

# The Arrival of Steppe and Iranian Related Ancestry in the Islands of the Western Mediterranean

Daniel M. Fernandes<sup>1,2,3\*</sup>, Alissa Mittnik<sup>4</sup>, Iñigo Olalde<sup>4</sup>, Iosif Lazaridis<sup>4</sup>, Olivia Cheronet<sup>1,2</sup>, Nadin Rohland<sup>4</sup>, Swapan Mallick<sup>4,5,6</sup>, Rebecca Bernardos<sup>4</sup>, Nasreen Broomandkhoshbacht<sup>4,5,†</sup>, Jens Carlsson<sup>7</sup>, Brendan J. Culleton<sup>8</sup>, Matthew Ferry<sup>4,5</sup>, Beatriz Gamarra<sup>2,9,10</sup>, Martina Lari<sup>11</sup>, Matthew Mah<sup>4,5,6</sup>, Megan Michel<sup>4,5,†</sup>, Alessandra Modi<sup>11</sup>, Mario Novak<sup>2,12</sup>, Jonas Oppenheimer<sup>4,5,†</sup>, Kendra A. Sirak<sup>2,4,†</sup>, Kirstin Stewardson<sup>4,5</sup>, Stefania Vai<sup>11</sup>, Edgard Camarós<sup>13</sup>, Carla Calo<sup>14</sup>, Giulio Catalano<sup>15</sup>, Marian Cueto<sup>13</sup>, Vincenza Forgia<sup>16</sup>, Marina Lozano<sup>9,10</sup>, Elisabetta Marini<sup>14</sup>, Margherita Micheletti<sup>17</sup>, Roberto M. Micciché<sup>15</sup>, Maria R. Palombo<sup>18</sup>, Damià Ramis<sup>19</sup>, Vittoria Schimmenti<sup>20</sup>, Pau Sureda<sup>21,22</sup>, Luís Teira<sup>13</sup>, Maria Teschler-Nicola<sup>1,23</sup>, Douglas J. Kennett<sup>24</sup>, Carles Lalueza-Fox<sup>25</sup>, Nick Patterson<sup>6,26</sup>, Luca Sineo<sup>15</sup>, David Caramelli<sup>11,\*</sup>, Ron Pinhasi<sup>1,2,\*</sup> and David Reich<sup>4,5,6,27,\*</sup>

<sup>1</sup> Department of Evolutionary Anthropology, University of Vienna, 1090 Vienna, Austria.

<sup>2</sup> Earth Institute and School of Archaeology, University College Dublin, Dublin 4, Republic of Ireland.

<sup>3</sup> CIAS, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal.

<sup>4</sup> Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

<sup>5</sup> Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.

<sup>6</sup> Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA.

<sup>7</sup> Area 52 Research Group, School of Biology and Environmental Science/Earth Institute, University College Dublin, Dublin 4, Republic of Ireland.

<sup>8</sup> Institutes for Energy and the Environment, The Pennsylvania State University, University Park, PA 16802, USA.

<sup>9</sup> Institut Català de Paleoecologia Humana i Evolució Social (IPHES), 43007 Tarragona, Spain.

<sup>10</sup> Department of History and History of Art, Universitat Rovira i Virgili (URV), 43002 Tarragona, Spain.

<sup>11</sup> Dipartimento di Biologia, Università di Firenze, 50122 Florence, Italy.

<sup>12</sup> Institute for Anthropological Research, 10000 Zagreb, Croatia.

<sup>13</sup> Instituto Internacional de Investigaciones Prehistóricas de Cantabria, Universidad de Cantabria-Gobierno de Cantabria-Banco Santander, 39005 Santander, Spain.

<sup>14</sup> Department of Life and Environmental Sciences, Section of Neuroscience and Anthropology, University of Cagliari, 09124 Cagliari, Italy.

<sup>15</sup> Dipartimento di Scienze e Tecnologie Biologiche Chimiche e Farmaceutiche, Università di Palermo, 90123 Palermo, Italy.

<sup>16</sup> Dipartimento di Cultura e Società, Università di Palermo, 90128 Palermo, Italy.

<sup>17</sup> Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università di Torino, 10123 Torino, Italy.

<sup>18</sup> Dipartimento di Scienze della Terra, Sapienza Università di Roma, 00185 Rome, Italy.

<sup>19</sup> Independent researcher, Moragues 34, 07006 Palma de Mallorca, Balearic Islands, Spain.

<sup>20</sup> Museo Archeologico Regionale Antonino Salinas, 90133 Palermo, Italy.

<sup>21</sup> Instituto de Ciencias del Patrimonio (Incipit-CSIC), 15705 Santiago de Compostela, Spain.

<sup>22</sup> McDonald Institute for Archaeological Research and Homerton College, University of Cambridge, Cambridge, United Kingdom.

<sup>23</sup> Department of Anthropology, Natural History Museum Vienna, 1010 Vienna, Austria.

<sup>24</sup> Department of Anthropology, University of California, Santa Barbara, CA 93106, USA.

<sup>25</sup> Institute of Evolutionary Biology, CSIC-Universitat Pompeu Fabra, 08003 Barcelona, Spain.

<sup>26</sup> Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA.

<sup>27</sup> Max Planck-Harvard Research Center for the Archaeoscience of the Ancient Mediterranean, Cambridge, MA, 02138.

‡ Present addresses: Department of Anthropology, University of California, Santa Cruz, CA 95064, USA (N.B.); Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA (Me.M.); Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA (J.O.); Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA (K.A.S.).

Correspondence to: daniel.fernandes@univie.ac.at (D.F.), david.caramelli@unifi.it (D.C.), ron.pinhasi@univie.ac.at (R.P.), reich@genetics.med.harvard.edu (D.R.)

A series of studies have documented how Steppe pastoralist-related ancestry reached central Europe by at least 2500 BCE, while Iranian farmer-related ancestry was present in Aegean Europe by at least 1900 BCE. However, the spread of these ancestries into the western Mediterranean where they have contributed to many populations living today remains poorly understood. We generated genome-wide ancient DNA from the Balearic Islands, Sicily, and Sardinia, increasing the number of individuals with reported data from these islands from 3 to 52. We obtained data from the oldest skeleton excavated from the Balearic islands (dating to ~2400 BCE), and show that this individual had substantial Steppe pastoralist-derived ancestry; however, later Balearic individuals had less Steppe heritage reflecting geographic heterogeneity or immigration from groups with more European first farmer-related ancestry. In Sicily, Steppe pastoralist ancestry arrived by ~2200 BCE and likely came at least in part from Spain as it was associated with Iberian-specific Y chromosomes. In Sicily, Iranian-related ancestry also arrived by the Middle Bronze Age, thus revealing that this ancestry type, which was ubiquitous in the Aegean by this time, also spread further west prior to the classical period of Greek expansion. In Sardinia, we find no evidence of either eastern ancestry type in the Nuragic Bronze Age, but show that Iranian-related ancestry arrived by at least ~300 BCE and Steppe ancestry arrived by ~300 CE, joined at that time or later by North African ancestry. These results falsify the view that the people of Sardinia are isolated descendants of Europe's first farmers. Instead, our results show that the island's admixture history since the Bronze Age is as complex as that in many other parts of Europe.

## Introduction

The advent of the European Bronze Age after 3000 BCE was marked by an increase in long-range human mobility. People with ancestry from the Steppe north of the Black and Caspian Seas made a profound demographic impact in central and eastern Europe, mixing with local farmers to contribute up to three quarters of the ancestry of peoples associated with the Corded Ware complex<sup>1-3</sup>. The expansion of the Beaker complex after around 2400 BCE from the west had a less straightforward correlation to genetic ancestry. In Iberia, most people buried with artifacts of the Beaker complex had little if any Steppe pastoralist-related ancestry (from here on denoted "Steppe ancestry"), but Beaker cultural practices were adopted by people in Central Europe were in part descended from Steppe pastoralists and then spread this material culture along with Steppe ancestry to northwestern Europe<sup>4</sup>. In Iberia, Steppe ancestry began to appear in outlier individuals by ~2500 BCE<sup>4</sup>, and became fully mixed into the Iberian population by 2000 BCE<sup>5</sup>. Meanwhile on Crete in the eastern Mediterranean, there was little if any Steppe ancestry identified in all published samples from the Middle to Late Bronze Age "Minoan" culture (individuals dating to 2400-1700 BCE), although these individuals derived about 15% of their ancestry from groups related to early Iranian farmers (from here on referred to as "Iranian-related ancestry")<sup>6</sup> (Fig. 1).

