ARTICLE

doi:10.1038/nature19310

Genomic insights into the origin of farming in the ancient Near East

Iosif Lazaridis[1,2], Dani Nadel[3], Gary Rollefson[4], Deborah C. Merrett[5], Nadin Rohland[1], Swapan Mallick[1,2,6], Daniel Fernandes[7,8], Mario Novak[7,9], Beatriz Gamarra[7], Kendra Sirak[7,10], Sarah Connell[7], Kristin Stewardson[1,6], Eadaoin Harney[1,6,11], Qiaomei Fu[1,12,13], Gloria Gonzalez-Fortes[14], Eppie R. Jones[15], Songül Alpaslan Roodenberg[16], György Lengyel[17], Fanny Bocquentin[18], Boris Gasparian[19], Janet M. Monge[20], Michael Gregg[20], Vered Eshed[21], Ahuva-Sivan Mizrahi[21], Christopher Meiklejohn[22], Fokke Gerritsen[23], Luminita Bejenaru[24], Matthias Blüher[25], Archie Campbell[26], Gianpiero Cavalleri[27], David Comas[28], Philippe Froguel[29,30], Edmund Gilbert[27], Shona M. Kerr[26], Peter Kovacs[31], Johannes Krause[32], Darren McGettigan[33], Michael Merrigan[34], D. Andrew Merriwether[35], Seamus O'Reilly[34], Martin B. Richards[36], Ornella Semino[37], Michel Shamoon-Pour[35], Gheorghe Stefanescu[38], Michael Stumvoll[25], Anke Tönjes[25], Antonio Torroni[37], James F. Wilson[39,40], Loic Yengo[29], Nelli A. Hovhannisyan[41], Nick Patterson[2], Ron Pinhasi[7]§ & David Reich[1,2,6]§

We report genome-wide ancient DNA from 44 ancient Near Easterners ranging in time between ~12,000 and 1,400 BC, from Natufian hunter–gatherers to Bronze Age farmers. We show that the earliest populations of the Near East derived around half their ancestry from a 'Basal Eurasian' lineage that had little if any Neanderthal admixture and that separated from other non-African lineages before their separation from each other. The first farmers of the southern Levant (Israel and Jordan) and Zagros Mountains (Iran) were strongly genetically differentiated, and each descended from local hunter–gatherers. By the time of the Bronze Age, these two populations and Anatolian-related farmers had mixed with each other and with the hunter–gatherers of Europe to greatly reduce genetic differentiation. The impact of the Near Eastern farmers extended beyond the Near East: farmers related to those of Anatolia spread westward into Europe; farmers related to those of the Levant spread southward into East Africa; farmers related to those of Iran spread northward into the Eurasian steppe; and people related to both the early farmers of Iran and to the pastoralists of the Eurasian steppe spread eastward into South Asia.

Between 10,000 and 9,000 BC, humans began practicing agriculture in the Near East[1]. In the ensuing five millennia, plants and animals domesticated in the Near East spread throughout West Eurasia (a vast region that also includes Europe) and beyond. The relative homogeneity of present-day West Eurasians in a world context[2] suggests the possibility of extensive migration and admixture that homogenized geographically and genetically disparate sources of ancestry. The spread of the world's first farmers from the Near East would have been a mechanism for such homogenization. To date, however, owing to the poor preservation of DNA in warm climates, it has been impossible to study the population structure and history of the first farmers and to trace their contribution to later populations.

In order to overcome the obstacle of poor DNA preservation, we took advantage of two methodological developments. First, we sampled from the inner ear region of the petrous bone[3,4] which can yield up to ~100 times more endogenous DNA than other skeletal elements[4]. Second, we used in-solution hybridization[5] to enrich extracted DNA for about 1.2 million single nucleotide polymorphism (SNP) targets[6,7], making efficient sequencing practical by filtering out microbial and non-informative human DNA. We merged all sequences extracted from each individual, and randomly sampled a single sequence with minimum mapping and sequence quality to represent each SNP, restricting our investigation to individuals with at least 9,000 SNPs covered at least once (Methods). We obtained genome-wide data that passed

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [3]The Zinman Institute of Archaeology, University of Haifa, Haifa 3498838, Israel. [4]Department of Anthropology, Whitman College, Walla Walla, Washington 99362, USA. [5]Department of Archaeology, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada. [6]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. [7]School of Archaeology and Earth Institute, Belfield, University College Dublin, Dublin 4, Ireland. [8]CIAS, Department of Life Sciences, University of Coimbra, Coimbra 3000-456, Portugal. [9]Institute for Anthropological Research, Zagreb 10000, Croatia. [10]Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA. [11]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [12]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany. [13]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing 100044, China. [14]Department of Biology and Evolution, University of Ferrara, Ferrara I-44121, Italy. [15]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. [16]J.M. van Nassaulaan 9, Santpoort-Noord 2071 VA, The Netherlands. [17]Department of Prehistory and Archaeology, University of Miskolc, Miskolc-Egyetemváros 3515, Hungary. [18]French National Centre for Scientific Research, UMR 7041, Nanterre Cedex 92023, France. [19]Institute of Archaeology and Ethnology, National Academy of Sciences of the Republic of Armenia, Yerevan 0025, Republic of Armenia. [20]University of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, Pennsylvania 19104, USA. [21]Israel Antiquities Authority, Jerusalem 91004, Israel. [22]Department of Anthropology, University of Winnipeg, Winnipeg, Manitoba R3B 2E9, Canada. [23]Netherlands Institute in Turkey, Istanbul 34433, Turkey. [24]Faculty of Biology, Alexandru Ioan Cuza University of Iasi, Iasi 700505, Romania. [25]Department of Internal Medicine and Dermatology, Clinic of Endocrinology and Nephrology, Leipzig 04103, Germany. [26]Generation Scotland, Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK. [27]RCSI Molecular & Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland. [28]Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona 08003, Spain. [29]Univ. Lille, CNRS, Institut Pasteur de Lille, UMR 8199 - EGID, Lille F-59000, France. [30]Imperial College London, Department of Genomics of Common Disease, London Hammersmith Hospital, London W12 0HS, UK. [31]Leipzig University Medical Center, IFB Adiposity Diseases, Leipzig 04103, Germany. [32]Max Planck Institute for the Science of Human History, Jena 07745, Germany. [33]School of History, Newman Building, University College Dublin, Belfield, Dublin 4, Ireland. [34]Genealogical Society of Ireland, Dún Laoghaire, County Dublin, Ireland. [35]Department of Anthropology, Binghamton University, State University of New York, New York 13902, USA. [36]Department of Biological Sciences, School of Applied Sciences, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK. [37]Dipartimento di Biologia e Biotecnologie "L. Spallanzani", Università di Pavia, Pavia 27100, Italy. [38]Institutul de Cercetari Biologice, Iaşi 700505, Romania. [39]Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh EH8 9AG, UK. [40]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. [41]Center of Excellence in Applied Biosciences, Yerevan State University, Yerevan 0025, Republic of Armenia. §These authors jointly supervised this work.

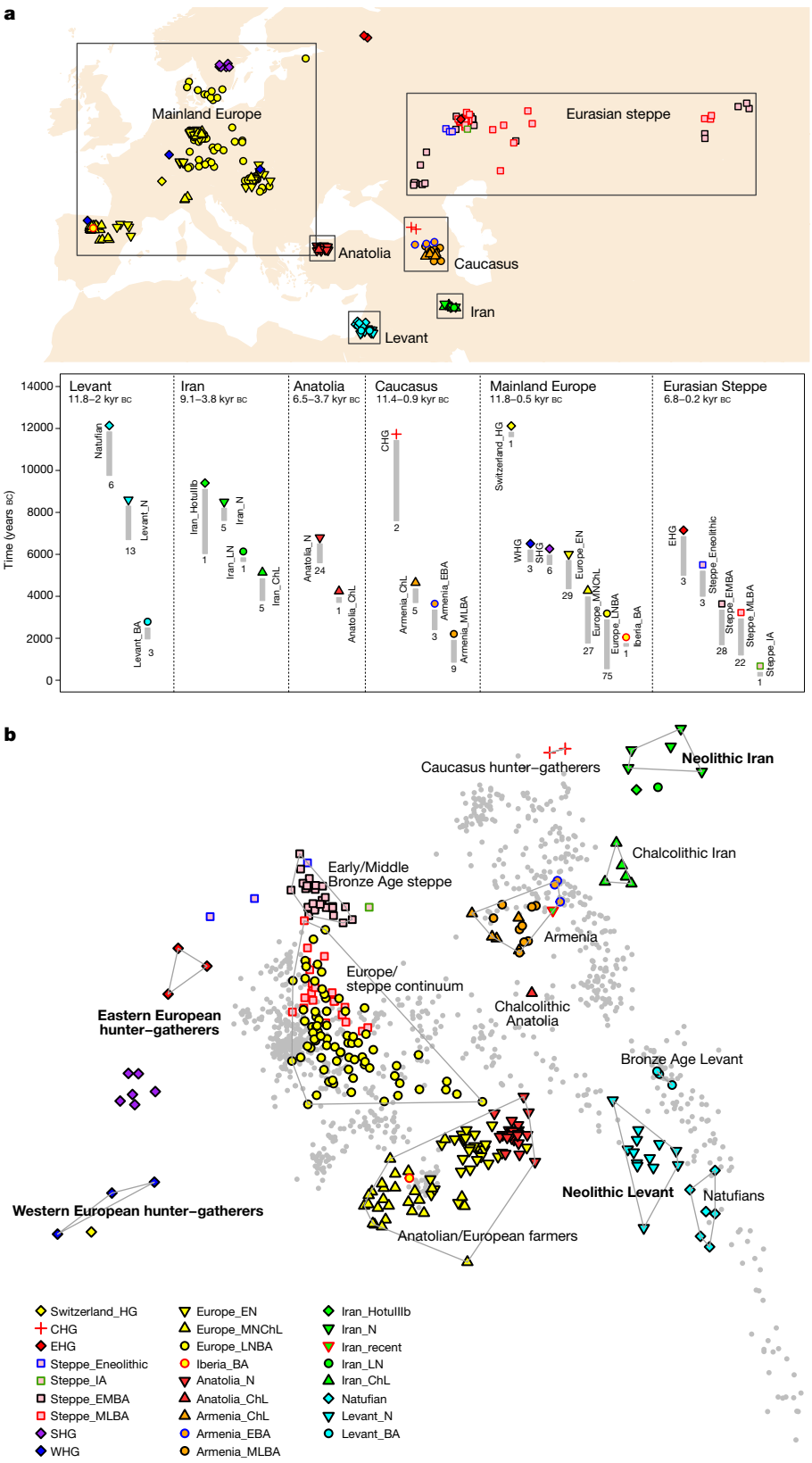


Figure 1 | Genetic structure of ancient West Eurasia. a, Sampling locations and times in six regions. Sample sizes for each population are given below each bar. E, Early; M, Middle; L, Late; HG, hunter–gatherer; N, Neolithic; ChL, Chalcolithic; BA, Bronze Age; IA, Iron Age. **b**, Principal components analysis of 991 present-day West Eurasians (grey points) with 278 projected ancient samples (excluding the Upper Palaeolithic Ust'-Ishim, Kostenki14, and MA1). To avoid visual clutter, population labels of present-day individuals are shown in Extended Data Fig. 1.

quality control for 45 individuals on whom we had a median coverage of 172,819 SNPs. We assembled direct radiocarbon dates on skeletal remains from 26 of these individuals (22 newly generated for this study) (Supplementary Table 1).

