The genomic history of southeastern Europe

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 Ganetsovski (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.

(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 Europe in the mid-7th millennium BCE–associated with migrants from Anatolia who settled in the Southeast before spreading throughout Europe. To understand the dynamics of this process, we analyzed genome-wide ancient DNA data from 225 individuals who lived in southeastern Europe and surrounding regions between 12,000 and 500 BCE. We document a West-East cline of ancestry in indigenous hunter-gatherers and–in far-eastern Europe–early stages in the formation of Bronze Age Steppe ancestry. We show that the first farmers of northern and western Europe passed through southeastern Europe with limited hunter-gatherer admixture, but that some groups that remained mixed extensively, without 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, with intermittent genetic contact with the Steppe up to 2000 years before the migrations 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 archaeological evidence documents cultural contacts with peoples of the Eurasian steppe.[13] However, the population movements that accompanied these events have been unknown due to the lack of ancient DNA.

149

Results

We generated genome-wide data from 225 ancient humans (216 reported for the first time), from the Balkan Peninsula, the Carpathian Basin, the North Pontic Steppe and neighboring regions, dated to 12,000-500 BCE (Figure 1, Supplementary Information Table 1, Supplementary Information Note 1). We extracted DNA from skeletal remains in dedicated clean rooms, built DNA libraries and enriched for DNA fragments overlapping 1.24 million single nucleotide polymorphisms (SNPs), then sequenced the product and restricted to libraries with evidence of authentic ancient DNA.[7,10,14] We filtered out individuals with fewer than 15,000 SNPs covered by at least one sequence, or that had unexpected ancestry for their archaeological context and were not directly dated. We report, but do not analyze, nine individuals that were first-degree relatives of others in the dataset, resulting in an analysis dataset of 216 individuals. We analyzed these data together with 274 previously reported ancient individuals,[9-11,15-27] 777 present-day individuals genotyped on the Illumina "Human Origins" array,[23] and 300 high coverage genomes from the Simons Genome Diversity Project (SGDP).[28] We used principal component analysis (PCA; Figure 1B, Extended Data Figure 1), supervised and unsupervised ADMIXTURE (Figure 1D, Extended Data Figures 2&3),[29] *D*-statistics, *qpAdm* and *qpGraph,*[30] along with archaeological and chronological information (including 137 newly reported AMS ^{14}C dates) to cluster the individuals into populations and investigate the relationships among them.

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

Hunter-Gatherer substructure and transitions

185 Of the 216 new individuals we report, 106 from Paleolithic, Mesolithic and eastern European

186 Neolithic contexts have almost entirely hunter-gatherer-related ancestry (in eastern Europe,

187 unlike western Europe, "Neolithic" refers to the presence of pottery,[31-33] not necessarily to

188 farming). These individuals form a cline from WHG to EHG that is correlated with geography

189 (Figure 1B), although it is neither geographically nor temporally uniform (Figure 2, Extended

190 Data Figure 4), and contains substructure in phenotypically important variants

191 (Supplementary Information Note 2).

192

193 From present-day Ukraine, our study reports new genome-wide data from seven Mesolithic

194 (~9500-6000 BCE) and 30 Neolithic (~6000-3500 BCE) individuals. On the cline from WHG-

195 to EHG-related ancestry, the Mesolithic individuals fall towards the East, intermediate

196 between EHG and Mesolithic hunter-gatherers from Scandinavia (Figure 1B).[7] The Neolithic

197 population has a significant difference in ancestry compared to the Mesolithic (Figures 1B,

198 Figure 2), with a shift towards WHG shown by the statistic D(Mbuti, WHG,

199 Ukraine_Mesolithic, Ukraine_Neolithic); Z=8.5 (Supplementary Information Table 2).

200 Unexpectedly, one Neolithic individual from Dereivka (I3719), which we directly date to

201 4949-4799 BCE, has entirely NW Anatolian Neolithic-related ancestry.

202

203 The pastoralist Bronze Age Yamnaya complex originated on the Eurasian steppe and is a

204 plausible source for the dispersal of steppe-related ancestry into central and western Europe

205 around 2500 BCE.[13] All previously reported Yamnaya individuals were from Samara[7] and

206 Kalmykia[15] in southwest Russia, and had entirely steppe-related ancestry. Here, we report

207 three Yamnaya individuals from further West – Ukraine and Bulgaria – and show that while

208 they all have high levels of steppe-related ancestry, one from Ozera in Ukraine and one from

209 Bulgaria (I1917 and Bul4, both dated to ~3000 BCE) have NW Anatolian Neolithic-related

210 admixture, the first evidence of such ancestry in Yamnaya–associated individuals (Figure

211 1B&D, Supplementary Data Table 2). Preceding the Yamnaya culture, four Copper Age

212 individuals (I4110, I5882, I5884 and I6561; Ukraine_Eneolithic) from Dereivka and

213 Alexandria dated to ~3600-3400 BCE have ancestry that is a mixture of hunter-gatherer-,

214 steppe- and NW Anatolian Neolithic-related (Figure 1D, Supplementary Data Table 2).