In the islands of the central and western Mediterranean, the Bronze Age transition has not been investigated with ancient DNA, despite the fact that archaeological evidence reveals that many of the same cultural changes that affected mainland Europe and the eastern Mediterranean also impacted this region<sup>7</sup>. The first evidence for a permanent human presence in the Balearic Islands is dated to just before the onset of the Bronze Age in this part of Europe, between ~2500-2300 BCE<sup>8,9</sup>. Early settlers initially relied on animal husbandry and their economy was focused on sheep, goat<sup>9,10</sup>, and cereal agriculture<sup>11</sup>, while exploitation of wild marine resources (fish, marine birds, mollusks) was central to subsistence on the small island of Formentera<sup>10,12</sup>. Around 1200 BCE, the development of the Talaiotic culture in Mallorca and Menorca (the easternmost Balearic Islands) was marked by intensified management of food resources and the appearance of monumental towers, the eponymous talaiots. These structures were similar in style to the Sardinian nuraghi<sup>10,13</sup>, raising the question of whether there was a cultural connection<sup>14</sup>, a scenario that would gain plausibility if there was substantial genetic exchange between the two regions. Nuragic Sardinians were also in cultural contact with groups from the eastern Mediterranean<sup>15</sup>, so an important question is whether they were admixed with either Steppe or Iranian-related ancestry. Meanwhile, the central Mediterranean island of Sicily was affected by the spread of Beaker cultural complex after around 2400 BCE, and by cultural influence from the Aegean in the Late Helladic Period ~1600-1200 BCE (the period of the “Mycenaean” culture)<sup>16-18</sup>. An unanswered question is whether these events or other cultural changes on the island involved substantial movements of people.

We increased the number of individuals from these islands with genome-wide data from 3 to 52, and analyzed the data to address three questions. First, to what extent did movements of people into these islands track the material culture exchanges documented in the archaeological record? Secondly, can we establish the source and minimum dates of arrival of Steppe ancestry in the central and western Mediterranean islands where this ancestry is present in variable proportions today? Thirdly, did Iranian-related ancestry reach the central and western Mediterranean prior to the period of Phoenician and Greek expansion?

## Results

### Samples and sequencing results

We prepared powder from petrous bones and teeth in dedicated ancient DNA clean rooms at University College Dublin, Harvard Medical School, and the University of Florence, extracted DNA using a method designed to retain short molecules<sup>19-22</sup>, and converted the extracted DNA into double-stranded libraries<sup>23</sup>. We treated all libraries with Uracil-DNA Glycosylase (UDG) to cleave the analyzed molecules at damaged uracil sites, thereby greatly reducing the rate of cytosine-to-thymine errors characteristic of ancient DNA. We enriched ancient DNA libraries for sequences overlapping approximately 1.24 million single nucleotide polymorphisms (SNPs)<sup>24,25</sup>, and obtained genome-wide data from a total of 49 individuals from the Balearic Islands, Sardinia, and Sicily while increasing the quality of data for a Bell Beaker culture associated individual from Sicily (adding three more libraries to the one previously generated) (Fig. 1, Online Table 1, Supplementary

**Materials**). We established chronology based on archaeological context and by assembling direct radiocarbon dates on bone for 28 of the individuals (direct dates for 26 individuals are reported for the first time here; **Online Table 2**). We removed from the analysis dataset eight individuals with fewer than 20,000 of the targeted SNPs covered by at least one sequence, and five with evidence of substantial contamination or less than 3% cytosine-to-thymine error in terminal cytosines. We also removed one individual who we detected as a first degree relative (a son) of another (his mother) that gave higher quality genetic data. This left 36 individuals for our modeling (however, since all the data are useful we fully report all individuals, **Online Table 1**). In the analysis dataset, the median coverage on targeted SNPs on chromosomes 1-22 was 3.20-fold (range 0.02-12.13), and the median number of SNPs covered by at least one sequence was 756709 (range 23600-1038409). All mitochondrial DNA point estimates for match rate to the consensus sequence had 95% confidence intervals with upper bounds from 0.96-1.00, while contamination estimates based on X chromosome variation (meaningful only in males) were all below 1.1% (**Online Table 1**). All individuals had data from at least one library with cytosine-to-thymine damage in the terminal nucleotides greater than 3% (the minimum suggested as a guideline for the plausible authentic DNA<sup>23</sup>). The qualitative patterns of ancestry in the data were unchanged when we restricted to transversion SNPs which are not affected by characteristic ancient DNA errors (**Supplementary Fig. 1**).

## Genetic affinities and population groupings

We carried out principal component analysis (PCA) of the ancient individuals merged with previously published ancient DNA<sup>1,2,4,6,26-44</sup>, projected onto genetic variation among 737 diverse present-day west Eurasians genotyped at ~600,000 SNPs (a subset of the positions on the ~1.24 million SNP set)<sup>27,39,45-47</sup> (**Fig. 2b**, **Online Table 3**). We also performed unsupervised clustering with ADMIXTURE<sup>48</sup> (**Fig. 2a**). The three Balearic Islands individuals in the analysis dataset fall between the European Neolithic and Bronze Age clusters on the PCA, consistent with harboring Steppe ancestry (**Fig. 2b**), a finding that is also supported by the finding in these individuals of an ADMIXTURE component maximized in Eastern European Hunter Gatherers (*EHG*) and Yamnaya Steppe Pastoralists. The eight Nuragic Sardinians cluster in PCA and ADMIXTURE with Middle Neolithic Europeans, with the exception of one individual (I10365: 1643-1263 calBCE) that shows a shift towards the Sicilian cluster. The Iron Age Sardinian (~400-200 BCE) and the four Late Antiquity Sardinians (~200-700 CE) deviate toward the Mycenaean cluster, while one of the Late Antiquity Sardinians also deviates toward Central European Bronze Age individuals. The 20 Sicilians cluster in PCA and ADMIXTURE mostly with the European Neolithic individuals, with the exception of two that have more affinity to the Central European Bronze Age individuals (**Fig. 2**). Relative to the Middle Neolithic Sicilians (*Sicily\_MN*), the main Bronze Age Sicilian cluster (after removing these outliers) deviates in a more subtle way toward eastern groups (either Steppe pastoralists or individuals from the Aegean Bronze Age), a pattern that is also evident in ADMIXTURE.

To formally cluster these individuals, we used *qpWave*<sup>2</sup> to test whether each individual in turn was consistent with being from the same group as others from the same time period and region (that is,

we tested whether they were consistent with forming a clade at a  $p < 0.010$  level) (**Supplementary Materials, Fig. 3**). In some instances where the *qpWave* results were ambiguous, we carried out more refined tests to split individuals into analysis groupings (**Supplementary Materials**).

In the Balearic Islands, *qpWave* revealed significant differences between the Early Bronze Age individual *Mallorca\_EBA* and the Late Bronze Age individual *Menorca\_LBA* ( $p = 0.002$ ) (**Supplementary Table 1**). While *qpWave* tests comparing the Middle Bronze Age individual *Formentera\_MBA* to the other two individuals were non-significant, the symmetry test statistic  $f_4(\text{Mbuti.DG, Iberia\_Chalcolithic; Formentera\_MBA, Menorca\_LBA})$  was  $Z = 2.6$  standard errors from zero (exceeding a threshold of  $|Z| > 2$ , which is approximately  $p < 0.05$ ), implying significantly different ancestry in *Formentera\_MBA* than in *Menorca\_LBA*. In light of this and the different dates and island sources of these three individuals, we treated all three separately for analysis.

For the Nuragic Sardinians (*Sardinia\_Nuragic\_BA*), only I10365 clearly did not form a clade with others of the same cultural affiliation (**Supplementary Tables 2 and 3**). Thus, we treated this individual whose radiocarbon date confirms it as contemporaneous with the others as an outlier (*Sardinia\_Nuragic\_BA10365*). The Iron Age Sardinian individual formed a clade with some from the *Sardinian\_LateAntiquity* cluster and some ancient Sicilians, but we treated it separately because of its distinctive time period and geographic location. The four *Sardinian\_LateAntiquity* individuals were consistent with forming a clade in *qpWave*, but one individual separated from the others in PCA (**Fig. 2**), and also showed distinct signals in admixture modeling, and hence we analyzed it separately as *Sardinian\_LateAntiquity12221* (**Supplementary Table 4**).

For Sicily, our analysis confirmed the two Early Bronze Age outliers *Sicily\_EBA11443* and *Sicily\_EBA8561* evident in PCA and ADMIXTURE (both  $p < 10^{-12}$  relative to the main cluster), while identifying a third outlier *Sicily\_EBA3123* ( $p = 0.004$ ) (**Supplementary Tables 5 and 6**). One Sicilian Middle Bronze Age individual was not consistent with being a clade with one of the other two, and we treated the three separately in subsequent analysis (*Sicily\_MBA3124*, *Sicily\_MBA3125*, and *Sicily\_MBA4109*) (**Supplementary Table 7**). All 5 Late Bronze Age individuals were consistent with being a clade at the  $p > 0.01$  threshold and we grouped them (*Sicily\_LBA*) (**Supplementary Table 8**).

