

 Open access • Posted Content • DOI:10.1101/584714

## The Arrival of Steppe and Iranian Related Ancestry in the Islands of the Western Mediterranean — [Source link](#)

[Daniel Fernandes](#), [Daniel Fernandes](#), [Daniel Fernandes](#), [Alissa Mittnik](#) ...+58 more authors

**Institutions:** [University of Vienna](#), [University of Coimbra](#), [University College Dublin](#), [Harvard University](#) ...+15 more institutions

**Published on:** 21 Mar 2019 - [bioRxiv](#) (Cold Spring Harbor Laboratory)

**Topics:** [Bronze Age](#)

Related papers:

- [The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean.](#)
- [Massive migration from the steppe was a source for Indo-European languages in Europe](#)
- [Ancient human genomes suggest three ancestral populations for present-day Europeans](#)
- [Genome-wide patterns of selection in 230 ancient Eurasians](#)
- [Genomic insights into the origin of farming in the ancient Near East](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/the-arrival-of-steppe-and-iranian-related-ancestry-in-the-4ql8dvhik1>

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

2 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  
3 Rohland<sup>4</sup>, Swapan Mallick<sup>4,5,6</sup>, Rebecca Bernardos<sup>4</sup>, Nasreen Broomandkshobacht<sup>4,5,‡</sup>, Jens Carlsson<sup>7</sup>,  
4 Brendan J. Cullen<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>,  
5 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>,  
6 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  
7 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.  
8 Miccichè<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>,  
9 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>,  
10 David Caramelli<sup>11,\*</sup>, Ron Pinhasi<sup>1,2,\*</sup> and David Reich<sup>4,5,6,27,\*</sup>

11

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

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

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

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

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

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

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

20 <sup>8</sup> Institutes for Energy and the Environment, The Pennsylvania State University, University Park, PA  
21 16802, USA.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

47 ‡ Present addresses: Department of Anthropology, University of California, Santa Cruz, CA 95064,  
48 USA (N.B.); Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138,  
49 USA (Me.M.); Department of Biomolecular Engineering, University of California, Santa Cruz, CA  
50 95064, USA (J.O.); Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115,  
51 USA (K.A.S.).

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

54

55

56

57

58

59 A series of studies have documented how Steppe pastoralist-related ancestry reached central  
60 Europe by at least 2500 BCE, while Iranian farmer-related ancestry was present in Aegean  
61 Europe by at least 1900 BCE. However, the spread of these ancestries into the western  
62 Mediterranean where they have contributed to many populations living today remains poorly  
63 understood. We generated genome-wide ancient DNA from the Balearic Islands, Sicily, and  
64 Sardinia, increasing the number of individuals with reported data from these islands from 3 to  
65 52. We obtained data from the oldest skeleton excavated from the Balearic islands (dating to  
66 ~2400 BCE), and show that this individual had substantial Steppe pastoralist-derived ancestry;  
67 however, later Balearic individuals had less Steppe heritage reflecting geographic  
68 heterogeneity or immigration from groups with more European first farmer-related ancestry. In  
69 Sicily, Steppe pastoralist ancestry arrived by ~2200 BCE and likely came at least in part from  
70 Spain as it was associated with Iberian-specific Y chromosomes. In Sicily, Iranian-related  
71 ancestry also arrived by the Middle Bronze Age, thus revealing that this ancestry type, which  
72 was ubiquitous in the Aegean by this time, also spread further west prior to the classical period  
73 of Greek expansion. In Sardinia, we find no evidence of either eastern ancestry type in the  
74 Nuragic Bronze Age, but show that Iranian-related ancestry arrived by at least ~300 BCE and  
75 Steppe ancestry arrived by ~300 CE, joined at that time or later by North African ancestry.  
76 These results falsify the view that the people of Sardinia are isolated descendants of Europe's  
77 first farmers. Instead, our results show that the island's admixture history since the Bronze Age  
78 is as complex as that in many other parts of Europe.

79

## 80 Introduction

81 The advent of the European Bronze Age after 3000 BCE was marked by an increase in long-range  
82 human mobility. People with ancestry from the Steppe north of the Black and Caspian Seas made a  
83 profound demographic impact in central and eastern Europe, mixing with local farmers to  
84 contribute up to three quarters of the ancestry of peoples associated with the Corded Ware  
85 complex<sup>1-3</sup>. The expansion of the Beaker complex after around 2400 BCE from the west had a less  
86 straightforward correlation to genetic ancestry. In Iberia, most people buried with artifacts of the  
87 Beaker complex had little if any Steppe pastoralist-related ancestry (from here on denoted “Steppe  
88 ancestry”), but Beaker cultural practices were adopted by people in Central Europe were in part  
89 descended from Steppe pastoralists and then spread this material culture along with Steppe  
90 ancestry to northwestern Europe<sup>4</sup>. In Iberia, Steppe ancestry began to appear in outlier individuals  
91 by ~2500 BCE<sup>4</sup>, and became fully mixed into the Iberian population by 2000 BCE<sup>5</sup>. Meanwhile on  
92 Crete in the eastern Mediterranean, there was little if any Steppe ancestry identified in all  
93 published samples from the Middle to Late Bronze Age “Minoan” culture (individuals dating to 2400-  
94 1700 BCE), although these individuals derived about 15% of their ancestry from groups related to  
95 early Iranian farmers (from here on referred to as “Iranian-related ancestry”)<sup>6</sup> (Fig. 1).

