

1 Population history from the Neolithic to present on the  
2 Mediterranean island of Sardinia: An ancient DNA perspective

3 Joseph H. Marcus<sup>1,\*</sup>, Cosimo Posth<sup>2,3,\*</sup>, Harald Ringbauer<sup>1,\*</sup>, Luca Lai<sup>4</sup>, Robin  
4 Skeates<sup>5</sup>, Carlo Sidore<sup>6,7</sup>, Jessica Beckett<sup>8</sup>, Anja Furtwängler<sup>3</sup>, Anna Olivieri<sup>9</sup>,  
5 Charleston Chiang<sup>10</sup>, Hussein Al-Asadi<sup>11,12</sup>, Kushal Dey<sup>13</sup>, Tyler A. Joseph<sup>14</sup>,  
6 Clio Der Sarkissian<sup>15</sup>, Rita Radzevičiūtė<sup>2</sup>, Maria Giuseppina Gradoli<sup>16</sup>, Wolfgang  
7 Haak<sup>2</sup>, David Reich<sup>17,18,19,20</sup>, David Schlessinger<sup>21</sup>, Francesco Cucca<sup>6,7,†</sup>, Johannes  
8 Krause<sup>2,3,20†</sup> and John Novembre<sup>1,22,†</sup>

9 <sup>1</sup>Department of Human Genetics, University of Chicago, Chicago, IL, USA.

10 <sup>2</sup>Max Planck Institute for the Science of Human History, Jena, Germany.

11 <sup>3</sup>Institute for Archaeological Sciences, University of Tübingen, Tübingen, Germany.

12 <sup>4</sup>Department of Anthropology, University of South Florida, Tampa, FL, USA.

13 <sup>5</sup>Department of Archaeology, Durham University, Durham, United Kingdom.

14 <sup>6</sup>Istituto di Ricerca Genetica e Biomedica - CNR, Cagliari, Italy.

15 <sup>7</sup>Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy.

16 <sup>8</sup>Private contractor, Cagliari, Sardinia.

17 <sup>9</sup>Dipartimento di Biologia e Biotecnologie "L. Spallanzani", Università di Pavia, Pavia, Italy.

18 <sup>10</sup>Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of  
19 Southern California, Los Angeles, CA, USA.

20 <sup>11</sup>Department of Statistics, University of Chicago, Chicago, IL, USA.

21 <sup>12</sup>Committee on Evolutionary Biology, University of Chicago, Chicago, IL, USA.

22 <sup>13</sup>Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA.

23 <sup>14</sup>Department of Computer Science, Columbia University, New York, NY, USA.

24 <sup>15</sup>Team AGES Laboratory AMIS Faculté de Médecine de Purpan, Toulouse, France.

25 <sup>16</sup>School of Archaeology and Ancient History, University of Leicester, Leicester, United Kingdom.

26 <sup>17</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

27 <sup>18</sup>Broad Institute of Harvard and MIT, Cambridge, MA, USA.

28 <sup>19</sup>Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA.

29 <sup>20</sup>Max Planck-Harvard Research Center for the Archaeoscience of the Ancient Mediterranean

30 <sup>21</sup>Laboratory of Genetics, NIA, NIH, Baltimore, MD, USA.

31 <sup>22</sup>Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA.

32 \*These authors contributed equally to this work.

33 †Corresponding authors.

## 34 Abstract

35 Recent ancient DNA studies of western Eurasia have revealed a dynamic history of admixture,  
36 with evidence for major migrations during the Neolithic and Bronze Age. The population of the  
37 Mediterranean island of Sardinia has been notable in these studies – Neolithic individuals from  
38 mainland Europe cluster more closely with Sardinian individuals than with all other present-day  
39 Europeans. The current model to explain this result is that Sardinia received an initial influx  
40 of Neolithic ancestry and then remained relatively isolated from expansions in the later Ne-  
41 olithic and Bronze Age that took place in continental Europe. To test this model, we generated  
42 genome-wide capture data (approximately 1.2 million variants) for 43 ancient Sardinian individu-  
43 als spanning the Neolithic through the Bronze Age, including individuals from Sardinia’s Nuragic  
44 culture, which is known for the construction of numerous large stone towers throughout the is-  
45 land. We analyze these new samples in the context of previously generated genome-wide ancient  
46 DNA data from 972 ancient individuals across western Eurasia and whole-genome sequence data  
47 from approximately 1,500 modern individuals from Sardinia. The ancient Sardinian individuals  
48 show a strong affinity to western Mediterranean Neolithic populations and we infer a high degree  
49 of genetic continuity on the island from the Neolithic (around fifth millennium BCE) through  
50 the Nuragic period (second millennium BCE). In particular, during the Bronze Age in Sardinia,  
51 we do not find significant levels of the “Steppe” ancestry that was spreading in many other parts  
52 of Europe at that time. We also characterize subsequent genetic influx between the Nuragic  
53 period and the present. We detect novel, modest signals of admixture between 1,000 BCE and  
54 present-day, from ancestry sources in the eastern and northern Mediterranean. Within Sardinia,  
55 we confirm that populations from the more geographically isolated mountainous provinces have  
56 experienced elevated levels of genetic drift and that northern and southwestern regions of the  
57 island received more gene flow from outside Sardinia. Overall, our genetic analysis sheds new  
58 light on the origin of Neolithic settlement on Sardinia, reinforces models of genetic continuity on  
59 the island, and provides enhanced power to detect post-Bronze-Age gene flow. Together, these  
60 findings offer a refined demographic model for future medical genetic studies in Sardinia.

## 61 Introduction

62 The whole-genome sequencing of Ötzi, a Neolithic individual who was preserved in ice for over  
63 5,000 years near the Italo-Austrian border, revealed a surprisingly high level of shared ancestry  
64 with present-day Sardinian individuals (ERMINI *et al.*, 2008; KELLER *et al.*, 2012; SIKORA *et al.*,  
65 2014). Subsequent work on genome-wide variation in ancient Europeans expanded upon this  
66 observation, finding that many “early European farmer” individuals have their highest genetic  
67 affinity with present-day Sardinian individuals, even when from geographically distant locales  
68 (e.g. from Hungary, Germany, Spain, Sweden) (e.g. SKOGLUND *et al.*, 2012, 2014).

69 Accumulating ancient DNA (aDNA) results have provided a potential framework for un-  
70 derstanding how early European farmers, such as Ötzi, show such genetic affinity to modern  
71 Sardinians. In this framework, Europe was first inhabited by Paleolithic hunter-gatherer groups.  
72 Then, starting about 7,000 BCE, farming peoples arrived from the Middle East as part of a Ne-  
73 olithic transition (AMMERMAN and CAVALLI-SFORZA, 2014; LAZARIDIS *et al.*, 2014), spreading  
74 through Anatolia and the Balkans (HOFMANOVÁ *et al.*, 2016; MATHIESON *et al.*, 2018) while  
75 progressively admixing with local hunter-gatherers (LIPSON *et al.*, 2017). Subsequently, major  
76 movements from the Eurasian Steppe, beginning about 3,000 BCE, resulted in further admixture  
77 throughout Europe (ALLENTOFT *et al.*, 2015; HAAK *et al.*, 2015; OLALDE *et al.*, 2018, 2019).  
78 These events are typically modeled in terms of three ancestry components, hunter-gatherers  
79 (and more specifically western hunter gatherers, “WHG”), early European farmers (“EEF”),  
80 and Steppe pastoralists (“Steppe”). Within this broad framework, the island of Sardinia is  
81 thought to have received a high level of EEF ancestry early on in its history and subsequently  
82 remained relatively isolated from the admixture occurring on mainland Europe (KELLER *et al.*,  
83 2012; SIKORA *et al.*, 2014). However, this specific model for Sardinian population history has  
84 not been tested with genome-wide aDNA samples from the island.

85 The oldest known human remains on Sardinia have been dated to be ~ 20,000 years old  
86 (MELIS, 2002), implying that humans first reached the island during the Paleolithic Age. Ar-  
87 chaeological evidence suggests that the island was not densely populated in the Mesolithic, with  
88 only irregular and episodic settlements, mostly concentrated near the coast (LUGLIÈ, 2018). The  
89 archaeological record shows that a population expansion coincided with a Neolithic transition in  
90 the sixth millennium BCE (FRANCALACCI *et al.*, 2013). At the same time, the early Neolithic  
91 “Cardial Impressed Ware” culture was spreading across the western Mediterranean (BARNETT,  
92 2000), with radio-carbon dates indicating a rapid maritime expansion about 5,500 BCE (ZILHÃO,  
93 2001; MARTINS *et al.*, 2015). Obsidian originating from Sardinia is found throughout many  
94 western Mediterranean archaeological sites associated with the middle Neolithic (TYKOT, 1996),  
95 indicating that the island was integrated into a maritime trade network. In the middle Bronze  
96 Age, about 1,600 BCE, the “Nuragic” culture emerged that derives its name from thousands of  
97 distinctive stone towers, Nuraghi, constructed across Sardinia’s landscape and in many instances  
98 still well preserved. More recently, the archaeological and historical record shows the influence  
99 of several major Mediterranean groups, such as Phoenicians, Carthaginians, the Roman and  
100 Byzantine empires, and later with North Africa, Tuscany, Genoa, Catalonia, Spain, Southern  
101 France, and Piedmont (ORTU, 2011; MASTINO, 2005).

102 The population genetics of Sardinia has long been studied (e.g. see CALÒ *et al.*, 2008) in  
103 part because of its importance as a population for medical genetics (LETTRE and HIRSCHHORN,  
104 2015). Pioneering studies, using classical genetic loci such as G6PD, HBB, and HLA and later  
105 maps of linkage disequilibrium, revealed that Sardinia is a genetic isolate with heterogeneous  
106 population sub-structure (e.g. SINISCALCO *et al.*, 1966; CONTU *et al.*, 1992; BARBUJANI and  
107 SOKAL, 1990; EAVES *et al.*, 2000; ZAVATTARI *et al.*, 2000; CAVALLI-SFORZA, 2005; SIDORE *et al.*,  
108 2015). Recently, CHIANG *et al.* (2018) analyzed the whole genome sequences of 3,514 individuals

109 from Sardinia to investigate the population genomic history of the island in finer resolution. In  
110 line with previous studies, they found substructure in which the mountainous Ogliastra region  
111 of central/eastern Sardinia carries a signature of relative isolation, presumably due to restricted  
112 gene flow across the rugged terrain. They also used a small sample of continental European  
113 aDNA to show suggestive evidence for differential contributions of ancestry from WHG, EEF,  
114 and Steppe to Sardinian genetic variation. This initial observation and the increased resolution  
115 of temporal, geographic, and cultural sampling in aDNA prompted us to investigate aDNA from  
116 Sardinia to gain further understanding.

117 Four previous studies have analyzed aDNA to provide an initial view of the genetics of pre-  
118 historic Sardinia, in each case, using mitochondrial DNA. [GHIROTTI \*et al.\* \(2009\)](#) contrasted  
119 patterns of continuity between Ogliastra (the mountainous and historically isolated central re-  
120 gion) and Gallura (a region in northern Sardinia with cultural and linguistic connections to  
121 Corsica), finding evidence for more genetic turnover in Gallura. [MODI \*et al.\* \(2017\)](#) provided  
122 the first complete mitogenomes of two Mesolithic individuals and found support for a model in  
123 which Mesolithic ancestry on the island was replaced by incoming populations in the Neolithic.  
124 [OLIVIERI \*et al.\* \(2017\)](#), in a companion project to the work described here, analyzed 21 an-  
125 cient mitogenomes from Sardinia as well as 3,491 mitogenomes from contemporary Sardinians  
126 and estimated the coalescent times of Sardinian-specific mtDNA haplogroups finding support for  
127 most of them originating in the Neolithic or later, but with a few coalescing earlier. Finally,  
128 [MATISOO-SMITH \*et al.\* \(2018\)](#) analyzed mitogenomes in a Phoenician colony on Sardinia and  
129 found evidence of continuity and exchange between the colony and broader Sardinia. Despite  
130 the initial insights these studies reveal, none of them analyze genome-wide autosomal data, which  
131 has proven to be of great use for studies of population history ([PICKRELL and REICH, 2014](#)).