We used *qpAdm*<sup>2,45</sup> to decompose the ancestry of each analysis grouping into four “distal” sources: *Anatolia\_Neolithic*, Western Hunter-Gatherers (*WHG*), *Iran\_Ganj\_Dareh\_Neolithic* and *Yamnaya\_Samara*. We first tested the model with *Anatolia\_Neolithic* and *WHG*, then added either *Iran\_Ganj\_Dareh\_Neolithic* or *Yamnaya\_Samara* as a potential third source, and finally combined all ancestry sources for a total of four sources. We quote the most parsimonious model (as measured by the lowest number of ancestry sources) that fits at  $p > 0.05$ . A unique parsimonious model fit for each analysis grouping (**Fig. 4b and Supplementary Table 9 and 10**).

## Formal modeling of the ancestry of Bronze Age Individual from the Balearic Islands

*Mallorca\_EBA* dates to the earliest period of permanent occupation of the islands at around 2400 BCE<sup>10,49</sup>. We parsimoniously modeled *Mallorca\_EBA* as deriving  $36.9 \pm 4.2\%$  of her ancestry from a source related to *Yamnaya\_Samara*; all fitting models require Steppe ancestry, whereas no Iranian-related ancestry is required to achieve a fit (Fig. 4, Supplementary Table 9). We next used *qpAdm* to identify “proximal” sources for *Mallorca\_EBA*’s ancestry that are more closely related to this individual in space and time, and found that she can be modeled as a clade with the (small) subset of Iberian Bell Beaker culture associated individuals who carried Steppe-derived ancestry<sup>4</sup> ( $p=0.442$ ). This suggests that the movements of people that brought Steppe ancestry into Iberia may have been related to those that first settled the Balearic islands. However, archaeological evidence for the Beaker complex in the Balearic islands during the 3rd millennium BCE is scarce<sup>9</sup>, so it is possible that a related non-Beaker using group spread this ancestry.

Our estimates of Steppe ancestry in the two later Balearic Islands individuals are lower than the earlier one:  $26.3 \pm 5.1\%$  for *Formentera\_MBA* and  $23.1 \pm 3.6\%$  for *Menorca\_LBA* (Supplementary Table 9), but the Middle to Late Bronze Age Balearic individuals are not a clade relative to non-Balearic groups. Specifically, we find that  $f_4(\text{Mbuti.DG}, X; \text{Formentera\_MBA}, \text{Menorca\_LBA})$  is positive when  $X=\text{Iberia\_Chalcolithic}$  ( $Z=2.6$ ) or  $X=\text{Sardinia\_Nuragic\_BA}$  ( $Z=2.7$ ). While it is tempting to interpret the latter statistic as suggesting a genetic link between peoples of the Talaiotic culture of the Balearic islands and the Nuragic culture of Sardinia, the attraction to *Iberia\\_Chalcolithic* is just as strong, and the mitochondrial haplogroup U5b1+16189+@16192 in *Menorca\_LBA* is not observed in *Sardinia\\_Nuragic\\_BA* but is observed in multiple *Iberia\\_Chalcolithic* individuals. A possible explanation is that both the ancestors of Nuragic Sardinians and the ancestors of Talaiotic people from the Balearic Islands received gene flow from an unsampled Iberian Chalcolithic-related group (perhaps a mainland group affiliated to both) that did not contribute to *Formentera\_MBA*.

During the Iron Age, Phoenician colonies were established in the Balearic islands. The Ibiza Phoenician individual published in <sup>50</sup> is not consistent with forming a clade with any of the Bronze Age individuals from the Balaeric islands newly reported in this study, and indeed we find that she can not be modeled even with our least parsimonious model of 4 distal sources. However, when we add in a North African source of ancestry, we can fit her as a two-way mix of  $18.8 \pm 7.9\%$  *Anatolia\_Neolithic* and  $81.2 \pm 7.9\%$  *Morocco\_LN* ancestry ( $p=0.141$ ) (Supplementary Materials). We also can fit the Ibiza Phoenician as two-way mixture of a variety of groups closer to her in time one of which is always *Morocco\_LN*. While several of these models include a Balaeric Island Bronze Age source, we cannot rule out the possibility that the Ibiza Phoenician individual has no local Balaeric ancestry at all. Specifically, we find that we can fit her with models that do not have a Balaeric source and that instead have Balaeric Bronze Age individuals in the outgroups (e.g.  $17.1 \pm 3.5\%$  *France\_Bell\_Beaker* and  $82.9 \pm 3.5\%$  *Morocco\_LN*,  $p=0.869$ ) (Supplementary Table 11).

Modern Balearic individuals also do not fit with the least parsimonious model of 4 distal sources, however, we can fit them as a mixture of Steppe, Iranian-related, and North African ancestry,

demonstrating the Balearic islands have been affected by significant admixture since the initial settlement.

# **Formal Modeling of Ancestry Changes Over Time in Sardinia**

We analyzed 13 individuals from Sardinia dated to ~2200 BCE - 700 CE (**Fig. 1, Online Table 1**).

In *qpAdm*, all eight Bronze Age Nuragic individuals fit as descending from the same two deep ancestral sources (*Anatolia\_Neolithic* and *WHG*), but mixed in different proportions:  $82.5 \pm 1.1\%$  *Anatolia\_Neolithic* for the main *Sardinia\_Nuragic\_BA* cluster ( $p=0.265$ ), and  $85.4 \pm 2.2\%$  for the *Sardinia\_Nuragic\_BA10365* outlier ( $p=0.064$ ) (**Supplementary Table 9**). We find no working models when we consider chronological or geographically more proximal sources (e.g. Beaker complex associated individuals from Iberia, France, Czech Republic, Germany; or Chalcolithic Iberians and Neolithic Sicilians), although we do not have access to early Neolithic Sardinians for this analysis.

Most Sardinians buried in a Nuragic Bronze Age context possessed uniparental haplogroups found in European hunter-gatherers and early farmers, including Y-haplogroup R1b1a[xR1b1a1a] which is different from the characteristic R1b1a1a2a1a2 spread in association with the Bell Beaker complex<sup>4</sup>. An exception is individual I10553 (1226-1056 calBCE) who carried Y-haplogroup J2b2a (**Online Table 1**), previously observed in a Croatian Middle Bronze Age individual bearing Steppe ancestry<sup>44</sup>, suggesting the possibility of genetic input from groups that arrived from the east after the spread of first farmers. This is consistent with the evidence of material culture exchange between Sardinians and mainland Mediterranean groups<sup>15</sup>, although genome-wide analyses find no significant evidence of Steppe ancestry so the quantitative demographic impact was minimal. *qpAdm* modeling of the ancestry of the *Sardinia\_Nuragic\_BA10365* outlier with respect to sources potentially more closely related in space and time does infer some ancestry in this individual from an eastern source (either carrying Steppe ancestry or Iranian-related ancestry) that we do not detect by modeling with sources more distant in space and time, consistent with the hypothesis of eastern influence (**Supplementary Table 12**).

We detect definitive evidence of Iranian-related ancestry in an Iron Age Sardinian I10366 (391-209 calBCE) with an estimate of  $11.9 \pm 3.7\%$  *Iran\_Ganj\_Dareh\_Neolithic* related ancestry, while rejecting the model with only *Anatolian\_Neolithic* and *WHG* at  $p=0.0066$  (**Supplementary Table 9**). The only model that we can fit for this individual using a pair of populations that are closer in time is as a mixture of *Iberia\_Chalcolithic* ( $11.9 \pm 3.2\%$ ) and *Mycenaean* ( $88.1 \pm 3.2\%$ ) ( $p=0.067$ ). This model fits even when including Nuragic Sardinians in the outgroups of the *qpAdm* analysis, which is consistent with the hypothesis that this individual had little if any ancestry from earlier Sardinians.

In the *Sardinian\_LateAntiquity* group (the earliest dating to 256-403 calCE), we detect even higher proportions of *Iran\_Ganj\_Dareh\_Neolithic*-related ancestry: an estimated  $29.6 \pm 4.6\%$  ( $p=0.000001$  for rejection of the alternative model that attempts to model its eastern ancestry as entirely Yamnaya-related, **Supplementary Table 9**). One possibility is the Iranian-related ancestry began to

be introduced in the Phoenician period, a scenario that is not only consistent with the historical evidence and our finding of this ancestry type in the Iron Age Sardinian, but is also supported by previously published mitochondrial DNA which has documented haplotypes in ancient Phoenician colonies in modern Sardinians<sup>51</sup>. In modeling using source populations that are temporally more plausible, this individual is consistent with being a clade with both *Myceanean* ( $p=0.241$ ) or *Ibiza\_Phoenician* ( $p=0.145$ ); importantly, both these models work with Nuragic Bronze Age Sardinians included in the outgroups, and so *Sardinian\_LateAntiquity* is consistent with having negligible ancestry from earlier Bronze Age groups to the limits of our resolution (**Supplementary Materials**). We also model the outlier *Sardinia\_LateAntiquity12221* as having  $33.3 \pm 5.5\%$  Yamnaya-related while confidently rejecting models with no Steppe ancestry (all  $p \leq 0.001$ ) (**Supplementary Table 9**), providing the earliest clear evidence of Steppe ancestry in Sardinia. However, we do not have sufficient resolution given the limited data from this single sample to determine the geographic source of the Steppe ancestry (**Supplementary Table 13**).