96 In the islands of the central and western Mediterranean, the Bronze Age transition has not been  
97 investigated with ancient DNA, despite the fact that archaeological evidence reveals that many of  
98 the same cultural changes that affected mainland Europe and the eastern Mediterranean also  
99 impacted this region<sup>7</sup>. The first evidence for a permanent human presence in the Balearic Islands is  
100 dated to just before the onset of the Bronze Age in this part of Europe, between ~2500-2300 BCE<sup>8,9</sup>.  
101 Early settlers initially relied on animal husbandry and their economy was focused on sheep, goat<sup>9,10</sup>,  
102 and cereal agriculture<sup>11</sup>, while exploitation of wild marine resources (fish, marine birds, mollusks)  
103 was central to subsistence on the small island of Formentera<sup>10,12</sup>. Around 1200 BCE, the  
104 development of the Talaiotic culture in Mallorca and Menorca (the easternmost Balearic Islands)  
105 was marked by intensified management of food resources and the appearance of monumental  
106 towers, the eponymous talaiots. These structures were similar in style to the Sardinian nuraghi<sup>10,13</sup>,  
107 raising the question of whether there was a cultural connection<sup>14</sup>, a scenario that would gain  
108 plausibility if there was substantial genetic exchange between the two regions. Nuragic Sardinians  
109 were also in cultural contact with groups from the eastern Mediterranean<sup>15</sup>, so an important  
110 question is whether they were admixed with either Steppe or Iranian-related ancestry. Meanwhile,  
111 the central Mediterranean island of Sicily was affected by the spread of Beaker cultural complex  
112 after around 2400 BCE, and by cultural influence from the Aegean in the Late Helladic Period  
113 ~1600-1200 BCE (the period of the “Mycenaean” culture)<sup>16-18</sup>. An unanswered question is whether  
114 these events or other cultural changes on the island involved substantial movements of people.

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

## 122 **Results**

### 123 **Samples and sequencing results**

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

134 **Materials**). We established chronology based on archaeological context and by assembling direct  
135 radiocarbon dates on bone for 28 of the individuals (direct dates for 26 individuals are reported for  
136 the first time here; **Online Table 2**). We removed from the analysis dataset eight individuals with  
137 fewer than 20,000 of the targeted SNPs covered by at least one sequence, and five with evidence of  
138 substantial contamination or less than 3% cytosine-to-thymine error in terminal cytosines. We also  
139 removed one individual who we detected as a first degree relative (a son) of another (his mother)  
140 that gave higher quality genetic data. This left 36 individuals for our modeling (however, since all  
141 the data are useful we fully report all individuals, **Online Table 1**). In the analysis dataset, the  
142 median coverage on targeted SNPs on chromosomes 1-22 was 3.20-fold (range 0.02-12.13), and the  
143 median number of SNPs covered by at least one sequence was 756709 (range 23600-1038409). All  
144 mitochondrial DNA point estimates for match rate to the consensus sequence had 95% confidence  
145 intervals with upper bounds from 0.96-1.00, while contamination estimates based on X chromosome  
146 variation (meaningful only in males) were all below 1.1% (**Online Table 1**). All individuals had data  
147 from at least one library with cytosine-to-thymine damage in the terminal nucleotides greater than  
148 3% (the minimum suggested as a guideline for the plausible authentic DNA<sup>23</sup>). The qualitative  
149 patterns of ancestry in the data were unchanged when we restricted to transversion SNPs which are  
150 not affected by characteristic ancient DNA errors (**Supplementary Fig. 1**).

#### 151 **Genetic affinities and population groupings**

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

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

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

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

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

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

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

206

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

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

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

231 During the Iron Age, Phoenician colonies were established in the Balearic islands. The Ibiza  
232 Phoenician individual published in <sup>50</sup> is not consistent with forming a clade with any of the Bronze  
233 Age individuals from the Balearic islands newly reported in this study, and indeed we find that she  
234 can not be modeled even with our least parsimonious model of 4 distal sources. However, when we  
235 add in a North African source of ancestry, we can fit her as a two-way mix of  $18.8 \pm 7.9\%$   
236 *Anatolia\_Neolithic* and  $81.2 \pm 7.9\%$  *Morocco\_LN* ancestry ( $p=0.141$ ) (**Supplementary Materials**). We  
237 also can fit the Ibiza Phoenician as two-way mixture of a variety of groups closer to her in time one  
238 of which is always *Morocco\_LN*. While several of these models include a Balearic Island Bronze Age  
239 source, we cannot rule out the possibility that the Ibiza Phoenician individual has no local Balearic  
240 ancestry at all. Specifically, we find that we can fit her with models that do not have a Balearic  
241 source and that instead have Balearic Bronze Age individuals in the outgroups (e.g. (e.g.  $17.1 \pm$   
242  $3.5\%$  *France\_Bell\_Beaker* and  $82.9 \pm 3.5\%$  *Morocco\_LN*,  $p=0.869$ ) (**Supplementary Table 11**).

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



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

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

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

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

256 Most Sardinians buried in a Nuragic Bronze Age context possessed uniparental haplogroups found in  
257 European hunter-gatherers and early farmers, including Y-haplogroup R1b1a[xR1b1a1a] which is  
258 different from the characteristic R1b1a1a2a1a2 spread in association with the Bell Beaker complex<sup>4</sup>.  
259 An exception is individual I10553 (1226-1056 calBCE) who carried Y-haplogroup J2b2a (**Online Table**  
260 **1**), previously observed in a Croatian Middle Bronze Age individual bearing Steppe ancestry<sup>44</sup>,  
261 suggesting the possibility of genetic input from groups that arrived from the east after the spread  
262 of first farmers. This is consistent with the evidence of material culture exchange between  
263 Sardinians and mainland Mediterranean groups<sup>15</sup>, although genome-wide analyses find no significant  
264 evidence of Steppe ancestry so the quantitative demographic impact was minimal. *qpAdm* modeling  
265 of the ancestry of the *Sardinia\_Nuragic\_BA10365* outlier with respect to sources potentially more  
266 closely related in space and time does infer some ancestry in this individual from an eastern source  
267 (either carrying Steppe ancestry or Iranian-related ancestry) that we do not detect by modeling  
268 with sources more distant in space and time, consistent with the hypothesis of eastern influence  
269 (**Supplementary Table 12**).