215

216 At Zvejnieki in Latvia (17 newly reported individuals, and additional data for 5 first reported

217 in Ref. 34) we observe a transition in hunter-gatherer-related ancestry that is opposite to that

218 seen in Ukraine. We find (Supplementary Data Table 3) that Mesolithic and Early Neolithic

219 individuals (Latvia_HG) associated with the Kunda and Narva cultures have ancestry

220 intermediate between WHG (~70%) and EHG (~30%), consistent with previous reports.[34-36]

221 We also detect a shift in ancestry between the Early Neolithic and individuals associated with

222 the Middle Neolithic Comb Ware Complex (Latvia_MN), who have more EHG-related

223 ancestry (we estimate 65% EHG, but two of four individuals appear to be 100% EHG in

224 PCA). The most recent individual, associated with the Final Neolithic Corded Ware Complex

225 (I4629, Latvia_LN), attests to another ancestry shift, clustering closely with Yamnaya from

226 Samara,[7] Kalmykia[15] and Ukraine (Figure 2).

227

228 We report new Upper Palaeolithic and Mesolithic data from southern and western Europe.[17]

229 Sicilian (I2158) and Croatian (I1875) individuals dating to ~12,000 and 6100 BCE cluster with

230 previously reported western hunter-gatherers (Figure 1B&D), including individuals from

231 Loschbour[23] (Luxembourg, 6100 BCE), Bichon[19] (Switzerland, 11,700 BCE), and Villabruna[17]

232 (Italy 12,000 BCE). These results demonstrate that WHG populations[23] were widely

233 distributed from the Atlantic seaboard of Europe in the West, to Sicily in the South, to the

234 Balkan Peninsula in the Southeast, for at least six thousand years.

235

236 A particularly important hunter-gatherer population that we report is from the Iron Gates

237 region that straddles the border of present-day Romania and Serbia. This population

238 (Iron_Gates_HG) is represented in our study by 40 individuals from five sites. Modeling Iron

239 Gates hunter-gatherers as a mixture of WHG and EHG (Supplementary Table 3) shows that

240 they are intermediate between WHG (~85%) and EHG (~15%). However, this *qpAdm* model

241 does not fit well (p=0.0003, Supplementary Table 3) and the Iron Gates hunter-gatherers

242 show an affinity towards Anatolian Neolithic, relative to WHG (Supplementary Table 2). In

243 addition, Iron Gates hunter-gatherers carry mitochondrial haplogroup K1 (7/40) as well as

244 other subclades of haplogroups U (32/40) and H (1/40) in contrast to WHG, EHG and

245 Scandinavian hunter-gatherers who almost all carry haplogroups U5 or U2. One interpretation

246 is that the Iron Gates hunter-gatherers have ancestry that is not present in either WHG or

247 EHG. Possible scenarios include genetic contact between the ancestors of the Iron Gates

248 population and a NW Anatolian-Neolithic-related population, or that the Iron Gates

249 population is related to the source population from which the WHG split during a re-

250 expansion into Europe from the Southeast after the Last Glacial Maximum.[17,37]

251

252 A notable finding from the Iron Gates concerns the four individuals from the site of Lepenski

253 Vir, two of whom (I4665 & I5405, 6200-5600 BCE), have entirely NW Anatolian Neolithic-

254 related ancestry. Strontium and Nitrogen isotope data[38] indicate that both these individuals

255 were migrants from outside the Iron Gates, and ate a primarily terrestrial diet (Supplementary

256 Information section 1). A third individual (I4666, 6070 BCE) has a mixture of NW Anatolian

257 Neolithic-related and hunter-gatherer-related ancestry and ate a primarily aquatic diet, while a

258 fourth, probably earlier, individual (I5407) had entirely hunter-gatherer-related ancestry

259 (Figure 1D, Supplementary Information section 1). We also identify one individual from
260 Padina (I5232), dated to 5950 BCE that had a mixture of NW Anatolian Neolithic-related and
261 hunter-gatherer-related ancestry. These results demonstrate that the Iron Gates was a region of
262 interaction between groups distinct in both ancestry and subsistence strategy.
263
264 **Population transformations in the first farmers**
265 Neolithic populations from present-day Bulgaria, Croatia, Macedonia, Serbia and Romania
266 cluster closely with the NW Anatolian Neolithic (Figure 1), consistent with archaeological
267 evidence.[39] Modeling Balkan Neolithic populations as a mixture of NW Anatolian Neolithic
268 and WHG, we estimate that 98% (95% confidence interval [CI]; 97-100%) of their ancestry is
269 NW Anatolian Neolithic-related. A striking exception is evident in 8 out of 9 individuals from
270 Malak Preslavets in present-day Bulgaria.[40] These individuals lived in the mid-6th millennium
271 BCE and have significantly more hunter-gatherer-related ancestry than other Balkan Neolithic
272 populations (Figure 1B,D, Extended Data Figures 1-3, Supplementary Tables 2-4); a model of
273 82% (CI: 77-86%) NW Anatolian Neolithic-related, 15% (CI: 12-17%) WHG-related, and 4%
274 (CI: 0-9%) EHG-related ancestry fits the data. This hunter-gatherer-related ancestry with a
275 ~4:1 WHG:EHG ratio plausibly represents a contribution from local Balkan hunter-gatherers
276 genetically similar to those of the Iron Gates. Late Mesolithic hunter-gatherers in the Balkans
277 were likely concentrated along the coast and major rivers such as the Danube,[41] which
278 directly connects the Iron Gates with Malak Preslavets. Thus, early farmer groups with the
279 most hunter-gatherer-related ancestry may have been those that lived close to the highest
280 densities of hunter-gatherers.
281
282 In the Balkans, Copper Age populations (Balkans_Chalcolithic) harbor significantly more
283 hunter-gatherer-related ancestry than Neolithic populations as shown, for example, by the
284 statistic D(Mbuti, WHG, Balkans_Neolithic, Balkans_Chalcolithic); Z=4.3 (Supplementary
285 Data Table 2). This is roughly contemporary with the "resurgence" of hunter-gatherer
286 ancestry previously reported in central Europe and Iberia[7,10,42] and is consistent with changes
287 in funeral rites, specifically the reappearance around 4500 BCE of the Mesolithic tradition of
288 extended supine burial – in contrast to the Early Neolithic tradition of flexed burial.[43] Four
289 individuals associated with the Copper Age Trypillian population have ~80% NW Anatolian-
290 related ancestry (Supplementary Table 3), confirming that the ancestry of the first farmers of
291 present-day Ukraine was largely derived from the same source as the farmers of Anatolia and
292 western Europe. Their ~20% hunter-gatherer ancestry is intermediate between WHG and
293 EHG, consistent with deriving from the Neolithic hunter-gatherers of the region.
294

