signal appears to be the legacy of the first farmers on the island, our results hint at the possibility that the situation might have been much more complex, both for Sardinia but also, by implication, for Europe as a whole. It now seems plausible that human mobility, inter-communication and gene flow around the Mediterranean from Late Glacial times onwards may well have left signatures that survive to this day.

Archeological evidence indicates that Mesolithic refugia persisted for many centuries in Italy and Iberia (Broodbank 2013), which, like the Near East and Caucasus, may have acted as a long-term refugial zone, as Gamble (2009) has suggested. These populations may have contributed to varying extents to the ancestry of the populations of subsequent millennia, not only around the Mediterranean but also into the heart of the continent. Although in the past the stress has often been on the spread of the Neolithic, genetic studies too are beginning to emphasize the complexity and mosaic nature of human ancestry in the Mediterranean, and indeed in Europe more widely. Future work on ancient DNA should be able to test directly to what extent this more complex model is supported by genetic evidence, and whether our predictions of Mesolithic ancestry in contemporary Sardinians can be sustained.

## Materials and Methods

All experimental and analytical procedures are described in the Supplementary method file, Supplementary Material online.

## Supplementary Material

Supplementary material is available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

## Acknowledgments

We are grateful to all the volunteers who generously participated in this study and made this research possible. This study received support from the University of Pavia strategic theme "Towards a governance model for international migration: an interdisciplinary and diachronic perspective" (MIGRAT-IN-G) (to A.O., A.A., O.S., A.T.), the British Academy Research Development Award n. 53097 (to R.S.), the Italian Ministry of Education, University and Research: Progetti Futuro in Ricerca 2012 (RBFR126B8I) (to A.O. and A.A.) and Progetti Ricerca Interesse Nazionale 2012 (to A.A., O.S. and A.T.). The sequencing data reported in the study were supported by the National Human Genome Research Institute grants HG005581, HG005552, HG006513, HG007022, and HG007089; the National Heart Lung and Blood Institute grant HL117626; the Intramural Research Program of the NIH, National Institute on Aging, with contracts N01-AG-1-2109 and HHSN271201100005C; the Sardinian Autonomous Region (L.R. no. 7/2009) grant cRP3-154. P.S. was supported by FCT, ESF and POPH through the FCT Investigator Programme (IF/01641/2013). P.S. and E.C.-S. acknowledge FCT IP and ERDF (COMPETE2020 – POCI) for the CBMA programme UID/BIA/04050/2013 (POCI-01-0145-FEDER-007569). T.R. was supported by an FCT grant (SFRH/BPD/108126/2015). M.P., P.S. and M.B.R. acknowledge FCT support through project PTDC/EPH-ARQ/4164/2014 partially funded by FEDER funds (COMPETE 2020 project 016899). M.B.R. received support from a Leverhulme Doctoral Scholarship programme. C.P. was supported by the Baden Württemberg Foundation and the Max Planck Society. The sequence data for the modern (N = 2,136) and ancient (N = 21) Sardinian mitogenomes are available in GenBank with

accession numbers KY408145–KY410236, KY399164–KY399207, and KY399143–KY399163.

## References

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, et al. 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 75:910–918.

Behar DM, Harmant C, Manry J, van Oven M, Haak W, Martinez-Cruz B, Salaberria J, Oyharçabal B, Bauduer F, Comas D, et al. 2012. The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times. *Am J Hum Genet* 90:486–493.

Berger J-F, Guilaine J. 2009. The 8200 cal BP abrupt environmental change and the Neolithic transition: a Mediterranean perspective. *Quaternary International* 200:31–49.

Broodbank C. 2013. The making of the middle sea: a history of the Mediterranean from the beginning to the emergence of the classical world. London: Thames & Hudson.

Costa MD, Pereira JB, Pala M, Fernandes V, Olivieri A, Achilli A, Perego UA, Rychkov S, Naumova O, Hatina J, et al. 2013. A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. *Nat Commun* 4:2543.

Cavalli-Sforza LL, Menozzi P, Piazza, A. 1994. The history and geography of human genes. Princeton Univ. Press. Princeton, New Jersey, USA.

Chiang CWK, Marcus JH, Sidore C, Al-Asadi H, Zoledziewska M, Pitzalis M, Busonero F, Maschio A, Pistis G, Steri M, et al. 2016. Population history of the Sardinian people inferred from whole-genome sequencing. bioRxiv. doi: https://doi.org/10.1101/092148.

Dyson SL, Rowland RJ. 2007. Archaeology and history in Sardinia from the Stone Age to the Middle Ages: Shepherds, sailors, and conquerors. Philadelphia: University of Pennsylvania Press.

Ermini L, Olivieri C, Rizzi E, Corti G, Bonnal R, Soares P, Luciani S, Marota I, De Bellis G, Richards MB, et al. 2008. Complete mitochondrial genome sequence of the Tyrolean Iceman. *Curr Biol* 18:1687–1693.

