

1 The Genomic History of Southeastern Europe

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

31
32 * These authors contributed equally to the manuscript

33 ⁺ Correspondence to I.M. (iain_mathieson@hms.harvard.edu) or D.R.
34 (reich@genetics.med.harvard.edu) or R.P. (ron.pinhasi@ucd.ie)
35

36 (1) Department of Genetics, Harvard Medical School, Boston 02115 MA USA (2) Department of Archaeogenetics,
37 Max Planck Institute for the Science of Human History, 07745 Jena, Germany (3) Institute for Scientific
38 Archaeology, University of Tuebingen, Germany (4) Laboratory of Archaeogenetics, Institute of Archaeology,
39 Research Centre for the Humanities, Hungarian Academy of Sciences, H-1097 Budapest, Hungary (5) Howard
40 Hughes Medical Institute, Harvard Medical School, Boston 02115 MA USA (6) School of Archaeology and Earth
41 Institute, University College Dublin, Belfield, Dublin 4, Republic of Ireland (7) CIAS, Department of Life
42 Sciences, University of Coimbra, 3000-456 Coimbra, Portugal (8) Department of Life Sciences and
43 Biotechnology, University of Ferrara, Via L. Borsari 46. Ferrara 44100 Italy (9) Australian Centre for Ancient
44 DNA, School of Biological Sciences, The University of Adelaide, SA-5005 Adelaide, Australia (10) Institute for
45 Anthropological Research, Ljudevita Gaja 32, 10000 Zagreb, Croatia (11) Broad Institute of Harvard and MIT,
46 Cambridge MA (12) Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA (13)
47 Dipartimento di Biologia, Università di Firenze, 50122 Florence, Italy (14) National Institute of Archaeology and
48 Museum, Bulgarian Academy of Sciences, 2 Saborna Str., BG-1000 Sofia, Bulgaria (15) Danube Private
49 University, A-3500 Krems, Austria (16) Department of Biomedical Engineering and Integrative Prehistory and
50 Archaeological Science, CH-4123 Basel-Allschwil, Switzerland (17) State Office for Heritage Management and
51 Archaeology Saxony-Anhalt and State Museum of Prehistory, D-06114 Halle, Germany (18) Romanian National
52 History Museum, Bucharest, Romania (19) Institute of Archaeology, Belgrade, Serbia (20) Institute of
53 Experimental Morphology, Pathology and Anthropology with Museum, Bulgarian Academy of Sciences, Sofia,
54 Bulgaria (21) Department of Geosciences, Biogeology, Universität Tübingen, Hölderlinstr. 12, 72074 Tübingen,
55 Germany (22) Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tuebingen, 72072

56 Tuebingen, Germany (23) Heidelberg Academy of Sciences and Humanities, Research Center “The Role of
57 Culture in Early Expansions of Humans” at the University of Tuebingen, Rümelinstraße 23, 72070 Tuebingen,
58 Germany (24) ‘Vasile Pârvan’ Institute of Archaeology, Romanian Academy (25) Human Biology Department,
59 Cardinal Stefan Wyszyński University, Warsaw, Poland (26) KADUCEJ d.o.o Papandopulova 27, 21000 Split,
60 Croatia (27) St. Cyril and Methodius University, Veliko Turnovo, Bulgaria (28) Department of Early Prehistory
61 and Quaternary Ecology, University of Tuebingen, Schloss HohenTuebingen, 72070 Tuebingen, Germany (29)
62 INRAP/UMR 8215 Trajectoires, 21 Allée de l’Université, 92023 Nanterre, France (30) Archaeological Museum of
63 Istria, Carrarina 3, 52100 Pula, Croatia (31) Service Régional de l’Archéologie de Bourgogne-Franche-Comté, 7
64 rue Charles Nodier, 25043 Besançon Cedex, France (32) Laboratoire Chronoenvironnement, UMR 6249 du
65 CNRS, UFR des Sciences et Techniques, 16 route de Gray, 25030 Besançon Cedex, France (33) Regional
66 Museum of History Veliko Tarnovo, Veliko Tarnovo, Bulgaria (34) Institute for Archaeological Sciences,
67 Paleoanthropology, University of Tuebingen, Rümelinstraße 23, 72070 Tuebingen, Germany (35) Laboratory for
68 human bio-archaeology, Bulgaria, 1202 Sofia, 42, George Washington str (36) Regional Museum of History,
69 Vratsa, Bulgaria (37) Service Régional de l’Archéologie d’Auvergne-Rhône-Alpes, Le Grenier d’Abondance, 6
70 quai Saint-Vincent, 69283 Lyon Cedex 01, France (38) Eötvös Loránd University, Faculty of Science, Institute of
71 Biology, Department of Biological Anthropology, H-1117 Pázmány Péter sétány 1/c. Budapest, Hungary (39)
72 Department of Archaeology, Sofia University St. Kliment Ohridski, Bulgaria (40) Oxford Radiocarbon
73 Accelerator Unit, Research Laboratory for Archaeology and the History of Art, University of Oxford, Dyson
74 Perrins Building, South Parks Road, OX1 3QY Oxford, UK (41) Regional Museum of History, Haskovo, Bulgaria
75 (42) Department of Anthropology, University of Wyoming, 1000 E. University Avenue, Laramie, WY 82071,
76 USA (43) Department of Archaeology, Faculty of Humanities and Social Sciences, University of Zagreb, Ivana
77 Lučića 3, 10000 Zagreb, Croatia (44) Department of Anthropology and Institutes for Energy and the Environment,
78 Pennsylvania State University, University Park, PA 16802 (45) Institute of Archaeology, National Academy of
79 Sciences of Ukraine, Kyiv, Ukraine (46) CHU Sainte-Justine Research Center, Pediatric Department, Université de
80 Montréal, Montreal, PQ, Canada, H3T 1C5 (47) National History Museum of Romania, Calea Victoriei, no. 12,
81 030026, Bucharest, Romania (48) Institute for Pre- and Protohistoric Archaeology and the Archaeology of the
82 Roman Provinces, Ludwig-Maximilians-University, Schellingstr. 12, 80799 Munich, Germany (49) Museo e
83 Istituto fiorentino di Preistoria, 50122 Florence, Italy (50) Dipartimento SAGAS - Sezione di Archeologia e Antico
84 Oriente, Università degli Studi di Firenze, 50122 Florence, Italy (51) School of History, Classics and Archaeology,
85 University of Edinburgh, Edinburgh EH8 9AG, United Kingdom (52) Conservation Department in Šibenik,
86 Ministry of Culture of the Republic of Croatia, Jurja Čulinovića 1, 22000 Šibenik, Croatia (53) Teleorman County
87 Museum, str. 1848, no. 1, 140033 Alexandria, Romania (54) Peter the Great Museum of Anthropology and
88 Ethnography (Kunstkamera) RAS, 199034 St. Petersburg, Russia (55) University of Wisconsin, Madison WI, USA
89 (56) Olga Necrasov Centre for Anthropological Research, Romanian Academy – Iași Branch, Theodor Codrescu
90 St. 2, P.C. 700481, Iași, Romania (57) Dipartimento di Scienze e tecnologie biologiche, chimiche e farmaceutiche,
91 Lab. of Anthropology, Università degli studi di Palermo, Italy (58) Anthropological Center, Croatian Academy of
92 Sciences and Arts, 10000 Zagreb, Croatia (59) Regional Historical Museum Varna, Maria Luiza Blvd. 41, BG-
93 9000 Varna, Bulgaria (60) National Museum in Belgrade, 1a Republic sq., Belgrade, Serbia (61) Department of
94 Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany (62) Department
95 of Anthropology, Natural History Museum Vienna, 1010 Vienna, Austria (63) INRAP/UMR 8215 Trajectoires, 21
96 Allée de l’Université, 92023 Nanterre, France (64) CNRS/UMR 7041 ArScAn MAE, 21 Allée de l’Université,
97 92023 Nanterre, France (65) Institute of Ethnology and Anthropology, Russian Academy of Sciences, Leninsky
98 Pr., 32a, Moscow, 119991, Russia (66) Archaeological Museum of Macedonia, Skopje (67) Regional museum of
99 history, Shumen, Bulgaria (68) Department of Anthropology, University of Toronto, Toronto, Ontario, M5S 2S2,
100 Canada (69) Institute of Archaeology & Ethnography, Siberian Branch, Russian Academy of Sciences, Lavrentiev
101 Pr. 17, Novosibirsk 630090, Russia (70) Satu Mare County Museum Archaeology Department, V. Lucaciu, nr.21,
102 Satu Mare, Romania (71) Municipal Museum Drniš, Domovinskog rata 54, 22320 Drniš, Croatia (72) anthropol -
103 Anthropologieservice, Schadenweilerstraße 80, 72379 Hechingen, Germany (73) Dipartimento di Scienze e
104 tecnologie biologiche, chimiche e farmaceutiche, Lab. of Anthropology, Università degli studi di Palermo, Italy (74)
105 Institute for Prehistory, Early History and Medieval Archaeology, University of Tuebingen, Germany (75) Institute
106 of Latvian History, University of Latvia, Kalpaka Bulvāris 4, Rīga 1050, Latvia (76) Department of Archaeology,
107 Durham University, UK (77) School of Environmental Sciences: Geography, University of Hull, Hull HU6 7RX,
108 UK (78) Department of Biology, Grand Valley State University, Allendale, Michigan, USA (79) Department of
109 Bioarchaeology, Institute of Archaeology, National Academy of Sciences of Ukraine (80) Ephorate of
110 Paleoanthropology and Speleology, Athens, Greece (81) Department of Archaeology, Cardiff University, Cardiff,
111 UK (82) Department of Anthropology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria.