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

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

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

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

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

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

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

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

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

342 We detect Iranian-related ancestry in Sicily by the Middle Bronze Age 1800-1500 BCE, consistent  
343 with the directional shift of these individuals toward Mycenaeans in PCA (**Fig. 2b**). Specifically, two  
344 of the Middle Bronze Age individuals can only be fit with models that in addition to  
345 *Anatolia\_Neolithic* and *WHG*, include *Iran\_Ganj\_Dareh\_Neolithic*. The most parsimonious model for  
346 *Sicily\_MBA3125* has  $18.0 \pm 3.6\%$  Iranian-related ancestry ( $p=0.032$  for rejecting the alternative  
347 model of Steppe rather than Iranian-related ancestry), and the most parsimonious model for  
348 *Sicily\_MBA4109* has  $14.9 \pm 3.9\%$  Iranian-related ancestry ( $p=0.037$  for rejecting the alternative  
349 model) (**Fig. 4a, Supplementary Table 9**). This inference is also supported by *qpAdm* using sources  
350 closer in geography and time that always identify a parsimonious model with *Minoan\_Lassithi* as a  
351 source for these two individuals (**Supplementary Table 15**). We also found evidence of Iranian-  
352 related ancestry in Sicily in an individual of the Early Bronze Age cluster, I11442, who could only be  
353 fit in a 3-way model with Iranian-related ancestry ( $19.3 \pm 3.8\%$  ancestry of this type,  $p=0.391$ ; the  
354 3-way model involving Steppe ancestry fails to a fit ( $p=0.010$ )) (**Supplementary Table 10**). However,  
355 this finding should be viewed with caution as *qpWave* clustered this individual with four other  
356 Sicilian Early Bronze Age individuals, so this finding could be an artifact of performing tests on our  
357 data beyond what is justified by our groupings. The modern southern Italian Caucasus-related signal

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

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

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

370

## 371 Discussion

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

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

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

393 Our co-analysis of modern and ancient Sardinians questions the commonly held view that Sardinians  
394 are well described as an isolated remnant of Europe's first farmers<sup>63</sup>. While Nuragic Bronze Age  
395 Sardinians are indeed well-modeled as having a typical early European farmer ancestry profile,  
396 modern Sardinians harbor substantial fractions of ancestry from several groups that arrived in  
397 Europe after the Neolithic, and we model modern Sardinians as harboring  $10.0 \pm 1.6\%$  Steppe  
398 ancestry and an even larger  $19.1 \pm 1.9\%$  Iranian-related ancestry. Both ancestry types are  
399 definitively required to model modern Sardinians, and we show that modern Sardinians have been  
400 substantially impacted by movement of ancestry from North Africa in the last two millennia. Thus,  
401 rather than being an island sheltered from admixture and migration since the Neolithic, Sardinia,  
402 like almost all other regions in Europe has, been a site for major movement and mixtures of people.  
403

## 404 **Materials and Methods**

### 405 **Laboratory work details**

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

### 436 **Radiocarbon dating and quality assurance**

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

442 (convervants/adhesives) by sonicating all bone samples in successive washes of ACS grade methanol,  
443 acetone, and dichloromethane for 30 minutes each at room temperature, followed by three washes  
444 in Nanopure water to rinse. We extracted bone collagen and purified using a modified Longin  
445 method with ultrafiltration (>30kDa gelatin<sup>70</sup>). If collagen yields were low and amino acids poorly  
446 preserved we used a modified XAD process (XAD Amino Acids<sup>71</sup>). For quality assurance, we  
447 measured carbon and nitrogen concentrations and C/N ratios of all extracted and purified  
448 collagen/amino acid samples with a Costech elemental analyzer (ECS 4010). We evaluated sample  
449 quality by % crude gelatin yield, %C, %N and C/N ratios before AMS 14C dating. C/N ratios for all  
450 directly radiocarbon samples fell between 2.9 and 3.6, indicating excellent preservation<sup>72</sup>. We  
451 combusted collagen/amino acid samples (~2.1 mg) for 3 h at 900 °C in vacuum-sealed quartz tubes  
452 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,  
453 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  
454 and loaded them onto a target wheel with OX-1 (oxalic acid) standards, known-age bone  
455 secondaries, and a 14C-free Pleistocene whale blank. We made all 14C measurements on a modified  
456 National Electronics Corporation compact spectrometer with a 0.5 MV accelerator (NEC 1.5SDH-1).  
457 We corrected the 14C ages for mass-dependent fractionation with measured  $\delta^{13}\text{C}$  values<sup>74</sup> and  
458 compared with samples of Pleistocene whale bone (backgrounds, 48,000 14C BP), late Holocene  
459 bison bone (~1,850 14C BP), late 1800s CE cow bone, and OX-2 oxalic acid standards. We calibrated  
460 14C ages with OxCal version 4.3<sup>75</sup> and the IntCal13 northern hemisphere curve<sup>76</sup>. The stable carbon  
461 and nitrogen isotope measurements we obtained do not indicate a large marine dietary component  
462 in these individuals despite their coming from island populations and hence we did not perform a  
463 correction of the dates for marine reservoir effect.

