The Genomic History of Southeastern Europe

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

* These authors contributed equally to the manuscript

[†] Present address; Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA 19104, USA

Correspondence to I.M. (mathi@upenn.edu) or D.R. (reich@genetics.med.harvard.edu) or R.P. (ron.pinhasi@ucd.ie)

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

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

Abstract

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

Introduction

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

151 archaeological evidence documents cultural contacts with peoples of the Eurasian steppe.[13]
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.[7,10,14] 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,[9-11,15-27] 799 present-day individuals genotyped on the Illumina "Human
168 Origins" array,[23] and 300 high coverage genomes from the Simons Genome Diversity Project
169 (SGDP).[28] We used principal component analysis (PCA; Figure 1B, Extended Data Figure 1),
170 supervised and unsupervised ADMIXTURE (Figure 1D, Extended Data Figure 2),[29] *D*-
171 statistics, *qpAdm* and *qpGraph,*[30] 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.[23] 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.[7] 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.[9,10]
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.[7,15] 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.[19,21,22]

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,[31-33] 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).[7] 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(Mbuti, WHG,
202 Ukraine_Mesolithic, 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.[13] All previously reported Yamnaya individuals were from Samara[7] and
209 Kalmykia[15] 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.[34-36] 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,[7] Kalmykia[15] and Ukraine (Figure 2).
230
231 We report new Upper Palaeolithic and Mesolithic data from southern and western Europe.[17]
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[23] (Luxembourg, 6100 BCE), Bichon[19] (Switzerland, 11,700 BCE), and Villabruna[17]
235 (Italy 12,000 BCE). These results demonstrate that WHG populations[23] 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.[17,37]
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[38] 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.[39] 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.[40] These individuals
272 lived in the mid-6th 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,[41] 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[7,10,42] 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.[43] 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[7] (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,[7,15] while in southeastern Europe, the Globular Amphora Complex
304 bordered populations with steppe-influenced material cultures for hundreds of years[44] 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 3rd 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.[7,15] 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[7], are closely related to Danubian populations represented by the
328 *Linearbandkeramik* (LBK) from central Europe[7,45] 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[26]) 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,[26] 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,[21] or the Aegean site of Kumtepe,[11] who are also shifted towards CHG relative
345 to NW Anatolian Neolithic samples, as are later Copper and Bronze Age Anatolians.[10,26]
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[7,42] 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)[46] are rare and within the intervals of genome-wide
359 ancestry proportions, but hunter-gatherer-associated Y chromosomes (haplogroups I, R1 and
360 C1)[17] 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.[47] 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,[26] they have neither the EHG-related ancestry characteristic of all steppe populations
374 sampled to date,[19] 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.[13] 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[7] and sent to Boston for subsequent screening,
402 1240k capture and sequencing.

403

404 Seven samples were processed[27] 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[48] and library preparation[49,50] 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[48] replacing the
417 column assembly with the column extenders from a Roche kit.[51] 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[52], but works inefficiently at the 5' ends,[50] 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.[53,54]

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[55] 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[56] 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.[7,14,23] 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*.[57] We then removed individuals with evidence of
448 contamination based on mitochondrial DNA polymorphism[58] or difference in PCA space
449 between damaged and undamaged reads[59], a high rate of heterozygosity on chromosome X
450 despite being male[59,60], 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[9-11,15-27], making pseudo-haploid calls in the same way
463 at the 1240k sites for individuals that were shotgun sequenced rather than captured.
464

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

Population genetic analysis

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

1. 300 high coverage genomes from a diverse worldwide set of 142 populations sequenced as part of the Simons Genome Diversity Project[28] (SGDP merge).

2. 799 West Eurasian individuals genotyped on the Human Origins array[23], with 597,573 sites in the merged dataset (HO merge).

We computed principal components of the present-day individuals in the HO merge and projected the ancient individuals onto the first two components using the "*lsqproject: YES*" option in *smartpca (v15100)*[61] (https://www.hsph.harvard.edu/alkes-price/software/).

