

# The genomic history of southeastern Europe

---

Mathieson, Iain; ...; Novak, Mario; ...; Čuka, Maja; ...; Janković, Ivor;  
Karavanić, Ivor; ...; Komšo, Darko; ...

Source / Izvornik: **Nature, 2018, 555, 197 - 203**

**Journal article, Submitted version**

**Rad u časopisu, Rukopis poslan na recenzijski postupak (preprint)**

<https://doi.org/10.1038/nature25778>

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:131:307624>

Rights / Prava: [Attribution 4.0 International](#) / [Imenovanje 4.0 međunarodna](#)

Download date / Datum preuzimanja: **2024-07-19**



Sveučilište u Zagrebu  
Filozofski fakultet  
University of Zagreb  
Faculty of Humanities  
and Social Sciences

Repository / Repozitorij:

[ODRAZ - open repository of the University of Zagreb  
Faculty of Humanities and Social Sciences](#)



# 1 The Genomic History of Southeastern Europe

---

2 Iain Mathieson<sup>†</sup> (1), Songül Alpaslan Roodenberg (1), Cosimo Posth (2,3), Anna Szécsényi-  
3 Nagy (4), Nadin Rohland (1), Swapan Mallick (1,5), Iñigo Olalde (1), Nasreen  
4 Broomandkoshbacht (1,5), Francesca Candilio (6), Olivia Cheronet (6,7), Daniel Fernandes  
5 (6,8), Matthew Ferry (1,5), Beatriz Gamarra (6), Gloria González Fortes (9), Wolfgang Haak  
6 (2,10), Eadaoin Harney (1,5), Eppie Jones (11,12), Denise Keating (6), Ben Krause-Kyora  
7 (2), Isil Kucukkalipci (3), Megan Michel (1,5), Alissa Mittnik (2,3), Kathrin Nägele (2),  
8 Mario Novak (6,13), Jonas Oppenheimer (1,5), Nick Patterson (14), Saskia Pfrenge (3),  
9 Kendra Sirak (6,15), Kristin Stewardson (1,5), Stefania Vai (16), Stefan Alexandrov (17),  
10 Kurt W. Alt (18,19,20), Radian Andreescu (21), Dragana Antonović (22), Abigail Ash (6),  
11 Nadezhda Atanassova (23), Krum Bacvarov (17), Mende Balázs Gusztáv (4), Hervé  
12 Bocherens (24,25), Michael Bolus (26), Adina Boroneanț (27), Yavor Boyadzhiev (17),  
13 Alicja Budnik (28), Josip Burmaz (29), Stefan Chohadzhiev (30), Nicholas J. Conard (31,25),  
14 Richard Cottiaux (32), Maja Čuka (33), Christophe Cupillard (34,35), Dorothée G. Drucker  
15 (25), Nedko Elenski (36), Michael Francken (37), Borislava Galabova (38), Georgi  
16 Ganetovski (39), Bernard Gély (40), Tamás Hajdu (41), Veneta Handzhyiska (42), Katerina  
17 Harvati (37,25), Thomas Higham (43), Stanislav Iliev (44), Ivor Janković (13,45), Ivor  
18 Karavanić (46,45), Douglas J. Kennett (47), Darko Komšo (33), Alexandra Kozak (48),  
19 Damian Labuda (49), Martina Lari (16), Catalin Lazar (50,51), Maleen Leppek (52),  
20 Krassimir Leshtakov (42), Domenico Lo Vetro (53,54), Dženi Los (29), Ivaylo Lozanov (42),  
21 Maria Malina (26), Fabio Martini (53,54), Kath McSweeney (55), Harald Meller (20), Marko  
22 Menđušić (56), Pavel Mirea (57), Vyacheslav Moiseyev (58), Vanya Petrova (42), T. Douglas  
23 Price (59), Angela Simalcsik (60), Luca Sineo (61), Mario Šlaus (62), Vladimir Slavchev  
24 (63), Petar Stanev (36), Andrej Starović (64), Tamás Szeniczey (41), Sahnra Talamo (65),  
25 Maria Teschler-Nicola (66,7), Corinne Thevenet (67), Ivan Valchev (42), Frédérique Valentin  
26 (68), Sergey Vasilyev (69), Fanica Veljanovska (70), Svetlana Venelinova (71), Elizaveta  
27 Veselovskaya (69), Bence Viola (72,73), Cristian Virag (74), Joško Zaninović (75), Steve  
28 Zäuner (76), Philipp W. Stockhammer (52,2), Giulio Catalano (61), Raiko Krauß (77), David  
29 Caramelli (16), Gunita Zariņa (78), Bisserka Gaydarska (79), Malcolm Lillie (80), Alexey G.  
30 Nikitin (81), Inna Potekhina (48), Anastasia Papatthanasiou (82), Dušan Borić (83), Clive  
31 Bonsall (55), Johannes Krause (2,3), Ron Pinhasi\* (6,7), David Reich\* (1,14,5)  
32

33 \* These authors contributed equally to the manuscript

34 <sup>†</sup> Present address; Department of Genetics, Perelman School of Medicine, University of  
35 Pennsylvania, Philadelphia PA 19104, USA

36 Correspondence to I.M. ([mathi@upenn.edu](mailto:mathi@upenn.edu)) or D.R. ([reich@genetics.med.harvard.edu](mailto:reich@genetics.med.harvard.edu)) or  
37 R.P. ([ron.pinhasi@ucd.ie](mailto:ron.pinhasi@ucd.ie))  
38

39 (1) Department of Genetics, Harvard Medical School, Boston 02115 MA USA (2) Department of Archaeogenetics,  
40 Max Planck Institute for the Science of Human History, 07745 Jena, Germany (3) Institute for Archaeological  
41 Sciences, University of Tuebingen, Germany (4) Laboratory of Archaeogenetics, Institute of Archaeology,  
42 Research Centre for the Humanities, Hungarian Academy of Sciences, H-1097 Budapest, Hungary (5) Howard  
43 Hughes Medical Institute, Harvard Medical School, Boston 02115 MA USA (6) Earth Institute and School of  
44 Archaeology, University College Dublin, Belfield, Dublin 4, Republic of Ireland (7) Department of Anthropology,  
45 University of Vienna, Althanstrasse 14, 1090 Vienna, Austria (8) CIAS, Department of Life Sciences, University  
46 of Coimbra, 3000-456 Coimbra, Portugal (9) Department of Life Sciences and Biotechnology, University of  
47 Ferrara, Via L. Borsari 46. Ferrara 44100 Italy (10) Australian Centre for Ancient DNA, School of Biological  
48 Sciences, The University of Adelaide, SA-5005 Adelaide, Australia (11) Smurfit Institute of Genetics, Trinity  
49 College Dublin, Dublin 2, Ireland (12) Department of Zoology, University of Cambridge, Downing Street,  
50 Cambridge CB2 3EJ, UK (13) Institute for Anthropological Research, Ljudevita Gaja 32, 10000 Zagreb, Croatia  
51 (14) Broad Institute of Harvard and MIT, Cambridge MA (15) Department of Anthropology, Emory University,  
52 Atlanta, Georgia 30322, USA (16) Dipartimento di Biologia, Università di Firenze, 50122 Florence, Italy (17)  
53 National Institute of Archaeology and Museum, Bulgarian Academy of Sciences, 2 Saborna Str., BG-1000 Sofia,  
54 Bulgaria (18) Danube Private University, A-3500 Krems, Austria (19) Department of Biomedical Engineering and

55 Integrative Prehistory and Archaeological Science, CH-4123 Basel-Allschwil, Switzerland (20) State Office for  
56 Heritage Management and Archaeology Saxony-Anhalt and State Museum of Prehistory, D-06114 Halle,  
57 Germany (21) Romanian National History Museum, Bucharest, Romania (22) Institute of Archaeology, Belgrade,  
58 Serbia (23) Institute of Experimental Morphology, Pathology and Anthropology with Museum, Bulgarian  
59 Academy of Sciences, Sofia, Bulgaria (24) Department of Geosciences, Biogeology, Universität Tübingen,  
60 Hölderlinstr. 12, 72074 Tübingen, Germany (25) Senckenberg Centre for Human Evolution and  
61 Palaeoenvironment, University of Tuebingen, 72072 Tuebingen, Germany (26) Heidelberg Academy of Sciences  
62 and Humanities, Research Center ‘‘The Role of Culture in Early Expansions of Humans’’ at the University of  
63 Tuebingen, Rümelinstraße 23, 72070 Tuebingen, Germany (27) ‘Vasile Pârvan’ Institute of Archaeology,  
64 Romanian Academy (28) Human Biology Department, Cardinal Stefan Wyszyński University, Warsaw, Poland  
65 (29) KADUCEJ d.o.o Papandopulova 27, 21000 Split, Croatia (30) St. Cyril and Methodius University, Veliko  
66 Turnovo, Bulgaria (31) Department of Early Prehistory and Quaternary Ecology, University of Tuebingen, Schloss  
67 Hohentübingen, 72070 Tuebingen, Germany (32) INRAP/UMR 8215 Trajectoires, 21 Allée de l’Université, 92023  
68 Nanterre, France (33) Archaeological Museum of Istria, Carrarina 3, 52100 Pula, Croatia (34) Service Régional de  
69 l’Archéologie de Bourgogne-Franche-Comté, 7 rue Charles Nodier, 25043 Besançon Cedex, France (35)  
70 Laboratoire Chronoenvironnement, UMR 6249 du CNRS, UFR des Sciences et Techniques, 16 route de Gray,  
71 25030 Besançon Cedex, France (36) Regional Museum of History Veliko Tarnovo, Veliko Tarnovo, Bulgaria (37)  
72 Institute for Archaeological Sciences, Paleoanthropology, University of Tuebingen, Rümelinstraße 23, 72070  
73 Tuebingen, Germany (38) Laboratory for human bio-archaeology, Bulgaria, 1202 Sofia, 42, George Washington  
74 str (39) Regional Museum of History, Vratsa, Bulgaria (40) DRAC Auvergne - Rhône Alpes, Ministère de la  
75 Culture, Le Grenier d’abondance 6, quai Saint Vincent 69283 LYON cedex 01 (41) Eötvös Loránd University,  
76 Faculty of Science, Institute of Biology, Department of Biological Anthropology, H-1117 Pázmány Péter sétány  
77 1/c. Budapest, Hungary (42) Department of Archaeology, Sofia University St. Kliment Ohridski, Bulgaria (43)  
78 Oxford Radiocarbon Accelerator Unit, Research Laboratory for Archaeology and the History of Art, University of  
79 Oxford, Dyson Perrins Building, South Parks Road, OX1 3QY Oxford, UK (44) Regional Museum of History,  
80 Haskovo, Bulgaria (45) Department of Anthropology, University of Wyoming, 1000 E. University Avenue,  
81 Laramie, WY 82071, USA (46) Department of Archaeology, Faculty of Humanities and Social Sciences,  
82 University of Zagreb, Ivana Lučića 3, 10000 Zagreb, Croatia (47) Department of Anthropology and Institutes for  
83 Energy and the Environment, Pennsylvania State University, University Park, PA 16802 (48) Department of  
84 Bioarchaeology, Institute of Archaeology, National Academy of Sciences of Ukraine (49) CHU Sainte-Justine  
85 Research Center, Pediatric Department, Université de Montréal, Montreal, PQ, Canada, H3T 1C5 (50) National  
86 History Museum of Romania, Calea Victoriei, no. 12, 030026, Bucharest, Romania (51) University of Bucharest,  
87 Mihail Kogalniceanu 36-46, 50107, Bucharest, Romania (52) Institute for Pre- and Protohistoric Archaeology and  
88 the Archaeology of the Roman Provinces, Ludwig-Maximilians-University, Schellingstr. 12, 80799 Munich,  
89 Germany (53) Dipartimento SAGAS - Sezione di Archeologia e Antico Oriente, Università degli Studi di Firenze,  
90 50122 Florence, Italy (54) Museo e Istituto fiorentino di Preistoria, 50122 Florence, Italy (55) School of History,  
91 Classics and Archaeology, University of Edinburgh, Edinburgh EH8 9AG, United Kingdom (56) Conservation  
92 Department in Šibenik, Ministry of Culture of the Republic of Croatia, Jurja Čulinovića 1, 22000 Šibenik, Croatia  
93 (57) Teleorman County Museum, str. 1848, no. 1, 140033 Alexandria, Romania (58) Peter the Great Museum of  
94 Anthropology and Ethnography (Kunstkamera) RAS, 199034 St. Petersburg, Russia (59) University of Wisconsin,  
95 Madison WI, USA (60) Olga Necrasov Centre for Anthropological Research, Romanian Academy – Iași Branch,  
96 Theodor Codrescu St. 2, P.C. 700481, Iași, Romania (61) Dipartimento di Scienze e tecnologie biologiche,  
97 chimiche e farmaceutiche, Lab. of Anthropology, Università degli studi di Palermo, Italy (62) Anthropological  
98 Center, Croatian Academy of Sciences and Arts, 10000 Zagreb, Croatia (63) Regional Historical Museum Varna,  
99 Maria Luiza Blvd. 41, BG-9000 Varna, Bulgaria (64) National Museum in Belgrade, 1a Republic sq., Belgrade,  
100 Serbia (65) Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig,  
101 Germany (66) Department of Anthropology, Natural History Museum Vienna, 1010 Vienna, Austria (67)  
102 INRAP/UMR 8215 Trajectoires, 21 Allée de l’Université, 92023 Nanterre, France (68) CNRS/UMR 7041 ArScAn  
103 MAE, 21 Allée de l’Université, 92023 Nanterre, France (69) Institute of Ethnology and Anthropology, Russian  
104 Academy of Sciences, Leninsky Pr., 32a, Moscow, 119991, Russia (70) Archaeological Museum of Macedonia,  
105 Skopje (71) Regional museum of history, Shumen, Bulgaria (72) Department of Anthropology, University of  
106 Toronto, Toronto, Ontario, M5S 2S2, Canada (73) Institute of Archaeology & Ethnography, Siberian Branch,  
107 Russian Academy of Sciences, Lavrentiev Pr. 17, Novosibirsk 630090, Russia (74) Satu Mare County Museum  
108 Archaeology Department, V. Lucaciu, nr.21, Satu Mare, Romania (75) Municipal Museum Drniš, Domovinskog  
109 rata 54, 22320 Drniš, Croatia (76) anthropol - Anthropologieservice, Schadenweilnerstraße 80, 72379 Hechingen,  
110 Germany (77) Institute for Prehistory, Early History and Medieval Archaeology, University of Tuebingen,  
111 Germany (78) Institute of Latvian History, University of Latvia, Kalpaka Bulvāris 4, Rīga 1050, Latvia (79)  
112 Department of Archaeology, Durham University, UK (80) School of Environmental Sciences: Geography,  
113 University of Hull, Hull HU6 7RX, UK (81) Department of Biology, Grand Valley State University, Allendale,  
114 Michigan, USA (82) Ephorate of Paleoanthropology and Speleology, Athens, Greece (83) The Italian Academy for  
115 Advanced Studies in America, Columbia University, 1161 Amsterdam Avenue, New York, NY 10027, USA.