#### 464 **Uniparental haplogroup determination**

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

#### 475 **Dataset assembly**

476 We assembled a base dataset and then subsetted for each analysis. This complete dataset included  
477 3310 individuals, of which 2191 were modern<sup>27,39,45-47</sup> and 1119 were ancient individuals from  
478 previous publications<sup>1,2,4,6,26-44,52</sup>, which we combined with the newly reported 49 samples (**Online  
479 Table 3**). We performed all subsequent analysis on autosomal data.

## 480 **Principal component analysis**

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

## 487 **Population structure analysis**

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

## 498 **$f_4$ -statistics**

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

## 506 ***qpWave/qpAdm***

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

## 514 **Data Availability**



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

### 518 **Acknowledgements**

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

### 531 **Author Contributions**

532 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.,  
533 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  
534 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.,  
535 M.N., J.O., K.A.S., K.S., and S.V. performed laboratory work, while N.R., D.C., and R.P. supervised  
536 this work. J.C. provided computing resources. B.J.C. performed radiocarbon analysis and D.J.K.  
537 supervised this work. D.M.F., I.O., R.B., S.M., and M.Ma. performed bioinformatic and population  
538 genetic analysis with input from A.Mi., I.L., N.P., and D.R.

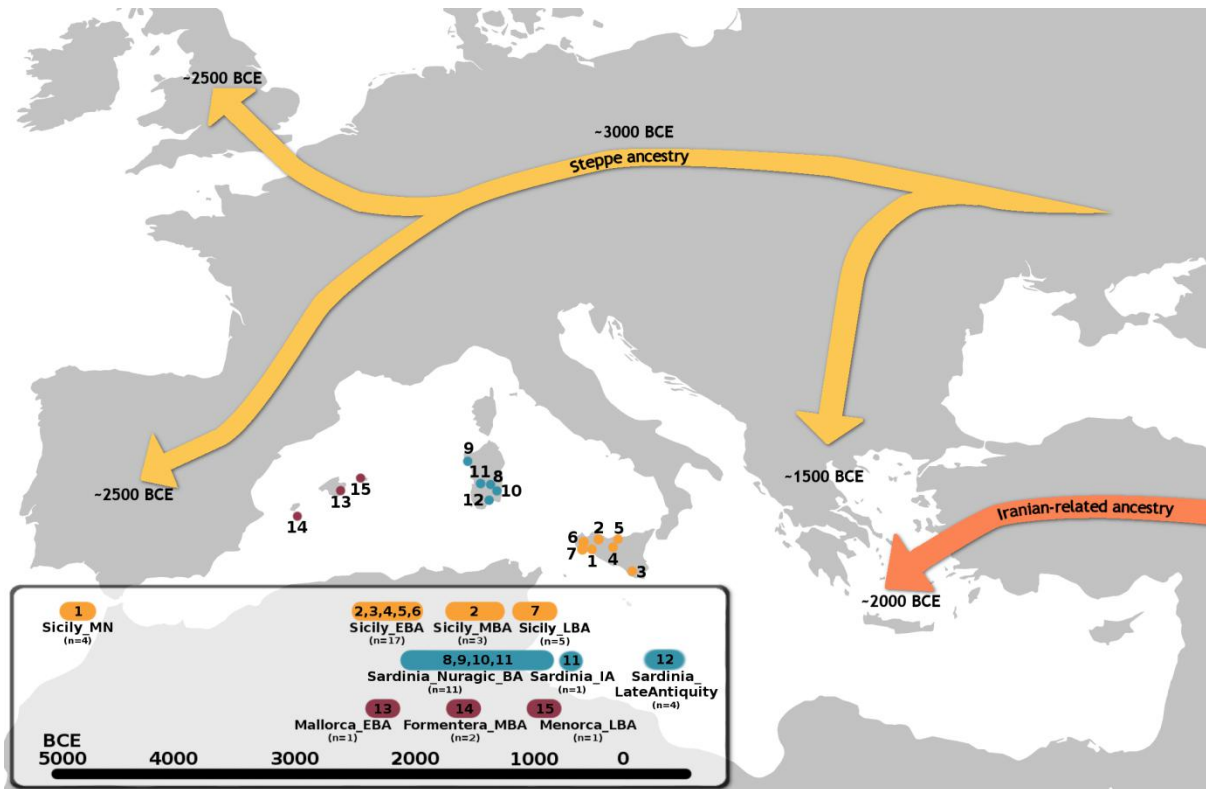
### 539 **Competing Interests**

540 The authors declare no competing financial interests.

541

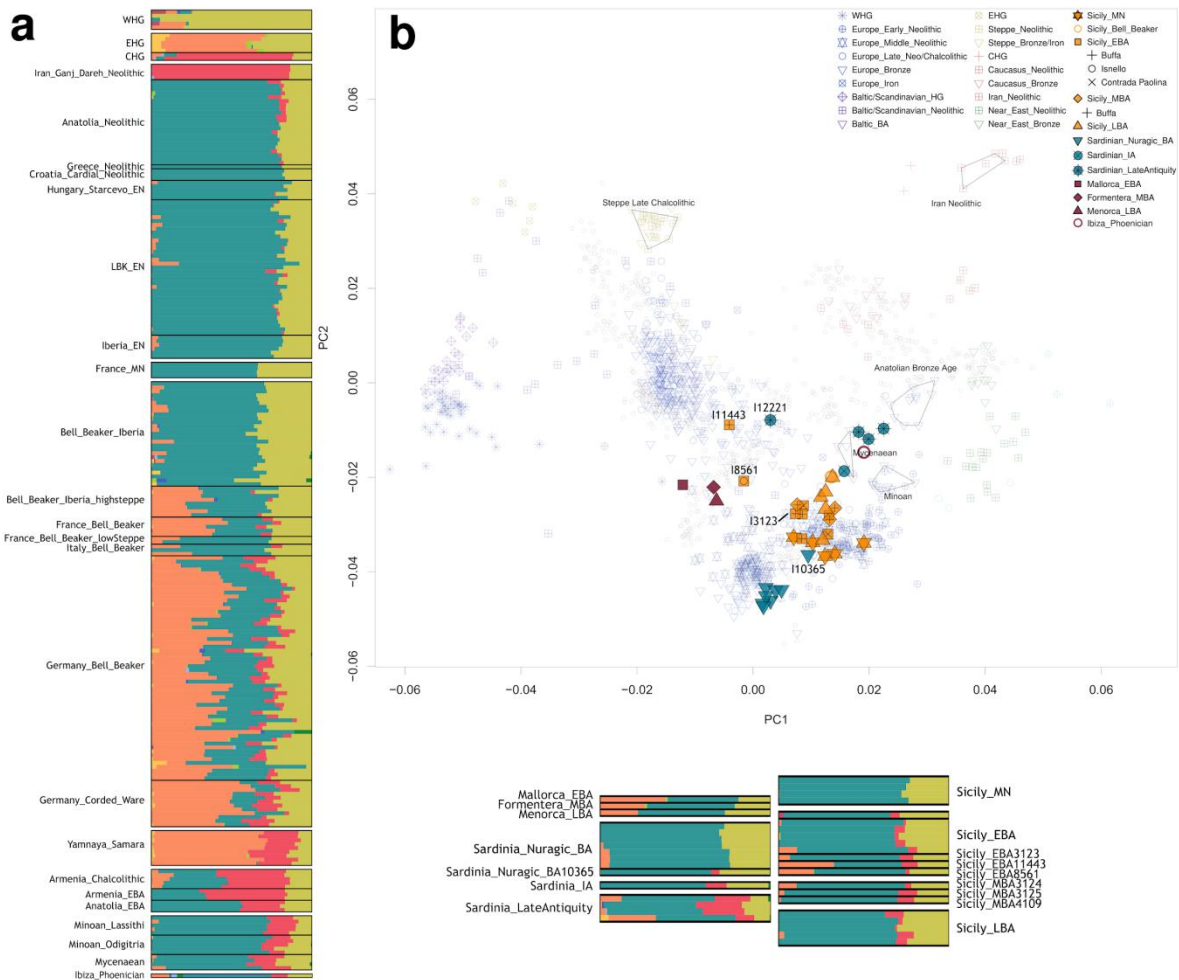
542

543 **Figures**



544

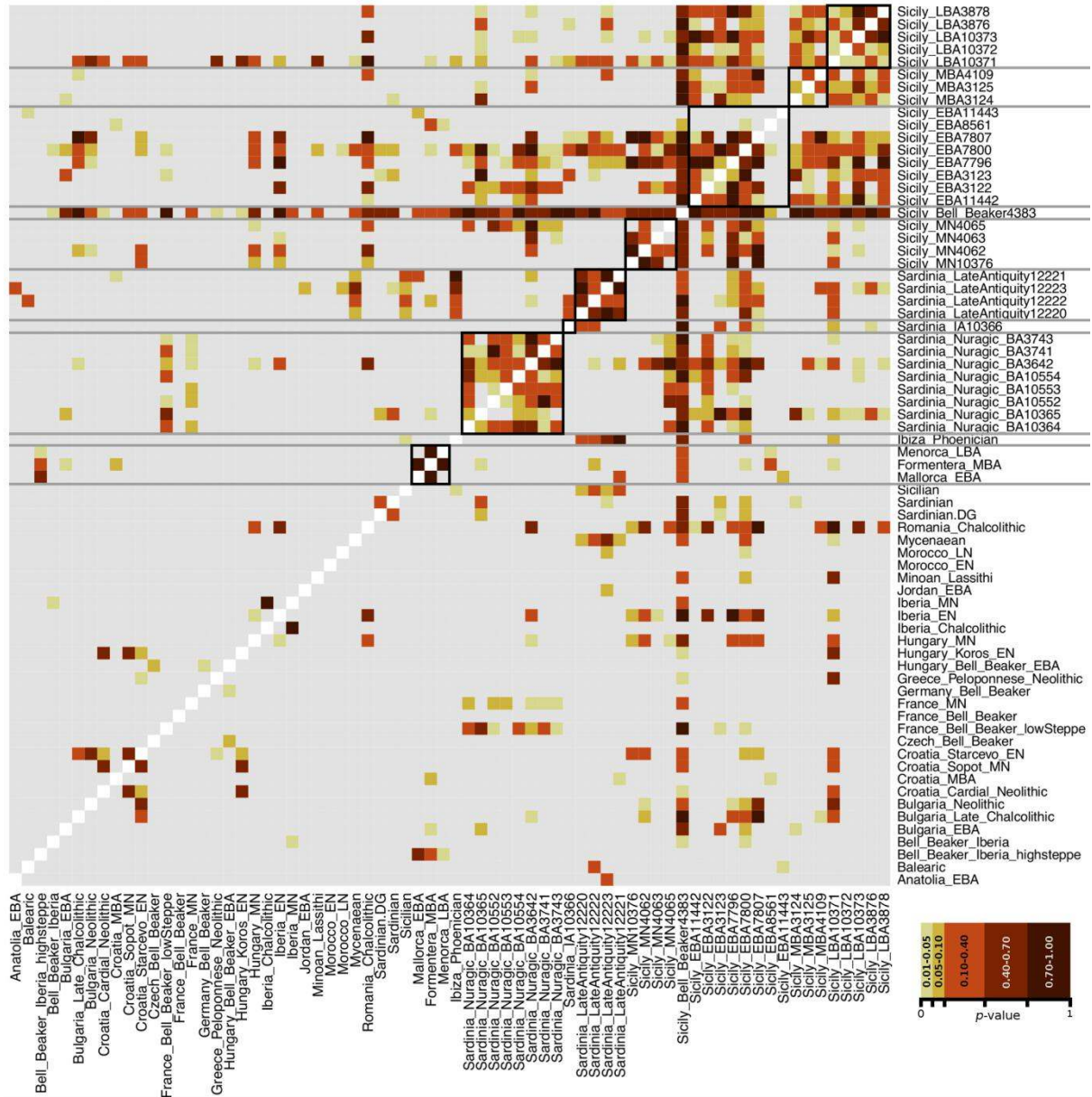
545 **Figure 1: Timeline and geographical origins of the 49 newly reported ancient individuals along**  
 546 **with the previously reported individual for whom we increase data quality. 1-Fossato di Stretto**  
 547 **Partana; 2-Bufa cave; 3-Contrada Paolina; 4-Isnello; 5-Vallone Inferno; 6-Marcita; 7-Salaparuta;**  
 548 **8-Seulo; 9-Alghero-Lu Maccioni cave; 10-Perdasdefogu; 11-Usellus; 12-Grotta Colombi; 13-Cova**  
 549 **des Moro; 14-Cap de Barbaria; 15-Naveta des Tudons.**