The newly reported ancient individuals date to ~12,000–1,400 BC and come from the southern Caucasus (Armenia), northwestern Anatolia (Turkey), Iran, and the southern Levant (Israel and Jordan) (Supplementary Table 1 and Fig. 1a). (One individual had a radiocarbon date that was not in agreement with the date of its archaeological context and was also a genetic outlier.) The samples include Epipalaeolithic Natufian hunter–gatherers from Raqefet Cave in the Levant (~12,000–9,800 BC); a likely Mesolithic individual (HotuIIIb) from Hotu Cave in the Alborz mountains of Iran (probable date of 9,100–8,600 BC); pre-pottery Neolithic farmers from 'Ain Ghazal and

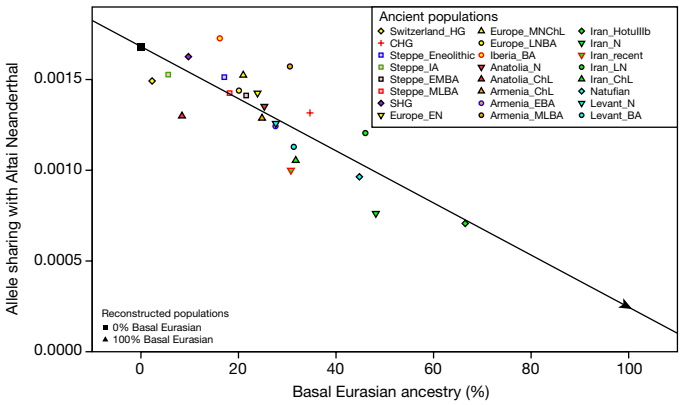


Figure 2 | Basal Eurasian ancestry explains the reduced Neanderthal admixture in West Eurasians. Basal Eurasian ancestry estimates are negatively correlated to a statistic measuring Neanderthal ancestry f_4(*Test*, Mbuti; Altai, Denisovan).

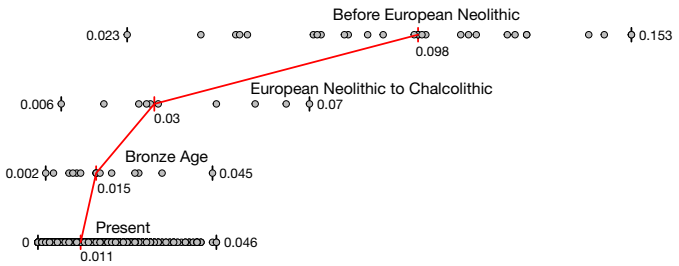


Figure 3 | Genetic differentiation and its marked decrease over time in West Eurasia. Pairwise F_{ST} distribution among populations belonging to four successive time slices in West Eurasia; the median (red) and range of F_{ST} is shown.

Motza in the southern Levant (~8,300–6,700 BC); and early farmers from Ganj Dareh in the Zagros mountains of western Iran (~8,200–7,600 BC). The samples also include later Neolithic, Chalcolithic (~4,800–3,700 BC), and Bronze Age (~3,350–1,400 BC) individuals (Supplementary Information, section 1). We combined our data with previously published ancient data[7–15] to form a dataset of 281 ancient individuals. We then further merged these data with 2,583 present-day people genotyped on the Affymetrix Human Origins array[13,16] (238 newly generated) (Supplementary Table 2 and Supplementary Information, section 2). We grouped the ancient individuals on the basis of archaeological culture and chronology (Fig. 1a and Supplementary Table 1). We refined the grouping on the basis of patterns evident in Principal Components Analysis (PCA)[17] (Fig. 1b and Extended Data Fig. 1), ADMIXTURE model-based clustering[18] (Extended Data Fig. 2a), and 'outgroup' f_3-analysis (Extended Data Fig. 3). We used f_4-statistics to identify outlier individuals and to cluster phylogenetically indistinguishable groups into 'Analysis Labels' (Supplementary Information, section 3).

We analysed these data to address six questions. (1) Previous work has shown that the first European farmers harboured ancestry from a Basal Eurasian lineage that diverged from the ancestors of north Eurasian hunter–gatherers and East Asians before they separated from each other[13]. What was the distribution of Basal Eurasian ancestry in the ancient Near East? (2) Were the first farmers of the Near East part of a single homogeneous population, or were they regionally differentiated? (3) Was there continuity between late pre-agricultural hunter–gatherers and early farming populations, or were the hunter–gatherers largely displaced by a single expansive population, as in early Neolithic Europe?[8] (4) What is the genetic contribution of these early Near Eastern farmers to later populations of the Near East? (5) What is the genetic contribution of the early Near Eastern farmers to later populations of mainland Europe, the Eurasian steppe, and to populations outside West Eurasia? (6) Do our data provide broader insights about population transformations in West Eurasia?

Basal Eurasian and Neanderthal ancestry

The 'Basal Eurasians' are a lineage hypothesized[13] to have split off before the differentiation of all other Eurasian lineages, including eastern non-African populations such as the Han Chinese, and even the early diverged lineage represented by the genome sequence of the ~45,000-year-old Upper Palaeolithic Siberian from Ust'-Ishim[11]. To test for Basal Eurasian ancestry, we computed the statistic f_4(*Test*, Han; Ust'-Ishim, Chimp) (Supplementary Information, section 4), which measures the excess of allele sharing of Ust'-Ishim with a variety of *Test* populations compared to Han as a baseline. This statistic is significantly negative ($Z < -3.7$) for all ancient Near Easterners as well as Neolithic and later Europeans, consistent with them having ancestry from a deeply divergent Eurasian lineage that separated from the ancestors of most Eurasians before the separation of Han and Ust'-Ishim. We used *qpAdm* (ref. 7) to estimate Basal Eurasian ancestry in each *Test* population. We obtained the highest estimates in the earliest populations from both Iran (66 ± 13% in the likely Mesolithic sample, 48 ± 6% in Neolithic samples), and the Levant (44 ± 8% in Epipalaeolithic Natufians) (Fig. 2), showing that Basal Eurasian ancestry was widespread across the ancient Near East.

West Eurasians harbour significantly less Neanderthal ancestry than East Asians[19–21], which could be explained if West Eurasians (but not East Asians) have partial ancestry from a source that diluted their Neanderthal inheritance[20]. Supporting this theory, we observe a negative correlation between Basal Eurasian ancestry and the rate of shared alleles with Neanderthals[19] (Supplementary Information, section 5 and Fig. 2). By extrapolation, we infer that the Basal Eurasian population had lower Neanderthal ancestry than non-Basal Eurasian populations and possibly none (95% confidence interval truncated at zero of 0–60%; Fig. 2; Methods). The finding of little if any Neanderthal ancestry in Basal Eurasians could be explained if the Neanderthal admixture into modern humans ~50,000–60,000 years ago[11] largely occurred after the splitting of the Basal Eurasians from other non-Africans.

It is striking that the highest estimates of Basal Eurasian ancestry are from the Near East, given the hypothesis that it was there that most admixture between Neanderthals and modern humans occurred[19,22]. This could be explained if Basal Eurasians thoroughly admixed into the Near East before the time of the samples we analysed but after the Neanderthal admixture. Alternatively, the ancestors of Basal Eurasians may have always lived in the Near East, but the lineage of which they were a part did not participate in the Neanderthal admixture.

A population without Neanderthal admixture, basal to other Eurasians, may have plausibly lived in Africa. Craniometric analyses have suggested an affinity between the Natufians and populations of north or sub-Saharan Africa[23,24], a result that finds some support from Y chromosome analysis showing that the Natufians and successor Levantine Neolithic populations carried haplogroup E, likely to be of ultimately African origin, which has not been detected in other ancient males from West Eurasia[7,8] (Supplementary Information, section 6). However, no affinity of Natufians to sub-Saharan Africans is evident in our genome-wide analysis, as present-day sub-Saharan Africans do not share more alleles with Natufians than with other ancient Eurasians (Extended Data Table 1). (We could not test for a link to present-day North Africans, who owe most of their ancestry to back-migration from Eurasia[25,26].) The idea of Natufians as a vector for the movement of Basal Eurasian ancestry into the Near East is also not supported by our data, as the Basal Eurasian ancestry in the Natufians (44 ± 8%) is consistent with stemming from the same population as that in the Neolithic and Mesolithic populations of Iran, and is not greater than in those populations (Supplementary Information, section 4). Further insight into the origins and legacy of the Natufians could come from comparison to Natufians from additional sites, and to ancient DNA from North Africa.

Extreme differentiation in the ancient Near East

PCA on present-day West Eurasian populations (Methods and Extended Data Fig. 1), on which we projected the ancient individuals

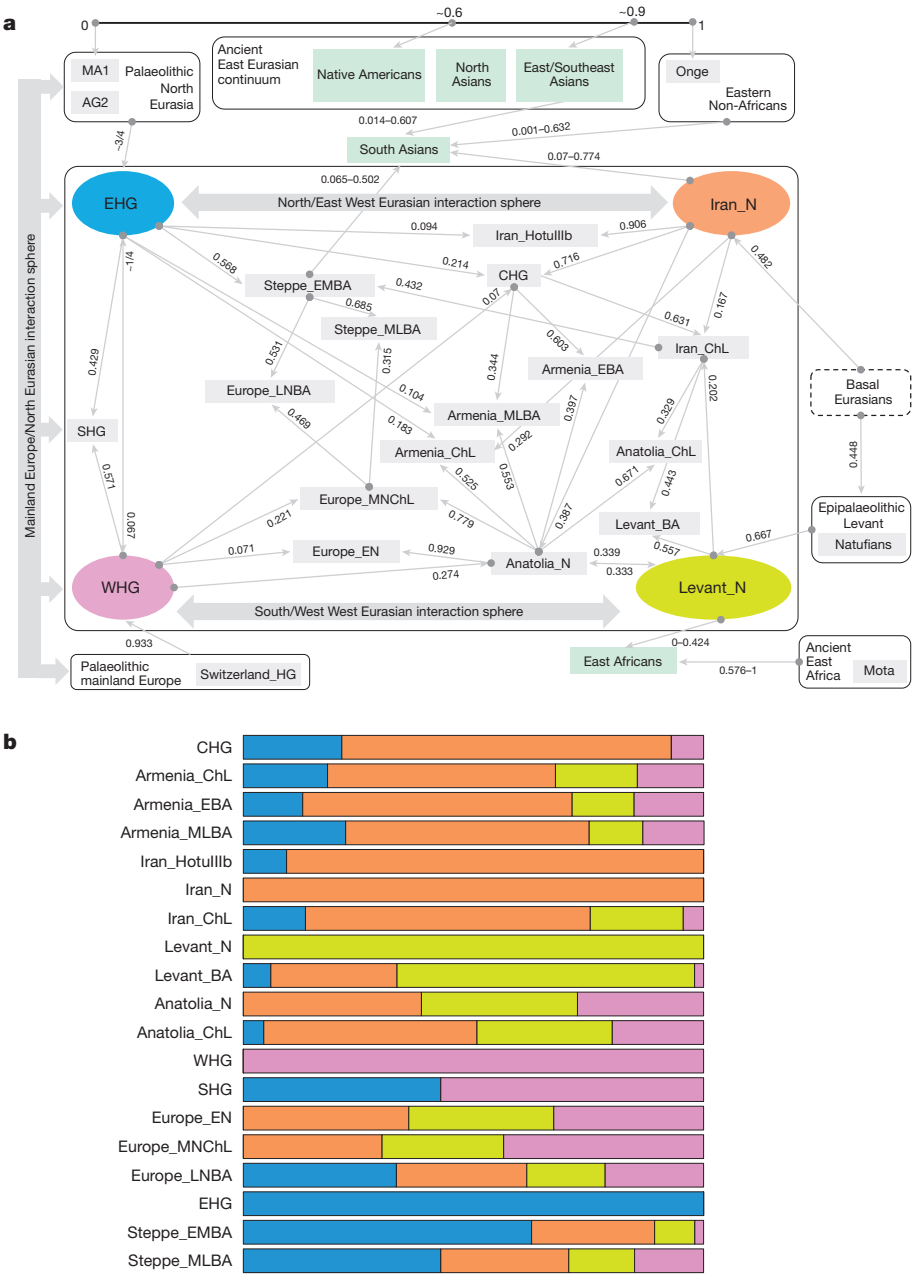


Figure 4 | Modelling ancient West Eurasians, East Africans, East Eurasians and South Asians. a, All the ancient populations can be modelled as mixtures of two or three other populations and up to four proximate sources (marked in colour). Mixture proportions inferred by *qpAdm* are indicated by the incoming arrows to each population. Green rectangles represent sets of more than one population. Multiple admixture solutions are consistent with the data for some populations, and while only one solution is shown here, Supplementary Information, section 7 presents the others. **b**, A flat representation of the graph showing mixture proportions from the four proximate sources.