## 116 **Abstract**

117 **Farming was first introduced to southeastern Europe in the mid-7<sup>th</sup> millennium BCE –**  
118 **brought by migrants from Anatolia who settled in the region before spreading**  
119 **throughout Europe. To clarify the dynamics of the interaction between the first farmers**  
120 **and indigenous hunter-gatherers where they first met, we analyze genome-wide ancient**  
121 **DNA data from 223 individuals who lived in southeastern Europe and surrounding**  
122 **regions between 12,000 and 500 BCE. We document previously uncharacterized genetic**  
123 **structure, showing a West-East cline of ancestry in hunter-gatherers, and show that**  
124 **some Aegean farmers had ancestry from a different lineage than the northwestern**  
125 **Anatolian lineage that formed the overwhelming ancestry of other European farmers.**  
126 **We show that the first farmers of northern and western Europe passed through**  
127 **southeastern Europe with limited admixture with local hunter-gatherers, but that some**  
128 **groups mixed extensively, with relatively sex-balanced admixture compared to the male-**  
129 **biased hunter-gatherer admixture that prevailed later in the North and West.**  
130 **Southeastern Europe continued to be a nexus between East and West after farming**  
131 **arrived, with intermittent genetic contact from the Steppe up to 2,000 years before the**  
132 **migration that replaced much of northern Europe’s population.**

133

## 134 **Introduction**

135 The southeastern quadrant of Europe was the beachhead in the spread of agriculture from its  
136 source in the Fertile Crescent of southwestern Asia. After the first appearance of agriculture  
137 in the mid-7<sup>th</sup> millennium BCE,<sup>1,2</sup> farming spread westward via a Mediterranean and  
138 northwestward via a Danubian route, and was established in both Iberia and Central Europe  
139 by 5600 BCE.<sup>3,4</sup> Ancient DNA studies have shown that the spread of farming across Europe  
140 was accompanied by a massive movement of people<sup>5-8</sup> closely related to the farmers of  
141 northwestern Anatolia<sup>9-11</sup> but nearly all the ancient DNA from Europe’s first farmers is from  
142 central and western Europe, with only three individuals reported from the southeast.<sup>9</sup> In the  
143 millennia following the establishment of agriculture in the Balkan Peninsula, a series of  
144 complex societies formed, culminating in sites such as the mid-5<sup>th</sup> millennium BCE necropolis  
145 at Varna, which has some of the earliest evidence of extreme inequality in wealth, with one  
146 individual (grave 43) from whom we extracted DNA buried with more gold than is known  
147 from any earlier site. By the end of the 6<sup>th</sup> millennium BCE, agriculture had reached eastern  
148 Europe, in the form of the Cucuteni-Trypillian complex in the area of present-day Moldova,  
149 Romania and Ukraine, including “mega-sites” that housed hundreds, perhaps thousands, of  
150 people.<sup>12</sup> After around 4000 BCE, these settlements were largely abandoned, and

151 archaeological evidence documents cultural contacts with peoples of the Eurasian steppe.<sup>13</sup>  
152 However, the population movements that accompanied these events have been unknown due  
153 to the lack of ancient DNA.

154

## 155 Results

156 We generated genome-wide data from 223 ancient humans (214 reported for the first time),  
157 from the Balkan Peninsula, the Carpathian Basin, the North Pontic Steppe and neighboring  
158 regions, dated to 12,000-500 BCE (Figure 1A, Supplementary Information Table 1,  
159 Supplementary Information Note 1). We extracted DNA from skeletal remains in dedicated  
160 clean rooms, built DNA libraries and enriched for DNA fragments overlapping 1.24 million  
161 single nucleotide polymorphisms (SNPs), then sequenced the product and restricted to  
162 libraries with evidence of authentic ancient DNA.<sup>7,10,14</sup> We filtered out individuals with fewer  
163 than 15,000 SNPs covered by at least one sequence, that had unexpected ancestry for their  
164 archaeological context and were not directly dated. We report, but do not analyze, nine  
165 individuals that were first-degree relatives of others in the dataset, resulting in an analysis  
166 dataset of 214 individuals. We analyzed these data together with 274 previously reported  
167 ancient individuals,<sup>9-11,15-27</sup> 799 present-day individuals genotyped on the Illumina “Human  
168 Origins” array,<sup>23</sup> and 300 high coverage genomes from the Simons Genome Diversity Project  
169 (SGDP).<sup>28</sup> We used principal component analysis (PCA; Figure 1B, Extended Data Figure 1),  
170 supervised and unsupervised ADMIXTURE (Figure 1D, Extended Data Figure 2),<sup>29</sup> *D*-  
171 statistics, *qpAdm* and *qpGraph*,<sup>30</sup> along with archaeological and chronological information to  
172 cluster the individuals into populations and investigate the relationships among them.

173

174 We described the individuals in our dataset in terms of their genetic relatedness to a  
175 hypothesized set of ancestral populations, which we refer to as their genetic ancestry. It has  
176 previously been shown that the great majority of European ancestry derives from three  
177 distinct sources.<sup>23</sup> First, there is “hunter-gatherer-related” ancestry that is more closely related  
178 to Mesolithic hunter-gatherers from Europe than to any other population, and that can be  
179 further subdivided into “Eastern” (EHG) and “Western” (WHG) hunter-gatherer-related  
180 ancestry.<sup>7</sup> Second, there is “NW Anatolian Neolithic-related” ancestry related to the  
181 Neolithic farmers of northwest Anatolia and tightly linked to the appearance of agriculture.<sup>9,10</sup>  
182 The third source, “steppe-related” ancestry, appears in Western Europe during the Late  
183 Neolithic to Bronze Age transition and is ultimately derived from a population related to  
184 Yamnaya steppe pastoralists.<sup>7,15</sup> Steppe-related ancestry itself can be modeled as a mixture of  
185 EHG-related ancestry, and ancestry related to Upper Palaeolithic hunter-gatherers of the  
186 Caucasus (CHG) and the first farmers of northern Iran.<sup>19,21,22</sup>

187 **Hunter-Gatherer substructure and transitions**

188 Of the 214 new individuals we report, 114 from Paleolithic, Mesolithic and eastern European  
189 Neolithic contexts have almost entirely hunter-gatherer-related ancestry (in eastern Europe,  
190 unlike western Europe, “Neolithic” refers to the presence of pottery,<sup>31-33</sup> not necessarily to  
191 farming). These individuals form a cline from WHG to EHG that is correlated with geography  
192 (Figure 1B), although it is neither geographically nor temporally uniform (Figure 2, Extended  
193 Data Figure 3), and there is also substructure in phenotypically important variants  
194 (Supplementary Information Note 2).

195

196 From present-day Ukraine, our study reports new genome-wide data from five Mesolithic  
197 individuals from ~9500-6000 BCE, and 31 Neolithic individuals from ~6000-3500 BCE. On the  
198 cline from WHG- to EHG-related ancestry, the Mesolithic individuals fall towards the East,  
199 intermediate between EHG and Mesolithic hunter-gatherers from Sweden (Figure 1B).<sup>7</sup> The  
200 Neolithic population has a significant difference in ancestry compared to the Mesolithic  
201 (Figures 1B, Figure 2), with a shift towards WHG shown by the statistic  $D(\text{Mbuti}, \text{WHG},$   
202  $\text{Ukraine\_Mesolithic}, \text{Ukraine\_Neolithic})$ ;  $Z=8.9$  (Supplementary Information Table 2).  
203 Unexpectedly, one Neolithic individual from Dereivka (I3719), which we directly date to  
204 4949-4799 BCE, has entirely NW Anatolian Neolithic-related ancestry.

205

206 The pastoralist Bronze Age Yamnaya complex originated on the Eurasian steppe and is a  
207 plausible source for the dispersal of steppe-related ancestry into central and western Europe  
208 around 2500 BCE.<sup>13</sup> All previously reported Yamnaya individuals were from Samara<sup>7</sup> and  
209 Kalmykia<sup>15</sup> in southwest Russia, and had entirely steppe-related ancestry. Here, we report  
210 three Yamnaya individuals from further West – from Ukraine and Bulgaria – and show that  
211 while they all have high levels of steppe-related ancestry, one from Ozera in Ukraine and one  
212 from Bulgaria (I1917 and Bul4, both dated to ~3000 BCE) have NW Anatolian Neolithic-  
213 related admixture, the first evidence of such ancestry in Yamnaya –associated individuals  
214 (Figure 1B,D, Supplementary Data Table 2). Two Copper Age individuals (I4110 and I6561,  
215 Ukraine\_Eneolithic) from Dereivka and Alexandria dated to ~3600-3400 BCE (and thus  
216 preceding the Yamnaya complex) also have mixtures of steppe- and NW Anatolian Neolithic-  
217 related ancestry (Figure 1D, Supplementary Data Table 2).