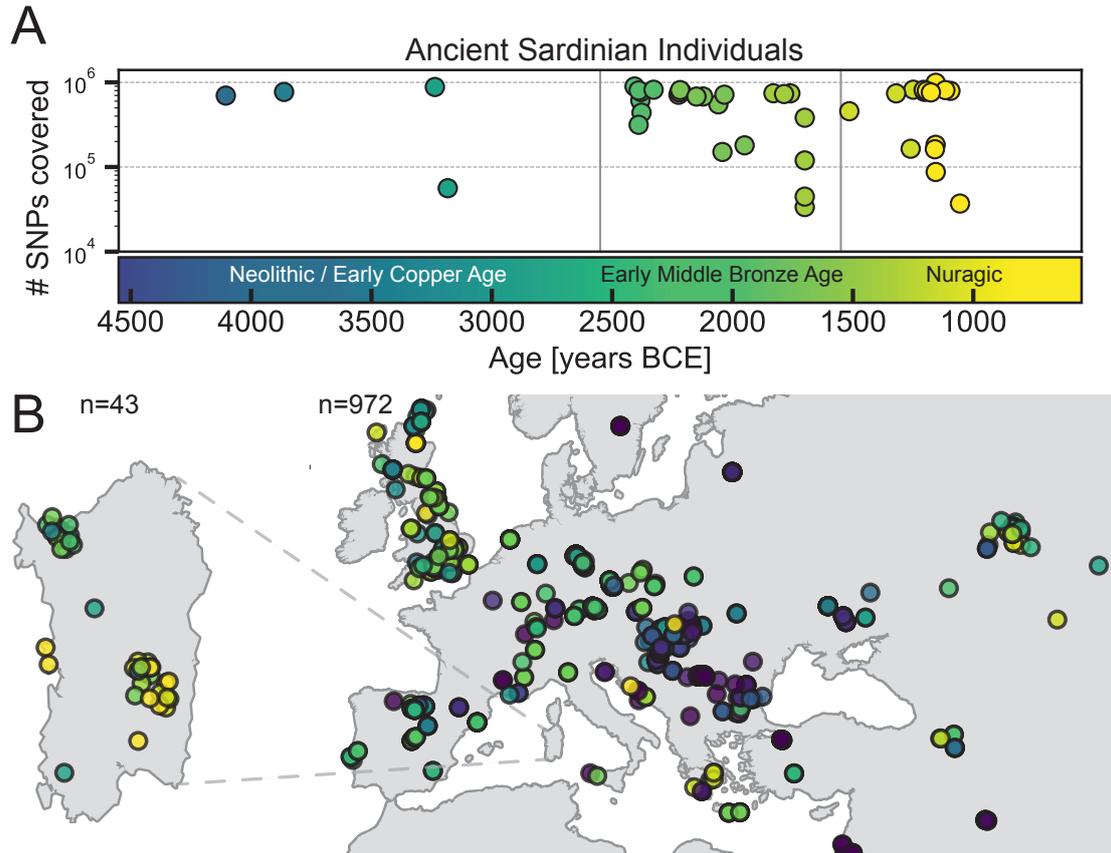
132 To provide a more detailed perspective on Sardinian population history, we generated genome-  
133 wide data from the skeletal remains of 43 Sardinian individuals radiocarbon dated to between  
134 4,100-1,000 BCE. We analyzed their genetic variation in the context of reference panels of ancient  
135 and contemporary individuals. Our goal was to investigate three aspects of Sardinian popula-  
136 tion history: First, the ancestry of Neolithic Sardinian individuals (ca. 5,700-3,400 BCE) – who  
137 were the early peoples expanding onto the island at this time? Second, the genetic structure  
138 through the Sardinian Chalcolithic (i.e. Copper Age, ca. 3,400-2,300 BCE) to the Bronze Age  
139 (ca. 2,300-1,000 BCE) – were there genetic turnover events through the different cultural tran-  
140 sitions observed in the archaeological record? And third, the post-Bronze Age contacts with  
141 major Mediterranean civilizations and more recent Italian populations – have they resulted in  
142 detectable gene flow? Our results revealed insights about each of these three periods of Sardinian  
143 history.

## 144 Results

### 145 Ancient DNA from Sardinia

146 We organized a collection of skeletal remains from: 1) previously excavated samples throughout  
147 Sardinia, in part drawing from samples initially used for isotopic analysis in [LAI \*et al.\* \(2013,](#)  
148 [Supp. Info. 1\)](#), and 2) the Seulo caves of central Sardinia ([SKEATES \*et al.\*, 2013,](#) [Supp. Info. 2\)](#).  
149 We generated and sequenced DNA libraries enriched for reads overlapping the complete mito-  
150 chondrial genome as well as a targeted set of 1.2 million single nucleotide polymorphisms (SNPs)  
151 ([FU \*et al.\*, 2015](#); [HAAK \*et al.\*, 2015](#); [MATHIESON \*et al.\*, 2015](#)). After applying several standard  
152 ancient DNA quality control filters, we arrived at a final set of 43 individuals with an average  
153 coverage of  $1.31\times$  (ranging from  $0.04\times$  to  $5.39\times$  per individual) and a median number of 715,737  
154 sites covered at least once per individual. We obtained age estimates for each individual by either

155 direct radiocarbon dating ( $n = 29$ ), using previously reported radiocarbon dates ( $n = 10$ ), or  
156 using a combination of archaeological context and radiocarbon dates from the same burial site  
157 ( $n = 4$ , Fig. 1). The estimated ages in our sample range from 4,100 years BCE to 1,000 years  
158 BCE (Fig. 1, Supp. Mat. 1A). To facilitate analyses of temporal structure within ancient Sar-  
159 dinia, we pragmatically grouped the data into three broad periods: Neolithic and Early Copper  
160 Age ('Sar-NECA', 4,100-3,000 BCE,  $n = 4$ ), Early Middle Bronze Age ('Sar-EMBA', 2,500-1,500  
161 BCE,  $n = 24$ ) and Nuragic ('Sar-Nur', 1,500-1,000 BCE,  $n = 15$ ). Figure 1 provides an overview  
162 of the sample.



**Figure 1: Average depth, sampling locations and ages of ancient individuals.** A: The number of SNPs covered at least once and age (mean of  $2\sigma$  radio-carbon age estimates) for the 43 ancient Sardinian individuals. B: The sampling locations of ancient Sardinian individuals and a reference dataset of 972 ancient individuals collected across western Eurasia, spanning a broad temporal period.

### 163 Uniparentally inherited markers

164 We were able to infer mitochondrial haplogroups for all 43 ancient Sardinian individuals (Supp. Mat. 1E),  
165 including a subset ( $n = 10$ ) previously reported by OLIVIERI *et al.* (2017). We confirm the obser-  
166 vation that ancient Sardinian mtDNA haplotypes belong almost exclusively to macro-haplogroups  
167 HV ( $n = 16$ ), JT ( $n = 17$ ) and U ( $n = 9$ ), a composition broadly similar to other European

168 Neolithic populations.

169 Our genome-wide data allowed us to assign Y haplogroups for 25 ancient Sardinian individ-  
170 uals. More than half of them consist of R1b-V88 ( $n = 10$ ) or I2-M223 ( $n = 7$ ) (Sup. Fig. 3,  
171 Supp. Mat. 1B). In our reference data set, these two Y-haplogroups appear first in Balkan  
172 hunter-gatherer and Balkan Neolithic individuals, and also in more recent western Neolithic  
173 populations. [FRANCALACCI \*et al.\* \(2013\)](#) identified three major Sardinia-specific founder clades  
174 based on present-day variation within the haplogroups I2-M26, G2-L91 and R1b-V88, and here  
175 we found each of those broader haplogroups in at least one ancient Sardinian individual. Two  
176 major present-day Sardinian haplogroups, R1b-M269 and E-M215, are absent (Sup. Fig. 3).  
177 Compared to other Neolithic and present-day European populations, the number of identified  
178 R1b-V88 carriers is relatively high (Supp. Info 4, Supp. Fig. 4). However, we observed cluster-  
179 ing of haplogroups by sample location, consistent with substructure (Supp. Mat. 1B); therefore  
180 some caution should be exercised with interpreting our results as estimates for island-wide Y  
181 haplogroup frequencies (see Supp. Mat. 1C).

## 182 Results from genome-wide aDNA

183 We then assessed the relationship of the ancient Sardinian individuals to other ancient and  
184 present-day west Eurasian populations using autosomal DNA data. For this purpose, we used:  
185 1) a subset of the Human Origins array dataset from contemporary human individuals ( $n =$   
186 1,963, [LAZARIDIS \*et al.\*, 2014](#)), 2) an extensive panel of genomic capture data from previously  
187 published ancient individuals ( $n = 972$ , [MATHIESON \*et al.\*, 2015](#); [LAZARIDIS \*et al.\*, 2016, 2017](#);  
188 [MATHIESON \*et al.\*, 2018, 2017](#); [LIPSON \*et al.\*, 2017](#); [OLALDE \*et al.\*, 2018](#)), and 3) a large sample  
189 of contemporary Sardinian individuals from our previous studies ( $n = 1,577$ , [SIDORE \*et al.\*,](#)  
190 [2015](#); [CHIANG \*et al.\*, 2018](#)). For some analyses, we grouped these individuals into those from the  
191 more isolated Sardinian province of Ogliastra ('Sar-Ogl',  $n = 419$ ) and the remainder ('Sar-non  
192 Ogl',  $n = 1,158$ ). For other analyses, we subset Sardinia into 8 geographic regions (see inset in  
193 panel C of Figure 2 for listing and abbreviations, also see Supp. Mat. 1G). Unless otherwise  
194 specified (see Materials and Methods), we refer to particular samples of individuals using the  
195 group labels used in the datasets they were derived from. As with other human genetic variation  
196 studies, consideration of the population annotations is important to consider in the interpretation  
197 of results.

## 198 Similarity to western mainland Neolithic populations

199 Importantly, we found low levels of differentiation between Neolithic Sardinian individuals and  
200 several Neolithic western mainland European populations, in particular, Cardial Ware-associated  
201 groups from Spain (Iberia-EN) and southern France (France-N). When projecting ancient indi-  
202 viduals onto the top two principal components (PCs) defined by modern variation, the Neolithic  
203 (and also later) ancient Sardinian individuals sit between early Neolithic Iberian and later Copper  
204 Age Iberian populations, roughly on an axis that differentiates WHG and EEF populations and  
205 embedded in a cluster that additionally includes Neolithic British individuals (Fig. 2). This result  
206 is also evident in terms of absolute genetic differentiation, with low pairwise  $F_{ST} \approx 0.005 \pm 0.002$ ,  
207 Fig. 3) between Neolithic Sardinian individuals and Neolithic western mainland European popu-  
208 lations. Pairwise outgroup- $f_3$  analysis shows a very similar pattern, with the highest values of  $f_3$   
209 (i.e. most shared drift) being with Neolithic and Copper Age Iberia (Fig. 3), gradually dropping  
210 off for temporally and geographically distant populations.