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

- *Anatolia_Neolithic*
- *WHG*
- *EHG*
- *Yamnaya*

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

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

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_4-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[17]:

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)*[30] (https://github.com/DReichLab/
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)[62]. 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 σ_A and σ_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[63,64] 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}C Bone Dates**
587 We report 113 new direct AMS ^{14}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 H_2O and gelatinized in HCL.[65] 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}C dated (Beta, AA) or further purified using
594 ultrafiltration (PSU, UCIAMS, OxA, Poz, MAMS).[66] 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}C dating.[67] C:N ratios for well-preserved samples fall between 2.9 and 3.6, indicating good
598 collagen preservation.[68] For 94 new samples, we also report δ^{13}C and δ^{15}N values
599 (Supplementary Table 6).
600
601 All ^{14}C ages were δ^{13}C-corrected for mass dependent fractionation with measured ^{13}C/^{12}C
602 values[69] and calibrated with OxCal version 4.2.3[70] using the IntCal13 northern hemisphere
603 calibration curve.[70] For hunter-gatherers from the Iron Gates, the direct ^{14}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),[71] assuming that 100% FRE = 545±70 yr, and
607 δ^{15}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}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}C work at Penn State); by MEN-UEFISCDI grant, Partnerships in Priority Areas Program –

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

Author Contributions

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

Figures

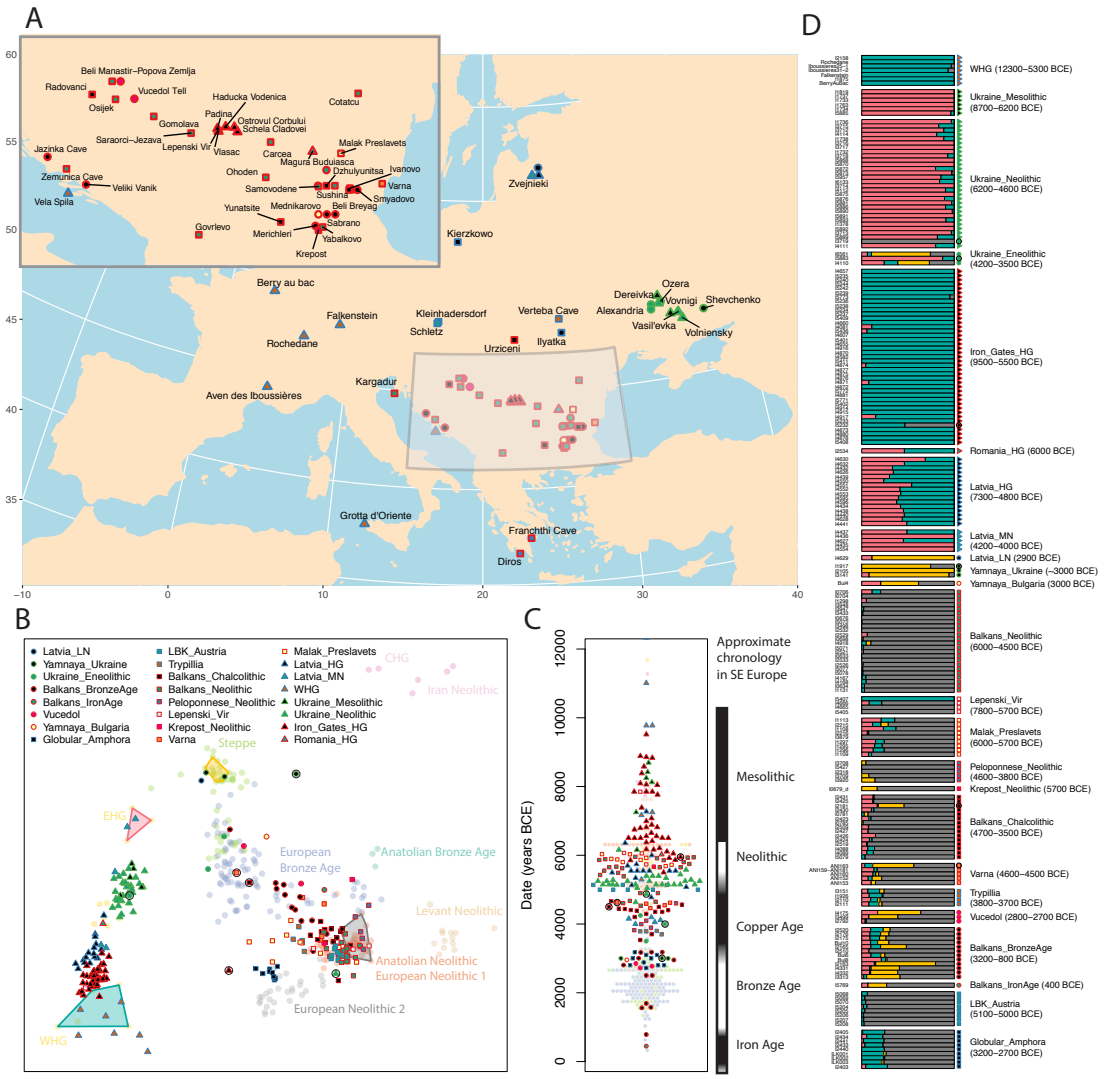


Figure 1: Geographic locations and genetic structure of newly reported individuals. **A**: Location and groupings of newly reported individuals. **B**: Individuals projected onto axes defined by the principal components of 799 present-day West Eurasians (not shown in this plot for clarity, but shown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

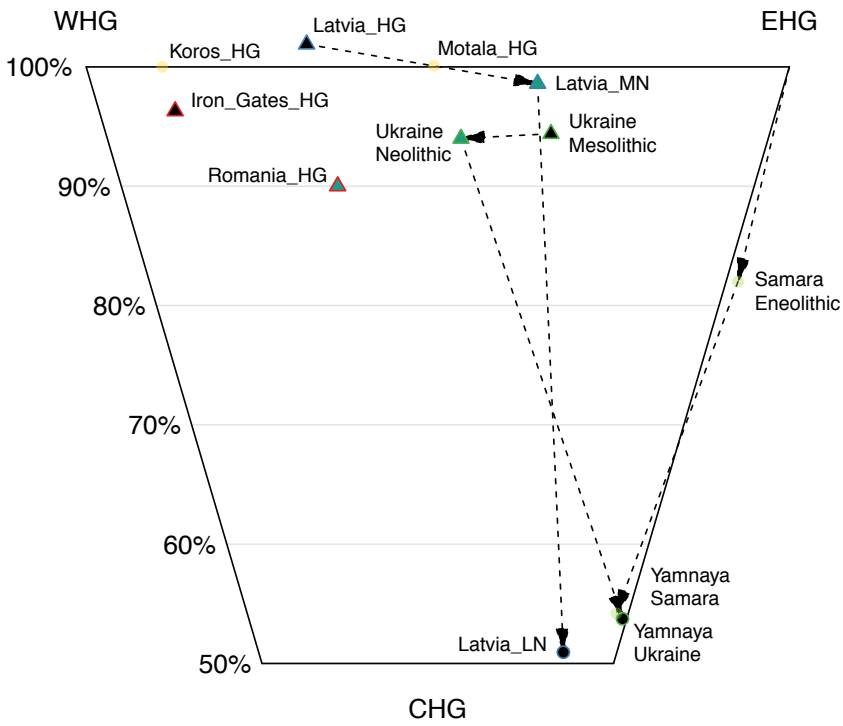


Figure 2: Structure and population change in European populations with hunter-gatherer-related ancestry. This figure shows inferred ancestry proportions for populations modeled as a mixture of WHG, EHG and CHG (Supplementary Table S3.1.3). Dashed lines show populations from the same geographic region.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