218

219 At Zvejnieki in Latvia (17 newly reported individuals, and additional data for 5 first reported  
220 in Ref. 34) we observe a transition in hunter-gatherer-related ancestry that is the opposite of  
221 that seen in Ukraine. We find (Supplementary Data Table 3) that Mesolithic and Early  
222 Neolithic individuals (Latvia\_HG) associated with the Kunda and Narva cultures have  
223 ancestry intermediate between WHG (~70%) and EHG (~30%), consistent with previous

224 reports.<sup>34-36</sup> We also detect a shift in ancestry between the Early Neolithic and individuals  
225 associated with the Middle Neolithic Comb Ware Complex (Latvia\_MN), who have more  
226 EHG-related ancestry (we estimate 65% EHG, but two of four individuals appear almost  
227 100% EHG in PCA). The most recent individual, associated with the Final Neolithic Corded  
228 Ware Complex (I4629, Latvia\_LN), attests to another ancestry shift, clustering closely with  
229 Yamnaya from Samara,<sup>7</sup> Kalmykia<sup>15</sup> and Ukraine (Figure 2).

230

231 We report new Upper Palaeolithic and Mesolithic data from southern and western Europe.<sup>17</sup>  
232 Sicilian (I2158) and Croatian (I1875) individuals dating to ~12,000 and 6100 BCE cluster with  
233 previously reported western hunter-gatherers (Figure 1B&D), including individuals from  
234 Loschbour<sup>23</sup> (Luxembourg, 6100 BCE), Bichon<sup>19</sup> (Switzerland, 11,700 BCE), and Villabruna<sup>17</sup>  
235 (Italy 12,000 BCE). These results demonstrate that WHG populations<sup>23</sup> were widely  
236 distributed from the Atlantic seaboard of Europe in the West, to Sicily in the South, to the  
237 Balkan Peninsula in the Southeast, for at least six thousand years.

238

239 A particularly important hunter-gatherer population that we report is from the Iron Gates  
240 region that straddles the border of present-day Romania and Serbia. This population  
241 (Iron\_Gates\_HG) is represented in our study by 40 individuals from five sites. Modeling Iron  
242 Gates hunter-gatherers as a mixture of WHG and EHG (Supplementary Table 3) shows that  
243 they are intermediate between WHG (~85%) and EHG (~15%). However, this *qpAdm* model  
244 does not fit well ( $p=0.0003$ , Supplementary table 3) and the Iron Gates hunter-gatherers carry  
245 mitochondrial haplogroup K1 (7/40) as well as other subclades of haplogroups U (32/40) and  
246 H (1/40). This contrasts with WHG, EHG and Scandinavian hunter-gatherers who almost all  
247 carry haplogroups U5 or U2. One interpretation is that the Iron Gates hunter-gatherers have  
248 ancestry that is not present in either WHG or EHG. Possible scenarios include genetic contact  
249 between the ancestors of the Iron Gates population and Anatolia, or that the Iron Gates  
250 population is related to the source population from which the WHG split during a re-  
251 expansion into Europe from the Southeast after the Last Glacial Maximum.<sup>17,37</sup>

252

253 A notable finding from the Iron Gates concerns the four individuals from the site of Lepenski  
254 Vir, two of whom (I4665 & I5405, 6200-5600 BCE), have entirely NW Anatolian Neolithic-  
255 related ancestry. Strontium and Nitrogen isotope data<sup>38</sup> indicate that both these individuals  
256 were migrants from outside the Iron Gates, and ate a primarily terrestrial diet (Supplementary  
257 Information section 1). A third individual (I4666, 6070 BCE) has a mixture of NW Anatolian  
258 Neolithic-related and hunter-gatherer-related ancestry and ate a primarily aquatic diet, while a  
259 fourth, probably earlier, individual (I5407) had entirely hunter-gatherer-related ancestry  
260 (Figure 1D, Supplementary Information section 1). We also identify one individual from

261 Padina (I5232), dated to 5950 BCE that had a mixture of NW Anatolian Neolithic-related and  
262 hunter-gatherer-related ancestry. These results demonstrate that the Iron Gates was a region of  
263 interaction between groups distinct in both ancestry and subsistence strategy.

264

### 265 **Population transformations in the first farmers**

266 Neolithic populations from present-day Bulgaria, Croatia, Macedonia, Serbia and Romania  
267 cluster closely with the NW Anatolian Neolithic farmers (Figure 1), consistent with  
268 archaeological evidence.<sup>39</sup> Modeling Balkan Neolithic populations as a mixture of NW  
269 Anatolian Neolithic and WHG, we estimate that 98% (95% confidence interval [CI]; 97-  
270 100%) of their ancestry is NW Anatolian Neolithic-related. A striking exception is evident in  
271 8 out of 9 individuals from Malak Preslavets in present-day Bulgaria.<sup>40</sup> These individuals  
272 lived in the mid-6<sup>th</sup> millennium BCE and have significantly more hunter-gatherer-related  
273 ancestry than other Balkan Neolithic populations (Figure 1B,D, Extended Data Figures 1-3,  
274 Supplementary Tables 2-4); a model of 82% (CI: 77-86%) NW Anatolian Neolithic-related,  
275 15% (CI: 12-17%) WHG-related, and 4% (CI: 0-9%) EHG-related ancestry is a fit to the data.  
276 This hunter-gatherer-related ancestry with a ~4:1 WHG:EHG ratio plausibly represents a  
277 contribution from local Balkan hunter-gatherers genetically similar to those of the Iron Gates.  
278 Late Mesolithic hunter-gatherers in the Balkans were likely concentrated along the coast and  
279 major rivers such as the Danube,<sup>41</sup> which directly connects the Iron Gates with Malak  
280 Preslavets. Thus, early farmer groups with the most hunter-gatherer-related ancestry may  
281 have been those that lived close to the highest densities of hunter-gatherers.

282

283 In the Balkans, Copper Age populations (Balkans\_Chalcolithic) harbor significantly more  
284 hunter-gatherer-related ancestry than Neolithic populations as shown, for example, by the  
285 statistic D(Mbuti, WHG, Balkans\_Neolithic, Balkans\_Chalcolithic); Z=4.3 ( Supplementary  
286 Data Table 2). This is roughly contemporary with the “resurgence” of hunter-gatherer  
287 ancestry previously reported in central Europe and Iberia<sup>7,10,42</sup> and is consistent with changes  
288 in funeral rites, specifically the reappearance around 4500 BCE of the Mesolithic tradition of  
289 extended supine burial – in contrast to the Early Neolithic tradition of flexed burial.<sup>43</sup> Four  
290 individuals associated with the Copper Age Trypillian population have ~80% NW Anatolian-  
291 related ancestry (Supplementary Table 3), confirming that the ancestry of the first farmers of  
292 present-day Ukraine was largely derived from the same source as the farmers of Anatolia and  
293 western Europe. Their ~20% hunter-gatherer ancestry is intermediate between WHG and  
294 EHG, consistent with deriving from the Neolithic hunter-gatherers of the region.

295

296 We also report the first genetic data associated with the Late Neolithic Globular Amphora  
297 Complex. Individuals from two Globular Amphora sites in Poland and Ukraine form a tight



298 cluster, showing high similarity over a large distance (Figure 1B,D). Both Globular Amphora  
299 Complex groups of samples had more hunter-gatherer-related ancestry than Middle Neolithic  
300 groups from Central Europe<sup>7</sup> (we estimate 25% [CI: 22-27%] WHG ancestry, similar to  
301 Chalcolithic Iberia, Supplementary Data Table 3). In east-central Europe, the Globular  
302 Amphora Complex preceded or abutted the Corded Ware Complex that marks the appearance  
303 of steppe-related ancestry,<sup>7,15</sup> while in southeastern Europe, the Globular Amphora Complex  
304 bordered populations with steppe-influenced material cultures for hundreds of years<sup>44</sup> and yet  
305 the individuals in our study have no evidence of steppe-related ancestry, providing support for  
306 the hypothesis that this material cultural frontier was also a barrier to gene flow.

307

308 The movements from the Pontic-Caspian steppe of individuals similar to those associated  
309 with the Yamnaya Cultural Complex in the 3<sup>rd</sup> millennium BCE contributed about 75% of the  
310 ancestry of individuals associated with the Corded Ware Complex and about 50% of the  
311 ancestry of succeeding material cultures such as the Bell Beaker Complex in central  
312 Europe.<sup>7,15</sup> In two directly dated individuals from southeastern Europe, one (ANI163) from  
313 the Varna I cemetery dated to 4711-4550 BCE and one (I2181) from nearby Smyadovo dated  
314 to 4550-4450 BCE, we find far earlier evidence of steppe-related ancestry (Figure 1B,D).  
315 These findings push back the first evidence of steppe-related ancestry this far West in Europe  
316 by almost 2,000 years, but it was sporadic as other Copper Age (~5000-4000 BCE) individuals  
317 from the Balkans have no evidence of it. Bronze Age (~3400-1100 BCE) individuals do have  
318 steppe-related ancestry (we estimate 30%; CI: 26-35%), with the highest proportions in the  
319 four latest Balkan Bronze Age individuals in our data (later than ~1700 BCE) and the least in  
320 earlier Bronze Age individuals (3400-2500 BCE; Figure 1D).

321

### 322 **A novel source of ancestry in Neolithic Europe**

323 An important question about the initial spread of farming into Europe is whether the first  
324 farmers that brought agriculture to northern Europe and to southern Europe were derived from  
325 a single population or instead represent distinct migrations. We confirm that Mediterranean  
326 populations, represented in our study by individuals associated with the Epicardial Early  
327 Neolithic from Iberia<sup>7</sup>, are closely related to Danubian populations represented by the  
328 *Linearbandkeramik* (LBK) from central Europe<sup>7,45</sup> and that both are closely related to the  
329 Balkan Neolithic population. These three populations form a clade with the NW Anatolian  
330 Neolithic individuals as an outgroup, consistent with a single migration into the Balkan  
331 peninsula, which then split into two (Supplementary Information Note 3).

332

333 In contrast, five southern Greek Neolithic individuals (Peloponnese\_Neolithic) – three (plus  
334 one previously published<sup>26</sup>) from Diros Cave and one from Franchthi Cave – are not

335 consistent with descending from the same source population as other European farmers. *D*-  
336 statistics (Supplementary Information Table 2) show that in fact, these “Peloponnese  
337 Neolithic” individuals dated to ~4000 BCE are shifted away from WHG and towards CHG,  
338 relative to Anatolian and Balkan Neolithic individuals. We see the same pattern in a single  
339 Neolithic individual from Krepost in present-day Bulgaria (I0679\_d, 5718-5626 BCE). An  
340 even more dramatic shift towards CHG has been observed in individuals associated with the  
341 Bronze Age Minoan and Mycenaean cultures,<sup>26</sup> and thus there was gene flow into the region  
342 from populations with CHG-rich ancestry throughout the Neolithic, Chalcolithic and Bronze  
343 Age. Possible sources are related to the Neolithic population from the central Anatolian site of  
344 Tepecik Ciftlik,<sup>21</sup> or the Aegean site of Kumtepe,<sup>11</sup> who are also shifted towards CHG relative  
345 to NW Anatolian Neolithic samples, as are later Copper and Bronze Age Anatolians.<sup>10,26</sup>

346

#### 347 **Sex-biased admixture between hunter-gatherers and farmers**

348 We provide the first evidence for sex-biased admixture between hunter-gatherers and farmers  
349 in Europe, showing that the Middle Neolithic “resurgence” of hunter-gatherer-related  
350 ancestry<sup>7,42</sup> in central Europe and Iberia was driven more by males than by females (Figure  
351 3B&C, Supplementary Data Table 5, Extended Data Figure 4). To document this we used  
352 *qpAdm* to compute ancestry proportions on the autosomes and the X chromosome; since  
353 males always inherit their X chromosome from their mothers, differences imply sex-biased  
354 mixture. In the Balkan Neolithic there is no evidence of sex bias ( $Z=0.27$  where a positive *Z*-  
355 score implies male hunter-gatherer bias), nor in the LBK and Iberian\_Early Neolithic ( $Z=-$   
356  $0.22$  and  $0.74$ ). In the Copper Age there is clear bias: weak in the Balkans ( $Z=1.66$ ), but  
357 stronger in Iberia ( $Z=3.08$ ) and Central Europe ( $Z=2.74$ ). Consistent with this, hunter-gatherer  
358 mitochondrial haplogroups (haplogroup U)<sup>46</sup> are rare and within the intervals of genome-wide  
359 ancestry proportions, but hunter-gatherer-associated Y chromosomes (haplogroups I, R1 and  
360 C1)<sup>17</sup> are more common: 7/9 in the Iberian Neolithic/Copper Age and 9/10 in Middle-Late  
361 Neolithic Central Europe (Central\_MN and Globular\_Amphora) (Figure 3C).