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. However, the dynamics of the interaction between the first farmers and the indigenous hunter-gatherers remain poorly understood because of the near absence of ancient DNA from the region. We report new genome-wide ancient DNA data from 204 individuals—65 Paleolithic and Mesolithic, 93 Neolithic, and 46 Copper, Bronze and Iron Age—who lived in southeastern Europe and surrounding regions between about 12,000 and 500 BCE. We document that the hunter-gatherer populations of southeastern Europe, the Baltic, and the North Pontic Steppe were distinctive from those of western Europe, with a West-East cline of ancestry. We show that the people who brought farming to Europe were not part of a single population, as early farmers from southern Greece are not descended from the Neolithic population of northwestern Anatolia that was ancestral to all other European farmers. The ancestors of the first farmers of northern and western Europe passed through southeastern Europe with limited admixture with local hunter-gatherers, but we show that some groups that remained in the region mixed extensively with local hunter-gatherers, with relatively sex-balanced admixture compared to the male-biased hunter-gatherer admixture that we show prevailed later in the North and West. After the spread of farming, southeastern Europe continued to be a nexus between East and West, with intermittent steppe ancestry, including in individuals from the Varna I cemetery and associated with the Cucuteni-Trypillian archaeological complex, up to 2,000 years before the Steppe migration that replaced much of northern Europe's population.

Introduction

The southeastern quadrant of Europe was the beachhead in the spread of agriculture into Europe from its source in the Fertile Crescent of southwestern Asia. After the first appearance of agriculture in the 7th millennium BCE^{1,2}, southeastern Europe incubated a succession of Early Neolithic cultures prior to the spread of farming westward via a Mediterranean route and northwestward via a Danubian route, reaching 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 people⁵⁻⁸ closely related to the farmers of northwestern Anatolia⁹⁻¹¹ but nearly all the evidence about the ancestry of the first farmers in Europe is from central and far western Europe, with only three individuals reported from northern Greece⁹. In the millennia following the establishment of agriculture in the Balkan

Peninsula, a series of complex societies formed, culminating in large tell settlements and sites like the mid 5th millennium BCE necropolis at Varna, which has some of the earliest evidence of extreme inequality in wealth, with one individual (ANI152/grave 43) from whom we extracted DNA buried with more gold than is known from any site prior to that time. 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, with densely settled “mega-sites” in Ukraine housing hundreds, and perhaps thousands, of people.¹² After around 4000 BCE, these societies further transformed, the tell settlements were largely abandoned, and there is archaeological evidence of contact with nomadic pastoralist populations from the Eurasian steppe. However, the population movements that accompanied these events are not immediately evident from the archaeological record and remain largely unknown.

Results

We generated new genome-wide data from 204 ancient humans (196 reported for the first time), from the Balkan Peninsula, the Carpathian Basin, the North Pontic Steppe, and neighboring regions mostly dated to between 12,000 and 1000 BCE (Figure 1A, Supplementary Data Table 1, Supplementary Information, section 1). To obtain genome-wide data in an efficient way, we enriched the DNA libraries in solution for sequences overlapping 1.24 million single nucleotide polymorphisms (SNPs) before sequencing.^{7,10,13} We filtered out individuals with low coverage (less 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 also report data for, but exclude from analysis, nine individuals that were first-degree relatives of others in the dataset, resulting in an analysis dataset of 195 individuals.

We applied principal component analysis (PCA; Figure 1B, Extended Data Figure 1), and supervised and unsupervised ADMIXTURE (Figure 1D, Extended Data Figures 2 and 3)¹⁴ to obtain an overview of this dataset in terms of population structure. We combined this genetic assessment of structure with archaeological and chronological information to cluster the individuals into populations. We used *D*-statistics to evaluate whether pairs of populations were consistent with being clades¹⁵ and the programs *qpAdm* and *qpGraph* to fit mixture proportions and admixture graphs to the data¹⁵. To investigate how these individuals fit into the wider context of European prehistory and present-day genetic diversity, we co-analyzed them with data from 265 previously reported ancient individuals^{9-11,16-28} as well as 799 present-day individuals genotyped on the Illumina “Human Origins” array²⁴, and 300 high coverage genomes from the Simons Genome Diversity Project²⁹ (SGDP).

Hunter-Gatherer substructure and transitions

We report new genome-wide data for 101 individuals from Paleolithic, Mesolithic and eastern European Neolithic contexts. In eastern Europe, the term “Neolithic” often refers to the presence of pottery³⁰⁻³² (and may not include a transition to agricultural subsistence which is the key criterion for the use of these terms in western Europe). Because of the differences in meaning across our region of study, we avoid the use of “Neolithic” as a general term and use terms corresponding to economic subsistence strategy (either “hunter-gatherer” or “farmer”), or genetic ancestry (we use “hunter-gatherer ancestry” to refer to genetic ancestry derived from a population closely related to Mesolithic Europeans, and “farmer ancestry” to refer to ancestry derived from a population like northwestern Anatolian farmers).

Hunter-gatherers from central Europe have both western and eastern European hunter-gatherer ancestry (WHG and EHG) – a cline that is clearly visible in PCA (Figure 1B). This motivated us to investigate whether genetic population structure of Mesolithic and Early Neolithic hunter-gatherers in Europe was determined purely by physical distance. We fit admixture proportions with *qpAdm* and also a model estimating a spatially continuous migration surface under an isolation-by-distance model,³³ and inferred a migration barrier separating populations that are predominantly WHG from EHG, but with some diffusion of ancestry across this boundary (Figure 2A). However, we also show that this frontier was not static, with dramatic local shifts in ancestry over time (Figure 2B) and substructure in phenotypically important variants (Supplementary Information, section 2).

From present-day Ukraine, our study reports new genome-wide data from 31 hunter-gatherers: five Mesolithic individuals dating from 9500-6000 BCE, and 26 Neolithic and Copper Age (“Eneolithic” or “Chalcolithic”) individuals dating from ~6000-3500 BCE. On the cline from western (WHG) to eastern hunter-gatherer ancestry (EHG, represented by individuals from Karelia and Samara), the Ukrainian Mesolithic individuals fall towards the eastern end, intermediate between EHG and hunter-gatherers from Sweden⁷. The Ukrainian Neolithic population has significant differences in ancestry compared to the Ukrainian Mesolithic population. A previous study of two individuals (for one of which we generated new data) from the Mesolithic Vasil’evka 3 and Neolithic Vovnigi 2 sites³⁴ suggested that between the Ukrainian Mesolithic and Neolithic there was an increase in Ancient North Eurasian (ANE) ancestry related to the 24,000-year old Siberian Mal’ta 1.²⁶ However, our larger sample shows the opposite – specifically that ANE ancestry decreases and WHG ancestry increases – as shown by the statistics $D(\text{Mbuti}, X, \text{Ukraine_Mesolithic}, \text{Ukraine_Neolithic})$, which is $Z=-4.9$ when X is the Mal’ta 1 individual and $Z=9.1$ for $X=\text{WHG}$ (Supplementary Data Table 2).

Individuals associated with the Bronze Age Yamnaya Complex from Ukraine, like previously reported Yamnaya individuals from Samara⁷ and Kalmykia¹⁶, have little evidence of WHG ancestry, but do have a third source of ancestry related to hunter-gatherers from the Caucasus²⁰ (CHG) and early Iranian farmers^{23,35} (Supplementary Data Table 3). Two Yamnaya individuals – one from Ozero in Ukraine and one from Bulgaria (I1917 and Bul4, both dated to ~3000 BCE) – in addition have evidence of early European farmer related admixture, which is the first evidence of such ancestry in Yamnaya individuals (Figure 1B,D, Supplementary Data Table 2). Similarly, one Copper Age individual (I4110) dated to ~3600-3400 BCE from Dereivka in Ukraine has both CHG and farmer ancestry (Figure 1D, Supplementary Data Table 2). This is by far the earliest appearance of farmer ancestry this far East in Eurasia, which was previously not known on the Steppe until the Srubnaya Complex after ~1800 BCE.

At Zvejnieki in Latvia (18 newly reported individuals, added to 4 first reported in Ref. 34 for which we report additional data here) there is a transition in hunter-gatherer ancestry that is the opposite of that seen in Ukraine. Consistent with similar data from the Baltic States^{34,36,37}, we find that Mesolithic and Early Neolithic individuals associated with the Kunda and Narva cultures had ancestry that was genetically intermediate between western and eastern hunter-gatherers (we estimate 70% WHG and 30% EHG, but perhaps with some differential relatedness to ANE, relative to EHG; Supplementary Data Table 3). However, there is a dramatic shift between the Early Neolithic and the Middle Neolithic Comb Ware Complex, who are almost entirely EHG in ancestry (we estimate 73% EHG, but two out of four individuals appear almost 100% EHG in PCA space). The most recent individual, associated with the Final Neolithic Corded Ware Complex, provides evidence of yet another transition, clustering closely with Yamnaya from Samara⁷, Kalmykia¹⁶ and Ukraine.