(Fig. 1b), replicates previous findings of a Europe–Near East contrast along the horizontal principal component 1 (PC1) and parallel clines (PC2) in both Europe and the Near East[7,8,13] (Extended Data Fig. 1). Ancient samples from the Levant clustered at one end of the Near Eastern cline, and ancient samples from Iran at the other. The two Caucasus hunter–gatherers (CHG)[9] are less extreme along PC1 than the Mesolithic and Neolithic individuals from Iran, while individuals from Chalcolithic Anatolia, Iran, Armenia, and Bronze Age Armenia occupy intermediate positions. Qualitatively, the PCA has the appearance of a quadrangle whose four corners are some of the oldest samples: bottom-left, Western hunter–gatherers (WHG); top-left, Eastern hunter–gatherers (EHG); bottom-right, Neolithic Levant and Natufians; top-right, Neolithic Iran. This suggests that diverse ancient West Eurasians can be modelled as mixtures of as few as four streams of ancestry related to these populations, which we confirmed using *qpWave* (ref. 7) (Supplementary Information, section 7).

We computed squared allele frequency differentiation between all pairs of ancient West Eurasians[27] (Methods; Fig. 3 and Extended Data Figs 2b and 4), and found that the populations at the four corners of the quadrangle had differentiation of $F_{ST} = 0.08–0.15$, comparable to the value of 0.09–0.13 seen between present-day West Eurasians and East Asians (Han) (Supplementary Table 3). By contrast, by the Bronze Age, genetic differentiation between pairs of West Eurasian populations had reached its present-day low levels (Fig. 3): today, F_{ST} is ≤ 0.025 for 95% of the pairs of West Eurasian populations and ≤ 0.046 for all pairs (Supplementary Table 3). These results point to a demographic process that established high differentiation across West Eurasia and then reduced this differentiation over time.

Continuity between hunter–gatherers and early farmers

Our data document continuity across the transition between hunter–gatherers and farmers, separately in the southern Levant and in the southern Caucasus–Iran highlands. The qualitative evidence for this is that PCA, ADMIXTURE, and outgroup f_3 analysis cluster Levantine hunter–gatherers (Natufians) with Levantine farmers, and Iranian and CHG with Iranian farmers (Fig. 1b and Extended Data Figs 1, 3). We confirm this in the Levant by showing that its early farmers share significantly more alleles with Natufians than with the early farmers of Iran:

the statistic f_4(Levant_N, Chimp; Natufian, Iran_N) is significantly positive ($Z = 13.6$). The early farmers of the Caucasus–Iran highlands similarly share significantly more alleles with the hunter–gatherers of this region than with the early farmers from the Levant: the statistic f_4(Iran_N, Chimp; Caucasus or Iran highland hunter–gatherers, Levant_N) is significantly positive ($Z > 6$).

Admixture in the ancient Near East

Almost all ancient and present-day West Eurasians have evidence of significant admixture between two or more ancestral populations, as documented by statistics of the form f_3(*Test*; *Reference*$_1$, *Reference*$_2$) which, if negative, show that a test population's allele frequencies tend to be an intermediate between two reference populations[16] (Extended Data Table 2). To better understand the admixture history beyond these patterns, we used *qpAdm* (ref. 7), which can evaluate whether a particular test population is consistent with being derived from a set of proposed source populations, and if so, infer mixture proportions (Methods). We used this approach to carry out a systematic survey of ancient West Eurasian populations to explore their possible sources of admixture (Fig. 4 and Supplementary Information, section 7).

Among first farmers, those of the Levant trace approximately two-thirds of their ancestry to people related to Natufian hunter–gatherers and about one-third to people related to Anatolian farmers (Supplementary Information, section 7). Western Iranian first farmers cluster with the likely Mesolithic HotuIIIb individual and more remotely with hunter–gatherers from the southern Caucasus (Fig. 1b), and share alleles at an equal rate with Anatolian and Levantine early farmers (Supplementary Information, section 7), highlighting the long-term isolation of western Iran.

During subsequent millennia, the early farmer populations of the Near East expanded in all directions and mixed, as we can model populations of the Chalcolithic and subsequent Bronze Age only as having ancestry from two or more sources. The Chalcolithic people of western Iran can be modelled as a mixture of the Neolithic people of western Iran, the Levant and CHG, consistent with their position in the PCA (Fig. 1b). Admixture from populations related to the Chalcolithic people of western Iran had a wide impact, consistent with contributing around 44% of the ancestry of Levantine Bronze Age populations in the south and about 33% of the ancestry of the Chalcolithic North-West Anatolians in the west. Our analysis shows that the ancient populations of Chalcolithic Iran, Chalcolithic Armenia, Bronze Age Armenia and Chalcolithic Anatolia were all composed of the same ancestral components, albeit in slightly different proportions (Fig. 4b and Supplementary Information, section 7).

Admixture into Europe, East Africa and South Asia

Admixture did not only occur within the Near East but also extended towards Europe. To the north, a population related to people of Chalcolithic Iran contributed about 43% of the ancestry of early Bronze Age populations of the steppe. The spread of Near Eastern ancestry into the Eurasian steppe was previously inferred[7] without access to ancient samples, with a population related to present-day Armenians as a suggested source[7,8]. To the west, the early farmers of mainland Europe were descended from a population related to Neolithic North-Western Anatolians[8]. This is consistent with an Anatolian origin of farming in Europe, but does not reject other sources, as the spatial distribution of the Anatolian/European-like farmer populations is unknown. We can rule out the hypothesis that European farmers stem directly from a population related to the ancient farmers of the southern Levant[28,29], however, because European farmers share more alleles with Anatolian Neolithic farmers than with Levantine farmers, as attested by the positive statistic f_4(Europe_EN, Chimp; Anatolia_N, Levant_N) ($Z = 15$).

Migration from the Near East also occurred towards the southwest into East African populations, which experienced West Eurasian admixture around 1,000 BC[30,31]. Previously, the West Eurasian population known to be the best proxy for this ancestry was present-day Sardinians[31], who resemble Neolithic Europeans genetically[13,32]. However, our analysis shows that East African ancestry is significantly better modelled by Levantine early farmers than by Anatolian or early European farmers, implying that the spread of this ancestry to East Africa was not from the same group that spread Near Eastern ancestry into Europe (Extended Data Fig. 5 and Supplementary Information, section 8).

In South Asia, our dataset provides insight into the sources of Ancestral North Indians (ANI), a West Eurasian-related population that no longer exists in unmixed form but contributes a variable amount of the ancestry of South Asians[33,34] (Supplementary Information, section 9 and Extended Data Fig. 5). We show that it is impossible to model the ANI as being derived from any single ancient population in our dataset. However, it can be modelled as a mix of ancestries related to both early farmers of western Iran and people of the Bronze Age Eurasian steppe; all sampled South Asian groups are inferred to have significant amounts of both ancestral types. The demographic impact of steppe-related populations on South Asia was substantial, as the Mala, a south Indian population with minimal ANI along the 'Indian Cline' of such ancestry[33,34], is inferred to have around 18% steppe-related ancestry, while the Kalash of Pakistan are inferred to have about 50%, similar to present-day northern Europeans[7].

Population transformations in West Eurasia and beyond

We were concerned that our conclusions might be biased by the particular populations we happened to sample, and that we would have obtained qualitatively different conclusions without data from some key populations. We tested our conclusions by plotting the inferred position of admixed populations in PCA against a weighted combination of their inferred source populations and obtained qualitatively consistent results (Extended Data Fig. 6).

To further assess the robustness of our inferences, we developed a method to infer the existence and genetic affinities of ancient populations from unobserved 'ghost' populations (Supplementary Information, section 10 and Extended Data Fig. 7). This method takes advantage of the insight that if an unsampled ghost population admixes with differentiated 'substratum' populations, it is possible to extrapolate its identity by intersecting clines of populations with variable proportions of ghost and substratum ancestry. Applying this approach while withholding major populations, we validated some of our key inferences, successfully inferring mixture proportions consistent with those obtained when the populations were included in the analysis. Application of this method highlights the impact of Ancient North Eurasian (ANE) ancestry related to the ~22,000 BC Mal'ta 1 and ~15,000 BC Afontova Gora 2 (ref. 15) on populations living in Europe, the Americas and Eastern Eurasia. Eastern Eurasians can be modelled as arrayed along a cline with different proportions of ANE ancestry (Supplementary Information, section 11 and Extended Data Fig. 8), ranging from about 40% ANE in Native Americans, matching previous findings[13,15], to no less than around 5–10% ANE in diverse East Asian groups including Han Chinese (Extended Data Figs 5, 7f). We also document a cline of ANE ancestry across the East–West extent of Eurasia. Eastern hunter-gatherers (EHG) derive about three-quarters of their ancestry from the ANE (Supplementary Information, section 11); Scandinavian hunter–gatherers[7,8,13] (SHG) are a mix of EHG and WHG; and WHG are a mix of EHG and populations related to the Upper Palaeolithic Bichon from Switzerland (Supplementary Information, section 7). Northwest Anatolians—with ancestry from a population related to European hunter–gatherers (Supplementary Information, section 7)—are better modelled if this ancestry is taken as more extreme than Bichon (Supplementary Information, section 10).

The population structure of the ancient Near East was not independent of that of Europe (Supplementary Information, section 4), as evidenced by the highly significant ($Z = -8.9$) statistic f_4(Iran_N, Natufian;WHG, EHG) which suggests gene flow in 'northeastern' (Neolithic Iran/EHG) and 'southwestern' (Levant/WHG) interaction spheres (Fig. 4d). This interdependence of the ancestry of Europe and

the Near East may have been mediated by unsampled geographically intermediate populations[35] that contributed ancestry to both regions.

Conclusions

By analysing genome-wide ancient DNA data from ancient individuals from the Levant, Anatolia, the southern Caucasus and Iran, we have provided a first glimpse into the demographic structure of the human populations that transitioned to farming. We reject the hypothesis that the spread of agriculture in the Near East was achieved by the dispersal of a single farming population displacing the hunter–gatherers they encountered. Instead, the spread of ideas and farming technology moved faster than the spread of people, as we can determine from the fact that the population structure of the Near East was maintained throughout the transition to agriculture. A priority for future ancient DNA studies should be to obtain data from older periods, which would reveal the deeper origins of the population structure in the Near East. It will also be important to obtain data from the ancient civilizations of the Near East to bridge the gap between the region's prehistoric inhabitants and those of the present.

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 8 March; accepted 18 July 2016.
Published online 25 July 2016.

1. Barker, G. & Goucher, C. *The Cambridge World History Volume II: A World with agriculture, 12,000 BCE–500 CE* (Cambridge Univ. Press, 2015).
2. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes.* (Princeton Univ. Press, 1994).
3. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5,** 5257 (2014).
4. Pinhasi, R. *et al.* Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS One* **10,** e0129102 (2015).
5. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110,** 2223–2227 (2013).
6. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524,** 216–219 (2015).
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. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528,** 499–503 (2015).
9. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6,** 8912 (2015).
10. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522,** 167–172 (2015).
11. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514,** 445–449 (2014).
12. Günther, T. *et al.* Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc. Nat. Acad Sci. USA* (2015).
13. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,** 409–413 (2014).
14. 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).
15. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505,** 87–91 (2014).
16. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192,** 1065–1093 (2012).
17. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
18. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).
19. Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–49 (2014).
20. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338,** 222–226 (2012).
21. Wall, J. D. *et al.* Higher levels of neanderthal ancestry in East Asians than in Europeans. *Genetics* **194,** 199–209 (2013).
22. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328,** 710–722 (2010).
23. Brace, C. L. *et al.* The questionable contribution of the Neolithic and the Bronze Age to European craniofacial form. *Proc. Natl Acad. Sci. USA* **103,** 242–247 (2006).
24. Ferembach, D. Squelettes du Natoufien d'Israel., etude anthropologique. *Anthropologie* **65,** 46–66 (1961).
25. Fadhlaoui-Zid, K. *et al.* Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa. *PLoS One* **8,** e80293 (2013).
26. Henn, B. M. *et al.* Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet.* **8,** e1002397 (2012).
27. Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting FST: the impact of rare variants. *Genome Res.* **23,** 1514–1521 (2013).
28. Fernández, E. *et al.* Ancient DNA analysis of 8000 B.C. near eastern farmers supports an early neolithic pioneer maritime colonization of Mainland Europe through Cyprus and the Aegean Islands. *PLoS Genet.* **10,** e1004401 (2014).
29. Ammerman, A. J., Pinhasi, R. & Banffy, E. Comment on Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* **312,** 1875; author reply 1875 (2006).
30. Pagani, L. *et al.* Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am. J. Hum. Genet.* **91,** 83–96 (2012).
31. Pickrell, J. K. *et al.* Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl Acad. Sci. USA* **111,** 2632–2637 (2014).
32. 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).
33. Moorjani, P. *et al.* Genetic evidence for recent population mixture in India. *Am. J. Hum. Genet.* **93,** 422–438 (2013).
34. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461,** 489–494 (2009).
35. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534,** 200–205 (2016).