362

#### 363 **No evidence that steppe-related ancestry moved through southeast Europe into Anatolia**

364 One version of the Steppe Hypothesis of Indo-European language origins suggests that Proto-  
365 Indo-European languages developed north of the Black and Caspian seas, and that the earliest  
366 known diverging branch – Anatolian – was spread into Asia Minor by movements of steppe  
367 peoples through the Balkan peninsula during the Copper Age around 4000 BCE.<sup>47</sup> If this were  
368 correct, then one way to detect evidence of it would be the appearance of large amounts of  
369 steppe-related ancestry first in the Balkan Peninsula, and then in Anatolia. However, our data  
370 show no evidence for this scenario. While we find sporadic examples of steppe-related  
371 ancestry in Balkan Copper and Bronze Age individuals, this ancestry is rare until the late

372 Bronze Age. Moreover, while Bronze Age Anatolian individuals have CHG-related  
373 ancestry,<sup>26</sup> they have neither the EHG-related ancestry characteristic of all steppe populations  
374 sampled to date,<sup>19</sup> nor the WHG-related ancestry that is ubiquitous in Neolithic southeastern  
375 Europe (Extended Data Figure 2, Supplementary Data Table 2). An alternative hypothesis is  
376 that the ultimate homeland of Proto-Indo-European languages was in the Caucasus or in Iran.  
377 In this scenario, westward movement contributed to the dispersal of Anatolian languages, and  
378 northward movement and mixture with EHG was responsible for the formation of a “Late  
379 Proto-Indo European”-speaking population associated with the Yamnaya Complex.<sup>13</sup> While  
380 this scenario gains plausibility from our results, it remains possible that Indo-European  
381 languages were spread through southeastern Europe into Anatolia without large-scale  
382 population movement or admixture.

### 383 Discussion

384 Our study shows that southeastern Europe consistently served as a genetic contact zone.  
385 Before the arrival of farming, the region saw interaction between diverged groups of hunter-  
386 gatherers, and this interaction continued after farming arrived. While this study has clarified  
387 the genomic history of southeastern Europe from the Mesolithic to the Bronze Age, the  
388 processes that connected these populations to the ones living today remain largely unknown.  
389 An important direction for future research will be to sample populations from the Bronze  
390 Age, Iron Age, Roman, and Medieval periods and to compare them to present-day  
391 populations to understand how these transitions occurred.

## 392 **Methods**

393

### 394 **Ancient DNA Analysis**

395 We extracted DNA and prepared next-generation sequencing libraries in four different  
396 dedicated ancient DNA laboratories (Adelaide, Boston, Budapest, and Tuebingen). We also  
397 prepared samples for extraction in a fifth laboratory (Dublin), from whence it was sent to  
398 Boston for DNA extraction and library preparation (Supplementary Table 1).

399

400 Two samples were processed at the Australian Centre for Ancient DNA, Adelaide, Australia,  
401 according to previously published methods<sup>7</sup> and sent to Boston for subsequent screening,  
402 1240k capture and sequencing.

403

404 Seven samples were processed<sup>27</sup> at the Institute of Archaeology RCH HAS, Budapest,  
405 Hungary, and amplified libraries were sent to Boston for screening, 1240k capture and  
406 sequencing.

407

408 Seventeen samples were processed at the Institute for Archaeological Sciences of the  
409 University of Tuebingen and at the Max Planck Institute for the Science of Human History in  
410 Jena, Germany. Extraction<sup>48</sup> and library preparation<sup>49,50</sup> followed established protocols. We  
411 performed in-solution capture as described below (“1240k capture”) and sequenced on an  
412 Illumina HiSeq 4000 or NextSeq 500 for 76bp using either single- or paired-end sequencing.

413

414 The remaining 197 samples were processed at Harvard Medical School, Boston, USA. From  
415 about 75mg of sample powder from each sample (extracted in Boston or University College  
416 Dublin, Dublin, Ireland), we extracted DNA following established methods<sup>48</sup> replacing the  
417 column assembly with the column extenders from a Roche kit.<sup>51</sup> We prepared double  
418 barcoded libraries with truncated adapters from between one ninth and one third of the DNA  
419 extract. Most libraries included in the nuclear genome analysis (90%) were subjected to  
420 partial (“half”) Uracil-DNA-glycosylase (UDG) treatment before blunt end repair. This  
421 treatment reduces by an order of magnitude the characteristic cytosine-to-thymine errors of  
422 ancient DNA data<sup>52</sup>, but works inefficiently at the 5’ ends,<sup>50</sup> thereby leaving a signal of  
423 characteristic damage at the terminal ends of ancient sequences. Some libraries were not  
424 UDG-treated (“minus”). For some samples we increased coverage by preparing additional  
425 libraries from the existing DNA extract using the partial UDG library preparation, but  
426 replacing the MinElute column cleanups in between enzymatic reactions with magnetic bead  
427 cleanups, and the final PCR cleanup with SPRI bead cleanup.<sup>53,54</sup>

428 We screened all libraries from Adelaide, Boston and Budapest by enriching for the  
429 mitochondrial genome plus about 3,000 (50 in an earlier, unpublished, version) nuclear SNPs  
430 using a bead-capture<sup>55</sup> but with the probes replaced by amplified oligonucleotides synthesized  
431 by CustomArray Inc. After the capture, we completed the adapter sites using PCR, attaching  
432 dual index combinations<sup>56</sup> to each enriched library. We sequenced the products of between  
433 100 and 200 libraries together with the non-enriched libraries (shotgun) on an Illumina  
434 NextSeq500 using v2 150 cycle kits for 2x76 cycles and 2x7 cycles.

435

436 In Boston, we performed two rounds of in-solution enrichment (“1240k capture”) for a  
437 targeted set of 1,237,207 SNPs using previously reported protocols.<sup>7,14,23</sup> For a total of 34  
438 individuals, we increased coverage by building one to eight additional libraries for the same  
439 sample. When we built multiple libraries from the same extract, we often pooled them in  
440 equimolar ratios before the capture. We performed all sequencing on an Illumina NextSeq500  
441 using v2 150 cycle kits for 2x76 cycles and 2x7 cycles. We attempted to sequence each  
442 enriched library up to the point where we estimated that it was economically inefficient to  
443 sequence further. Specifically, we iteratively sequenced more and more from each individual  
444 and only stopped when we estimated that the expected increase in the number of targeted  
445 SNPs hit at least once would be less than about one for every 100 new read pairs generated.  
446 After sequencing, we trimmed two bases from the end of each read and aligned to the human  
447 genome (b37/hg19) using *bwa*.<sup>57</sup> We then removed individuals with evidence of  
448 contamination based on mitochondrial DNA polymorphism<sup>58</sup> or difference in PCA space  
449 between damaged and undamaged reads<sup>59</sup>, a high rate of heterozygosity on chromosome X  
450 despite being male<sup>59,60</sup>, or an atypical ratio of X-to-Y sequences. We also removed individuals  
451 that had low coverage (fewer than 15,000 SNPs hit on the autosomes). We report, but do not  
452 analyze, data from nine individuals that were first-degree relatives of others in the dataset  
453 (determined by comparing rates of allele sharing between pairs of individuals).

454

455 After removing a small number of sites that failed to capture, we were left with a total of  
456 1,233,013 sites of which 32,670 were on chromosome X and 49,704 were on chromosome Y,  
457 with a median coverage at targeted SNPs on the 214 newly reported individuals of 0.90  
458 (range 0.007-9.2; Supplementary Table 1). We generated “pseudo-haploid” calls by selecting  
459 a single read randomly for each individual at each SNP. Thus, there is only a single allele  
460 from each individual at each site, but adjacent alleles might come from either of the two  
461 haplotypes of the individual. We merged the newly reported data with previously reported  
462 data from 274 other ancient individuals<sup>9-11,15-27</sup>, making pseudo-haploid calls in the same way  
463 at the 1240k sites for individuals that were shotgun sequenced rather than captured.

464

465 Using the captured mitochondrial sequence from the screening process, we called  
466 mitochondrial haplotypes. Using the captured SNPs on the Y chromosome, we called Y  
467 chromosome haplogroups for males by restricting to sequences with mapping quality  $\geq 30$  and  
468 bases with base quality  $\geq 30$ . We determined the most derived mutation for each individual,  
469 using the nomenclature of the International Society of Genetic Genealogy  
470 (<http://www.isogg.org>) version 11.110 (21 April 2016).

471

## 472 **Population genetic analysis**

473 To analyze these ancient individuals in the context of present day genetic diversity, we  
474 merged them with the following two datasets:

475

476 1. 300 high coverage genomes from a diverse worldwide set of 142 populations  
477 sequenced as part of the Simons Genome Diversity Project<sup>28</sup> (SGDP merge).

478

479 2. 799 West Eurasian individuals genotyped on the Human Origins array<sup>23</sup>, with  
480 597,573 sites in the merged dataset (HO merge).

481

482 We computed principal components of the present-day individuals in the HO merge and  
483 projected the ancient individuals onto the first two components using the “*lsqproject: YES*”  
484 option in *smartpca* (*v15100*)<sup>61</sup> (<https://www.hsph.harvard.edu/alkes-price/software/>).

485

486 We ran *ADMIXTURE* (*v1.3.0*) in both supervised and unsupervised mode. In supervised mode  
487 we used only the ancient individuals, on the full set of SNPs, and the following population  
488 labels fixed:

- 489 • *Anatolia\_Neolithic*
- 490 • *WHG*
- 491 • *EHG*
- 492 • *Yamnaya*

493

494 For unsupervised mode we used the HO merge, including 799 present-day individuals. We  
495 flagged individuals that were genetic outliers based on PCA and *ADMIXTURE*, relative to  
496 other individuals from the same time period and archaeological culture.

497

498 We computed *D*-statistics using *qpDstat* (*v710*). *D*-statistics of the form  $D(A,B,X,Y)$  test the  
499 null hypothesis of the unrooted tree topology  $((A,B),(X,Y))$ . A positive value indicates that  
500 either A and X, or B and Y, share more drift than expected under the null hypothesis. We  
501 quote *D*-statistics as the *Z*-score computed using default block jackknife parameters.

502

503 We fitted admixture proportions with *qpAdm* (v610) using the SGDP merge. Given a set of  
504 outgroup (“right”) populations, *qpAdm* models one of a set of source (“left”) populations (the  
505 “test” population) as a mixture of the other sources by fitting admixture proportions to match  
506 the observed matrix of  $f_d$ -statistics as closely as possible. We report a p-value for the null  
507 hypothesis that the test population does not have ancestry from another source that is  
508 differentially related to the right populations. We computed standard errors for the mixture  
509 proportions using a block jackknife. Importantly, *qpAdm* does not require that the source  
510 populations are actually the admixing populations, only that they are a clade with the correct  
511 admixing populations, relative to the other sources. Infeasible coefficient estimates (i.e.  
512 outside [0,1]) are usually a sign of poor model fit, but in the case where the source with a  
513 negative coefficient is itself admixed, could be interpreted as implying that the true source is a  
514 population with different admixture proportions. We used the following set of seven  
515 populations as outgroups or “right populations”:

- 516 • *Mbuti.DG*
- 517 • *Ust\_Ishim\_HG\_published.DG*
- 518 • *Mota.SG*
- 519 • *MA1\_HG.SG*
- 520 • *Villabruna*
- 521 • *Papuan.DG*
- 522 • *Onge.DG*
- 523 • *Han.DG*

524

525 For some analyses where we required extra resolution (Extended Data Table 4) we used an  
526 extended set of 14 right (outgroup) populations, including additional Upper Paleolithic  
527 European individuals<sup>17</sup>:

- 528 • *ELMiron*
- 529 • *Mota.SG*
- 530 • *Mbuti.DG*
- 531 • *Ust\_Ishim\_HG\_published.DG*
- 532 • *MA1\_HG.SG*
- 533 • *AfontovaGora3*
- 534 • *GoyetQ116-1\_published*
- 535 • *Villabruna*
- 536 • *Kostenki14*
- 537 • *Vestonice16*
- 538 • *Karitiana.DG*
- 539 • *Papuan.DG*
- 540 • *Onge.DG*
- 541 • *Han.DG*

542

543 We also fitted admixture graphs with *qpGraph* (v6021)<sup>30</sup> ([https://github.com/DReichLab/](https://github.com/DReichLab/AdmixTools)  
544 *AdmixTools*, Supplementary Information, section 3). Like *qpAdm*, *qpGraph* also tries to  
545 match a matrix of  $f$ -statistics, but rather than fitting one population as a mixture of other,

546 specified, populations, it fits the relationship between all tested populations simultaneously,  
547 potentially incorporating multiple admixture events. However, *qpGraph* requires the graph  
548 relating populations to be specified in advance. We tested goodness-of-fit by computing the  
549 expected *D*-statistics under the fitted model, finding the largest *D*-statistic outlier between the  
550 fitted and observed model, and computing a *Z*-score using a block jackknife.