550

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

555



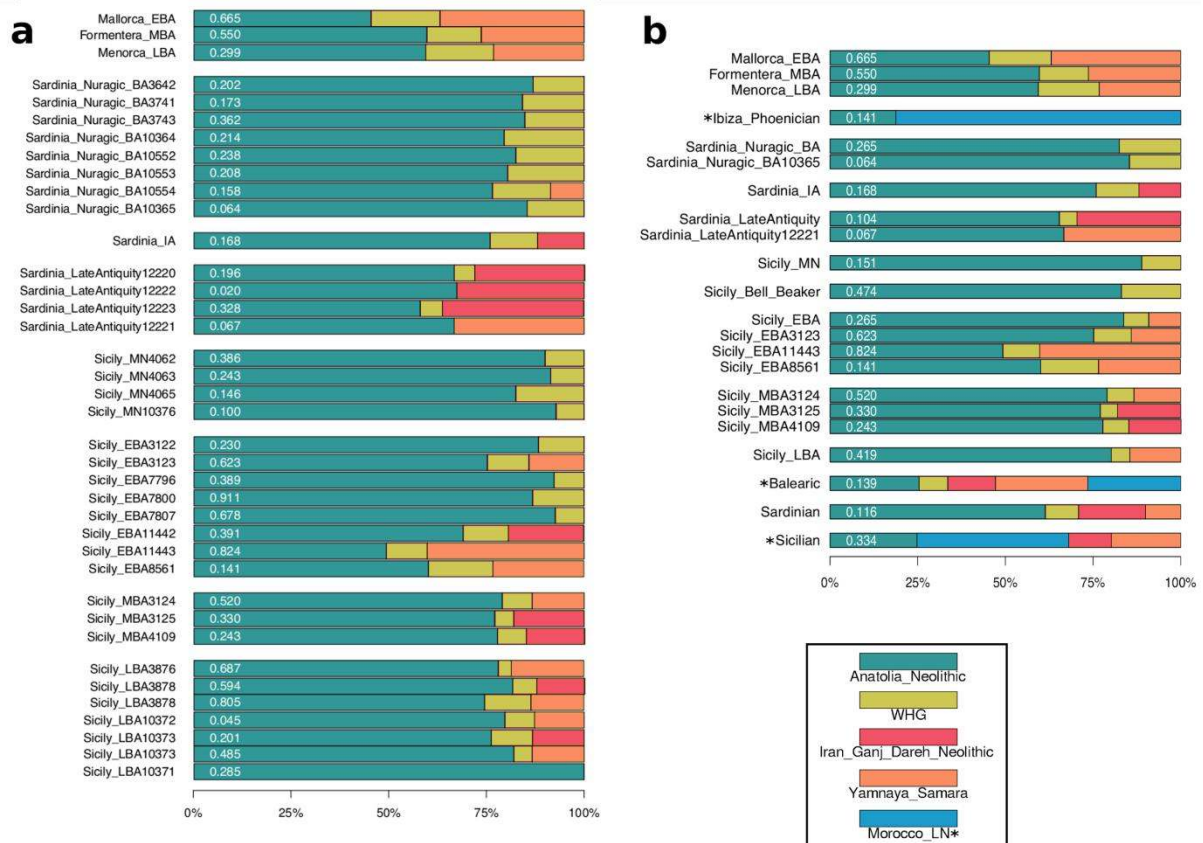
556

557 **Figure 3: Pairwise *qpWave* testing to group individuals. Black lines represent the initial clusters**  
 558 **of individuals from this study by location and/or period. Gray-coloured models have a p-value**  
 559 **below 0.010 and are rejected.**