We also report the first genetic data associated with the Late Neolithic Globular Amphora Complex. Individuals from two Globular Amphora sites in Poland and Ukraine form a tight cluster, showing high similarity over a large distance (Figure 1B,D). Both groups of Globular Amphora Complex samples had more hunter-gatherer-related ancestry than Middle Neolithic groups from Central Europe[7] (we estimate 25% [CI: 22-27%] WHG ancestry, similar to Chalcolithic Iberia, Supplementary Data Table 3). In east-central Europe, the Globular Amphora Complex preceded or abutted the Corded Ware Complex that marks the appearance of steppe-related ancestry,[7,15] while in southeastern Europe, the Globular Amphora Complex bordered populations with steppe-influenced material cultures for hundreds of years[44] and yet the individuals in our study have no evidence of steppe-related ancestry, supporting the hypothesis that this material cultural frontier was also a barrier to gene flow.

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

A new source of ancestry in Neolithic Europe

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

332 In contrast, five southern Greek Neolithic individuals (Peloponnese_Neolithic) – three (plus

333 one from Ref. 26) from Diros Cave and one from Franchthi Cave – are not consistent with

334 descending from the same source population as other European farmers. *D*-statistics

335 (Supplementary Information Table 2) show that in fact, these "Peloponnese Neolithic"

336 individuals dated to ~4000 BCE are shifted away from WHG and towards CHG, relative to

337 Anatolian and Balkan Neolithic individuals. We see the same pattern in a single Neolithic

338 individual from Krepost in present-day Bulgaria (I0679_d, 5718-5626 BCE). An even more

339 dramatic shift towards CHG has been observed in individuals associated with the Bronze Age

340 Minoan and Mycenaean cultures,[26] suggesting gene flow into the region from populations

341 with CHG-rich ancestry throughout the Neolithic, Chalcolithic and Bronze Age. Possible

342 sources are related to the Neolithic population from the central Anatolian site of Tepecik

343 Çiftlik,[21] or the Aegean site of Kumtepe,[11] who are also shifted towards CHG relative to NW

344 Anatolian Neolithic samples, as are later Copper and Bronze Age Anatolians.[10,26]

345

346 **Sex-biased admixture between hunter-gatherers and farmers**

347 We provide the first evidence for sex-biased admixture between hunter-gatherers and farmers

348 in Europe, showing that the Middle Neolithic "resurgence" of hunter-gatherer-related

349 ancestry[7,42] in central Europe and Iberia was driven more by males than by females (Figure

350 3B&C, Supplementary Data Table 5, Extended Data Figure 5). To document this we used

351 *qpAdm* to compute ancestry proportions on the autosomes and the X chromosome; since

352 males always inherit a maternal X chromosome, differences imply sex-biased mixture. In the

353 Balkan Neolithic there is no evidence of sex bias (Z=0.27 where a positive Z-score implies

354 male hunter-gatherer bias), nor in the LBK and Iberian_Early Neolithic (Z=-0.22 and 1.09). In

355 the Copper Age there is clear bias: weak in the Balkans (Z=1.66), but stronger in Iberia

356 (Z=3.08) and Central Europe (Z=2.74). Consistent with this, hunter-gatherer mitochondrial

357 haplogroups (haplogroup U)[46] are rare and within the intervals of genome-wide ancestry

358 proportions, but hunter-gatherer-associated Y chromosomes (haplogroups I2, R1 and C1)[17]

359 are more common: 7/9 in the Iberian Neolithic/Copper Age and 9/10 in Middle-Late Neolithic

360 Central Europe (Central_MN and Globular_Amphora) (Figure 3C).

361

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

363 One version of the Steppe Hypothesis of Indo-European language origins suggests that Proto-

364 Indo-European languages developed north of the Black and Caspian seas, and that the earliest

365 known diverging branch – Anatolian – was spread into Asia Minor by movements of steppe

366 peoples through the Balkan peninsula during the Copper Age around 4000 BCE.[47] If this were

367 correct, then one way to detect evidence of it would be the appearance of large amounts of

368 steppe-related ancestry first in the Balkan Peninsula, and then in Anatolia. However, our data

369   show no evidence for this scenario. While we find sporadic steppe-related ancestry in Balkan

370   Copper and Bronze Age individuals, this ancestry is rare until the late Bronze Age. Moreover,

371   while Bronze Age Anatolian individuals have CHG-related ancestry,[26] they have neither the

372   EHG-related ancestry characteristic of all steppe populations sampled to date,[19] nor the

373   WHG-related ancestry that is ubiquitous in Neolithic southeastern Europe (Extended Data

374   Figure 2&3, Supplementary Data Table 2). An alternative hypothesis is that the ultimate

375   homeland of Proto-Indo-European languages was in the Caucasus or in Iran. In this scenario,

376   westward movement contributed to the dispersal of Anatolian languages, and northward

377   movement and mixture with EHG was responsible for the formation of a "Late Proto-Indo

378   European"-speaking population associated with the Yamnaya Complex.[13] While this scenario

379   gains plausibility from our results, it remains possible that Indo-European languages were

380   spread through southeastern Europe into Anatolia without large-scale population movement

381   or admixture.

