



Universiteit  
Leiden  
The Netherlands

## **Ancient genomics support deep divergence between Eastern and Western Mediterranean Indo-European languages**

Yediay, F.E.; Kroonen, G.J.; Sabatini, S.; Frei, K.M.; Frank, A. B.; Pinotti, T.; ... ; Willerslev, E.

### **Citation**

Yediay, F. E., Kroonen, G. J., Sabatini, S., Frei, K. M., Frank, A. B., Pinotti, T., ... Willerslev, E. (2024). Ancient genomics support deep divergence between Eastern and Western Mediterranean Indo-European languages. *Biorxiv*. doi:10.1101/2024.12.02.626332

Version: Accepted Manuscript

License: [Creative Commons CC BY-NC-ND 4.0 license](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Downloaded from: <https://hdl.handle.net/1887/4171691>

**Note:** To cite this publication please use the final published version (if applicable).

1 **Ancient genomics support deep divergence between Eastern and Western**  
2 **Mediterranean Indo-European languages**

3

4 Fulya Eylem Yediay<sup>1,2\*</sup>, Guus Kroonen<sup>3,4\*</sup>, Serena Sabatini<sup>2#</sup>, Karin Margarita Frei<sup>5#</sup>, Anja B.  
5 Frank<sup>6,7</sup>, Thomaz Pinotti<sup>1,8</sup>, Andrew Wigman<sup>3</sup>, Rasmus Thorsø<sup>3</sup>, Tharsika Vimala<sup>1</sup>, Hugh  
6 McColl<sup>1,2</sup>, Ioanna Moutafi<sup>9,10</sup>, Isin Altinkaya<sup>1</sup>, Abigail Ramsøe<sup>1</sup>, Charleen Gaunitz<sup>1</sup>, Gabriel  
7 Renaud<sup>11</sup>, Alfredo Mederos Martin<sup>12</sup>, Fabrice Demeter<sup>1,13</sup>, Gabriele Scorrano<sup>1,14</sup>, Alessandro  
8 Canci<sup>15</sup>, Peter Fischer<sup>2</sup>, Izzet Duyar<sup>16</sup>, Claude Serhal<sup>17</sup>, Alexander Varzari<sup>18,19</sup>, Murat  
9 Türkteki<sup>20</sup>, John O'Shea<sup>21</sup>, Lorenz Rahmstorf<sup>22</sup>, Gürcan Polat<sup>23</sup>, Derya Atamtürk<sup>24</sup>, Lasse  
10 Vinner<sup>1</sup>, Sachihito Omura<sup>25</sup>, Kimiyoshi Matsumura<sup>25</sup>, Jialu Cao<sup>1</sup>, Frederik Valeur  
11 Seersholm<sup>1</sup>, Jose Miguel Morillo Leon<sup>26</sup>, Sofia Voutsaki<sup>27</sup>, Raphaël Orgeolet<sup>28,29</sup>, Brendan  
12 Burke<sup>30</sup>, Nicholas P Herrmann<sup>31</sup>, Giulia Recchia<sup>32</sup>, Susi Corazza<sup>15</sup>, Elisabetta  
13 Borgna<sup>15</sup>, Mirella Cipolloni Sampò<sup>33</sup>, Flavia Trucco<sup>34</sup>, Ana Pajuelo Pando<sup>35</sup>, Marie Louise  
14 Schjellerup Jørkov<sup>36</sup>, Patrice Courtaud<sup>37</sup>, Rebecca Peake<sup>38,39</sup>, Juan Francisco Gibaja  
15 Bao<sup>40</sup>, Györgyi Parditka<sup>21</sup>, Jesper Stenderup<sup>1</sup>, Karl-Göran Sjögren<sup>2</sup>, Jacqueline Staring<sup>1</sup>, Line  
16 Olsen<sup>1</sup>, Igor V. Deyneko<sup>41</sup>, György Pálfi<sup>42</sup>, Pedro Manuel López Aldana<sup>35</sup>, Bryan  
17 Burns<sup>43</sup>, László Paja<sup>42</sup>, Christian Mühlenbock<sup>44</sup>, Claudio Cavazzuti<sup>45</sup>, Alberto  
18 Cazzella<sup>32</sup>, Anna Lagia<sup>46</sup>, Vassilis Lambrinoudakis<sup>47</sup>, Lazaros Kolonas<sup>48</sup>, Jörg  
19 Rambach<sup>49,50</sup>, Eugen Sava<sup>18</sup>, Sergey Agulnikov<sup>51</sup>, Vicente Castañeda Fernández<sup>52</sup>, Mia  
20 Broné<sup>53</sup>, Victoria Peña Romo<sup>54</sup>, Fernando Molina González<sup>55</sup>, Juan Antonio Cámara  
21 Serrano<sup>55</sup>, Sylvia Jiménez Brobeil<sup>56</sup>, Trinidad Nájera Molino<sup>55</sup>, María Oliva Rodríguez  
22 Ariza<sup>57</sup>, Catalina Galán Saulnier<sup>58</sup>, Armando González Martín<sup>59</sup>, Nicolas Cauwe<sup>60</sup>, Claude  
23 Mordant<sup>61</sup>, Mafalda Roscio<sup>62</sup>, Luc Staniaszek<sup>38,61</sup>, Mary Anne Tafuri<sup>32</sup>, Tayfun  
24 Yıldırım<sup>63</sup>, Luciano Salzani<sup>64</sup>, Thorfinn Sand Korneliusen<sup>1</sup>, J. Víctor Moreno-  
25 Mayar<sup>1</sup>, Morten Erik Allentoft<sup>1,65</sup>, Martin Sikora<sup>1</sup>, Rasmus Nielsen<sup>1,66</sup>, Kristian  
26 Kristiansen<sup>1,2\*</sup> & Eske Willerslev<sup>1,67\*</sup>

27

28 Affiliations

29 <sup>1</sup>Lundbeck Foundation GeoGenetics Center, Globe Institute, University of Copenhagen,  
30 Copenhagen, Denmark, <sup>2</sup>Department of Historical Studies, University of Gothenburg,  
31 Gothenburg, Sweden, <sup>3</sup>Leiden University Centre for Linguistics, Leiden University, Leiden,  
32 The Netherlands, <sup>4</sup>Department of Nordic Studies and Linguistics, University of Copenhagen,  
33 Copenhagen, Denmark, <sup>5</sup>Department of Environmental Archaeology and Materials Science,  
34 National Museum of Denmark, Kongens Lyngby, Denmark, <sup>6</sup>Institute for Geology,  
35 University of Hamburg, Hamburg, Germany, <sup>7</sup>Department of Research, Collections and  
36 Conservation, Environmental Archaeology and Materials Science, National Museum of  
37 Denmark, Kongens Lyngby, Denmark, <sup>8</sup>Laboratório de Biodiversidade e Evolução Molecular  
38 (LBEM), Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, <sup>9</sup>Institute for Pre-  
39 and Protohistoric Archaeology and Archaeology of the Roman Provinces, Ludwig-  
40 Maximilians-Universität München, Germany, <sup>10</sup>The M.H. Wiener Laboratory for  
41 Archaeological Science, American School of Classical Studies at Athens, <sup>11</sup>Department of  
42 Health Technology, Technical University of Denmark, Copenhagen, Denmark, <sup>12</sup>Universidad  
43 Autonoma de Madrid, Spain, <sup>13</sup>Eco-anthropologie (EA), Dpt ABBA, Muséum national  
44 d'Histoire naturelle, CNRS, Université Paris Cité, Musée de l'Homme 17 place du  
45 Trocadéro, 75016 Paris, France, <sup>14</sup>Center for Molecular Anthropology for the study of ancient  
46 DNA, Department of Biology, University of Rome Tor Vergata, Rome Italy, <sup>15</sup>University of  
47 Udine, Italy, <sup>16</sup>Istanbul University, <sup>17</sup>British Museum, London (UK) and University College  
48 London (UK), <sup>18</sup>The National Museum of History of Moldova, Chişinău, Republic of  
49 Moldova, <sup>19</sup>Laboratory of Human Genetics, Chiril Draganiuc Institute of

50 Phthisiopneumology, Chişinău, Republic of Moldova, <sup>20</sup>Bilecik Şeyh Edebali University,  
51 Turkey, <sup>21</sup>Museum of Anthropological Archaeology, University of Michigan,  
52 USA, <sup>22</sup>University of Göttingen, Germany, <sup>23</sup>Ege University, Turkey, <sup>24</sup>Istanbul University,  
53 Turkey, <sup>25</sup>Japanese Institute of Anatolian Archaeology, Turkey, <sup>26</sup>Universidad Christian  
54 Albrecht de Kiel, Germany, <sup>27</sup>Groningen Institute of Archaeology, University of Groningen,  
55 The Netherlands, <sup>28</sup>Aix Marseille Univ, CNRS, CCJ, Aix-en-Provence, France, <sup>29</sup>École  
56 française d'Athènes, <sup>30</sup>University of Victoria, Canada, <sup>31</sup>Texas State University,  
57 USA, <sup>32</sup>Rome University La Sapienza, Italy, <sup>33</sup>National Research Council, Italy, <sup>34</sup>Italian  
58 Heritage board, Italy, <sup>35</sup>Grupo de Investigación TELLUS. Departamento de Prehistoria y  
59 Arqueología. Universidad de Sevilla. España, <sup>36</sup>Laboratory of Biological Anthropology,  
60 ArchaeoScience, Globe Institute, University of Copenhagen, Denmark, <sup>37</sup>University of  
61 Bordeaux, France, <sup>38</sup>Institut national de recherches archéologiques préventives (Inrap),  
62 France, <sup>39</sup>UMR 6298, Université de Bourgogne, France, <sup>40</sup>Spanish National Research  
63 Council, Spain, <sup>41</sup>Laboratory of Human Genetics, Chiril Draganiuc Institute of  
64 Phthisiopneumology, Chişinău, Republic of Moldova, <sup>42</sup>Department of Biological  
65 Anthropology, University of Szeged, Szeged, Hungary, <sup>43</sup>Classical Studies, Wellesley  
66 College, Boston, USA, <sup>44</sup>Lund University, Sweden, <sup>45</sup>University of Bologna, Italy, <sup>46</sup>Ghent  
67 University, Department of Archaeology, The Netherlands, <sup>47</sup>National and Kapodistrian  
68 University of Athens, <sup>48</sup>Ministry of Culture (Greece), <sup>49</sup>Greek Archaeological  
69 Society, <sup>50</sup>German Archaeological Institute, Athens, Greece, <sup>51</sup>National Agency for  
70 Archaeology, Chişinău, Republic of Moldova, <sup>52</sup>Universidad de Cádiz,  
71 Spain, <sup>53</sup>Världskulturmuseerna, Sweden, <sup>54</sup>Complutense University, Madrid,  
72 Spain, <sup>55</sup>Department of Prehistory and Archaeology, University of Granada,  
73 Spain, <sup>56</sup>Department of Legal Medicine, Toxicology and Anthropology University of  
74 Granada, Spain, <sup>57</sup>Branch of Prehistory, Department of Historical Heritage, University of  
75 Jaen, Spain, <sup>58</sup>Department of Prehistory and Archaeology, Autonomous University of  
76 Madrid, Spain, <sup>59</sup>Department of Biology, Autonomous University of Madrid, Spain, <sup>60</sup>Royal  
77 Museums of Art and History Brussels, Belgium, <sup>61</sup>UMR 6298, Université de Bourgogne,  
78 France, <sup>62</sup>UMR 8164, Université de Lille, France, <sup>63</sup>Department of Archaeology, Ankara  
79 University, Ankara, Turkey, <sup>64</sup>Archaeological Superintendence of Verona, Italy, <sup>65</sup>Trace and  
80 Environmental DNA (TrEnD) Laboratory, School of Molecular and Life Sciences, Curtin  
81 University, Perth, Australia, <sup>66</sup>Departments of Integrative Biology and Statistics, UC  
82 Berkeley, USA, <sup>67</sup>GeoGenetics Group, Department of Zoology, University of Cambridge,  
83 Cambridge, UK

84

85 \*Corresponding author, #Shared second co-authors

86

## 87 **Summary**

88 The Indo-European languages are among the most widely spoken in the world, yet their early  
89 diversification remains contentious<sup>1-5</sup>. It is widely accepted that the spread of this language  
90 family across Europe from the 5th millennium BP correlates with the expansion and  
91 diversification of steppe-related genetic ancestry from the onset of the Bronze Age<sup>6,7</sup>.

92 However, multiple steppe-derived populations co-existed in Europe during this period, and it  
93 remains unclear how these populations diverged and which provided the demographic  
94 channels for the ancestral forms of the Italic, Celtic, Greek, and Armenian languages<sup>8,9</sup>. To  
95 investigate the ancestral histories of Indo-European-speaking groups in Southern Europe, we  
96 sequenced genomes from 314 ancient individuals from the Mediterranean and surrounding  
97 regions, spanning from 5,200 BP to 2,100 BP, and co-analysed these with published genome

98 data. We additionally conducted strontium isotope analyses on 224 of these individuals. We  
99 find a deep east-west divide of steppe ancestry in Southern Europe during the Bronze Age.  
100 Specifically, we show that the arrival of steppe ancestry in Spain, France, and Italy was  
101 mediated by Bell Beaker (BB) populations of Western Europe, likely contributing to the  
102 emergence of the Italic and Celtic languages. In contrast, Armenian and Greek populations  
103 acquired steppe ancestry directly from Yamnaya groups of Eastern Europe. These results are  
104 consistent with the linguistic Italo-Celtic<sup>10,11</sup> and Graeco-Armenian<sup>1,12,13</sup> hypotheses  
105 accounting for the origins of most Mediterranean Indo-European languages of Classical  
106 Antiquity. Our findings thus align with specific linguistic divergence models for the Indo-  
107 European language family while contradicting others. This underlines the power of ancient  
108 DNA in uncovering prehistoric diversifications of human populations and language  
109 communities.