211 In explicit admixture models (using qpAdm, see Methods) the southern French Neolithic  
212 individuals (France-N) are the most consistent with being a single source for Neolithic Sardinia

213 ( $p \approx 0.074$  to reject the model of one population being the direct source of the other); followed by  
214 other populations associated with the western Mediterranean Neolithic Cardial Ware expansion  
215 (Supp. Tab. 3). As we discuss below, caution is necessary for interpreting the result as there is  
216 a lack of aDNA from other relevant populations of the same period (such as neighboring islands  
217 and mainland Italian Cardial Ware cultures).

## 218 **Constancy of Western Hunter Gatherer ancestry**

219 Similar to western European Neolithic and central European Late Neolithic populations, ancient  
220 Sardinian individuals are shifted towards WHG individuals in the top two PCs relative to early  
221 Neolithic Anatolians (Fig. 2). Admixture analysis using qpAdm infers that ancient Sardinian  
222 individuals harbour HG ancestry ( $\approx 17\%$ ) that is higher than early Neolithic mainland popu-  
223 lations (including Iberia,  $\approx 8\%$ ), but lower than Copper Age Iberians ( $\approx 25\%$ ) and about the  
224 same as Southern French Middle-Neolithic individuals ( $\approx 21\%$ ) (Tab. 1, Supp. Fig. 9). A null  
225 model of a two-way admixture between WHG and Neolithic Anatolian populations is inferred to  
226 be consistent with our data ( $p \approx 0.22$ , Tab. 1). This  $p$ -value, describing the power to reject the  
227 null model of two-way admixture, is similar to the value observed for other western European  
228 populations of the early Neolithic (Supp. Tab. 2).

## 229 **Continuity from the Sardinian Neolithic through the Nuragic**

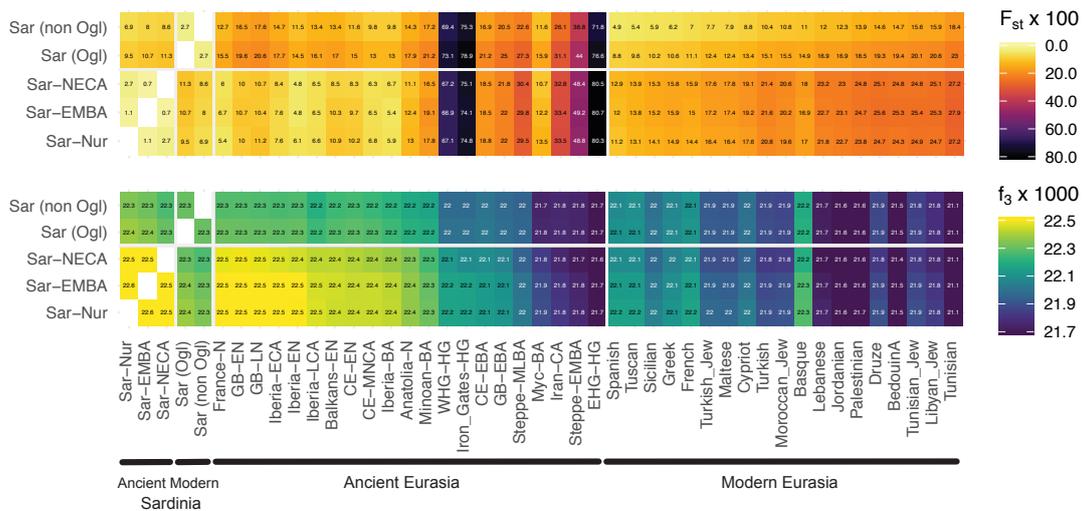
230 We found several lines of evidence supporting genetic continuity from the Sardinian Neolithic  
231 into the Bronze Age and Nuragic times. Importantly, we observed low genetic differentiation  
232 between ancient Sardinian individuals from various time periods. We estimated  $F_{ST}$  to be  
233  $0.0027 \pm 0.0014$  between Neolithic and late Bronze Age (mostly Nuragic) Sardinian individuals  
234 (Fig. 3). Furthermore, we did not observe temporal substructure within the ancient Sardinian  
235 individuals in the top two PCs – they form a coherent cluster (Fig. 2). In stark contrast,  
236 ancient individuals from many mainland geographic regions, such as central Europe, show larger  
237 movements over the first two PCs from the Late Neolithic to the Bronze Age, and also have  
238 higher pairwise differentiation ( $F_{ST} = 0.0194 \pm 0.0003$ ).

239 In the presence of significant influx, differential genetic affinity of a test population  $x$  would  
240 cause  $f_4$  statistics of the form  $f(\text{Sard Period 1} - \text{Sard Period 2}; \text{Pop } x - \text{Ancestral Allele})$   
241 to be non-zero (where “Ancestral Allele” is an inferred ancestral allelic state from a multi-species  
242 alignment). However, we observe that no such statistic differs significantly from zero for all test  
243 populations  $x$  (Supp. Mat. 2D). A qpAdm analysis, which is based on simultaneously testing  
244  $f$ -statistics with a number of outgroups and adjusts for correlations, cannot reject a model of  
245 Neolithic Sardinian individuals being a direct predecessor of Nuragic Sardinian individuals either  
246 ( $p = 0.54$ , Supp. Tab. 3). Our qpAdm analysis further shows that the WHG ancestry proportion,  
247 in a model of admixture with Neolithic Anatolia, remains stable at  $\sim 17\%$  throughout three  
248 ancient time-periods (Tab. 1A). When using a three-way admixture model, we do not detect  
249 significant Steppe ancestry in any ancient Sardinian individual, as is inferred, for example, in  
250 later Bronze Age Iberians (Tab. 1B, Supp. Fig. 9).

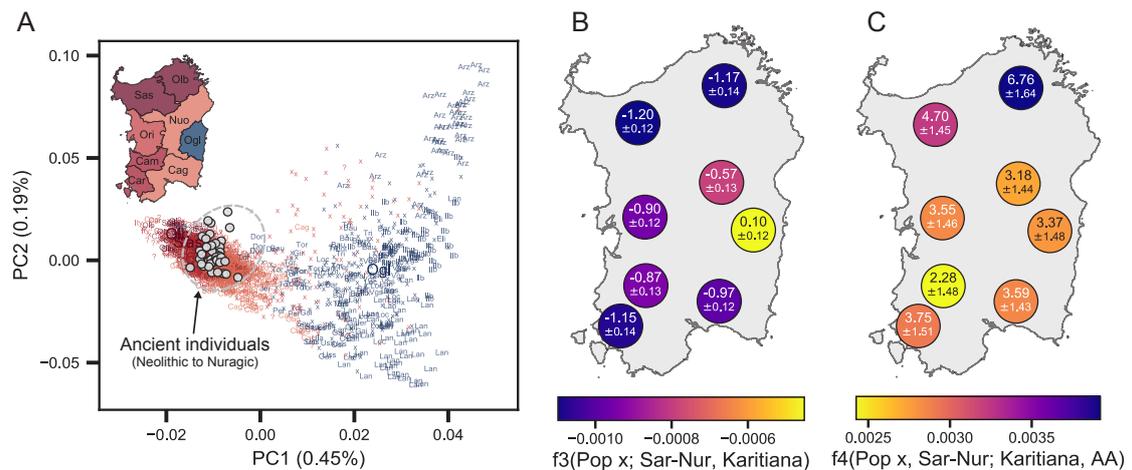
## 251 **From the Nuragic to present-day Sardinia**

252 Our results demonstrate that ancient Sardinian individuals are genetically closest to contem-  
253 porary Sardinian individuals among all the ancient individuals analyzed (Fig. 3), and relative  
254 to other European populations, there is lower differentiation between present-day and ancient  
255 individuals (Supp. Fig. 7). However, we also find multiple lines of evidence for appreciable gene  
256 flow into Sardinia after the Nuragic period.





**Figure 3:** Genetic similarity matrices. We calculated  $F_{ST}$  (upper panel) and outgroup- $f_3$  (lower panel) of ancient Sardinian and modern Sardinian individuals (grouped into within and outside the Ogliastra region) with each other (left), various ancient (middle), and modern populations (right) of interest.



**Figure 4:** Present-day genetic structure in Sardinia reanalyzed with aDNA. A: Scatter plot of the first two principal components trained on 1577 present-day individuals with grand-parental ancestry from Sardinia. Each individual is labeled with a location if at least 3 of the 4 grand-parents were born in the same geographical location (“small” three letter abbreviations); otherwise with “x” or if grand-parental ancestry is missing with “?”. We calculated median PC values for each Sardinian province (large abbreviations). We also projected each ancient Sardinian individual on to the top two PCs (gray points). B/C: We plot  $f$ -statistics that test for admixture of modern Sardinian individuals (grouped into provinces) when using Nuragic Sardinian individuals as one source population. Uncertainty ranges depict one standard error (calculated from block bootstrap). Karitiana are used in the  $f$ -statistic calculation as a proxy for ANE/Steppe ancestry (PATTERSON *et al.*, 2012).

257 Firstly, present-day Sardinian individuals are shifted from the ancients towards more eastern  
258 Mediterranean populations on the western Eurasian PCA (Fig. 2). We observe a corresponding  
259 signal in our  $f_4$  analysis, in that we see significantly higher affinity of many present-day and some  
260 ancient populations to modern Sardinian versus Nuragic Sardinian individuals ( $f_4$  of the form  
261  $f(\text{Mod Sard} - \text{Ancient Sard}; \text{Pop } x - \text{Ancestral Allele})$ , see Fig. 4 and Supp. Mat. 2D). Similarly,  
262  $f_3$  statistics that directly test for admixture of present-day Sardinians, with Nuragic Sardinian  
263 individuals as one source, yield highly significant negative values, indicating admixture (Fig. 4).  
264 Using qpAdm we find that models of continuity from Nuragic Sardinia to present-day Sardinian  
265 populations (e.g. Cagliari) without influx are rejected ( $p < 10^{-40}$ , Tab. 1C). Moreover, genetic  
266 differentiation between the Nuragic and present is higher than across ancient periods (between  
267 Nuragic and present-day non-Ogliastra individuals pairwise  $F_{ST} = 0.00695 \pm 0.00041$ ; compared  
268 to  $F_{ST} = 0.0027 \pm 0.0014$  between Late Neolithic and Nuragic individuals.)

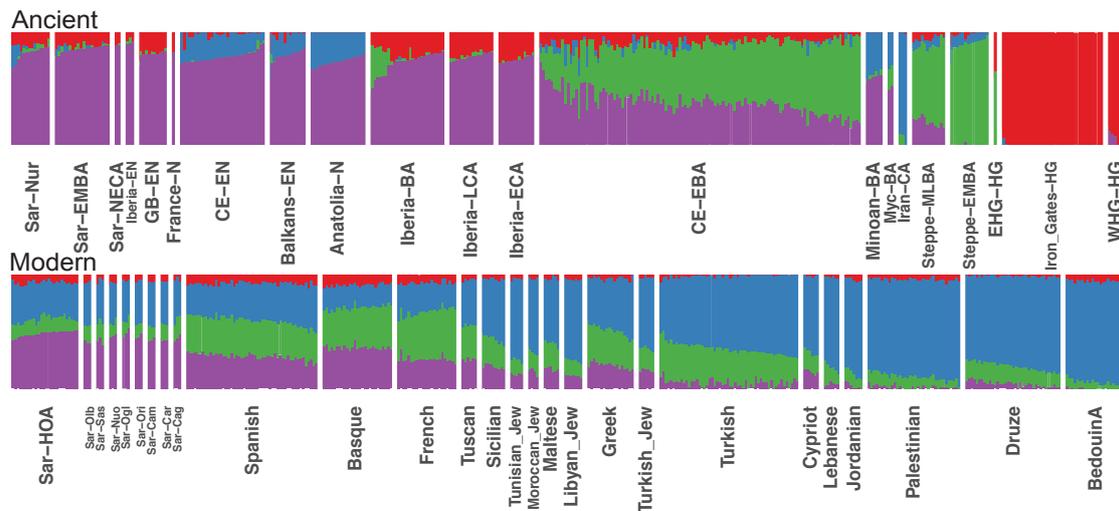
269 Second, we find many populations that can produce significant  $f_4$  and  $f_3$  statistics consistent  
270 with admixture (Supp. Mat. 2C and D). Many of these populations carry high levels of Ancestral  
271 North Eurasian (ANE) ancestry, and likely serve as a proxy for ancient Eurasian ancestry  
272 that entered Europe after the Neolithic with Steppe expansions, as similarly observed for many  
273 present-day mainland Europeans (PATTERSON *et al.*, 2012).