560

561

562



563

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

573

574

575

576 **References**

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

- 606 14. Lull, V., Mico, R., Rihuete, C. I. & Risch, R. Los cambios sociales en las islas Baleares a lo largo  
607 del II milenio. *Cypsela* **15**, 123-148 (2004).
- 608 15. Holt, E. Mobility and meaning in the Nuragic culture of Bronze Age Sardinia (ca. 1700-900BC).  
609 in *Forging Identities. The Mobility of Culture in Bronze Age Europe* (eds. Suchowska-Ducke, P.,  
610 Reiter, S. S. & Vandkilde, H.) **1**, 193-202 (British Archaeological Reports, 2015).
- 611 16. Sestieri, A. M. B. The Bronze Age in Sicily. in *The Oxford Handbook of European Bronze Age*  
612 (eds. Harding, A. & Fokkens, H.) 653-667 (Oxford University Press, 2013).
- 613 17. Sarno, S. *et al.* Ancient and recent admixture layers in Sicily and Southern Italy trace multiple  
614 migration routes along the Mediterranean. *Sci. Rep.* **7**, 1984 (2017).
- 615 18. Holloway, R. R. *The Archaeology of Ancient Sicily*. (Routledge, 2000).
- 616 19. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear  
617 reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15758-15763  
618 (2013).
- 619 20. Damgaard, P. B. *et al.* Improving access to endogenous DNA in ancient bones and teeth. *Sci.*  
620 *Rep.* **5**, 11184 (2015).
- 621 21. Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions from  
622 ancient bones and teeth. *Biotechniques* **59**, 87-93 (2015).
- 623 22. Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded DNA from  
624 ancient bones, teeth and sediments for high-throughput sequencing. *Nature Protocols* **13**,  
625 2447-2461 (2018).
- 626 23. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase  
627 treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130624  
628 (2015).
- 629 24. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes.  
630 *Curr. Biol.* **23**, 553-559 (2013).
- 631 25. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature*  
632 **524**, 216-219 (2015).
- 633 26. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by  
634 whole-genome sequencing. *Nat. Commun.* **3**, 698 (2012).
- 635 27. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day

- 636 Europeans. *Nature* **513**, 409-413 (2014).
- 637 28. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory.  
638 *Nat. Commun.* **5**, 5257 (2014).
- 639 29. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old  
640 Mesolithic European. *Nature* **507**, 225-228 (2014).
- 641 30. Skoglund, P. *et al.* Genomic diversity and admixture differs for Stone-Age Scandinavian  
642 foragers and farmers. *Science* **344**, 747-750 (2014).
- 643 31. Günther, T. *et al.* Ancient genomes link early farmers from Atapuerca in Spain to modern-day  
644 Basques. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11917-11922 (2015).
- 645 32. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat.*  
646 *Commun.* **6**, 8912 (2015).
- 647 33. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**,  
648 499-503 (2015).
- 649 34. Olalde, I. *et al.* A Common Genetic Origin for Early Farmers from Mediterranean Cardial and  
650 Central European LBK Cultures. *Mol. Biol. Evol.* **32**, 3132-3142 (2015).
- 651 35. Broushaki, F. *et al.* Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**,  
652 499-503 (2016).
- 653 36. Cassidy, L. M. *et al.* Neolithic and Bronze Age migration to Ireland and establishment of the  
654 insular Atlantic genome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 368-373 (2016).
- 655 37. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205 (2016).
- 656 38. Hofmanová, Z. *et al.* Early farmers from across Europe directly descended from Neolithic  
657 Aegeans. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6886-6891 (2016).
- 658 39. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature*  
659 **536**, 419-424 (2016).
- 660 40. Martiniano, R. *et al.* Genomic signals of migration and continuity in Britain before the Anglo-  
661 Saxons. *Nat. Commun.* **7**, 10326 (2016).
- 662 41. Schiffels, S. *et al.* Iron Age and Anglo-Saxon genomes from East England reveal British  
663 migration history. *Nat. Commun.* **7**, 10408 (2016).
- 664 42. Jones, E. R. *et al.* The Neolithic Transition in the Baltic Was Not Driven by Admixture with  
665 Early European Farmers. *Current Biology* **27**, 576-582 (2017).