110

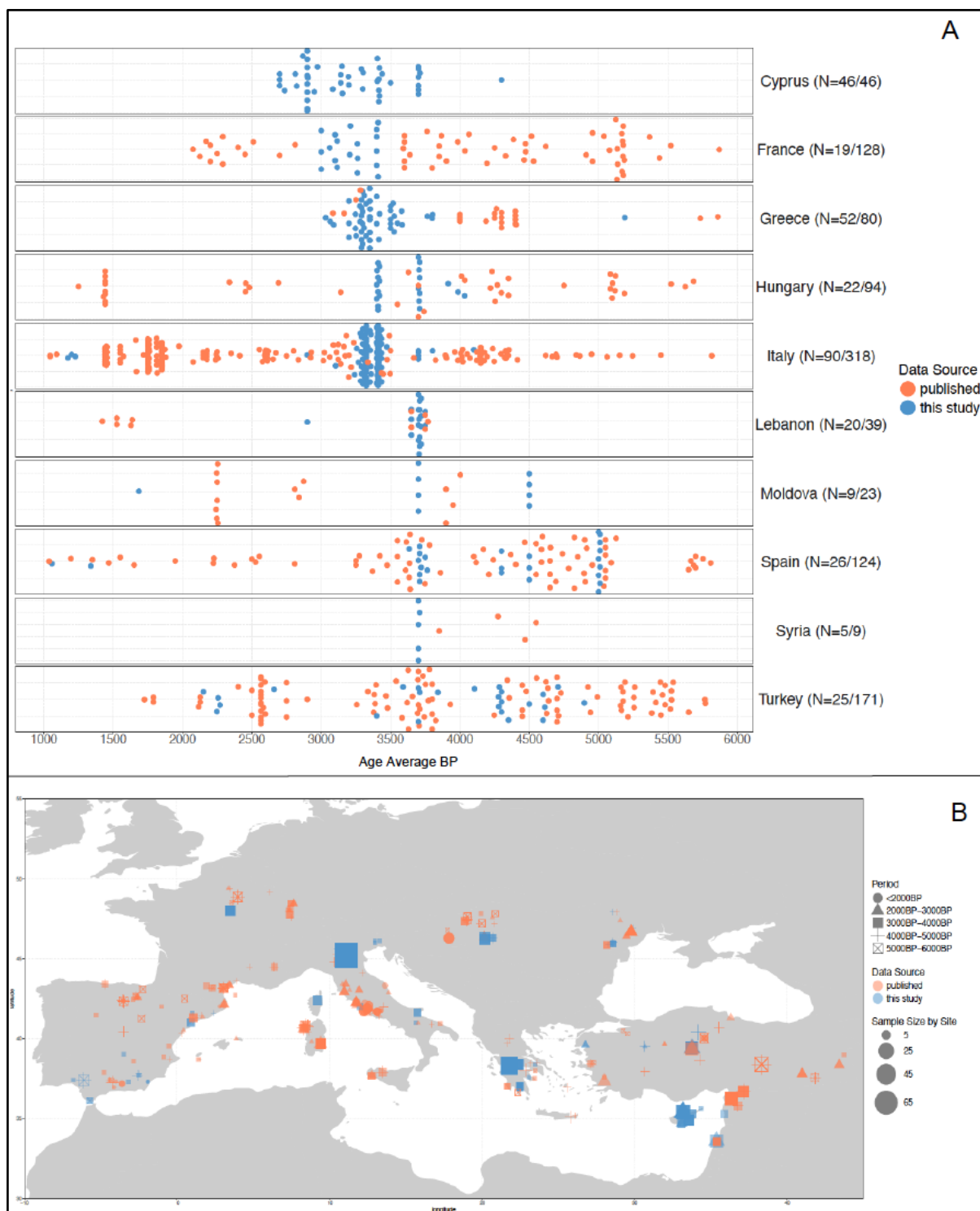
## 111 **Introduction**

112 From 5,000 BP, large-scale human migrations significantly restructured the genetic makeup  
113 of human populations across Eurasia<sup>6,7</sup>. Various pulses of steppe-related ancestry, associated  
114 with the mobile pastoralists of the Early Bronze Age Yamnaya culture, spread across vast  
115 distances, leaving distinct cultural and genetic signatures. These migrations likely also played  
116 a key role in the prehistoric dispersal of the Indo-European language family. However, steppe  
117 ancestry reached the various regions of Western Eurasia by different mechanisms and at  
118 different times. In various historically Indo-European-speaking regions of Europe, the arrival  
119 of steppe ancestry was mediated by populations associated with the archaeological complexes  
120 of the Corded Ware (CWC) (5,000–4,350 BP)<sup>6,7,14</sup> and Bell Beakers (4,800–4,300/3,800  
121 BP)<sup>15–17</sup>. However, the extent to which similar dynamics occurred in the Eastern  
122 Mediterranean and adjacent Western Asian populations remains less well-understood.  
123 Moreover, while steppe ancestry has previously been detected in prehistoric Iberians<sup>18,19</sup>,  
124 Italians<sup>20–23</sup>, Greeks<sup>24–27</sup>, and Caucasians<sup>24</sup>, questions remain regarding the interrelatedness of  
125 the proximal source populations in the context of the Mediterranean region at large.

126

127 The spread of steppe ancestries is closely tied to the diversification of the Indo-European  
128 protolanguage into its historically attested subgroups<sup>28</sup>. In the Mediterranean, important Indo-  
129 European languages from Classical Antiquity include Gaulish, Latin, Greek, and Armenian,  
130 the latter being indigenous to the South Caucasus and Eastern Anatolia. For these, multiple  
131 competing phylogenetic linguistic models have been proposed<sup>29</sup> (Linguistic Supplementary  
132 2–4). The so-called Indo-Greek hypothesis groups Greek as well as the closely related  
133 Phrygian with Indo-Iranian<sup>30</sup>, while the competing Graeco-Armenian<sup>12,13</sup> hypothesis posits  
134 that Greek forms a subclade with Armenian, possibly also including Albanian (“Balkan Indo-  
135 European”). Similarly, the Italic Indo-European subgroup, which is ancestral to Latin, has  
136 been variously grouped with Celtic and Germanic, giving rise to the traditionally popular  
137 Italo-Germanic<sup>31</sup> and contrastive Italo-Celtic<sup>10,32</sup> hypotheses. While relative linguistic  
138 consensus exists on the Graeco-Armenian and Italo-Celtic hypotheses, these are not  
139 unchallenged<sup>33,34</sup>. More fundamentally, the lack of a definitive linguistic consensus model for  
140 the Indo-European diversification constitutes a key obstacle to the interdisciplinary study of  
141 Indo-European language dispersal.

142 Here, we investigate the various sources of steppe ancestry along the entire northern border  
143 of the Mediterranean region, so as to establish the most parsimonious divergence model for  
144 the Indo-European languages in this area. We present new whole-genome data from 314  
145 ancient individuals (>0.1X genomic coverage) from Southern and Central-Eastern Europe, as  
146 well as from the Eastern Mediterranean comprising Spain, France, Italy, Hungary, Moldova,  
147 Greece, Cyprus, Turkey, Syria, and Lebanon. These individuals mostly belong to the Bronze  
148 Age but span a time frame from 5,200 BP to 2,100 BP (Fig. 1) (Genetics and Strontium  
149 Supplementary S3; Supplementary Table S1). We also provide strontium isotope signatures  
150 from 224 individuals and radiocarbon-dating of 144 individuals (Supplementary Tables S8  
151 and S9). Utilizing identity-by-descent (IBD) inferred admixture modelling<sup>17,35</sup> with specific  
152 source populations, we obtained enhanced resolution of genetic ancestries, enabling us to  
153 differentiate diverse or common ancestries. Furthermore, we combined our strontium isotope  
154 data with our genetic results to deepen the understanding of mobility patterns over time.  
155



156  
157 Fig. 1. Distribution of ancient individuals distributed by country (A), (N= the number of  
158 individuals in this study/total number of individuals in dataset), and locality shown on the  
159 map (B). We only demonstrate the individuals limited by time frame (6,000–1,000 cal BP) to  
160 avoid overlapping.

### 161 **Population structure and overview of IBD clusters**

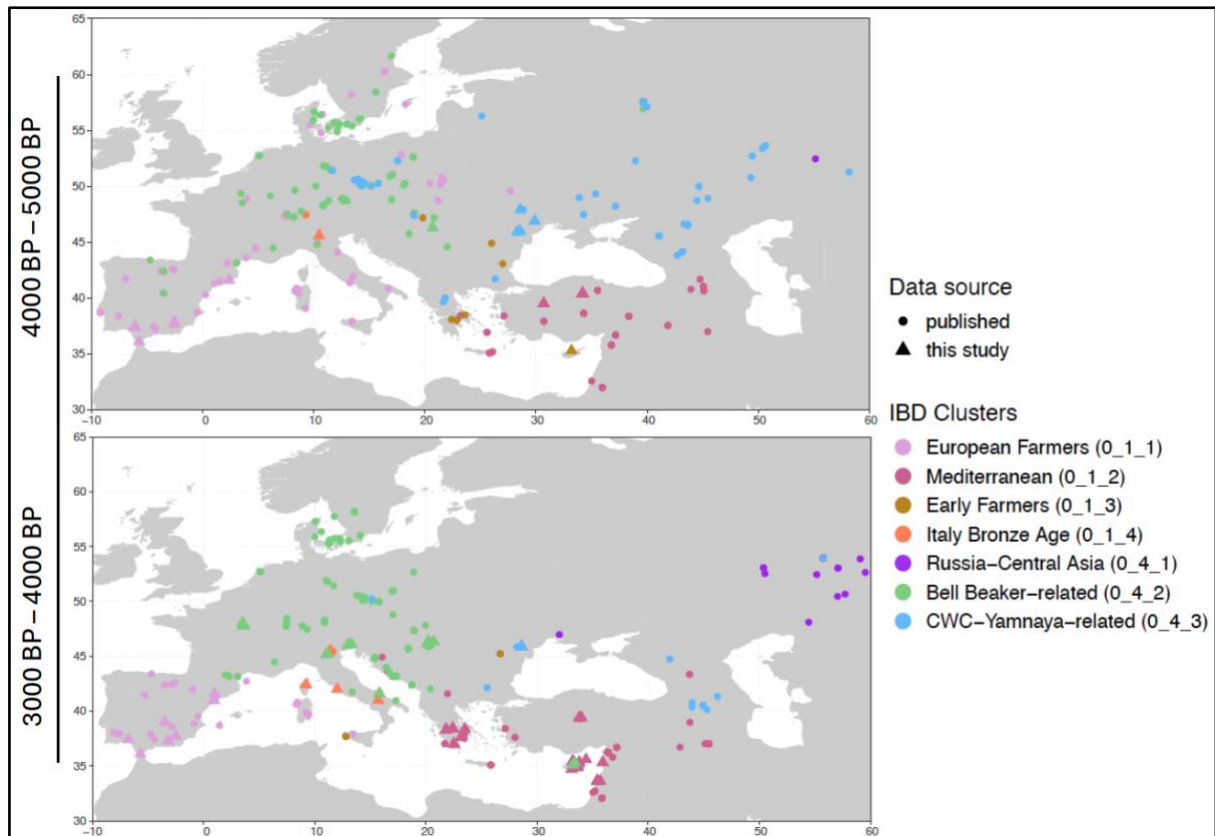
162 We generated the dataset by shotgun-sequencing genomic DNA extracted from 314 ancient  
163 individuals (Genetics and Strontium Supplementary S2; S3). This was combined with 992  
164 published shotgun-sequenced genomes and 1,097 genome-wide SNP-captured ancient

165 individuals (Supplementary Table S2). These were all imputed using established cut-offs<sup>17,35–</sup>  
166 <sup>37</sup> of 1X capture and 0.1X shotgun data resulting in a total combined dataset of 2,403 imputed  
167 genomes across 643,430 SNP (single nucleotide polymorphism) sites (Genetics and  
168 Strontium Supplementary S3). We called IBD segments between pairs of individuals using  
169 IBDseq<sup>38</sup>, and employed a network-based hierarchical approach<sup>39</sup> to obtain fine-scale  
170 population clusters based on the total length of shared IBD segments between pairs of  
171 individuals (Genetics and Strontium Supplementary S5).

172  
173 Through this clustering approach, we detected six deep clusters corresponding to distinct  
174 ancestries previously related to the prehistoric formation of Eurasia: “Farmer-related (Cluster  
175 0\_1)”, “European-Hunter-Gatherers (Cluster 0\_2)”, “Caucasus – Iran (Cluster 0\_3)”,  
176 “Steppe-related (Cluster 0\_4)”, “Central Asia – Siberia (Cluster 0\_5)”, “Moroccan-Hunter-  
177 Gatherers (Cluster 0\_6)” (Extended Fig. 1 and Fig. 2; Supplementary Table S3; Genetics and  
178 Strontium Supplementary Table S5.1). We deeply investigated “Farmer-related (0\_1)” and  
179 “Steppe-related (0\_4)” clusters, since the ancient individuals reported here fell within a  
180 number of subclusters of these main clusters. The “Farmer-related (0\_1)” cluster is further  
181 divided into four subclusters, distinguishing between East (0\_1\_2 and 0\_1\_3) and West  
182 Mediterranean (0\_1\_1 and 0\_1\_4). The “Steppe-related (0\_4)” cluster comprises three  
183 subclusters with individuals from Central Asia and Europe, “Russia-Central Asia (0\_4\_1)”,  
184 “Bell Beaker-related (0\_4\_2)”, and “CWC-Yamnaya-related (0\_4\_3)” (Fig. 2; Genetics and  
185 Strontium Supplementary Table S5.2; Extended Data Fig. 3). The occurrence of these  
186 subclusters refers to the spread of Yamnaya-related ancestry both eastward and westward, as  
187 well as the formation of new genetic profiles throughout Europe and Central Asia<sup>6,7,40–42</sup>.  
188 When focusing specifically on the time frame of this study (between 5,000 BP and 3,000 BP),  
189 we observe a distinct pattern emerging between the 5th and 4th millennia in terms of the  
190 distribution of these Steppe-related clusters (Fig. 2). The distribution of the “CWC-Yamnaya-  
191 related (0\_4\_3)” cluster originating from the Pontic Steppe, extends into Central Eastern  
192 Europe and Northern Greece before the 4th millennium BP. This cluster persisted west of the  
193 Black Sea and in Armenia after the 4th millennium BP, whereas the “Bell Beaker-related  
194 (0\_4\_2)” cluster became prevalent in Europe. Further south to Greece, during this time we  
195 see a shift of the border between the “Steppe-related (0\_4)” cluster and the “Farmer-related  
196 (0\_1)” cluster. The first instance we detect of cross-border interactions occurs during an  
197 advanced phase of the 4th millennium BP. Here we find that two published Greece Bronze  
198 Age individuals<sup>43,44</sup> (Log04 and G23) fall within the “CWC-Yamnaya-related (0\_4\_3)”  
199 cluster, specifically within subclusters corresponding to the “Yamnaya-related (0\_4\_3\_1)”  
200 culture and the “Armenian-Caucasus (0\_4\_3\_3)”, respectively (Genetics and Strontium  
201 Supplementary Fig. S5.9). At the east end of this border zone, individuals from Moldova fall  
202 within the subclusters of the “Steppe-related (0\_4)” cluster, whereas all Eastern  
203 Mediterranean individuals including those from Greece fall within the subclusters of  
204 “Farmer-related (0\_1)” cluster. Moreover, we detect the newly sequenced Early Bronze age  
205 individuals from Moldova within the subcluster of “Yamnaya-related (0\_4\_3\_1)”, while the  
206 Middle Bronze Age individuals within the subcluster corresponding to “Corded Ware  
207 (0\_4\_3\_2)” individuals like previously published Moldovan individuals<sup>24</sup> (Supplementary  
208 Table S3). Additionally, we found all individuals from Anatolia, Cyprus, and Levant within

209 the “Mediterranean (0\_1\_2)” cluster during the 3rd and 4th millennia, except one Early  
210 Bronze age individual and two European outliers from Cyprus are found within “Early  
211 farmers (0\_1\_3)” and “Bell Beaker-related (0\_4\_2)” throughout the 4th and 3rd millennia  
212 BP.

213



214  
215 Fig. 2. Geographical distribution of the IBD clusters of individuals in the 5th and 4th  
216 millennia BP.

### 217 **Southern and Central Eastern Europe**

218 The spread of steppe-related ancestry in Europe has been traced to the Yamnaya populations  
219 and their subsequent admixture with local European populations after admixing with  
220 Globular Amphora culture-related (GAC-related) populations in the east<sup>6,7</sup>. To differentiate  
221 between closely-related steppe ancestries in Southern European populations, we first  
222 generated an ancestry “palette” for every ancient individual in the database, representing the  
223 individual sharing patterns with all clusters in the dataset (Genetics and Strontium  
224 Supplementary S6.2). By designating specific clusters as sources, we are able to employ IBD  
225 Mixture Modelling<sup>17,35</sup> to model the palettes of target individuals as best-fitting combinations  
226 of the source palettes. Of relevance here were various steppe-related populations from the  
227 Pontic Steppe and Europe (Genetics and Strontium Supplementary S6.2). We built on from a  
228 basic source set<sup>35</sup> that includes the Hunter-Gatherer and early Farming-related populations  
229 involved in the formation of European population structure, including a series of outgroups  
230 from Eurasia and Africa (Supplementary Table S4). We then progressively added a series of  
231 more recent populations as sources. By including two steppe sources, early Corded Ware

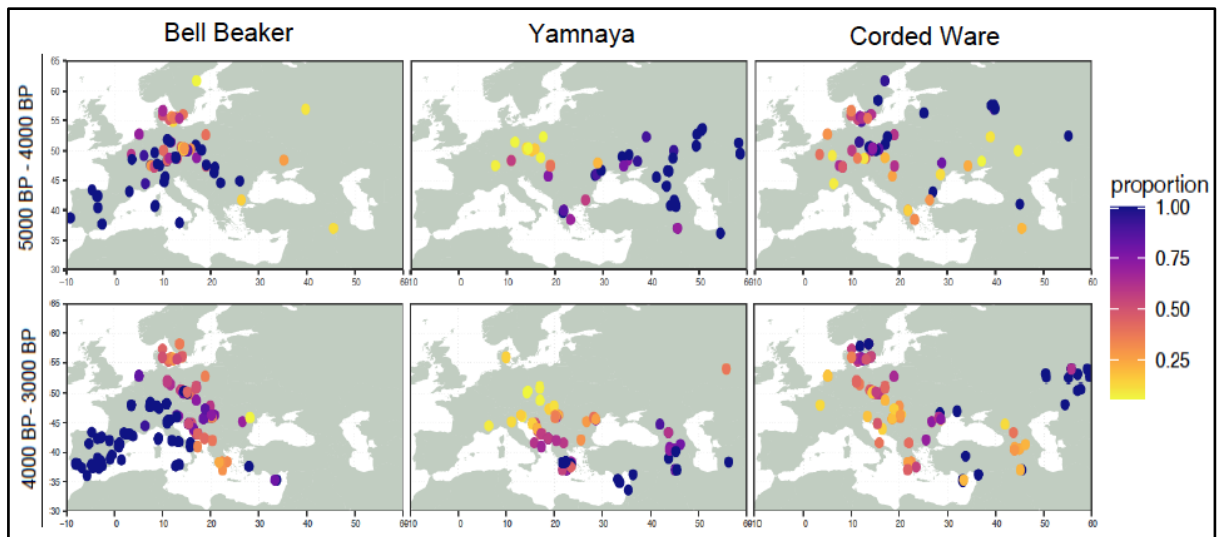
232 culture (CWC) and Yamnaya (Genetics and Strontium Supplementary S6.2; Supplementary  
233 Table S4), we revealed distinctly dissimilar patterns, contrasting Greece with France, Spain,  
234 and Italy, reflecting the expansion of two separate steppe ancestries—one mediated through  
235 Corded Ware groups and the other derived from Yamnaya populations (Extended Data Fig. 4;  
236 Genetics and Strontium Supplementary S6.2). The distinction between these two expansions  
237 is also distinguishable by the Farming-related ancestry they carry. The arrival of steppe  
238 ancestry in Europe was accompanied by GAC-related Farming ancestry: this farming  
239 ancestry is present in the Corded Ware carrying GAC-related ancestry, but absent in Greece  
240 (Genetics and Strontium Supplementary S6.2; Fig. S6.18; Supplementary Table S5). We thus  
241 detect a clear division in steppe ancestry between the Eastern and Western Mediterranean.