274 ADMIXTURE analysis gives further insight into this signal of gene flow. While contemporary  
275 Sardinian individuals show the highest affinity towards EEF-associated populations among all of  
276 the modern populations, they also display membership with other clusters (Fig. 5). In contrast  
277 to ancient Sardinian individuals, present-day Sardinian individuals carry a modest “Steppe-like”  
278 ancestry component (but generally less than continental present-day European populations),  
279 and an appreciable broadly “eastern Mediterranean” ancestry component (also inferred at a  
280 high fraction in other present-day Mediterranean populations, such as Sicily and Greece).

281 To further characterize signatures of admixture, we used qpAdm to test the fit of a model  
282 of present-day Sardinian populations as a simple two-way admixture between Nuragic Sardinian  
283 individuals and potential other source populations (Table 1D, Supp. Tab. 4). A model of admix-  
284 ture with modern Sicilians ( $p = 0.031$ ) had the best support, followed by Maltese ( $p = 0.0128$ ),  
285 Turkish ( $p = 0.0086$ ) and Greeks ( $p = 0.00071$ ). For the model of a mixture of Sicilians and  
286 ancient Sardinian individuals, we infer an admixture proportion of  $43.5 \pm 2.1$  percent Sicilian  
287 admixture (Tab. 1, Sup. Tab. 4, Supp. Fig. 10).

288 We also considered three-way models of admixture with qpAdm to further refine the geo-  
289 graphic origins of this recent admixture signal (Supp. Info 6). Indeed, we find models with ad-  
290 mixture between Nuragic Sardinia, one northern Mediterranean source and one eastern Mediter-  
291 ranean source fit well ( $p > 0.01$  for several combinations, Table 1E,F). For a representative  
292 sample from Sardinia (Cagliari), across various proxies (excluding Sicily and Malta) the ad-  
293 mixture fractions range 10-30% for the “northern Mediterranean” component, 13-33% for the  
294 “eastern Mediterranean”, with the remaining 52-57% coming from Nuragic Sardinia. For models  
295 with Sicilian or Maltese as proxy sources, the estimates of ancestry for the N. Mediterranean  
296 component shrink to small values (8.6% and 6.5% for Sicilian and Maltese, respectively), essen-  
297 tially bringing the fitted parameters towards the two-way mixture models. Maltese and Sicilian  
298 individuals appear to reflect a mixture of N. Mediterranean and E. Mediterranean ancestries,  
299 and as such they can serve as single-source proxies in two-way admixture models with Nuragic  
300 Sardinia (Table 1D).

301 Caution is warranted when interpreting inferred admixture fractions with each of these simple  
302 models; however, the signal across multiple analyses indicates that complex post-Nuragic gene  
303 flow, partly from sources originating in the eastern Mediterranean and partly from the northern  
304 Mediterranean, has likely played a role in the population genetic history of Sardinia.



**Figure 5:** Admixture coefficients estimated by ADMIXTURE ( $K = 4$ ). Each stacked bar represents one individual and color fractions depict the fraction of the given individual’s ancestry coming from a given “cluster”. For  $K = 4$  (depicted here), ancient Sardinian individuals share similar admixture proportions as other western European Neolithic individuals. Present-day Sardinian individuals additionally have elevated Steppe-like ancestry (but less than other European populations), and an additional ancestry component prevalent in Near Eastern / Levant populations. ADMIXTURE results for all  $K=2, \dots, 11$  are depicted in the supplement (Supp. Fig. 14)

## 305 Fine-scale structure in contemporary Sardinia

306 Ancient DNA can shed new light on present-day genetic variation. We, therefore, re-assessed  
 307 spatial substructure previously observed in a dense geographic modern sampling (1,577 whole  
 308 genome sequences) from Sardinia (CHIANG *et al.*, 2018).

309 In a PCA of the modern Sardinian variation, individuals from Ogliastra fall furthest away  
 310 from the ancient Sardinian individuals (CHIANG *et al.*, 2018) (Fig. 4). In stark contrast, in the  
 311 PCA of modern Western Eurasian variation, the pattern reverses: Ogliastra is placed closest  
 312 of all provinces to the ancient Sardinian individuals (Fig. 2). Direct tests for admixture using  
 313  $f_3$  statistics with Nuragic Sardinian individuals as one source yielded highly significant results  
 314 for all present-day provinces except Ogliastra (Fig. 4). The non-significant value of Ogliastra  
 315 can have two causes: An actual lack of admixture or high levels of drift that mask admixture  
 316  $f_3$ . However, the  $f_4$  statistics and admixture proportions of qpAdm are robust to recent drift  
 317 of the admixed population, and in both analyses Ogliastra shows an admixture signal that is  
 318 only slightly weaker than most other provinces (Fig. 4, Supp. Fig. 12). Together, these results  
 319 suggest high levels of drift specific to Ogliastra (likely also driving the first two PCs of present-day  
 320 Sardinian variation), but simultaneously also less admixture than other Sardinian provinces.

321 In the previous section, we reported finding that many non-Sardinian modern populations  
 322 have a higher affinity to present-day Sardinian individuals than to Nuragic Sardinian individuals  
 323 (using a  $f_4$  statistic of the form  $f_4(\text{Mod. Sard Pop } y - \text{Sar-Nur; Pop } x - \text{Ancestral Allele})$  where  
 324  $x$  are test non-Sardinian modern populations, Fig. 4, Sup. Mat. 2D). Interestingly, the northern  
 325 province Olbia (north-east) and to some degree also Sassari (north-west) have the highest affinity  
 326 to most tested populations (Fig. 4). A three-way admixture model fit with qpAdm finds a similar  
 327 signal. In a model with Tuscan as a proxy for northern Mediterranean immigration and Lebanese

328 as a proxy for a second additional, more eastern Mediterranean source, the inferred admixture  
329 fractions vary across Sardinia, with the highest eastern Mediterranean ancestry in the southwest  
330 (Carbonia, Campidano) and the highest northern Mediterranean ancestry in the northeast of  
331 the island (Olbia, Sassari, Supp. Fig. 12). In addition, we observed a marked shift of individuals  
332 from Olbia and Sassari towards continental populations in the PCA (Fig. 4).

Target	Proxy Source Populations				p-value	Admixture Fractions			Standard Error		
	a	b	c			a	b	c	a	b	c
A	Sar-NECA	WHG	Anatolia-N	-	<b>0.677</b>	0.173	0.827	-	0.014	0.014	-
	Sar-EMBA	WHG	Anatolia-N	-	<b>0.062</b>	0.163	0.837	-	0.007	0.007	-
	Sar-Nur	WHG	Anatolia-N	-	<b>0.182</b>	0.166	0.834	-	0.009	0.009	-
B	Sar-NECA	WHG	Anatolia-N	Steppe	<b>0.588</b>	0.172	0.824	0.003	0.016	0.023	0.026
	Sar-EMBA	WHG	Anatolia-N	Steppe	<b>0.041</b>	0.163	0.837	0.000	0.009	0.012	0.014
	Sar-Nur	WHG	Anatolia-N	Steppe	<b>0.123</b>	0.167	0.833	0.000	0.010	0.014	0.016
	Iberia-EN	WHG	Anatolia-N	Steppe	<b>0.276</b>	0.081	0.919	0.000	0.013	0.017	0.019
	Iberia-BA	WHG	Anatolia-N	Steppe	$6.0 \cdot 10^{-3}$	0.239	0.689	0.072	0.010	0.014	0.016
	CE-EN	WHG	Anatolia-N	Steppe	<b>0.663</b>	0.046	0.954	0.000	0.007	0.010	0.012
	CE-LBA	WHG	Anatolia-N	Steppe	<b>0.079</b>	0.128	0.403	0.469	0.008	0.011	0.013
C	Sar-EMBA	Sar-NECA	-	-	<b>0.548</b>	-	-	-	-	-	-
	Sar-Nur	Sar-EMBA	-	-	<b>0.385</b>	-	-	-	-	-	-
	Cagliari	Sar-Nur	-	-	$3.2 \cdot 10^{-37}$	-	-	-	-	-	-
D	Cagliari	Sar-Nur	Sicilian	-	<b>0.031</b>	0.565	0.435	-	0.021	0.021	-
	Cagliari	Sar-Nur	Maltese	-	<b>0.013</b>	0.590	0.410	-	0.022	0.022	-
	Cagliari	Sar-Nur	Tuscan	-	$2.8 \cdot 10^{-6}$	0.540	0.460	-	0.026	0.026	-
	Cagliari	Sar-Nur	Lebanese	-	$1.2 \cdot 10^{-6}$	0.724	0.276	-	0.016	0.016	-
	Cagliari	Sar-Nur	Spanish	-	$8.0 \cdot 10^{-26}$	0.680	0.320	-	0.027	0.027	-
E	Cagliari	Sar-Nur	N Mediterranean	Turkish-Jew	<b>0.186</b>	0.529	0.212	0.259	0.023	0.043	0.035
	Cagliari	Sar-Nur	N Mediterranean	Libyan-Jew	<b>0.086</b>	0.532	0.272	0.196	0.024	0.039	0.028
	Cagliari	Sar-Nur	N Mediterranean	Maltese	<b>0.07</b>	0.562	0.086	0.351	0.027	0.074	0.060
	Cagliari	Sar-Nur	N Mediterranean	Tunisian-Jew	<b>0.064</b>	0.515	0.285	0.200	0.023	0.037	0.029
	Cagliari	Sar-Nur	N Mediterranean	Sicilian	<b>0.06</b>	0.546	0.065	0.389	0.024	0.068	0.060
	Cagliari	Sar-Nur	N Mediterranean	Moroccan-Jew	<b>0.05</b>	0.536	0.247	0.217	0.024	0.044	0.033
	Cagliari	Sar-Nur	N Mediterranean	Lebanese	<b>0.049</b>	0.571	0.269	0.160	0.026	0.040	0.023
	Cagliari	Sar-Nur	N Mediterranean	Druze	<b>0.04</b>	0.562	0.260	0.178	0.024	0.040	0.025
	Cagliari	Sar-Nur	N Mediterranean	Cypriot	<b>0.024</b>	0.538	0.270	0.192	0.023	0.037	0.027
	Cagliari	Sar-Nur	N Mediterranean	Jordanian	<b>0.018</b>	0.563	0.291	0.146	0.025	0.037	0.022
	Cagliari	Sar-Nur	N Mediterranean	Palestinian	<b>0.014</b>	0.563	0.301	0.135	0.026	0.037	0.020
	Cagliari	Sar-Nur	N Mediterranean	Turkish	$8.6 \cdot 10^{-3}$	0.668	0.089	0.243	0.037	0.074	0.042
	Cagliari	Sar-Nur	N Mediterranean	Tunisian	$1.0 \cdot 10^{-4}$	0.547	0.375	0.079	0.028	0.034	0.014
F	Cagliari	Sar-Nur	E Mediterranean	Lombardy	<b>0.186</b>	0.529	0.259	0.212	0.023	0.035	0.043
	Cagliari	Sar-Nur	E Mediterranean	Greek	<b>0.153</b>	0.560	0.165	0.274	0.020	0.054	0.057
	Cagliari	Sar-Nur	E Mediterranean	Tuscan	<b>0.115</b>	0.544	0.215	0.241	0.022	0.048	0.056
	Cagliari	Sar-Nur	E Mediterranean	French	<b>0.094</b>	0.573	0.321	0.105	0.021	0.028	0.024
	Cagliari	Sar-Nur	E Mediterranean	Basque	<b>0.066</b>	0.548	0.337	0.115	0.023	0.026	0.026
	Cagliari	Sar-Nur	E Mediterranean	Spanish	<b>0.052</b>	0.555	0.309	0.137	0.022	0.030	0.032