We report new data from hunter-gatherers from France, Sicily and Croatia, as well as higher coverage data from three previously published hunter-gatherers from France and Germany.¹⁸ The Sicilian and Croatian individuals dating to 12,000 and 6100 BCE cluster closely with western hunter-gatherers, including individuals from Loschbour²⁴ (Luxembourg, 6100 BCE), Bichon²⁰ (Switzerland, 11,700 BCE), and Villabruna¹⁸ (Italy 12,000 BCE). These results demonstrate that the “western hunter-gatherer” population²⁴ was widely distributed from the Atlantic seaboard of Europe in the West, to Sicily in the South, to the Balkan Peninsula in the Southeast, for at least six thousand years, strengthening the evidence that the western hunter-gatherers represent a population that expanded from a southeastern European refugium following the last Ice Age around 15,000 years ago—in the process displacing or admixing with the existing population of western Europe.^{18 38}

A particularly important hunter-gatherer population that we newly report in this study is from the Iron Gates region that straddles the border of present-day Romania and Serbia. This region is close to the route taken by farmer migrants on their way from the Balkans to central Europe, and the population is represented in our study by 35 individuals from the Iron Gates sites of Hajdučka Vodenica, Ostrovul Corbului, Padina, Schela Cladovei and Vlasac, one individual (I2534) with similar (but more EHG) hunter-gatherer ancestry from the Early Neolithic site of Măgura Buduiasca 250km east of the Iron Gates, and two individuals with farmer ancestry from the Mesolithic-Neolithic site of Lepenski Vir. Modeling Iron Gates hunter-gatherers as a mixture of WHG and EHG (Supplementary Table 3) shows that they are – as expected given their geographic location and the hunter-gatherer ancestry cline – intermediate between WHG (87%) and EHG (13%). However, this *qpAdm* model does not fit well ($p=0.0003$, Supplementary table 3) and we note that the Iron Gates hunter-gatherers carry mitochondrial haplogroups K1 (8/36) as well as other subclades of haplogroup U (27/36) and haplogroup H (1/36). This contrasts with WHG, EHG and Scandinavian hunter-gatherers who almost all carry haplogroup U5 or U2. Therefore the Iron Gates hunter-gatherers have ancestry that is not present in WHG or EHG. This suggests either genetic contact between the ancestors of the Iron Gates population and hunter-gatherers from Anatolia, or that the Iron Gates population is related to the source population from which the WHG split off during a post-LGM re-expansion into Europe.

In contrast, the two individuals (I4665 and I4666, dated to 6205-5907 calBCE and 6222-5912 calBCE respectively) that we sampled from Iron Gates site of Lepenski Vir are genetically farmers rather than hunter-gatherers, despite having been buried in the local Mesolithic tradition. Strontium isotope data shows that many of the individuals buried after ~6100 BCE at Lepenski Vir—including one of the two that we sampled (I4665, burial 54E)—were not originally from the Danube Gorges.³⁹ These observations, combined with one individual from Padina (I5232), dated to 6061-5841 calBCE that has both farmer and hunter-gatherer ancestry, demonstrates that the Iron Gates region was one where farmer and hunter-gatherer groups interacted both genetically and culturally, and provides a window into the first few generations of interactions between these disparate groups.

Population structure and transformation of farmers in southeast Europe

Neolithic populations from present-day Bulgaria, Croatia, Macedonia, Serbia and Romania cluster closely with the Neolithic populations of northwestern Anatolia (Figure 1). Modeling Balkan Neolithic populations as a mixture of Anatolia Neolithic, western hunter-gatherer and Ukraine Mesolithic, we estimate that Balkan Neolithic populations derive 98% (95% confidence interval [CI]; 97-100%) of their ancestry from populations related to those of the

northwestern Anatolian Neolithic. A striking exception to the pattern of limited hunter-gatherer admixture in Balkan Neolithic populations is evident in 8 out of 9 individuals from Malak Preslavets, a site in present-day Bulgaria close to the Danube river.⁴⁰ These individuals likely lived in the mid-6th millennium BCE and have significantly more hunter-gatherer ancestry than other Balkan Neolithic individuals as shown by PCA and ADMIXTURE as well as *D*-statistics and *qpAdm* modeling (Figure 1B,D, Extended Data Figures 1-3, Supplementary Data Tables 2-4). We find that a model of 82% (CI: 77-86%) Anatolian Neolithic, 15% (CI: 12-17%) WHG, and 4% (CI: 0-9%) EHG ancestry is a good fit to the data. This hunter-gatherer ancestry with a ~4:1 WHG:EHG ratio plausibly represents a contribution from local Balkan hunter-gatherers like those that we sampled from the Iron Gates. By the Late Mesolithic, hunter-gatherers populations in the Balkans were likely concentrated in sites along the coast and major rivers such as the Danube,⁴¹ which directly connects the Iron Gates with Malak Preslavets. This suggests a heterogeneous landscape of farmer populations with different proportions of hunter-gatherer ancestry during the early Neolithic, and that farmer groups with the most hunter-gatherer ancestry, like those at Malak Preslavets and, possibly, Lepenski Vir were those that lived close to the highest densities of hunter-gatherers.

In the Balkans, Copper Age populations have significantly more hunter-gatherer ancestry than Neolithic populations as shown, for example, by the statistic *D*(Mbuti, WHG, Balkans_Neolithic, Balkans_Chalcolithic); *Z*=5.18 (Supplementary Data Table 2). This is consistent with changes in funeral rites⁴² and roughly contemporary with the “resurgence” of hunter-gatherer ancestry previously reported in central Europe and Iberia^{7,10,43}.

We also report the first data from the Late Neolithic Globular Amphora Complex. Globular Amphora individuals from two sites in Poland and Ukraine form a tight genetic cluster, showing genetic homogeneity over a large distance (Figure 1B,D). We find that this population had more hunter-gatherer ancestry than Middle Neolithic groups from Central Europe⁷ (we estimate 25% [CI: 22-27%] WHG ancestry, similar to Chalcolithic Iberia). This finding further extends our knowledge of the variable landscape of hunter-gatherer and farmer admixture proportions in Europe (Supplementary Data Table 3). In east-central Europe, the Globular Amphora Complex immediately precedes the Corded Ware Complex that marks the first appearance of steppe ancestry in the region.^{7,16} The Globular Amphora abutted populations with steppe-influenced material cultures for hundreds of years and yet the individuals in our study have no evidence of steppe ancestry, suggesting that this persistent culture frontier corresponded to a genetic barrier.

The migrations from the Pontic-Caspian steppe associated with the Yamnaya Cultural Complex in the 3rd millennium BCE made a profound contribution to the genetic ancestry of central Europe, contributing 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.^{7,16} In a few individuals from southeastern Europe, we find evidence of steppe-related ancestry far earlier (defined here as a mixture of EHG and CHG similar to the genetic signature of individuals of the later Yamnaya; Figure 1B,D). One individual (ANI163) from the Varna I cemetery dates to 4711-4550 BCE, one (I2181) from nearby Smyadovo dates to 4550-4450 BCE, and a third individual (I1927) from Verteba cave, associated with the Cucuteni-Trypillian complex, dates to 3619-2936 BCE. These findings push back by almost 2000 years the first evidence of steppe ancestry this far West in Europe, demonstrating the resumption of genetic contact between southeastern Europe and the Steppe that also occurred in the Mesolithic. Other Copper Age (~5000-4000 BCE) individuals from the Balkans have little evidence of steppe ancestry, but Bronze Age (~3400-1100 BCE) individuals do (we estimate 30%; CI: 26-35%). The four latest Balkan Bronze Age individuals in our data (later than ~1700 BCE) all have more steppe ancestry than earlier Bronze Age individuals (3200-2500 BCE; Figure 1D), showing that the contribution of the Steppe to southeast European populations increased further during the Bronze Age.

New evidence about the spread of farming into, and throughout, Europe

This study resolves two open questions about the initial spread of farming into Europe. The first is the question of whether the first farmers of the Danubian Route that brought agriculture to northern Europe along the Danube River valley, and those that spread along the Mediterranean coast to Iberia and other southern European locations, were derived from a single ancestral population or instead represent separate migrations from different Anatolian sources. A challenge in studying the relationship among Early Neolithic populations is the different proportions of hunter-gatherer ancestry they carry, which obscures the more subtle differences between their farmer ancestries. One approach to this problem is to explicitly model both sources of ancestry using an admixture graph (Supplementary Information, section 3). We confirm that Mediterranean populations, represented in our study by individuals of the Impressa complex from Croatia and the Epicardial Early Neolithic from Spain⁷, are closely related to the Danubian population represented by the *Linearbandkeramik* (LBK) from central Europe^{7,44} and show that both groups are closely related to the Balkan Neolithic population. These three populations form a clade with Northwest Anatolians as an outgroup, consistent with a single migration from a population closely related to the northwestern Anatolian Neolithic farming population into the Balkan peninsula, which then split into two populations that followed the Danubian and Mediterranean routes.