242 By introducing a third steppe-related source from the “Bell Beaker-related (0\_4\_2)” cluster,  
243 Southwestern Europe shows ancestry that is BB-related, while that of Greece is Yamnaya-  
244 related, particularly in the Peloponnese (Figs. 3 and 4). The Balkans, however, exhibit mixed  
245 ancestries of CWC/BB and Yamnaya, suggestive of interactions between migrants from the  
246 Pontic Steppe and with local populations associated with CWC/BB-related ancestry, or an  
247 unsampled steppe source emerging west of the Black Sea (Fig. 4). During the Mycenaean  
248 period in Greece (or Late Helladic period, c. 3,700–3,200 BP), Yamnaya ancestry became  
249 widespread, even extending into the Eastern Mediterranean (Fig. 3). We interpret this as  
250 evidence of a direct Yamnaya migration into the Peloponnese, supported also by the detection  
251 of the typical Yamnaya Y-chromosome haplogroup lineage R1b-Z2103 (Z2108)<sup>24,45</sup> in two  
252 individuals from Ayios Vasileios, one from the Mycenaean cemetery Voudeni, and one from  
253 Kirrha (Fig. 5, Supplementary Table S7). However, the dominant Y-haplogroup lineage in  
254 Greece, J2a-Y7011, is also detected in individuals from Lapithos, Iron Age Cyprus, who are  
255 additionally characterized by non-local strontium isotope signatures. This haplogroup is also  
256 found in Szőreg, Hungary, suggesting a connection between the Balkans and the Eastern  
257 Mediterranean, as this lineage is present in early Balkan populations. To identify a potential  
258 connection with the Balkans, we present new data from Middle and Late Bronze Age  
259 Hungary revealing a complex population structure. We observe a mixed pattern of BB and  
260 Yamnaya components in Hungary during the Bronze Age (Fig. 4; Genetics and Strontium  
261 Supplementary S6.3). This suggests either another source population for those individuals or  
262 an admixture between two groups.

263 Italian Bronze Age individuals show at least three distinct admixture patterns. The first is  
264 similar to the Bronze Age Bell Beaker-related groups of France and Spain. Interestingly, that  
265 group includes all individuals from Central Italy and Corsica, a few individuals from Olmo,  
266 and one published Bell Beaker-context individual from Northern Italy<sup>15</sup> (I2478, 4,006 BP). A  
267 second group, primarily from Olmo, shows increased Neolithic farmer ancestry with lower  
268 amounts of steppe ancestry, suggesting an ongoing local admixture process. A third limited  
269 group with increased Yamnaya ancestry, similar to Balkan and Greek Bronze Age  
270 populations, is observed in individuals from the Adriatic coast of Italy (Fig. 4; Genetics and  
271 Strontium Supplementary S6.2, CGG\_2\_100646, NEO806, R1)<sup>6,20</sup>. Interestingly, three Early  
272 Bronze Age individuals from Northeastern Italy carry an additional CWC or Yamnaya  
273 component, and such a profile is also observed in the individuals from Hungary (Fig. 4;

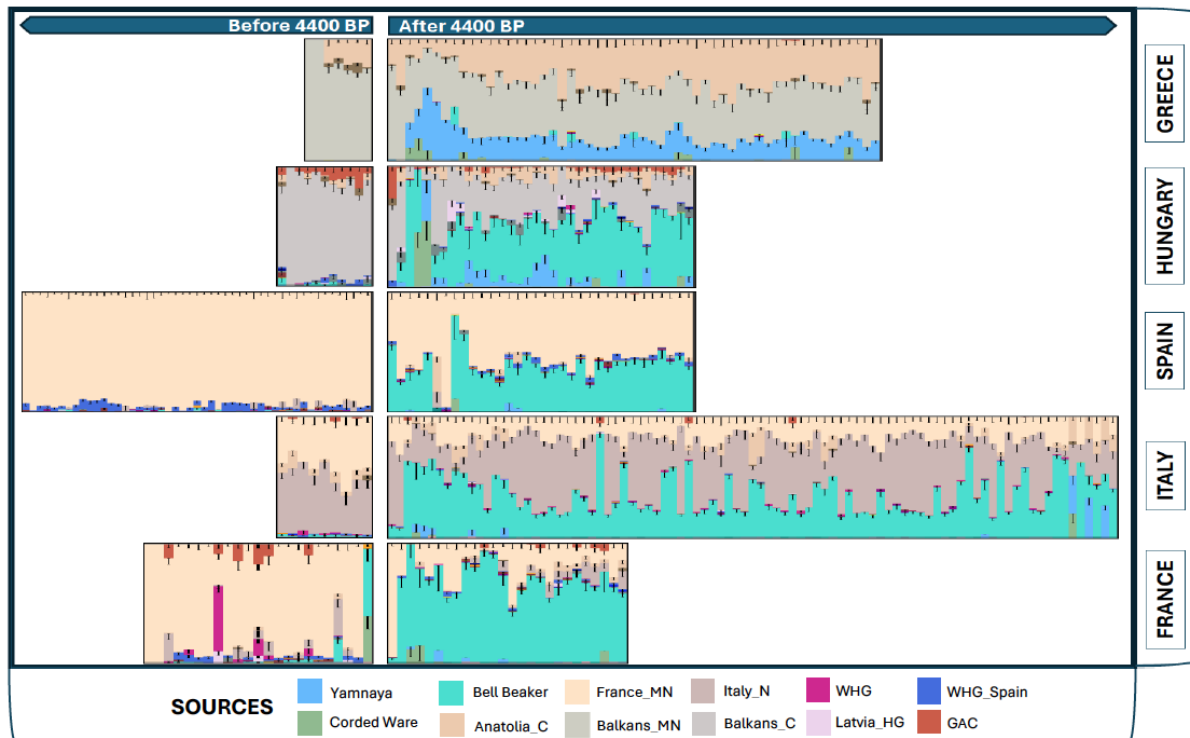
274 Genetics and Strontium Supplementary S6.3). These individuals, found in monumental  
275 tumulus graves and a coffin burial (CGG\_2\_022653), likely held a special social or cultural  
276 status, as suggested by the complexity of the tumuli and the ritual activities linked to those  
277 monuments<sup>46</sup> (Archaeological Supplementary 2.7.4; 2.7.9–11).

278 To compare the BB source with local contribution, we added two Early Bronze Age  
279 individuals from Northeastern Italy as an ancestral source to model Late Bronze Age Italian  
280 individuals (Genetics and Strontium Supplementary S6.2, Fig. S6.13; Supplementary Table  
281 S5). We found this new, local source to better model the ancestry component previously  
282 modelled by the BB source, except in a few individuals who show a mixed pattern of Bell  
283 Beaker and local ancestry, rejecting the model. This local source also replaced the steppe  
284 proportion in Corsica and Sicily, suggesting a maritime spread of steppe ancestry through the  
285 Mediterranean (Genetics and Strontium Supplementary Fig. S6.15). This ancestry also  
286 increased in Late Bronze Age individuals from Migennes, France (Genetics and Strontium  
287 Supplementary S6.2; Fig. S6.15), suggesting later formation in steppe ancestry in Western  
288 Europe could also connect the Balkans and Italy.



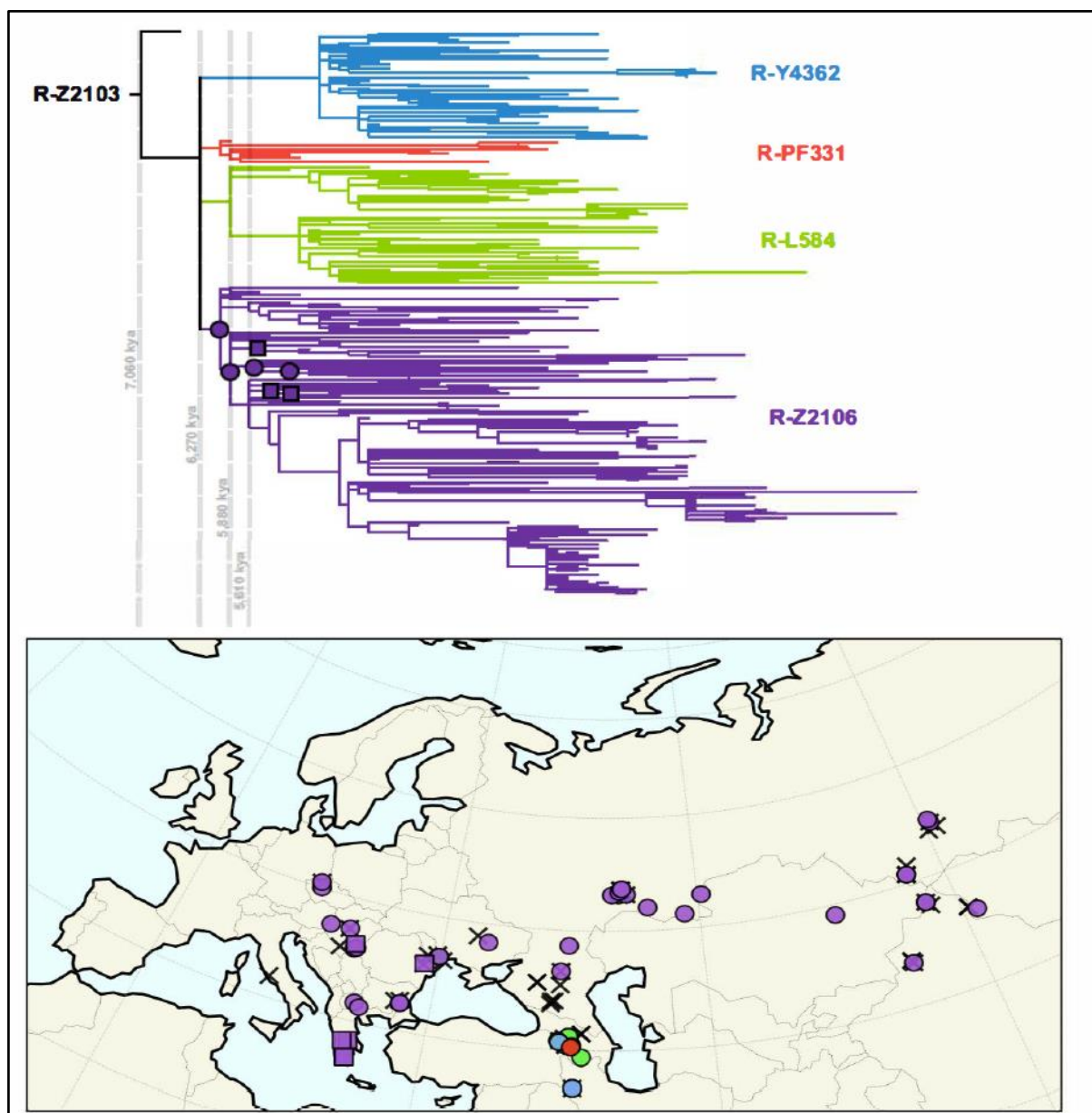
289  
290 Fig. 3. Distribution of Bell Beaker-derived and Yamnaya-derived ancestry proportions  
291 obtained from the IBD admixture model. The proportion of each steppe source is  
292 standardized by the total steppe contributions, i.e. the sum of Corded Ware, Bell Beaker and  
293 Yamnaya\_Samara contributions.

294



295  
296  
297  
298

Fig. 4. Ancestry bar plots generated for each individual using source population proportions of IBD admixture modelling sorted by time BP and divided into two time series, before and after 4,400 BP, illustrating a Southern Europe split (Italy, France, and Spain vs Greece).



299  
300 Fig. 5. Phylogeography of Y-chromosome haplogroup R-Z2103. Phylogram of haplogroup  
301 R-Z2103, with branch lengths proportional to SNP number, built from a dataset of unique  
302 variants in private datasets. The haplogroup resolves in a four-way polytomy. We plotted all  
303 occurrences older than 2,000 years ago in the ancient DNA record in Western Eurasia on an  
304 Albers Equal Area map, colour-coded by clades downstream of R-Z2103. Only haplogroup  
305 R-Z2106 extends beyond the Caucasus and Northern Iran, and we indicate phylogenetic  
306 position of individuals in the tree. Crosses mark R-Z2103 individuals with uncertain clade  
307 assignment due to low coverage, and squares indicate individuals from Greece, Moldova and  
308 Hungary generated in this study. Dotted lines denote split time estimates of key haplogroups,  
309 calculated using rho statistics.

310

### 311 **Eastern Mediterranean**

312 The Caucasus region has experienced a complex history of human migrations, interactions,  
313 and cultural exchanges, marked by admixture among diverse population groups. With the  
314 spread of farming to Iran and the Caucasus, populations in the region were shaped by  
315 admixture between early Neolithic farmers from the Fertile Crescent and two genetically

316 similar population groups, Iran Neolithic and Caucasus Hunter-Gatherers (CHG)<sup>24,41</sup>.  
317 Additionally, the expansion of the Kura-Araxes culture during the 5th millennium BP  
318 connected the Caucasus with the Levant and Mesopotamia through extensive trade  
319 networks<sup>47</sup>. Interaction between Anatolia and the Caucasus increased during the Chalcolithic  
320 and the Bronze Ages, leading to the spread of CHG ancestry. It also diffused into the  
321 Mediterranean, an early indication of which is found in Anatolian farmer groups from  
322 Tepecik-Çiftlik<sup>21,25,43,44</sup>.