551

552 For 116 individuals with hunter-gatherer-related ancestry we estimated an effective migration  
553 surface using the software *EEMS* (<https://github.com/dipetkov/eems>)<sup>62</sup>. We computed  
554 pairwise differences between individuals using the *bed2diffs2* program provided with *EEMS*.  
555 We set the number of demes to 400 and defined the outer boundary of the region by the  
556 polygon (in latitude-longitude co-ordinates) [(66,60), (60,10), (45,-15), (35,-10), (35,60)]. We  
557 ran the MCMC ten times with different random seeds, each time with one million burn-in and  
558 four million regular iterations, thinned to one in ten thousand.

559

560 To analyze potential sex bias in admixture, we used *qpAdm* to estimate admixture proportions  
561 on the autosomes (default option) and on the X chromosome (option “*chrom: 23*”). We  
562 computed *Z*-scores for the difference between the autosomes and the X chromosome as  $Z =$   
563  $\frac{p_A - p_X}{\sqrt{\sigma_A^2 + \sigma_X^2}}$  where  $p_A$  and  $p_X$  are the hunter-gatherer admixture proportions on the autosomes and

564 the X chromosome, and  $\sigma_A$  and  $\sigma_X$  are the corresponding jackknife standard deviations. Thus,  
565 a positive *Z*-score means that there is more hunter-gatherer admixture on the autosomes than  
566 on the X chromosome, indicating that the hunter-gatherer admixture was male-biased.

567 Because X chromosome standard errors are high and *qpAdm* results can be sensitive to which  
568 population is first in the list of outgroup populations, we checked that the patterns we observe  
569 were robust to cyclic permutation of the outgroups. To compare frequencies of hunter-  
570 gatherer uniparental markers, we counted the individuals with mitochondrial haplogroup U  
571 and Y chromosome haplogroups C2, I2 and R1, which are all common in Mesolithic hunter-  
572 gatherers but rare or absent in Anatolian Neolithic individuals. The Iron Gates hunter-  
573 gatherers also carry H and K1 mitochondrial haplogroups so the proportion of haplogroup U  
574 represents the minimum maternal hunter-gatherer contribution. We computed binomial  
575 confidence intervals for the proportion of haplogroups associated with each ancestry type  
576 using the Agresti-Coull method<sup>63,64</sup> implemented in the *binom* package in *R*.

577

578 Given autosomal and X chromosome admixture proportions, we estimated the proportion of  
579 male and female hunter-gatherer ancestors by assuming a single-pulse model of admixture. If  
580 the proportions of male and female ancestors that are hunter-gatherer-related are given by  $m$   
581 and  $f$ , respectively, then the proportions of hunter-gatherer-related ancestry on the autosomes



582 and the X chromosome are given by  $\frac{m+f}{2}$  and  $\frac{m+2f}{3}$ . We approximated the sampling error in  
583 the observed admixture proportions by the estimated jackknife error and computed the  
584 likelihood surface for  $(m,f)$  over a grid ranging from (0,0) to (1,1).

585

### 586 **Direct AMS $^{14}\text{C}$ Bone Dates**

587 We report 113 new direct AMS  $^{14}\text{C}$  bone dates for 112 individuals from multiple AMS  
588 radiocarbon laboratories. In general, bone samples were manually cleaned and demineralized  
589 in weak HCl and, in most cases (PSU, UCIAMS, OxA), soaked in an alkali bath (NaOH) at  
590 room temperature to remove contaminating soil humates. Samples were then rinsed to  
591 neutrality in Nanopure  $\text{H}_2\text{O}$  and gelatinized in HCL.<sup>65</sup> The resulting gelatin was lyophilized  
592 and weighed to determine percent yield as a measure of collagen preservation (% crude  
593 gelatin yield). Collagen was then directly AMS  $^{14}\text{C}$  dated (Beta, AA) or further purified using  
594 ultrafiltration (PSU, UCIAMS, OxA, Poz, MAMS).<sup>66</sup> It is standard in some laboratories  
595 (PSU/UCIAMS, OxA) to use stable carbon and nitrogen isotopes as an additional quality  
596 control measure. For these samples, the %C, %N and C:N ratios were evaluated before AMS  
597  $^{14}\text{C}$  dating.<sup>67</sup> C:N ratios for well-preserved samples fall between 2.9 and 3.6, indicating good  
598 collagen preservation.<sup>68</sup> For 94 new samples, we also report  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values  
599 (Supplementary Table 6).

600

601 All  $^{14}\text{C}$  ages were  $\delta^{13}\text{C}$ -corrected for mass dependent fractionation with measured  $^{13}\text{C}/^{12}\text{C}$   
602 values<sup>69</sup> and calibrated with OxCal version 4.2.3<sup>70</sup> using the IntCal13 northern hemisphere  
603 calibration curve.<sup>70</sup> For hunter-gatherers from the Iron Gates, the direct  $^{14}\text{C}$  dates tend to be  
604 overestimates because of the freshwater reservoir effect (FRE), which arises because of a diet  
605 including fish that consumed ancient carbon, and for these individuals we performed a  
606 correction (Supplementary Information Note 1),<sup>71</sup> assuming that 100% FRE =  $545 \pm 70$  yr, and  
607  $\delta^{15}\text{N}$  values of 8.3% and 17.0% for 100% terrestrial and aquatic diets, respectively.

608

### 609 **Acknowledgments**

610 We thank David Anthony, Iosif Lazaridis, and Mark Lipson for comments on the manuscript,  
611 Bastien Llamas and Alan Cooper for contributions to laboratory work, Richard Evershed for  
612 contributing  $^{14}\text{C}$  dates and Friederike Novotny for assistance with samples. Support for this  
613 project was provided by the Human Frontier Science Program fellowship LT001095/2014-L  
614 to I.M.; by DFG grant AL 287 / 14-1 to K.W.A.; by Irish Research Council grant  
615 GOIPG/2013/36 to D.F.; by the NSF Archaeometry program BCS-1460369 to DJK (for AMS  
616  $^{14}\text{C}$  work at Penn State); by MEN-UEFISCDI grant, Partnerships in Priority Areas Program –

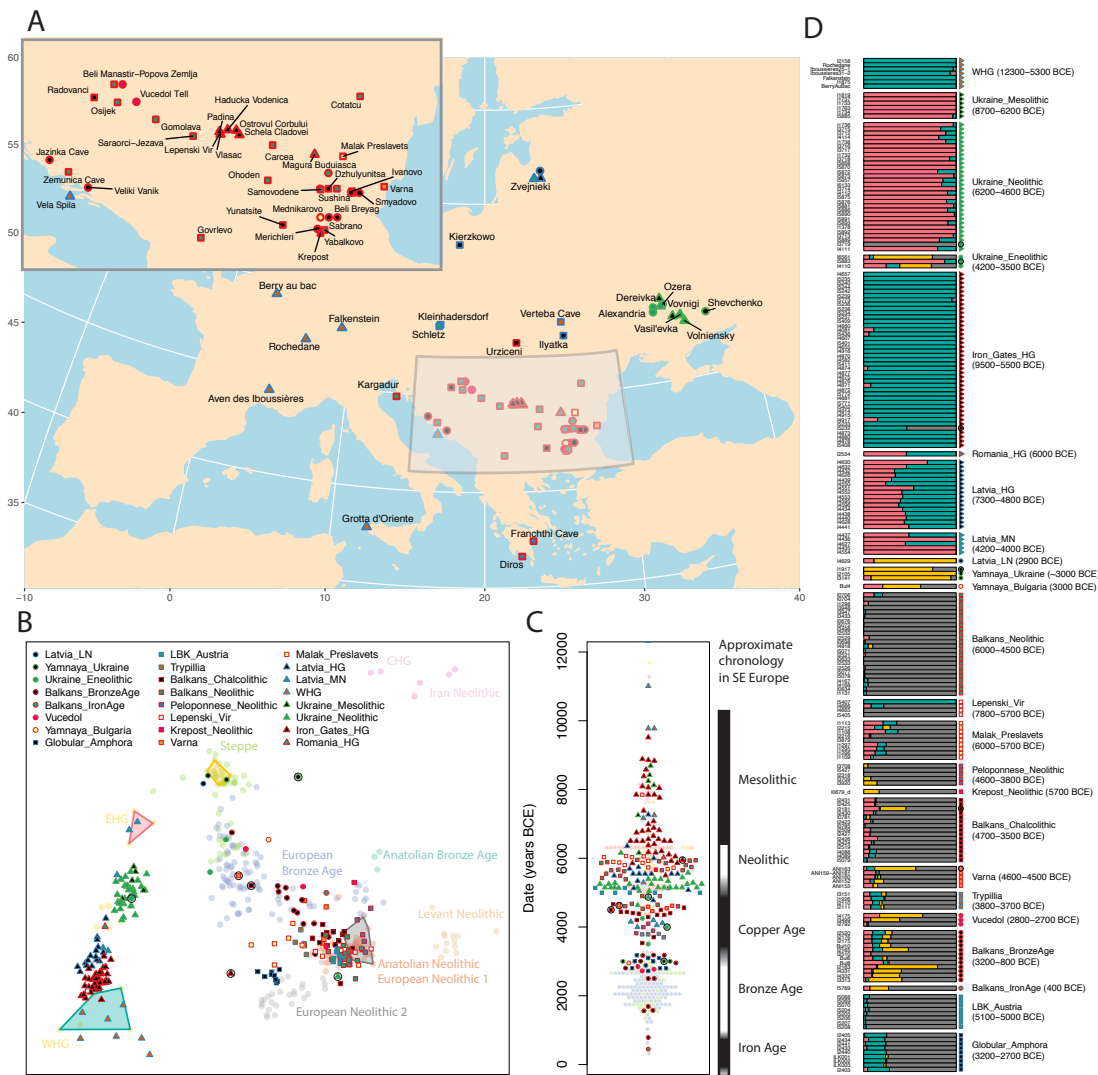
617 PN II (PN-II-PT-PCCA-2013-4-2302) to C.L.; by Croatian Science Foundation grant IP-  
618 2016-06-1450 to M.N.; by European Research Council grant ERC StG 283503 and Deutsche  
619 Forschungsgemeinschaft DFG FOR2237 to K.H.; by ERC starting grant ADNABIOARC  
620 (263441) to R.P.; and by US National Science Foundation HOMINID grant BCS-1032255,  
621 US National Institutes of Health grant GM100233, and the Howard Hughes Medical Institute  
622 to D.R.