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

Acknowledgements We thank the 238 human subjects who donated samples for genome-wide analysis, and D. Labuda and P. Zalloua for sharing samples from Poland and Lebanon. The Fig. 1a map was plotted in R using the worldHiRes map of the 'mapdata' package (using public domain data from the CIA World Data Bank II). We thank O. Bar-Yosef, M. Bonogofsky, I. Hershkowitz, M. Lipson, I. Mathieson, H. May, R. Meadow, I. Olalde, S. Paabo, P. Skoglund, and N. Nakatsuka for comments and critiques, and D. Bradley, M. Dallakyan, S. Esoyan, M. Ferry and M. Michel, and A. Yesayan, for contributions to bone preparation and ancient DNA work. D.F. and M.N. were supported by Irish Research Council grants GOIPG/2013/36 and GOIPD/2013/1, respectively. S.C. was funded by the Irish Research Council for Humanities and Social Sciences (IRCHSS) ERC Support Programme. Q.F. was funded by the Bureau of International Cooperation of the Chinese Academy of Sciences, the National Natural Science Foundation of China (L1524016) and the Chinese Academy of Sciences Discipline Development Strategy Project (2015-DX-C-03). The Scottish diversity data was funded by the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6), the Scottish Funding Council (HR03006), and a project grant from the Scottish Executive Health Department, Chief Scientist Office (CZB/4/285). M.S., A.Tön., M.B. and P.K. were supported by the German Research Foundation (CRC 1052; B01, B03, C01). M.S.-P. was funded by a Wenner-Gren Foundation Dissertation Fieldwork grant (9005), and by the National Science Foundation DDRIG (BCS-1455744). P.K. was funded by the Federal Ministry of Education and Research, Germany (FKZ: 01EO1501). J.F.W. acknowledge the MRC 'QTL in Health and Disease' programme grant. The Romanian diversity data was supported by the EC Commission, Directorate General XII (Supplementary Agreement ERBCIPDCT 940038 to the Contract ERBCHRXCT 920032, coordinated by A. Piazza, Turin, Italy). M.R. received support from the Leverhulme Trust's Doctoral Scholarship programme. O.S. and A.Tor. were supported by the University of Pavia (MIGRAT-IN-G) and the Italian Ministry of Education, University and Research: Progetti Ricerca Interesse Nazionale 2012. The Raqefet Cave Natufian project was supported by funds from the National Geographic Society (grant 8915-11), the Wenner-Gren Foundation (grant 7481) and the Irene Levi-Sala CARE Foundation, while radiocarbon dating on the samples was funded by the Israel Science Foundation (grant 475/10; E. Boaretto). R.P. was supported by ERC starting grant ADNABIOARC (263441). D.R. was supported by NIH grant GM100233, by NSF HOMINID BCS-1032255, and is a Howard Hughes Medical Institute investigator.

Author Contributions R.P. and D.R. conceived the idea for the study. D.N., G.R., D.C.M., S.C., S.A.R., G.L., F.B., B.Gas., J.M.M., M.G., V.E., A.M., C.M., F.G., N.A.H. and R.P. assembled skeletal material. N.R., D.F., M.N., B.Gam., K.Si., S.C., K.St., E.H., Q.F., G.G.-F., E.R.J., R.P. and D.R. performed or supervised ancient DNA wet laboratory work. L.B, M.B., A.C., G.C., D.C., P.F., E.G., S.M.K., P.K., J.K., D.M., M.M., D.A.M., S.O., M.B.R., O.S., M.S.-P., G.S., M.S., A.Tön., A.Tor., J.F.W., L.Y. and D.R. assembled present-day samples for genotyping. I.L, N.P. and D.R. developed methods for data analysis. I.L., S.M., Q.F., N.P. and D.R. analysed data. I.L., R.P. and D.R. wrote the manuscript and supplements.

Author Information The aligned sequences are available through the European Nucleotide Archive under accession number PRJEB14455. Fully public subsets of the analysis datasets are at http://genetics.med.harvard.edu/reichlab/Reich_Lab/Datasets.html. The complete dataset (including present-day humans for which the informed consent is not consistent with public posting of data) is available to researchers who send a signed letter to D.R. indicating that they will abide by specified usage conditions (Supplementary Information, section 2). 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. Correspondence and requests for materials should be addressed to I.L. (lazaridis@genetics.med.harvard.edu) or R.P. (ron.pinhasi@ucd.ie) or D.R. (reich@genetics.med.harvard.edu).

Reviewer Information *Nature* thanks O. Bar-Yosef, G. Coop and the other anonymous reviewer(s) for their contribution to the peer review of 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 data. In a dedicated ancient DNA laboratory at University College Dublin, we prepared powder from 132 ancient Near Eastern samples, either by dissecting the inner ear region of the petrous bone using a sandblaster (Renfert), or by drilling using a Dremel tool and single-use drill bits and selecting the best preserved bone fragments based on anatomical criteria. These fragments were then powdered using a mixer mill (Retsch Mixer Mill 400)[4].

We performed all subsequent processing steps in a dedicated ancient DNA laboratory at Harvard Medical School, where we extracted DNA from the powder (usually 75 mg, range 14–81 mg) using an optimized ancient DNA extraction protocol[36], but replaced the assembly of Qiagen MinElute columns and extension reservoirs from Zymo Research with a High Pure Extender Assembly from the High Pure Viral Nucleic Acid Large Volume Kit (Roche Applied Science). We built a total of 170 barcoded double-stranded Illumina sequencing libraries for these samples[37], of which we treated 167 with uracil-DNA glycosylase (UDG) to remove the characteristic C-to-T errors of ancient DNA[38]. The UDG treatment strategy is (by–design) inefficient at removing terminal uracils, allowing the mismatch rate to the human genome at the terminal nucleotide to be used for authentication[37]. We updated this library preparation protocol in two ways compared to the original publication: first, we used 16U Bst2.0 Polymerase, Large Fragment (NEB) and 1× Isothermal amplification buffer (NEB) in a final volume of 25 μl fill-in reaction, and second, we used the entire inactivated 25 μl fill-in reaction in a total volume of 100 μl PCR mix with 1 μM of each primer[39]. We included extraction negative controls (where no sample powder was used) and library negative controls (where extract was supplemented by water) in every batch of samples processed and carried them through the entire wet laboratory processing to test for reagent contamination.

We screened the libraries by hybridizing them in solution to a set of oligonucleotide probes tiling the mitochondrial genome[40], using the protocol described previously[7]. We sequenced the enriched libraries using an Illumina NextSeq 500 instrument using 2× 76 bp reads, trimmed identifying sequences (seven base pair molecular barcodes at either end) and any trailing adapters, merged read pairs that overlapped by at least 15 base pairs, and mapped the merged sequences to the RSRS mitochondrial DNA reference genome[41], using the Burrows Wheeler Aligner[42] (*bwa*) and the command *samse* (v0.6.1).

We enriched promising libraries for a targeted set of ∼1.2 million SNPs[8] as in ref. 5, and adjusted the blocking oligonucleotide and primers to be appropriate for our libraries. The specific probe sequences are given in supplementary data 2 of ref. 7. and supplementary data 1 of ref. 6. We sequenced the libraries on an Illumina NextSeq 500 using 2× 76 bp reads. We trimmed identifying sequences (molecular barcodes) and any trailing adapters, merged pairs that overlapped by at least 15 base pairs (allowing up to one mismatch), and mapped the merged sequences to *hg19* using the single-ended aligner *samse* in bwa (v0.6.1). We removed duplicated sequences by identifying sets of sequences with the same orientation and start and end positions after alignment to *hg19*; we picked the highest quality sequence to represent each set. For each sample, we represented each SNP position by a randomly chosen sequence, restricting to sequences with a minimum mapping quality (MAPQ ≥ 10), sites with a minimum sequencing quality (≥20), and removing two bases at the ends of reads. We sequenced the enriched products up to the point that we estimated that generating a hundred new sequences was expected to add data on less than about one new SNP[8].

Testing for contamination and quality control. For each ancient DNA library, we evaluated authenticity in several ways. First, we estimated the rate of matching to the consensus sequence for mitochondrial genomes sequenced to a coverage of at least tenfold from the initial screening data. Of the 76 libraries that contributed to our dataset (coming from 45 samples), 70 had an estimated rate of sequencing matching to the consensus of >95% according to contamMix[5] (the remaining libraries had estimated match rates of 75–92%, but gave no sign of being outliers in principal component analysis or X-chromosome contamination analysis so we retained them for analysis) (Supplementary Table 1). We quantified the rate of C-to-T substitution in the final nucleotide of the sequences analysed, relative to the human reference genome sequence, and found that all the libraries analysed had rates of at least 3% (ref. 37), consistent with genuine ancient DNA. For the nuclear data from males, we used the ANGSD software[43] to obtain a conservative X-chromosome estimate of contamination. We determined that all libraries that passed our quality control and for which we had sufficient X-chromosome data to make an assessment, had contamination rates of 0–1.5%. Finally, we merged data for samples for which we had multiple libraries to produce an analysis dataset.

Affymetrix Human Origins genotyping data. We genotyped 238 present-day individuals from 17 diverse West Eurasian populations on the Affymetrix Human Origins array[16], and applied quality control analyses as previously described[13] (Supplementary Table 2). We merged the newly generated data with data from 2,345 individuals previously genotyped on the same array[13]. All individuals that were genotyped provided individual informed consent consistent with studies of population history, following protocols approved by the ethical review committees of the institutions of the researchers who collected the samples. The collection and analysis of genome-wide data on anonymized samples at Harvard Medical School for the purpose of studying population history was approved by the Harvard Human Research Protection Program, protocol 11681, re-reviewed on 12 July 2016. Anonymized aliquots of DNA from all individuals were sent to the core facility of the Center for Applied Genomics at the Children's Hospital of Philadelphia for genotyping and data processing. For 127 of the individuals with newly reported data, the informed consent was consistent with public distribution of data, and the data can be downloaded at http://genetics.med.harvard.edu/reich/Reich_Lab/Datasets.html. To access data for the remaining 111 newly reported samples, researchers should send a signed letter to D.R. containing the following text: "(a) I will not distribute the data outside my collaboration; (b) I will not post the data publicly; (c) I will make no attempt to connect the genetic data to personal identifiers for the samples; (d) I will use the data only for studies of population history; (e) I will not use the data for any selection studies; (f) I will not use the data for medical or disease-related analyses; (g) I will not use the data for commercial purposes." Supplementary Table 2 specifies which samples are consistent with which type of data distribution.

Datasets. We carried out population genetic analysis on two datasets: (i) *HO* includes 2,583 present-day humans genotyped on the Human Origins array[13,16] including 238 newly reported, (Supplementary Table 2; Supplementary Information, section 2), and 281 ancient individuals on a total of 592,146 autosomal SNPs. (ii) *HOIll* includes the 281 ancient individuals on a total of 1,055,186 autosomal SNPs, including those present in both the Human Origins and Illumina genotyping platforms, but excluding SNPs on the sex chromosomes or additional SNPs of the 1,240k capture array that were included because of their potential functional importance[8]. We used *HO* for analyses that involve both ancient and present-day individuals, and *HOIll* for analysis on ancient individuals alone. We also used 235 individuals from Pagani *et al.*[30] genotyped at 418,700 autosomal SNPs to study admixture in East Africans (Supplementary Information, section 8). Ancient individuals are represented in 'pseudo-haploid' form by randomly choosing one allele for each position of the array.

Principal components analysis. We carried out principal components analysis in the *smartpca* program of EIGENSOFT[17], using default parameters and the lsqproject: YES[13] and numoutlieriter: 0 options. We carried out PCA on the *HO* dataset for 991 present-day West Eurasians (Extended Data Fig. 1), and projected the 278 ancient individuals (Fig. 1b).

ADMIXTURE analysis. We carried out ADMIXTURE analysis[18] of the *HO* dataset after pruning for linkage disequilibrium in PLINK[44,45] with parameters indep-pairwise 200 25 0.4, which retained 296,309 SNPs. We performed analysis in 20 replicates with different random seeds, and retained the highest likelihood replicate for each value of *K*. We show the $K = 11$ results for the 281 ancient samples in Extended Data Fig. 2a (this is the lowest *K* for which components maximized in European hunter–gatherers, ancient Levant, and ancient Iran appear).