**Table 1: Results from fitting models of admixture with qpAdm.** A) Two-way models of admixture for ancient Sardinia using Western Hunter-Gatherer (WHG) and Neolithic Anatolia (Anatolia-N) individuals as proxy sources. B) Three-way models of admixture for ancient Sardinia using Western Hunter-Gatherer (WHG), Neolithic Anatolia (Anatolia-N), and Early Middle Bronze Age Steppe (Steppe-EMBA, abbreviated Steppe in table), individuals as proxy sources. C) Single-source models to assess continuity of each Sardinian period with the previous one (see main text for guide to abbreviations). E) Results of three-way models showing multiple eastern Mediterranean populations that can produce viable models (Results shown with individuals from Lombardy [Bergamo in the HOA dataset] as one of several possible proxies for north Mediterranean ancestry, see part F). F) Results of three-way models showing multiple north Mediterranean populations that can produce viable models (Results shown with Jewish individuals from Turkey [‘Turkish-Jew’ in the HOA dataset] used as one of several possible proxies for east Mediterranean ancestry, see part E). As a visual aid, p-values greater than 0.01 are bolded. Full results are reported in Supp. Info. 6.

## 333 Discussion

334 Our analysis of genome-wide data from 43 ancient Sardinian individuals generated new direct  
335 evidence regarding the population history of Sardinia and the Mediterranean. Importantly,  
336 we detected a strong genetic affinity of Neolithic Sardinian individuals to other early Neolithic  
337 western Mediterranean populations. This signal is especially interesting in light of archaeological  
338 evidence for a rapid maritime spread of Neolithic Cardial Impressed Ware culture through the  
339 western Mediterranean occurring around 5,500 BCE (ZILHÃO, 2001; MARTINS *et al.*, 2015).  
340 While we lack aDNA data from Neolithic mainland Italy, a putative center of this spread to  
341 Sardinia, we find that Neolithic Sardinian individuals are closely related to other hypothesized  
342 populations of this initial wave, in particular, Neolithic Spanish and southern French populations.  
343 This new evidence points towards Neolithic Sardinian individuals descending principally from  
344 mainland Neolithic populations. This hypothesis is also consistent with a signal of population  
345 turnover associated with the Neolithic transition observed in Sardinian ancient mtDNA (MODI  
346 *et al.*, 2017), and with a gap in the Sardinian archaeological record before its Neolithic transition  
347 (LUGLIÈ, 2018).

348 Neolithic Sardinian individuals fit well as a two-way admixture between EEF and WHG  
349 sources, similar to other EEF populations including Linear Pottery and Cardial cultures. Recent  
350 evidence based on aDNA indicates that traces of WHG ancestry were already part of the initial  
351 wave of the Mediterranean Neolithic transition and that similar to other EEF populations, subse-  
352 quent local admixture increased WHG ancestry substantially over time in Iberia (LIPSON *et al.*,  
353 2017). In stark contrast, in Sardinia, we observed remarkable constancy of WHG ancestry close  
354 to 20% throughout our sampling periods, well into the second millennium BCE. This reflects the  
355 continuity of Sardinia through this time period and is consistent with a model of a low density  
356 or even absence of local Mesolithic hunter-gatherers at the time of arrival of EEF individuals  
357 (LUGLIÈ, 2018). However, we can not rule out a model with an initial pulse of local admixture.  
358 Genome-wide data from a Mesolithic or a very early Neolithic individual from Sardinia could  
359 help settle this question.

360 Additional insight into the origins of Neolithic populations of Sardinia comes from Y chro-  
361 mosome variation in the ancient samples. We detected Y haplogroups R1b-V88 and I2-M223 in  
362 the majority of the ancient Sardinian males. In our reference dataset, both haplogroups appear  
363 earliest in Mesolithic hunter-gatherers and then Neolithic groups of the Balkans (MATHIESON  
364 *et al.*, 2018) and also EEF Iberians, but not in Neolithic Anatolians or more western WHG  
365 individuals. Further sampling is necessary, but the current data are consistent with hypotheses  
366 of the expansion of Cardial Impressed Ware related cultures through the Mediterranean via the  
367 Balkans. Future studies, including ancient DNA from early Neolithic sites in mainland Italy, will  
368 help to further resolve details of these putative migrations.

369 From the Neolithic onwards, Sardinia appears to have been relatively isolated until at least  
370 the late second millennium BC, unlike many other parts of Europe which had experienced sub-  
371 stantial gene flow from central Eurasian Steppe ancestry starting about 3,000 years BCE (HAAK  
372 *et al.*, 2015; ALLENTOFT *et al.*, 2015). While we cannot exclude influx from genetically similar  
373 populations such as early Iberian Bell Beakers, the absence of Steppe ancestry suggests genetic  
374 isolation from many Bronze Age mainland populations - including later Iberian Bell Beakers,  
375 who would already have carried substantial Steppe ancestry (OLALDE *et al.*, 2018). As fur-  
376 ther support, the Y haplogroup R1b-M269, the most frequent present-day western European  
377 haplogroup and the haplogroup associated with expansions that brought Steppe ancestry into  
378 Britain (OLALDE *et al.*, 2018) and Iberia (OLALDE *et al.*, 2019) about 2,500-2,000 BCE, remains  
379 absent in our sample of ancient Sardinian individuals through the end of our sampling period  
380 (1,200-1,000 BCE).

381 The genetic continuity inferred throughout our ancient sampling period does not continue  
382 fully into the present. Previously, admixture tests based on  $f$ -statistics did not provide significant  
383 evidence for gene flow (CHIANG *et al.*, 2018), likely because no suitable proxy for the Nuragic  
384 Sardinian ancestry component was available. Here, including direct aDNA data increased the  
385 power of admixture tests, which resulted in uncovering multiple lines of evidence of moderate  
386 gene flow into Sardinia. This post-Nuragic admixture likely brought additional Y chromosome  
387 haplotype diversity to Sardinia (Sup. Fig. 3), such as R1b-M269 and also E-M215 (now prevalent  
388 in northern Africa).

389 We find evidence for at least two phases of post-Nuragic gene flow. First, there is a general  
390 shift towards central and eastern Mediterranean sources, demonstrated by the direction of the  
391 overall change in the PCA and ADMIXTURE, and the results of modeling population relation-  
392 ships using qpAdm. Second, we detected variation in the signals in the PCA and qpAdm analysis  
393 suggesting that the northern provinces of Olbia, and to a lesser degree Sassari, have received  
394 more northern Mediterranean immigration after the Bronze Age than the other provinces; mean-  
395 while the southwestern provinces of Campidano and Carbonia show more eastern Mediterranean  
396 ancestry. Together, these signals suggest temporally and geographically complex post-Nuragic  
397 gene flow into Sardinia. Ultimately, aDNA data from these historical periods will be needed to  
398 clarify and refine the interpretation.

399 A preliminary hypothesis would be that an influx from eastern Mediterranean sources is  
400 overlaid by more recent influx from the Italian mainland. Historically, both of these seem  
401 plausible. Sardinia hosted major Phoenician colonies in the first millennium BCE, principally  
402 along the south and west coasts of the island, and previous studies based on uni-parentally  
403 inherited markers have found evidence for Phoenician contact and gene flow (ZALLOUA *et al.*,  
404 2008; MATISOO-SMITH *et al.*, 2018). Sardinia was also an important Roman province and then  
405 was later under occupation by the Vandals and the Byzantine Empire. There are also more recent  
406 sources of immigration in the last few hundred years from Italy, Spain, and Corsica. Shepherds  
407 from Corsica immigrated to occupy large pastures left largely empty since the late Middle Ages,  
408 bringing an Italian-Corsican dialect (Gallurese) now prevalent in the northeastern part of Sardinia  
409 (LANNOU, 1941). The differing historical impacts of these external contacts in different regions  
410 of Sardinia is supported in the patterns we observe, with more northern Mediterranean ancestry  
411 inferred in the north (where Gallurese is prevalent), eastern Mediterranean ancestry inferred in  
412 the south and west of Sardinia (where more Punic colonies existed), and more isolation in central  
413 regions of Ogliastra and Nuoro.

414 The evidence for gene flow after the second millennium BCE seems to contradict previous  
415 models emphasizing Sardinian isolation, but we confirm that contemporary Sardinian individuals  
416 have retained an exceptionally high degree of EEF ancestry (HAAK *et al.*, 2015). Compared to  
417 other European populations, Sardinia experienced relative genetic isolation through the Bronze  
418 age, and our models also fit the majority of modern Sardinian ancestry being retained from the  
419 Nuragic period. The subsequent post-Nuragic admixture appears to derive from Mediterranean  
420 sources that have relatively little Steppe ancestry (SARNO *et al.*, 2017; LAZARIDIS *et al.*, 2017).  
421 Therefore, contemporary Sardinians still cluster with several mainland European Copper Age  
422 individuals such as Ötzi (SIKORA *et al.*, 2014), even as they are shifted from ancient Sardinian  
423 individuals of a similar time period (Fig. 2).

424 The history of gene flow into Sardinia is also relevant to understanding its relationship to  
425 the Basque populations of Iberia. Previous studies have suggested both present-day and ancient  
426 Basque individuals share a genetic connection with modern Sardinian individuals (GÜNTHER  
427 *et al.*, 2015; CHIANG *et al.*, 2018). We detected a similar signal, with modern Basque having, of  
428 all modern samples, the largest pairwise outgroup- $f_3$  with Sardinians in each of our time periods  
429 (Fig. 3). A plausible explanation arises from the observation that both Basque and Sardinians

430 have remained relatively isolated since the Neolithic transition (e.g. see [OLALDE \*et al.\*, 2019](#),  
431 for novel aDNA evidence on the Basque). While both Basque and Sardinians have received some  
432 immigration, apparently from different sources, both populations also retained an exceptionally  
433 high fraction of EEF ancestry (e.g., Fig. 5). This shared ancestry component likely contributes  
434 to the high pairwise outgroup- $f_3$  (Fig. 3) between Basque and Sardinians, and explains how both  
435 populations share a genetic affinity despite their geographic separation.