Discussion

383   Our study shows that southeastern Europe consistently served as a genetic contact zone.

384   Before the arrival of farming, the region saw interaction between diverged groups of hunter-

385   gatherers, and this interaction continued after farming arrived. While this study has clarified

386   the genomic history of southeastern Europe from the Mesolithic to the Bronze Age, the

387   processes that connected these populations to the ones living today remain largely unknown.

388   An important direction for future research will be to sample populations from the Bronze

389   Age, Iron Age, Roman, and Medieval periods and to compare them to present-day

390   populations to understand how these transitions occurred.

Methods

Ancient DNA Analysis

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

Two samples were processed at the Australian Centre for Ancient DNA, Adelaide, Australia, according to previously published methods[7] and sent to Boston for subsequent screening, 1240k capture and sequencing.

Seven samples were processed[27] at the Institute of Archaeology RCH HAS, Budapest, Hungary, and amplified libraries were sent to Boston for screening, 1240k capture and sequencing.

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

The remaining 199 samples were processed at Harvard Medical School, Boston, USA. From about 75mg of sample powder from each sample (extracted in Boston or University College Dublin, Dublin, Ireland), we extracted DNA following established methods[48] replacing the column assembly with the column extenders from a Roche kit.[51] We prepared double barcoded libraries with truncated adapters from between one ninth and one third of the DNA extract. Most libraries included in the nuclear genome analysis (90%) were subjected to partial ("half") Uracil-DNA-glycosylase (UDG) treatment before blunt end repair. This treatment reduces by an order of magnitude the characteristic cytosine-to-thymine errors of ancient DNA data[52], but works inefficiently at the 5' ends,[50] thereby leaving a signal of characteristic damage at the terminal ends of ancient sequences. Some libraries were not UDG-treated ("minus"). For some samples we increased coverage by preparing additional libraries from the existing DNA extract using the partial UDG library preparation, but replacing the MinElute column cleanups in between enzymatic reactions with magnetic bead cleanups, and the final PCR cleanup with SPRI bead cleanup.[53,54]

427   We screened all libraries from Adelaide, Boston and Budapest by enriching for the

428   mitochondrial genome plus about 3,000 (50 in an earlier, unpublished, version) nuclear SNPs

429   using a bead-capture[55] but with the probes replaced by amplified oligonucleotides synthesized

430   by CustomArray Inc. After the capture, we completed the adapter sites using PCR, attaching

431   dual index combinations[56] to each enriched library. We sequenced the products of between

432   100 and 200 libraries together with the non-enriched libraries (shotgun) on an Illumina

433   NextSeq500 using v2 150 cycle kits for 2x76 cycles and 2x7 cycles.

434

435   In Boston, we performed two rounds of in-solution enrichment ("1240k capture") for a

436   targeted set of 1,237,207 SNPs using previously reported protocols.[7,14,23] For a total of 34

437   individuals, we increased coverage by building one to eight additional libraries for the same

438   sample. When we built multiple libraries from the same extract, we often pooled them in

439   equimolar ratios before the capture. We performed all sequencing on an Illumina NextSeq500

440   using v2 150 cycle kits for 2x76 cycles and 2x7 cycles. We attempted to sequence each

441   enriched library up to the point where we estimated that it was economically inefficient to

442   sequence further. Specifically, we iteratively sequenced more and more from each individual

443   and only stopped when we estimated that the expected increase in the number of targeted

444   SNPs hit at least once would be less than about one for every 100 new read pairs generated.

445   After sequencing, we trimmed two bases from the end of each read and aligned to the human

446   genome (b37/hg19) using *bwa*.[57] We then removed individuals with evidence of

447   contamination based on mitochondrial DNA polymorphism[58] or difference in PCA space

448   between damaged and undamaged reads[59], a high rate of heterozygosity on chromosome X

449   despite being male[59,60], or an atypical ratio of X-to-Y sequences. We also removed individuals

450   that had low coverage (fewer than 15,000 SNPs hit on the autosomes). We report, but do not

451   analyze, data from nine individuals that were first-degree relatives of others in the dataset

452   (determined by comparing rates of allele sharing between pairs of individuals).

453

454   After removing a small number of sites that failed to capture, we were left with a total of

455   1,233,013 sites of which 32,670 were on chromosome X and 49,704 were on chromosome Y,

456   with a median coverage at targeted SNPs on the 216 newly reported individuals of 0.90

457   (range 0.007-9.2; Supplementary Table 1). We generated "pseudo-haploid" calls by selecting

458   a single read randomly for each individual at each SNP. Thus, there is only a single allele

459   from each individual at each site, but adjacent alleles might come from either of the two

460   haplotypes of the individual. We merged the newly reported data with previously reported

461   data from 274 other ancient individuals[9-11,15-27], making pseudo-haploid calls in the same way

462   at the 1240k sites for individuals that were shotgun sequenced rather than captured.

463

464   Using the captured mitochondrial sequence from the screening process, we called

465   mitochondrial haplotypes. Using the captured SNPs on the Y chromosome, we called Y

466   chromosome haplogroups for males by restricting to sequences with mapping quality ≥30 and

467   bases with base quality ≥30. We determined the most derived mutation for each individual,

468   using the nomenclature of the International Society of Genetic Genealogy

469   (http://www.isogg.org) version 11.110 (21 April 2016).

470

471   **Population genetic analysis**

472   To analyze these ancient individuals in the context of present day genetic diversity, we

473   merged them with the following two datasets:

474

475   1.   300 high coverage genomes from a diverse worldwide set of 142 populations

476   sequenced as part of the Simons Genome Diversity Project[28] (SGDP merge).

477

478   2.   777 West Eurasian individuals genotyped on the Human Origins array[23], with

479   597,573 sites in the merged dataset (HO merge).