A related question about the spread of farming into Europe concerns whether its initial arrival in present-day Greece and subsequent expansion was mediated by a single population migrating from Anatolia – as has been consistent with genetic data up until now⁹ – or whether there were multiple initial groups, as suggested by the archaeological record.^{45,46} We find that four southern Greek (Peloponnese) Neolithic individuals – three from Diros Cave and one from Franchthi Cave, plus one previously published individual from Diros²⁷ – are not consistent with descending from the same source population as other European farmers. In PCA these individuals are outliers; shifted away from northwestern Anatolian and European Early Neolithic individuals, in a direction opposite from WHG. *D*-statistics (Supplementary Data Table 2) show that in fact, these “Peloponnese Neolithic” individuals have less WHG-related ancestry than Anatolia Neolithic ones, and that they form an outgroup relative to Anatolian and Balkans Neolithic populations, suggesting an independent migration into Europe from a population that split off from the ancestors of the northwest Anatolian individuals from which we have data. Admixture graph modeling (Supplementary Information, section 3) supports this interpretation, confirming that their Near Eastern ancestry is derived from a lineage that is close, or basal, to the non-WHG component of Anatolian Neolithic ancestry. One possibility is that this independent migration is related to an earlier Aceramic Neolithic in Greece that was derived from the pre-pottery Neolithic (PPNB) of Cyprus and the Levant⁴⁶. Under this model, the earliest Neolithic populations in Greece migrated from the Levant, perhaps via the southwestern Anatolian coast as early as 7000 BCE,^{45,46} but the majority of Neolithic ancestry arrived around 500 years later via a route that passed through northwestern Anatolia. The predictions of this hypothesis could be further tested with genome-wide data of Early Neolithic individuals from Cyprus, Crete and southwest Anatolia. Populations related to the Peloponnese Neolithic potentially made a small contribution to the ancestry of other Mediterranean Neolithic populations like Early Neolithic Iberia and Neolithic farmers from northern Greece⁹ but we do not strongly reject models without such a contribution (Supplementary Information, section 3).

Sex-biased admixture between hunter-gatherers and farmers

We provide the first evidence for sex-biased admixture between hunter-gatherers and farmers in Europe, showing that the Middle Neolithic “resurgence” of hunter-gatherer ancestry^{7,43} in central Europe and Iberia was driven more by male than by female hunter-gatherers (Figure 3B&C, Supplementary Data Table 5). One way of detecting this is to compare ancestry proportions on the autosomes and the X chromosome. Since males always inherit their X chromosome from their mothers, differences between ancestry on the autosomes and chromosome X imply sex-biased mixture. In the Balkan Neolithic there is no evidence of sex

bias using ancestry estimates obtained from *qpAdm* ($Z=-0.65$ where a positive Z -score implies male hunter-gatherer bias), nor in the LBK and Iberian_Early Neolithic ($Z=-0.24$ and 1.04). However, in the Middle Neolithic and later populations, this effect reverses. In the Balkan Copper Age there is weak evidence of bias ($Z=1.77$) but in Iberia and central Europe Middle Neolithic there is clear bias is in favor of male hunter-gatherer ancestry ($Z=2.98$, and 2.77 in Iberia Copper Age and central European Middle Neolithic). This result is independently supported by uniparental markers (Figure 3C). Proportions of typically hunter-gatherer mitochondrial haplogroups (haplogroup U)⁴⁷ are low in all populations and within the intervals of genome-wide ancestry proportions. On the other hand, hunter-gatherer Y chromosomes (haplogroups I2, R1 and C1)¹⁸ are much more common: 6/7 in the Iberian Neolithic/Copper Age and 7/8 in Middle-Late Neolithic central Europe (Central_MN and Globular_Amphora). Under a single pulse model of admixture, the autosomal/X chromosome ancestry proportions imply that in the central European Middle Neolithic population that shows the strongest evidence of sex bias, 35-50% of the male ancestors were hunter-gatherers, compared to 0-5% of the female ancestors (Extended Data Figure 4).

The merging of hunter-gatherer and farmer populations was a dynamic process that unfolded over thousands of years, and proceeded in a profoundly different way in different parts of Europe. Our analysis shows that in some places – for example at Malak Preslavets in Bulgaria – there was extensive mixing between hunter-gatherers and farmers, likely driven by the high local hunter-gatherer population density. In other places—in particular in western, central and northern Europe—hunter-gatherers and farmers lived in close proximity for long periods of time with minimal mixture^{6,43,48}. When they did finally mix, we find evidence that the hunter-gather admixture was male-biased, implying a different dynamic. Farming was initially unable to expand widely in central and northern Europe because early farming techniques were only suitable for specific regions within the loess belt of the northern European plain. Thus, northern and central European hunter-gatherers were protected from the demographic impact of farming migrations, resulting in persistent frontiers between farmers and hunter-gatherers.^{49,50} This may have given hunter-gatherers and farmers time to learn from each other and interact in a different way than during the more rapid expansion of the first farmers in the South.

No evidence of Copper Age Balkans-to-Anatolia migration

One version of the Steppe Hypothesis of Indo-European language origins suggests that Proto-Indo European languages developed in the steppe north of the Black and Caspian seas, and that the earliest known diverging branch – Anatolian – was spread into Asia Minor by movements of steppe peoples through the Balkan peninsula during the Copper Age around

4000 BCE, as part of the same incursions from the steppe that coincided with the decline of the tell settlements.⁵¹ If this were correct, then one way to detect evidence of it would be the appearance of large amounts of characteristic steppe ancestry first in the Balkan Peninsula, and then in Anatolia. However, our genetic data do not support this scenario. While we find steppe ancestry in Balkan Copper Age and Bronze Age individuals, this ancestry is sporadic across individuals in the Copper Age, and at low levels in the Bronze Age. Moreover, while Bronze Age Anatolian individuals²⁷ have CHG / Iran Neolithic related ancestry, they have neither the EHG ancestry characteristic of all steppe populations sampled to date²⁰, nor the WHG ancestry that is ubiquitous in southeastern Europe in the Neolithic (Figure 1A, Supplementary Data Table 2, Supplementary Information section 1). This pattern is consistent with that seen in northwestern Anatolia¹¹ and later in Copper Age Anatolia²³, suggesting continuing migration into Anatolia from the East rather than from Europe.

An alternative hypothesis is that the ultimate homeland of Proto-Indo European languages was in the Caucasus or in Iran. In this scenario, westward movement contributed to the dispersal of Anatolian languages, and northward movement and mixture with EHG was responsible for the formation of the population associated with the Yamnaya complex. These steppe pastoralists plausibly spoke a “Late Proto-Indo European” language that is ancestral to many of the non-Anatolian branches of the Indo-European language family.⁵² On the other hand, our data could still be consistent with the Steppe-Balkans-Anatolia route hypothesis model, albeit with constraints. It remains possible that populations dating to around 1600 BCE in the regions where the Indo-European Luwian, Hittite and Palaic languages were spoken did have European hunter-gatherer ancestry. However, our results would require that such ancestry was not ubiquitous in Bronze Age Anatolia, and was perhaps tightly linked to Indo-European speaking groups. We predict that additional insight about the genetic origins of the potential speakers of early Indo-European languages will be obtained when ancient DNA data become available from additional sites in this key period in Anatolia and the Caucasus.

Discussion

Our study shows that southeastern Europe consistently served as a genetic contact zone between different populations. This role likely contributed to the extraordinary series of cultural innovations that characterize the region, from the elegant figurines of the Neolithic to the ornaments and precious metalwork of Varna. Before the arrival of farming, this region saw constant interaction between highly diverged groups of hunter-gatherers, and this interaction continued, perhaps accelerating, after the arrival of farming. We find evidence that some early farmers from Greece derived ancestry from a different source compared to the one

that contributed the majority of ancestry of all other farmers in Europe. In eastern Europe we document the appearance of CHG/Iranian Neolithic ancestry north of the Black Sea, and its eventual extension as far north as the Baltic. In some ways, this expansion parallels the expansion of Anatolian farmer ancestry into western Europe although it is less dramatic, and several thousand years later. These expansions set up the two, largely separate, populations in western and eastern Europe that would come together in the Final Neolithic and Early Bronze Age to form the ancestry of present-day Europe.

This study describes key ancestral components that contributed to present-day West Eurasian genetic diversity. However, the more recent processes that created present-day southeastern European populations are unknown, and understanding this will require dense sampling of Bronze Age, Iron Age, Roman, and Medieval groups and comparison to present-day populations. At the most ancient end of our time series, while information about hunter-gatherer population structure in northern and western Europe now extends back throughout the Upper Paleolithic,¹⁸ we have little data about how these populations fit into a wider Eurasian context, and more data from hunter-gatherer populations in Anatolia, the Near East and East Asia will be needed to resolve that question. Finally, many questions about the nature of the interactions between populations remain unresolved. For example, we report evidence for sex-bias in one particular set of interactions between hunter-gatherers and farmers, and other interactions may have had similar dynamics^{37,53}. However, many more examples of such interactions need to be collected before it will become possible to make generalizable claims about the patterns of sex-biased interactions among human populations as they came into contact and mixed during prehistory.