436 Overall, we find that genome-wide ancient DNA provides unique insights into the population  
437 history of Sardinia. We do not detect any significant admixture from the the Neolithic period  
438 of Sardinia through the Nuragic. From the Nuragic to the present we observe a significant shift  
439 in ancestry tied to northern and eastern Mediterranean sources. Genetic analyses that include  
440 Sardinian individuals spanning the post-Nuragic period to the present, as well as individuals from  
441 plausible sources of this gene flow, such as Punic, Roman, and other Mediterranean groups, will  
442 help to more precisely date these events and to relate them to the historical and archaeological  
443 record. Ultimately, having a more refined model of post-Nuragic demographic history will provide  
444 a better framework to understand the evolutionary history of genetic disease-variants prevalent  
445 in Sardinia and throughout the Mediterranean, such as beta-thalassemia and G6PD deficiency.

## 446 **Materials and Methods**

### 447 **Archaeological sampling**

448 The archaeological samples used in this project derive from two major collection avenues. The  
449 first was a sampling effort led by co-author Luca Lai, leveraging a broad base of samples from  
450 different existing collections in Sardinia, a subset of which were previously used in isotopic analy-  
451 ses to understand dietary composition and change in prehistoric Sardinia (LAI *et al.*, 2013). The  
452 second was from the Seulo Caves project (SKEATES *et al.*, 2013), an on-going project on a series  
453 of caves that span the Middle Neolithic to late Bronze Age near the town of Seulo. The project  
454 focuses on the diverse forms and uses of caves in the prehistoric culture of Sardinia. All samples  
455 were handled in collaboration with local scientists and with the approval of the local Sardinian  
456 authorities for the handling of archaeological samples (Ministero per i Beni e le Attivita Cultur-  
457 ali, Direzione Generale per i beni Archeologici, request dated 11 August 2009; Soprintendenza  
458 per le Beni Archeologici per le province di Sassari e Nuoro, prot. 12278 dated 05 Dec. 2014;  
459 Soprintendenza ai Beni Archeologici per le Province di Cagliari e Oristano, prot. 62, dated 08  
460 Jan 2015; Soprintendenza Archeologia, Belle arti e Paesaggio per le province di Sassari, Olbia-  
461 Tempio e Nuoro, prot. 4247 dated 14 March 2017; Soprintendenza per i Beni Archeologici per le  
462 Province di Sassari e Nuoro, prot. 12930 dated 30 Dec. 2014; Soprintendenza Archeologia, belle  
463 arti e paesaggio per le province di Sassari e Nuoro, prot. 7378 dated 9 May, 2017; Soprintendenza  
464 Archeologia, belle arti e paesaggio per le province di Sassari e Nuoro, prot. 16258 dated 26 Nov.  
465 2017). For more, detailed description of the sites please see Supplemental Information Sections  
466 1 and 2.

### 467 **Initial sample screening and sequencing**

468 The ancient DNA (aDNA) workflow was implemented in dedicated facilities at the Palaeogenetic  
469 Laboratory of the University of Tübingen and at the Department of Archaeogenetics of the Max  
470 Planck Institute for the Science of Human History in Jena. The only exception was for four sam-  
471 ples from the Seulo Cave Project which had DNA isolated at the Australian Centre for Ancient  
472 DNA and capture and sequencing carried out in the Reich lab at Harvard University. Different  
473 skeletal elements were sampled using a dentist drill to generate bone and tooth powder respec-  
474 tively. DNA was extracted following an established aDNA protocol (DABNEY *et al.*, 2013) and  
475 then converted into double-stranded libraries retaining (MEYER and KIRCHER, 2010) or partially  
476 reducing (ROHLAND *et al.*, 2015) the typical aDNA substitution pattern resulting from deami-  
477 nated cytosines that accumulate towards the molecule's termini. After indexing PCR (MEYER  
478 and KIRCHER, 2010) and differential amplification cycles, the DNA was shotgun sequenced on Il-  
479 lumina platforms. Samples showing sufficient aDNA preservation were captured for mtDNA and  
480  $\approx 1.24$  million SNPs across the human genome chosen to intersect with the Affymetrix Human  
481 Origins array and Illumina 610-Quad array (FU *et al.*, 2015). The resulting enriched libraries  
482 were also sequenced on Illumina machines in single-end or paired-end mode. Sequenced data  
483 were pre-processed using the EAGER pipeline (PELTZER *et al.*, 2016). Specifically, DNA adapters  
484 were trimmed using AdapterRemoval v2 (SCHUBERT *et al.*, 2016) and paired-end sequenced li-  
485 braries were merged. Sequence alignment to the mtDNA (RSRS) and nuclear (hg19) reference  
486 genomes was performed with BWA (LI and DURBIN, 2009) (parameters  $-n$  0.01, seeding disabled),  
487 duplicates were removed with DeDup (PELTZER *et al.*, 2016) and a mapping quality filter was  
488 applied ( $MQ \geq 30$ ). For genetic sexing, we compared relative X and Y-chromosome coverage to  
489 the autosomal coverage with a custom script. For males, nuclear contamination levels were esti-  
490 mated based on heterozygosity on the X-chromosome with the software ANGSD (KORNELIUSSEN  
491 *et al.*, 2014). Data originating from mtDNA capture was processed with schmutzi (RENAUD

492 *et al.*, 2015), which jointly estimates mtDNA contamination and reconstructs mtDNA consen-  
493 sus sequences that were assigned to the corresponding mtDNA haplogroups using *Haplofind*  
494 (*VIANELLO et al.*, 2013) (Supp. Mat. 1D). We applied several standard ancient DNA quality  
495 control metrics: We retained endogenous DNA content in shotgun sequencing  $>0.2\%$ , evidence  
496 of an average damage pattern present at the molecule termini, mtDNA contamination  $<4\%$   
497 (average  $1.6\%$ ) and nuclear contamination  $<6\%$  (average  $1.1\%$ ).

498 We next generated genotype calls that were used for downstream population genetic analyses.  
499 To account for sequencing errors we first removed any reads that overlapped a SNP on the capture  
500 array with a base quality score less than 20. We also removed the last 3-bp on both sides of  
501 every read to reduce the effect of DNA damage on the resulting genotype calls (*AL-ASADI*  
502 *et al.*, 2018). With these filtered aligned reads in hand, we used custom python scripts (<https://github.com/mathii/gdc3>)  
503 to generate pseudo-haploid genotypes by sampling a random read  
504 for each SNP on the capture array and setting the genotype to be homozygous for the allele  
505 present on the randomly sampled read.

## 506 Merging newly generated data with published data

507 **Ancient DNA datasets from Western Eurasia.** To provide context for the study of an-  
508 cestry of the ancient individuals from Sardinia, we downloaded and processed several ancient  
509 datasets from continental Europe and the Middle-east (*MATHIESON et al.*, 2015; *LAZARIDIS*  
510 *et al.*, 2016, 2017; *MATHIESON et al.*, 2018, 2017; *LIPSON et al.*, 2017; *OLALDE et al.*, 2018).  
511 To minimize technology-specific batch effects in genotype calls and thus downstream population  
512 genetic inference, we focused on previously published ancient samples that had undergone the  
513 capture protocol on the same set of SNPs targeted in our study. We processed these samples  
514 through the same pipeline and filters described above, resulting in a dataset of 972 ancient sam-  
515 ples. Throughout our analysis, we used a subset of  $n = 1,013,439$  variants that was created by  
516 removing SNPs missing in more than 80% of all ancient individuals (Sardinian and reference  
517 dataset) with at least 60% of all captured SNPs covered.

518 This ancient dataset spans a wide geographic distribution and temporal range. Ancient  
519 individuals are associated with a variety of different cultures, which provides rich context for  
520 interpreting downstream results. Our reference ancient dataset is comprised of many individuals  
521 sampled from a particular geographic locale, such as Germany or Hungary, in a transect of  
522 multiple cultural changes through time (Fig. 2). For the PCA (Fig. 2), we additionally included  
523 a single low-coverage ancient individual (label “Pun”) dated to 361-178 BCE from a Punic  
524 necropolis on the west Mediterranean island of Ibiza (*ZALLOUA et al.*, 2018).

525 We merged individuals into groups (Supp. Mat. 1F,G). For ancient samples, these groups were  
526 chosen manually, trying to strike a balance between reducing overlap in the PCA and keeping  
527 culturally distinct populations separate. We used geographic location to first broadly group  
528 samples into geographic areas (such as Iberia, Central Europe and Balkans), and then further  
529 annotated each of these groups by different time periods.

530 **Contemporary DNA datasets from Western Eurasia.** We downloaded and processed the  
531 Human Origins dataset to characterize a subset of Eurasian human genetic diversity at 594,924  
532 autosomal SNPs (*LAZARIDIS et al.*, 2014). To be consistent with previous studies (*LAZARIDIS*  
533 *et al.*, 2014; *MATHIESON et al.*, 2015), we focused on a subset of 777 individuals from Western  
534 Eurasia.

535 **Contemporary DNA dataset from Sardinia.** We merged in a whole-genome sequence  
536 dataset which was described and previously analyzed by *CHIANG et al.* (2018). It consists of

537 1,577 unrelated individuals, grouped into multiple geographic regions within Sardinia (Fig. 2C).

### 538 Principal Components Analysis

539 We performed Principal Components Analysis (PCA) on two large-scale datasets of modern  
540 genotypes from Western Eurasian (777 individuals from the Human Origins dataset) and Sardinia  
541 (1,577 individuals from the SARDINIA project). For both datasets, we normalized the genotype  
542 matrix by mean-centering and scaling the genotypes at each SNP using the inverse of the square-  
543 root of heterozygosity (PATTERSON *et al.*, 2006). We additionally filtered out rare variants with  
544 minor allele frequency ( $p_{\min} < 0.05$ ).

545 To assess population structure in the ancient individuals, we projected them onto our pre-  
546 computed principal axes using only the non-missing SNPs via a least-squares approach, and  
547 correcting for the shrinkage effect observed in high-dimensional PC score prediction (REICH  
548 *et al.*, 2008; LEE *et al.*, 2010). More details on how we corrected the biased PC scores are  
549 discussed in Supp. Info. 8.

550 We also projected a number of out-sample sub-populations from Sardinia onto our PCs.  
551 Reassuringly, these out-of-sample Sardinian individuals project very close to Humans Origins  
552 Sardinian individuals (Fig. 2). Moreover, the test-set Sardinia individuals with grand-parental  
553 ancestry from Southern Italy cluster with reference individuals with ancestry from Sicily.

### 554 ADMIXTURE Analysis

555 We applied ADMIXTURE to an un-normalized genotype matrix of ancient and modern samples  
556 (ALEXANDER *et al.*, 2009). ADMIXTURE is a maximum-likelihood based method for fitting the  
557 Pritchard, Stephens and Donnelly model PRITCHARD *et al.* (2000) using sequential quadratic  
558 programming. We first LD pruned the data matrix based off the modern Western Eurasian  
559 genotypes, using plink1.9 with parameters [--indep-pairwise 200 25 0.4]. We then ran 5  
560 replicates of ADMIXTURE for values of  $K = 2, \dots, 11$ . We display results for the replicate that  
561 reached the highest log-likelihood after the algorithm converged (Supp. Fig. 8).

### 562 Estimation of $f$ -statistics

563 We measured similarity between groups of individuals through computing an outgroup- $f_3$  statis-  
564 tic. The outgroup- $f_3$  statistic can be interpreted as a measure of the internal branch length of a  
565 three-taxa population phylogeny and thus does not depend strongly on genetic drift or systematic  
566 error in the focal pair of populations that are being compared (PATTERSON *et al.*, 2012).