480

481   We computed principal components of the present-day individuals in the HO merge and

482   projected the ancient individuals onto the first two components using the "*lsqproject: YES*"

483   option in *smartpca (v15100)*[61] (https://www.hsph.harvard.edu/alkes-price/software/).

484

485   We ran *ADMIXTURE (v1.3.0)* in both supervised and unsupervised mode. In supervised mode

486   we used only the ancient individuals, on the full set of SNPs, and the following population

487   labels fixed:

488   • *Anatolia_Neolithic*
489   • *WHG*
490   • *EHG*
491   • *Yamnaya*
492

493   For unsupervised mode we used the HO merge, including 777 present-day individuals. We

494   flagged individuals that were genetic outliers based on PCA and ADMIXTURE, relative to

495   other individuals from the same time period and archaeological culture.

496

497   We computed *D*-statistics using *qpDstat (v710)*. *D*-statistics of the form D(A,B,X,Y) test the

498   null hypothesis of the unrooted tree topology ((A,B),(X,Y)). A positive value indicates that

499   either A and X, or B and Y, share more drift than expected under the null hypothesis. We

500   quote *D*-statistics as the *Z*-score computed using default block jackknife parameters.

501

502   We fitted admixture proportions with *qpAdm (v610)* using the SGDP merge. Given a set of

503   outgroup ("right") populations, *qpAdm* models one of a set of source ("left") populations (the

504   "test" population) as a mixture of the other sources by fitting admixture proportions to match

505   the observed matrix of f_4-statistics as closely as possible. We report a p-value for the null

506   hypothesis that the test population does not have ancestry from another source that is

507   differentially related to the right populations. We computed standard errors for the mixture

508   proportions using a block jackknife. Importantly, *qpAdm* does not require that the source

509   populations are actually the admixing populations, only that they are a clade with the correct

510   admixing populations, relative to the other sources. Infeasible coefficient estimates (i.e.

511   outside [0,1]) are usually a sign of poor model fit, but in the case where the source with a

512   negative coefficient is itself admixed, could be interpreted as implying that the true source is a

513   population with different admixture proportions. We used the following set of seven

514   populations as outgroups or "right populations":

515   • *Mbuti.DG*
516   • *Ust_Ishim_HG_published.DG*
517   • *Mota.SG*
518   • *MA1_HG.SG*
519   • *Villabruna*
520   • *Papuan.DG*
521   • *Onge.DG*
522   • *Han.DG*

523

524   For some analyses where we required extra resolution (Supplementary Data Table 4) we used

525   an extended set of 14 right (outgroup) populations, including additional Upper Paleolithic

526   European individuals[17]:

527   • *ElMiron*
528   • *Mota.SG*
529   • *Mbuti.DG*
530   • *Ust_Ishim_HG_published.DG*
531   • *MA1_HG.SG*
532   • *AfontovaGora3*
533   • *GoyetQ116-1_published*
534   • *Villabruna*
535   • *Kostenki14*
536   • *Vestonice16*
537   • *Karitiana.DG*
538   • *Papuan.DG*
539   • *Onge.DG*
540   • *Han.DG*
541

542   We also fitted admixture graphs with *qpGraph (v6021)*[30] (https://github.com/DReichLab/

543   AdmixTools, Supplementary Information Note 3). Like *qpAdm*, *qpGraph* also tries to match a

544   matrix of *f*-statistics, but rather than fitting one population as a mixture of other, specified,

545 populations, it fits the relationship between all tested populations simultaneously, potentially

546 incorporating multiple admixture events. However, *qpGraph* requires the graph relating

547 populations to be specified in advance. We tested goodness-of-fit by computing the expected

548 *D*-statistics under the fitted model, finding the largest *D*-statistic outlier between the fitted and

549 observed model, and computing a *Z*-score using a block jackknife.

550

551 For 114 individuals with hunter-gatherer-related ancestry we estimated an effective migration

552 surface using the software *EEMS* (https://github.com/dipetkov/eems)[62]. We computed

553 pairwise differences between individuals using the *bed2diffs2* program provided with *EEMS*.

554 We set the number of demes to 400 and defined the outer boundary of the region by the

555 polygon (in latitude-longitude co-ordinates) [(66,60), (60,10), (45,-15), (35,-10), (35,60)]. We

556 ran the MCMC ten times with different random seeds, each time with one million burn-in and

557 four million regular iterations, thinned to one in ten thousand.

558

559 To analyze potential sex bias in admixture, we used *qpAdm* to estimate admixture proportions

560 on the autosomes (default option) and on the X chromosome (option "*chrom: 23*"). We

561 computed Z-scores for the difference between the autosomes and the X chromosome as $Z =$

562 $\frac{p_A - p_X}{\sqrt{\sigma_A^2 + \sigma_X^2}}$ where p_A and p_X are the hunter-gatherer admixture proportions on the autosomes and

563 the X chromosome, and σ_A and σ_X are the corresponding jackknife standard deviations. Thus,

564 a positive Z-score means that there is more hunter-gatherer admixture on the autosomes than

565 on the X chromosome, indicating that the hunter-gatherer admixture was male-biased.

566 Because X chromosome standard errors are high and *qpAdm* results can be sensitive to which

567 population is first in the list of outgroup populations, we checked that the patterns we observe

568 were robust to cyclic permutation of the outgroups. To compare frequencies of hunter-

569 gatherer uniparental markers, we counted the individuals with mitochondrial haplogroup U

570 and Y chromosome haplogroups C1, I2 and R1, which are all common in Mesolithic hunter-

571 gatherers but rare or absent in Anatolian Neolithic individuals. The Iron Gates hunter-

572 gatherers also carry H and K1 mitochondrial haplogroups so the proportion of haplogroup U

573 represents the minimum maternal hunter-gatherer contribution. We computed binomial

574 confidence intervals for the proportion of haplogroups associated with each ancestry type

575 using the Agresti-Coull method[63,64] implemented in the *binom* package in *R*.