323

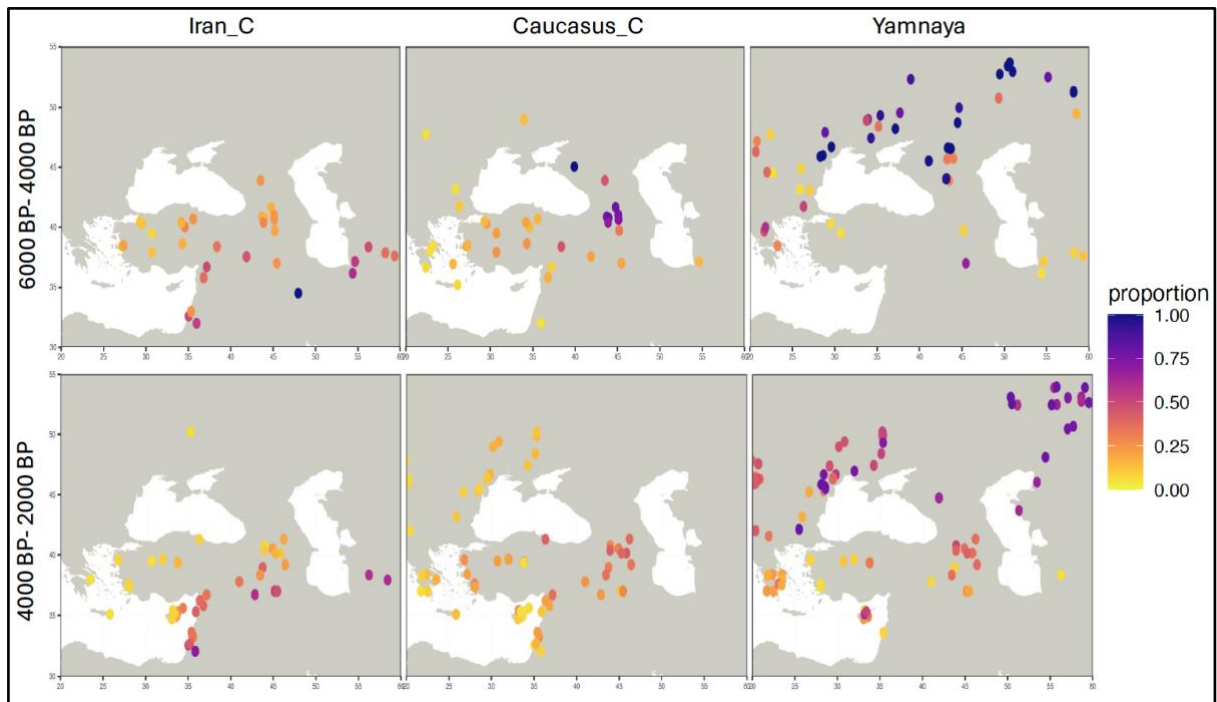
324 Here, we report new genomic data from 25 individuals from Anatolia, including individuals  
325 from Resuloğlu, an Early Bronze Age cemetery associated with Hattian culture, from  
326 Western Anatolia (Küllüoba, Keçiçayırı and Antandros), and from Central Anatolia (Kaman  
327 Kalehöyük), representing both the Bronze and Iron Ages. To assess the population structure  
328 of Anatolian individuals, we applied IBD mixture modelling, removing the CWC and BB  
329 sources, and instead progressively added a series of eastern ancestry sources from Iran and  
330 Caucasus Chalcolithic individuals, also including a Central Anatolian farmer source from  
331 Tepecik (Genetics and Strontium Supplementary S6.4; Supplementary Table S4). In our base  
332 modelling, all Chalcolithic and Bronze Age genomes from Anatolia appeared as mixtures of  
333 three components: local Neolithic farmers (from Tepecik), CHG (Caucasus Hunter-Gatherer)  
334 and a small proportion of Iranian Neolithic ancestry, with the exception of a minor proportion  
335 of EHG (Eastern Hunter-Gatherer) found in some individuals from Kalehöyük, Arslantepe,  
336 and Western Anatolia (Genetics and Strontium Supplementary Fig. S6.25; Supplementary  
337 Table S5). Together with CHG, this proportion was slightly elevated in the Iron Age. To  
338 understand the origin of the CHG and EHG contributions, we added Chalcolithic individuals  
339 from Iran and the Southern Caucasus as geographically proximal sources (Supplementary  
340 Table S4). The model revealed that Anatolian Chalcolithic and Bronze Age individuals  
341 received various proportions of ancestry from both Iran and the Southern Caucasus, with a  
342 higher amount of Iranian ancestry in Eastern than Western and Central Anatolia (Fig. 6).  
343 Moreover, the Chalcolithic Caucasus proportion increased in Central Anatolia, such as in  
344 Kaman Kalehöyük, in the Middle Bronze Age, and Iranian ancestry is minimal or absent in  
345 Western Anatolia. Admixture patterns in Eastern Anatolian individuals resemble those of  
346 Bronze Age individuals from Lebanon. Notably, Chalcolithic and Bronze Age individuals  
347 from Arslantepe, who received both Caucasus and Iranian ancestries, played a pivotal role  
348 within the Kura-Araxes region, spanning from Southern Caucasus to Levant.

349

350 To distinguish increased proportions of steppe ancestry in the Iron Age, we included multiple  
351 steppe sources (Yamnaya, CWC, BB) that revealed different signatures depending on the  
352 geographical location. In the newly sequenced Iron Age samples from Central and  
353 Northwestern Anatolia (Kalehöyük, Antandros and Keçiçayırı), we observed minor  
354 proportions of steppe ancestry with the pattern found in Balkans/Greek Late Bronze Age and  
355 probably reflects migrations from the Balkans (Genetics and Strontium Supplementary Fig.  
356 S6.37; S6.38; S6.39; Supplementary Table S5). Given that the individual from Keçiçayırı  
357 (CGG\_2\_022162) was unearthed from the Phrygian valley, the appearance of this ancestry  
358 may be associated with the emergence of the Phrygian state during the late 4th millennium  
359 BP<sup>48</sup> (Archaeology Supplementary 2.12.5; Linguistic Supplementary 3.3).

360

361 To trace the interaction between Eastern Mediterranean and Southern Europe, we sequenced  
362 Bronze Age and Iron Age individuals from Cyprus and Lebanon. Cypriot individuals make  
363 up a highly diverse group connecting all of the Eastern Mediterranean (Fig. 6). Our results  
364 suggest that Cyprus, and in particular its coastal towns, were a genetic and cultural “melting  
365 pot” during the Bronze Age. The Middle and Late Bronze Age individuals from Cyprus show  
366 a genetic pattern similar to that of individuals from Lebanon and Eastern Anatolia Bronze  
367 Age, while one (the earliest) individual from Karavas (CGG\_2\_022531), dated to 5,000–  
368 4,500 BP (Archaeology Supplementary 2.2.3), show extra early Anatolian farmer ancestry. A  
369 long-distance genetic connection is observed at Hala Sultan Tekke, one genetic outlier  
370 carrying Balkan/Aegean ancestry similar to that of Late Bronze Age Greece individuals.  
371 Thus, our dataset suggests close contacts of Cyprus with the Levant and the Aegean during  
372 the Bronze Age and even earlier periods. During the subsequent Iron Age, data from both  
373 Lapithos and Amathus suggest an increasingly uniform population across the island.  
374 Moreover, Iron Age genomes show a formation similar to populations of the Aegean and  
375 Western Anatolia Iron Age, carrying a small proportion of Yamnaya ancestry which reflects  
376 Greece ancestry (Fig. 6). Additionally, there is genetic evidence of long-distance interaction  
377 with Northern Europe, as seen in a Scandinavian genetic outlier (CGG\_2\_022535) from a  
378 rock-cut tomb at Vounous Bellapais, excavated by the Swedish-Cyprus expedition and dated  
379 to c. 4,000–3,800 BP. (Genetics and Strontium Supplementary Fig. S6.45, Archaeology  
380 Supplementary 2.2.7). This outlier clusters with Scandinavian Bronze Age individuals and,  
381 intriguingly, this origin is also supported by the Y-haplogroup I1 and by a non-local highly  
382 radiogenic strontium isotope signature compatible with some parts of Scandinavia (Genetics  
383 and Strontium Supplementary S10; Supplementary Table S8). The implications of this  
384 observation are not conclusive, since we do not have radiocarbon dating from this individual.  
385 Due to its rich copper sources, Cyprus maintained extensive trading networks with most of  
386 the Mediterranean; this situation is well mirrored by the presence of Anatolian, Levantine,  
387 Greek, and European ancestries in the Cypriot genomes.  
388



389

390 Fig. 6. Distribution of Iran Chalcolithic (Iran\_C), Caucasus Chalcolithic (Caucasus\_C) and  
391 Yamnaya-related ancestry proportions obtained from the IBD admixture model.

392

### 393 **Increasing mobility towards the end of the Bronze Age**

394 Considering the highly dynamic Bronze Age period in the Mediterranean, we analysed  
395 strontium isotope ratios from 224 ancient individuals across Cyprus, Greece, Italy, and Spain  
396 to trace the mobility of individuals (Genetics and Strontium Supplementary S10;  
397 Supplementary Table S8). We identified 56 individuals as potential non-locals and compared  
398 them to their genetic profiles.

399 In Greece, nine individuals from Kirrha, Voudeni, Eleon, and Apollo Maleatas returned non-  
400 local strontium isotope signatures. These might have travelled from regions within Greece  
401 where strontium isotope baselines are slightly more radiogenic. Among these nine  
402 individuals, we only have genetic data for four individuals, as several closely-related  
403 individuals were removed to avoid skewing the IBD admixture analysis. Genetically, these  
404 four individuals are similar to others from the same area, except for one individual from  
405 Eleon, who carries a small proportion of Bell Beaker ancestry reflecting a connection with  
406 Central Eastern Europe. Interestingly, two individuals (from Kirrha and Apollo Maleatas,  
407 Genetics and Strontium Supplementary S10; Supplementary Table S6) with non-local  
408 signatures have second-degree relatives who fall within the local baseline, indicating different  
409 mobility patterns within the same family. However, the relatives falling within the local  
410 baseline could also have originated from a different region given the large overlap in  
411 strontium signatures across Greece (Genetics and Strontium Supplementary S10;  
412 Supplementary Table S8). Finally, the father of a non-local individual from Apollo Maleatas  
413 (CGG\_2\_023933) differs from other Late Bronze Age Greece individuals by having a higher  
414 Yamnaya proportion compare to Late Bronze Age individuals, similar to the Middle Bronze  
415 Age individuals (>~3,800 BP).

416

417 In Cyprus, despite the high levels of genetic variation, only four Cypriot samples are  
418 characterized by non-local strontium isotope signatures (Supplementary Table S8). The  
419 genetic outlier from Vounous-Bellapais (CGG\_2\_022535) with unusual Scandinavian  
420 ancestry is also confirmed by the strontium isotope results. Another non-local Iron Age  
421 individual (CGG\_2\_022526) exhibits an admixture signature consistent with Greece Late  
422 Bronze Age (Genetics and Strontium Supplementary Fig. S6.45; S6.46). In contrast, the  
423 remainder of the outliers resemble populations from the Anatolia/Levant Bronze Age  
424 (Genetics and Strontium Supplementary S10). For many genetic outliers in which strontium  
425 data is also presented, long-term mobility was not detected (Genetics and Strontium  
426 Supplementary S10).

427

428 We additionally identified several individuals with non-local strontium isotope signatures  
429 within Italy, consistent with a highly dynamic society. These individuals come from Olmo,  
430 Pian Sultano, Scalvinetto, and from Coppa Nevigata. One individual (CGG\_2\_101264) from  
431 Pian Sultano had a bone and a tooth sample investigated for strontium isotopes, returning one  
432 signature (tooth) within their local strontium isotope baseline and another (petrous) outside of  
433 it. These results suggest that she did not originate from the area of Pian Sultano, but moved  
434 there in her early childhood/adolescence (Genetics and Strontium Supplementary S10). Yet  
435 another interesting non-local (CGG\_2\_022591), from Olmo, carries the highest proportion of  
436 steppe-related ancestry found among Italian Bronze Age individuals, displaying a different  
437 composition farmer and steppe proportion than other Italian Bronze Age individuals, carrying  
438 no Italian Neolithic ancestry. Another non-local from Coppa Nevigata (CGG\_2\_100646)  
439 genetically can be distinguished from the other Italian individuals by carrying Yamnaya-  
440 related steppe ancestry and Mediterranean farmer proportion similar to the Aegean/Balkans.  
441 Overall, non-local individuals of Bronze Age Italy cannot be distinguished from each other in  
442 terms of the steppe source obtained from IBD admixture modelling except for two cases from  
443 Coppa Nevigata and Olmo (CGG\_2\_100646 and CGG\_2\_022591; Genetics and Strontium  
444 Supplementary Fig. S6.13). Moreover, most of the non-local individuals in Olmo come from  
445 a group among Bronze Age Italians that exhibit a lower proportion of steppe-related ancestry,  
446 thus resembling individuals from Mediterranean islands (Sicily, Corsica)(Genetics and  
447 Strontium Supplementary Fig. S6.13; S6.15).

448

449 In Spain, we identified individuals with non-local strontium isotope signatures genetically  
450 belonging to two different groups: one with local farmer-related ancestry, and the other group  
451 dated to the Early Bronze Age (~4,300 BP) showing Bell Beaker-related ancestry. Since we  
452 observe only small-scale variation among farmer populations in Spain, the outliers with  
453 farmer ancestry may have originated from other farmer sites within the region.

454

## 455 **Discussion**

456 To elucidate the genetic formation and infer the divergence of the Indo-European language  
457 family, we investigated the contribution of potential ancestral source populations to the wider  
458 Mediterranean region, including from areas in which the Italic, Celtic, Greek, and Armenian  
459 languages are historically spoken. The genetic results obtained by IBD admixture modelling

460 with putative steppe-related source populations support a deep divergence between Eastern  
461 and Western Mediterranean Indo-European-speaking groups through the detection of  
462 Yamnaya-related and Bell Beaker-related steppe ancestry respectively (Fig. 4). This  
463 divergence supports the linguistic hypotheses on the existence of the so-called Graeco-  
464 Armenian and Italo-Celtic subclades. In contrast, it disqualifies the rival cladistic hypotheses  
465 known as Indo-Greek and Italo-Germanic, the steppe ancestry among the populations of  
466 historically Germanic- and Indo-Iranian-speaking areas previously having been characterized  
467 as primarily Corded Ware-related<sup>35</sup>.

#### 468 **South-central Europe: Italic**

469 Prior to its Romanization, Italy was characterized by marked linguistic complexity,  
470 harbouring multiple Indo-European and non-Indo-European language groups<sup>49</sup>. The Italic  
471 languages, including Latin, Oscan, Umbrian, and possibly Venetic, are Indo-European in  
472 origin, just like the more distantly related Cisalpine Gaulish, Messapic, and Greek. Etruscan  
473 and Rhaetic, on the other hand, belong to the so-called Tyrrhenian family (Linguistic  
474 Supplementary 2). Due to this complexity, tracing the spread of the Italic group, which  
475 eventually came to dominate the Italian Peninsula, is notoriously difficult<sup>50</sup>. Archaeological  
476 interpretations have variously linked the spread to the Remedello culture, the Rinaldone  
477 culture, the Terramare culture, and the Proto-Villanovan as well as the Villanovan cultures<sup>51</sup>.  
478 However, the oldest direct linguistic evidence for the presence of Italic is first found in Old  
479 Latin and Umbrian inscriptions from around 2,650 BP.

480  
481 Previous genetic studies have dated the arrival of steppe ancestry in Northern Italy to around  
482 4,000 BP and in Central Italy by 3,600 BP<sup>21</sup>. In our data, steppe ancestry appears in Central  
483 Italy a century earlier, in two newly reported individuals from Latium (Pian Sultano,  
484 CGG\_2\_101264, CGG\_2\_101266), the eventual epicentre of the Latin language. According  
485 to our IBD admixture modelling, the steppe ancestry of these individuals, along with that of  
486 published Late Bronze Age individuals from Grotta Regina Margherita and Toppo Daguzzo,  
487 is characterized as genetically Bell Beaker rather than Yamnaya-related, similar to  
488 historically Celtic-speaking populations in Western Europe. This, as well as the prevalence of  
489 this ancestry throughout the Bronze and Iron Age, is consistent with the linguistic Italo-Celtic  
490 hypothesis.

491

#### 492 **The Eastern Mediterranean: Greek**

493 The timing and trajectory of the entry of the Greek language into Greece are traditionally  
494 much-debated topics<sup>2,25</sup> (Linguistic Supplementary 4.1). In Late Bronze Age Greece, writing  
495 emerges in the form of the Linear A Minoan (c. 3,800–3,400 BP) and Linear B Mycenaean  
496 (c. 3,350–3,150 BP) syllabary scripts<sup>52</sup>. In Cyprus, it appears with the Cypro-Minoan  
497 syllabary (c. 3,500–3,300 BP), a local variant of Linear A, and the derived Cypriot syllabary  
498 (c. 3,000–2,300 BP)<sup>53</sup>. While the Linear A text corpus remains largely undeciphered, Linear  
499 B as well as the Cypriot syllabary have been shown to represent the earliest written evidence  
500 of Greek<sup>54</sup>.