In a dataset of 27 modern Sardinians for whom we have genotyping data at about 600,000 SNPs<sup>45</sup>, we obtain a fit for a model of  $61.4 \pm 1.6\%$  *Anatolia\_Neolithic*,  $9.5 \pm 1.0\%$  *WHG*,  $19.1 \pm 1.9\%$  *Iran\_Ganj\_Dareh\_Neolithic* and  $10.0 \pm 1.6\%$  *Yamnaya\_Samara* related ancestry and definitively reject models without all four ancestries (all models  $p < 10^{-6}$  in **Supplementary Table 9**). We replicate the finding of *Iran\_Ganj\_Dareh\_Neolithic*-related ancestry (and not just Steppe ancestry) in a subset of four of the modern Sardinian individuals with whole genome shotgun sequencing data (**Supplementary Table 9**). Even the four-way model is not comprehensive for modern Sardinians, however, as when we add Late Neolithic North Africans from Morocco to the outgroup set<sup>52</sup>, we reject the four-way mixture model ( $p < 10^{-12}$ ) (adding the Neolithic Moroccans to the outgroup set does not cause model rejection for any of the ancient samples in our dataset, showing that it may reflect events taking place after the times our individuals lived; **Supplementary Table 9**). Modeling modern Sardinians with this fifth source produces a fit with an estimate of  $16.1 \pm 8.4\%$  *Morocco\_LN*-related ancestry ( $p=0.235$ ). Our signal of North African-related mixture in Sardinians may reflect the same process that introduced sub-Saharan African ancestry into Sardinians<sup>53-55</sup> which was argued in<sup>56</sup> to reflect North African-related admixture with an average date of ~630 CE.

An important question is how much ancestry modern Sardinians have inherited from people related to those of the Nuragic Bronze Age. We could parsimoniously model our modern Sardinian sample as a 2-way mixture of  $13.6 \pm 3.4\%$  *Sardinia\_Nuragic\_BA* and  $86.4 \pm 3.4\%$  *Sardinia\_LateAntiquity12221*. It is striking that most of the ancestry in modern Sardinians is inferred in this analysis to come from a *Sardinia\_LateAntiquity12221*-related group, which can itself be modeled as closely related to Mycenaeans or Phoenicians with no evidence of specific shared ancestry with Bronze Age Sardinians. The group of modern Sardinians we are modeling has often been interpreted as an isolated lineage that derives from early Sardinian farmers with little subsequent immigration into the islands. Our finding that a large fraction of this group's ancestry is consistent with deriving from a group that was present in Sardinia in Late Antiquity and that had no evidence of a contribution from earlier Sardinian groups is therefore surprising (although we caution that this inference is tentative as the

amount of data we have for *Sardinia\_LateAntiquity12221* is limited; **Online Table 1**). Modern Sardinian populations are geographically highly substructured for example among different valleys and coastal and inland sites.<sup>55</sup> Analyses of more geographically diverse modern and ancient Sardinians will provide additional insight into the population turnovers.

# **Formal Modeling of the Neolithic to Bronze Age transition in Sicily**

In the Middle Neolithic, Sicilians harbored ancestry typical of early European farmers, well modeled as a mixture of *Anatolia\_Neolithic* and *WHG* (**Fig. 2, Fig. 4, Supplementary Table 9**).

Steppe ancestry arrived in Sicily by the Early Bronze Age. While a previously reported Bell Beaker culture-associated individual from Sicily had no evidence of Steppe ancestry<sup>4</sup>, a result we confirm by more than tripling the number of sequences for this individual who previously had marginal quality data, we find evidence of Steppe ancestry in the Early Bronze Age by ~2200 BCE. In distal *qpAdm*, the outlier *Sicily\_EBA11443* is parsimoniously modeled as harboring  $40.2 \pm 3.5\%$  Steppe ancestry, and the outlier *Sicily\_EBA8561* is parsimoniously modeled as harboring  $23.3 \pm 3.5\%$  Steppe ancestry (**Fig. 4a, Supplementary Table 9**). The main *Sicily\_EBA* cluster also can only be fit with Steppe ancestry albeit at a lower proportion of  $9.1 \pm 2.3\%$ , and models without Steppe ancestry can be rejected ( $p=0.001$ ) (**Supplementary Table 9**). The presence of Steppe ancestry in Early Bronze Age Sicily is also evident in Y chromosome analysis, which reveals that 4 of the 5 Early Bronze Age males had Steppe-associated Y-haplogroup R1b1a1a2a1a2. (**Online Table 1**). Two of these were Y-haplogroup R1b1a1a2a1a2a1 (Z195) which today is largely restricted to Iberia and has been hypothesized to have originated there 2500-2000 BCE<sup>57</sup>. This evidence of west-to-east gene flow from Iberia is also suggested by *qpAdm* modeling where the only parsimonious proximate source for the Steppe ancestry we found in the main *Sicily\_EBA* cluster is Iberians (**Supplementary Table 14**).

We detect Iranian-related ancestry in Sicily by the Middle Bronze Age 1800-1500 BCE, consistent with the directional shift of these individuals toward Mycenaeans in PCA (**Fig. 2b**). Specifically, two of the Middle Bronze Age individuals can only be fit with models that in addition to *Anatolia\_Neolithic* and *WHG*, include *Iran\_Ganj\_Dareh\_Neolithic*. The most parsimonious model for *Sicily\_MBA3125* has  $18.0 \pm 3.6\%$  Iranian-related ancestry ( $p=0.032$  for rejecting the alternative model of Steppe rather than Iranian-related ancestry), and the most parsimonious model for *Sicily\_MBA4109* has  $14.9 \pm 3.9\%$  Iranian-related ancestry ( $p=0.037$  for rejecting the alternative model) (**Fig. 4a, Supplementary Table 9**). This inference is also supported by *qpAdm* using sources closer in geography and time that always identify a parsimonious model with *Minoan\_Lassithi* as a source for these two individuals (**Supplementary Table 15**). We also found evidence of Iranian-related ancestry in Sicily in an individual of the Early Bronze Age cluster, I11442, who could only be fit in a 3-way model with Iranian-related ancestry ( $19.3 \pm 3.8\%$  ancestry of this type,  $p=0.391$ ; the 3-way model involving Steppe ancestry fails to a fit ( $p=0.010$ )) (**Supplementary Table 10**). However, this finding should be viewed with caution as *qpWave* clustered this individual with four other Sicilian Early Bronze Age individuals, so this finding could be an artifact of performing tests on our data beyond what is justified by our groupings. The modern southern Italian Caucasus-related signal

identified in <sup>58</sup> is plausibly related to the same Iranian-related spread of ancestry into Sicily that we observe in the Middle Bronze Age (and possibly the Early Bronze Age).

For the Late Bronze Age group of individuals, *qpAdm* documented Steppe-related ancestry, modeling this group as  $80.2 \pm 1.8\%$  *Anatolia\_Neolithic*,  $5.3 \pm 1.6\%$  *WHG*, and  $14.5 \pm 2.2\%$  *Yamnaya\_Samara* (Fig. 4b, Supplementary Table 9). Our modeling using sources more closely related in space and time also supports *Sicily\_LBA* having Minoan-related ancestry or being derived from local preceding populations or individuals with ancestries similar to those of *Sicily\_EBA3123* ( $p=0.527$ ), *Sicily\_MBA3124* ( $p=0.352$ ), and *Sicily\_MBA3125* ( $p=0.095$ ) (Supplementary Table 15).

Finally, when we model modern Sicilians, we find that they require not only Steppe and Iranian-related ancestries but also North African ancestry, confirming the ample historical and archaeological evidence of major cultural impacts on the island from North Africa after the Bronze Age (Supplementary Materials).

## Discussion

The islands of the western Mediterranean have been among the most poorly studied regions of Europe from the perspective of genome-wide ancient DNA. Here we increase by about 17-fold the number of individuals with data from the Neolithic onward in these islands to document the arrival of both Steppe and Iranian-related ancestry.