576

577 Given autosomal and X chromosome admixture proportions, we estimated the proportion of

578 male and female hunter-gatherer ancestors by assuming a single-pulse model of admixture. If

579 the proportions of male and female ancestors that are hunter-gatherer-related are given by *m*

580 and *f*, respectively, then the proportions of hunter-gatherer-related ancestry on the autosomes

581 and the X chromosome are given by $\frac{m+f}{2}$ and $\frac{m+2f}{3}$. We approximated the sampling error in

582 the observed admixture proportions by the estimated jackknife error and computed the

583 likelihood surface for *(m,f)* over a grid ranging from (0,0) to (1,1).

584

585 **Direct AMS ^{14}C Bone Dates**

586 We report 137 new direct AMS ^{14}C bone dates for 136 individuals from multiple AMS

587 radiocarbon laboratories. In general, bone samples were manually cleaned and demineralized

588 in weak HCl and, in most cases (PSU, UCIAMS, OxA), soaked in an alkali bath (NaOH) at

589 room temperature to remove contaminating soil humates. Samples were then rinsed to

590 neutrality in Nanopure H_2O and gelatinized in HCL.[65] The resulting gelatin was lyophilized

591 and weighed to determine percent yield as a measure of collagen preservation (% crude

592 gelatin yield). Collagen was then directly AMS ^{14}C dated (Beta, AA) or further purified using

593 ultrafiltration (PSU, UCIAMS, OxA, Poz, MAMS).[66] It is standard in some laboratories

594 (PSU/UCIAMS, OxA) to use stable carbon and nitrogen isotopes as an additional quality

595 control measure. For these samples, the %C, %N and C:N ratios were evaluated before AMS

596 ^{14}C dating.[67] C:N ratios for well-preserved samples fall between 2.9 and 3.6, indicating good

597 collagen preservation.[68] For 119 of the new dates, we also report δ^{13}C and δ^{15}N values

598 (Supplementary Table 6).

599

600 All ^{14}C ages were δ^{13}C-corrected for mass dependent fractionation with measured ^{13}C/^{12}C

601 values[69] and calibrated with OxCal version 4.2.3[70] using the IntCal13 northern hemisphere

602 calibration curve.[70] For hunter-gatherers from the Iron Gates, the direct ^{14}C dates tend to be

603 overestimates because of the freshwater reservoir effect (FRE), which arises because of a diet

604 including fish that consumed ancient carbon, and for these individuals we performed a

605 correction (Supplementary Information Note 1),[71] assuming that 100% FRE = 545±70 yr, and

606 δ^{15}N values of 8.3% and 17.0% for 100% terrestrial and aquatic diets, respectively.

607

608 ## Data Availability

609 The aligned sequences are available through the European Nucleotide Archive under

610 accession number PRJEB22652. The pseudo-haploid genotype dataset used in analysis and

611 consensus mitochondrial genomes are available at https://reich.hms.harvard.edu/datasets.

612 ## Code Availability

613 Software used to analyze the data is available from the following sources:

614 *smartpca, qpAdm, qpDstat, qpGraph:* https://github.com/DReichLab/AdmixTools/

615 *ADMIXTURE*: https://www.genetics.ucla.edu/software/admixture/

616   *EEMS:* https://github.com/dipetkov/eems/

617   *bwa*: http://bio-bwa.sourceforge.net

618   OxCal: https://c14.arch.ox.ac.uk/oxcal.html

619

635

Author Contributions

637   SAR, AS-N, SVai, SA, KWA, RA, DA, AA, NA, KB, MBG, HB, MB, ABo, YB, ABu, JB,

638   SC, NC, RC, MC, CC, DD, NE, MFr, BGal, GG, BGe, THa, VH, KH, THi, SI, IJ, IKa, DKa,

639   AK, DLa, MLa, CL, MLe, KL, DLV, DLo, IL, MMa, FM, KM, HM, MMe, PM, VM, VP,

640   TDP, ASi, LS, MŠ, VS, PS, ASt, TS, MT-N, CT, IV, FVa, SVas, FVe, SV, EV, BV, CV, JZ,

641   SZ, PWS, GC, RK, DC, GZ, BGay, MLi, AGN, IP, AP, DB, CB, JK, RP & DR assembled

642   and interpreted archaeological material. CP, AS-N, NR, NB, FC, OC, DF, MFe, BGam, GGF,

643   WH, EH, EJ, DKe, BK-K, IKu, MMi, AM, KN, MN, JO, SP, KSi, KSt & SVai performed

644   laboratory work. IM, CP, AS-N, SM, IO, NP & DR analyzed data. DJK, ST, DB, CB

645   interpreted ^{14}C dates. JK, RP & DR supervised analysis or laboratory work. IM & DR wrote

646   the paper, with input from all co-authors.

647

654

655 **Figure 1: Geographic and genetic structure of 216 newly reported individuals**. **A**:

656 Locations of newly reported individuals. **B**: Ancient individuals projected onto principal

657 components defined by 777 present-day West Eurasians (shown in Extended Data Figure 1).