501

502 The arrival of steppe ancestry has previously been documented in individuals from Northern  
503 Greece as early as ~4,200 BP, suggesting a connection to the Pontic Steppe<sup>44,55,25,43</sup>. Using  
504 IBD admixture modelling, we establish that the Steppe source in Greece originates directly  
505 from Yamnaya-related populations, and differs from CWC populations who are formed with  
506 Yamnaya and GAC and widespread across most of Europe. Notably, the proportion of this  
507 Yamnaya-related ancestry is higher in the individuals before ~3,800 BP, including a  
508 previously unpublished male from Argolis (Apollo Maleatas, CGG\_2\_022928-23929), dated  
509 to around 3,800 BP. This provides the hitherto clearest evidence for the intrusion of steppe-  
510 derived, potentially Greek-speaking groups into the Peloponnese. The appearance of steppe  
511 ancestry here thus predates the oldest direct evidence of Greek in the form of the Linear B  
512 script, by which time it had stabilized at lower levels<sup>43</sup>. At the same time, the preliterate  
513 interactions with local populations find a possible analogue in the absorption of so-called  
514 “Pre-Greek” vocabulary<sup>56</sup>, reflecting contact of Greek with non-Indo-European language  
515 varieties.

516  
517 In Cyprus, the Arcadocypriot Greek dialect was spoken from at least the Early Iron Age, next  
518 to Phoenician and one or more unknown languages attested through the Cypro-Minoan and  
519 Cypriot syllabaries<sup>57</sup>. A diverse population is also suggested by genetic links between LBA  
520 and EIA Cyprus with the Levant and the Aegean. Steppe ancestry is identified in a number of  
521 unpublished individuals from Hala Sultan Tekke (CGG\_2\_022123, CGG\_2\_022924) and  
522 from Lapithos (CGG\_2\_022517), reflecting an affiliation with Late Helladic (i.e. Mycenaean  
523 Age) populations of the Peloponnese. This aligns with the appearance of Mycenaean pottery  
524 imports in the Late Bronze Age<sup>58</sup> and with the linguistic classification of Arcadocypriot as a  
525 descendant of the same South Greek dialect as Mycenaean, in contrast to other Greek dialects  
526 such as Doric and Ionic<sup>54,59</sup>. However, one other individual from Lapithos (CGG\_2\_22488),  
527 although buried 4,100–4,000 BP, already clusters with LBA Greeks, suggesting Pre-  
528 Mycenaean connectivity with steppe-impacted populations of the Aegean.

### 529 **The Caucasus and Eastern Anatolia: Armenian**

531 The Armenian language, attested from c. 1,550 BP, is historically spoken in the South  
532 Caucasus and Eastern Anatolia (Linguistic Supplementary 4.3). Its earliest presence there has  
533 previously been estimated to date to around 3,100 BP<sup>51,60</sup>. During the Late Iron Age, most of  
534 the region was part of the Urartian Kingdom. This state was culturally diverse<sup>61</sup> and likely  
535 also contained an Armenian linguistic element, as evidenced by the exchange of vocabulary  
536 between Urartian and an early form of Armenian<sup>62</sup>. This is potentially supported by the  
537 presence of steppe ancestry among the published individuals from Urartian and pre-Urartian  
538 contexts<sup>26</sup>.

539 Steppe ancestry has previously been detected in the South Caucasus from the Middle Bronze  
540 Age<sup>24</sup>, coinciding with the transition from the Kura-Araxes culture to the Trialeti culture by  
541 the end of the 5th millennium BP. We can now demonstrate that these individuals, as well as  
542 those from Urartian contexts, received steppe ancestry from the same, western Yamnaya  
543 population as 4th millennium BP individuals from the Aegean (Extended Data Fig. 6,  
544 Genetics and Strontium Supplementary Fig. S6.19; S6.21). These findings support the

545 linguistic Graeco-Armenian hypothesis and suggest that the linguistic precursor of Armenian  
546 was introduced to the Caucasus by the end of the 5th millennium BP<sup>24,63</sup>.

### 547 **Archaeological implications**

548 Bell Beaker populations exerted a pronounced genetic and cultural impact in Iberia, whereas  
549 the situation in Italy is more complex. In Northeastern Italy, Bell Beaker groups appear to  
550 have arrived in relatively small numbers, with some individuals receiving prominent tumulus  
551 burials. However, during the Early Bronze Age and at the transition to the Middle Bronze  
552 Age, evidence emerges of a connection between Central Europe and Italy in the distribution  
553 of triangular daggers, often found in hoards<sup>64,65,66</sup>. These daggers are distributed across Italy,  
554 with their appearance coinciding with the widespread diffusion of Bell Beaker-related  
555 ancestry across the Italian Peninsula between 3,800 and 3,500 BP.

556

557 In the Terramare region, population growth significantly exceeded local development,  
558 suggesting an influx from surrounding areas<sup>67,68</sup>. Archaeological evidence suggests  
559 connections with Hungary and regions north of the Alps<sup>69–72</sup>, although these links only  
560 become evident in genetic ancestries later, during the Iron Age. This suggests that Northern  
561 Italy functioned as a cultural “melting pot”, which would later impact other Italian regions.  
562 After 3,200 BP, a noticeable exodus is archaeologically documented, particularly from the  
563 southern part of the Terramare region in the Po Valley<sup>67,73</sup>.

564

565 Further south along the Adriatic coast, three genomes from the Central and Southern Italian  
566 littoral (CGG\_2\_100646; NEO806; R1) reveal affinities with Balkan populations, reflecting  
567 sustained contacts between both sides of the Adriatic Sea during the 5th and 4th millennia  
568 BP<sup>74–76</sup>. These contacts extended beyond the Adriatic, reaching as far as the Aegean, as  
569 indicated by the presence of Mycenaean pottery and other goods in settlements along the  
570 Italian coast<sup>77</sup>. Genetic ancestries further suggest small-scale movements of people, possibly  
571 involving specialized craftspeople and traders, or potentially driven by exogamy practices,  
572 contributing to the observed genetic admixture in the region.

573

574 Greece reveals a different picture. A distinct steppe proportion was introduced into both the  
575 mainland and the Peloponnese around 3,800 BP, subsequently becoming prevalent in burial  
576 sites over the ensuing centuries. The arrival of Mycenaean culture, along with the Greek  
577 language, has been subject to divergent interpretations, ranging from gradual, peaceful  
578 infiltration to more rapid acquisition of political control<sup>78–80</sup>. Neither of these models are  
579 contradicted by the new genetic evidence: following the collapse of Early Helladic society,  
580 the brief Middle Helladic period saw new migrants from the north. These eventually  
581 consolidated political control with the advent of the Late Helladic period and the formation of  
582 Early Mycenaean Culture around 3,700 BP<sup>81</sup>. The process also entailed a cultural and  
583 linguistic encounter between newcomers and the residing population. Over time the original  
584 steppe signal diminished due to admixture with local populations with farmer ancestry.  
585 Archaeological evidence documents sustained contacts with the genetic source region in  
586 Moldova and the surrounding Carpathian region, as reflected in the introduction of steppe  
587 horses and chariots<sup>82</sup>, as well as the trade in silver from Carpathian mines<sup>83</sup>.

588

589 Thus, a long-standing debate regarding the origins of Mycenaean culture can be resolved, at  
590 least in part. Genetic links with Anatolia and Crete persisted, mirroring the cultural influences  
591 that shaped the formation of Mycenaean society. Moreover, Mycenaean culture exhibits  
592 striking similarities with the slightly earlier Trialeti culture of the Southern Caucasus<sup>68</sup>, which  
593 likely contributed to processes behind the subsequent Armenian ethnogenesis. In the South  
594 Caucasus, we see the rise of new Bronze Age elites in the mineral-rich regions of present-day  
595 Georgia and Armenia. These elites had links to both steppe chariot traditions and the Hittite  
596 civilization in Anatolia. Trialeti burial inventories, characterized by monumental tumulus  
597 chamber burials featuring precious imports, exhibit parallels with both Hittite city-states and  
598 early Mycenaean shaft graves. This suggests the formation of new commercial and military  
599 networks linking steppe societies and Near Eastern civilizations<sup>68</sup>, as now also corroborated  
600 by the genetics.

601

602 The genetic evidence from Cyprus underscores its significant role as a trade hub owing to its  
603 abundant copper resources from Chalcolithic until the Iron Age. Moreover, this evidence  
604 highlights the island's strong ties to Western Anatolia and the Aegean. The Late Bronze Age  
605 marked the peak of Cypriot copper mining and trade across the Mediterranean, fostering a  
606 flourishing international culture<sup>84,85</sup>. Also, these connections are mirrored in the genetic  
607 evidence, which reveals links between Cyprus and both the Eastern and Western  
608 Mediterranean regions. By the early 4th millennium BP, an archaeological connection  
609 additionally existed between the Únětice culture and the Eastern Mediterranean, including  
610 Cyprus, as exemplified by the presence of Únětice ring ingots and dress pins<sup>81</sup>. This  
611 connection is potentially supported by genetic evidence from a single individual, although its  
612 historical significance remains enigmatic.

613

## 614 **Conclusion**

615 For the first time, we are able to coherently combine evidence from ancient DNA, strontium  
616 isotopes, linguistics and archaeology to support a dual model for the divergence and dispersal  
617 of the Italic, Greek and Armenian languages during the Bronze Age. Specifically, we show  
618 that the arrival of steppe ancestry in Spain, France and Italy was mediated by Bell Beaker  
619 populations of Western Europe, likely contributing to the emergence of the Italic and Celtic  
620 languages. In contrast, Armenian and Greek populations acquired steppe ancestry directly  
621 from Yamnaya groups of Eastern Europe. These results are consistent with the linguistic  
622 Italo-Celtic and Graeco-Armenian hypotheses accounting for the origins of most  
623 Mediterranean Indo-European languages of Classical Antiquity. In contrast, however, our  
624 results fail to support the competing Italo-Germanic and Indo-Greek hypotheses, as the  
625 steppe ancestry among the populations of historically Germanic- and Indo-Iranian-speaking  
626 areas has been characterized as primarily Corded Ware-related. Our findings thus align with  
627 specific linguistic divergence models for the Indo-European language family while  
628 contradicting others. This underlines the power of ancient DNA in uncovering prehistoric  
629 diversifications of human populations and language communities.

630

631

632 **References**

- 633 1. Ringe, D., Warnow, T. & Taylor, A. Indo-European and computational cladistics. *Trans. Philol.*  
634 *Soc.* **100**, 59–129 (2002).
- 635 2. Anthony, D. W. *The Horse, the Wheel, and Language*. (Princeton University Press, 2010).
- 636 3. Chang, W., Hall, D., Cathcart, C. & Garrett, A. ANCESTRY-CONSTRAINED  
637 PHYLOGENETIC ANALYSIS SUPPORTS THE INDO-EUROPEAN STEPPE HYPOTHESIS.  
638 *Language* **91**, 194–244 (2015).
- 639 4. Bouckaert, R. *et al.* Mapping the origins and expansion of the Indo-European language family.  
640 *Science* **337**, 957–960 (2012).
- 641 5. Heggarty, P. *et al.* Language trees with sampled ancestors support a hybrid model for the origin  
642 of Indo-European languages. *Science* **381**, eabg0818 (2023).
- 643 6. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
- 644 7. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in  
645 Europe. *Nature* **522**, 207–211 (2015).
- 646 8. Blažek, V. From August Schleicher to Sergei Starostin: On the development of the tree-diagram  
647 models of the Indo-European languages. *Journal of Indo-European Studies* **35**, 82–110 (2007).
- 648 9. Goldstein, D. Indo-European phylogenetics with R: A tutorial introduction. *Indo-European*  
649 *Linguistics* **8**, 110–180 (2020).
- 650 10. Weiss, M. Italo-Celtic. in *The Indo-European Language Family* (ed. Olander, T.) 102–113  
651 (Cambridge University Press, 2022).
- 652 11. Schrijver, P. Sound change, the Italo-Celtic linguistic unity, and the Italian homeland of Celtic.  
653 in *Celtic from the West 3: Atlantic Europe in the Metal Ages — questions of shared language*  
654 (eds. Koch, J. T. & Cunliffe, B.) 489–502 (Oxbow Books, 2016).
- 655 12. Pedersen, H. Armenier. B. Sprache. in *Reallexikon der Vorgeschichte* (ed. Ebert, M.) vol. 1 219–  
656 226 (1924).
- 657 13. Olsen, B. A. & Thorsø, R. Armenian. in *The Indo-European Language Family: A Phylogenetic*  
658 *Perspective* (ed. Olander, T.) 202–222 (Cambridge University Press, 2022).

- 659 14. Malmström, H. *et al.* The genomic ancestry of the Scandinavian Battle Axe Culture people and  
660 their relation to the broader Corded Ware horizon. *Proc. Biol. Sci.* **286**, 20191528 (2019).
- 661 15. Olalde, I. *et al.* The Beaker phenomenon and the genomic transformation of northwest Europe.  
662 *Nature* **555**, 190–196 (2018).
- 663 16. Patterson, N. *et al.* Large-scale migration into Britain during the Middle to Late Bronze Age.  
664 *Nature* **601**, 588–594 (2022).
- 665 17. Allentoft, M. E. *et al.* Population genomics of post-glacial western Eurasia. *Nature* **625**, 301–311  
666 (2024).
- 667 18. Olalde, I. *et al.* The genomic history of the Iberian Peninsula over the past 8000 years. *Science*  
668 **363**, 1230–1234 (2019).
- 669 19. Villalba-Mouco, V. *et al.* Genomic transformation and social organization during the Copper  
670 Age–Bronze Age transition in southern Iberia. *Science Advances* **7**, eabi7038 (2021).
- 671 20. Antonio, M. L. *et al.* Ancient Rome: A genetic crossroads of Europe and the Mediterranean.  
672 *Science* **366**, 708–714 (2019).
- 673 21. Saube, T. *et al.* Ancient genomes reveal structural shifts after the arrival of Steppe-related  
674 ancestry in the Italian Peninsula. *Curr. Biol.* **31**, 2576–2591.e12 (2021).
- 675 22. Posth, C. *et al.* The origin and legacy of the Etruscans through a 2000-year archeogenomic time  
676 transect. *Sci Adv* **7**, eabi7673 (2021).
- 677 23. Moots, H. M. *et al.* A genetic history of continuity and mobility in the Iron Age central  
678 Mediterranean. *Nat. Ecol. Evol.* **7**, 1515–1524 (2023).
- 679 24. Lazaridis, I. *et al.* The genetic history of the Southern Arc: A bridge between West Asia and  
680 Europe. *Science* **377**, eabm4247 (2022).
- 681 25. Lazaridis, I. *et al.* Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218 (2017).
- 682 26. Lazaridis, I. *et al.* A genetic probe into the ancient and medieval history of Southern Europe and  
683 West Asia. *Science* **377**, 940–951 (2022).
- 684 27. Skourtanioti, E. *et al.* Ancient DNA reveals admixture history and endogamy in the prehistoric  
685 Aegean. *Nat Ecol Evol* **7**, 290–303 (2023).
- 686 28. *The Indo-European Puzzle Revisited: Integrating Archaeology, Genetics, and Linguistics.*

- 687 (Cambridge University Press, Cambridge, England, 2023).
- 688 29. Olander, T. *The Indo-European Language Family*. (Cambridge University Press, 2022).
- 689 30. Euler, W. *Indoiranisch-Griechische Gemeinsamkeiten Der Nominalbildung Und Deren*  
690 *Indogermanische Grundlagen*. (Institut für Sprachwissenschaft der Universität Innsbruck,  
691 Innsbruck, 1979).
- 692 31. Porzig, W. *Die Gliederung Des Indogermanischen Sprachgebiets*. (Carl Winter, Heidelberg,  
693 1954).
- 694 32. Schrijver, P. Sound change, the Italo-Celtic linguistic unity, and the Italian homeland of Celtic.  
695 in *Celtic from the West 3* (eds. Koch, J. T. & Cunliffe, B.) 489–502 (Oxbow, Oxford, 2016).
- 696 33. Isaac, G. The nature and origins of the celtic languages : Atlantic seaways, italo-celtic and other  
697 paralinguistic misapprehensions. *Studia Celtica* **38**, 49–58 (2004).
- 698 34. Kim, R. I. Greco-Armenian: The persistence of a myth. *Indogermanische Forschungen* **123**,  
699 247–272 (2018).
- 700 35. McColl, H. *et al.* Steppe Ancestry in western Eurasia and the spread of the Germanic Languages.  
701 *bioRxiv* (2024) doi:10.1101/2024.03.13.584607.
- 702 36. Sousa da Mota, B. *et al.* Imputation of ancient human genomes. *Nat. Commun.* **14**, 3660 (2023).
- 703 37. Ringbauer, H. *et al.* Accurate detection of identity-by-descent segments in human ancient DNA.  
704 *Nat. Genet.* **56**, 143–151 (2024).
- 705 38. Browning, B. L. & Browning, S. R. Detecting identity by descent and estimating genotype error  
706 rates in sequence data. *Am. J. Hum. Genet.* **93**, 840–851 (2013).
- 707 39. Greenbaum, G., Rubin, A., Templeton, A. R. & Rosenberg, N. A. Network-based hierarchical  
708 population structure analysis for large genomic data sets. *Genome Res.* **29**, 2020–2033 (2019).
- 709 40. Narasimhan, V. M. *et al.* The formation of human populations in South and Central Asia.  
710 *Science* **365**, eaat7487 (2019).
- 711 41. Wang, C.-C. *et al.* Ancient human genome-wide data from a 3000-year interval in the Caucasus  
712 corresponds with eco-geographic regions. *Nat. Commun.* **10**, 590 (2019).
- 713 42. de Barros Damgaard, P. *et al.* The first horse herders and the impact of early Bronze Age steppe  
714 expansions into Asia. *Science* **360**, eaar7711 (2018).