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). Sample powder was also generated in a fifth laboratory (Dublin) and 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 methods⁷ and sent to Boston for subsequent screening, 1240k capture and sequencing.

Seven samples were processed as previously described²⁸ 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. Extraction⁵⁴ and library preparation^{55,56} 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 either single or paired-end.

The remaining 195 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 methods⁵⁴ replacing the column assembly with the column extenders from a Roche kit.⁵⁷ We prepared double barcoded libraries with truncated adapters from between a ninth to a 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 data⁵⁸, but works inefficiently at 5’ ends,⁵⁶ and thereby leaves 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.^{59,60}

537

538 We screened all libraries from Adelaide, Boston and Budapest by enriching for the
539 mitochondrial genome plus about 3000 (50 in an earlier, unpublished, version) nuclear SNPs
540 using a bead-capture⁶¹ but with the probes replaced by amplified oligonucleotides synthesized
541 by CustomArray Inc. After the capture, the adapter sites were completed by PCR, and thereby
542 dual index combinations⁶² were attached to each enriched library. We sequenced the products
543 of between 100 and 200 libraries together with the non-enriched libraries (shotgun) on an
544 Illumina NextSeq500 using v2 150 cycle kits for 2x76 cycles and 2x7 cycles.

545

546 In Boston, We performed two rounds of in-solution enrichment (“1240k capture”) for a
547 targeted set of 1,237,207 SNPs using previously reported protocols.^{7,13,24} When we enriched
548 additional libraries to increase coverage, multiple libraries from the same sample were pooled
549 in equimolar ratios before the capture. All sequencing was performed on an Illumina
550 NextSeq500 using v2 150 cycle kits for 2x76 cycles and 2x7 cycles. We attempted to
551 sequence each enriched library up to the point where we estimated that it was economically
552 inefficient to sequence further. Specifically, we iteratively sequenced more and more from
553 each individual and only stopped when we estimated that the expected increase in the number
554 of targeted SNPs hit at least once would be less than about one for every 100 new read pairs
555 generated. After sequencing, we removed individuals with evidence of contamination based
556 on mitochondrial DNA polymorphism⁶³ or difference in PCA space between damaged and
557 undamaged reads⁶⁴, a high rate of heterozygosity on chromosome X despite being male^{64,65}, or
558 an atypical ratio of X to Y sequences. We report, but do not analyze, data from 17 individuals
559 that had low coverage (less than 15,000 SNPs hit on the autosomes), were first-degree
560 relatives of others in the dataset, or were undated and had unusual ancestry for their
561 archaeological context.

562

563 After removing a small number of sites that failed to capture, we were left with a total of
564 1,233,013 sites of which 32,670 were on chromosome X and 49,704 were on chromosome Y,
565 with a median coverage at targeted SNPs on the 138 new individuals of 0.73 (range 0.017-
566 9.2; Supplementary Table 1). We generated “pseudo-haploid” calls by selecting a single read
567 randomly for each individual at each SNP. Thus, there is only a single allele from each
568 individual at each site, but adjacent alleles might come from either of the two haplotypes of
569 the individual. We merged the newly reported data with previously reported data from 266
570 other ancient individuals^{9-11,16-28}, making pseudo haploid calls in the same way at the 1240k
571 sites for individuals that were shotgun sequenced rather than captured.

572

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 Project²⁹ (SGDP merge).
2. 799 West Eurasian individuals genotyped on the Human Origins array²⁴, 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* (v1.5100)⁶⁶ (<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 (so, for $k=4$, we used labels 1-4 in the list below, for example):

1. *Anatolia_Neolithic*
2. *WHG*
3. *EHG*
4. *Yamnaya*
5. *Ukraine_Mesolithic*
6. *SHG*

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), and fitted admixture proportions with *qpAdm* (v610 and admixture graphs with *qpGraph* (v6021)¹⁵ (<https://github.com/DReichLab/> *AdmixTools*), in each case using the SGDP merge. We computed standard errors for *Z* scores with the default block jackknife parameters. For *qpAdm* we used the following set of seven populations as outgroups or “right populations”:

- *Mbuti.DG*
- *Ust_Ishim_HG_published.DG*
- *Mota.SG*
- *MA1_HG.SG*
- *Villabruna*
- *Papuan.DG*
- *Onge.DG*
- *Han.DG*

For some analyses (Extended Data Table 4) we used an extended set of 14 right populations, including additional Upper Paleolithic European individuals¹⁸:

- *ElMiron*
- *Mota.SG*
- *Mbuti.DG*
- *Ust_Ishim_HG_published.DG*
- *MA1_HG.SG*
- *AfontovaGora3*
- *GoyetQ116-1_published*
- *Villabruna*
- *Kostenki14*
- *Vestonice16*
- *Karitiana.DG*
- *Papuan.DG*
- *Onge.DG*
- *Han.DG*

For 40 Mesolithic individuals we estimated an effective migration surface using the software *EEMS* (<https://github.com/dipetkov/eems>)³³. We computed pairwise differences between individuals using the *bed2diffs2* program provided with *EEMS*. We set the number of demes to 400 and defined the outer boundary of the region by the polygon (in latitude-longitude coordinates) [(66,60), (60,10), (45,-15), (35,-10), (35,60)]. We ran the MCMC ten times with different random seeds, each time with one million burn-in and four million regular iterations, thinned to one in ten thousand.

To analyze potential sex bias in admixture, we used *qpAdm* to estimate admixture proportions on the autosomes (default option) and on the X chromosome (option “*chrom: 23*”). We computed Z scores for the difference between the autosomes and the X chromosome as

$$Z = \frac{p_A - p_X}{\sqrt{\sigma_A^2 + \sigma_X^2}}$$

where p_A and p_X are the hunter-gatherer admixture proportions on the autosomes and the X chromosome, and σ_A and σ_X are the corresponding jackknife standard deviations. Thus, a positive Z-score means that there is more hunter-gatherer admixture on the autosomes than on the X chromosome and thus the hunter-gatherer admixture was male-biased. Because X chromosome standard errors are high and *qpAdm* results can be sensitive to which

population is first in the list of outgroup populations, we checked that the patterns we observe were robust to cyclic permutation of the outgroups. To compare frequencies of hunter-gatherer uniparental markers we counted the individuals with mitochondrial haplogroup U and Y chromosome haplogroups C2, I2 and R1 which are all common in Mesolithic hunter-gatherers but rare or absent in the Anatolian Neolithic. The Iron Gates hunter-gatherers also carry H and K1 mitochondrial haplogroups so the proportion of haplogroup U represents the minimum maternal hunter-gatherer contribution. We computed binomial confidence intervals for the proportion of markers using the Agresti-Coull method^{67,68} implemented in the *binom* package in R.

Given autosomal and X chromosome admixture proportions, we estimated the proportion of male and female hunter-gatherer ancestors by assuming a single pulse model of admixture. If the proportions of male and female ancestors that are hunter-gatherer are given by m and f , respectively, then the proportions of hunter-gatherer ancestry on the autosomes and the X chromosome are given by $\frac{m+f}{2}$ and $\frac{m+2f}{3}$. We approximated the sampling error in the observed admixture proportions by the estimated jackknife sampling error, computed the likelihood surface for (m, f) over a grid ranging from (0,0) to (1,1).

Direct AMS ¹⁴C Bone Dates

We report new direct AMS ¹⁴C bone dates in this study from multiple AMS radiocarbon laboratories. In general, bone samples were manually cleaned and demineralized in weak HCl and, in most cases (PSU, UCIAMS, OxA), soaked in an alkali bath (NaOH) at room temperature to remove contaminating soil humates. Samples were then rinsed to neutrality in Nanopure H₂O and gelatinized in HCL.⁶⁹ The resulting gelatin was lyophilized and weighed to determine percent yield as a measure of collagen preservation (% crude gelatin yield). Collagen was then directly AMS ¹⁴C dated (Beta, AA) or further purified using ultrafiltration (PSU/UCIAMS, OxA, Poz, Wk, MAMS).⁷⁰ It is standard in some laboratories (PSU/UCIAMS, OxA, Wk) to use stable carbon and nitrogen isotopes as an additional quality control measure. For these samples, the %C, %N and C:N ratios were evaluated before AMS ¹⁴C dating. C/N ratios for well-preserved samples fall between 2.9 and 3.6, indicating good collagen preservation.⁷¹

All ¹⁴C ages were $\delta^{13}\text{C}$ -corrected for mass dependent fractionation with measured ¹³C/¹²C values⁷² and calibrated with OxCal version 4.2.3⁷³ using the IntCal13 northern hemisphere calibration curve.⁷³

Acknowledgments

We thank David Anthony, Iosif Lazaridis, and Mark Lipson for comments on the manuscript; Bastien Llamas and Alan Cooper for supporting laboratory work. Support for this project was provided by the Human Frontier Science Program fellowship LT001095/2014-L to I.M.; by DFG grant AL 287 / 14-1 to K.W.A.; by Irish Research Council grant GOIPG/2013/36 to D.F.; 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.