Francalacci P, Morelli L, Angius A, Berutti R, Reinier F, Atzeni R, Pilu R, Busonero F, Maschio A, Zara I, et al. 2013. Low-Pass DNA Sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. *Science* 341:565–569.

Francalacci P, Sanna D, Useli A, Berutti R, Barbato M, Whalen MB, Angius A, Sidore C, Alonso S, Tofanelli S, et al. 2015. Detection of phylogenetically informative polymorphisms in the entire euchromatic portion of human Y chromosome from a Sardinian sample. *BMC Res Notes* 8:174.

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.

Gandini F, Achilli A, Pala M, Bodner M, Brandini S, Huber G, Egyed B, Ferretti L, Gómez-Carballa A, Salas A, et al. 2016. Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes. *Sci Rep* 6:25472.

Gamble C, Davies W, Pettitt P, Richards M. 2004. Climate change and evolving human diversity in Europe during the last glacial. *Philos Trans R Soc Lond B Biol Sci* 359:243–253.

Günther T, Valdiosera C, Malmström H, Ureña I, Rodriguez-Varela R, Sverrisdóttir ÓO, Daskalaki EA, Skoglund P, Naidoo T, Svensson EM, et al. 2015. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc Natl Acad Sci U S A* 112:11917–11922.

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.

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 U S A* 113:6886–6891.

Hofmeijer KG, Alderliesten C, van der Borg K, Houston CM, de Jong AFM, Martini F, Sanges M, Sondaar PY, de Visser JA. 1989. Dating of the Upper Pleistocene lithic industry of Sardinia. *Radiocarbon*. 31:986–991.

Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.

Lazaridis I, Nadel D, Rollefson G, Merrett DC, 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* 536:419–24.

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

Keller A, Graefen A, Ball M, Matzas M, Boisguerin V, Maixner F, Leidinger P, Backes C, Khairat R, Forster M, et al. 2012. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat Commun* 3:698.

Macaulay V, Richards MB. 2013. Mitochondrial genome sequences and their phylogeographic interpretation. In *Encyclopedia of Life Sciences* (ELS) (Chichester, John Wiley & Sons, Ltd.

Manning K, Timpson A, Colledge S, Crema E, Edinborough K, Kerig T, Shennan S. 2014. The chronology of culture: a comparative assessment of European Neolithic dating approaches. *Antiquity* 88:1065–1080.

Olivieri A, Pala M, Gandini F, Hooshiar Kashani B, Perego UA, Woodward SR, Grugni V, Battaglia V, Semino O, Achilli A, et al. 2013. Mitogenomes from two uncommon haplogroups mark late glacial/postglacial expansions from the near east and neolithic dispersals within Europe. *Plos One* 8:e70492.

Omrak A, Günther T, Valdiosera C, Svensson EM, Malmström H, Kiesewetter H, Aylward W, Storå J, Jakobsson M, Götherström A. 2016. Genomic evidence establishes Anatolia as the source of the European Neolithic gene pool. *Curr Biol* 26:270–275.

Pala M, Achilli A, Olivieri A, Hooshiar Kashani B, Perego UA, Sanna D, Metspalu E, Tambets K, Tamm E, Accetturo M, et al. 2009. 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.

Pala M, Olivieri A, Achilli A, Accetturo M, Metspalu E, Reidla M, Tamm E, Karmin M, Reisberg T, Hooshiar Kashani B, et al. 2012. Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am J Hum Genet* 90:915–924.

Pereira JB, Costa MD, Vieira D, Pala M, Bamford L, Harrich N, Cherni L, Alshamali F, Hatina J, Rychkov S, Stefanescu G, et al. 2016. Reconciling evidence from ancient and contemporary genomes: a major source for the European Neolithic within Mediterranean Europe. Submitted.

Pinhasi R, Thomas MG, Hofreiter M, Currat M, Burger J. 2012. The genetic history of Europeans. *Trends Genet* 28:496–505.

Posth C, Renaud G, Mittnik A, Drucker DG, Rougier H, Cupillard C, Valentin F, Thevenet C, Furtwängler A, Wißing C, et al. 2016. Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a late glacial population turnover in Europe. *Curr Biol* 26:827–833.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67:1251–1276.

Richards MB, Soares P, Torroni A. 2016. Palaeogenomics: mitogenomes and migrations in Europe's past. *Curr Biol* 26:R243–246.

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barać L, Peričić M, Balanovsky O, et al. 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet* 75:128–137.

Shackleton JC, van Andel TH, Runnels CN. 1984. Coastal paleogeography of the Central and Western Mediterranean during the last 125,000 years and its archaeological implications. *J Field Archaeol* 11:307–314.

Sidore C, Busonero F, Maschio A, Porcu E, Naitza S, Zoledziewska M, Mulas A, Pistis G, Steri M, Danjou F, et al. 2015. Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nat Genet* 47:1272-1281.

Sikora M, Carpenter ML, Moreno-Estrada A, Henn BM, Underhill PA, Sánchez-Quinto F, Zara I, Pitzalis M, Sidore C, Busonero F, et al. 2014. Population genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of the Tyrolean Iceman and the Genetic structure of Europe. *PLoS Genet* 10:e1004353.