- 715 43. Clemente, F. *et al.* The genomic history of the Aegean palatial civilizations. *Cell* **184**, 2565–  
716 2586.e21 (2021).
- 717 44. Koptekin, D. *et al.* Spatial and temporal heterogeneity in human mobility patterns in Holocene  
718 Southwest Asia and the East Mediterranean. *Curr. Biol.* **33**, 41–57.e15 (2023).
- 719 45. Scorrano, G., Yediay, F. E., Pinotti, T., Feizabadifarahani, M. & Kristiansen, K. The genetic and  
720 cultural impact of the Steppe migration into Europe. *Ann. Hum. Biol.* **48**, 223–233 (2021).
- 721 46. Borgna, E., Simeoni, G., Vinci, F. & Nicosia, C. Origin and evolution of a bronze age funerary  
722 landscape in Friuli: The ‘Lower Context’ of the Tumulus of Mereto di Tomba (Udine) and the  
723 3rd-2nd Millennium Transition in the Northern Adriatic. *Origini* **53**, 113–140 (2019).
- 724 47. Sagona, A. *The Archaeology of the Caucasus: From Earliest Settlements to the Iron Age*.  
725 (Cambridge University Press, 2017).
- 726 48. Bang, P. F. & Scheidel, W. *The Oxford Handbook of the State in the Ancient Near East and*  
727 *Mediterranean*. (Oxford University Press, 2013).
- 728 49. Farney, G. D. & Bradley, G. *The Peoples of Ancient Italy*. (Walter de Gruyter GmbH & Co KG,  
729 2017).
- 730 50. Mallory, J. P. *In Search of the Indo-Europeans: Language, Archaeology, and Myth*. (Thames and  
731 Hudson, 1989).
- 732 51. Mallory, J. P. & Adams, D. Q. *Encyclopedia of Indo-European Culture*. (Taylor & Francis,  
733 1997).
- 734 52. Bartoněk, A. *Handbuch Des Mykenischen Griechisch*. (C. Winter, Heidelberg, 2003).
- 735 53. Ferrara, S. *Cypro-Minoan Inscriptions: Volume 1: Analysis*. (2012).
- 736 54. Gary Miller, D. *Ancient Greek Dialects and Early Authors: Introduction to the Dialect Mixture*  
737 *in Homer, with Notes on Lyric and Herodotus*. (Walter de Gruyter, 2013).
- 738 55. Anthony, D. W. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the*  
739 *Eurasian Steppes Shaped the Modern World*. (Princeton University Press, 2010).
- 740 56. Verhasselt, G. The Pre-Greek linguistic substratum: An overview of current research. *Les Études*  
741 *Classiques* **77**, 211–239 (2009).
- 742 57. Steele, P. M. *A Linguistic History of Ancient Cyprus: The Non-Greek Languages and Their*

- 743 *Relations with Greek, c.1600–300 BC*. (Cambridge University Press, 2013).
- 744 58. Fischer, P. M. Interregional trade at Hala Sultan Tekke, Cyprus: Analysis and chronology of  
745 imports. *J. Archaeol. Sci. Rep.* **47**, 103722 (2023).
- 746 59. Giannakis, G. K., Crespo, E. & Filos, P. *Studies in Ancient Greek Dialects: From Central Greece*  
747 *to the Black Sea*. (Walter de Gruyter GmbH & Co KG, 2017).
- 748 60. Diakonoff, I. *The Pre-History of the Armenian People*. (Caravan Books, Delmar, NY, 1984).
- 749 61. Zimansky, P. Archaeological inquiries into ethno-linguistic diversity in Urartu. in *Greater*  
750 *Anatolia and the Indo-Hittite Language Family: Papers Presented at the Colloquium Hosted by*  
751 *the University of Richmond, March 18–19, 2000* (ed. Drews, R.) 15–27 (Institute for the Study of  
752 Man, Washington, DC, 2001).
- 753 62. Thorsø, R. *Prehistoric Loanwords in Armenian: Hurro-Urartian, Kartvelian, and the*  
754 *Unclassified Substrate*. (Leiden University, Leiden, 2023).
- 755 63. Anthony, D. W. Ten constraints that limit the Late PIE homeland to the steppes. in *Proceedings*  
756 *of the 33rd Annual UCLA Indo-European Conference: Los Angeles, November 12th and 13th,*  
757 *2022* (eds. Goldstein, D. M., Jamison, S. W. & Yates, A. D.) 1–25 (Helmut Buske Verlag, 2024).
- 758 64. Schwenzer, S. *Frühbronzezeitliche Vollgriffdolche: typologische, chronologische und technische*  
759 *Studien auf der Grundlage einer Materialaufnahme von Hans-Jürgen Hundt*. (Verlag des  
760 Römisch-Germanischen Zentralmuseums, 2004).
- 761 65. Van Rosenberg, E. *Cultural Landscapes, Social Networks and Historical Trajectories: A Data-*  
762 *Rich Synthesis of Early Bronze Age Networks (C. 2200-1700 BC) in Abruzzo and Lazio (Central*  
763 *Italy)*. (Sidestone Press, 2015).
- 764 66. Peroni, V. B. *I pugnali nell'Italia continentale*. (Franz Steiner Verlag Wiesbaden GmbH, 1994).
- 765 67. Cardarelli, A. The collapse of the Terramare culture and growth of new economic and social  
766 systems during the Late Bronze Age in Italy. *Scienze dell'Antichità: 15, 2009* 449–520 (2009).
- 767 68. Kristiansen, K. The rise of bronze age peripheries and the expansion of international trade 1950–  
768 1100 BC. in *Trade and Civilisation: Economic Networks and Cultural Ties, from Prehistory to*  
769 *the Early Modern Era* (ed. Kristiansen, K., Lindkvist, T. & Myrdal, J.) 87–112 (Cambridge  
770 University Press, Cambridge, 2018).

- 771 69. Cremaschi, M. *et al.* Climate change versus land management in the Po Plain (Northern Italy)  
772 during the Bronze Age: New insights from the VP/VG sequence of the Terramara Santa Rosa di  
773 Poviglio. *Quat. Sci. Rev.* **136**, 153–172 (2016).
- 774 70. Nicolis, F. Northern Italy. in *The Oxford Handbook of the European Bronze Age* (eds. Harding,  
775 A. & Fokkens, H.) 692–705 (OUP Oxford, 2013).
- 776 71. Vanzetti, A. The rise of the Terramara system (Northern Italy). *Cultural change in the shadow of*  
777 *Thera-Eruption* (2013).
- 778 72. Cardarelli, A., Cavazzuti, C., Fritzl, M., Gavranovic, M., Hajdu, T., Kiss, V., Köhler, K.,  
779 Kulcsár, G., Melis, E., Rebay-Salisbury, K., Szabó, G., Szeverényi, V. The connections between  
780 the plains of the Po and the Danube during the Bronze Age seen through the spread of the  
781 urnfield model. *Rivista di Scienze Preistoriche* **LXX**, 231–243 (2020).
- 782 73. Bettelli, M., Cardarelli, A., Damiani, I. & Others. Le ultime terramare e la penisola. Circolazione  
783 di modelli o diaspora? *STUDI DI PREISTORIA E PROTOSTORIA* **3**, 187–198 (2018).
- 784 74. Arena, A. & Barbarić, V., Cardarelli, A., Govedarica, B., Radić Rossi, I., Tunzi, A.M. The  
785 Adriatic Sea and the interactions between its two shores during the late Early and Middle Bronze  
786 Age. *Rivista di scienze preistoriche* 245–257 (2020).
- 787 75. Cavazzuti, C. *et al.* The first ‘urnfields’ in the Plains of the Danube and the Po. *J. World Prehist.*  
788 **35**, 45–86 (2022).
- 789 76. Gori, M., Recchia, G., Pacciarelli, M. & Cazzella, A. 2500-2000 BC: connectivity phenomena  
790 between the Balkans, Greece, Southern Italy, Eastern Sicily, the Aeolian Islands and Malta,  
791 RIVISTA DI SCIENZE PREISTORICHE numero speciale Italia tra Mediterraneo ed Europa:  
792 mobilità, interazioni e scambi LXX S1 2020. *Rivista di Scienze Preistoriche* 181–197 (2020).
- 793 77. Iacono, F. *et al.* Establishing the Middle Sea: The Late Bronze Age of Mediterranean Europe  
794 (1700–900 BC). *Journal of Archaeological Research* **30**, 371–445 (2022).
- 795 78. Voutsaki, S. Social place space early Mycenaean Greece. *International Discussions Mycenaean*  
796 *Archaeology* (Mykenische Studien 35). in *Social change and human agency. The Argolid at the*  
797 *onset of the Mycenaean era. In Eder, B. and Zavadil, M. (eds) (Social) place and space in early*  
798 *Mycenaean Greece. International Discussions in Mycenaean Archaeology* (Mykenische Studien

- 799 35) (Austrian Academy Sciences Press, 2021).
- 800 79. Drews, R. The coming of the Greeks: Indo-European conquests in the Aegean and the Near East.  
801 (1988).
- 802 80. Wright, J. C. Early Mycenaean Greece. in *The Cambridge Companion to the Aegean Bronze Age*  
803 (ed. Shelmerdine, C. W.) 230–257 (Cambridge University Press, Cambridge, 2008).
- 804 81. Kristiansen, K. & Larsson, T. B. *The Rise of Bronze Age Society: Travels, Transmissions and*  
805 *Transformations*. (Cambridge University Press, 2005).
- 806 82. Librado, P. *et al.* The origins and spread of domestic horses from the Western Eurasian steppes.  
807 *Nature* **598**, 634–640 (2021).
- 808 83. Stos-Gale, Z. A. Silver vessels in the Mycenaean Shaft Graves and their origin in the context of  
809 the metal supply in the Bronze Age Aegean. in *Metals of power – Early gold and silver* (ed.  
810 Meller, H., Risch, R. & Pernicka, E.) 183–208 (Landesmuseum für Vorgeschichte, Halle (Saale),  
811 2014).
- 812 84. Steel, L. *Oxford Handbook Archaeology Levant: C. 8000-332 BCE. 577–591. 577–591* (Oxford  
813 University Press, Oxford, 2013).
- 814 85. Knapp, A. B. *Cambridge World Archaeology: The Archaeology of Cyprus: From Earliest*  
815 *Prehistory through the Bronze Age*. (Cambridge University Press, Cambridge, England, 2013).
- 816 86. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target  
817 capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448 (2010).
- 818 87. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast  
819 approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–  
820 1684 (2013).
- 821 88. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo methylation in  
822 ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
- 823 89. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.  
824 *Bioinformatics* **25**, 1754–1760 (2009).
- 825 90. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming,  
826 identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).

- 827 91. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*  
828 **514**, 445–449 (2014).
- 829 92. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and  
830 endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
- 831 93. Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals into  
832 Asia. *Science* **334**, 94–98 (2011).
- 833 94. Hofmanová, Z. *et al.* Early farmers from across Europe directly descended from Neolithic  
834 Aegeans. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6886–6891 (2016).
- 835 95. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using  
836 dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).
- 837 96. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete  
838 sequence classes. *Nature* **423**, 825–837 (2003).
- 839 97. Poznik, G. D. *et al.* Sequencing Y chromosomes resolves discrepancy in time to common  
840 ancestor of males versus females. *Science* **341**, 562–565 (2013).
- 841 98. Jobling, M. A. & Tyler-Smith, C. Human Y-chromosome variation in the genome-sequencing  
842 era. *Nat. Rev. Genet.* **18**, 485–497 (2017).
- 843 99. Hanghøj, K., Moltke, I., Andersen, P. A., Manica, A. & Korneliussen, T. S. Fast and accurate  
844 relatedness estimation from high-throughput sequencing data in the presence of inbreeding.  
845 *Gigascience* **8**, (2019).
- 846 100. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079  
847 (2009).
- 848 101. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and  
849 population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993  
850 (2011).
- 851 102. Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. MAFFT: a novel method for rapid multiple  
852 sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
- 853 103. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:  
854 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

855 104. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable  
856 and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–  
857 4455 (2019).

858 105. Frei, K. M. *et al.* Tracing the dynamic life story of a Bronze Age Female. *Sci. Rep.* **5**, 10431  
859 (2015).

## 860 **Methods**

861

### 862 **Data generation**

863 All steps of generating the ancient genomes were conducted in dedicated ancient DNA lab  
864 facilities at Lundbeck Centre for GeoGenetics, University of Copenhagen, using well  
865 established protocols within ancient DNA<sup>17,35</sup> (Genetics and Strontium Supplementary S1).  
866 Drilling was performed manually, if possible, on both petrous bone and tooth. DNA  
867 extractions and library builds were carried out in the ancient DNA clean lab both manually  
868 and with automatisation. Double stranded “USER” and “non-USER” libraries<sup>86</sup> were built  
869 and sequenced on the Illumina Hiseq 4000 and the Novaseq 6000. non-USER were evaluated  
870 for the purpose of authenticating ancient reads, to allow for the detection of post-mortem  
871 damage<sup>87</sup>. Where possible, USER treated libraries were built from the authenticated extracts  
872 to minimise the effects of post-mortem damage on downstream analysis<sup>88</sup> (Supplementary  
873 Table S1). We merged the USER and non-USER libraries as described in Genetics and  
874 Strontium Supplementary S2. All reads were mapped to the human reference genomes (build  
875 hs37d5) using bwa aln (0.7.17)<sup>89</sup>, and removed the adapters using AdapterRemoval (2.3.2)<sup>90</sup>,  
876 duplicates by using picard MarkDuplicates (2.25.0). to ensure the authenticity of the  
877 sequenced data, we performed mapDamage2.0 (v2.2.1)<sup>87</sup> (Supplementary Table S1).  
878 Contaminated libraries were identified using contamix<sup>91</sup>, schmutzi<sup>92</sup> and for the X  
879 chromosome contamination level, we ran ANGSD<sup>93</sup>. The total 380 samples yielded DNA  
880 ranging from 0.004X to 15.5X in autosomal coverage, 314 samples out of total samples  
881 provided DNA coverage >0.1X which were used for downstream analysis (Supplementary  
882 Table S1).

883

### 884 **Imputation**

885 We imputed<sup>36</sup> a total 2,403 ancient samples (merged whole genome shotgun and capture),  
886 314 newly sequenced whole genome shotgun samples by following McColl et al. 2024<sup>35</sup> and  
887 Allentoft et al 2024<sup>17</sup> (Genetics and Strontium Supplementary S3).