623

## 624 **Author Contributions**

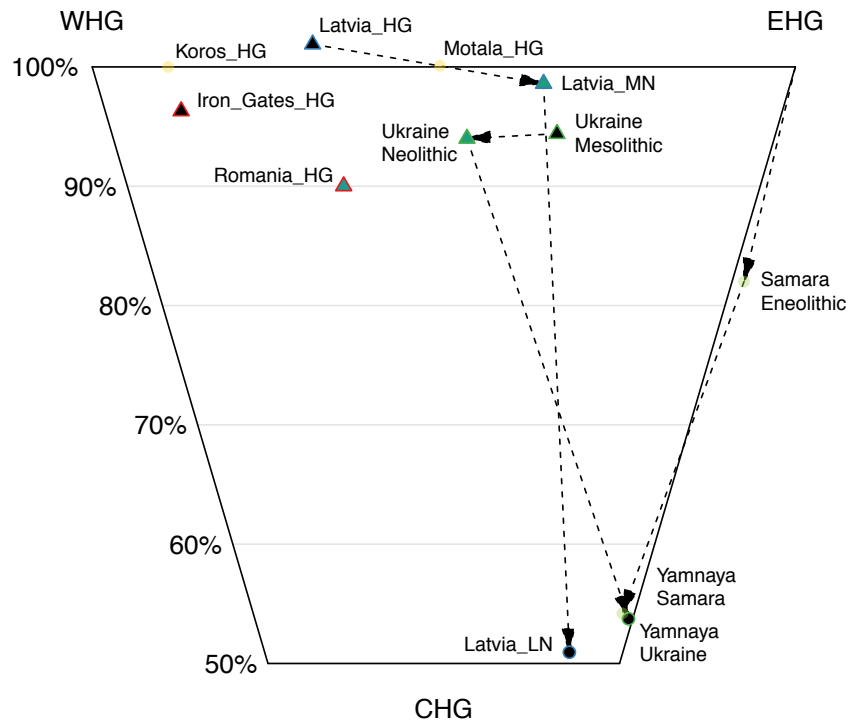
625 SAR, AS-N, SVai, SA, KWA, RA, DA, AA, NA, KB, MBG, HB, MB, ABo, YB, ABu, JB,  
626 SC, NC, RC, MC, CC, DD, NE, MFr, BGal, GG, BGe, THa, VH, KH, THi, SI, IJ, IKa, DKa,  
627 AK, DLa, MLa, CL, MLe, KL, DLV, DLo, IL, MMa, FM, KM, HM, MMe, PM, VM, VP,  
628 TDP, ASi, LS, MŠ, VS, PS, ASt, TS, MT-N, CT, IV, FVa, SVas, FVe, SV, EV, BV, CV, JZ,  
629 SZ, PWS, GC, RK, DC, GZ, BGay, MLi, AGN, IP, AP, DB, CB, JK, RP & DR assembled  
630 and interpreted archaeological material. CP, AS-N, NR, NB, FC, OC, DF, MFe, BGam, GGF,  
631 WH, EH, EJ, DKe, BK-K, IKu, MMi, AM, KN, MN, JO, SP, KSi, KSt & SVai performed  
632 laboratory work. IM, CP, AS-N, SM, IO, NP & DR analyzed data. DJK, ST, DB, CB  
633 interpreted <sup>14</sup>C dates. JK, RP & DR supervised analysis or laboratory work. IM & DR wrote  
634 the paper, with input from all co-authors.

635 **Figures**



636

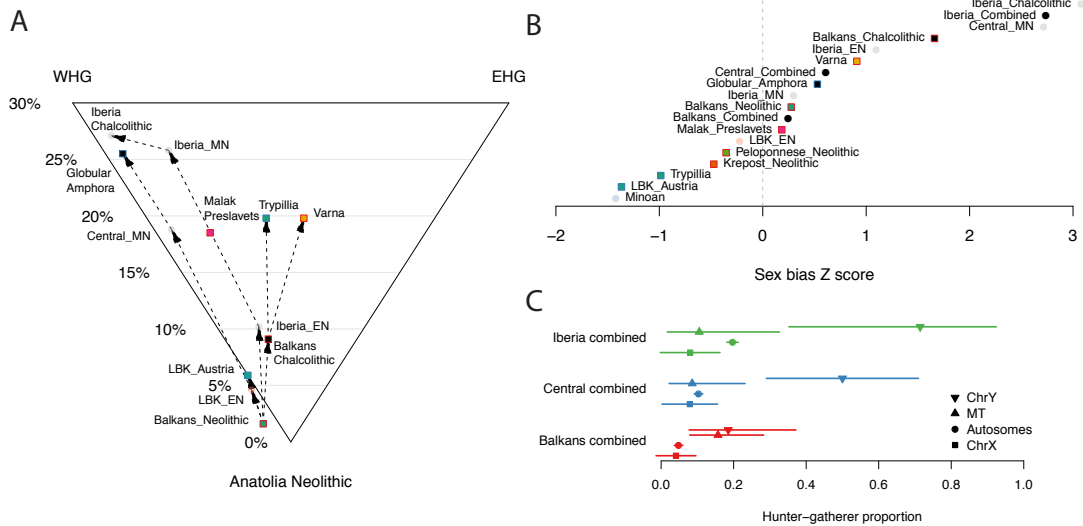
637 **Figure 1:** Geographic locations and genetic structure of newly reported individuals. **A:**  
 638 Location and groupings of newly reported individuals. **B:** Individuals projected onto axes  
 639 defined by the principal components of 799 present-day West Eurasians (not shown in this  
 640 plot for clarity, but shown in Extended Data Figure 1). Projected points include selected  
 641 published individuals (faded colored circles, labeled) and newly reported individuals (other  
 642 symbols; outliers shown by additional black circles). Colored polygons indicate the  
 643 individuals that had cluster memberships fixed at 100% for the supervised admixture analysis  
 644 in **D**. **C:** Estimated age (direct or contextual) for each sample. Approximate chronology used  
 645 in southeastern Europe shown to the right **D:** Supervised ADMIXTURE plot, modeling each  
 646 ancient individual (one per row), as a mixture of populations represented by clusters  
 647 containing Anatolian Neolithic (grey), Yamnaya from Samara (yellow), EHG (pink) and  
 648 WHG (green). Dates indicate approximate range of individuals in each population. Map data  
 649 in **A** from the *R* package *mapdata*.



650

651 **Figure 2:** Structure and population change in European populations with hunter-gatherer-  
652 related ancestry. This figure shows inferred ancestry proportions for populations modeled as a  
653 mixture of WHG, EHG and CHG (Supplementary Table S3.1.3). Dashed lines show  
654 populations from the same geographic region. Standard errors range from 1.5-8.3%  
655 (Supplementary Table S3.1.3).

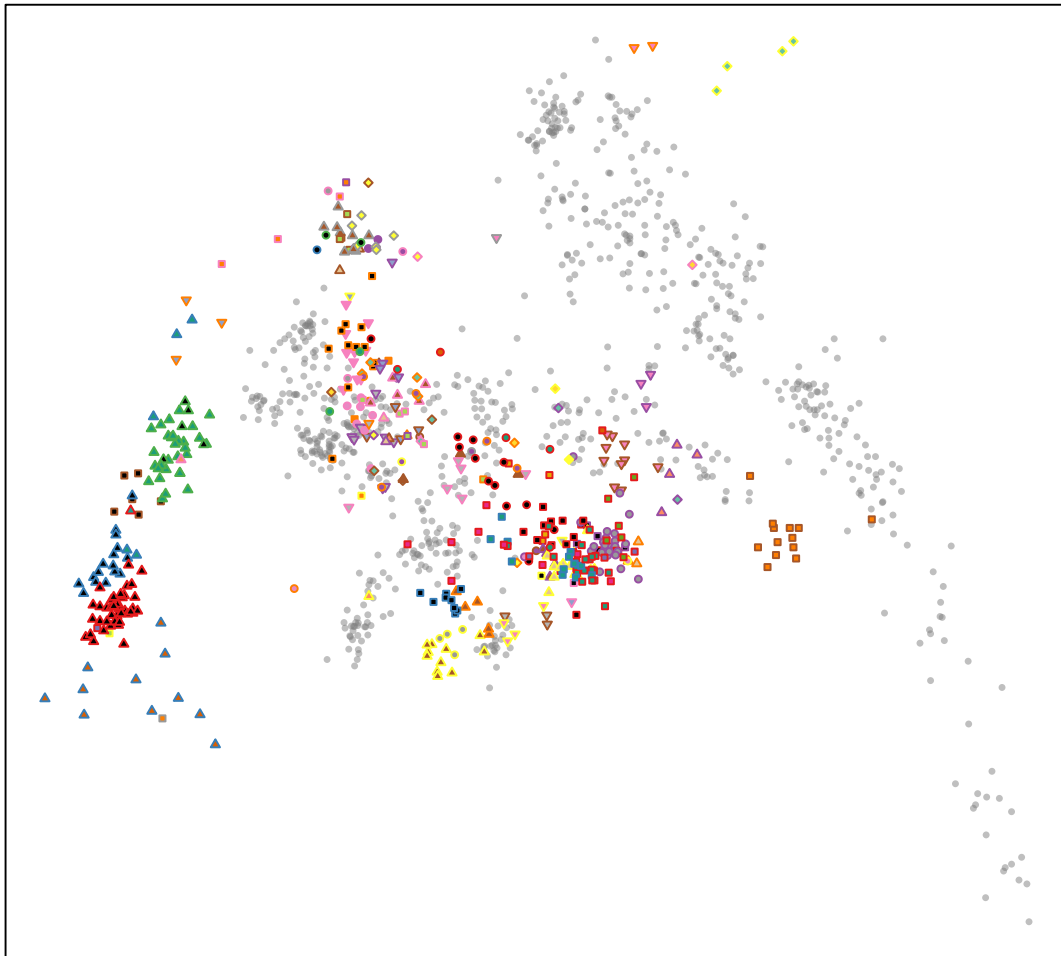
656



657

658 **Figure 3:** Structure and population change in European populations with NW Anatolian  
 659 Neolithic-related ancestry. **A:** each population is modeled as a mixture of NW Anatolia  
 660 Neolithic, WHG, and EHG. Dashed lines show temporal relationships between populations  
 661 from the same geographic areas with similar ancestries. Standard errors range from 0.7-6.0%  
 662 (Supplementary Table S3.2.2). **B:** Z-scores for the difference in hunter-gatherer-related  
 663 ancestry on the autosomes compared to the X chromosome when populations are modeled as  
 664 a mixture of NW Anatolia Neolithic and WHG. A positive score indicates that there is more  
 665 hunter-gatherer-related ancestry on the autosomes and therefore the hunter-gatherer-related  
 666 ancestry is male-biased. **C:** Hunter-gatherer-related ancestry proportions on the autosomes, X  
 667 chromosome, mitochondrial DNA (i.e. mt haplogroup U), and the Y chromosome (i.e. Y  
 668 chromosome haplogroups I2, R1 and C2). Bars show approximate 95% confidence intervals.  
 669 “Combined” populations merge all individuals from different times from a geographic area.  
 670