***f*-statistics.** We carried out analysis of f_3-statistics, f_4-ratio, and f_4-statistics statistics using the ADMIXTOOLS[16] programs *qp3Pop*, *qpF4ratio* with default parameters, and *qpDstat* with f4mode: YES, and computed standard errors with a block jack-knife[46]. For computing f_3-statistics with an ancient population as a target, we set the inbreed: YES parameter. We computed *f*-statistics on the *HOIll* dataset when no present-day humans were involved and on the *HO* dataset when they were. We computed the statistic f_4(*Test*, Mbuti; Altai, Denisovan) in Fig. 2 on the *HOIll* dataset after merging with whole genome data on 3 Mbuti individuals from Panel C of the Simons Genome Diversity Project[47]. We computed the dendrogram of Extended Data Fig. 3 showing hierarchical clustering of populations with outgroup f_3-statistics using the open source *heatmap.2* function of the *gplots* package in *R*.

Negative correlation of Basal Eurasian ancestry with Neanderthal ancestry. We used the *lm* function of R to fit a linear regression of the rate of allele sharing of a *Test* population with the Altai Neanderthal as measured by f_4(Test, Mbuti; Altai, Denisovan) as the dependent variable, and the proportion of Basal Eurasian ancestry (Supplementary Information, section 4) as the predictor variable. Extrapolating from the fitted line, we obtain the value of the statistic expected if *Test* is a population of 0% or 100% Basal Eurasian ancestry. We then compute the ratio of the Neanderthal ancestry estimate in Basal Eurasians relative to non-Basal Eurasians as f_4(100% Basal Eurasian, Mbuti; Altai, Denisovan)/ f_4(0% Basal Eurasian, Mbuti; Altai, Denisovan). We use a block jack-knife[46], dropping one of 100 contiguous blocks of the genome at a time, to estimate the value and standard error of this quantity (9 ± 26%). We compute a 95% confidence interval based on the point estimate ± 1.96-times the standard error: −42 to 60%. We truncated to 0–60%

on the assumption that Basal Eurasians had no less Neanderthal admixture than Mbuti from sub-Saharan Africa.

Estimation of F_{ST} coefficients. We estimated F_{ST} in *smartpca*[17] with default parameters, inbreed: YES, and fstonly: YES.

Admixture graph modelling. We carried out Admixture Graph modelling with the *qpGraph* software[16] using Mbuti as an outgroup unless otherwise specified.

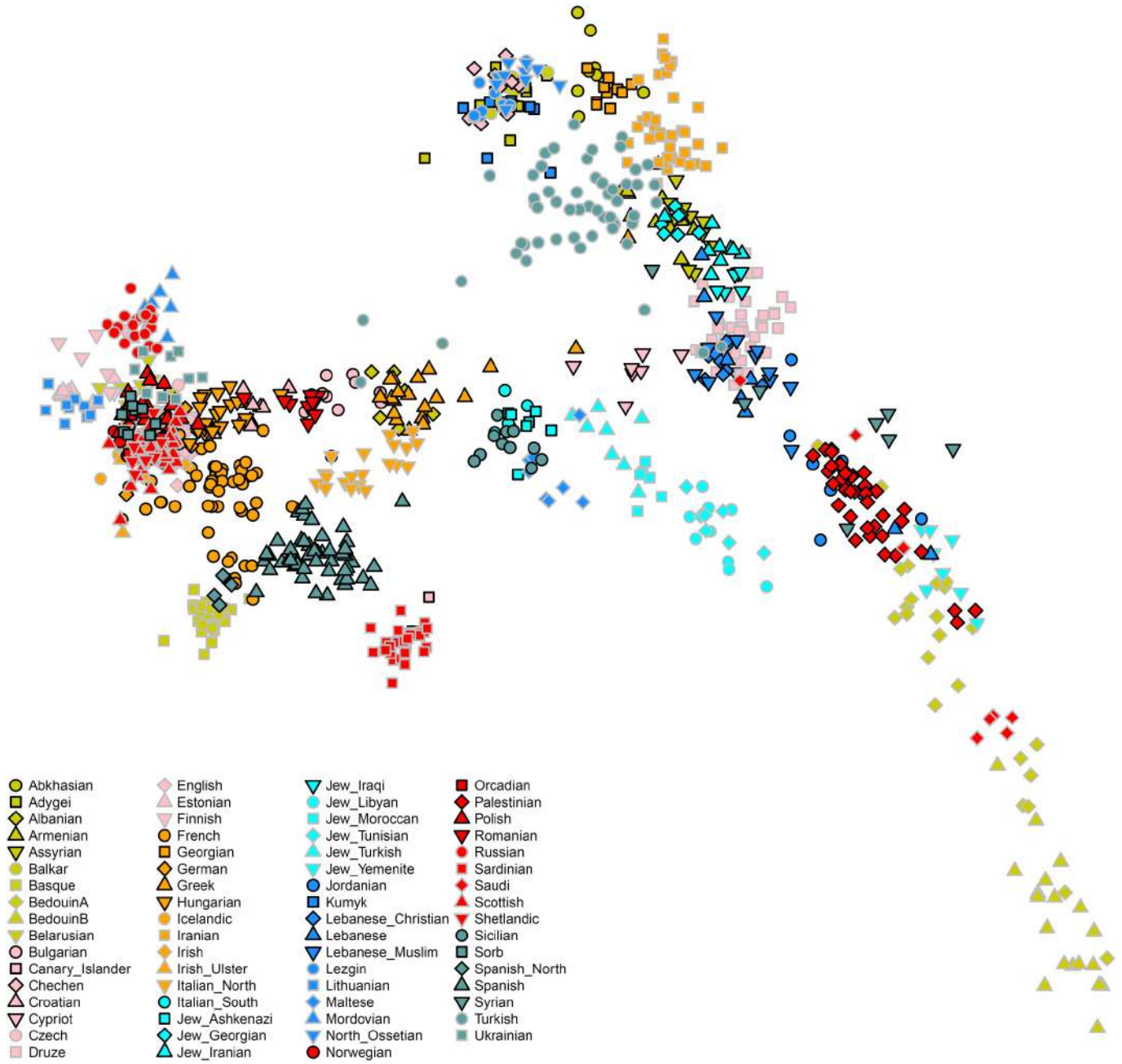
Testing for the number of streams of ancestry. We used the *qpWave*[33,48] software, described in Supplementary Information, section 10 of ref. 7, to test whether a set of 'Left' populations is consistent with being related via as few as N streams of ancestry to a set of 'Right' populations by studying statistics of the form $X(u, v) = F_4(u_0, u; v_0, v)$ where u_0, v_0 are basis populations chosen from the 'Left' and 'Right' sets and u, v are other populations from these sets. We use a Hotelling's T^2 test[48] to evaluate whether the matrix of size $(L-1)*(R-1)$, where L, R are the sizes of the 'Left' and 'Right' sets has rank m. If this is the case, we can conclude that the 'Left' set is related via at least $N = m+1$ streams of ancestry differently to the 'Right' set. We use the parameter allsnps: YES which computes each f_4-statistic based on the full set of SNPs with coverage among the four populations used in the statistic (without regard to whether the SNPs are covered in the other populations in the 'Left' and 'Right' sets).

Inferring mixture proportions without an explicit phylogeny. We used the *qpAdm* methodology described in Supplementary Information, section 10 of ref. 7 to estimate the proportions of ancestry in a *Test* population deriving from a mixture of N 'reference' populations by exploiting (but not explicitly modelling) shared genetic drift with a set of 'Outgroup' populations (Supplementary Information, section 7). We set the details: YES parameter, which reports a normally distributed Z-score estimated with a block jack-knife for the difference between the statistics $f_4(u_0, Test; v_0, v)$ and $f_4(u_0, Estimated\ Test; v_0, v)$ where *Estimated Test* is $\sum_{i=1}^{N} \alpha_i f_4(u_0, Ref_i; v_0, v)$, the average of these f_4-statistics weighed by the mixture proportions α_i from the N reference populations. We use the allsnps: YES parameter.

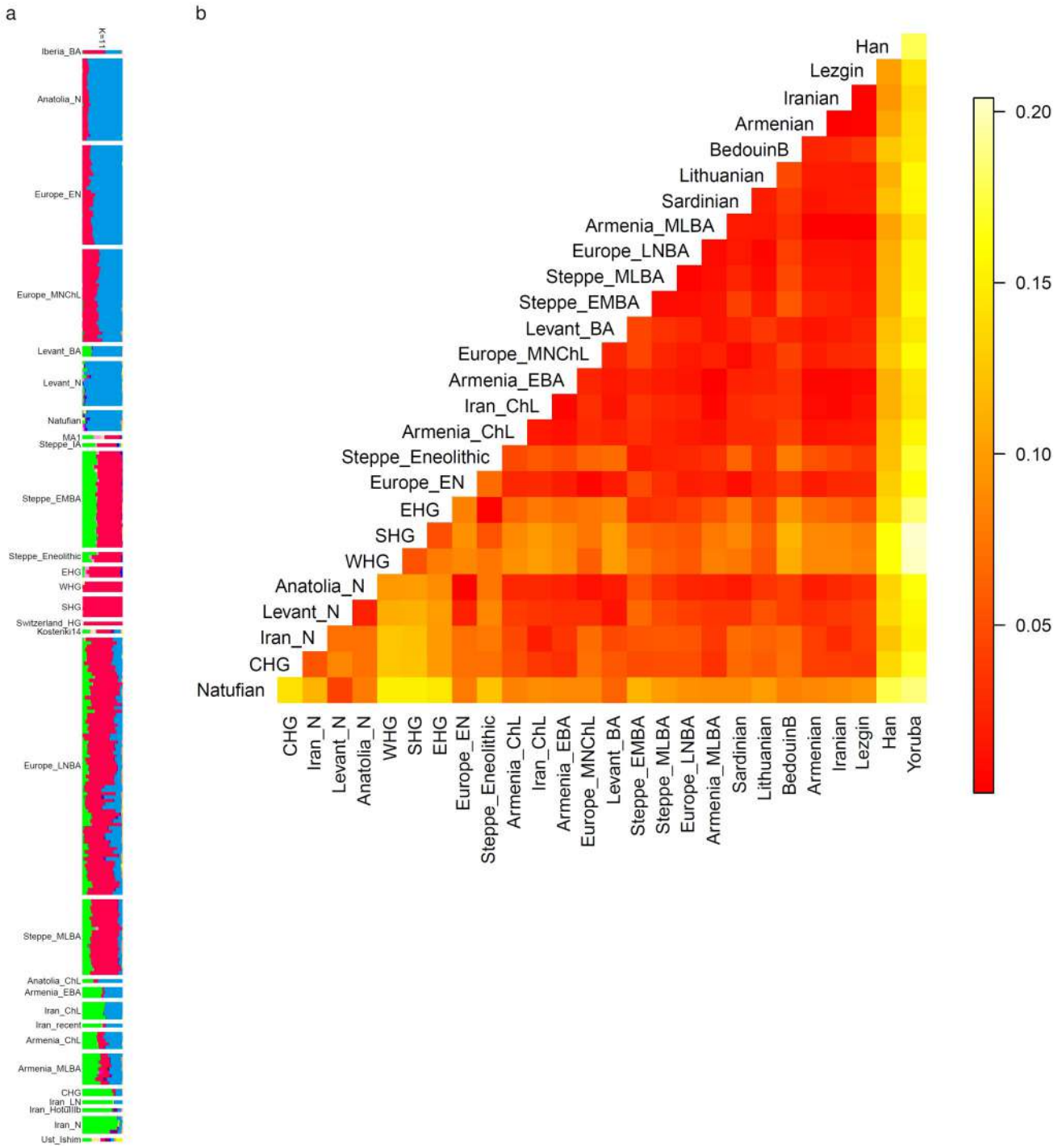
Modelling admixture from ghost populations. We model admixture from a 'ghost' (unobserved) population X in the specific case that X has part of its ancestry from two unobserved ancestral populations p and q. Any population X composed of the same populations p and q resides on a line defined by two observed reference populations r_1 and r_2 composed of the same elements p and q according to a parametric equation $x = r_1 + \lambda(r_2 - r_1)$ with real-valued parameter λ. We define and solve the optimization problem of fitting λ and obtain mixture proportions (Supplementary Information, section 10).