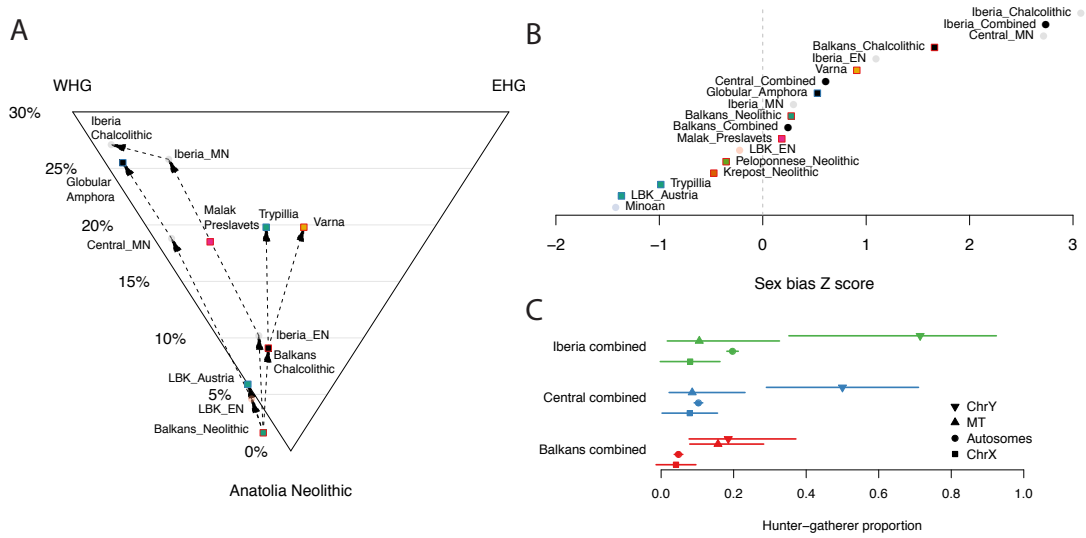
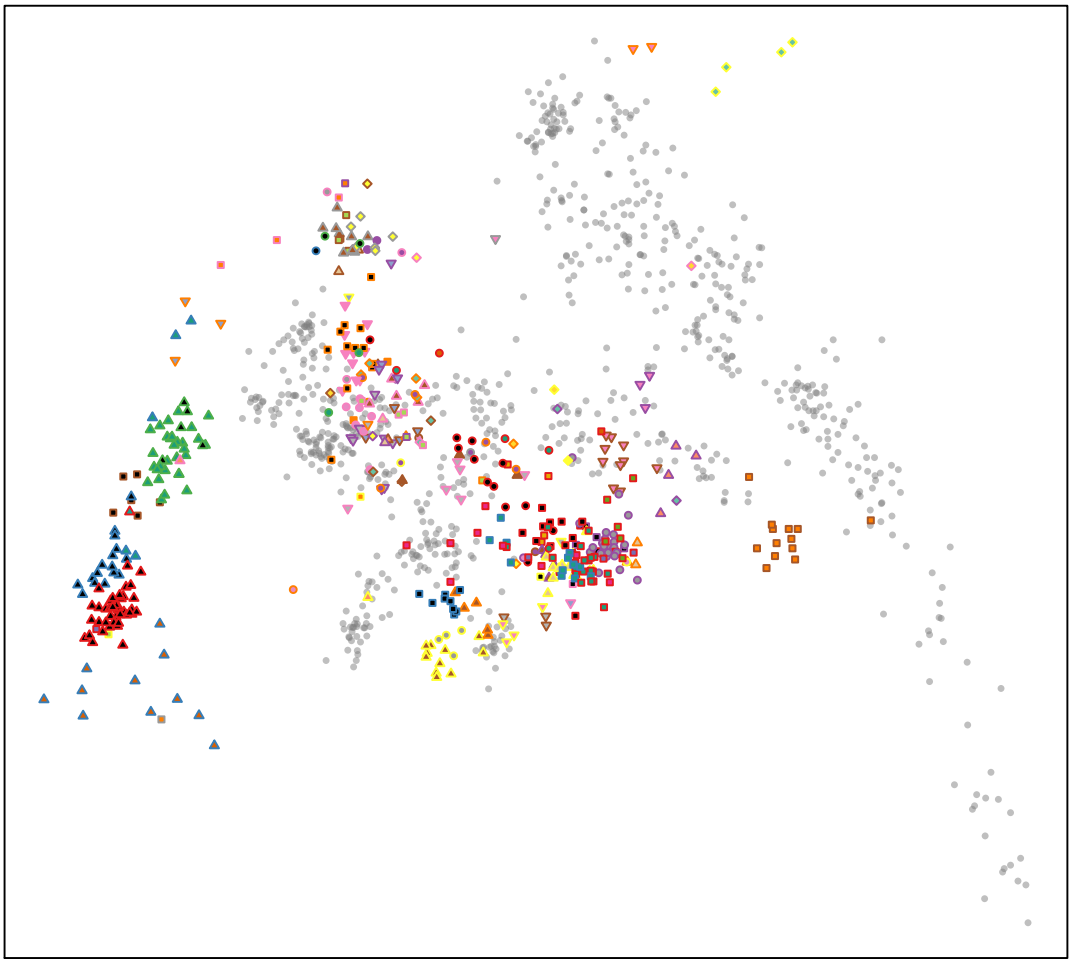


