ARTICLE

doi:10.1038/nature25778

The genomic history of southeastern Europe

A list of authors and affiliations appears at the end of the paper.

Farming was first introduced to Europe in the mid-seventh millennium BC, and was associated with migrants from Anatolia who settled in the southeast before spreading throughout Europe. Here, to understand the dynamics of this process, we analysed genome-wide ancient DNA data from 225 individuals who lived in southeastern Europe and surrounding regions between 12000 and 500 BC. We document a west–east cline of ancestry in indigenous hunter-gatherers and, in eastern Europe, the early stages in the formation of Bronze Age steppe ancestry. We show that the first farmers of northern and western Europe dispersed through southeastern Europe with limited hunter-gatherer admixture, but that some early groups in the southeast mixed extensively with hunter-gatherers without the sex-biased admixture that prevailed later in the north and west. We also show that southeastern Europe continued to be a nexus between east and west after the arrival of farmers, with intermittent genetic contact with steppe populations occurring up to 2,000 years earlier than the migrations from the steppe that ultimately replaced much of the population of northern Europe.

Southeastern 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 Europe in the mid-seventh millennium BC[1,2], farming spread westward along a Mediterranean route, and northwestward via a Danubian route, and was established in Iberia and central Europe by 5600 BC[3,4]. Previous studies have shown that the spread of farming across Europe was accompanied by a massive movement of people[5–8] who were closely related to the farmers of northwestern Anatolia[9–11], but nearly all the ancient DNA on which these studies are based derives from first farmers in central and western Europe, with only three individuals reported from the southeast[9]. In the two millennia after the establishment of agriculture in the Balkan Peninsula a series of complex societies formed, which culminated in sites such as the mid-fifth millennium BC necropolis at Varna. The Varna necropolis has some of the earliest evidence for extreme inequality in wealth; one individual there (grave 43), from whom we extracted DNA, was buried with more gold than is known from all other Neolithic and Copper Age burials, combined. By the end of the sixth millennium BC, agriculture had reached eastern Europe, where it is associated with the Cucuteni–Trypillian complex in the area of present-day Moldova, Romania and Ukraine; this complex was characterized by 'mega-sites' that housed hundreds, or perhaps even thousands, of people[12]. After around 4000 BC, 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 previously been unknown, owing to the lack of ancient DNA evidence.

Results

We generated genome-wide data from 225 ancient humans dated to 12000–500 BC, from the Balkan Peninsula, the Carpathian Basin, the North Pontic steppe and neighbouring regions (Fig. 1, Supplementary Table 1, Supplementary Note 1); 215 of these individuals are reported here for the first time. 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), and 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 those that had unexpected ancestry for their archaeological context and were not directly dated. We report, but do not analyse, nine individuals that were first-degree relatives of others in the dataset, resulting in an analysis dataset of 216 individuals. We analysed these data together with data from 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[28]. We used principal component analysis (Fig. 1b, Extended Data Fig. 1), supervised and unsupervised ADMIXTURE[29] (Fig. 1d, Extended Data Figs 2, 3), *D* statistics, qpAdm and qpGraph[30], along with archaeological and chronological information (including 137 newly reported accelerator mass spectrometry (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 that can be further subdivided into 'eastern' (EHG) and 'western' (WHG) hunter-gatherer-related ancestry[7]; second, 'northwestern-Anatolian-Neolithic-related' ancestry related to the Neolithic farmers of northwest Anatolia and tightly linked to the appearance of agriculture[9,10]; and third, 'steppe-related' ancestry that appears in western Europe during the Late Neolithic-to-Bronze Age transition, and which is ultimately derived from a population related to Yamnaya steppe pastoralists[7,15]. Steppe-related ancestry itself can be modelled 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

Of the 215 new individuals that we report, 105 from Palaeolithic, Mesolithic and eastern European Neolithic contexts (unlike in western Europe, the eastern European Neolithic refers to the presence

A list of authors and affiliations appears at the end of the paper.

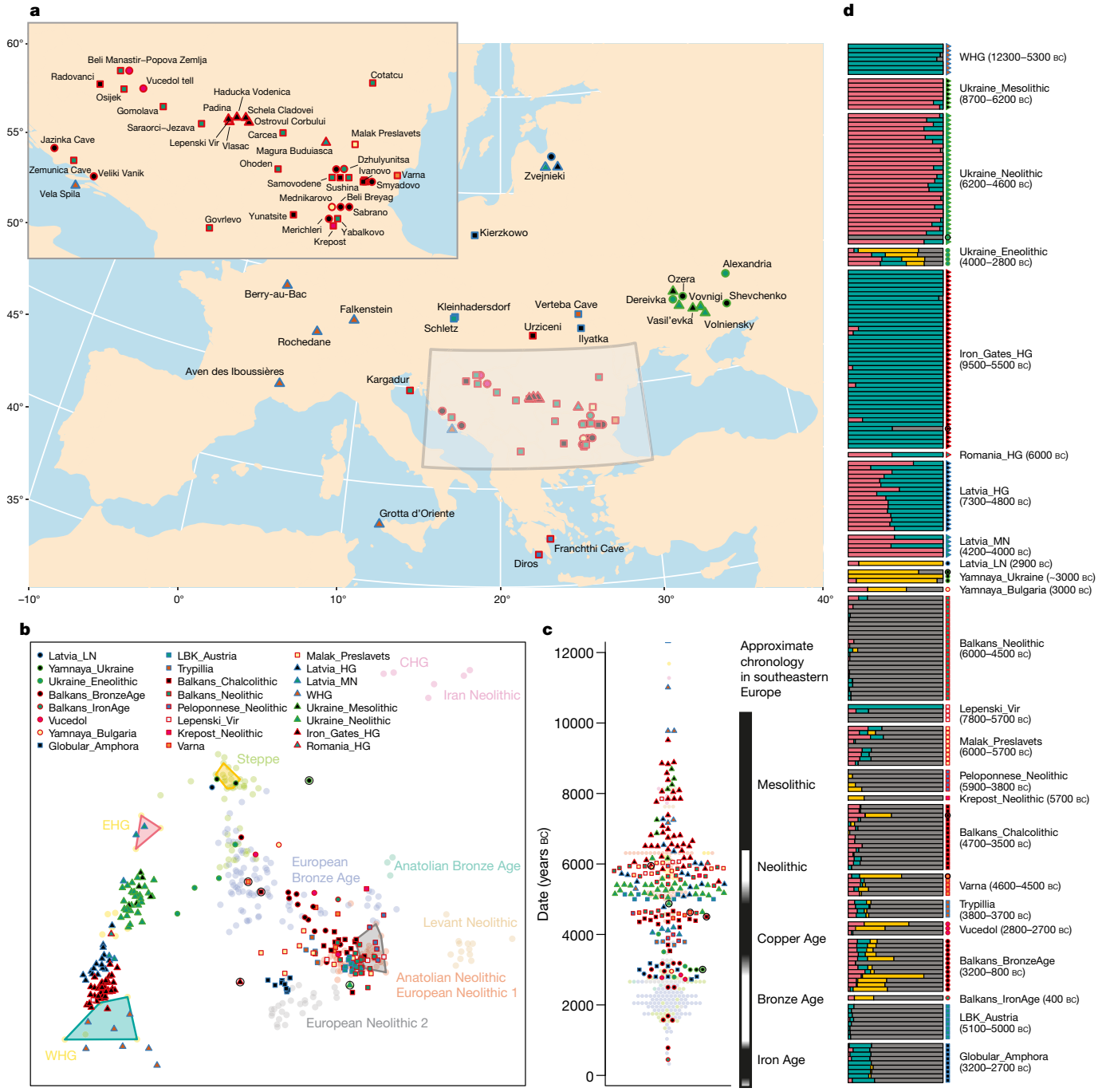


Figure 1 | Geographic and genetic structure of 216 analysed individuals. a, Locations of newly reported individuals. **b**, Ancient individuals projected onto principal components defined by 777 present-day west Eurasians (shown in Extended Data Fig. 1); data include selected published individuals (faded circles, labelled) and newly reported individuals (other symbols, outliers enclosed in black circles). Coloured polygons cover individuals that had cluster memberships fixed at 100% for supervised ADMIXTURE analysis. **c**, Direct or contextual dates for each sample and approximate chronology of southeastern Europe. **d**, Supervised ADMIXTURE analysis, modelling each ancient individual (one per row) as a mixture of population clusters constrained to contain northwestern-Anatolian Neolithic (grey), Yamnaya from Samara (yellow), EHG (pink) and WHG (green) populations. Dates in parentheses indicate approximate range of individuals in each population. See Extended Data Fig. 2 for individual sample identification numbers. Map data in **a** from the R package 'maps'.

of pottery[31–33], not necessarily to farming) have almost entirely hunter-gatherer-related ancestry. These individuals form a cline from WHG to EHG that is correlated with geography (Fig. 1b)—although it is neither geographically nor temporally uniform (Fig. 2, Extended Data Fig. 4)—and contains substructure in phenotypically important variants (Supplementary Note 2).