### 888 **Principal component analysis (PCA)**

889 To get an overview of the basic structure of our data, we carried out a principal component  
890 analysis (PCA) on an imputed dataset, comprising a whole set (2,228 ancient individuals) and  
891 a subset (1,837 ancient individuals) from both shotgun and capture (1,240K) data using Plink  
892 (v1.90b6.21) (Genetics and Strontium Supplementary S4). We first restricted the SNPs to  
893 capture sites, and applied MAF (0.05) filtering, resulting in a dataset with 580,130 SNPs. We

894 visualized the main IBD clusters by colouring for all ancient genomes used for clustering  
895 (Extended Data Fig. 2) and focused on the subclusters of “Farmer-related (0\_1)” and  
896 “Steppe-related (0\_4)” in the PCA plot of a subset individuals (Extended Data Fig. 3).

### 897 **IBD clustering and modelling**

898 filtered the merged VCF files INFO>0.5, MAF (minor allele frequency) 0.01 and restricted to  
899 1,240K capture SNP sites. To run IBDseq, we included the samples with >0.1x coverage for  
900 shotgun, >1x coverage for capture, and >0.90 average genotype probability across 643,430  
901 SNP (single nucleotide polymorphism) sites<sup>38</sup>. We ran IBDseq following the implementation  
902 in McColl et al. 2024 on ancient samples within a broad geographical area and period (300  
903 BP–45,000 BP)<sup>17,35,37</sup>. After removing close relatives, we ran IBD admixture models with a  
904 total of 2,228 ancient individuals, of which 274 were newly sequenced Bronze Age  
905 Mediterranean individuals

906 Prior to running Leiden network-based hierarchical clustering, we applied filters by removing  
907 the IBD segments of less than 1 cM, LOD score of less than 3, and hotspot regions. We also  
908 removed one pair of the first and second-degree relatives to minimize small clusters formed  
909 only with close relatives. We then ran the clustering with 2,228 ancient individuals by setting  
910 a minimum total shared IBD of 5 cM and a permutation of 200.

911 IBD admixture modelling is a method combining allele-matching profiling<sup>94</sup> and  
912 chromopainting<sup>95</sup> using the shared IBD length instead of allele frequency to get a better  
913 resolution for distinguishing especially genetically similar populations. Before aggregating  
914 the IBD sharing, we filtered out IBD segments of less than 1 cM, with no limit to the upper  
915 bound of total shared segments<sup>17,35</sup>. The models and the source individuals for IBD  
916 Admixture modelling were given in the Supplementary Table S4.

### 917 **Genetic sex determination**

918 We estimated the depth of coverage for the individuals with newly generated data using pysam,  
919 by counting and measuring the length of the reads (MQ > 30) and dividing the sum by the  
920 reference contig length of chromosomes 1–22, X and Y. Because the Y-chromosome presents  
921 large regions of repetitive sequence not mappable using short-read sequencing technologies<sup>96–</sup>  
922 <sup>98</sup>, we restricted all analyses to the 10 Mb single-copy region defined in Poznik et al 2013<sup>97</sup>.  
923 We called chromosomal sex for all individuals in the dataset by calculating the ratio of the  
924 depth of coverage of X to the autosomes, Y to the autosomes, and Y and X chromosomes.

925

### 926 **Relatedness**

927 To identify relatedness, we ran NGSRelate (v2)<sup>99</sup> on the imputed dataset, calculating allele  
928 frequency by using only our samples since we have a dense population structure from Eurasia  
929 (Genetics and Strontium Supplementary S7).

930

### 931 **Y-chromosome analysis**

932 We used bcftools<sup>100</sup> mpileup and call functions to call genotypes within the 10 Mb accessible  
933 region of the Y-chromosome<sup>97</sup>. We excluded indels, triallelic positions, and genotypes that  
934 were not called in more than 95% of the population of non-clonal reads. To determine

935 haplogroups, we matched ancestral and derived calls to the ISOGG 2019–2020 database using  
936 an in-house script that generates haplogroup paths in a root-to-tip manner. Those paths are  
937 ranked by the number of supporting variants – while also distinguishing C to T in forward and  
938 G to A in reverse strands – and then manually verified.

939

#### 940 **Mitochondrial DNA analyses**

941 We re-aligned the newly sequenced ancient DNA reads to the revised Cambridge Reference  
942 Sequence (rCRS) for the human mitochondrial DNA sequence using *bwa aln* v. 0.7.17<sup>89</sup>  
943 (*options: -l10000*) and filtered for reads with a mapping quality of minimum 30 using  
944 SAMtools v. 1.17<sup>100</sup>. We then reconstructed consensus sequences of the mitogenomes with  
945 *bcftools*<sup>101</sup> v. 1.18 using *mpileup* (*options: --no-BAQ*) to obtain read pileups along the  
946 reference sequence, which were then inputted to *bcftools call* (*options: --multiallelic-caller --*  
947 *ploidy 1*) for haploid genotype calling. We kept variants covered by at least five reads and a  
948 genotype quality above 25. We generated the final consensus sequences with *bcftools*  
949 *consensus*. The reconstructed mitogenomes were aligned with *mafft*<sup>102,103</sup> v.7.490, while we  
950 restricted the phylogenetic analysis to the coding region located at 577–16,023 base pairs  
951 (rCRS coordinates). We carried out a Maximum Likelihood (ML)-based phylogenetic tree  
952 analysis with *RAxML-NG*<sup>104</sup> v. 1.2.2 under the substitution model GTR+I+G4 (*options: --all -*  
953 *-bs-trees 100*).

954

#### 955 **Sr isotope analysis**

956 We performed Sr isotope analysis on 232 skeletal samples (139 teeth and 93 petrous bones)  
957 from 224 ancient individuals from Cyprus, Greece, Italy, and Spain (Supplementary Table  
958 S8). A diamond-tipped dental drill was used to cut a clean enamel sample (1–2 mg) from the  
959 tooth samples and to drill 1–2 mg of sample from the densest part of the otic capsule of the  
960 petrous bones. The tooth and bone samples were dissolved using a 1:1 solution of 0.5 ml 6M  
961 HCl and 0.5 ml 30% H<sub>2</sub>O<sub>2</sub>. Selected samples were spiked with a <sup>84</sup>Sr-enriched tracer to  
962 determine Sr concentration via isotope dilution (ID).

963 The Sr column separation was done according to the methods of Frei et al<sup>105</sup>, using  
964 disposable 1 ml pipette tips fitted with pre-cleaned filters and charged using 200 µl pre-  
965 cleaned SrSpec™ resin (50–100 mesh; Eichrome Inc./Tristchem) as disposable extraction  
966 columns. The prepared samples were dissolved, loaded onto the columns, and washed using  
967 3M HNO<sub>3</sub>, before the Sr was collected using mq. All Sr concentrations and isotope  
968 measurements were performed at the University of Copenhagen using A VG Sector 54 IT  
969 mass spectrometer equipped with eight Faraday detectors.

#### 970 **Data availability**

971 Sequence data were deposited in the ENA under accession: xxxxxxxx

972

#### 973 **Methods References**

974 References 17, 35–36, 86–105

975

#### 976 **Acknowledgements**

977 The Lundbeck Foundation GeoGenetics Centre is supported by grants from the Lundbeck  
978 Foundation (R302-2018-2155, R155-2013-16338), the Novo Nordisk Foundation  
979 (NNF18SA0035006), the Wellcome Trust (WT214300), Carlsberg Foundation (CF18-0024),  
980 the Danish National Research Foundation (DNRF94, DNRF174), the University of  
981 Copenhagen (KU2016 programme) and Ferring Pharmaceuticals A/S, to E.W. This work was  
982 further supported by the Swedish Foundation for Humanities and Social Sciences grant  
983 (Riksbankens Jubileumsfond M16-0455:1) and funded by the European Research Council  
984 (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant  
985 agreement no. 95138) to K.K., G.K., K.G.S., K.M.F. and F.E.Y.  
986 A.M.W., F.E.Y., G.J.K. and R.T. received support from the European Research Council  
987 (Starting Grant no. 716732). We are grateful for the support of the Carlsberg Foundation  
988 Semper Ardens research grant (CF18-0005) “Tales of Bronze Age People” and the research  
989 grant “Tales of Bronze Age Women” (CF-15 0878) both to K.M.F. and F.E.Y.  
990 S.S. and G.S. were supported by Riksbankens Jubileumsfond, grant P22-0641 Italy before  
991 Rome. J.V.M. is supported by the European Research Council (101078151) and VILLUM  
992 FONDEN (VIL53099). R.N. received support from NIH R35 GM153400-01.  
993 We would like to thank to Sapienza Università di Roma - progetti di Ricerca e Grandi Scavi,  
994 Regione Friuli Venezia Giulia; Fondazione Friuli; Università di Udine (projects “Dai tumuli  
995 ai castellieri: 1500 anni di storia in Friuli”, “Protostoria della media pianura friulana”), The  
996 M.H. Wiener Laboratory for Archaeological Science, American School of Classical Studies  
997 at Athens, Sapienza Università di Roma; Soprintendenza Archeologia Belle Arti Paesaggio  
998 per le province di Foggia, Andria – Barletta, Soprintendenza Archeologia Belle Arti  
999 Paesaggio Friuli Venezia Giulia, Japanese Institute of Anatolian Archaeology, Kaman  
1000 Kalehoyük Archaeology Museum, Balıkesir Kuva-yi Milliye Archaeology Museum,  
1001 Eskisehir Eti Archaeology Museum, Çorum Archaeology Museum, Ministry of Culture and  
1002 Tourism in Republic of Turkey for providing access to archeological samples. We thank  
1003 Henrik Bjerre Hansen, Maria Madrona, Lærke Daniela Kjærsgaard Hansen, Tina-Louise  
1004 Marie Mortensen, Marcus Andreas Behnke Hjorth, Andreas Bak Pørksen and Nils August  
1005 Thomasen for their assistance.

1006  
1007 **Author contributions**

1008 K.K., E.W., F.E.Y., K.M.F. initiated the study. F.E.Y., K.M.F. (Strontium part), K.K. and  
1009 E.W. led the study. F.E.Y., G.K., K.M.F., K.K. and E.W. conceptualised the study. G.K.,  
1010 L.V.I., K.M.F., L.O., R.N., M.S., K.K. and E.W. supervised the research. G.K., K.M.F., R.N.,  
1011 K.K. and E.W. acquired funding for research. F.E.Y., S.S., M.E.A., F.D., I.M., A.M.M., P.F.,  
1012 I.D., A.V., M.T., G.P., J.M.O., S.O., S.V., R.O., N.P.H., Gi.R., E.B., M.L.J., G.Pa., K.G.S.,  
1013 K.M.F., G.F., A.L., A.Ca., E.S., S.A., M.A.T., T.Y., A.C., C.S., L.R., J.M.M.L., Gi.R., S.C.,  
1014 E.B., M.C.S., F.T., A.P.P., P.C., R.P., J.F.G.B., G.Pa., P.M.L.A., B.Bu., Ch.M., C.C., A.Ca.,  
1015 E.S., S.A., V.C.F., A.M., C.M., M.R., L.S., M.A.T., V.P.R., F.M.G., J.A.C.S., S.J.B., T.N.M.,  
1016 M.O.R.A., C.G.S., A.G.M., N.C. and L.Sa. were involved in sample collection. S.S., I.M.,  
1017 A.M.M., P.F., I.D., A.V., M.T., D.A., G.P., S.O., K.M., S.V., R.O., N.P.H., M.L.J., K.M.F.,  
1018 I.V.D., G.F., A.L., V.L., J.R., L.K., L.P., T.Y., S.S., A.C., C.S., L.R., J.M.M.L., B.B., Gi.R.,  
1019 S.C., E.B., M.C.S., F.T., A.P.P., P.C., R.P., J.F.G.B., G.Pa., P.M.L.A., B.Bu., Ch.M., C.C.,  
1020 A.Ca., E.S., S.A., V.C.F., A.M., C.M., M.R., L.S., M.A.T. and L.Sa. curated  
1021 bioarchaeological data. F.E.Y., H.M., G.S., L.V.I., J.C., J.S., M.E.A., F.V.S., J.T.S. and C.G.  
1022 were involved in genetic data generation. F.E.Y., T.P., H.M., M.S., T.V., I.A., T.S.K., A.R.,  
1023 G.R. and R.N. were involved in genetic data analysis. I.M., A.B.F., K.G.S., K.M.F., S.S. were  
1024 involved in other data generation and analysis (Calibrating 14C-dating, diet, Strontium etc.).  
1025 F.E.Y., G.K., A.W., R.T., K.K. and E.W. drafted the main text. F.E.Y., G.K., S.S., A.B.F.,  
1026 A.W., R.T., I.M., S.V., R.O. and N.P.H. drafted supplementary notes and materials. F.E.Y.,

1027 G.K., S.S., A.B.F., A.W., R.T., I.M., I.A., H.M., G.S., A.V., J.M.O., L.V.I., S.V., R.O.,  
1028 N.P.H., Gi.R., E.B., M.L.J., G.Pa., K.G.S., K.M.F., R.N., L.O., A.Ca., M.E.A., M.S., J.V.M.,  
1029 K.K. and E.W. involved in reviewing and editing drafts. All co-authors read, commented on  
1030 and agreed on the submitted manuscript.

1031

### 1032 **Ethics declarations**

1033 The authors declare no competing interests

1034

### 1035 **Additional Information**

1036 Supplementary Information is available for this paper.

1037

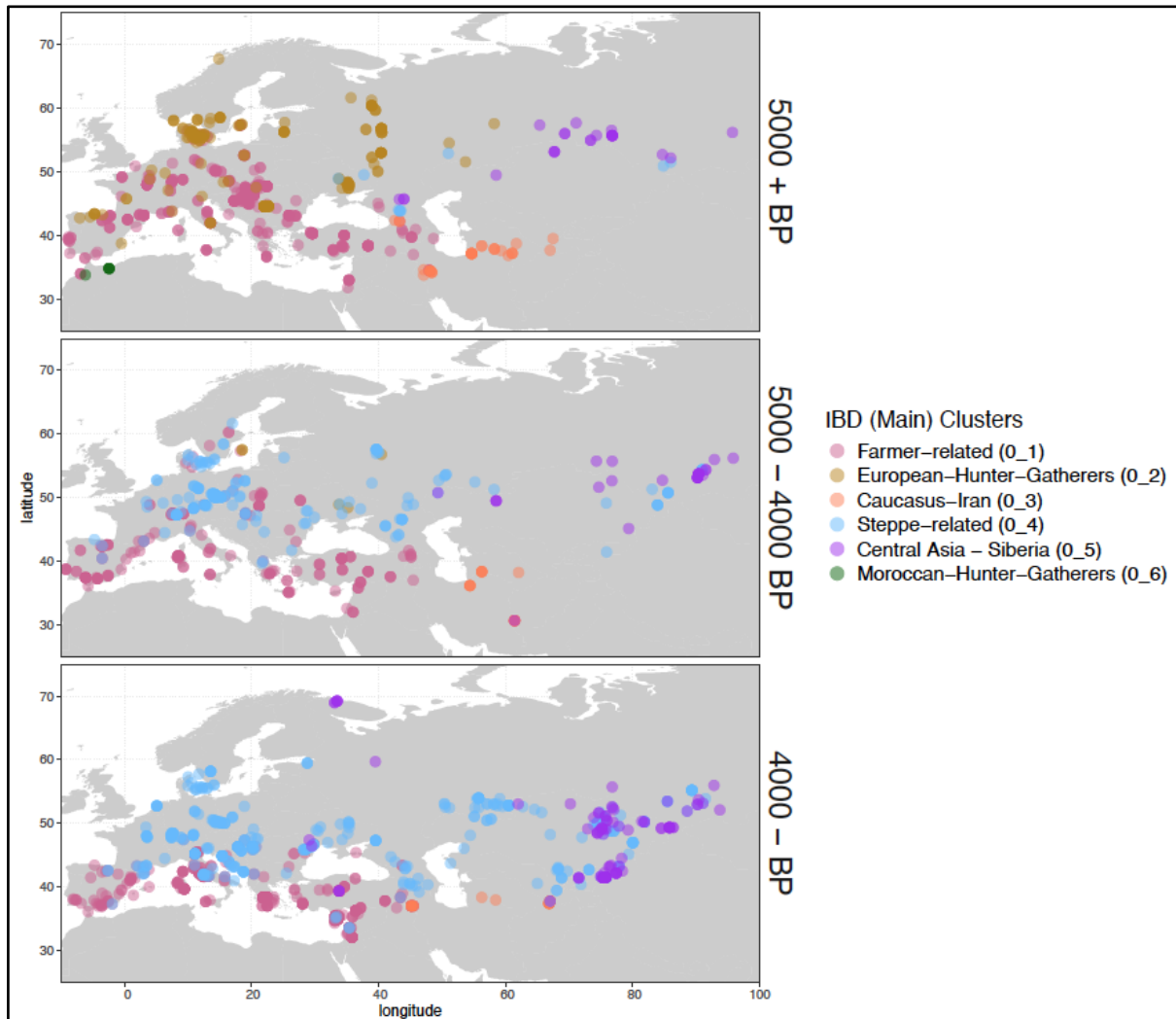
1038 Correspondence and requests for materials should be addressed to: ewillerslev@sund.ku.dk,