- 666 43. Unterländer, M. *et al.* Ancestry and demography and descendants of Iron Age nomads of the  
667 Eurasian Steppe. *Nat. Commun.* **8**, 14615 (2017).
- 668 44. Mathieson, I. *et al.* The genomic history of southeastern Europe. *Nature* **555**, 197-203 (2018).
- 669 45. Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192**, 1065-1093 (2012).
- 670 46. Pickrell, J. K. *et al.* The genetic prehistory of southern Africa. *Nat. Commun.* **3**, 1143 (2012).
- 671 47. Qin, P. & Stoneking, M. Denisovan Ancestry in East Eurasian and Native American Populations.  
672 *Mol. Biol. Evol.* **32**, 2665-2674 (2015).
- 673 48. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in  
674 unrelated individuals. *Genome Res.* **19**, 1655-1664 (2009).
- 675 49. Ramis, D., Alcover, J. A., Coll, J. & Trias, M. The Chronology of the First Settlement of the  
676 Balearic Islands. *Journal of Mediterranean Archaeology* **15**, 3-24 (2002).
- 677 50. Zalloua, P. *et al.* Ancient DNA of Phoenician remains indicates discontinuity in the settlement  
678 history of Ibiza. *Scientific Reports* **8**, 17567 (2018).
- 679 51. Matisoo-Smith, E. *et al.* Ancient mitogenomes of Phoenicians from Sardinia and Lebanon: A  
680 story of settlement, integration, and female mobility. *PLoS One* **13**, e0190169 (2018).
- 681 52. Fregel, R. *et al.* Ancient genomes from North Africa evidence prehistoric migrations to the  
682 Maghreb from both the Levant and Europe. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6774-6779  
683 (2018).
- 684 53. Moorjani, P. *et al.* The history of African gene flow into Southern Europeans, Levantines, and  
685 Jews. *PLoS Genet.* **7**, e1001373 (2011).
- 686 54. Loh, M. *et al.* Can population differences in chemotherapy outcomes be inferred from  
687 differences in pharmacogenetic frequencies? *Pharmacogenomics J.* **13**, 423-429 (2013).
- 688 55. Chiang, C. W. K. *et al.* Genomic history of the Sardinian population. *Nature Genetics* **50**, 1426-  
689 1434 (2018).
- 690 56. Hellenthal, G. *et al.* A Genetic Atlas of Human Admixture History. *Science* **343**, 747-751 (2014).
- 691 57. Solé-Morata, N. *et al.* Analysis of the R1b-DF27 haplogroup shows that a large fraction of  
692 Iberian Y-chromosome lineages originated recently in situ. *Sci. Rep.* **7**, 7341 (2017).
- 693 58. Raveane, A. *et al.* Population structure of modern-day Italians reveals patterns of ancient and  
694 archaic ancestries in Southern Europe. *bioRxiv* (2018). doi:10.1101/494898
- 695 59. Sangmeister, E. Die Datierung des Rickstroms der Glockenbecker und ihre Auswirkung auf die

- 696 Chronologie der Kupferzeit in Portugal. *Palaeohistoria* **12**, 395-407 (1966).
- 697 60. D'Agata, A. L. Interactions between Aegean groups and local communities in Sicily in the  
698 Bronze Age: The evidence from pottery. *Studi micenei ed egeo-anatolici* **42**, 61-83 (2000).
- 699 61. Shelton, K. Mainland Greece. in *The Oxford Handbook of the Bronze Age Aegean* (ed. Cline, E.  
700 H.) 139-148 (Oxford University Press, 2012).
- 701 62. Alberti, G. Issues in the absolute chronology of the Early-Middle Bronze Age transition in Sicily  
702 and southern Italy: a Bayesian radiocarbon view. *Journal of Quaternary Science* **28**, 630-640  
703 (2013).
- 704 63. Sikora, M. *et al.* Population genomic analysis of ancient and modern genomes yields new  
705 insights into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe.  
706 *PLoS Genet.* **10**, e1004353 (2014).
- 707 64. Pinhasi, R. *et al.* Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous  
708 Bone. *PLoS One* **10**, e0129102 (2015).
- 709 65. Pinhasi, R., Fernandes, D. M., Sirak, K. & Cheronet, O. Isolating the human cochlea to generate  
710 bone powder for ancient DNA analysis. *Nature Protocols* (2019). doi:10.1038/s41596-019-0137-7
- 711 66. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA Sequence Capture of Mitochondrial  
712 Genomes Using PCR Products. *PLoS ONE* **5**, e14004 (2010).
- 713 67. Behar, D. M. *et al.* A 'Copernican' reassessment of the human mitochondrial DNA tree from its  
714 root. *Am. J. Hum. Genet.* **90**, 675-684 (2012).
- 715 68. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform.  
716 *Bioinformatics* **26**, 589-595 (2010).
- 717 69. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation  
718 Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
- 719 70. Kennett, D. J. *et al.* Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat.*  
720 *Commun.* **8**, 14115 (2017).
- 721 71. Lohse, J. C., Culleton, B. J., Black, S. L. & D. J. Kennett, A. A Precise Chronology of Middle to  
722 Late Holocene Bison Exploitation in the Far Southern Great Plains. *Journal of Texas Archeology*  
723 *and History* **1**, 94-126 (2014).
- 724 72. van Klinken, G. J. Bone Collagen Quality Indicators for Palaeodietary and Radiocarbon  
725 Measurements. *Journal of Archaeological Science* **26**, 687-695 (1999).

- 726 73. Santos, G. M., Southon, J. R., Druffel-Rodriguez, K. C., Griffin, S. & Mazon, M. Magnesium  
727 Perchlorate as an Alternative Water Trap in AMS Graphite Sample Preparation: A Report On  
728 Sample Preparation at Kccams at the University of California, Irvine. *Radiocarbon* **46**, 165-173  
729 (2004).
- 730 74. Stuiver, M. & Polach, H. A. Discussion Reporting of <sup>14</sup>C Data. *Radiocarbon* **19**, 355-363 (1977).
- 731 75. Ramsey, C. B. & Lee, S. Recent and Planned Developments of the Program OxCal. *Radiocarbon*  
732 **55**, 720-730 (2013).
- 733 76. Reimer, P. J. *et al.* IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0-50,000 Years  
734 cal BP. *Radiocarbon* **55**, 1869-1887 (2013).
- 735 77. Kloss-Brandstätter, A. *et al.* HaploGrep: a fast and reliable algorithm for automatic  
736 classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25-32 (2011).
- 737 78. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human  
738 mitochondrial DNA variation. *Hum. Mutat.* **30**, E386-94 (2009).
- 739 79. Weissensteiner, H. *et al.* HaploGrep 2: mitochondrial haplogroup classification in the era of  
740 high-throughput sequencing. *Nucleic Acids Res.* **44**, W58-63 (2016).
- 741 80. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079  
742 (2009).
- 743 81. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genetics* **2**,  
744 e190 (2006).
- 745 82. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer  
746 datasets. *Gigascience* **4**, 7 (2015).

747