From present-day Ukraine, our study reports genome-wide data from 7 Mesolithic (approximately 9500–6000 BC) and 30 Neolithic (approximately 6000–3500 BC) individuals. On the cline from WHG-to-EHG-related ancestry, the Mesolithic individuals fall towards the east, intermediate between EHG and Mesolithic hunter-gatherers from Scandinavia[7] (Fig. 1b). The Neolithic population has a significant difference in ancestry compared to the Mesolithic (Figs 1b, 2), with a shift towards WHG shown by the statistic D(Mbuti, WHG, Ukraine_Mesolithic, Ukraine_Neolithic); $Z = 8.5$ (Supplementary Table 2). Unexpectedly, one Neolithic individual from Dereivka (I3719)

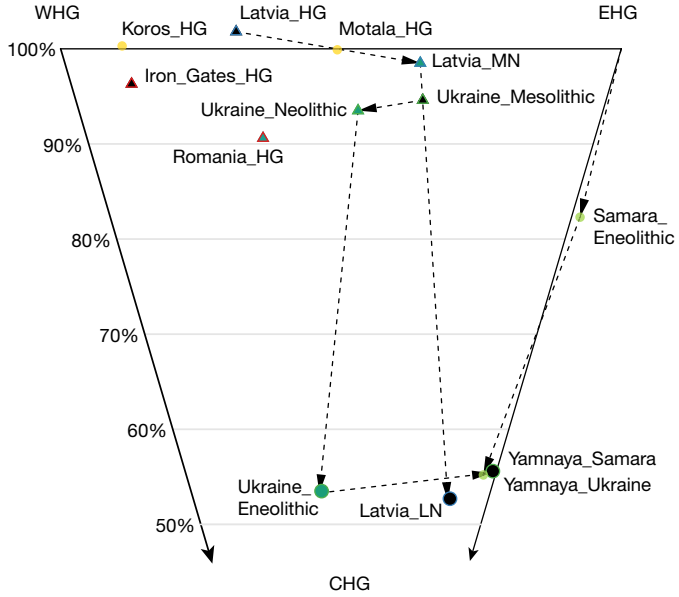


Figure 2 | Structure and change in hunter-gatherer-related populations. Inferred ancestry proportions for populations modelled as a mixture of WHG, EHG and CHG (Supplementary Table 3.1.3). Dashed lines show populations from the same geographic region. Percentages indicate proportion of WHG + EHG ancestry. Standard errors range from 1.5 to 8.3% (Supplementary Table 3.1.3).

that we directly dated to 4949–4799 BC has entirely northwestern-Anatolian-Neolithic-related ancestry.

The pastoralist Bronze Age Yamnaya complex originated on the Eurasian steppe and is a plausible source for the dispersal of steppe-related ancestry into central and western Europe from around 2500 BC[13]. All previously reported Yamnaya individuals were from Samara[7] and Kalmykia[15] in southwest Russia, and had entirely steppe-related ancestry. Here, we report three Yamnaya individuals from further west, in Ukraine and Bulgaria, and show that although they all have high levels of steppe-related ancestry, one individual from Ozera in Ukraine and one from Bulgaria (I1917 and Bul4, respectively, both dated to approximately 3000 BC) have northwestern-Anatolian-Neolithic-related admixture, the first evidence of such ancestry in Yamnaya-associated individuals (Fig. 1b, d, Supplementary Table 2). Preceding the Yamnaya complex, four Copper Age individuals (I4110, I5882, I5884 and I6561; labelled as 'Ukraine_Eneolithic') from Dereivka and Alexandria dated to approximately 3600–3400 BC have a mixture of hunter-gatherer-, steppe- and northwestern-Anatolian-Neolithic-related ancestry (Fig. 1d, Supplementary Table 2).

At Zvejnieki in Latvia, using 17 newly reported individuals and additional data for 5 previously reported[34] individuals, we observe a transition in hunter-gatherer-related ancestry that is opposite to that seen in Ukraine. We find that Mesolithic and Early Neolithic individuals (labelled 'Latvia_HG') associated with the Kunda and Narva cultures have ancestry that is intermediate between WHG (approximately 70%) and EHG (approximately 30%), consistent with previous reports[34–36] (Supplementary Table 3). We also detect a shift in ancestry between Early Neolithic individuals and those associated with the Middle Neolithic Comb Ware complex (labelled 'Latvia_MN'), who have more EHG-related ancestry; we estimate that the ancestry of Latvia_MN individuals comprises 65% EHG-related ancestry, but two of the four individuals appear to be 100% EHG in principal component space (Fig. 1b). The most recent individual, associated with the Final Neolithic Corded Ware complex (I4629, labelled 'Latvia_LN'), attests to another ancestry shift, clustering closely with Yamnaya from Samara[7], Kalmykia[15] and Ukraine (Fig. 2).

We report Upper Palaeolithic and Mesolithic data from southern and western Europe[17]. Sicilian (I2158) and Croatian (I1875) individuals dating to approximately 12000 and 6100 BC cluster with previously reported WHG (Fig. 1b, d), including individuals from Loschbour[23] (Luxembourg, 6100 BC), Bichon[19] (Switzerland, 11700 BC), and Villabruna[17] (Italy, 12000 BC). These results demonstrate that, for at least six thousand years, WHG populations[23] were widely distributed from the Atlantic seaboard of Europe in the west, to Sicily in the south, and to the Balkan Peninsula in the southeast.

A particularly important hunter-gatherer population that we report is from the Iron Gates region, which straddles the border of present-day Romania and Serbia. This population (labelled 'Iron_Gates_HG') is represented in our study by 40 individuals from 5 sites. We modelled Iron Gates hunter-gatherers as a mixture of WHG- and EHG-related ancestry (Supplementary Table 3), which showed that they are intermediate between the two (WHG contributing approximately 85%, and EHG approximately 15%, of ancestry). However, this qpAdm model does not fit well ($P = 0.0003$, Supplementary Table 3) and the Iron Gates hunter-gatherers show an affinity towards northwestern-Anatolian-Neolithic-, relative to WHG-, ancestry populations (Supplementary Table 2). In addition, Iron Gates hunter-gatherers carry mitochondrial haplogroup K1 (7/40) as well as other subclades of haplogroups U (32/40) and H (1/40), in contrast to WHG, EHG and Scandinavian hunter-gatherers—almost all of whom carry haplogroups U5 or U2. One interpretation is that the Iron Gates hunter-gatherers have ancestry that is not present in either WHG or EHG. Possible scenarios include genetic contact between the ancestors of the Iron Gates population and a northwestern-Anatolian-Neolithic-related population, or that the Iron Gates population is related to the source population from which the WHG split during a re-expansion into Europe from the southeast after the Last Glacial Maximum[17,37].

A notable finding from the Iron Gates concerns the four individuals from the site of Lepenski Vir, two of whom (I4665 and I5405, 6200–5600 BC), have entirely northwestern-Anatolian-Neolithic-related ancestry. Strontium and nitrogen isotope data[38] indicate that both these individuals were migrants from outside the Iron Gates region and ate a primarily terrestrial diet (Supplementary Information section 1). A third individual (I4666, 6070 BC) has a mixture of northwestern-Anatolian-Neolithic-related and hunter-gatherer-related ancestry and consumed aquatic foods, and a fourth and probably earlier individual (I5407) had entirely hunter-gatherer-related ancestry (Fig. 1d, Supplementary Information section 1). We also identify one individual from Padina (I5232), dated to 5950 BC, who had a mixture of northwestern-Anatolian-Neolithic-related and hunter-gatherer-related ancestry. These results provide genetic confirmation that the Iron Gates was a region of interaction between groups distinct in both ancestry and subsistence strategy.

Population transformations in the first farmers

Neolithic populations from present-day Bulgaria, Croatia, the former Yugoslav Republic of Macedonia, Serbia and Romania cluster closely with the northwestern-Anatolian-Neolithic individuals (Fig. 1), consistent with archaeological evidence[39]. Modelling Balkan Neolithic populations as a mixture of northwestern-Anatolian-Neolithic and WHG, we estimate that 98% (95% confidence interval; 97–100%) of their ancestry is northwestern-Anatolian-Neolithic-related. A notable exception is evident in eight out of nine individuals from Malak Preslavets in present-day Bulgaria[40]. These individuals lived in the mid-sixth millennium BC and have significantly more hunter-gatherer-related ancestry than other Balkan Neolithic populations (Fig. 1b, d, Extended Data Figs 1–3, Supplementary Tables 2–4); a model of 82% (confidence interval: 77–86%) northwestern-Anatolian-Neolithic-related, 15% (confidence interval: 12–17%) WHG-related and 4% (confidence interval: 0–9%) EHG-related ancestry fits the data. This hunter-gatherer-related ancestry, with an approximately 4:1 WHG:EHG ratio, plausibly represents a contribution from local Balkan hunter-gatherers genetically similar to those of the Iron Gates region. Late Mesolithic hunter-gatherers in the Balkans were probably