1039 fulyay@sund.ku.dk, guus@hum.ku.dk, kristian.kristiansen@archaeology.gu.se

1040

1041

### 1042 **Extended Data Figures**

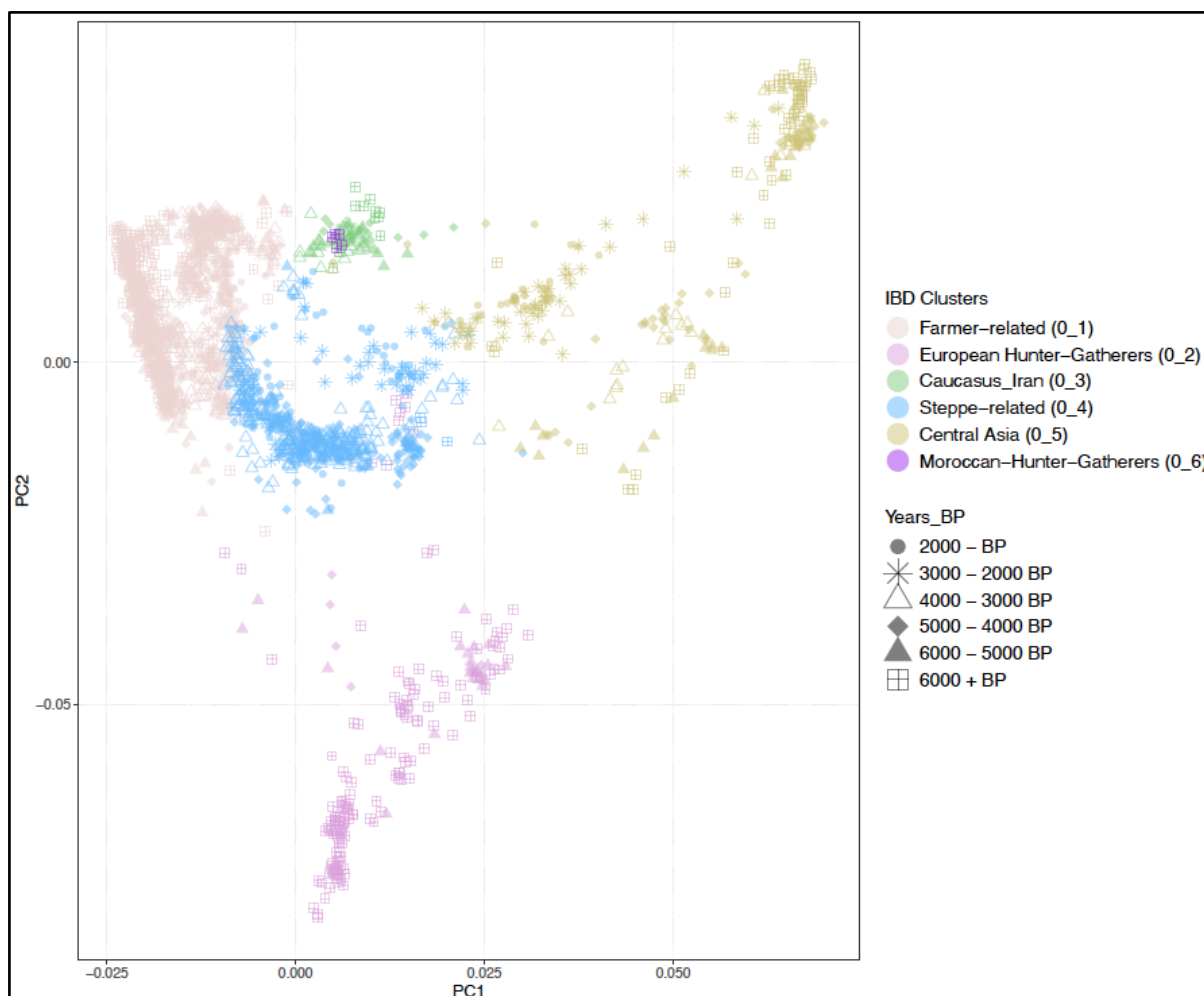


1043

1044 Extended Data Fig.1. Geographical distribution of the main IBD clusters, split into time

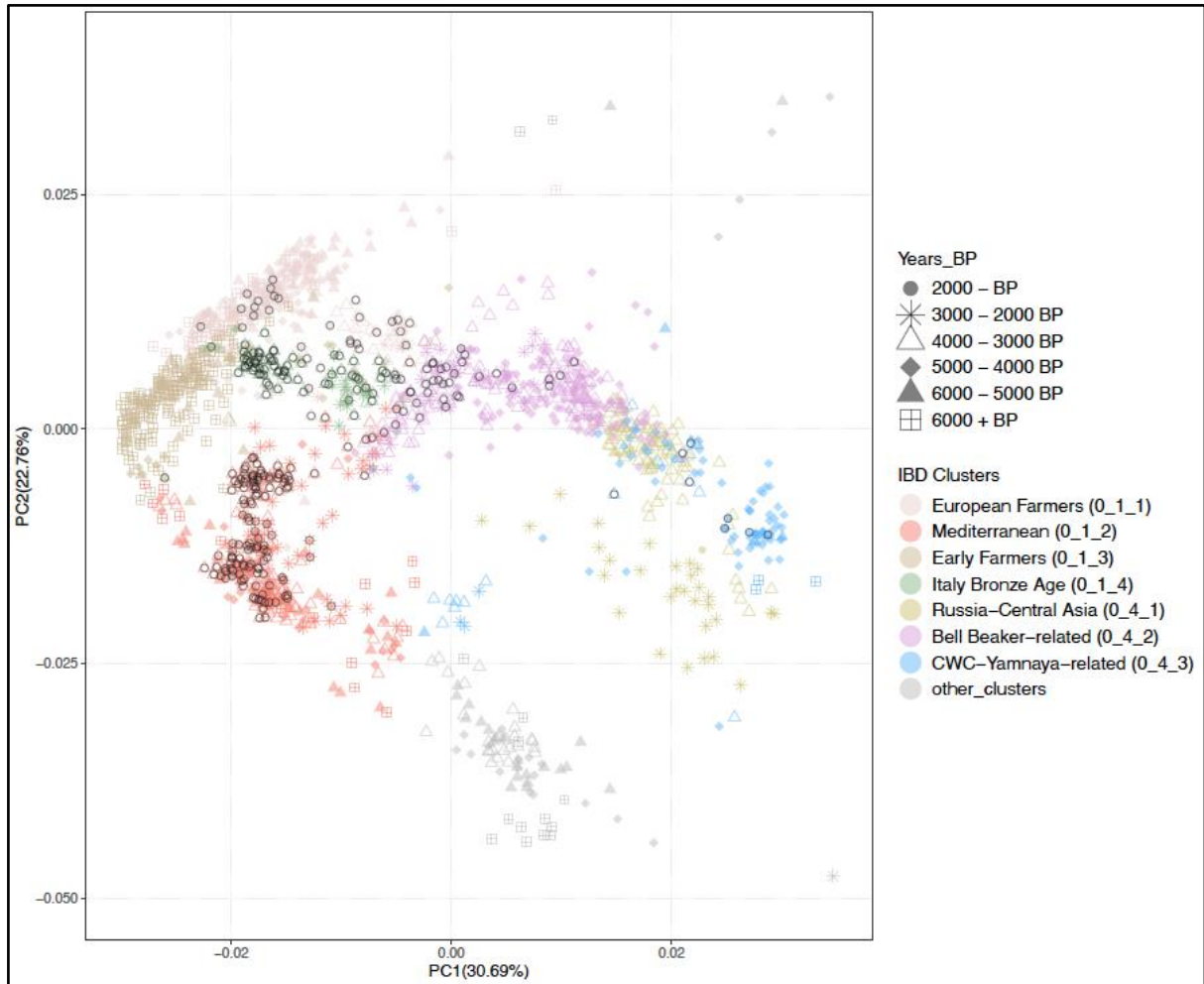
1045 ranges, pre 5,000 BP, 5,000–4,000 BP, and post 4,000 BP.

1046



1047  
1048  
1049  
1050

Extended Data Fig. 2. The PCA plot demonstrates the distribution of main IBD clusters on 2,228 ancient individuals.



1051

1052

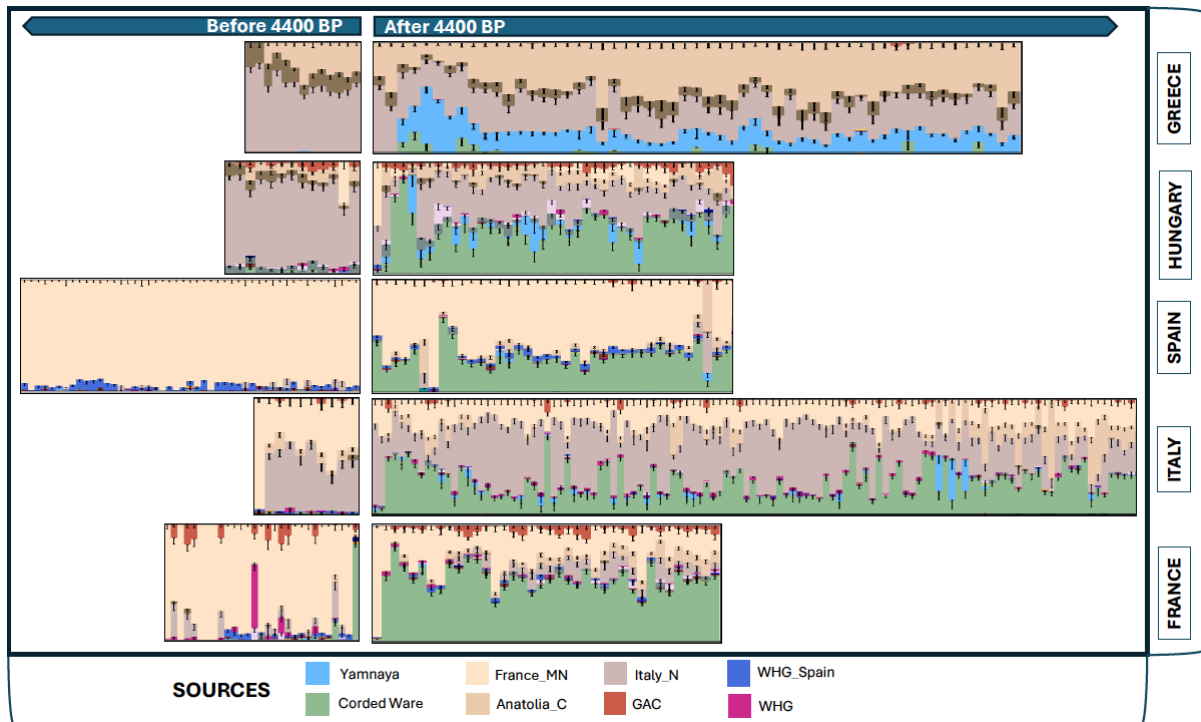
1053

1054

1055

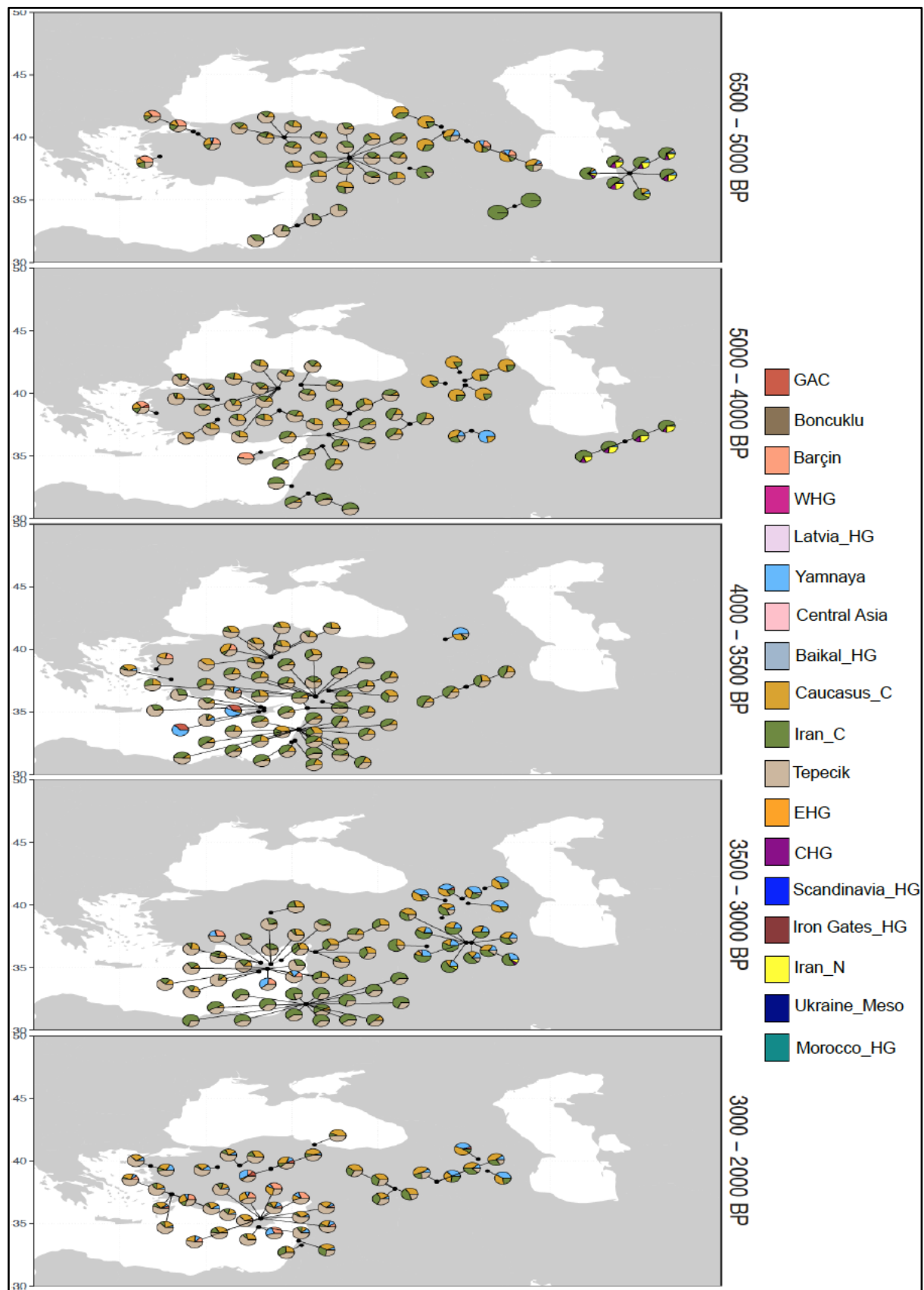
1056

Extended Data Fig. 3. The PCA plot demonstrates the distribution of subclusters of “Farmer-related (0\_1)” and “Steppe-related (0\_4)”. New genomes presented in this study were marked with black circle legends.



1057  
1058  
1059  
1060  
1061

Extended Data Fig. 4. Ancestry bar plots generated for each individual using source population proportions of IBD admixture modelling sorted by time BP and divided into two time series, before and after 4,400 BP, illustrating a Southern and Central Eastern Europe split (Italy, France, Spain and Hungary vs Greece).



1062

1063

1064

1065

Extended Data Fig. 5. Pie charts generated by using the proportions of the applied IBD admixture model for each individual from Anatolia, Cyprus, Iran, Caucasus and Levant, divided into five time periods to avoid overlapping.