567 Here we used the ancestral allelic states as an outgroup, inferred from a multi-species align-  
568 ment from Ensembl Compara release 59, as annotated in the 1000 Genomes Phase3 sites vcf  
569 ([ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ALL.wgs.phase3\\_shapeit2\\_](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ALL.wgs.phase3_shapeit2_mvncall_integrated_v5b.20130502.sites.vcf.gz)  
570 [mvncall\\_integrated\\_v5b.20130502.sites.vcf.gz](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ALL.wgs.phase3_shapeit2_mvncall_integrated_v5b.20130502.sites.vcf.gz)) (1000 GENOMES PROJECT CONSORTIUM  
571 *et al.*, 2015). We fixed the ancestral allele counts to  $n = 10^6$  to avoid finite sample size correction  
572 when calculating outgroup  $f_3$ .

573 The  $f_3$ - and  $f_4$ -statistics that test for admixture were computed with scikit-allel using  
574 average\_patterson\_f3 and average\_patterson\_d. We estimated standard errors with a block  
575 jack-knife over 1000 markers (blen=1000). When analyzing ancient individuals that were repre-  
576 sented as pseudo-haploid genotypes, we analyzed only one allele to avoid an artificial appearance  
577 of genetic drift - that could for instance mask a negative  $f_3$  signal of admixture.

### 578 Estimation of $F_{ST}$ -coefficients

579 To measure pairwise genetic differentiation between two populations, we estimated average pair-  
580 wise  $F_{ST}$  and standard error via block-jackknife over 1000 markers, using `average_patterson_fst`  
581 from the package `scikit-allel`. When analyzing ancient individuals that were represented  
582 as pseudo-haploid genotypes, we analyzed only one allele to avoid artificial genetic drift. For  
583 this analysis, we removed first degree relatives within each population. Another estimator,  
584 `average_hudson_fst` gave highly correlated results ( $r^2 = 0.89$ ), differing mostly for populations  
585 with very low sample size ( $n \leq 5$ ) and did not change any qualitative conclusions.

### 586 Estimation of admixture proportions with qpAdm

587 We estimated admixture fractions of a selected target population as well as model consistency for  
588 one-, two- or three-source models. We used the framework of [HAAK \*et al.\* \(2015\)](#) as implemented  
589 in `qpAdm`, which relates a set of “left” populations (the population of interest and candidate  
590 ancestral sources) to a set of “right” populations (diverse out-groups). To test the robustness of  
591 our results to the choice of right populations, we ran one analysis with a previously used set of  
592 modern populations as outgroup ([HAAK \*et al.\*, 2015](#)), and another analysis with a set of ancient  
593 Europeans that have been previously used to disentangle divergent strains of ancestry present in  
594 Europe ([LAZARIDIS \*et al.\*, 2017](#)). The full qpAdm results are discussed in Supp. Info. 7.

### 595 Inference of Y haplogroups

596 To determine the haplotype branch of the Y chromosome of male ancient individuals, we analyzed  
597 informative SNPs on the Y-haplotype tree. For reference, we used markers from [https://  
598 isogg.org/tree/](https://isogg.org/tree/) (Version: 13.238, 2018). We merged this data-set with our set of calls and  
599 identified markers available in both to create groups of equivalent markers for sub-haplogroups.  
600 Our targeted sequencing approach yielded calls for up to 32,681 such Y-linked markers per  
601 individual. As the conventions for naming of haplogroups are subject to change, we annotate  
602 them in terms of carrying the derived state at a defining SNP. We analyzed the number of  
603 derived and ancestral calls for each informative marker for all ancient Sardinian individuals, and  
604 additionally reanalyzed male ancient West Eurasians in our reference data set.

## 605 Data Availability

606 The raw reads and aligned sequences of the data generated from this study will available through  
607 the European Nucleotide Archive (ENA) before publication under accession number [TBD prior  
608 to publication]. Processed genotype calls will be posted to the Novembre Lab website and on  
609 Data Dryad. The contemporary Sardinia data used to support this study have allele frequency  
610 summary data deposited to EGA under accession number EGAS00001002212. The disaggregated  
611 individual-level sequence data for 1,577 used in this study is a subset of 2,105 samples (adult  
612 volunteers of the SardiNIA cohort longitudinal study) from Sidore *et al* (2015) and are available  
613 from dbGAP under project identifier phs000313 (v4.p2). The remaining individual-level sequence  
614 data are from a case-control study of autoimmunity from across Sardinia, and per the obtained  
615 consent and local IRB, these data are only available for collaboration by request from the project  
616 leader (Francesco Cucca, Consiglio Nazionale delle Ricerche, Italy)

## 617 Acknowledgements

618 We thank Maanasa Rhagavan for in-depth feedback on drafts, Anna Di Rienzo and Goncalo  
619 Abecasis for helpful discussions, and Magdalena Zoledziewska for useful comments and early  
620 assistance. We would like to thank Antje Wissgott, Cécilia Freund and other members of the  
621 wet laboratory and computationally teams at MPI-SHH in Jena. We thank Nadin Rohland,  
622 Éadaoin Harney, Shop Mallick, and Alan Cooper for contributing to generating the data for the  
623 four samples processed at the Australian Centre for Ancient DNA and in D.R.'s ancient DNA  
624 laboratory. We also thank Dan Rice, Chi-chun Liu, and members of the Novembre lab for helpful  
625 discussion and feedback. This study was supported in part by the National Science Foundation  
626 fellowship DGE-1746045 and the National Institute of General Medical Sciences under training  
627 grant award number T32GM007197 to J.H.M. and by funding from the MacArthur Foundation  
628 to J.N.. The contribution of D.R. to this study was supported by NSF HOMINID grant BCS-  
629 1032255 and by the Howard Hughes Medical Institute. We also thank Alan Cooper for a role in  
630 the collection of those 4 samples.

## 631 Author contributions

632 We annotate author contributions using the CRediT Taxonomy labels (<https://casrai.org/credit/>).  
633 Where multiple individuals serve in the same role, the degree of contribution is  
634 specified as 'lead', 'equal', or 'supporting'.

- 635 • Conceptualization (Design of study) – lead: FC, JN, JK, LL; supporting: CS, CP, DS,  
636 JHM, GA
- 637 • Investigation (Collection of skeletal samples) – lead: LL, RS; supporting: JB, MGG, CDS,  
638 CP (minor contribution from CS, JN)
- 639 • Investigation (Ancient DNA isolation and sequencing) – lead: CP, AF; supporting: CDS,  
640 JK, DR\*, RR
- 641 • Data Curation (Data quality control and initial analysis) – lead: JHM, CP; supporting:  
642 HR, CS, CC, KD, HA, AO
- 643 • Formal Analysis (General population genetics) – lead: JHM, HR; supporting: TAJ
- 644 • Writing (original draft preparation) – lead: JHM, HR, JN; supporting: CP, RS, LL, FC
- 645 • Writing (review and editing) – input from all authors\*
- 646 • Supervision – equal: FC, JK, JN
- 647 • Funding acquisition – lead: JK, FC, JN; supporting: RS

648 \*: D.R. contributed data for four samples and reviewed the description of the data generation  
649 for these samples. As he is also senior author on a separate manuscript that reports data on a  
650 non-overlapping set of ancient Sardinians and his group and ours wished to keep the two studies  
651 intellectually independent, he did not review any other parts of this manuscript.

## References

- 652
- 653 1000 GENOMES PROJECT CONSORTIUM, *et al.*, 2015 A global reference for human genetic vari-  
654 ation. *Nature* **526**: 68.
- 655 AL-ASADI, H., K. DEY, J. NOVEMBRE, and M. STEPHENS, 2018 Inference and visualization of  
656 dna damage patterns using a grade of membership model. *bioRxiv* : 327684.
- 657 ALEXANDER, D. H., J. NOVEMBRE, and K. LANGE, 2009 Fast model-based estimation of  
658 ancestry in unrelated individuals. *Genome Research* **19**: 1655–1664.
- 659 ALLENTOFT, M. E., M. SIKORA, K.-G. SJÖGREN, S. RASMUSSEN, M. RASMUSSEN, *et al.*, 2015  
660 Population genomics of Bronze Age Eurasia. *Nature* **522**: 167–172.
- 661 AMMERMAN, A. J., and L. L. CAVALLI-SFORZA, 2014 *The Neolithic transition and the genetics*  
662 *of populations in Europe*, volume 836. Princeton University Press.
- 663 BARBUJANI, G., and R. R. SOKAL, 1990 Zones of sharp genetic change in europe are also  
664 linguistic boundaries. *Proceedings of the National Academy of Sciences* **87**: 1816–1819.
- 665 BARNETT, W. K., 2000 Cardial pottery and the agricultural transition in mediterranean europe.  
666 Europe’s first farmers : 93–116.
- 667 CALÒ, C., A. MELIS, G. VONA, and I. PIRAS, 2008 Review synthetic article: Sardinian popu-  
668 lation (italy): A genetic review. *International Journal of Modern Anthropology* **1**: 39–64.
- 669 CAVALLI-SFORZA, L. L., 2005 The human genome diversity project: past, present and future.  
670 *Nature Reviews Genetics* **6**: 333.
- 671 CHIANG, C. W., J. H. MARCUS, C. SIDORE, A. BIDDANDA, H. AL-ASADI, *et al.*, 2018 Genomic  
672 history of the Sardinian population. *Nature Genetics* : 1.
- 673 CONTU, L., M. ARRAS, C. CARCASSI, G. L. NASA, and M. MULARGIA, 1992 Hla structure of  
674 the sardinian population: a haplotype study of 551 families. *Tissue Antigens* **40**: 165–174.
- 675 DABNEY, J., M. KNAPP, I. GLOCKE, M.-T. GANSAUGE, A. WEIHMANN, *et al.*, 2013 Complete  
676 mitochondrial genome sequence of a middle pleistocene cave bear reconstructed from ultrashort  
677 dna fragments. *Proceedings of the National Academy of Sciences* : 201314445.
- 678 EAVES, I. A., T. R. MERRIMAN, R. A. BARBER, S. NUTLAND, E. TUOMILEHTO-WOLF, *et al.*,  
679 2000 The genetically isolated populations of finland and sardinia may not be a panacea for  
680 linkage disequilibrium mapping of common disease genes. *Nature Genetics* **25**: 320.
- 681 ERMINI, L., C. OLIVIERI, E. RIZZI, G. CORTI, R. BONNAL, *et al.*, 2008 Complete mitochondrial  
682 genome sequence of the Tyrolean Iceman. *Current Biology* **18**: 1687–1693.
- 683 FRANCALACCI, P., L. MORELLI, A. ANGIUS, R. BERUTTI, F. REINIER, *et al.*, 2013 Low-pass  
684 DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. *Science*  
685 **341**: 565–569.
- 686 FU, Q., M. HAJDINJAK, O. T. MOLDOVAN, S. CONSTANTIN, S. MALLICK, *et al.*, 2015 An early  
687 modern human from romania with a recent neanderthal ancestor. *Nature* **524**: 216.
- 688 GHIROTTI, S., S. MONA, A. BENAZZO, F. PAPAARAZZO, D. CARAMELLI, *et al.*, 2009 Infer-  
689 ring genealogical processes from patterns of bronze-age and modern dna variation in sardinia.  
690 *Molecular biology and evolution* **27**: 875–886.