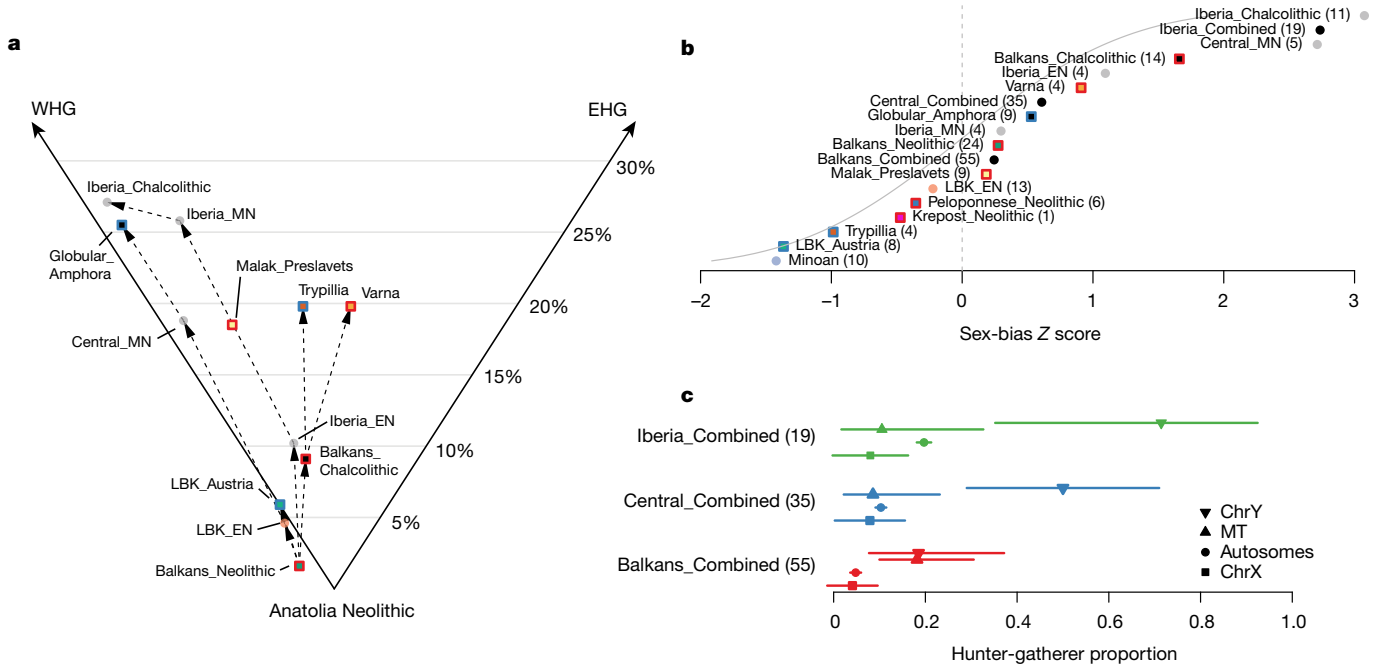


Figure 3 | Structure and change in northwestern-Anatolian-Neolithic-related populations. a, Populations modelled as a mixture of northwestern Anatolia Neolithic, WHG and EHG. Dashed lines show temporal relationships between populations from the same geographic region. Percentages indicate proportion of WHG + EHG ancestry. Standard errors range from 0.7 to 6.0% (Supplementary Table 3.2.2). **b**, *Z* scores for the difference between hunter-gatherer-related ancestry on the autosomes and that on the X chromosome, when populations are modelled as a mixture of northwestern Anatolia Neolithic and WHG ($n = 126$ individuals, group sizes in parentheses). Positive values indicate more hunter-gatherer-related ancestry on the autosomes, and thus male-biased hunter-gatherer ancestry. 'Combined' populations merge all individuals from different times from a geographic area. **c**, Hunter-gatherer-related ancestry proportions on the autosomes, X chromosome (ChrX), mitochondrial DNA (MT; mitochondrial haplogroup U), and the Y chromosome (ChrY; Y chromosome haplogroups I2, R1 and C1). Points show qpAdm (autosomes and X chromosome) or maximum likelihood (mitochondrial DNA and Y chromosome) estimates and bars show approximate 95% confidence intervals ($n = 109$ individuals, group sizes in parentheses).

concentrated along the coast and major rivers such as the Danube[41], which directly connects the Iron Gates with Malak Preslavets. Early farmer groups with the highest percentages of hunter-gatherer-related ancestry may, therefore, have been those that lived close to the highest densities of hunter-gatherers.

In the Balkans, Copper Age populations (labelled 'Balkans_Chalcolithic') contain significantly more hunter-gatherer-related ancestry than Neolithic populations as shown, for example, by the statistic *D*(Mbuti, WHG, Balkans_Neolithic, Balkans_Chalcolithic); $Z = 4.3$ (Supplementary Table 2). This is roughly contemporary with the 'resurgence' of hunter-gatherer ancestry previously reported in central Europe and Iberia[7,10,42] and is consistent with changes in funeral rites, specifically the reappearance at around 4500 BC of the Mesolithic tradition of extended supine burial in contrast to the Early Neolithic tradition of flexed burial[43]. Four individuals associated with the Copper Age Trypillian population have approximately 80% northwestern-Anatolian-Neolithic-related ancestry (Supplementary Table 3), confirming that the ancestry of the first farmers of present-day Ukraine was largely derived from the same source as the farmers of Anatolia and western Europe. The roughly 20% of their ancestry from hunter-gatherers is intermediate between WHG and EHG, consistent with it deriving from the Neolithic hunter-gatherers of the region.

We also report 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 (Fig. 1b, d). Both groups of Globular Amphora complex samples had more hunter-gatherer-related ancestry than did Middle Neolithic groups from central Europe[7]; we estimate that the Globular Amphora individuals harboured 25% (confidence interval: 22–27%) WHG-related ancestry, similar to the level seen in Chalcolithic Iberian individuals (Supplementary Table 3). In east-central Europe, the Globular Amphora complex preceded the Corded Ware complex that marks the appearance of steppe-related ancestry[7,15], and in southeastern Europe, the Globular Amphora complex bordered populations with steppe-influenced material cultures for hundreds of years.[44] Despite this, the Globular Amphora complex individuals in our study show no evidence of steppe-related ancestry, supporting the hypothesis that this material cultural frontier was also a barrier to gene flow.

About 75% of the ancestry of individuals associated with the Corded Ware complex—and, in central Europe, about 50% of the ancestry of people associated with succeeding archaeological cultures such as the Bell Beaker complex[7,15]—can be traced to a population that probably moved from the Pontic–Caspian steppe in the third millennium BC and that had ancestry similar to that of individuals linked with the Yamnaya complex. However, in two directly dated individuals from southeastern Europe, we find far-earlier evidence of steppe-related ancestry (Fig. 1b, d). One (ANI163) from the Varna I cemetery was dated to 4711–4550 BC and another (I2181) from nearby Smyadovo was dated to 4550–4450 BC. These findings push back the first evidence for steppe-related ancestry this far west in Europe by almost 2,000 years, but it must have been sporadic because other Copper Age (approximately 5000–4000 BC) individuals from the Balkans have no evidence for such ancestry. Bronze Age (approximately 3400–1100 BC) individuals do have steppe-related ancestry: we estimate that they have about 30% (confidence interval: 26–35%), with the highest proportions in the four latest Balkan Bronze Age individuals in our data (later than roughly 1700 BC) and the least in earlier Bronze Age individuals (3400–2500 BC; Fig. 1d).

Two sources of Asian 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 and

southern Europe were derived from a single source population or instead derived from multiple sources. 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 complex from central Europe[7,45], and that both Mediterranean and Danubian populations are closely related to the Balkan Neolithic population. These three populations form a clade with the northwestern-Anatolian Neolithic individuals as an outgroup, consistent with a single migration into the Balkan Peninsula that then split into two (Supplementary Note 3).

By contrast, data from five southern Greek Neolithic individuals (labelled 'Peloponnese_Neolithic')—three (plus one that has previously been published[26]) from Diros Cave and one from Franchthi Cave—are not consistent with descent from the same source population as other European farmers. *D* statistics (Supplementary Table 2) show that these 'Peloponnese Neolithic' individuals, dated to around 4000 BC, are shifted away from WHG, and towards CHG, relative to northwestern-Anatolian Neolithic and Balkan Neolithic individuals. We detect the same pattern in a single Neolithic individual from Krepost in present-day Bulgaria (I0679_d, 5718–5626 BC). An even more marked shift towards CHG has previously been observed in individuals associated with the Bronze Age Minoan and Mycenaean cultures[26], suggesting gene flow into the region from populations with CHG-rich ancestry throughout the Neolithic, Chalcolithic and Bronze Age. Possible sources are from people related to the Neolithic population of the central Anatolian site of Tepecik Çiftlik[21], or the Aegean site of Kumtepe[11], who are also shifted towards CHG relative to northwestern-Anatolian Neolithic samples, as are later Copper and Bronze Age Anatolians[10,26].

Sex bias in Neolithic hunter–gatherer admixture

We provide evidence for sex-biased admixture between hunter-gatherers and farmers in Europe, showing that the Middle Neolithic resurgence of hunter-gatherer-related ancestry[7,42] in central Europe and Iberia was driven more by males than by females (Fig. 3b, c, Extended Data Fig. 5, Supplementary Table 5). To document this, we used qpAdm to compute ancestry proportions on the autosomes and the X chromosome; because males always inherit a maternal X chromosome, differences in these proportions imply sex-biased mixture. There is no evidence of sex bias in the Balkan Neolithic ($Z = 0.27$; a positive Z score implies bias towards male hunter-gatherer ancestry) or in the Linearbankeramik and Iberian Early Neolithic ($Z = -0.22$ and 1.09). In the Copper Age there is clear bias towards male hunter-gatherer ancestry that is weak in the Balkans ($Z = 1.66$), but stronger in Iberia ($Z = 3.08$) and central Europe ($Z = 2.74$). Consistent with this, hunter-gatherer mitochondrial haplogroups (haplogroup U)[46] are rare and within the intervals of genome-wide ancestry proportions, but hunter-gatherer-associated Y chromosomes[17] are more common (Fig. 3c): seven out of nine male individuals in the Iberian Neolithic and Copper Age and nine out of ten male individuals in Middle–Late Neolithic central Europe ('Central_MN' and 'Globular_Amphora') carried haplogroups I2, R1 or C1.

No steppe migration to Anatolia via southeast Europe

One version of the steppe hypothesis of Indo-European language origins suggests that Proto-Indo-European languages developed north of the Black and Caspian seas, and that the earliest-known diverging branch, the Anatolian branch, was spread into Asia Minor by the movements of steppe peoples through the Balkan Peninsula during the Copper Age at around 4000 BC[47]. If this were correct, then one way to detect evidence of the spread of Indo-European languages would be the appearance of large amounts of steppe-related ancestry first in the Balkan Peninsula, and later in Anatolia. However, our data provide no evidence for this scenario. Although we find sporadic steppe-related ancestry in Balkan Copper and Bronze Age individuals, this ancestry is rare until the late Bronze Age. Furthermore, although Bronze Age Anatolian individuals have CHG-related ancestry[26], they do not have the EHG-related ancestry characteristic of all steppe populations sampled to date[19] or the WHG-related ancestry that is ubiquitous in Neolithic southeastern Europe (Extended Data Figs 2, 3, Supplementary Table 2). We caution, however, that at present we only have data from a small number of Bronze Age Anatolian individuals, none of whom are associated with known Indo-European-speaking populations. An alternative hypothesis is that the homeland of Proto-Indo-European languages was in the Caucasus or in Iran. In this scenario, westward population movement contributed to the dispersal of Anatolian languages, and northward movement and mixture with EHG was responsible for the formation of a 'Late Proto-Indo European'-speaking population associated with the Yamnaya complex[13]. Although this scenario gains plausibility from our results, it remains possible that Indo-European languages were spread through southeastern Europe into Anatolia without large-scale population movement or admixture.