In the Balearic islands, we show that Steppe ancestry arrived almost simultaneously with the first permanent human occupation of the islands in the Early Bronze Age, while the North African ancestry that arrived at least by the time of the Phoenicians<sup>50</sup> still is present today. In Sicily, Steppe ancestry arrived by ~2200 BCE, and likely came at least in part from the west as it was associated with the Iberian-specific Y haplogroup R1b1a1a2a1a2a1 (Z195),<sup>57</sup> thus documenting how Iberia was not just a destination of east-to-west human movement in Europe, but also an important source for west-to-east Steppe ancestry reflux<sup>59</sup>. In Sardinia, we find no convincing evidence of Steppe ancestry in the Bronze Age, but we detect it by ~200-700 CE.

We find no evidence of Iranian-related ancestry in the Balearic Islands individuals until the Phoenician period, around the same time as we detect it in Sardinia. In Sicily, Iranian-related ancestry was present during the Middle Bronze Age, showing that this ancestry which was widespread in the Aegean around this time (in association with the Minoan and Mycenaean cultures), also reached further west. Based on our analysis of modern individuals, it is possible that this ancestry first spread west in substantial amounts during the Late Helladic period of the Mycenaean expansion when strong cultural interactions between Sicily and the Aegean are documented<sup>18,60-62</sup>. However, if our signal of such ancestry in an Early Bronze Age Sicilian individual is correct then some of this spread began even earlier.

Our co-analysis of modern and ancient Sardinians questions the commonly held view that Sardinians are well described as an isolated remnant of Europe's first farmers<sup>63</sup>. While Nuragic Bronze Age Sardinians are indeed well-modeled as having a typical early European farmer ancestry profile, modern Sardinians harbor substantial fractions of ancestry from several groups that arrived in Europe after the Neolithic, and we model modern Sardinians as harboring  $10.0 \pm 1.6\%$  Steppe ancestry and an even larger  $19.1 \pm 1.9\%$  Iranian-related ancestry. Both ancestry types are definitively required to model modern Sardinians, and we show that modern Sardinians have been substantially impacted by movement of ancestry from North Africa in the last two millennia. Thus, rather than being an island sheltered from admixture and migration since the Neolithic, Sardinia, like almost all other regions in Europe has, been a site for major movement and mixtures of people.

## Materials and Methods

### Laboratory work details

We ground skeletal samples to powder in dedicated ancient DNA facilities at the University College Dublin in Ireland, at the University of Florence in Italy, at the University of Palermo in Italy, and at Harvard Medical School in Boston USA (**Online Table 1**)<sup>22,64,65</sup>. We treated all DNA extracts with Uracil-DNA Glycosylase (UDG) to remove characteristic ancient DNA damage to cleave the molecules at 5' Uracils, thus reducing the rate of damage-induced errors<sup>23</sup>. For two of the samples, we performed DNA extraction<sup>19,20</sup> and double-indexed library preparation in Florence<sup>23</sup>. For all other samples, we performed DNA extraction at Harvard Medical School, sometimes using silica coated magnetic beads to support robotic cleanups (instead of silica column cleanups that were used for manual DNA extraction)<sup>19,21</sup>. We converted these DNA extracts to individually barcoded libraries, in some cases assisted by a robotic liquid handler<sup>23</sup> (see **Online Table 1** for details). We initially screened libraries by enriching the libraries for the human mitochondrial genome<sup>66</sup> and about 3000 nuclear SNPs using synthesized baits (CustomArray Inc.), and sequencing on an Illumina NextSeq500 instrument, using different index pairs to distinguish between them. We merged read pairs that overlapped by at least 15 base pairs allowing up to one mismatch (and representing each overlapping base by the higher quality base), and computationally trimmed adapters and barcodes. We mapped the merged sequences to the reconstructed human mitochondrial DNA consensus sequence<sup>67</sup> using bwa (v.0.6.1)<sup>68</sup>, and removed duplicate sequences that had the same orientation, same start and stop positions, and the same barcodes. We assessed the data for authenticity by computing the damage rate at the terminal cytosines (which we required to be at least 3% for at least one library for each individual following published recommendations for libraries of this type<sup>23</sup>), and by estimating the rate of mismatches to the consensus mitochondrial sequence using contamMix<sup>24</sup>. We next enriched the samples with promising quality for 1233013 SNPs ('1240K SNP capture')<sup>2,25</sup>, and sequenced and processed them as for the mitochondrial DNA with the difference that we mapped to the human reference genome *hg19*. We assessed authenticity as for the mitochondrial DNA data, while also estimating contamination based on the ratio of Y to X chromosome sequences (filtering out individuals that had a ratio unexpected for a male or a female) as well as the rate of heterozygosity at X-chromosome positions (only valid as an estimate of contamination in males who should have no X chromosome variation<sup>69</sup>). For some libraries we co-enriched samples for the mitochondrial genome together with the 1240k targets ("1240k+" enrichment).

### Radiocarbon dating and quality assurance

We performed 25 accelerator mass spectrometry (AMS) radiocarbon dates (<sup>14</sup>C) on samples from 24 skeletons at the Pennsylvania State University (PSU) Radiocarbon Laboratory, as well as an additional 4 direct dates on an additional 3 samples. Here we give a detailed description of the samples processing at PSU, as it is the source of most of our dates (for the other samples, we refer readers to the published protocols). As precaution at PSU, we removed possible contaminants

(convervants/adhesives) by sonicating all bone samples in successive washes of ACS grade methanol, acetone, and dichloromethane for 30 minutes each at room temperature, followed by three washes in Nanopure water to rinse. We extracted bone collagen and purified using a modified Longin method with ultrafiltration (>30kDa gelatin<sup>70</sup>). If collagen yields were low and amino acids poorly preserved we used a modified XAD process (XAD Amino Acids<sup>71</sup>). For quality assurance, we measured carbon and nitrogen concentrations and C/N ratios of all extracted and purified collagen/amino acid samples with a Costech elemental analyzer (ECS 4010). We evaluated sample quality by % crude gelatin yield, %C, %N and C/N ratios before AMS 14C dating. C/N ratios for all directly radiocarbon samples fell between 2.9 and 3.6, indicating excellent preservation<sup>72</sup>. We combusted collagen/amino acid samples (~2.1 mg) for 3 h at 900°C in vacuum-sealed quartz tubes with CuO and Ag wires. Sample CO<sub>2</sub> was reduced to graphite at 550°C using H<sub>2</sub> and a Fe catalyst, and drew off reaction water with Mg(ClO<sub>4</sub>)<sub>2</sub><sup>73</sup>. We pressed graphite samples into targets in Al boats and loaded them onto a target wheel with OX-1 (oxalic acid) standards, known-age bone secondaries, and a 14C-free Pleistocene whale blank. We made all 14C measurements on a modified National Electronics Corporation compact spectrometer with a 0.5 MV accelerator (NEC 1.5SDH-1). We corrected the 14C ages for mass-dependent fractionation with measured  $\delta^{13}\text{C}$  values<sup>74</sup> and compared with samples of Pleistocene whale bone (backgrounds, 48,000 14C BP), late Holocene bison bone (~1,850 14C BP), late 1800s CE cow bone, and OX-2 oxalic acid standards. We calibrated 14C ages with OxCal version 4.3<sup>75</sup> and the IntCal13 northern hemisphere curve<sup>76</sup>. The stable carbon and nitrogen isotope measurements we obtained do not indicate a large marine dietary component in these individuals despite their coming from island populations and hence we did not perform a correction of the dates for marine reservoir effect.

#### Uniparental haplogroup determination

We determined mitochondrial haplogroups using HaploGrep<sup>77</sup> and phylotree<sup>78</sup> (build 17) on the data from the mitochondrial enrichment experiment<sup>79</sup>. We restricted sequences and base qualities to values of  $\geq 30$ , and built a consensus sequence with *samtools* and *bcftools*<sup>80</sup>, using a majority rule and minimum coverage of 1, trimming 2 basepairs from the end of each sequence. We further restricted the data for each sample to the damaged reads as determined by *pmdtools* (using a minimum *pmdscore* of 3) and repeated the calling. In almost every case where there was sufficient post-damage restricted coverage to give a confident haplogroup call, the calls matched the non-restricted read sample. We restricted sequences for Y-chromosome haplogroup assessment to qualities  $\geq 30$ , and identified the most derived mutations using the nomenclature of the International Society of Genetic Genealogy (<http://www.isogg.org>) version 11.110.

#### Dataset assembly

We assembled a base dataset and then subsetted for each analysis. This complete dataset included 3310 individuals, of which 2191 were modern<sup>27,39,45-47</sup> and 1119 were ancient individuals from previous publications<sup>1,2,4,6,26-44,52</sup>, which we combined with the newly reported 49 samples (Online Table 3). We performed all subsequent analysis on autosomal data.

## Principal component analysis

We used a subset of 736 modern and 1123 ancient West Eurasians for principal component analysis (PCA) using *smartpca* from the EIGENSOFT package<sup>81</sup>. We modified the standard parameter file with the options shrinkmode: YES, and lsqproject: YES to project all ancient individuals onto the eigenvectors computed from modern vectors. We used a dataset containing only transversions to assess the robustness of our qualitative inferences to bias due to ancient DNA damage-induced errors (Supplementary Fig. 1).