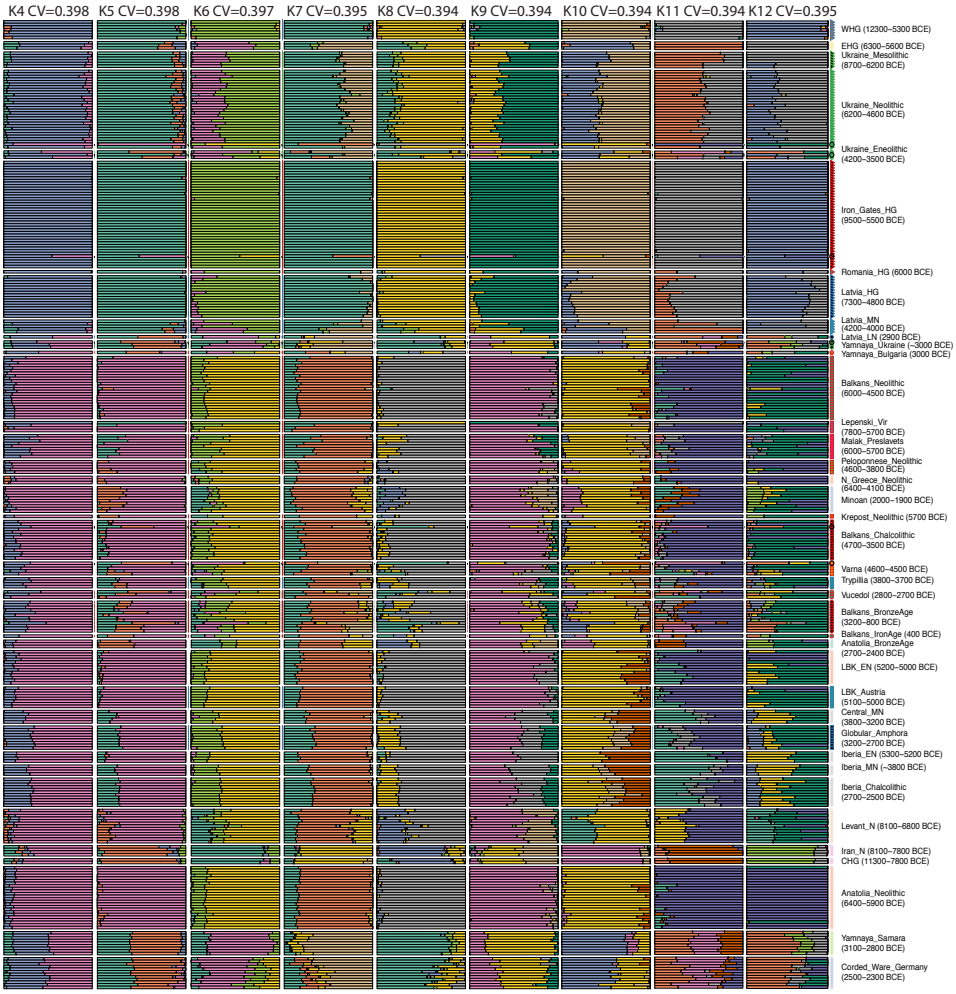
Figure 3: Structure and population change in European populations with NW Anatolian Neolithic-related ancestry. **A**: each population is modeled as a mixture of NW Anatolia Neolithic, WHG, and EHG. Dashed lines show temporal relationships between populations from the same geographic areas with similar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