Discussion

Our study shows that southeastern Europe served as a genetic contact zone between east and west over thousands of years. Before the arrival of farming, the region saw interaction between diverged groups of hunter-gatherers, and this interaction continued after farming arrived. Although this study has clarified the genomic history of the region from the Mesolithic to the Bronze Age, the processes that connected these populations to ones living today remain largely unknown. An important priority for future research should be to sample populations from the Bronze Age, Iron Age, Roman and Medieval periods and to compare them to present-day populations to understand how these transitions occurred.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 May 2017; accepted 16 January 2018.
Published online 21 February 2018.

1. Tringham, R. E. in *The Transition to Agriculture in Prehistoric Europe* (ed. Price, D.) 19–56 (Cambridge Univ. 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 Bogucki, P. & Crabtree, P. J.) 259–266 (Charles Scribners & Sons, 2003).
4. VanderLinden, M. in *Investigating Archaeological Cultures: Material Culture, Variability, and Transmission* (eds Roberts, B. W. & VanderLinden, M.) 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. USA* **113,** 368–373 (2016).
9. Hofmanová, Z. *et al.* Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl Acad. Sci. USA* **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 Univ. Press, 2007).
14. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524,** 216–219 (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. *Nat. Commun.* **6,** 8912 (2015).
20. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* **3,** 698 (2012).
21. Kılınç, G. M. *et al.* The demographic development of the first farmers in Anatolia. *Curr. Biol.* **26,** 2659–2666 (2016).
22. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536,** 419–424 (2016).
23. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,** 409–413 (2014).
24. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507,** 225–228 (2014).
25. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505,** 87–91 (2014).
26. Lazaridis, I. *et al.* Genetic origins of the Minoans and Mycenaeans. *Nature* **548,** 214–218 (2017).
27. Lipson, M. *et al.* Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551,** 368–372 (2017).
28. Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538,** 201–206 (2016).
29. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).
30. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192,** 1065–1093 (2012).
31. Gronenborn, D. & Dolukhanov, P. in *The Oxford Handbook of Neolithic Europe* (eds Fowler, C., Harding, J. & Hofmann, D.) 195–214 (Oxford Univ. Press, 2015).
32. Telegin, D. Y. Neolithic cultures of Ukraine and adjacent areas and their chronology. *J. World Prehist.* **1,** 307–331 (1987).
33. Telegin, D. Y. & Potekhina, I. D. *Neolithic Cemeteries and Populations in the Dnieper Basin* (British Archaeological Reports, 1987).
34. Jones, E. R. *et al.* The Neolithic transition in the Baltic was not driven by admixture with early European farmers. *Curr. Biol.* **27,** 576–582 (2017).
35. Mittnik, A. *et al.* The genetic prehistory of the Baltic Sea region. *Nat. Commun.* **9,** 443 (2018).
36. Saag, L. *et al.* Extensive farming in Estonia started through a sex-biased migration from the Steppe. *Curr. Biol.* **27,** 2185–2193.e6 (2017).
37. Maier, A. *The Central European Magdalenian: Regional Diversity and Internal Variability* (Springer, 2015).
38. Borić, D. & Price, T. D. Strontium isotopes document greater human mobility at the start of the Balkan Neolithic. *Proc. Natl Acad. Sci. USA* **110,** 3298–3303 (2013).
39. Krauß, R., Marinova, E., De Brue, H. & Weninger, B. The rapid spread of early farming from the Aegean into the Balkans via the Sub-Mediterranean–Aegean vegetation Zone. *Quat. Int.* https://doi.org/10.1016/j.quaint.2017.01.019 (2017).
40. Bacvarov, K. in *Moments in Time: Papers Presented to Pál Raczky on his 60th Birthday* (eds Anders, A. & Kulcsár, G.) 29–34 (L'Harmattan, 2013).
41. Gurova, M. & Bonsall, C. 'Pre-Neolithic' in Southeast Europe: a Bulgarian perspective. *Documenta Praehistorica* **41,** 95–109 (2014).
42. Brandt, G. *et al.* Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* **342,** 257–261 (2013).
43. Borić, D. in *The Oxford Handbook of Neolithic Europe* (eds Fowler C., Harding, J. & Hofmann, D.) 927–957 (Oxford Univ. Press, 2015).
44. Szmyt, M. in *Transition to the Bronze Age (Archaeolingua 30)* (eds Heyd, V., Kulcsár, G. & Szeverényi, V.) 93–111 (Archaeolingua, 2013).
45. Olalde, I. *et al.* A common genetic origin for early farmers from Mediterranean Cardial and central European LBK cultures. *Mol. Biol. Evol.* **32,** 3132–3142 (2015).
46. Posth, C. *et al.* Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a Late Glacial population turnover in Europe. *Curr. Biol.* **26,** 827–833 (2016).
47. Anthony, D. W. & Ringe, D. The Indo-European homeland from linguistic and archaeological perspectives. *Annu. Rev. Linguist.* **1,** 199–219 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Anthony, I. Lazaridis and M. Lipson for comments on the manuscript, B. Llamas, A. Cooper and A. Furtwängler for contributions to laboratory work, R. Evershed for contributing ^{14}C dates and F. Novotny for assistance with samples. 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 the NSF Archaeometry program BCS-1460369 to D.J.K. for AMS ^{14}C work; by MEN-UEFISCDI grant, Partnerships in Priority Areas Program – PN II (PN-II-PT-PCCA-2013-4-2302) to C.L.; by Croatian Science Foundation grant IP-2016-06-1450 to M.N. and I.J.; by European Research Council grant ERC CoG 724703 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, the Howard Hughes Medical Institute and an Allen Discovery Center grant from the Paul Allen Foundation to D.R.

Author Contributions S.A.-R., A.S.-N., S.Vai., S.A., K.W.A., R.A., D.A., A.A., N.A., K.B., M.B.G., H.B., M.B., A.Bo., Y.B., A.Bu., J.B., S.C., N.J.C., R.C., M.C., C.C., D.G.D., N.E., M.Fr., B.Gal., G.G., B.Ge., T.Ha., V.H., K.H., T.Hi., S.I., I.J., I.Ka., D.Ko., A.K., D.La., M.La., C.L., M.Le., K.L., D.L.V., D.Lo., I.L., M.Ma., F.M., K.M., H.M., M.Me., P.M., V.M., V.P., T.D.P., A.Si., L.S., M.Š., V.S., P.S., A.St., T.S., M.T.-N., C.T., I.V., F.Va., S.Vas., F.Ve., S.Ve., E.V., B.V., C.V., J.Z., S.Z., P.W.S., G.C., R.K., D.C., G.Z., B.Gay., M.Li., A.G.N., I.P., A.P., D.B., C.B., J.K., R.P. and D.R. assembled and interpreted archaeological material. C.P., A.S.-N., N.R., N.B., F.C., O.C., D.F., M.Fe., B.Gam., G.G.F., W.H., E.H., E.J., D.Ke., B.K.-K., I.Ku., M.Mi., A.M., K.N., M.N., J.O., S.P., K.Si., K.St. and S.Vai. performed laboratory work. I.M., C.P., A.S.-N., S.M., I.O., N.P. and D.R. analysed data. D.J.K., S.T., D.B. and C.B. interpreted ^{14}C dates. J.K., R.P. and D.R. supervised analysis or laboratory work. I.M. and D.R. wrote the paper with input from all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to D.R. (reich@genetics.med.harvard.edu), R.P. (ron.pinhasi@univie.ac.at) or I.M. (mathi@pennmedicine.upenn.edu).

Reviewer Information *Nature* thanks C. Renfrew, A. Scally and the other anonymous reviewer(s) for their contribution to the peer review of this work.

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[25,31], 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[25,37], Thomas Higham[43], Stanislav Iliev[44], Ivor Janković[13,45], Ivor Karavanić[45,46], Douglas J. Kennett[47], Darko Komšo[33], Alexandra Kozak[48], Damian Labuda[49], Martina Lari[16], Catalin Lazar[21,50], Maleen Leppek[51], Krassimir Leshtakov[42], Domenico Lo Vetro[52,53], Dženi Los[29], Ivaylo Lozanov[42], Maria Malina[26], Fabio Martini[52,53], Kath McSweeney[54], Harald Meller[20], Marko Menđušić[55], Pavel Mirea[56], Vyacheslav Moiseyev[57], Vanya Petrova[42], T. Douglas Price[58], Angela Simalcsik[59], Luca Sineo[60], Mario Šlaus[61], Vladimir Slavchev[62], Petar Stanev[36], Andrej Starović[63], Tamás Szeniczey[41], Sahra Talamo[64], Maria Teschler-Nicola[7,65], Corinne Thevenet[66], Ivan Valchev[42], Frédérique Valentin[67], Sergey Vasilyev[68], Fanica Veljanovska[69], Svetlana Venelinova[70], Elizaveta Veselovskaya[68], Bence Viola[71,72], Cristian Virag[73], Joško Zaninović[74], Steve Zäuner[75], Philipp W. Stockhammer[2,51], Giulio Catalano[60], Raiko Krauß[76], David Caramelli[16], Gunita Zariņa[77], Bisserka Gaydarska[78], Malcolm Lillie[79], Alexey G. Nikitin[80], Inna Potekhina[48], Anastasia Papathanasiou[81], Dušan Borić[82], Clive Bonsall[54], Johannes Krause[2,3], Ron Pinhasi[6,7]* & David Reich[1,5,14]*

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, 07745 Jena, Germany. [3]Institute for Archaeological Sciences, University of Tübingen, Tübingen, 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, Massachusetts 02115, USA. [6]Earth Institute, University College Dublin, Belfield, Dublin 4, Ireland. [7]Department of Anthropology, University of Vienna, 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, Ferrara 44100, Italy. [10]Australian Centre for Ancient DNA, School of Biological Sciences, The University of Adelaide, SA-5005 Adelaide, South Australia, Australia. [11]Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland. [12]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. [13]Institute for Anthropological Research, 10000 Zagreb, Croatia. [14]Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. [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, 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, 06114 Halle, Germany. [21]National History Museum of Romania, 030026, Bucharest, Romania. [22]Institute of Archaeology, Belgrade, Serbia. [23]Institute of Experimental Morphology, Pathology and Anthropology with Museum, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.