## Population structure analysis

We ran ADMIXTURE<sup>48</sup> after pruning to remove one SNP each in pairs of SNPs in linkage disequilibrium, using PLINK1.9<sup>82</sup> and the option `--indep-pairwise 200 25 0.4`, leaving 321518 SNPs. We ran ADMIXTURE from K=5 to K=15, with 5 random-seeded replicates for each value of K. We used cross validation by adding the option `--cv` to find the runs with the lowest errors. For each value of K, we kept the replicate with lowest error. We present results for K=10, as we empirically found that this is the value of K with lowest cross-validation error that also showed clear distinctions between ancient Western, Eastern, and Caucasus Hunter-Gatherer backgrounds, while having a maximized Early Neolithic Anatolian component. We also performed ADMIXTURE restricting to transversion SNPs and obtained qualitatively similar results suggesting that ancient DNA damage is unlikely to be strongly biasing our findings (Supplementary Fig. 1).

## $f_4$ -statistics

We used ADMIXTOOLS<sup>45</sup> to compute  $f_4$ -statistics (*qpDstat*). We used *Mbuti.DG* as our outgroup, and computed statistics of the form  $f_4(Mbuti.DG, X; Y, Z)$ , where X is our test population/individual and Y/Z are pairs to test against. We used the options `f4mode: YES` and `prints: YES`. We used  $f_4$ -statistics to assess overall population affinities and changes in ancestry through time either by direct comparison of the test populations with the desired pairs or by using symmetry tests, where the populations Y and Z are the populations being tested for consistent with descent from a common ancestral population.

## *qpWave/qpAdm*

We used *qpWave/qpAdm* from ADMIXTOOLS<sup>45</sup> to estimate admixture coefficients and to model our individuals/populations as result of groups related to different proxies for the true source population. We used a base outgroup set including the following individuals/populations: *Mbuti.DG*, *Ust\_Ishim*, *CHG*, *EHG*, *ElMiron*, *Vestonice16*, *MA1*, *Israel\_Natufian*, *Jordan\_PPNB*. Extra populations were included in each test to improve accuracy when using populations with similar ancestries (see Supplementary Materials for a detailed description). When analyzing the results we present the most parsimonious model with the highest probability. We used the option `allsnps: YES`.

## Data Availability

All raw data are available at the European Nucleotide Archive and the National Center for Biotechnology Information under the accession number [to be included upon paper acceptance] and at <https://reich.hms.harvard.edu/datasets>.

## Acknowledgements

This manuscript is dedicated to the memory of Sebastiano Tusa of the Soprintendenza del Mare in Palermo, who would have been an author of this study had he not tragically died in the crash of Ethiopia Airlines flight 302 on March 10. We thank Zhao Zhang for database support. We thank the Soprintendenza BBCCAA Palermo and Rosario Schicchi (director of Museum of Castelbuono) for facilitating access to important skeletal materials. D.F. was supported by an Irish Research Council grant GOIPG/2013/36. Radiocarbon work was supported in part by the NSF Archaeometry program BCS-1460369 to D.J.K. and B.J.C. C.L.-F. was supported by Obra Social La Caixa and by FEDER-MINECO (BFU2015- 64699-P). D.C. was supported by the grant 20177PJ9XF MIUR PRIN 2017. D.Re. is an Investigator of the Howard Hughes Medical Institute and his ancient DNA laboratory work was supported by National Science Foundation HOMINID grant BCS-1032255, by National Institutes of Health grant GM100233, by an Allen Discovery Center grant, and by grant 61220 from the John Templeton Foundation.

## Author Contributions

D.M.F., D.Re., and R.P. conceived the study. D.M.F., E.C., C.C., G.C., M.C., V.F., M.Lo., E.M., Ma.M., R.M.M., D.Ra., M.R.P., V.S., P.S., L.T, M.T-N., C.L-F, L.S., D.C., R.P. excavated, assembled and/or studied the osteological material. D.M.F., O.C., N.R., N.B., M.F., B.G., M.La., Me.M., A.Mo., M.N., J.O., K.A.S., K.S., and S.V. performed laboratory work, while N.R., D.C., and R.P. supervised this work. J.C. provided computing resources. B.J.C. performed radiocarbon analysis and D.J.K. supervised this work. D.M.F., I.O., R.B., S.M., and M.Ma. performed bioinformatic and population genetic analysis with input from A.Mi., I.L., N.P., and D.R.

## Competing Interests

The authors declare no competing financial interests.