Extended Data Figures

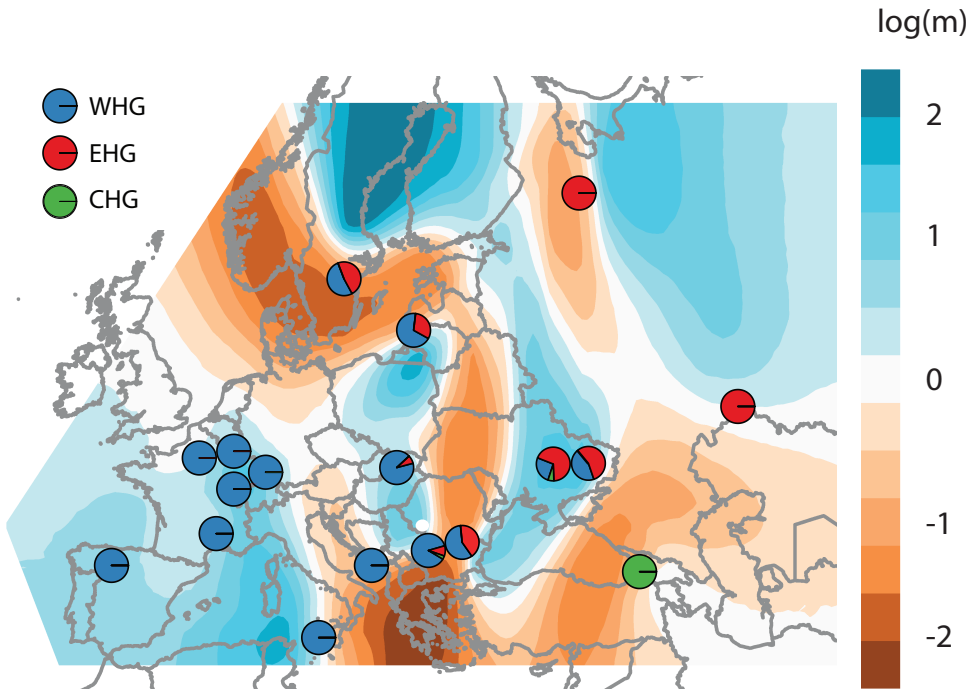


Afanasievo.SG
AfontovaGora3
Alberstedt_LN
ALPc_MN
Anatolia_BronzeAge
Anatolia_Neolithic
Anatolia_Neolithic_Boncuklu.SG
Anatolia_Neolithic_Kumtepe.SG
Anatolia_Neolithic_Tepecik_Ciftlik.SG
Andronovo.SG
Baden_LCA
Balkans_Chalcolithic_outlier
BattleAxe_Sweden.SG
Bell_Beaker_Czech.SG
Bell_Beaker_Germany
Bell_Beaker_Germany.SG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