671 **Extended Data Figures**

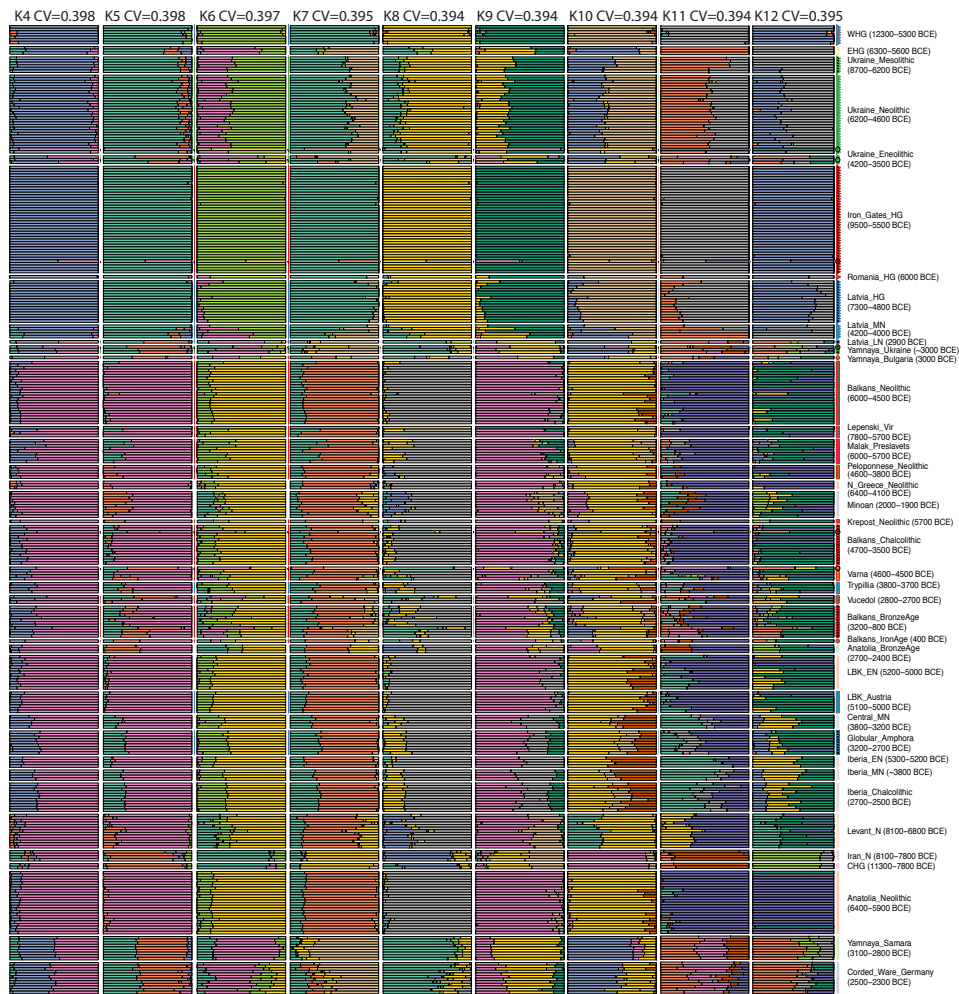


672

- |   |                         |                              |                           |
|---|-------------------------|------------------------------|---------------------------|
| ● Afanasievo.SG                         | ■ Germany_Bronze_Age.SG | ◆ Nordic_BA.SG               | ▲ Yamnaya_Samara          |
| ■ AfontovaGora3                         | ◆ GoyetQ116-1_published | ▲ Nordic_LBA.SG              | ▼ Yamnaya_Ukraine_outlier |
| ◆ Alberstedt_LN                         | ▲ Greece_Neolithic      | ● Nordic_LN.SG               | ● Latvia_LN               |
| ▼ ALPc_MN                               | ▼ Halberstadt_LBA       | ● Nordic_MN_B.SG             | ● Yamnaya_Ukraine         |
| ▲ Anatolia_BronzeAge                    | ● Hungary_LBA           | ● Poltavka                   | ● Ukraine_Eneolithic      |
| ● Anatolia_Neolithic                    | ● Hungary_Mako_EBA      | ▲ Poltavka_outlier           | ● Balkans_BronzeAge       |
| ■ Anatolia_Neolithic_Boncuklu.SG        | ● Hungary_MBA.SG        | ▲ Potapovka                  | ● Balkans_IronAge         |
| ◆ Anatolia_Neolithic_Kumtepe.SG         | ▲ Iberia_Chalcolithic   | ▼ Remedello_BA.SG            | ● Vucedol                 |
| ▲ Anatolia_Neolithic_Tepecik_Ciftlik.SG | ▼ Iberia_EN             | ● Russia_EBA.SG              | ● Yamnaya_Bulgaria        |
| ▼ Andronovo.SG                          | ● Iberia_MN             | ■ Samara_Eneolithic          | ■ Globular_Amphora        |
| ● Baden_LCA                             | ● Iceman_MN.SG          | ◆ Scythian_IA                | ■ LBK_Austria             |
| ■ Balkans_Chalcolithic_outlier          | ● Iran_N                | ▲ Sintashta_MBA_RISE.SG      | ■ Trypillia               |
| ◆ BattleAxe_Sweden.SG                   | ▲ Iron_Gates_HG_outlier | ▼ Srubnaya                   | ■ Balkans_Chalcolithic    |
| ▲ Bell_Beaker_Czech.SG                  | ▼ Karsdorf_LN           | ● Srubnaya_Outlier           | ■ Balkans_Neolithic       |
| ▼ Bell_Beaker_Germany                   | ● Koros_EN              | ■ Starcevo                   | ■ Peloponnese_Neolithic   |
| ● Bell_Beaker_Germany.SG                | ● Koros_HG              | ◆ Starouetice_EBA.SG         | ■ Lepenski_Vir            |
| ■ BenzigerodeHeimbürg_LN                | ◆ Kostenki14            | ▲ Ukraine_Eneolithic_outlier | ■ Krepost_Neolithic       |
| ● Buekk_MN                              | ▲ LBK_EN                | ▼ Ukraine_Neolithic_outlier  | ■ Varna                   |
| ▲ Central_MN                            | ▼ LBKT_MN               | ● Unetice_EBA                | ■ Malak_Preslavets        |
| ▼ CHG                                   | ● Lengyel_LN            | ■ Unetice_EBA.SG             | ▲ Latvia_HG               |
| ● Corded_Ware_Estonia.SG                | ● Levant_N              | ◆ Ust_Ishim_HG_published.DG  | ▲ Latvia_MN               |
| ■ Corded_Ware_Germany                   | ◆ MA1_HG.SG             | ▲ Varna_outlier              | ▲ WHG                     |
| ◆ Corded_Ware_Germany.SG                | ▲ Maros.SG              | ▼ Vatia.SG                   | ▲ Ukraine_Mesolithic      |
| ▲ Corded_Ware_Proto_Unetice_Poland.SG   | ▼ Minoan                | ● Vestonice16                | ▲ Ukraine_Neolithic       |
| ▼ EHG                                   | ● Mota.SG               | ■ Villabruna                 | ▲ Iron_Gates_HG           |
| ● ElMiron                               | ■ Motala_HG             | ◆ Yamnaya_Kalmykia.SG        | ▲ Romania_HG              |

673

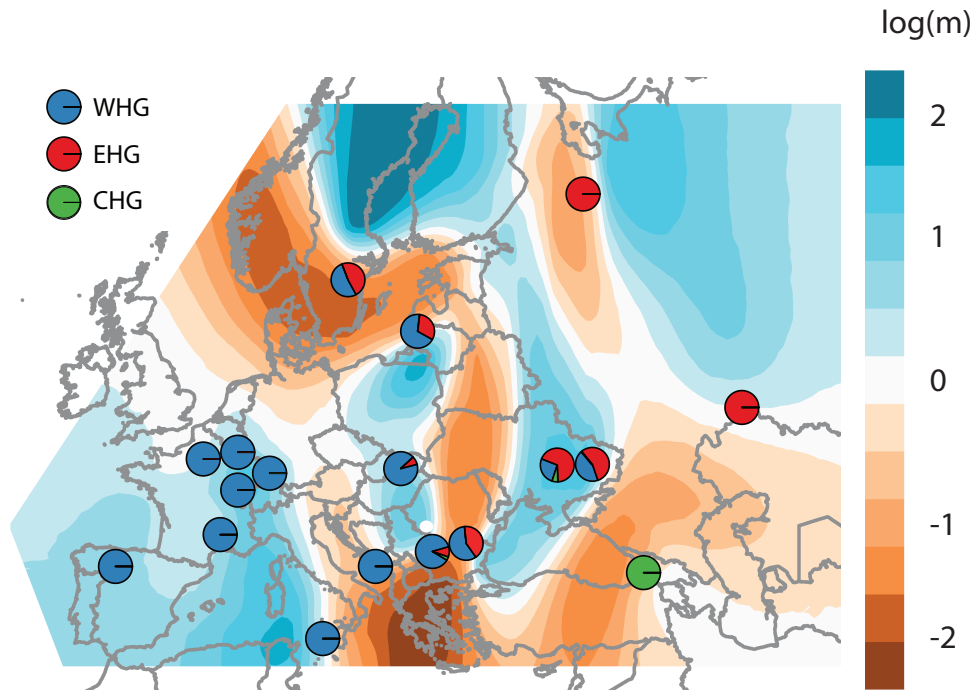
674 **Extended Data Figure 1: PCA of all ancient individuals, projected onto principal**  
 675 **components defined by 799 present-day West Eurasian individuals. (This differs from Figure**  
 676 **1B in that the plot is not cropped and the present-day individuals are shown.)**



677

678 **Extended Data Figure 2:** Unsupervised ADMIXTURE plot from k=4 to 12, on a dataset  
679 consisting of 1099 present-day individuals and 476 ancient individuals. We show newly  
680 reported ancient individuals and some previously published individuals for comparison.

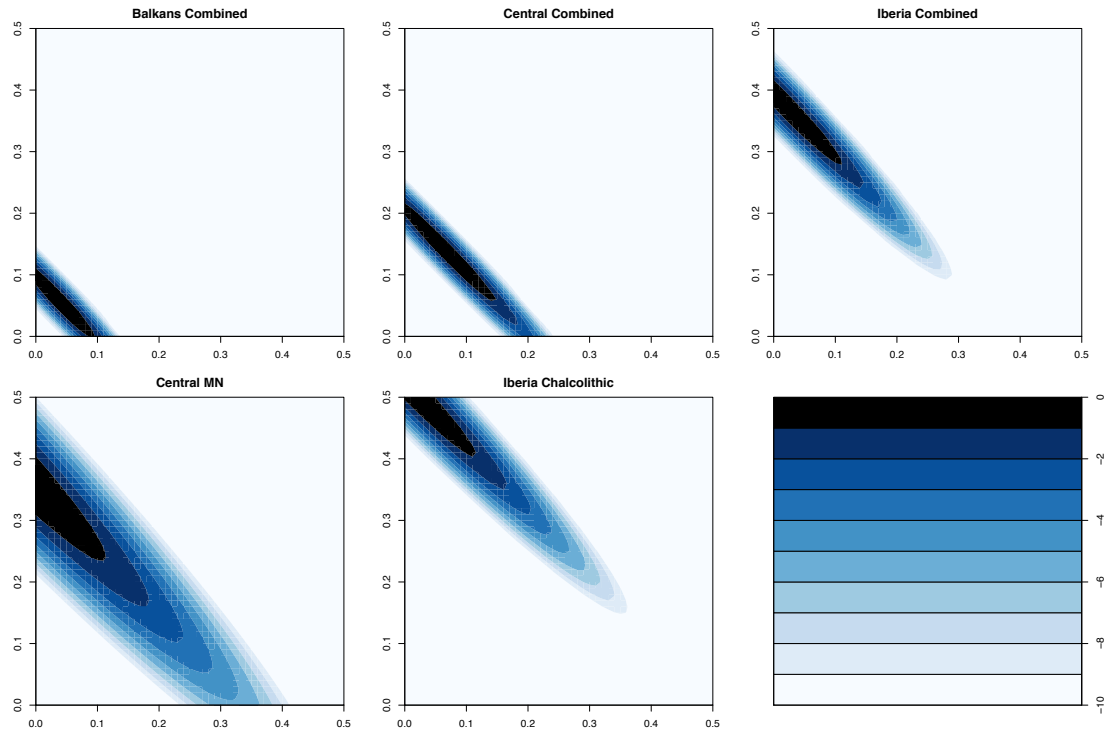
681



682

683 **Extended Data Figure 3:** Spatial structure in hunter-gatherers. Estimated effective migration  
684 surface (EEMS).<sup>62</sup> This fits a model of genetic relatedness where individuals move (in a  
685 random direction) from generation to generation on an underlying grid so that genetic  
686 relatedness is determined by distance. The migration parameter  $m$  defines the local rate of  
687 migration, varies on the grid and is inferred. This plot shows  $\log_{10} m$ , scaled relative to the  
688 average migration rate (which is arbitrary). Thus  $\log_{10}(m)=2$ , for example, implies that the  
689 rate of migration at this point on the grid is 100 times higher than average. To restrict as much  
690 as possible to hunter-gatherer structure, the migration surface is inferred using data from 116  
691 individuals from populations that date earlier than ~5000 BCE and have no NW Anatolian-  
692 related ancestry. Though the migration surface is sensitive to sampling, and fine-scale  
693 features may not be interpretable, the migration “barrier” (region of low migration) running  
694 north-south and separating populations with primarily WHG from primarily EHG ancestry  
695 seems to be robust, and consistent with inferred admixture proportions. This analysis suggests  
696 that Mesolithic hunter-gatherer population structure was clustered and not smoothly clinal, in  
697 the sense that genetic differentiation did not vary consistently with distance. Superimposed  
698 on this background, pies show the WHG, EHG and CHG ancestry proportions inferred for  
699 populations used to construct the migration surface (another way of visualizing the data in  
700 show in Figure 2, Supplementary Table 3.1.3 – we use two population models if they fit with  
701  $p>0.1$ , and three population models otherwise). Pies with only a single color have been fixed  
702 to be the source populations.