## Figures

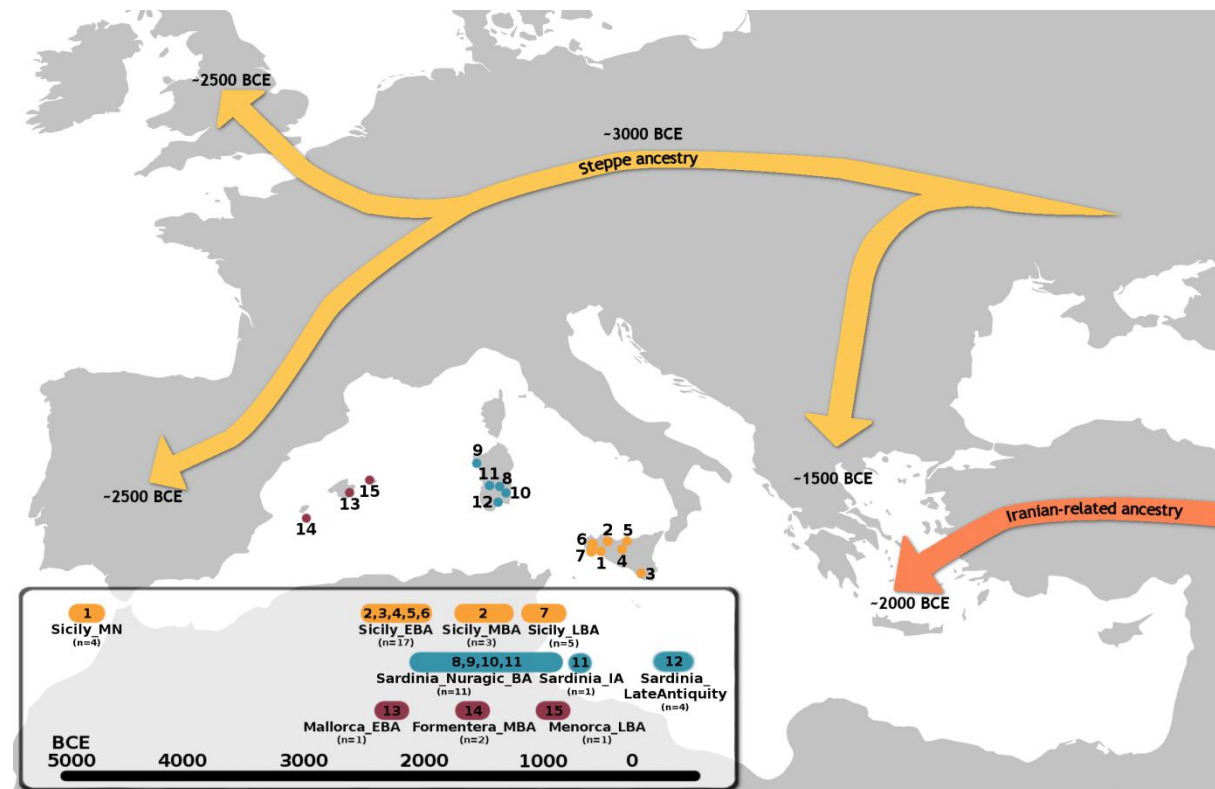
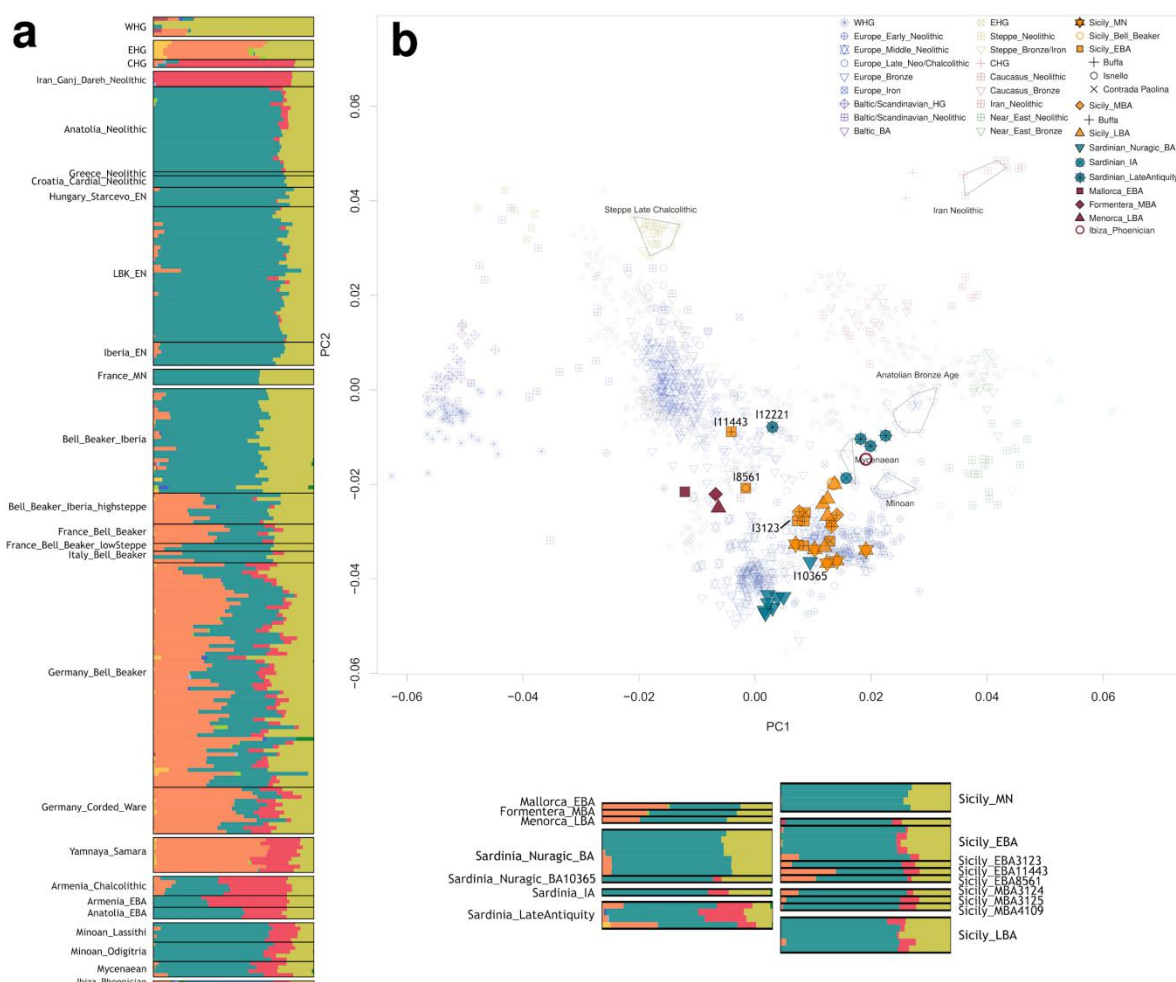
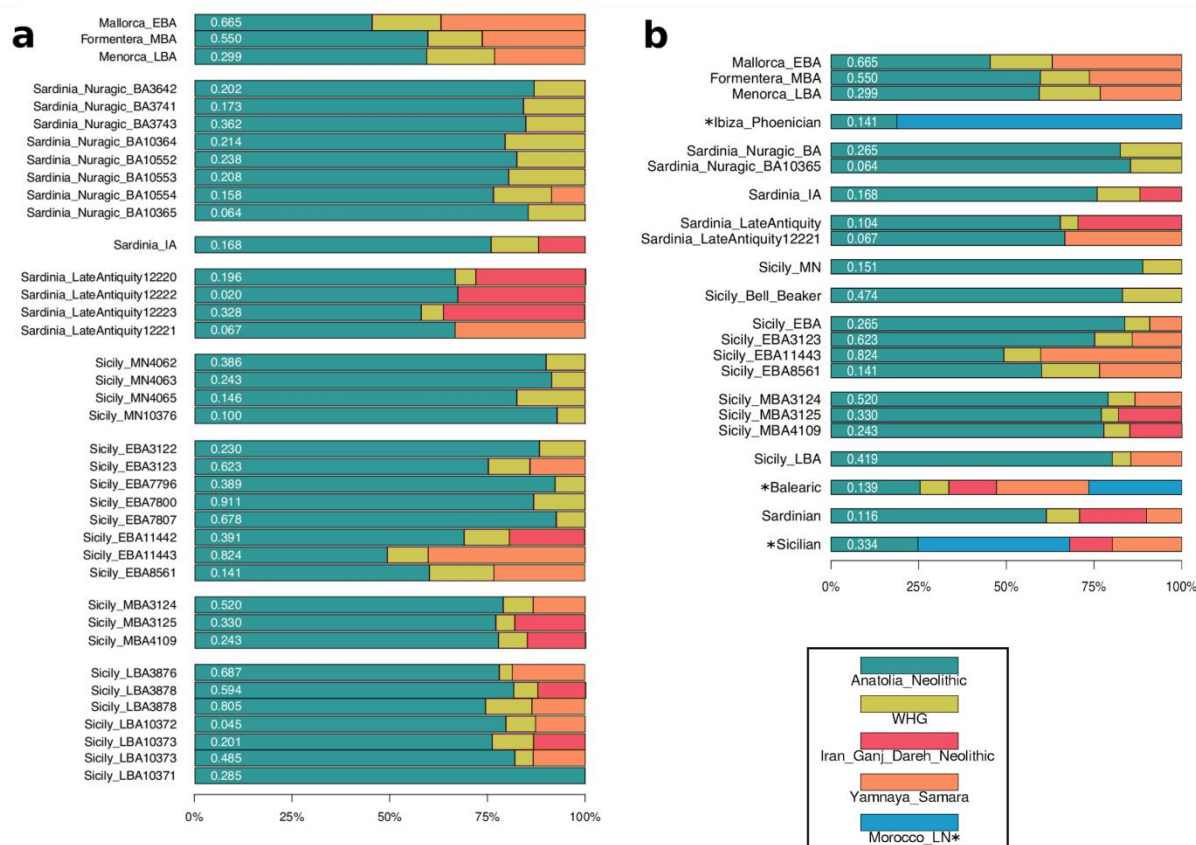


Figure 1: Timeline and geographical origins of the 49 newly reported ancient individuals along with the previously reported individual for whom we increase data quality. 1-Fossato di Stretto Partana; 2-Buffa cave; 3-Contrada Paolina; 4-Isnello; 5-Vallone Inferno; 6-Marcita; 7-Salaparuta; 8-Seulo; 9-Alghero-Lu Maccioni cave; 10-Perdasdefogu; 11-Usellus; 12-Grotta Colombi; 13-Cova des Moro; 14-Cap de Barbaria; 15-Naveta des Tudons.



**Figure 2: Ancestry of ancient Sardinians, Sicilians and Balearic islanders and other ancient and present-day populations according to a) unsupervised ADMIXTURE analysis with K=10 clusters; and b) PCA with previously published ancient individuals (non-filled symbols), projected onto variation from present-day populations (gray squares).**





**Figure 4: Proportions of ancestry using a distal *qpAdm* framework on an individual basis (a), and based on *qpWave* clusters (Fig. 3) (b). We show all valid models ( $p > 0.05$ ) for the lowest possible ranks. Some individuals produced two valid models at  $p > 0.05$ , but for b), there is only a single parsimonious model for each analysis grouping. In panel b) relevant published individuals (*Ibiza\_Phoenician*, *Sicily\_Bell\_Beaker*) and modern populations (*Balearic*, *Sardinian*, *Sicilian*) are also presented. Some of these did not produce valid models with the four base ancestries so we show the most parsimonious working models after including *Morocco\_LN* (denoted by an asterisk). P-values in white are within bars (Supplementary Materials and Supplementary Tables 9 and 10 give all numbers underlying this figure).**

# References

1. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167-172 (2015).
2. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211 (2015).
3. Kristiansen, K. *et al.* Re-theorising mobility and the formation of culture and language among the Corded Ware Culture in Europe. *Antiquity* **91**, 334-347 (2017).
4. Olalde, I. *et al.* The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190-196 (2018).
5. Martiniano, R. *et al.* The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genet.* **13**, e1006852 (2017).
6. Lazaridis, I. *et al.* Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214-218 (2017).
7. Knapp, A. & Van Dommelen, P. The Cambridge Prehistory of the Bronze and Iron Age Mediterranean. in *The Cambridge Prehistory of the Bronze and Iron Age Mediterranean* (eds. Knapp, A. & Van Dommelen, P.) I-II (Cambridge University Press, 2015).
8. Alcover, J. A. The First Mallorcans: Prehistoric Colonization in the Western Mediterranean. *Journal of World Prehistory* **21**, 19-84 (2008).
9. Ramis, D. Early Island Exploitations: Productive and Subsistence Strategies on the Prehistoric Balearic Islands. in *The Cambridge Prehistory of the Bronze and Iron Age Mediterranean* (eds. Knapp, A. B. & van Dommelen, P.) 40-56 (Cambridge University Press, 2014).
10. Ramis, D. Animal Exploitation in the Early Prehistory of the Balearic Islands. *The Journal of Island and Coastal Archaeology* **13**, 269-282 (2018).
11. López-Garí, J. M., Pérez-Jordà, G., Marlasca-Martín, R., Farrera-Fernández, V. & Enrich-Hoja, J. La primera agricultura Pitiusa y Balear: las evidencias de la Cova des Riuets. *SAGVNTVM. Papeles del Laboratorio de Arqueología de Valencia* **45**, 65-77 (2014).
12. Sureda, P. *et al.* Surviving on the isle of Formentera (Balearic Islands): Adaptation of economic behaviour by Bronze Age first settlers to an extreme insular environment. *Journal of Archaeological Science* **12**, 860-875 (2017).
13. Plantalamor, L. & van Strydonck, M. *La cronologia de la prehistòria de Menorca: noves datacions de 14C.* (Govern Balear, Conselleria d'Educació, Cultura i Esports, 1997).