Code availability. Code implementing the newly developed method for modelling admixture from ghost populations is available on request from I.L.

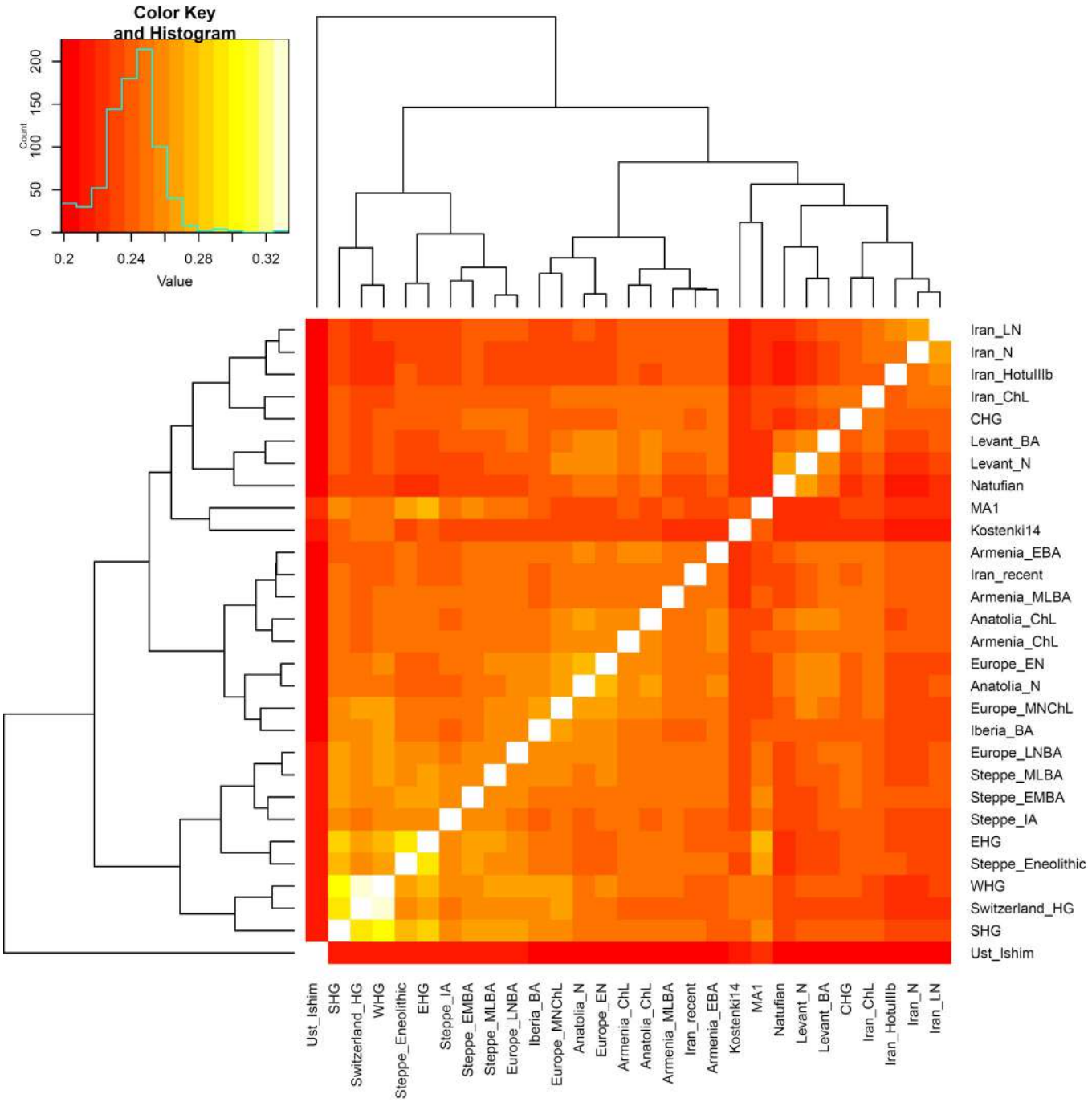
36. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110,** 15758–15763 (2013).
37. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. Lond. B* **370,** 20130624 (2015).
38. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38,** e87 (2010).
39. Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59,** 87–93 (2015).
40. Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505,** 403–406 (2014).
41. Behar, D. M. *et al.* A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90,** 675–684 (2012).
42. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
43. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15,** 356 (2014).
44. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
45. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4,** 7 (2015).
46. Busing, F. T. A., Meijer, E. & Leeden, R. Delete-m Jackknife for Unequal m. *Stat. Comput.* **9,** 3–8 (1999).
47. Sudmant, P. H. *et al.* Global diversity, population stratification, and selection of human copy-number variation. *Science* **349,** aab3761 (2015).
48. Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488,** 370–374 (2012).
49. Gallego Llorente, M. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* **350,** 820–822 (2015).



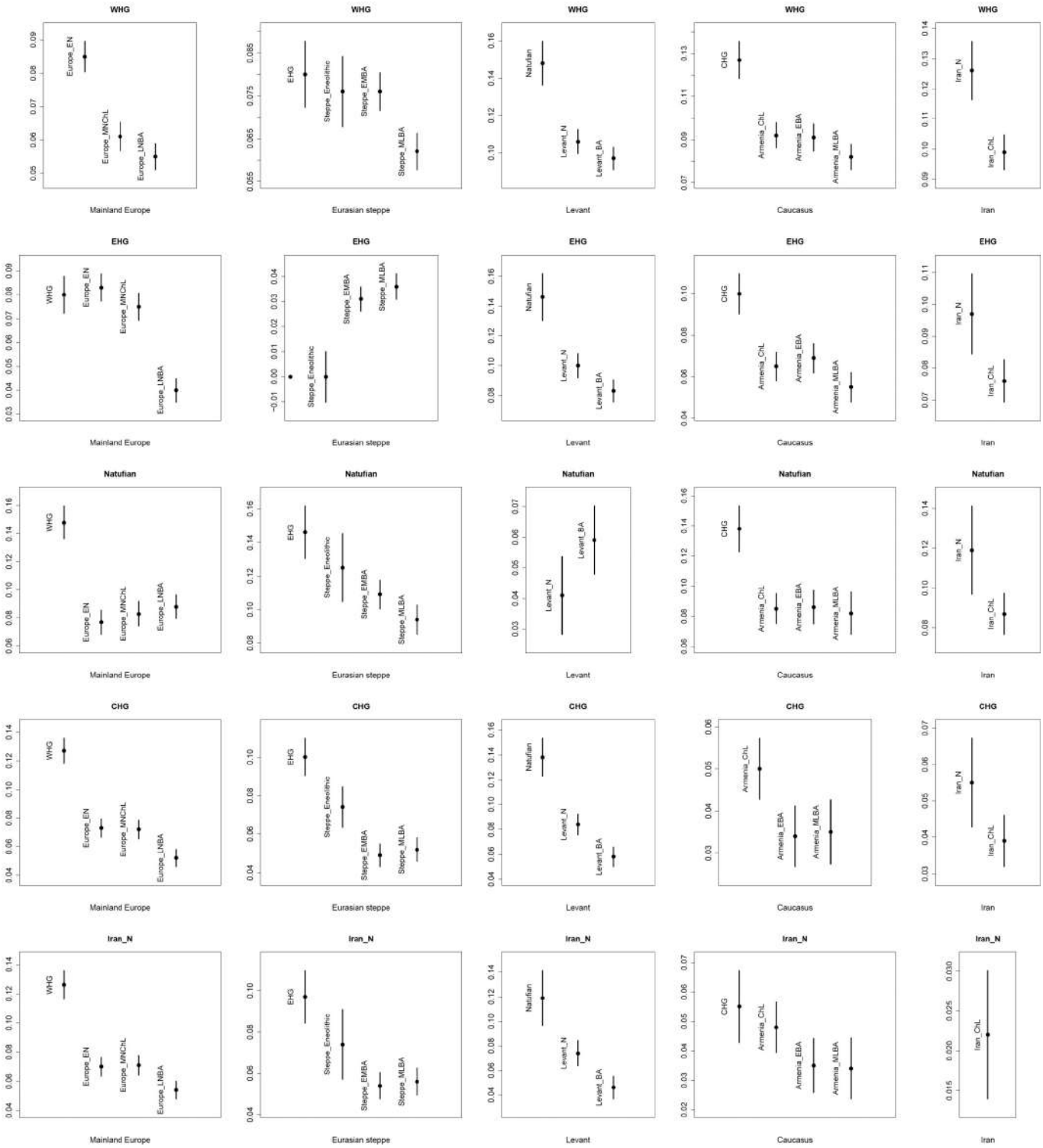
Extended Data Figure 1 | Principal components analysis of 991 present-day West Eurasians. The PCA analysis is performed on the same set of individuals as are reported in Fig. 1b, using EIGENSOFT. Here, we colour the samples by population (to highlight the present-day populations) instead of using grey points as in Fig. 1b (where the goal is to highlight ancient samples).



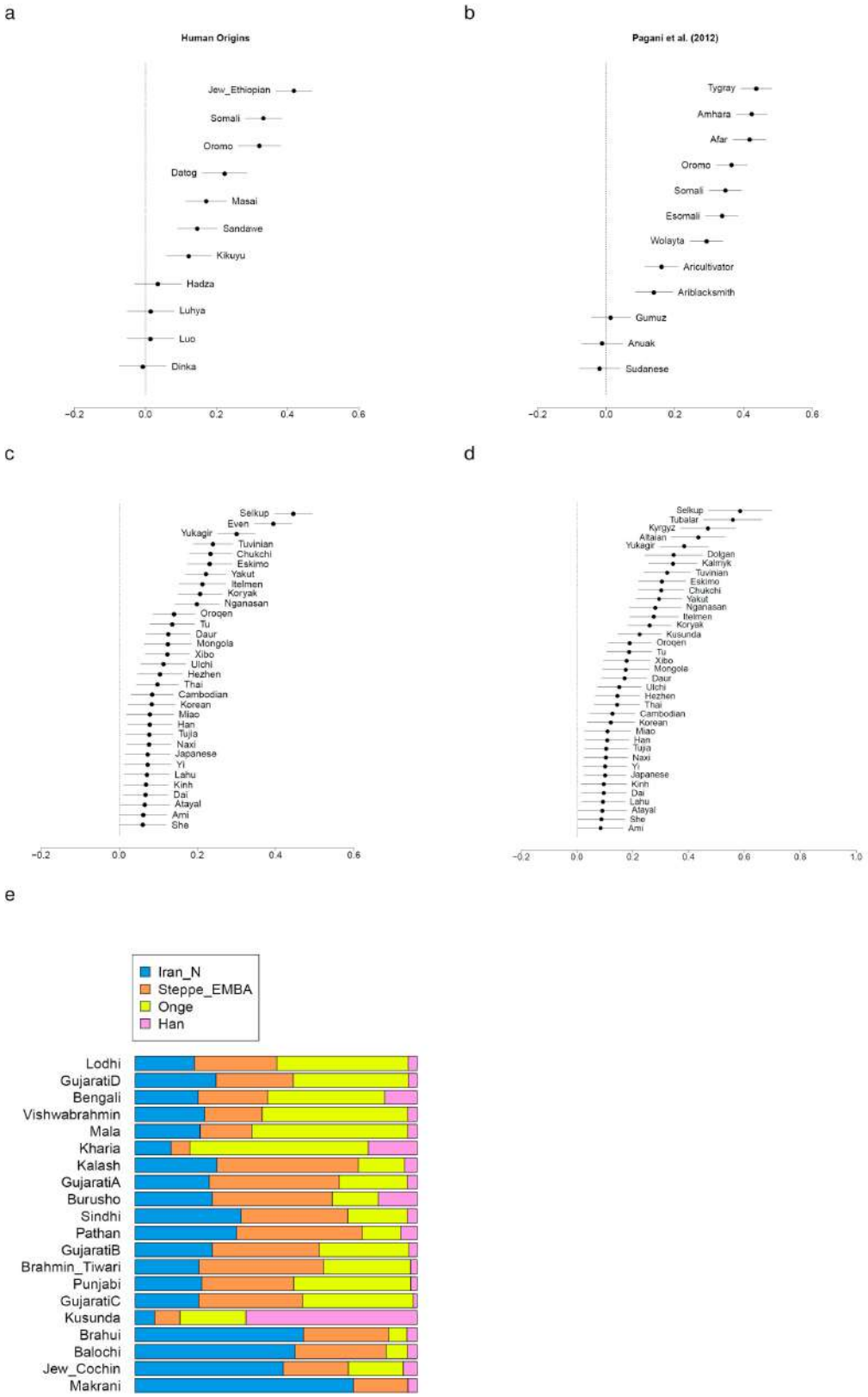
Extended Data Figure 2 | Genetic structure in ancient West Eurasian populations across time and decline of genetic differentiation over time. a, ADMIXTURE model-based clustering analysis of 2,583 present-day humans and 281 ancient samples; we show the results only for ancient samples for $K = 11$ clusters. **b**, Pairwise F_{ST} between 19 Ancient West Eurasian populations (arranged in approximate chronological order), and select present-day populations.