- 691 GÜNTHER, T., C. VALDIOSERA, H. MALMSTRÖM, I. UREÑA, R. RODRIGUEZ-VARELA, *et al.*,  
692 2015 Ancient genomes link early farmers from atapuerca in spain to modern-day basques.  
693 *Proceedings of the National Academy of Sciences* **112**: 11917–11922.
- 694 HAAK, W., I. LAZARIDIS, N. PATTERSON, N. ROHLAND, S. MALLICK, *et al.*, 2015 Massive  
695 migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**:  
696 207–211.
- 697 HOFMANOVÁ, Z., S. KREUTZER, G. HELLENTHAL, C. SELL, Y. DIEKMANN, *et al.*, 2016 Early  
698 farmers from across Europe directly descended from Neolithic Aegeans. *Proceedings of the*  
699 *National Academy of Sciences* **113**: 6886–6891.
- 700 KELLER, A., A. GRAEFEN, M. BALL, M. MATZAS, V. BOISGUERIN, *et al.*, 2012 New insights  
701 into the tyrolean iceman’s origin and phenotype as inferred by whole-genome sequencing.  
702 *Nature communications* **3**: 698.
- 703 KORNELIUSSEN, T. S., A. ALBRECHTSEN, and R. NIELSEN, 2014 Angsd: analysis of next  
704 generation sequencing data. *BMC bioinformatics* **15**: 356.
- 705 LAI, L., R. H. TYKOT, E. USAI, J. F. BECKETT, R. FLORIS, *et al.*, 2013 Diet in the sar-  
706 dinian bronze age: models, collagen isotopic data, issues and perspectives. *Préhistoires*  
707 *Méditerranéennes* .
- 708 LANNOU, M. L., 1941 Pâtres et Paysans de la Sardaigne. *Tours* **8**: 364.
- 709 LAZARIDIS, I., A. MITTNIK, N. PATTERSON, S. MALLICK, N. ROHLAND, *et al.*, 2017 Genetic  
710 origins of the Minoans and Mycenaeans. *Nature* **548**: 214–218.
- 711 LAZARIDIS, I., D. NADEL, G. ROLLEFSON, D. C. MERRETT, N. ROHLAND, *et al.*, 2016 Genomic  
712 insights into the origin of farming in the ancient Near East. *Nature* **536**: 419–424.
- 713 LAZARIDIS, I., N. PATTERSON, A. MITTNIK, G. RENAUD, S. MALLICK, *et al.*, 2014 Ancient  
714 human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**:  
715 409–413.
- 716 LEE, S., F. ZOU, and F. A. WRIGHT, 2010 Convergence and prediction of principal component  
717 scores in high-dimensional settings. *Annals of Statistics* **38**: 3605.
- 718 LETTRE, G., and J. N. HIRSCHHORN, 2015 Small island, big genetic discoveries. *Nature Genetics*  
719 **47**: 1224–1225.
- 720 LI, H., and R. DURBIN, 2009 Fast and accurate short read alignment with burrows–wheeler  
721 transform. *bioinformatics* **25**: 1754–1760.
- 722 LIPSON, M., A. SZÉCSÉNYI-NAGY, S. MALLICK, A. PÓSA, B. STÉGMÁR, *et al.*, 2017 Parallel  
723 palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* .
- 724 LUGLIÈ, C., 2018 Your path led trough the sea... the emergence of Neolithic in Sardinia and  
725 Corsica. *Quaternary International* **470**: 285–300.
- 726 MARTINS, H., F. X. OMS, L. PEREIRA, A. W. PIKE, K. ROWSELL, *et al.*, 2015 Radiocar-  
727 bon dating the beginning of the Neolithic in Iberia: new results, new problems. *Journal of*  
728 *Mediterranean Archaeology* **28**: 105–131.
- 729 MASTINO, A., 2005 *Storia della Sardegna antica*, volume 2. Il Maestrale.

- 730 MATHIESON, I., S. ALPASLAN-ROODENBERG, C. POSTH, A. SZÉCSÉNYI-NAGY, N. ROHLAND,  
731 *et al.*, 2018 The genomic history of southeastern Europe. *Nature* **555**: 197.
- 732 MATHIESON, I., I. LAZARIDIS, N. ROHLAND, S. MALLICK, N. PATTERSON, *et al.*, 2015 Genome-  
733 wide patterns of selection in 230 ancient Eurasians. *Nature* **528**: 499–503.
- 734 MATHIESON, I., S. A. ROODENBERG, C. POSTH, A. SZÉCSÉNYI-NAGY, N. ROHLAND, *et al.*,  
735 2017 The genomic history of Southeastern Europe. *bioRxiv* : 135616.
- 736 MATISOO-SMITH, E., A. GOSLING, D. PLATT, O. KARDAILSKY, S. PROST, *et al.*, 2018 Ancient  
737 mitogenomes of Phoenicians from Sardinia and Lebanon: A story of settlement, integration,  
738 and female mobility. *PloS one* **13**: e0190169.
- 739 MELIS, P., 2002 Un Approdo della costa di Castelsardo, fra età nuragica e romana. *Atti del*  
740 *XIV Congresso L’Africa Romana – Sassari 7-10 dicembre 2000* : 1331–1343.
- 741 MEYER, M., and M. KIRCHER, 2010 Illumina sequencing library preparation for highly multi-  
742 plexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010**: pdb-prot5448.
- 743 MODI, A., F. TASSI, R. R. SUSCA, S. VAI, E. RIZZI, *et al.*, 2017 Complete mitochondrial  
744 sequences from Mesolithic Sardinia. *Scientific reports* **7**: 42869.
- 745 OLALDE, I., S. BRACE, M. E. ALLENTOFT, I. ARMIT, K. KRISTIANSEN, *et al.*, 2018 The Beaker  
746 phenomenon and the genomic transformation of Northwest Europe. *Nature* **555**: 190.
- 747 OLALDE, I., S. MALLICK, N. PATTERSON, N. ROHLAND, V. VILLALBA-MOUCO, *et al.*, 2019 The  
748 genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**: 1230–1234.
- 749 OLIVIERI, A., C. SIDORE, A. ACHILLI, A. ANGIUS, C. POSTH, *et al.*, 2017 Mitogenome diversity  
750 in Sardinians: a genetic window onto an island’s past. *Molecular Biology and Evolution* **34**:  
751 1230–1239.
- 752 ORTU, L., 2011 *Storia della Sardegna dal Medioevo all’età contemporanea*. Cucc.
- 753 PATTERSON, N., P. MOORJANI, Y. LUO, S. MALLICK, N. ROHLAND, *et al.*, 2012 Ancient  
754 admixture in human history. *Genetics* **192**: 1065–1093.
- 755 PATTERSON, N., A. L. PRICE, and D. REICH, 2006 Population structure and eigenanalysis.  
756 *PLoS Genetics* **2**: e190.
- 757 PELTZER, A., G. JÄGER, A. HERBIG, A. SEITZ, C. KNIEP, *et al.*, 2016 Eager: efficient ancient  
758 genome reconstruction. *Genome Biology* **17**: 60.
- 759 PICKRELL, J. K., and D. REICH, 2014 Toward a new history and geography of human genes  
760 informed by ancient dna. *Trends in Genetics* **30**: 377–389.
- 761 PRITCHARD, J. K., M. STEPHENS, and P. DONNELLY, 2000 Inference of population structure  
762 using multilocus genotype data. *Genetics* **155**: 945–959.
- 763 REICH, D., A. L. PRICE, and N. PATTERSON, 2008 Principal component analysis of genetic  
764 data. *Nature Genetics* **40**: 491.
- 765 RENAUD, G., V. SLON, A. T. DUGGAN, and J. KELSO, 2015 Schmutzi: estimation of contami-  
766 nation and endogenous mitochondrial consensus calling for ancient dna. *Genome Biology* **16**:  
767 224.

- 768 ROHLAND, N., E. HARNEY, S. MALLICK, S. NORDENFELT, and D. REICH, 2015 Partial uracil-  
769 dna-glycosylase treatment for screening of ancient dna. *Philosophical Transactions of the*  
770 *Royal Society B: Biological Sciences* **370**: 20130624.
- 771 SARNO, S., A. BOATTINI, L. PAGANI, M. SAZZINI, S. DE FANTI, *et al.*, 2017 Ancient and  
772 recent admixture layers in Sicily and Southern Italy trace multiple migration routes along the  
773 Mediterranean. *Scientific reports* **7**: 1984.
- 774 SCHUBERT, M., S. LINDGREEN, and L. ORLANDO, 2016 Adapterremoval v2: rapid adapter  
775 trimming, identification, and read merging. *BMC research notes* **9**: 88.
- 776 SIDORE, C., F. BUSONERO, A. MASCHIO, E. PORCU, S. NAITZA, *et al.*, 2015 Genome sequenc-  
777 ing elucidates Sardinian genetic architecture and augments association analyses for lipid and  
778 blood inflammatory markers. *Nature Genetics* **47**: 1272–1281.
- 779 SIKORA, M., M. L. CARPENTER, A. MORENO-ESTRADA, B. M. HENN, P. A. UNDERHILL,  
780 *et al.*, 2014 Population genomic analysis of ancient and modern genomes yields new insights  
781 into the genetic ancestry of the Tyrolean Iceman and the genetic structure of Europe. *PLoS*  
782 *Genetics* **10**: e1004353.
- 783 SINISCALCO, M., L. BERNINI, G. FILIPPI, B. LATTE, P. M. KHAN, *et al.*, 1966 Population  
784 genetics of haemoglobin variants, thalassaemia and glucose-6-phosphate dehydrogenase de-  
785 ficiency, with particular reference to the malaria hypothesis. *Bulletin of the World Health*  
786 *Organization* **34**: 379.
- 787 SKEATES, R., M. G. GRADOLI, and J. BECKETT, 2013 The cultural life of caves in Seulo, central  
788 Sardinia. *Journal of Mediterranean Archaeology* **26**.
- 789 SKOGLUND, P., H. MALMSTRÖM, A. OMRÄK, M. RAGHAVAN, C. VALDIOSERA, *et al.*, 2014  
790 Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers.  
791 *Science* **344**: 747–750.
- 792 SKOGLUND, P., H. MALMSTRÖM, M. RAGHAVAN, J. STORÅ, P. HALL, *et al.*, 2012 Origins and  
793 genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**: 466–469.
- 794 TYKOT, R. H., 1996 Obsidian procurement and distribution in the central and western Mediter-  
795 ranean. *Journal of Mediterranean Archaeology* **9**: 39–82.
- 796 VIANELLO, D., F. SEVINI, G. CASTELLANI, L. LOMARTIRE, M. CAPRI, *et al.*, 2013 Haplofind:  
797 A new method for high-throughput mt dna haplogroup assignment. *Human mutation* **34**:  
798 1189–1194.
- 799 ZALLOUA, P., C. J. COLLINS, A. GOSLING, S. A. BIAGINI, B. COSTA, *et al.*, 2018 Ancient  
800 DNA of Phoenician remains indicates discontinuity in the settlement history of Ibiza. *Scientific*  
801 *reports* **8**: 17567.
- 802 ZALLOUA, P. A., D. E. PLATT, M. EL SIBAI, J. KHALIFE, N. MAKHOUL, *et al.*, 2008 Identifi-  
803 cing genetic traces of historical expansions: Phoenician footprints in the Mediterranean. *The*  
804 *American Journal of Human Genetics* **83**: 633–642.
- 805 ZAVATTARI, P., E. DEIDDA, M. WHALEN, R. LAMPIS, A. MULARGIA, *et al.*, 2000 Major factors  
806 influencing linkage disequilibrium by analysis of different chromosome regions in distinct pop-  
807 ulations: demography, chromosome recombination frequency and selection. *Human molecular*  
808 *genetics* **9**: 2947–2957.

809 ZILHÃO, J., 2001 Radiocarbon evidence for maritime pioneer colonization at the origins of farm-  
810 ing in west Mediterranean Europe. *Proceedings of the national Academy of Sciences* **98**:  
811 14180–14185.