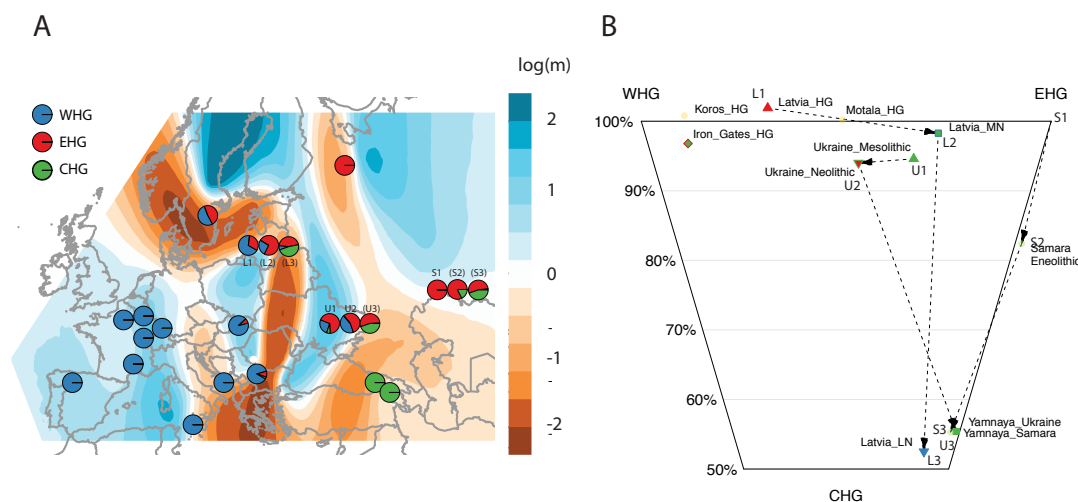


Figure 2: Structure and population change in European populations with mainly hunter-gatherer ancestry. **Background:** Effective migration surface³³ inferred using 109 Mesolithic and Early Neolithic genomes. Blue indicates regions of high migration and red regions of low migration (relative migration rate m). **Pies:** Each pie represents one or more individuals. Blue, red and green segments pies show, respectively, WHG, EHG and CHG ancestry – we chose a 2-population model if it fitted in *qpAdm* with $p > 0.01$ (Supplementary Data Table 3). Pies with 100% ancestry are those that were fixed as source populations. In Latvia, Samara and Ukraine, sets of three pies show the transition in ancestries, approximately from Mesolithic-Neolithic-Yamnaya (Exact labels in panel B). Populations with labels in brackets were not used to fit the migration surface. **B:** Ancestry transitions for hunter-gatherer populations in the 3-population WHG+EHG+CHG model. Dashed lines show populations with geographic continuity.

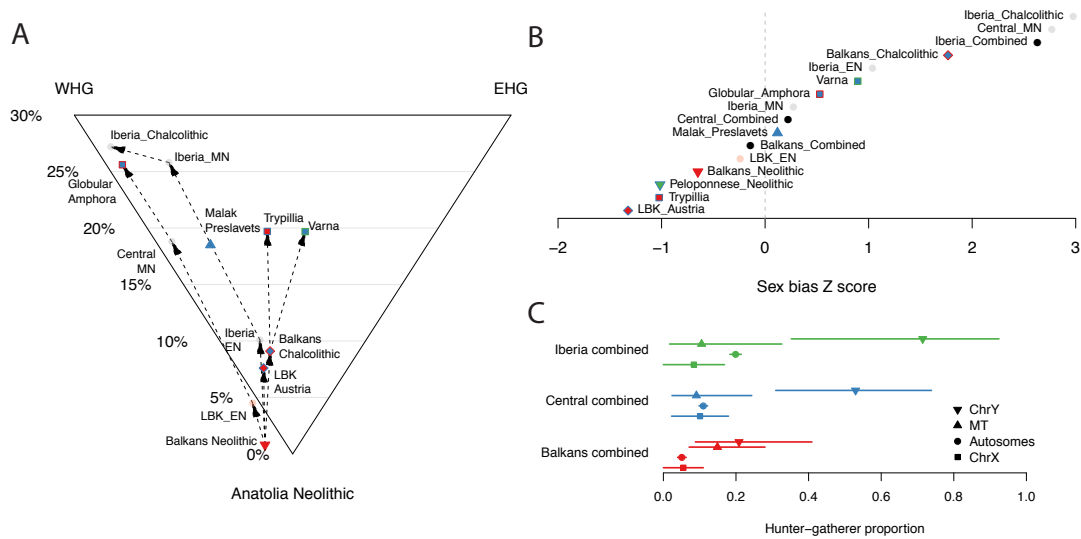
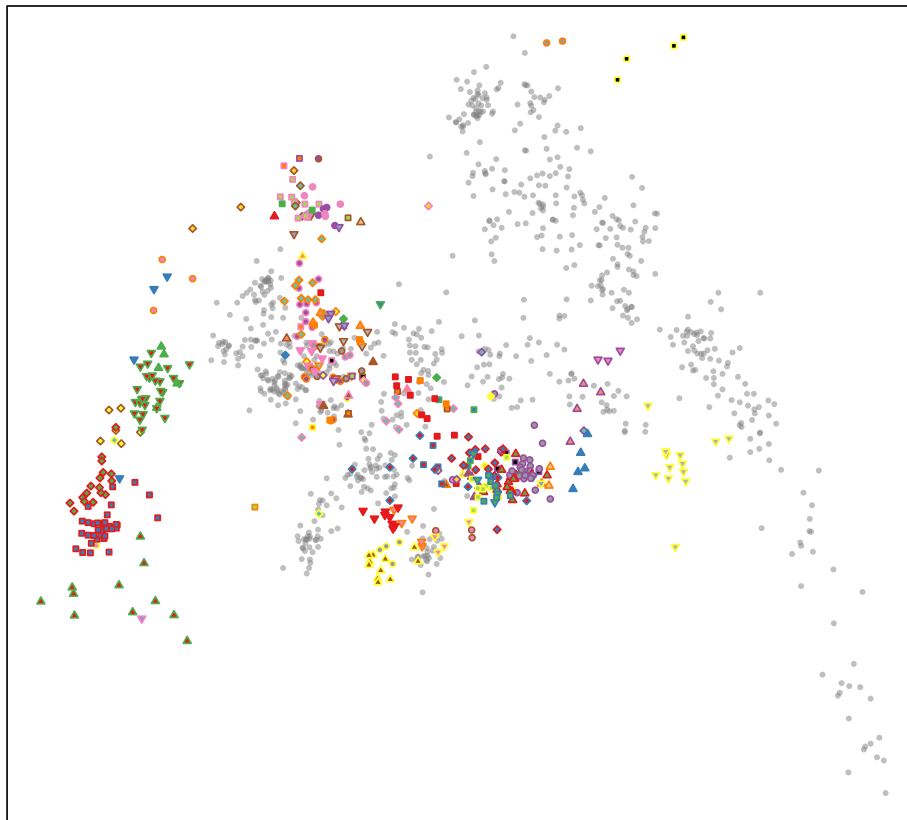


Figure 3: Structure and population change in European populations with farmer ancestry. **A:** each population is modeled as a mixture of Anatolia Neolithic, WHG, and EHG. Dashed lines show temporal relationships between populations from the same geographic areas with similar ancestries. **B:** Z-scores for the difference in hunter-gatherer ancestry on the autosomes compared to the X chromosome when populations are modeled as a mixture of Anatolia Neolithic and WHG. A positive score indicates that there is more hunter-gatherer ancestry on the autosomes and therefore the hunter-gatherer ancestry is male-biased. **C:** Hunter-gatherer ancestry proportions on the autosomes, X chromosome, mitochondria (i.e. mt haplogroup U), and Y chromosome (i.e. Y chromosome haplogroups I2, R1 and C2). Bars show approximate 95% confidence intervals.