[24]Department of Geosciences, Biogeology, Universität Tübingen, 72074 Tübingen, Germany. [25]Senckenberg Centre for Human Evolution and Palaeoenvironment at the University of Tübingen, 72076 Tübingen, Germany. [26]ROCEEH Research Center, Heidelberg Academy of Sciences and Humanities, University of Tübingen, 72070 Tübingen, Germany. [27]Vasile Pârvan Institute of Archaeology, Romanian Academy, 010667 Bucharest, Romania. [28]Human Biology Department, Cardinal Stefan Wyszyński University, 01-938 Warsaw, Poland. [29]KADUCEJ d.o.o., 21000 Split, Croatia. [30]St. Cyril and Methodius University, 5000 Veliko Turnovo, Bulgaria. [31]Department of Early Prehistory and Quaternary Ecology, University of Tübingen, 72070 Tübingen, Germany. [32]INRAP/UMR 8215 Trajectoires, 92023 Nanterre, France. [33]Archaeological Museum of Istria, 52100 Pula, Croatia. [34]Service Régional de l'Archéologie de Bourgogne-Franche-Comté, 25043 Besançon Cedex, France. [35]Laboratoire Chronoenvironnement, UMR 6249 du CNRS, UFR des Sciences et Techniques, 25030 Besançon Cedex, France. [36]Regional Museum of History Veliko Tarnovo, 5000 Veliko Tarnovo, Bulgaria. [37]Institute for Archaeological Sciences, Paleoanthropology, University of Tübingen, 72070 Tübingen, Germany. [38]Laboratory for Human Bioarchaeology, 1202 Sofia, Bulgaria. [39]Regional Museum of History, 3000 Vratsa, Bulgaria. [40]DRAC Auvergne - Rhône Alpes, Ministère de la Culture, Lyon Cedex 01, France. [41]Eötvös Loránd University, Faculty of Science, Institute of Biology, Department of Biological Anthropology, H-1117 Budapest, Hungary. [42]Department of Archaeology, Sofia University St. Kliment Ohridski, 1504 Sofia, Bulgaria. [43]Oxford Radiocarbon Accelerator Unit, Research Laboratory for Archaeology and the History of Art, University of Oxford, Dyson Perrins Building Oxford OX1 3QY, UK. [44]Regional Museum of History, 6300 Haskovo, Bulgaria. [45]Department of Anthropology, University of Wyoming, Laramie, Wyoming 82071, USA. [46]Department of Archaeology, Faculty of Humanities and Social Sciences, University of Zagreb, 10000 Zagreb, Croatia. [47]Department of Anthropology and Institutes for Energy and the Environment, Pennsylvania State University, University Park, Pennsylvania 16802, USA. [48]Department of Bioarchaeology, Institute of Archaeology, National Academy of Sciences of Ukraine, 04210 Kiev, Ukraine. [49]CHU Sainte-Justine Research Center, Pediatric Department, Université de Montréal, Montreal, Québec H3T 1C5, Canada. [50]Department of Ancient History, Archaeology and History of Art, Faculty of History, University of Bucharest, 50107 Bucharest, Romania. [51]Institute for Pre- and Protohistoric Archaeology and the Archaeology of the Roman Provinces, Ludwig-Maximilians-University, 80799 Munich, Germany. [52]Dipartimento SAGAS - Sezione di Archeologia e Antico Oriente, Università degli Studi di Firenze, 50122 Florence, Italy. [53]Museo e Istituto fiorentino di Preistoria, 50122 Florence, Italy. [54]School of History, Classics and Archaeology, University of Edinburgh, Edinburgh EH8 9AG, UK. [55]Conservation Department in Šibenik, Ministry of Culture of the Republic of Croatia, 22000 Šibenik, Croatia. [56]Teleorman County Museum, 140033 Alexandria, Romania. [57]Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS, 199034 St. Petersburg, Russia. [58]Department of Anthropology, University of Wisconsin, Madison, Wisconsin 53706, USA. [59]Olga Necrasov Centre for Anthropological Research, Romanian Academy – Iaşi Branch, 700481 Iaşi, Romania. [60]Dipartimento di Scienze e tecnologie biologiche, chimiche e farmaceutiche, Lab. of Anthropology, Università degli studi di Palermo, 90133 Palermo, Italy. [61]Anthropological Center, Croatian Academy of Sciences and Arts, 10000 Zagreb, Croatia. [62]Regional Historical Museum Varna, BG-9000 Varna, Bulgaria. [63]National Museum in Belgrade, Belgrade, Serbia. [64]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [65]Department of Anthropology, Natural History Museum Vienna, 1010 Vienna, Austria. [66]INRAP/UMR 8215 Trajectoires, 92023 Nanterre, France. [67]CNRS/UMR 7041 ArScAn MAE, 92023 Nanterre, France. [68]Institute of Ethnology and Anthropology, Russian Academy of Sciences, Moscow, 119991, Russia. [69]Archaeological Museum of Macedonia, 1000 Skopje, the former Yugoslav Republic of Macedonia. [70]Regional Museum of History, 9700 Shumen, Bulgaria. [71]Department of Anthropology, University of Toronto, Toronto, Ontario, M5S 2S2, Canada. [72]Institute of Archaeology & Ethnography, Siberian Branch, Russian Academy of Sciences, Novosibirsk 630090, Russia. [73]Satu Mare County Museum Archaeology Department, 440026 Satu Mare, Romania. [74]Municipal Museum Drniš, 22320 Drniš, Croatia. [75]anthropol - Anthropologieservice, 72379 Hechingen, Germany. [76]Institute for Prehistory, Early History and Medieval Archaeology, University of Tübingen, 72070 Tübingen, Germany. [77]Institute of Latvian History, University of Latvia, Rīga 1050, Latvia. [78]Department of Archaeology, Durham University, Durham DH1 3LE, UK. [79]School of Environmental Sciences: Geography, University of Hull, Hull HU6 7RX, UK. [80]Department of Biology, Grand Valley State University, Allendale, Michigan 49401, USA. [81]Ephorate of Paleoanthropology and Speleology, 11636 Athens, Greece. [82]The Italian Academy for Advanced Studies in America, Columbia University, New York, New York 10027, USA.
†Present address; Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.
*These authors contributed equally to this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Ancient DNA analysis. We extracted DNA and prepared next-generation sequencing libraries in four different dedicated ancient DNA laboratories (Adelaide, Boston, Budapest and Tübingen). We also prepared powder in a fifth laboratory (Dublin), which 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, 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, 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 Tübingen and at the Max Planck Institute for the Science of Human History in Jena. Extraction[48] and library preparation[49,50] followed established protocols. We performed in-solution enrichment for sequences overlapping about 1.24 million SNPs ('1240k capture'), and sequenced on an Illumina HiSeq 4000 or NextSeq 500 for 76 bp using either single- or paired-end sequencing.

The remaining 199 samples were processed at Harvard Medical School, Boston. Using about 75 mg of sample powder from each sample (prepared from skeletal samples either in Boston or University College Dublin, Dublin), 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 a subset of samples, we increased coverage by preparing additional libraries from the existing DNA extract using 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]. We screened all libraries from Adelaide, Boston and Budapest by enriching for the mitochondrial genome, plus about 3,000 (50 in an earlier and unpublished version) nuclear SNPs using a bead-based capture[55] but with the probes replaced by amplified oligonucleotides synthesized by CustomArray. After the capture, we completed the adapter sites using PCR, attaching dual index combinations[56] to each enriched library. We sequenced the products of between 100 and 200 libraries together with the non-enriched libraries (shotgun) on an Illumina NextSeq500 using v.2 150 cycle kits for 2×76 cycles and 2×7 cycles.

In Boston, we performed two rounds of in-solution enrichment ('1240k capture') for a targeted set of 1,237,207 SNPs using previously reported protocols[7,14,23]. For a total of 41 individuals, we increased coverage by building one to nine additional libraries for the same sample. When we built multiple libraries from the same extract, we often pooled them in equimolar ratios before the capture. We performed all sequencing on an Illumina NextSeq500 using v.2 150 cycle kits for 2×76 cycles and 2×7 cycles. We attempted to sequence each enriched library up to the point at which we estimated that it was economically inefficient to sequence further. Specifically, we iteratively sequenced more from each individual and only stopped when we estimated that the expected increase in the number of targeted SNPs hit at least once would be less than about one for every 100 new read pairs generated. After sequencing and merging the paired reads into a single sequence, we trimmed two bases from the end of each sequence and aligned to the human genome (b37/hg19) using bwa[57]. We then removed individuals with evidence of contamination based on mitochondrial DNA polymorphism[58] or difference in principal component analysis space between damaged and undamaged reads[59], a high rate of heterozygosity on the chromosome X despite being male[59,60] or a ratio of X-to-Y chromosome sequences different than would be expected from either an XX or XY karyotype. We also removed individuals that had low coverage (fewer than 15,000 SNPs hit on the autosomes). We report, but do not analyse, data from nine individuals that were first-degree relatives of others in the dataset (determined by comparing rates of allele sharing between pairs of individuals).

After removing a small number of sites that failed to capture, we were left with a total of 1,233,013 sites of which 32,670 were on chromosome X and 49,704 were on chromosome Y, with a median coverage at targeted SNPs on the 215 newly reported individuals of 0.90 (range 0.007–9.2; Supplementary Table 1). We generated 'pseudo-haploid' calls by selecting a single sequence randomly for each individual at each SNP. Thus, there is only a single allele from each individual at each site, but adjacent alleles might come from either of the two haplotypes of the individual. We merged the newly reported data with previously reported data from 274 other ancient individuals[9–11,15–27], making pseudo-haploid calls in the same way at the 1240k sites for individuals that were shotgun-sequenced rather than captured.

Using the captured mitochondrial sequence from the screening process, we called mitochondrial haplogroups. Using the 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 (accessed 21 April 2016).

Population genetic analysis. To analyse these ancient individuals in the context of present-day genetic diversity, we merged them with the following two datasets: (1) 300 high-coverage genomes from a diverse worldwide set of 142 populations sequenced as part of the Simons Genome Diversity Project[28] ('SGDP merge') and (2) 777 west Eurasian individuals genotyped on the Human Origins array[23], with 597,573 sites in the merged dataset ('HO merge').