658 Includes selected published individuals (faded circles, labeled) and newly reported individuals

659 (other symbols, outliers enclosed in black circles). Colored polygons cover individuals that

660 had cluster memberships fixed at 100% for supervised admixture analysis. **C**: Date (direct or

661 contextual) for each sample and approximate chronology of southeastern Europe. **D**:

662 Supervised ADMIXTURE analysis, modeling each ancient individual (one per row), as a

663 mixture of population clusters constrained to contain Anatolian Neolithic (grey), Yamnaya

664 from Samara (yellow), EHG (pink) and WHG (green) populations. Dates in parentheses

665 indicate approximate range of individuals in each population. See Extended Data Figure 2 for

666 individual sample IDs. Map data in **A** from the *R* package *maps*.

667

668 **Figure 2: Structure and change in hunter-gatherer-related populations**. Inferred ancestry

669 proportions for populations modeled as a mixture of WHG, EHG and CHG (Supplementary

670 Table S3.1.3). Dashed lines show populations from the same geographic region. Percentages

671 indicate proportion of WHG+EHG ancestry. Standard errors range from 1.5-8.3%

672 (Supplementary Table S3.1.3).

673

674 **Figure 3**: **Structure and change in NW Anatolian Neolithic-related populations**. **A**:

675 Populations modeled as a mixture of NW Anatolia Neolithic, WHG, and EHG. Dashed lines

676 show temporal relationships between populations from the same geographic region.

677 Percentages indicate proportion of WHG+EHG ancestry. Standard errors range from 0.7-

678 6.0% (Supplementary Table S3.2.2). **B**: Z-scores for the difference in hunter-gatherer-related

679 ancestry on the autosomes compared to the X chromosome when populations are modeled as

680 a mixture of NW Anatolia Neolithic and WHG (N=126 individuals, group sizes in

681 parentheses). Positive values indicate more hunter-gatherer-related ancestry on the autosomes

682 and thus male-biased hunter-gatherer ancestry. "Combined" populations merge all individuals

683 from different times from a geographic area. **C**: Hunter-gatherer-related ancestry proportions

684 on the autosomes, X chromosome, mitochondrial DNA (i.e. mt haplogroup U), and the Y

685 chromosome (i.e. Y chromosome haplogroups I2, R1 and C1). Points show *qpAdm*

686 (autosomes and X chromosome) or maximum likelihood (MT and Y chromosome) estimates

687 and bars show approximate 95% confidence intervals (N=109 individuals, group sizes in

688 parentheses).

689

Extended Data Figure Captions

Extended Data Figure 1: PCA of 486 ancient individuals, projected onto principal components defined by 777 present-day West Eurasian individuals (grey points). This differs from Figure 1B in that the plot is not cropped and the present-day individuals are shown.

Extended Data Figure 2: Supervised ADMIXTURE analysis modeling each ancient individual (one per row), as a mixture of populations represented by clusters that are constrained to contain Anatolian Neolithic (grey), Yamnaya from Samara (yellow), EHG (pink) and WHG (green) populations. Dates in parentheses indicate approximate range of individuals in each population. This differs from Figure 1D in that it contains some previously published samples, and includes sample IDs.

Extended Data Figure 3: 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.

Extended Data Figure 4: Spatial structure in hunter-gatherers. Estimated effective migration surface (EEMS).[62] This fits a model of genetic relatedness where individuals move (in a random direction) from generation to generation on an underlying grid so that genetic relatedness is determined by distance. The migration parameter m defines the local rate of migration, varies on the grid and is inferred. This plot shows $log_{10}(m)$, scaled relative to the average migration rate (which is arbitrary). Thus $log_{10}(m)=2$, for example, implies that the rate of migration at this point on the grid is 100 times higher than average. To restrict as much as possible to hunter-gatherer structure, the migration surface is inferred using data from 116 individuals that date to earlier than ~5000 BCE and have no NW Anatolian-related ancestry. Though the migration surface is sensitive to sampling, and fine-scale features may not be interpretable, the migration "barrier" (region of low migration) running north-south and separating populations with primarily WHG from primarily EHG ancestry seems to be robust, and consistent with inferred admixture proportions. This analysis suggests that Mesolithic hunter-gatherer population structure was clustered and not smoothly clinal, in the sense that genetic differentiation did not vary consistently with distance. Superimposed on this background, pies show the WHG, EHG and CHG ancestry proportions inferred for populations used to construct the migration surface (another way of visualizing the data in Figure 2, Supplementary Table 3.1.3 – we use two population models if they fit with p>0.01, and three population models otherwise). Pies with only a single color are those that were fixed to be the source populations.

726

Extended Data Figure 5: 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. Numbers in parentheses give the number of individuals in each group. Log-likelihood scale ranges from 0 to -10, where 0 is the feasible point with the highest likelihood.

Main Text 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).