Extended Data Figure 3 | Outgroup f_3(Mbuti; X, Y) for pairs of ancient populations. The dendrogram is plotted for convenience and should not be interpreted as a phylogenetic tree. Areas of high shared genetic drift are 'yellow' and include from top-right to bottom-left along the diagonal: early Anatolian and European farmers; European hunter–gatherers, Steppe populations and populations admixed with steppe ancestry; populations from the Levant from the Epipalaeolithic (Natufians) to the Bronze Age; populations from Iran from the Mesolithic to the Late Neolithic.

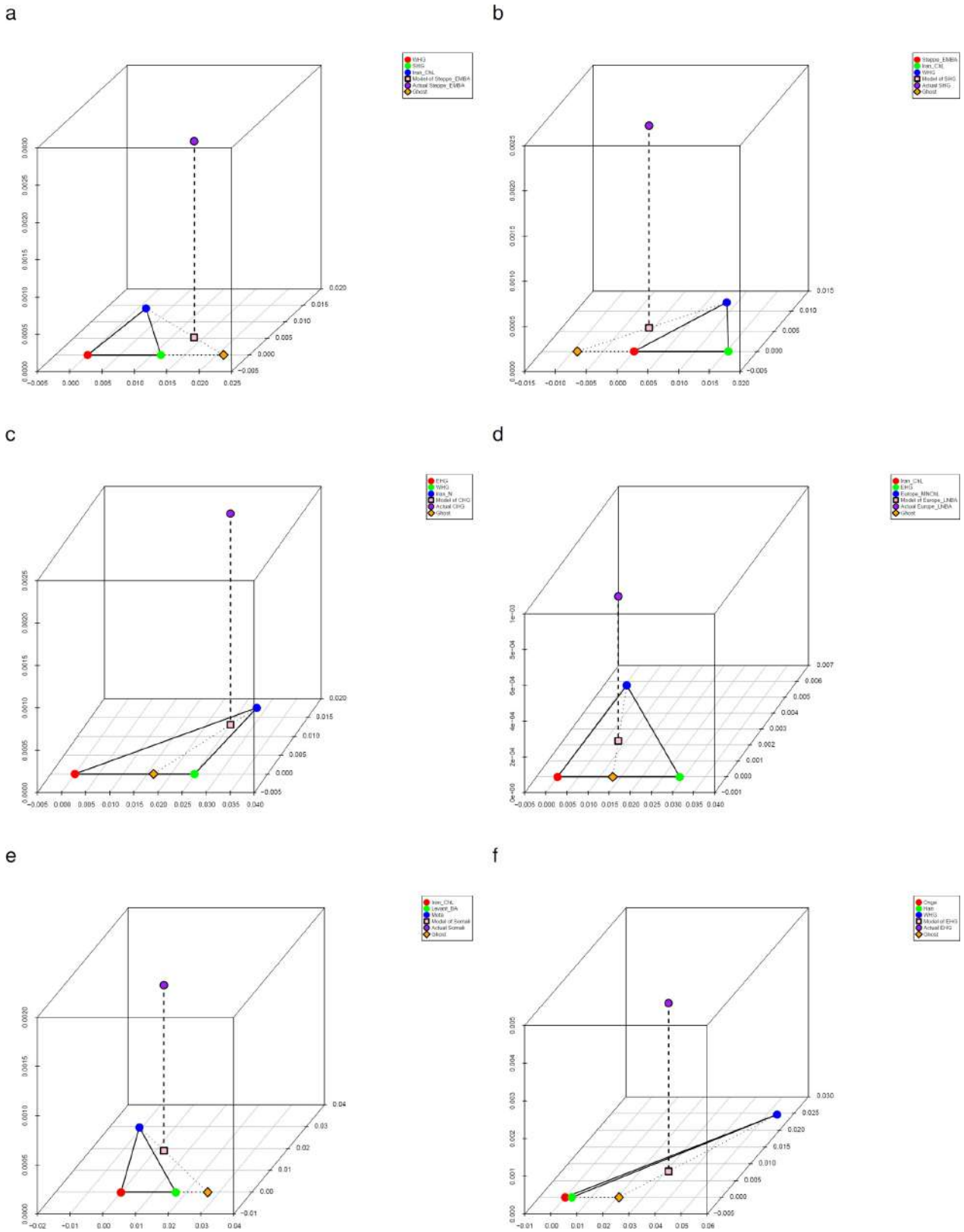


Extended Data Figure 4 | Reduction of genetic differentiation in West Eurasia over time. We measure differentiation by F_{ST}. Each column of the 5×5 matrix of plots represents a major region and each row the earliest population with at least two individuals from each major region.

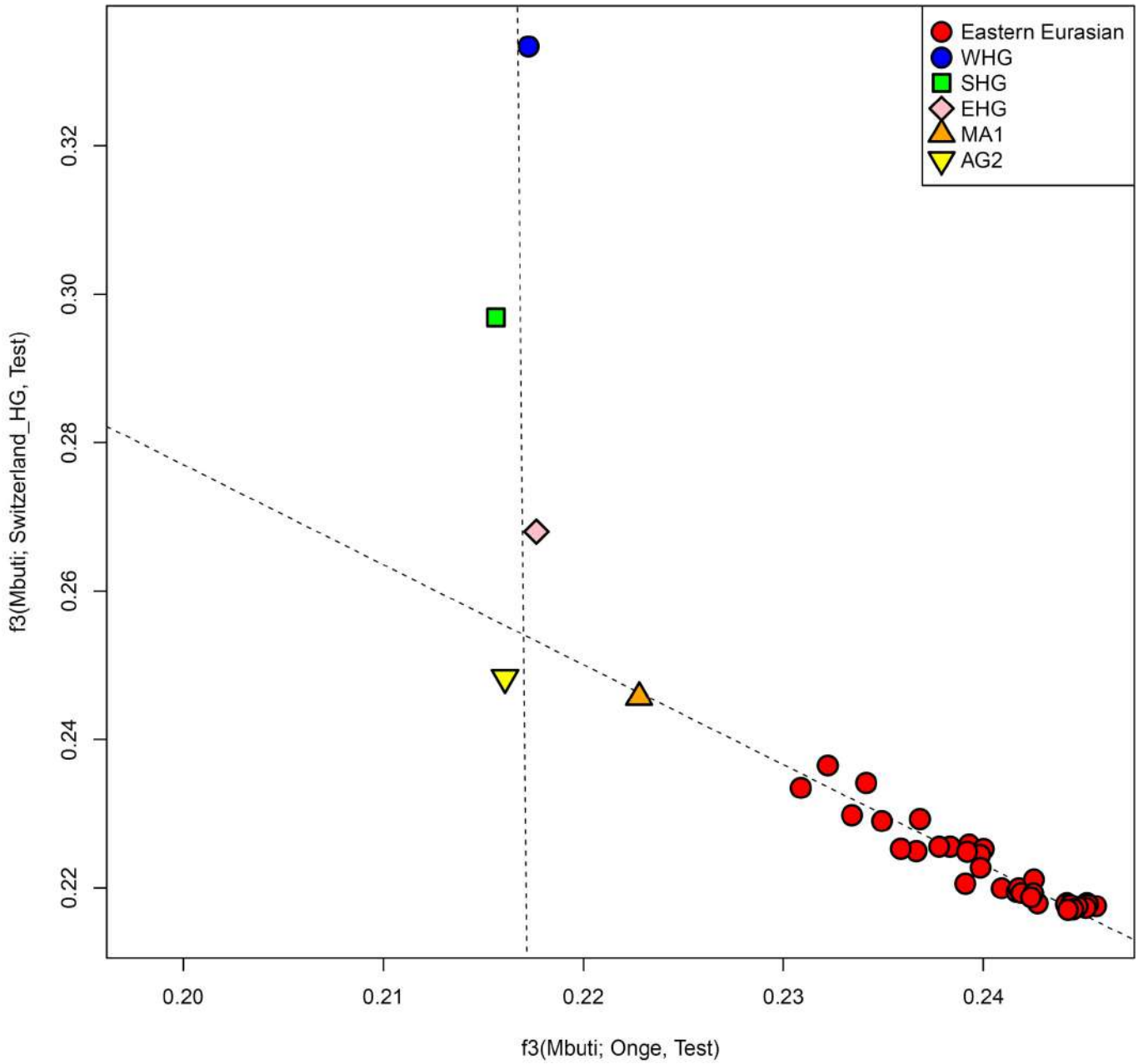


Extended Data Figure 5 | West Eurasian related admixture in East Africa, Eastern Eurasia and South Asia. a, Levantine ancestry in Eastern Africa in the Human Origins dataset. **b**, Levantine ancestry in different Eastern African population in the dataset from Pagani *et al.* (2012); the remainder of the ancestry is a clade with Mota, a ~4,500 year old sample from Ethiopia[49]. **c**, EHG ancestry in Eastern Eurasians. **d**, Afontova Gora (AG2)-related ancestry in Eastern Eurasians; the remainder of their ancestry is a clade with Onge. **e**, Mixture proportions for South Asian populations showing that they can be modelled as having West Eurasian-related ancestry similar to that in populations from both the Eurasian steppe and Iran.

Extended Data Figure 6 | Inferred position of ancient populations in West Eurasian PCA according to the model of Fig. 4.



Extended Data Figure 7 | Admixture from ghost populations using 'cline intersection'. **a**–**f**, We model each *Test* population (purple) as a mixture (pink) of a fixed reference population (blue) and a ghost population (orange) residing on the cline defined by two other populations (red and green) according to the visualization method of Supplementary Information, section 10. **a**, Early/Middle Bronze Age steppe populations are a mixture of Iran_ChL and a population on the WHG→SHG cline. **b**, Scandinavian hunter–gatherers (SHG) are a mixture of WHG and a population on the Iran_ChL→Steppe_EMBA cline. **c**, Caucasus hunter–gatherers (CHG) are a mixture of Iran_N and both WHG and EHG. **d**, Late Neolithic/Bronze Age Europeans are a mixture of the preceding Europe_MNChL population and a population with both EHG and Iran_ChL ancestry. **e**, Somali are a mixture of Mota[49] and a population on the Iran_ChL→Levant_BA cline. **f**, Eastern European hunter–gatherers (EHG) are a mixture of WHG and a population on the Onge→Han cline.



Extended Data Figure 8 | Admixture from a 'ghost' ANE population into both European and Eastern Eurasian ancestry. EHG, and Upper Palaeolithic Siberians Mal'ta 1 (MA1) and Afontova Gora 2 (AG2) are positioned near the intersection of clines formed by European hunter–gatherers (WHG, SHG, EHG) and Eastern non-Africans in the space of outgroup f_3-statistics of the form f_3(Mbuti; Papuan, *Test*) and f_3(Mbuti; Switzerland_HG, *Test*).