14. Lull, V., Mico, R., Rihuete, C. I. & Risch, R. Los cambios sociales en las islas Baleares a lo largo del II milenio. *Cypsela* **15**, 123-148 (2004).
15. Holt, E. Mobility and meaning in the Nuragic culture of Bronze Age Sardinia (ca. 1700-900BC). in *Forging Identities. The Mobility of Culture in Bronze Age Europe* (eds. Suchowska-Ducke, P., Reiter, S. S. & Vandkilde, H.) **1**, 193-202 (British Archaeological Reports, 2015).
16. Sestieri, A. M. B. The Bronze Age in Sicily. in *The Oxford Handbook of European Bronze Age* (eds. Harding, A. & Fokkens, H.) 653-667 (Oxford University Press, 2013).
17. Sarno, S. *et al.* Ancient and recent admixture layers in Sicily and Southern Italy trace multiple migration routes along the Mediterranean. *Sci. Rep.* **7**, 1984 (2017).
18. Holloway, R. R. *The Archaeology of Ancient Sicily*. (Routledge, 2000).
19. Dabney, J. *et al.* 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).
20. Damgaard, P. B. *et al.* Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.* **5**, 11184 (2015).
21. Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87-93 (2015).
22. Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nature Protocols* **13**, 2447-2461 (2018).
23. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130624 (2015).
24. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553-559 (2013).
25. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216-219 (2015).
26. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* **3**, 698 (2012).
27. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day

- Europeans. *Nature* **513**, 409-413 (2014).
28. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
29. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225-228 (2014).
30. Skoglund, P. *et al.* Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344**, 747-750 (2014).
31. Günther, T. *et al.* Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11917-11922 (2015).
32. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
33. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503 (2015).
34. Olalde, I. *et al.* A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. *Mol. Biol. Evol.* **32**, 3132-3142 (2015).
35. Broushaki, F. *et al.* Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499-503 (2016).
36. Cassidy, L. M. *et al.* Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 368-373 (2016).
37. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205 (2016).
38. Hofmanová, Z. *et al.* Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6886-6891 (2016).
39. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419-424 (2016).
40. Martiniano, R. *et al.* Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nat. Commun.* **7**, 10326 (2016).
41. Schiffels, S. *et al.* Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nat. Commun.* **7**, 10408 (2016).
42. Jones, E. R. *et al.* The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. *Current Biology* **27**, 576-582 (2017).

43. Unterländer, M. *et al.* Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. *Nat. Commun.* **8**, 14615 (2017).
44. Mathieson, I. *et al.* The genomic history of southeastern Europe. *Nature* **555**, 197-203 (2018).
45. Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192**, 1065-1093 (2012).
46. Pickrell, J. K. *et al.* The genetic prehistory of southern Africa. *Nat. Commun.* **3**, 1143 (2012).
47. Qin, P. & Stoneking, M. Denisovan Ancestry in East Eurasian and Native American Populations. *Mol. Biol. Evol.* **32**, 2665-2674 (2015).
48. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655-1664 (2009).
49. Ramis, D., Alcover, J. A., Coll, J. & Trias, M. The Chronology of the First Settlement of the Balearic Islands. *Journal of Mediterranean Archaeology* **15**, 3-24 (2002).
50. Zalloua, P. *et al.* Ancient DNA of Phoenician remains indicates discontinuity in the settlement history of Ibiza. *Scientific Reports* **8**, 17567 (2018).
51. Matisoo-Smith, E. *et al.* Ancient mitogenomes of Phoenicians from Sardinia and Lebanon: A story of settlement, integration, and female mobility. *PLoS One* **13**, e0190169 (2018).
52. Fregel, R. *et al.* 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**, 6774-6779 (2018).
53. Moorjani, P. *et al.* The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet.* **7**, e1001373 (2011).
54. Loh, M. *et al.* Can population differences in chemotherapy outcomes be inferred from differences in pharmacogenetic frequencies? *Pharmacogenomics J.* **13**, 423-429 (2013).
55. Chiang, C. W. K. *et al.* Genomic history of the Sardinian population. *Nature Genetics* **50**, 1426-1434 (2018).
56. Hellenthal, G. *et al.* A Genetic Atlas of Human Admixture History. *Science* **343**, 747-751 (2014).
57. Solé-Morata, N. *et al.* Analysis of the R1b-DF27 haplogroup shows that a large fraction of Iberian Y-chromosome lineages originated recently in situ. *Sci. Rep.* **7**, 7341 (2017).
58. Raveane, A. *et al.* Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe. *bioRxiv* (2018). doi:10.1101/494898
59. Sangmeister, E. Die Datierung des Rickstroms der Glockenbecker und ihre Auswirkung auf die

- Chronologie der Kupferzeit in Portugal. *Palaeohistoria* **12**, 395-407 (1966).
60. D'Agata, A. L. Interactions between Aegean groups and local communities in Sicily in the Bronze Age: The evidence from pottery. *Studi micenei ed egeo-anatolici* **42**, 61-83 (2000).
61. Shelton, K. Mainland Greece. in *The Oxford Handbook of the Bronze Age Aegean* (ed. Cline, E. H.) 139-148 (Oxford University Press, 2012).
62. Alberti, G. Issues in the absolute chronology of the Early-Middle Bronze Age transition in Sicily and southern Italy: a Bayesian radiocarbon view. *Journal of Quaternary Science* **28**, 630-640 (2013).
63. Sikora, M. *et al.* 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 (2014).
64. Pinhasi, R. *et al.* Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PLoS One* **10**, e0129102 (2015).
65. Pinhasi, R., Fernandes, D. M., Sirak, K. & Cheronet, O. Isolating the human cochlea to generate bone powder for ancient DNA analysis. *Nature Protocols* (2019). doi:10.1038/s41596-019-0137-7
66. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLoS ONE* **5**, e14004 (2010).
67. Behar, D. M. *et al.* A 'Copernican' reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675-684 (2012).
68. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).
69. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
70. Kennett, D. J. *et al.* Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017).
71. Lohse, J. C., Culleton, B. J., Black, S. L. & D. J. Kennett, A. A Precise Chronology of Middle to Late Holocene Bison Exploitation in the Far Southern Great Plains. *Journal of Texas Archeology and History* **1**, 94-126 (2014).
72. van Klinken, G. J. Bone Collagen Quality Indicators for Palaeodietary and Radiocarbon Measurements. *Journal of Archaeological Science* **26**, 687-695 (1999).

73. Santos, G. M., Southon, J. R., Druffel-Rodriguez, K. C., Griffin, S. & Mazon, M. Magnesium  
Perchlorate as an Alternative Water Trap in AMS Graphite Sample Preparation: A Report On  
Sample Preparation at Kccams at the University of California, Irvine. *Radiocarbon* **46**, 165-173  
(2004).
74. Stuiver, M. & Polach, H. A. Discussion Reporting of <sup>14</sup>C Data. *Radiocarbon* **19**, 355-363 (1977).
75. Ramsey, C. B. & Lee, S. Recent and Planned Developments of the Program OxCal. *Radiocarbon*  
**55**, 720-730 (2013).
76. Reimer, P. J. *et al.* IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0-50,000 Years  
cal BP. *Radiocarbon* **55**, 1869-1887 (2013).
77. Kloss-Brandstätter, A. *et al.* HaploGrep: a fast and reliable algorithm for automatic  
classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25-32 (2011).
78. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human  
mitochondrial DNA variation. *Hum. Mutat.* **30**, E386-94 (2009).
79. Weissensteiner, H. *et al.* HaploGrep 2: mitochondrial haplogroup classification in the era of  
high-throughput sequencing. *Nucleic Acids Res.* **44**, W58-63 (2016).
80. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079  
(2009).
81. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genetics* **2**,  
e190 (2006).
82. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer  
datasets. *Gigascience* **4**, 7 (2015).