741 Extended Data Figures



742

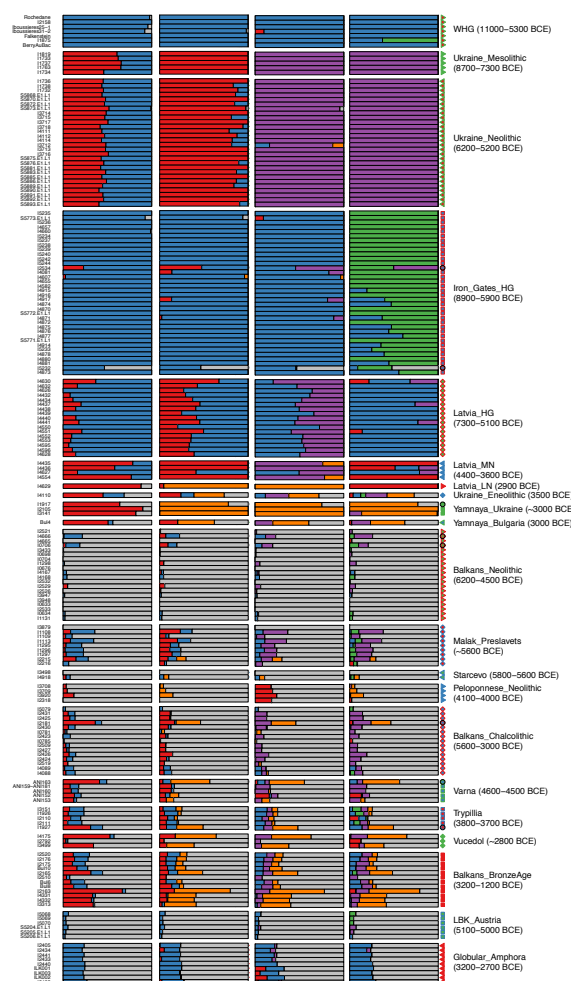


743

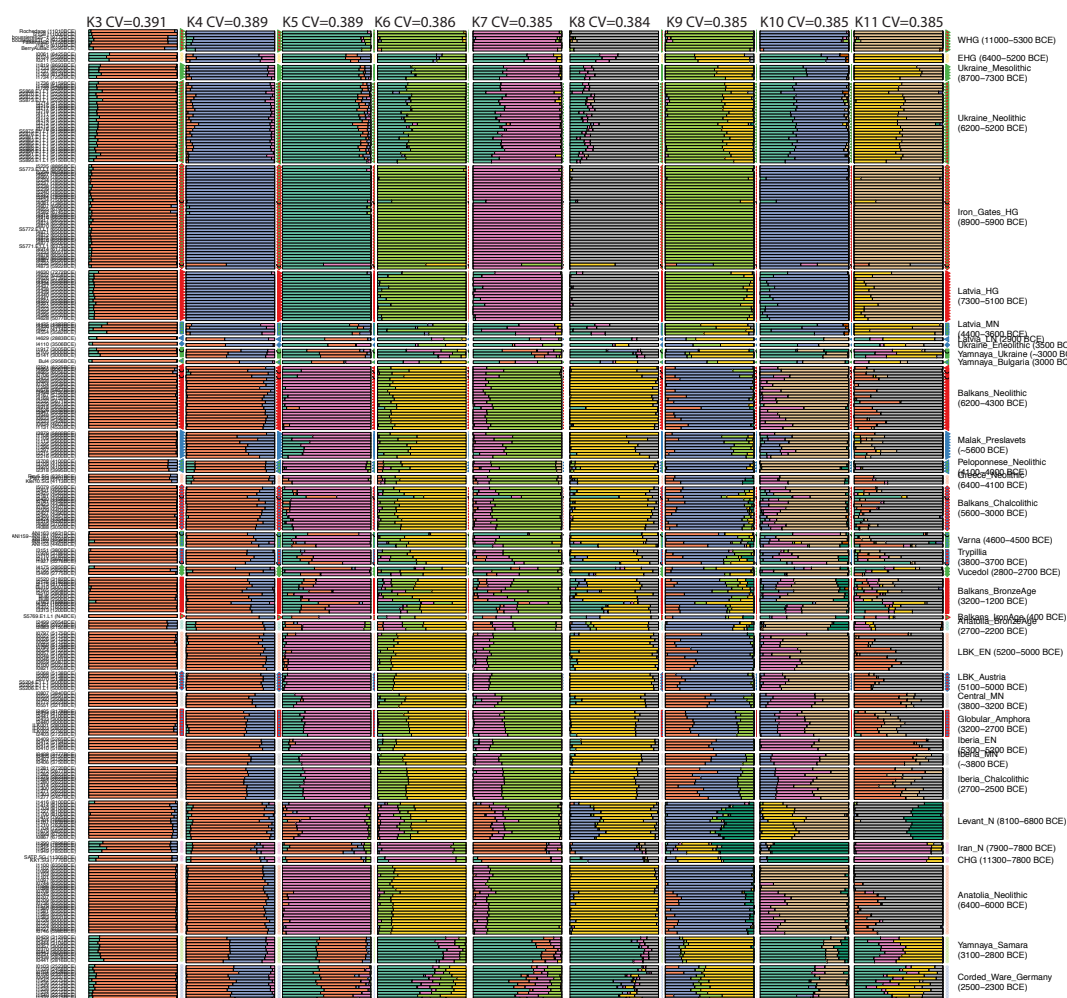
744

745 **Extended Data Figure 1: PCA of all ancient individuals, projected onto principal**
 746 **components defined by 799 present-day West Eurasian individuals. (This differs from Figure**
 747 **1B in that the plot is not cropped and the present-day individuals are shown.)**

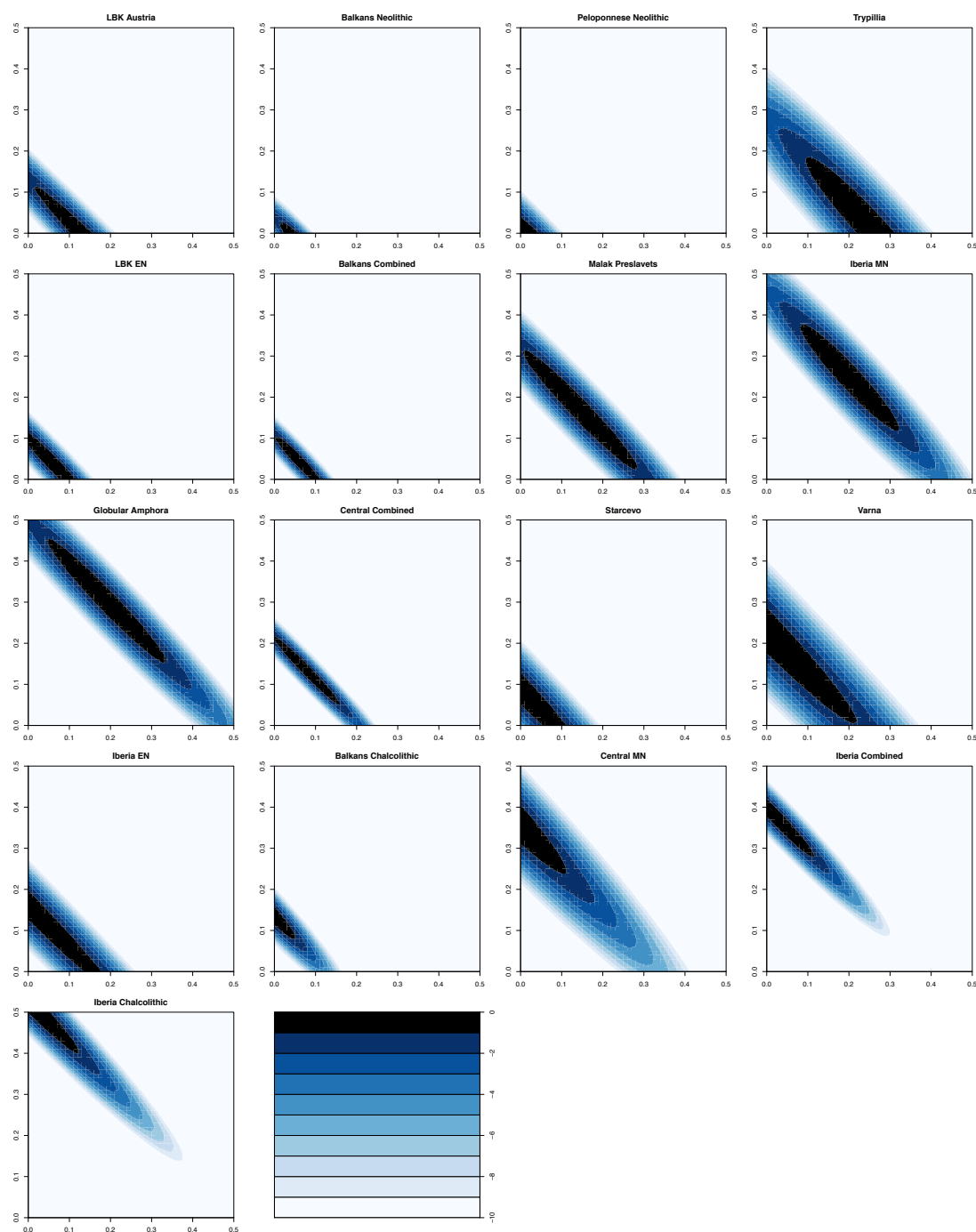
748



749
 750 **Extended Data Figure 2:** Supervised admixture plot for K=3 to 6. We ran on 357 ancient
 751 individuals and show newly reported ancient individuals with some previously published
 752 ancient individuals for comparison. K=4 is shown in Figure 1D. Clusters constrained to
 753 include individuals from Anatolia_Neolithic (grey), WHG (blue), EHG (red),
 754 Yamnaya_Samara (orange), Ukraine_Mesolithic (purple), Motala_HG (green).
 755



Extended Data Figure 3: Unsupervised ADMIXTURE plot from k=3 to 11, on a dataset consisting of 799 present-day individuals and 405 ancient individuals. We show newly reported ancient individuals and some previously published individuals for comparison.



Extended Data Figure 4: log-likelihood surfaces for the proportion of female (x-axis) and male (y-axis) ancestors that are hunter-gatherers for the populations analysed in Figure 3. 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.