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84:740–759.

Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt H-J, Torroni A, Richards MB. 2010. The archaeogenetics of Europe. *Curr Biol* 20:R174–183.

Soares PA, Trejaut JA, Rito T, Cavadas B, Hill C, Eng KK, Mormina M, Brandão A, Fraser RM, Wang TY, et al. 2016. Resolving the ancestry of Austronesian-speaking populations. *Hum Genet* 135:309–326.

Sondaar PY. 1998. Palaeolithic Sardinians: paleontological evidence and methods. In Sardinian and Aegean chronology. Eds. Balmuth MS and Tykot RH, Oxbow Books, Oxford, UK. p. 45–51.

Torroni A, Achilli A, Macaulay V, Richards M, Bandelt H-J. 2006. Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22:339–345.

## Figure Legends

**Fig. 1.** Schematic representation of possible scenarios (A, B, C, and D) for the differentiation of a founder mtDNA haplotype into a Sardinian-Specific Haplogroup (SSH) and implications for the SSH age estimate. A. The founder haplotype (from outside Sardinia) acquired one novel mutation (or more) in Sardinia (*in situ*) giving rise to the SSH. At the present time, the founder haplotype could have been lost, still be present, and/or have differentiated into a new haplotype(s) outside Sardinia. B. The founder haplotype arrived to Sardinia and gave rise to Sardinian-specific haplotypes, but it was lost outside Sardinia. C. The founder haplotype arrived to Sardinia and gave rise to Sardinian-specific haplotypes and is still present and/or differentiated outside Sardinia. D. The founder haplotype diverged outside Sardinia and both the founder and the derived haplotypes arrived in Sardinia where they both differentiated into Sardinian-specific haplotypes, whilst both were lost outside Sardinia. Scenarios A, B and D would give rise to what we defined as "Sardinian-specific haplogroups", but only scenario D would lead to an overestimation of the SSH presence /arrival time on the island.

**Fig. 2.** Specular schematic trees encompassing the three Sardinian-specific haplogroups (N1b1a9, U5b1i1, K1a2d) whose age estimates might predate the Neolithic (>7.8 Kya) and the Sardinian haplogroups H1 and H3. Age estimates were calculated by employing two mutation rates, by Soares et al. (2009) (tree on the left) and by Posth et al. (2016) (tree on the right). Triangles and continuous lines indicate ML estimates. Circles and dashed lines indicate BEAST estimates. Ages are according to the (non-linear) time scale on the bottom. Coloured shadings show the largest confidential intervals of age estimates. For details concerning the age estimates of all SSHs see Supplementary fig. S2.

**Fig. 3.** Schematic representation of K1a2 (panel a), U5b1i (panel b), and N1b1a (panel c) phylogenies. Subclades are represented by triangles, while singletons by lines. The width of triangles is proportional to the number of both modern and ancient mitogenomes, while the height to the age of the clades (Kya) estimated with ML and the molecular clock proposed by Soares et al. (2009)**.** These and the ages obtained with other methods/rates are listed in

Supplementary table S3. Colours indicate the geographical origin of samples according to the legend. Ancient samples (whose codes are those reported in Supplementary table S7 and Supplementary fig. S5) are placed in correspondence of their radiocarbon calibrated ages. The name of Sardinian-specific haplogroups is underlined and in a purple field.

**Fig. 4**. Spatial frequency distribution map of haplogroup H3. Dots indicate the geographic locations of the surveyed populations. Population frequencies (%) are provided in Supplementary table S9. We constructed spatial frequency distribution plots with the program Surfer 9 (Golden Software, http://www.goldensoftware.com/products/surfer). The inset shows the Bayesian skyline plot (BSP) showing effective population size trends of Sardinian H3 mtDNAs. The black and white lines are the median estimates obtained by employing the mutation rates proposed by Soares et al. (2009) and Posth et al. (2016), respectively; the grey shading shows the highest posterior density limits.

**Table 1.** Maximum Likelihood (ML) and Bayesian age estimates for the three Sardinian-specific haplogroups (SSHs) whose age estimates are >7.8 Ky.

| **SSH** | **N[a]** | **ML Age Estimates (Ky)** | | | | **BEAST Age Estimates (Ky)** | | | | **Ancestral Geographic Source** |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Soares et al. 2009 | | Posth et al. 2016 | | Soares et al. 2009 | | Posth et al. 2016 | | |
| | | T | SE | T | SE | T | SE | T | SE | |
| K1a2d | 9 | 16.02 | 1.98 | 13.10 | 1.55 | 12.79 | 2.71 | 10.99 | 2.35 | Near East |
| U5b1i1 | 41 | 12.98 | 6.07 | 10.70 | 4.82 | 11.32 | 2.69 | 9.69 | 2.37 | Western Europe |
| N1b1a9 | 4 | 9.39 | 1.89 | 7.83 | 1.53 | 8.46 | 2.39 | 7.25 | 2.01 | Near East |

[a] Number of mitogenomes included in the corresponding SSH.