We computed principal components of the present-day individuals in the HO merge and projected the ancient individuals onto the first two components using the 'lsqproject: YES' option in smartpca (v.15100)[61] (https://www.hsph.harvard.edu/alkes-price/software/).

We ran ADMIXTURE (v.1.3.0) in both supervised and unsupervised mode. In supervised mode we used only the ancient individuals, on the full set of SNPs, and with the following population labels fixed: Anatolia_Neolithic, WHG, EHG, and Yamnaya.

For unsupervised mode we used the HO merge dataset, including 777 present-day individuals. We flagged individuals that were genetic outliers based on principal components analysis and ADMIXTURE relative to other individuals from the same time period and archaeological culture.

We computed D statistics using qpDstat (v.710). D statistics of the form $D(A,B,X,Y)$ test the null hypothesis of the unrooted tree topology $((A,B),(X,Y))$. A positive value indicates that either A and X, or B and Y, share more drift than expected under the null hypothesis. We quote D statistics as the Z score computed using default block jackknife parameters.

We fitted admixture proportions with qpAdm (v.610) using the SGDP merge. Given a set of outgroup ('right') populations, qpAdm models one of a set of source ('left') populations (the 'test' population) as a mixture of the other sources by fitting admixture proportions to match the observed matrix of f_4 statistics as closely as possible. We report a P value for the null hypothesis that the test population does not have ancestry from another source that is differentially related to the right populations. We computed standard errors for the mixture proportions using a block jackknife. Importantly, qpAdm does not require that the source populations are actually the admixing populations—nor does it require that they are unadmixed themselves. Instead, qpAdm only requires that source populations are a clade with the correct admixing populations, relative to the other sources. Infeasible coefficient estimates (that is, outside [0, 1]) are usually a sign of poor model fit, but in the case where the source with a negative coefficient is itself admixed, could be interpreted as implying that the true source is a population with different admixture proportions. 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 and Han.DG.

For some analyses for which we required extra resolution (Supplementary Table 4) we used an extended set of 14 right (outgroup) populations, including additional Upper Palaeolithic European individuals[17]: 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 and Han.DG.

We also fitted admixture graphs with qpGraph (v.6021)[30] (https://github.com/DReichLab/AdmixTools, Supplementary Note 3). As with qpAdm, qpGraph also tries to match a matrix of f statistics, but rather than fitting one population as a mixture of other, specified, populations, it fits the relationship between all tested populations simultaneously, potentially incorporating multiple admixture events. However, qpGraph requires the graph relating the populations to be specified in advance. We tested goodness-of-fit by computing the expected D statistics under the fitted model, finding the largest D statistic outlier between the fitted and observed model, and then computing a Z score using a block jackknife.

For 114 individuals with hunter-gatherer-related ancestry, we estimated an effective migration surface using the software EEMS (https://github.com/dipetkov/eems)[62]. 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 Markov chain Monte Carlo 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 analyse 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 $Z = \frac{p_A - p_X}{\sqrt{\sigma_A^2 + \sigma_X^2}}$ in which 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, indicating that 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 observed 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 C1, I2 and R1, which are all common in Mesolithic hunter-gatherers but rare or absent in northwestern-Anatolian Neolithic individuals. 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 haplogroups associated with each ancestry type using the Agresti–Coull method[63,64] 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-related are given by m and f, respectively, then the proportions of hunter-gatherer-related 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 error and computed the likelihood surface for (m,f) over a grid ranging from (0,0) to (1,1).

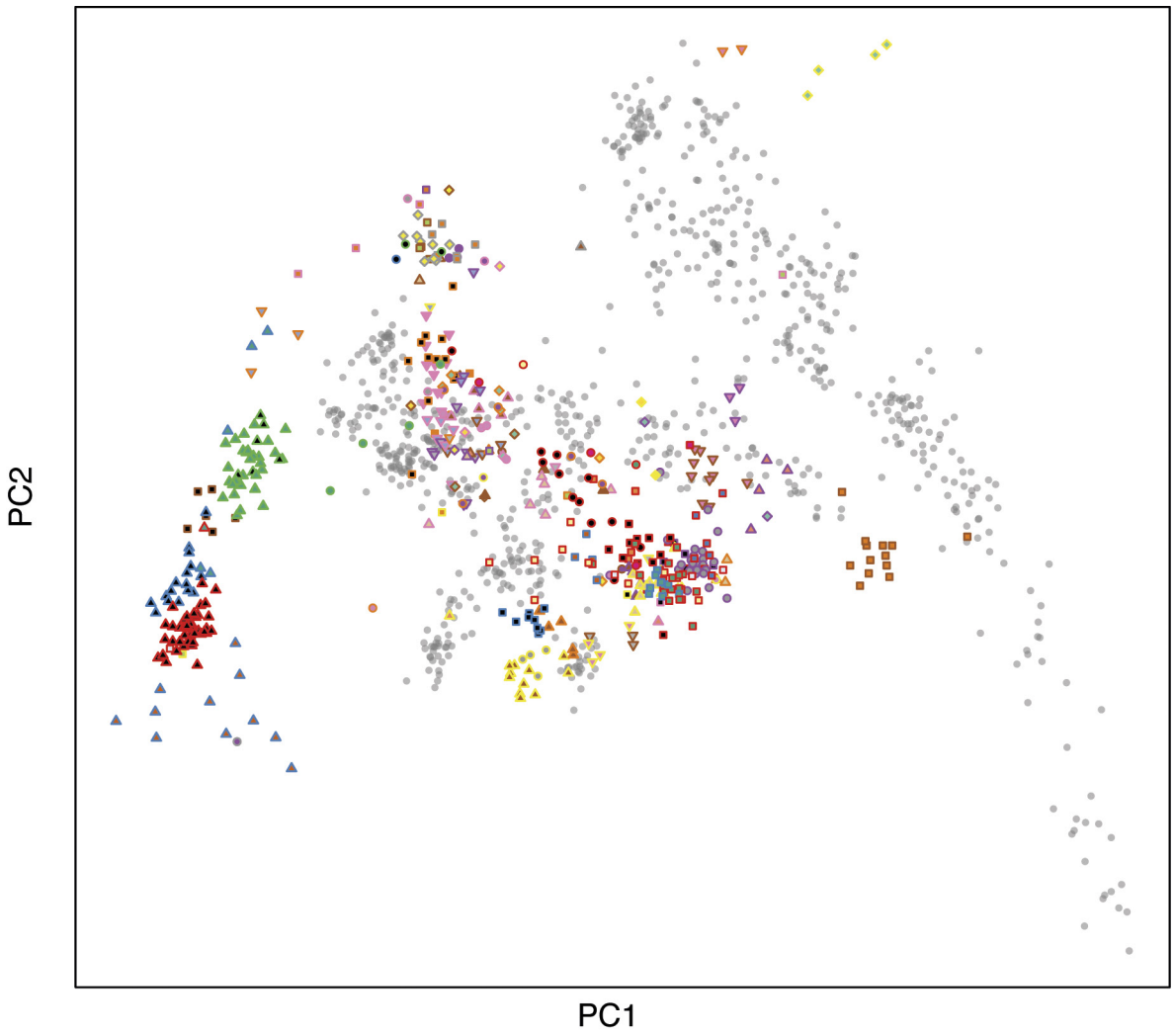
Direct AMS ^{14}C bone dates. We report 137 direct AMS ^{14}C bone dates for 136 individuals from multiple AMS radiocarbon laboratories. In general, bone samples were manually cleaned and demineralized in weak HCl and, in most cases (Radiocarbon laboratory codes: UCIAMS and OxA), soaked in an alkali bath (NaOH) at room temperature to remove contaminating soil humates. Samples were then rinsed to neutrality in Nanopure H_2O and gelatinized in HCl[65]. The resulting gelatin was lyophilized and weighed to determine per cent yield as a measure of collagen preservation (percentage crude gelatin yield). Collagen was then directly AMS ^{14}C-dated (Radiocarbon laboratory codes: Beta, AA) or further purified using ultrafiltration (Radiocarbon laboratory codes: PSU, UCIAMS, OxA, Poz, and MAMS)[66]. It is standard in some laboratories (Radiocarbon laboratory codes: PSU, UCIAMS and OxA) to use stable carbon and nitrogen isotopes as an additional quality control measure. For these samples, the percentage of C, percentage of N and C:N ratios were evaluated before AMS ^{14}C dating[67]. C:N ratios for well-preserved samples fall between 2.9 and 3.6, indicating good collagen preservation[68]. For 119 of the new dates, we also report δ^{13}C and δ^{15}N values (Supplementary Table 6).

All ^{14}C ages were δ^{13}C-corrected for mass-dependent fractionation with measured ^{13}C/^{12}C values[69] and calibrated with OxCal version 4.2.3[70] using the IntCal13 northern hemisphere calibration curve[70]. For hunter-gatherers from the Iron Gates, the direct ^{14}C dates tend to be overestimates because of the freshwater reservoir effect (FRE), which arises because of a diet including fish that consumed ancient carbon. For these individuals, we performed a correction[71] (Supplementary Note 1) that assumed that 100% FRE = 545 ± 70 years, and δ^{15}N values of 8.3% and 17.0% for 100% terrestrial and aquatic diets, respectively.

Code availability. The software used to analyse the data is available from the following sources: smartpca, qpAdm, qpDstat and qpGraph (https://github.com/DReichLab/AdmixTools/), ADMIXTURE (https://www.genetics.ucla.edu/software/admixture/), EEMS (https://github.com/dipetkov/eems/), bwa (http://bio-bwa.sourceforge.net) and OxCal (https://c14.arch.ox.ac.uk/oxcal.html).