References

- 1 Tringham, R. E. Southeastern Europe in the Transition to Agriculture in Europe: Bridge, Buffer or Mosaic. 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* (Wiley-Blackwell, 2004).
- 3 Golitko, M. First farmers of Central Europe. 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. To Tame a Land: Archaeological Cultures and the Spread of the Neolithic in Western Europe. 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. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 368-373 (2016).
- 9 Hofmanova, Z. *et al.* Early farmers from across Europe directly descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences of the United States of America* **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. *Current Biology* **26**, 270-275 (2016).
- 12 Müller, J., Rassmann, K. & Videiko, M. *Trypillia Mega-Sites and European Prehistory: 4100–3400 BCE*. (Routledge, 2016).
- 13 Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* (2015).
- 14 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655-1664 (2009).
- 15 Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093 (2012).
- 16 Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167-172 (2015).
- 17 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449 (2014).

816 18 Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205 (2016).
817 19 Gallego Llorente, M. *et al.* Ancient Ethiopian genome reveals extensive Eurasian
818 admixture in Eastern Africa. *Science* **350**, 820-822 (2015).
819 20 Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians.
820 *Nat Commun* **6**, 8912 (2015).
821 21 Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as
822 inferred by whole-genome sequencing. *Nat Commun* **3**, 698 (2012).
823 22 Kilinc, G. M. *et al.* The Demographic Development of the First Farmers in Anatolia.
824 *Current Biology* (2016).
825 23 Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near
826 East. *Nature* **536**, 419-424 (2016).
827 24 Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for
828 present-day Europeans. *Nature* **513**, 409-413 (2014).
829 25 Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-
830 old Mesolithic European. *Nature* **507**, 225-228 (2014).
831 26 Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of
832 Native Americans. *Nature* **505**, 87-91 (2014).
833 27 Lazaridis, I. *et al.* Manuscript in revision. (2017).
834 28 Lipson, M. *et al.* Parallel ancient genomic transects reveal complex population
835 history of early European farmers. *bioRxiv*, <https://doi.org/10.1101/114488> (2017).
836 29 Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142
837 diverse populations. *Nature* (2016).
838 30 Gronenborn, D. & Dolukhanov, P. Early Neolithic Manifestations in Central and
839 Eastern Europe. In: *The Oxford Handbook of Neolithic Europe* (eds C. Fowler, J.
840 Harding, & D. Hofmann) (Oxford University Press, 2015).
841 31 Telegin, D. Y. Neolithic cultures of Ukraine and their chronology. *Journal of World*
842 *Prehistory* **I**, 307-331 (1987).
843 32 Telegin, D. Y. & Potekhina, I. D. *Neolithic cemeteries and populations in the*
844 *Dnieper Basin*. . BAR International Series 383 (British Archaeological Reports,
845 1987).
846 33 Petkova, D., Novembre, J. & Stephens, M. Visualizing spatial population structure
847 with estimated effective migration surfaces. *Nature Genetics* **48**, 94-100 (2016).
848 34 Jones, E. R. *et al.* The Neolithic Transition in the Baltic Was Not Driven by
849 Admixture with Early European Farmers. *Current Biology* (2017).
850 35 Broushaki, F. *et al.* Early Neolithic genomes from the eastern Fertile Crescent.
851 *Science* **353**, 499-503 (2016).
852 36 Mitnik, A. *et al.* The Genetic History of Northern Europe. *bioRxiv*,
853 <https://doi.org/10.1101/113241> (2017).
854 37 Saag, L. *et al.* Extensive farming in Estonia started through a sex-biased migration
855 from the Steppe. *bioRxiv* (2017).
856 38 Maier, A. *The Central European Magdalenian: Regional Diversity and Internal*
857 *Variability*. (Springer, 2015).
858 39 Borić, D. & Price, T. D. Strontium isotopes document greater human mobility at the
859 start of the Balkan Neolithic. *PNAS* **110**, 3298-3303 (2013).
860 40 Bacvarov, K. Malak Preslavets Revisited: The Early Neolithic Burials. In: *Moments*
861 *in time: Papers Presented to Pál Raczky on His 60th Birthday* (eds A. Anders & G
862 Kulcsár) 29-34 (L'Harmattan, 2013).
863 41 Gurova, M. & Bonsall, C. 'Pre-Neolithic' in Southeast Europe: a Bulgarian
864 perspective. *Documenta Praehistorica* **XLI**, 95-109 (2014).
865 42 Borić, D. Mortuary practices, bodies and persons in the Neolithic and Early-Middle
866 Copper Age of southeast Europe. In: *The Oxford Handbook of Neolithic Europe* (eds
867 C. Fowler, J. Harding, & D. Hofmann) 927-957 (Oxford University Press, 2015).
868 43 Brandt, G. *et al.* Ancient DNA reveals key stages in the formation of central
869 European mitochondrial genetic diversity. *Science* **342**, 257-261 (2013).

870 44 Olalde, I. *et al.* A Common Genetic Origin for Early Farmers from Mediterranean
871 Cardial and Central European LBK Cultures. *Molecular Biology and Evolution* **32**,
872 3132-3142 (2015).

873 45 Krauß, R., Marinova, E., De Brue, H. & Weninger, B. The rapid spread of early
874 farming from the Aegean into the Balkans via the Sub-Mediterranean-Aegean
875 Vegetation Zone. *Quaternary International* **XXX**, 1-18 (2017).

876 46 Perlès, C. *The early Neolithic in Greece* (Cambridge University Press, 2001).

877 47 Posth, C. *et al.* Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal
878 of Non-Africans and a Late Glacial Population Turnover in Europe. *Current Biology*
879 **26**, 827-833 (2016).

880 48 Skoglund, P. *et al.* Genomic diversity and admixture differs for Stone-Age
881 Scandinavian foragers and farmers. *Science* **344**, 747-750 (2014).

882 49 Rowley-Conwy, P. Westward Ho! The Spread of Agriculturalism from Central
883 Europe to the Atlantic. *Current Anthropology* **52**, S431-S451 (2011).

884 50 Bollongino, R. *et al.* 2000 years of parallel societies in Stone Age Central Europe.
885 *Science* **342**, 479-481 (2013).

886 51 Anthony, D. W. & Ringe, D. The Indo-European Homeland from Linguistic and
887 Archaeological Perspectives. *Annual Review of Linguistics* **1**, 199-219 (2015).

888 52 Anthony, D. W. *The horse the wheel and language* (Princeton University Press,
889 2007).

890 53 Goldberg, A., Gunther, T., Rosenberg, N. A. & Jakobsson, M. Ancient X
891 chromosomes reveal contrasting sex bias in Neolithic and Bronze Age Eurasian
892 migrations. *Proceedings of the National Academy of Sciences of the United States of*
893 *America* **114**, 2657-2662 (2017).

894 54 Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene
895 cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National*
896 *Academy of Sciences of the United States of America* **110**, 15758-15763 (2013).

897 55 Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly
898 multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**, pdb
899 prot5448 (2010).

900 56 Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-
901 glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of*
902 *the Royal Society of London. Series B: Biological Sciences* **370**, 20130624 (2015).

903 57 Korlevic, P. *et al.* Reducing microbial and human contamination in DNA extractions
904 from ancient bones and teeth. *BioTechniques* **59**, 87-93 (2015).

905 58 Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo
906 methylation in ancient DNA. *Nucleic Acids Research* **38**, e87 (2010).

907 59 DeAngelis, M. M., Wang, D. G. & Hawkins, T. L. Solid-phase reversible
908 immobilization for the isolation of PCR products. *Nucleic Acids Research* **23**, 4742-
909 4743 (1995).

910 60 Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries
911 for multiplexed target capture. *Genome Research* **22**, 939-946 (2012).

912 61 Maricic, T., Whitten, M. & Paabo, S. Multiplexed DNA sequence capture of
913 mitochondrial genomes using PCR products. *PloS One* **5**, e14004 (2010).

914 62 Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in
915 multiplex sequencing on the Illumina platform. *Nucleic Acids Research* **40**, e3
916 (2012).

917 63 Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial
918 genomes. *Current Biology* **23**, 553-559 (2013).

919 64 Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day
920 contamination in a Siberian Neandertal. *Proceedings of the National Academy of*
921 *Sciences of the United States of America* **111**, 2229-2234 (2014).

922 65 Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next
923 Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).

924 66 Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-
925 wide association studies. *Nature Genetics* **38**, 904-909 (2006).
926 67 Brown, L., Cai, T. & DasGupta, A. Interval Estimation for a Binomial Proportion.
927 *Statistical Science* **16**, 101-133 (2001).
928 68 Agresti, A. & Coull, B. Approximate is better than "exact" for interval estimation of
929 binomial proportions. *The American Statistician* **52**, 119-126 (1998).
930 69 Longin, R. New method of collagen extraction for radiocarbon analysis. *Nature* **230**,
931 241-242 (1971).
932 70 Brown, T. A., Nelson, D. E., Vogel, J. S. & Southon, J. R. Improved collagen
933 extraction by modified longin method. *Radiocarbon* **30**, 171-177 (1988).
934 71 van Klinken, G. J. Bone collagen quality indicators for paleodietary and radiocarbon
935 measurements. *Journal of Archaeological Science* **26**, 687-695 (1999).
936 72 Stuiver, M. & Polach, H. A. Discussion: Reporting of ¹⁴C data. *Radiocarbon* **19**,
937 355-363 (1977).
938 73 Bronk Ramsey, C. *OxCal 4.23 Online Manual*,
939 <https://c14.arch.ox.ac.uk/oxcalhelp/hlp_contents.html> (2013).
940