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

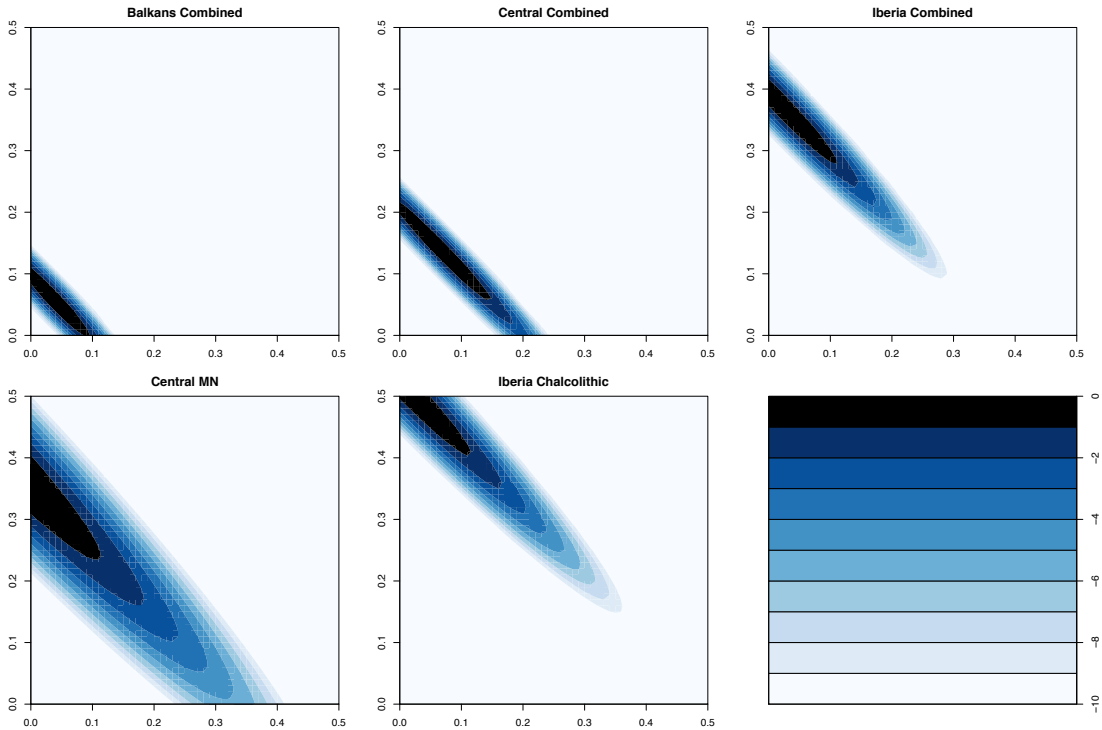


Extended Data Figure 2: Unsupervised ADMIXTURE plot from k=4 to 12, on a dataset consisting of 1099 present-day individuals and 476 ancient individuals. We show newly reported ancient individuals and some previously published individuals for comparison.



682

683 **Extended Data Figure 3**: Spatial structure in hunter-gatherers. Estimated effective migration
684 surface (EEMS).[62] 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 *log10(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.



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

Supplementary Tables

Supplementary Table 1: Details of ancient individuals analyzed in this study.

Supplementary Table 2: Key *D*-statistics to support statements about population history.

Supplementary Table 3: *qpAdm* models with 7-population outgroup set.

Supplementary Table 4: *qpAdm* models with extended 14-population outgroup set.

Supplementary Table 5: *qpAdm* models for Neolithic populations for chromosome X.

Supplementary Table 6: Additional ^{14}C dating information.

References

1 Tringham, R. E. in *The Transition to Agriculture in Prehistoric Europe* (ed D. Price) 19-56 (Cambridge University Press, 2000).

2 Bellwood, P. *First Farmers: The Origins of Agricultural Societies*. 2nd edn, (Wiley-Blackwell, 2004).

3 Golitko, M. in *Ancient Europe, 8000 B.C. to A.D. 1000: An Encyclopedia of the Barbarian World* (eds P. Bogucki & P.J. Crabtree) 259-266 (Charles Scribners & Sons, 2003).

4 Vander Linden, M. in *Investigating Archaeological Cultures: Material Culture, Variability, and Transmission* (eds B.W. Roberts & M. Vander Linden) 289-319 (Springer, 2012).

5 Bramanti, B. *et al.* Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326**, 137-140 (2009).

6 Skoglund, P. *et al.* Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466-469 (2012).

7 Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211 (2015).

8 Cassidy, L. M. *et al.* Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 368-373 (2016).

9 Hofmanova, Z. *et al.* Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6886-6891 (2016).

10 Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503 (2015).

11 Omrak, A. *et al.* Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. *Curr. Biol.* **26**, 270-275 (2016).

12 Müller, J., Rassmann, K. & Videiko, M. *Trypillia Mega-Sites and European Prehistory: 4100–3400 BCE.*, (Routledge, 2016).

13 Anthony, D. W. *The horse the wheel and language*. (Princeton University Press, 2007).

14 Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* (2015).

15 Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167-172 (2015).

16 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449 (2014).

17 Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205 (2016).

18 Gallego Llorente, M. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* **350**, 820-822 (2015).

19 Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature communications* **6**, 8912 (2015).

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

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

70 Bronk Ramsey, C. *OxCal 4.23 Online Manual*, <https://c14.arch.ox.ac.uk/oxcalhelp/hlp_contents.html> (2013).

71 Cook, G. T. *et al.* A freshwater diet-derived ^{14}C reservoir effect at the Stone Age sites in the Iron Gates gorge. *Radiocarbon* **43**, 453-460 (2001).