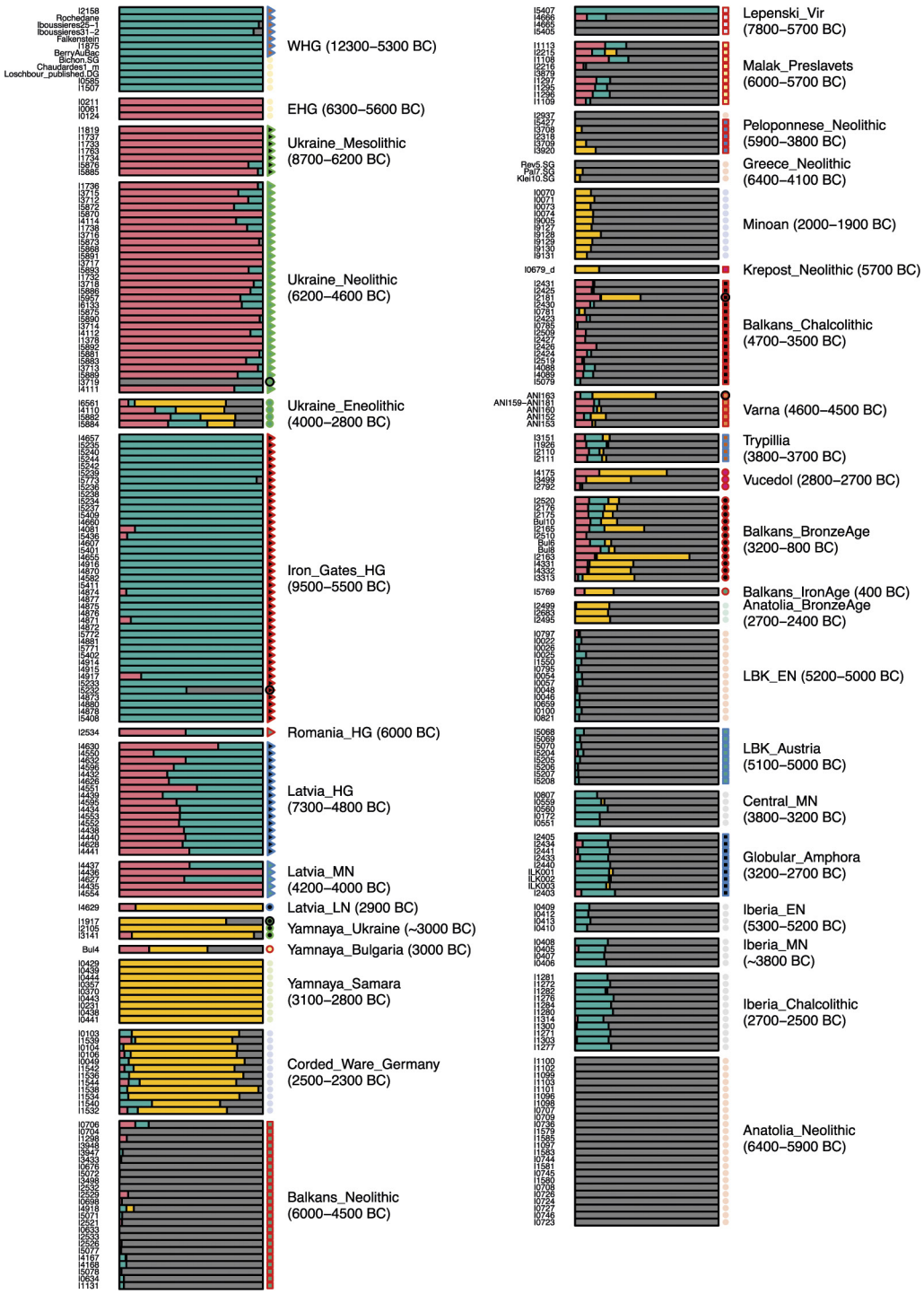
Data availability. The aligned sequences are available through the European Nucleotide Archive under accession number PRJEB22652. The pseudo-haploid genotype dataset used in analysis and in consensus mitochondrial genomes is available at https://reich.hms.harvard.edu/datasets.

48. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110,** 15758–15763 (2013).
49. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* https://doi.org/10.1101/pdb.prot5448 (2010).
50. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. Lond. B* **370,** 20130624 (2015).
51. Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59,** 87–93 (2015).
52. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Res.* **38,** e87 (2010).
53. DeAngelis, M. M., Wang, D. G. & Hawkins, T. L. Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res.* **23,** 4742–4743 (1995).
54. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22,** 939–946 (2012).
55. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5,** e14004 (2010).
56. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40,** e3 (2012).
57. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26,** 589–595 (2010).
58. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23,** 553–559 (2013).
59. Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* **111,** 2229–2234 (2014).
60. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15,** 356 (2014).
61. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).
62. Petkova, D., Novembre, J. & Stephens, M. Visualizing spatial population structure with estimated effective migration surfaces. *Nat. Genet.* **48,** 94–100 (2016).
63. Brown, L. D., Cai, T. T. & DasGupta, A. Interval estimation for a binomial proportion. *Stat. Sci.* **16,** 101–133 (2001).
64. Agresti, A. & Coull, B. A. Approximate is better than 'exact' for interval estimation of binomial proportions. *Am. Stat.* **52,** 119–126 (1998).
65. Longin, R. New method of collagen extraction for radiocarbon dating. *Nature* **230,** 241–242 (1971).
66. Brown, T. A., Nelson, D. E., Vogel, J. S. & Southon, J. R. Improved collagen extraction by modified Longin method. *Radiocarbon* **30,** 171–177 (1988).
67. Kennett, D. J. *et al.* Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8,** 14115 (2017).
68. van Klinken, G. J. Bone collagen quality indicators for palaeodietary and radiocarbon measurements. *J. Archaeol. Sci.* **26,** 687–695 (1999).
69. Stuiver, M. & Polach, H. A. Discussion: reporting of ^{14}C data. *Radiocarbon* **19,** 355–363 (1977).
70. Bronk Ramsey, C. *OxCal 4.23 Online Manual* https://c14.arch.ox.ac.uk/oxcalhelp/hlp_contents.html (2013).
71. Cook, G. T. *et al.* A freshwater diet-derived ^{14}C reservoir effect at the Stone Age sites in the Iron Gates gorge. *Radiocarbon* **43,** 453–460 (2001).

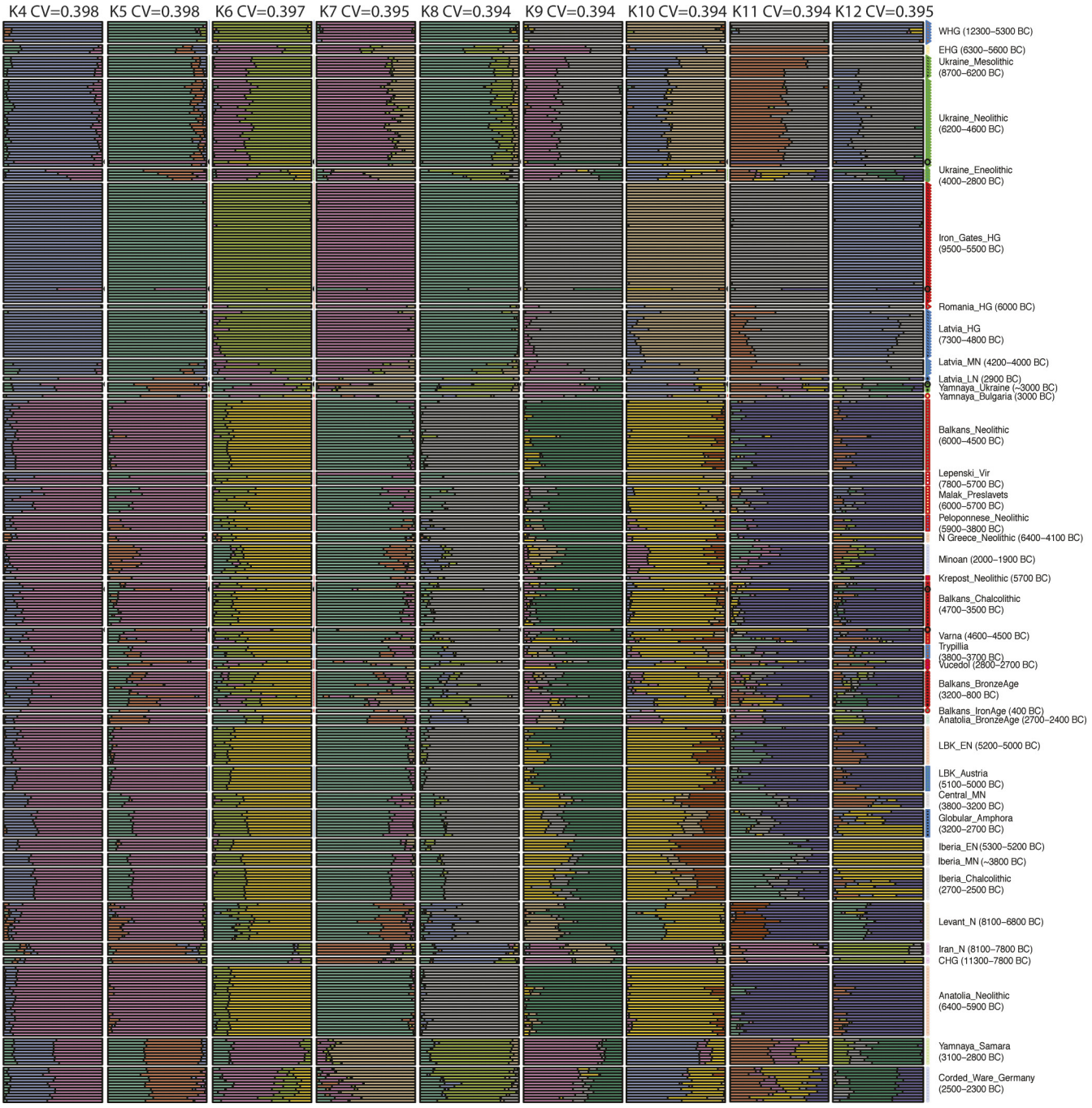


- Afanasievo.SG
- AfontovaGora3
- Alberstedt_LN
- ALPc_MN
- Anatolia_BronzeAge
- Anatolia_Neolithic
- Anatolia_Neolithic_Boncuklu.SG
- Anatolia_Neolithic_Kumtepe.SG
- Anatolia_Neolithic_Tepecik_Ciftlik.SG
- Andronovo.SG
- Baden_LCA
- Balkans_Chalcolithic_outlier
- BattleAxe_Sweden.SG
- Bell_Beaker_Czech.SG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