785 26 Lazaridis, I. *et al.* Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214-
786   218 (2017).
787 27 Lipson, M. *et al.* Parallel ancient genomic transects reveal complex population
788   history of early European farmers. *Nature* **551**, 368-372 (2017).
789 28 Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142
790   diverse populations. *Nature* **538**, 201-206 (2016).
791 29 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry
792   in unrelated individuals. *Genome Res.* **19**, 1655-1664 (2009).
793 30 Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093
794   (2012).
795 31 Gronenborn, D. & Dolukhanov, P. in *The Oxford Handbook of Neolithic Europe*
796   (eds C. Fowler, J. Harding, & D. Hofmann) 195-214 (Oxford University Press,
797   2015).
798 32 Telegin, D. Ya. Neolithic cultures of Ukraine and their chronology. *Journal of World*
799   *Prehistory* **I**, 307-331 (1987).
800 33 Telegin, D. Ya. & Potekhina, I. D. *Neolithic cemeteries and populations in the*
801   *Dnieper Basin*. (British Archaeological Reports, 1987).
802 34 Jones, E. R. *et al.* The Neolithic Transition in the Baltic Was Not Driven by
803   Admixture with Early European Farmers. *Curr. Biol.*, 2185–2193 (2017).
804 35 Mittnik, A. *et al.* The Genetic History of Northern Europe. *bioRxiv*,
805   https://doi.org/10.1101/113241 (2017).
806 36 Saag, L. *et al.* Extensive farming in Estonia started through a sex-biased migration
807   from the Steppe. *Curr. Biol.* **27**, 2185-2193 (2017).
808 37 Maier, A. *The Central European Magdalenian: Regional Diversity and Internal*
809   *Variability.*, (Springer, 2015).
810 38 Borić, D. & Price, T. D. Strontium isotopes document greater human mobility at the
811   start of the Balkan Neolithic. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 3298-3303 (2013).
812 39 Krauß, R., Marinova, E., De Brue, H. & Weninger, B. The rapid spread of early
813   farming from the Aegean into the Balkans via the Sub-Mediterranean-Aegean
814   Vegetation Zone. *Quaternary International* **XXX**, 1-18 (2017).
815 40 Bacvarov, K. in *Moments in time: Papers Presented to Pál Raczky on His 60th*
816   *Birthday* (eds A. Anders & G Kulcsár) 29-34 (L'Harmattan, 2013).
817 41 Gurova, M. & Bonsall, C. 'Pre-Neolithic' in Southeast Europe: a Bulgarian
818   perspective. *Documenta Praehistorica* **XLI**, 95-109 (2014).
819 42 Brandt, G. *et al.* Ancient DNA reveals key stages in the formation of central
820   European mitochondrial genetic diversity. *Science* **342**, 257-261 (2013).
821 43 Borić, D. in *The Oxford Handbook of Neolithic Europe* (eds C. Fowler, J. Harding,
822   & D. Hofmann) 927–957 (Oxford University Press, 2015).
823 44 Szmyt, M. in *Transition to the Bronze Age (Archaeolingua 30)* (eds V. Heyd, G.
824   Kulcsár, & V. Szeverényi) 93-111 (Archaeolingua, 2013).
825 45 Olalde, I. *et al.* A Common Genetic Origin for Early Farmers from Mediterranean
826   Cardial and Central European LBK Cultures. *Mol. Biol. Evol.* **32**, 3132-3142 (2015).
827 46 Posth, C. *et al.* Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal
828   of Non-Africans and a Late Glacial Population Turnover in Europe. *Curr. Biol.* **26**,
829   827-833 (2016).
830 47 Anthony, D. W. & Ringe, D. The Indo-European Homeland from Linguistic and
831   Archaeological Perspectives. *Annual Review of Linguistics* **1**, 199-219 (2015).
832

833 **Additional Methods References**

834
835 48 Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene
836   cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S.*
837   *A.* **110**, 15758-15763 (2013).

838   49   Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly
839   multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**, pdb
840   prot5448 (2010).

841   50   Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-
842   glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B*
843   *Biol. Sci.* **370**, 20130624 (2015).

844   51   Korlevic, P. *et al.* Reducing microbial and human contamination in DNA extractions
845   from ancient bones and teeth. *BioTechniques* **59**, 87-93 (2015).

846   52   Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo
847   methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).

848   53   DeAngelis, M. M., Wang, D. G. & Hawkins, T. L. Solid-phase reversible
849   immobilization for the isolation of PCR products. *Nucleic Acids Res.* **23**, 4742-4743
850   (1995).

851   54   Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries
852   for multiplexed target capture. *Genome Res.* **22**, 939-946 (2012).

853   55   Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of
854   mitochondrial genomes using PCR products. *PLoS One* **5**, e14004 (2010).

855   56   Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in
856   multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).

857   57   Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler
858   transform. *Bioinformatics* **26**, 589-595 (2010).

859   58   Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial
860   genomes. *Curr. Biol.* **23**, 553-559 (2013).

861   59   Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day
862   contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2229-
863   2234 (2014).

864   60   Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next
865   Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).

866   61   Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-
867   wide association studies. *Nat. Genet.* **38**, 904-909 (2006).

868   62   Petkova, D., Novembre, J. & Stephens, M. Visualizing spatial population structure
869   with estimated effective migration surfaces. *Nat. Genet.* **48**, 94-100 (2016).

870   63   Brown, L., Cai, T. & DasGupta, A. Interval Estimation for a Binomial Proportion.
871   *Statistical Science* **16**, 101-133 (2001).

872   64   Agresti, A. & Coull, B. Approximate is better than "exact" for interval estimation of
873   binomial proportions. *The American Statistician* **52**, 119-126 (1998).

874   65   Longin, R. New method of collagen extraction for radiocarbon analysis. *Nature* **230**,
875   241-242 (1971).

876   66   Brown, T. A., Nelson, D. E., Vogel, J. S. & Southon, J. R. Improved collagen
877   extraction by modified longin method. *Radiocarbon* **30**, 171-177 (1988).

878   67   Kennett, D. J. *et al.* Archaeogenomic evidence reveals prehistoric matrilineal dynasty.
879   *Nature communications* **8**, 14115 (2017).

880   68   van Klinken, G. J. Bone collagen quality indicators for paleodietary and radiocarbon
881   measurements. *JAS* **26**, 687-695 (1999).

882   69   Stuiver, M. & Polach, H. A. Discussion: Reporting of ^{14}C data. *Radiocarbon* **19**, 355-
883   363 (1977).

884   70   Bronk Ramsey, C. *OxCal 4.23 Online Manual*,
885   https://c14.arch.ox.ac.uk/oxcalhelp/hlp_contents.html (2013).

886   71   Cook, G. T. *et al.* A freshwater diet-derived ^{14}C reservoir effect at the Stone Age sites
887   in the Iron Gates gorge. *Radiocarbon* **43**, 453-460 (2001).

888