703

704 **Extended Data Figure 4:** log-likelihood surfaces for the proportion of female (x-axis) and  
705 male (y-axis) ancestors that are hunter-gatherer-related for the combined populations  
706 analyzed in Figure 3C, and the two populations with the strongest evidence for sex-bias. Log-  
707 likelihood scale ranges from 0 to -10, where 0 is the feasible point with the highest likelihood.

708

709

## 710 **Supplementary Tables**

- 711 **Supplementary Table 1:** Details of ancient individuals analyzed in this study.  
712 **Supplementary Table 2:** Key *D*-statistics to support statements about population history.  
713 **Supplementary Table 3:** *qpAdm* models with 7-population outgroup set.  
714 **Supplementary Table 4:** *qpAdm* models with extended 14-population outgroup set.  
715 **Supplementary Table 5:** *qpAdm* models for Neolithic populations for chromosome X.  
716 **Supplementary Table 6:** Additional <sup>14</sup>C dating information.

## 717 **References**

- 718 1 Tringham, R. E. in *The Transition to Agriculture in Prehistoric Europe* (ed D. Price)  
719 19-56 (Cambridge University Press, 2000).  
720 2 Bellwood, P. *First Farmers: The Origins of Agricultural Societies*. 2nd edn, (Wiley-  
721 Blackwell, 2004).  
722 3 Golitko, M. in *Ancient Europe, 8000 B.C. to A.D. 1000: An Encyclopedia of the*  
723 *Barbarian World* (eds P. Bogucki & P.J. Crabtree) 259-266 (Charles Scribners &  
724 Sons, 2003).  
725 4 Vander Linden, M. in *Investigating Archaeological Cultures: Material Culture,*  
726 *Variability, and Transmission* (eds B.W. Roberts & M. Vander Linden) 289-319  
727 (Springer, 2012).  
728 5 Bramanti, B. *et al.* Genetic discontinuity between local hunter-gatherers and central  
729 Europe's first farmers. *Science* **326**, 137-140 (2009).  
730 6 Skoglund, P. *et al.* Origins and genetic legacy of Neolithic farmers and hunter-  
731 gatherers in Europe. *Science* **336**, 466-469 (2012).  
732 7 Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European  
733 languages in Europe. *Nature* **522**, 207-211 (2015).  
734 8 Cassidy, L. M. *et al.* Neolithic and Bronze Age migration to Ireland and  
735 establishment of the insular Atlantic genome. *Proc. Natl. Acad. Sci. U. S. A.* **113**,  
736 368-373 (2016).  
737 9 Hofmanova, Z. *et al.* Early farmers from across Europe directly descended from  
738 Neolithic Aegeans. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6886-6891 (2016).  
739 10 Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians.  
740 *Nature* **528**, 499-503 (2015).  
741 11 Omrak, A. *et al.* Genomic Evidence Establishes Anatolia as the Source of the  
742 European Neolithic Gene Pool. *Curr. Biol.* **26**, 270-275 (2016).  
743 12 Müller, J., Rassmann, K. & Videiko, M. *Trypillia Mega-Sites and European*  
744 *Prehistory: 4100–3400 BCE.*, (Routledge, 2016).  
745 13 Anthony, D. W. *The horse the wheel and language*. (Princeton University Press,  
746 2007).  
747 14 Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal  
748 ancestor. *Nature* (2015).  
749 15 Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167-  
750 172 (2015).  
751 16 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western  
752 Siberia. *Nature* **514**, 445-449 (2014).  
753 17 Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205 (2016).  
754 18 Gallego Llorente, M. *et al.* Ancient Ethiopian genome reveals extensive Eurasian  
755 admixture in Eastern Africa. *Science* **350**, 820-822 (2015).  
756 19 Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians.  
757 *Nature communications* **6**, 8912 (2015).

- 758 20 Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as  
759 inferred by whole-genome sequencing. *Nature communications* **3**, 698 (2012).
- 760 21 Kilinc, G. M. *et al.* The Demographic Development of the First Farmers in Anatolia.  
761 *Curr. Biol.* (2016).
- 762 22 Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near  
763 East. *Nature* **536**, 419-424 (2016).
- 764 23 Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for  
765 present-day Europeans. *Nature* **513**, 409-413 (2014).
- 766 24 Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-  
767 old Mesolithic European. *Nature* **507**, 225-228 (2014).
- 768 25 Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of  
769 Native Americans. *Nature* **505**, 87-91 (2014).
- 770 26 Lazaridis, I. *et al.* Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214-  
771 218 (2017).
- 772 27 Lipson, M. *et al.* Parallel ancient genomic transects reveal complex population  
773 history of early European farmers. *bioRxiv*, <https://doi.org/10.1101/114488> (2017).
- 774 28 Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142  
775 diverse populations. *Nature* **538**, 201-206 (2016).
- 776 29 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry  
777 in unrelated individuals. *Genome Res.* **19**, 1655-1664 (2009).
- 778 30 Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093  
779 (2012).
- 780 31 Gronenborn, D. & Dolukhanov, P. in *The Oxford Handbook of Neolithic Europe*  
781 (eds C. Fowler, J. Harding, & D. Hofmann) 195-214 (Oxford University Press,  
782 2015).
- 783 32 Telegin, D. Ya. Neolithic cultures of Ukraine and their chronology. *Journal of World*  
784 *Prehistory* **I**, 307-331 (1987).
- 785 33 Telegin, D. Ya. & Potekhina, I. D. *Neolithic cemeteries and populations in the*  
786 *Dnieper Basin*. (British Archaeological Reports, 1987).
- 787 34 Jones, E. R. *et al.* The Neolithic Transition in the Baltic Was Not Driven by  
788 Admixture with Early European Farmers. *Curr. Biol.*, 2185-2193 (2017).
- 789 35 Mittnik, A. *et al.* The Genetic History of Northern Europe. *bioRxiv*,  
790 <https://doi.org/10.1101/113241> (2017).
- 791 36 Saag, L. *et al.* Extensive farming in Estonia started through a sex-biased migration  
792 from the Steppe. *Curr. Biol.* **27**, 2185-2193 (2017).
- 793 37 Maier, A. *The Central European Magdalenian: Regional Diversity and Internal*  
794 *Variability.*, (Springer, 2015).
- 795 38 Borić, D. & Price, T. D. Strontium isotopes document greater human mobility at the  
796 start of the Balkan Neolithic. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 3298-3303 (2013).
- 797 39 Krauß, R., Marinova, E., De Brue, H. & Weninger, B. The rapid spread of early  
798 farming from the Aegean into the Balkans via the Sub-Mediterranean-Aegean  
799 Vegetation Zone. *Quaternary International* **XXX**, 1-18 (2017).
- 800 40 Bacvarov, K. in *Moments in time: Papers Presented to Pál Raczky on His 60th*  
801 *Birthday* (eds A. Anders & G. Kulcsár) 29-34 (L'Harmattan, 2013).
- 802 41 Gurova, M. & Bonsall, C. 'Pre-Neolithic' in Southeast Europe: a Bulgarian  
803 perspective. *Documenta Praehistorica* **XLI**, 95-109 (2014).
- 804 42 Brandt, G. *et al.* Ancient DNA reveals key stages in the formation of central  
805 European mitochondrial genetic diversity. *Science* **342**, 257-261 (2013).
- 806 43 Borić, D. in *The Oxford Handbook of Neolithic Europe* (eds C. Fowler, J. Harding,  
807 & D. Hofmann) 927-957 (Oxford University Press, 2015).
- 808 44 Szymt, M. in *Transition to the Bronze Age (Archaeolingua 30)* (eds V. Heyd, G.  
809 Kulcsár, & V. Szeverényi) 93-111 (Archaeolingua, 2013).
- 810 45 Olalde, I. *et al.* A Common Genetic Origin for Early Farmers from Mediterranean  
811 Cardial and Central European LBK Cultures. *Mol. Biol. Evol.* **32**, 3132-3142 (2015).

- 812 46 Posth, C. *et al.* Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal  
813 of Non-Africans and a Late Glacial Population Turnover in Europe. *Curr. Biol.* **26**,  
814 827-833 (2016).
- 815 47 Anthony, D. W. & Ringe, D. The Indo-European Homeland from Linguistic and  
816 Archaeological Perspectives. *Annual Review of Linguistics* **1**, 199-219 (2015).
- 817 48 Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene  
818 cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S.*  
819 *A.* **110**, 15758-15763 (2013).
- 820 49 Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly  
821 multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**, pdb  
822 prot5448 (2010).
- 823 50 Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-  
824 glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B*  
825 *Biol. Sci.* **370**, 20130624 (2015).
- 826 51 Korlevic, P. *et al.* Reducing microbial and human contamination in DNA extractions  
827 from ancient bones and teeth. *BioTechniques* **59**, 87-93 (2015).
- 828 52 Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo  
829 methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
- 830 53 DeAngelis, M. M., Wang, D. G. & Hawkins, T. L. Solid-phase reversible  
831 immobilization for the isolation of PCR products. *Nucleic Acids Res.* **23**, 4742-4743  
832 (1995).
- 833 54 Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries  
834 for multiplexed target capture. *Genome Res.* **22**, 939-946 (2012).
- 835 55 Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of  
836 mitochondrial genomes using PCR products. *PLoS One* **5**, e14004 (2010).
- 837 56 Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in  
838 multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
- 839 57 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler  
840 transform. *Bioinformatics* **26**, 589-595 (2010).
- 841 58 Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial  
842 genomes. *Curr. Biol.* **23**, 553-559 (2013).
- 843 59 Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day  
844 contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2229-  
845 2234 (2014).
- 846 60 Korneliusson, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next  
847 Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
- 848 61 Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-  
849 wide association studies. *Nat. Genet.* **38**, 904-909 (2006).
- 850 62 Petkova, D., Novembre, J. & Stephens, M. Visualizing spatial population structure  
851 with estimated effective migration surfaces. *Nat. Genet.* **48**, 94-100 (2016).
- 852 63 Brown, L., Cai, T. & DasGupta, A. Interval Estimation for a Binomial Proportion.  
853 *Statistical Science* **16**, 101-133 (2001).
- 854 64 Agresti, A. & Coull, B. Approximate is better than "exact" for interval estimation of  
855 binomial proportions. *The American Statistician* **52**, 119-126 (1998).
- 856 65 Longin, R. New method of collagen extraction for radiocarbon analysis. *Nature* **230**,  
857 241-242 (1971).
- 858 66 Brown, T. A., Nelson, D. E., Vogel, J. S. & Southon, J. R. Improved collagen  
859 extraction by modified longin method. *Radiocarbon* **30**, 171-177 (1988).
- 860 67 Kennett, D. J. *et al.* Archaeogenomic evidence reveals prehistoric matrilineal dynasty.  
861 *Nature communications* **8**, 14115 (2017).
- 862 68 van Klinken, G. J. Bone collagen quality indicators for paleodietary and radiocarbon  
863 measurements. *JAS* **26**, 687-695 (1999).
- 864 69 Stuiver, M. & Polach, H. A. Discussion: Reporting of <sup>14</sup>C data. *Radiocarbon* **19**, 355-  
865 363 (1977).

866 70 Bronk Ramsey, C. *OxCal 4.23 Online Manual*,  
867 <[https://c14.arch.ox.ac.uk/oxcalhelp/hlp\\_contents.html](https://c14.arch.ox.ac.uk/oxcalhelp/hlp_contents.html)> (2013).  
868 71 Cook, G. T. *et al.* A freshwater diet-derived <sup>14</sup>C reservoir effect at the Stone Age sites  
869 in the Iron Gates gorge. *Radiocarbon* **43**, 453-460 (2001).  
870