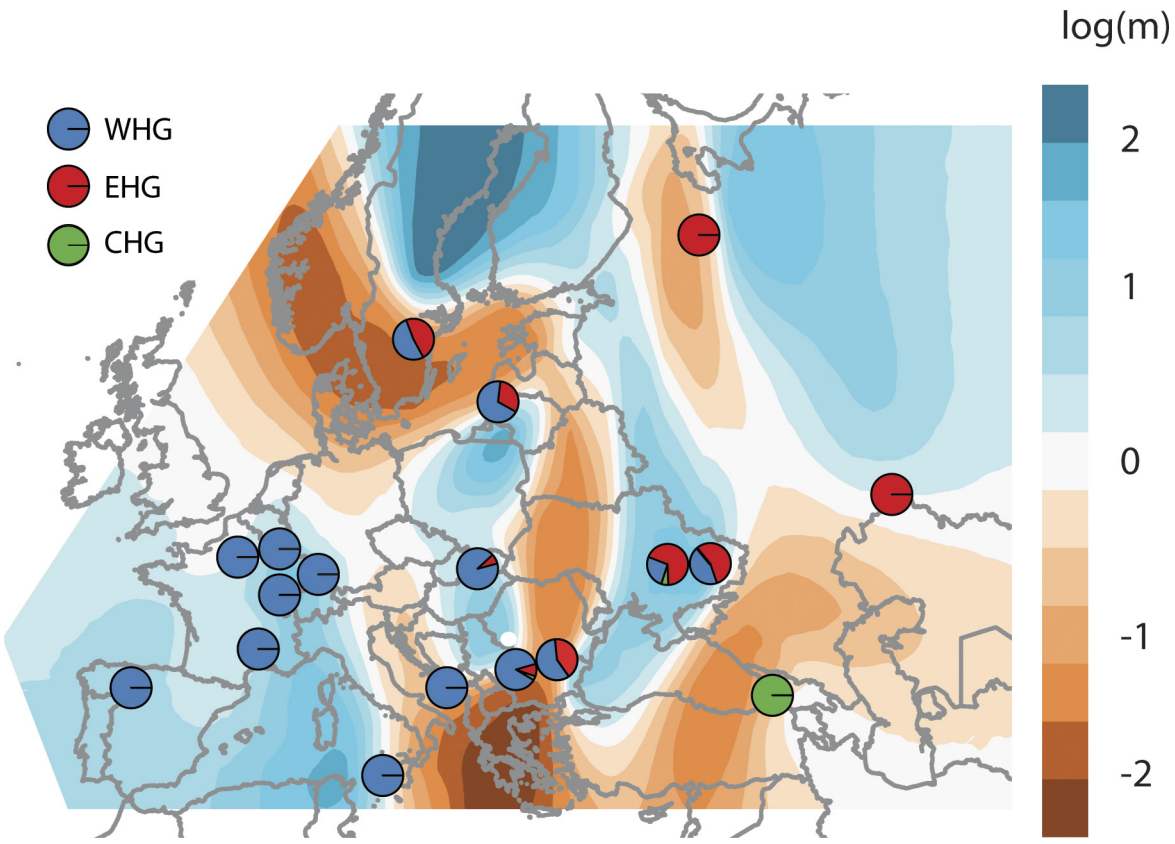
Extended Data Figure 1 | Principal components analysis of ancient individuals. Points for 486 ancient individuals are projected onto principal components defined by 777 present-day west Eurasian individuals (grey points). This differs from Fig. 1b in that the plot is not cropped and the present-day individuals are shown.



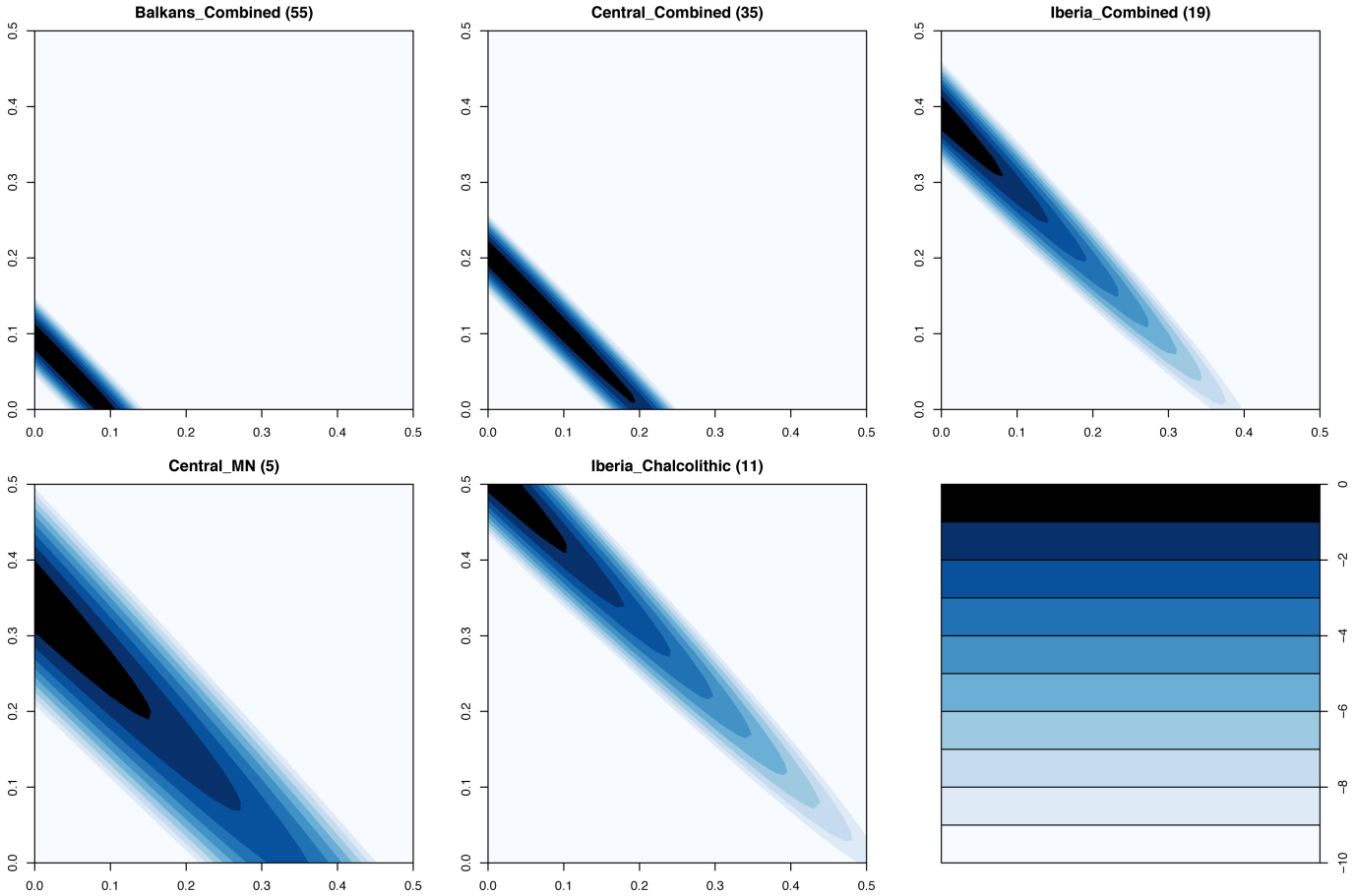
Extended Data Figure 2 | Supervised ADMIXTURE analysis. Supervised ADMIXTURE analysis modelling each ancient individual (one per row), as a mixture of populations represented by clusters that are constrained to contain northwestern-Anatolian Neolithic (grey), Yamnaya from Samara (yellow), EHG (pink) and WHG (green) populations. Dates in parentheses indicate approximate range of individuals in each population. This differs from Fig. 1d in that it contains some previously published samples[7,9,10,19,23,26] and includes sample identification numbers.



Extended Data Figure 3 | Unsupervised ADMIXTURE analysis. Unsupervised ADMIXTURE plot from $k = 4$ to 12 on a dataset consisting of 1,099 present-day individuals and 476 ancient individuals. We show newly reported ancient individuals and some previously published individuals[7,10,19,22,23,26] for comparison.



Extended Data Figure 4 | Genetic spatial structure in hunter-gatherers. We infer the estimated effective migration surface[62], a model of genetic relatedness in which individuals move in a random direction from generation to generation on an underlying grid, such that genetic relatedness is determined by distance. The migration parameter, m, defining 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 the model as much as possible to hunter-gatherer populations, the migration surface is inferred using data from 116 individuals that date to earlier than approximately 5000 BC and have no northwestern-Anatolian-Neolithic-related ancestry. Although the migration surface is sensitive to sampling and fine-scale features may not be interpretable, the migration 'barrier' (region of low migration) running north-to-south and separating populations with primarily WHG ancestry from those with 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 (that is, genetic differentiation did not vary consistently with distance). Superimposed on this background, pie charts show the WHG, EHG and CHG ancestry proportions inferred for populations used to construct the migration surface. This represents another way of visualizing the data in Fig. 2, Supplementary Table 3.1.3; we use two population models if they fit with $P > 0.01$, and three population models otherwise. Pie charts with only a single colour are those that were fixed to be the source populations.



Extended Data Figure 5 | Sex bias in hunter-gatherer admixture. The log-likelihood surfaces for the proportion of female (*x* axis) and male (*y* axis) ancestors that are hunter-gatherer-related for the combined populations analysed in Fig. 3c, and the two populations with the strongest evidence for sex bias. Numbers in parentheses, number of individuals in each group. The log-likelihood scale ranges from 0 to −10, in which 0 is the feasible point with the highest likelihood.

Web summary	Genome-wide ancient DNA data from 225 individuals who lived in southeastern Europe between 12000 and 500 BC reveals that the region acted as a genetic crossroads before and after the arrival of farming.