Extended Data Table 1 | No evidence for admixture related to sub-Saharan Africans in Natufians

Other Ancient	*African*	f_4(Natufian, *Other Ancient*; *African*, Chimp)	Z	Number of SNPs
EHG	Mbuti	-0.00044	-1.0	254033
EHG	Yoruba	0.00029	0.7	254033
EHG	Ju_hoan_North	-0.00015	-0.4	254033
EHG	Mota	-0.00022	-0.4	253986
WHG	Mbuti	-0.00067	-1.7	261514
WHG	Yoruba	-0.00045	-1.1	261514
WHG	Ju_hoan_North	-0.00046	-1.2	261514
WHG	Mota	-0.00129	-2.3	261461
SHG	Mbuti	-0.00076	-2.0	255686
SHG	Yoruba	-0.00039	-1.0	255686
SHG	Ju_hoan_North	-0.00052	-1.4	255686
SHG	Mota	-0.00091	-1.7	255641
Switzerland_HG	Mbuti	-0.00018	-0.4	261322
Switzerland_HG	Yoruba	0.00019	0.4	261322
Switzerland_HG	Ju_hoan_North	0.00009	0.2	261322
Switzerland_HG	Mota	-0.00062	-0.9	261276
Kostenki14	Mbuti	0.00034	0.7	246765
Kostenki14	Yoruba	0.00120	2.3	246765
Kostenki14	Ju_hoan_North	0.00069	1.4	246765
Kostenki14	Mota	0.00036	0.5	246719
MA1	Mbuti	-0.00038	-0.7	191819
MA1	Yoruba	0.00009	0.2	191819
MA1	Ju_hoan_North	-0.00010	-0.2	191819
MA1	Mota	-0.00038	-0.5	191782
CHG	Mbuti	-0.00051	-1.2	261505
CHG	Yoruba	-0.00012	-0.3	261505
CHG	Ju_hoan_North	-0.00013	-0.3	261505
CHG	Mota	-0.00042	-0.7	261456
Iran_N	Mbuti	-0.00018	-0.4	232927
Iran_N	Yoruba	0.00036	0.8	232927
Iran_N	Ju_hoan_North	0.00041	0.9	232927
Iran_N	Mota	0.00006	0.1	232880

We computed the statistic f_4(Natufian, *Other Ancient*; *African*, Chimp) varying *African* to be Mbuti, Yoruba, Ju_hoan_North, or the ancient Mota individual. Gene flow between Natufians and African populations would be expected to bias these statistics positive. However, we find most of them to be negative in sign and all of them to be non-significant ($|Z| < 3$), providing no evidence that Natufians differ from other ancient samples with respect to African populations.

Extended Data Table 2 | Admixture f_3-statistics

Test	*Reference*$_1$	*Reference*$_2$	f_3(*Test*; *Reference*$_1$, *Refrence*$_2$)	Z-score	Number of SNPs
Anatolia_N	Iberia_BA	Levant_N	-0.00034	-0.2	111632
Armenia_ChL	EHG	Levant_N	-0.00249	-1.5	167020
Armenia_EBA	Anatolia_N	CHG	-0.01017	-7.9	195596
Armenia_MLBA	Anatolia_N	Steppe_EMBA	-0.00809	-7.3	203796
CHG	Anatolia_ChL	Iran_HotuIIIb	0.02612	3.6	9884
EHG	Steppe_Eneolithic	Switzerland_HG	-0.00282	-0.9	67938
Europe_EN	Anatolia_N	WHG	-0.00494	-11.2	380684
Europe_LNBA	Europe_MNChL	Steppe_EMBA	-0.00920	-41.8	414782
Europe_MNChL	Anatolia_N	WHG	-0.01351	-26.8	363672
Iran_ChL	Anatolia_N	Iran_N	-0.01285	-10.6	167941
Iran_N	Iran_LN	Gana	-0.00462	-1.1	17804
Levant_BA	Iran_N	Levant_N	-0.00853	-4.7	118269
Levant_N	Europe_MNChL	Natufian	-0.00671	-3.6	61845
Natufian	Iberia_BA	Iran_HotuIIIb	0.07613	3.4	1054
SHG	Steppe_Eneolithic	Switzerland_HG	0.00728	3.2	154825
Steppe_EMBA	EHG	Abkhasian	-0.00756	-11.2	349359
Steppe_Eneolithic	EHG	Iran_LN	-0.01637	-4.2	25100
Steppe_MLBA	Europe_MNChL	Steppe_EMBA	-0.00573	-18.0	378298
WHG	Switzerland_HG	Saudi	-0.01562	-7.7	218758
Abkhasian	CHG	Sardinian	-0.00754	-13.1	387956
Adygei	Anatolia_N	Eskimo	-0.00699	-14.4	413128
Albanian	Europe_EN	Burusho	-0.00650	-16.8	395851
Armenian	Anatolia_N	Sindhi	-0.00603	-19.5	406021
Assyrian	Iran_N	Sardinian	-0.00672	-11.8	309055
Balkar	Anatolia_N	Chukchi	-0.00975	-18.8	401928
Basque	Switzerland_HG	Druze	-0.00726	-12.6	416070
BedouinA	Europe_EN	Yoruba	-0.01584	-42.8	460762
BedouinB	Iran_HotuIIIb	Natufian	0.01384	4.1	32266
Belarusian	WHG	Iranian	-0.00974	-19.8	392363
Bulgarian	Anatolia_N	Steppe_EMBA	-0.00807	-26.7	400263
Canary_Islander	Europe_MNChL	Mende	-0.00829	-5.9	353172
Chechen	Anatolia_N	Eskimo	-0.00440	-7.9	396678
Croatian	WHG	Druze	-0.00871	-18.6	394032
Cypriot	Anatolia_N	Sindhi	-0.00562	-16.1	401141
Czech	SHG	Druze	-0.00919	-21.7	374705
Druze	Iran_N	Sardinian	-0.00269	-5.8	343813
English	Steppe_EMBA	Sardinian	-0.00628	-20.6	402502
Estonian	SHG	Druze	-0.00789	-17.6	371575
Finnish	SHG	Assyrian	-0.00716	-12.6	355744
French	Steppe_EMBA	Sardinian	-0.00669	-37.9	441807
Georgian	CHG	Sardinian	-0.00782	-13.7	390744
German	WHG	Druze	-0.01103	-22.9	391302
Greek	Europe_EN	Pathan	-0.00600	-30.0	421984
Hungarian	Steppe_EMBA	Sardinian	-0.00644	-31.2	420017
Icelandic	WHG	Abkhasian	-0.00974	-17.0	394625
Iranian	Anatolia_N	Sindhi	-0.00594	-30.9	443011
Irish	Steppe_EMBA	Sardinian	-0.00590	-22.8	416663
Irish_Ulster	SHG	Assyrian	-0.00909	-15.6	350547
Italian_North	Europe_EN	Steppe_EMBA	-0.00627	-26.4	419169
Italian_South	Iberia_BA	Iran_HotuIIIb	0.01224	2.6	17678
Jew_Ashkenazi	Anatolia_N	Koryak	-0.00532	-9.4	389012
Jew_Georgian	Iran_N	Sardinian	-0.00306	-4.2	292410
Jew_Iranian	Iran_N	Sardinian	-0.00385	-5.8	302446
Jew_Iraqi	Iran_N	Sardinian	-0.00486	-6.5	287673
Jew_Libyan	Europe_EN	Yoruba	-0.00397	-7.2	415797
Jew_Moroccan	Europe_EN	Yoruba	-0.00649	-10.9	405193
Jew_Tunisian	Anatolia_N	Mende	-0.00276	-4.1	399354
Jew_Turkish	Anatolia_N	Burusho	-0.00571	-16.4	405254
Jew_Yemenite	Natufian	Kalash	-0.00341	-3.8	174052
Jordanian	Europe_EN	Yoruba	-0.01283	-26.7	423649
Kumyk	Anatolia_N	Chukchi	-0.01025	-19.6	396439
Lebanese	Anatolia_N	Yoruba	-0.01022	-19.5	414854
Lebanese_Christian	Anatolia_N	Sindhi	-0.00504	-15.7	404858
Lebanese_Muslim	Anatolia_N	Brahmin_Tiwari	-0.00616	-20.4	415129
Lezgin	Steppe_EMBA	Jew_Yemenite	-0.00481	-13.1	398974
Lithuanian	WHG	Abkhasian	-0.00999	-17.7	386718
Maltese	Anatolia_N	Brahmin_Tiwari	-0.00518	-14.5	404438
Mordovian	WHG	Iranian	-0.00912	-18.4	395230
North_Ossetian	Anatolia_N	Chukchi	-0.00894	-17.2	401729
Norwegian	WHG	Abkhasian	-0.00957	-16.5	393546
Orcadian	SHG	Druze	-0.00662	-15.8	379656
Palestinian	Europe_EN	Yoruba	-0.01129	-31.3	464066
Polish	SHG	Druze	-0.00924	-27.8	394654
Romanian	Europe_EN	Steppe_EMBA	-0.00549	-16.9	397119
Russian	SHG	Turkish	-0.00731	-25.0	398393
Sardinian	Anatolia_N	Switzerland_HG	-0.00587	-9.6	417931
Saudi	Anatolia_N	Dinka	-0.00326	-5.1	404923
Scottish	Steppe_EMBA	Sardinian	-0.00622	-26.6	426660
Shetlandic	WHG	Abkhasian	-0.00868	-14.6	386562
Sicilian	Anatolia_N	Brahmin_Tiwari	-0.00646	-22.2	411481
Sorb	SHG	Palestinian	-0.00787	-16.8	366924
Spanish	Steppe_EMBA	Sardinian	-0.00557	-32.2	447735
Spanish_North	WHG	Armenian	-0.00825	-10.9	356832
Syrian	Europe_EN	Dinka	-0.01002	-17.3	410920
Turkish	Europe_EN	Sindhi	-0.00709	-41.1	448975
Ukrainian	WHG	Abkhasian	-0.01183	-21.4	388282

We show the lowest *Z*-score of the statistic f_3(*Test*; *Reference*$_1$, *Refrence*$_2$) for *Test* populations with at least 2 individuals and every pair (*Reference*$_1$, *Refrence*$_2$) of ancient or present-day source populations. *Z*-scores lower than −3 are highlighted and indicate that the *Test* population is admixed from sources related to (but not identical to) the reference populations. *Z*-scores greater than −3 are consistent with the population either being admixed or not.

Genetic ID	Alternative ID	Latitude	Longitude	Genetic Sex Determination	Analysis Label	Detailed Label	Source (Library components if this study)	UDG treatment (minus =untreated; half=treated except	Data type	Date: One of two formats: (Format 1) 95.4% CI calibrated radiocarbon age (Conventional Radiocarbon Age, Lab number) e.g. 5983-5747 calBCE (6980±50 BP, Beta-226472) or (Format 2) Archaeological context date, e.g. 2500-1700 BCE	Location	Country	Coverage	SNPs	mitochondrial DNA haplogroup determination (for newly reported sample, only given for individuals
RISE500	6652-38	53,46	85,45	F	Steppe_	Andronovo	AllentoftNa	minus	Shotgun	..	Kytmanovo	Russia	1,818	904483	U4d1
RISE471	burial 1	48,17	10,81	M	Europe_	Germany_Bronze_A	AllentoftNa	minus	Shotgun	..	Untermeiti	Germany	0,101	116341	J1c1b
RISE559	F0174, gr 4	48,33	10,9	F	Europe_	Bell_Beaker_Germa	AllentoftNa	minus	Shotgun	..	Augsburg	Germany	0,127	146427	H46
RISE560	F0187, gr 3	48,33	10,9	M	Europe_	Bell_Beaker_Germa	AllentoftNa	minus	Shotgun	..	Augsburg	Germany	0,048	58027	U5a1a1
RISE562	F0228, obj.	48,66	12,71	F	Europe_	Bell_Beaker_Germa	AllentoftNa	minus	Shotgun	..	Landau an	Germany	0,057	68118	H2a1e
RISE563	F0234, obj.	48,69	13,02	M	Europe_	Bell_Beaker_Germa	AllentoftNa	minus	Shotgun	..	Osterhofen	Germany	0,329	340319	K1c1
RISE564	F0241, obj.	48,69	13,02	M	Europe_	Bell_Beaker_Germa	AllentoftNa	minus	Shotgun	..	Osterhofen	Germany	0,079	93003	H-T16311C
RISE566	F0521, A01	50,12	14,26	M	Europe_	Bell_Beaker_Czech	AllentoftNa	minus	Shotgun	..	Knezeves	Czech Repu	0,103	119789	H
RISE568	F0525, A01	50,19	14,16	F	Europe_	Bell_Beaker_Czech	AllentoftNa	minus	Shotgun	..	Brandysek	Czech Repu	0,053	63391	H44a
RISE569	F0527, A01	50,19	14,16	F	Europe_	Bell_Beaker_Czech	AllentoftNa	minus	Shotgun	..	Brandysek	Czech Repu	0,98	736039	H1af
RISE577	F0565, gr. 2	50,16	14,31	F	Europe_	Unetice_EBA	AllentoftNa	minus	Shotgun	..	Velke Prile	Czech Repu	1,048	738363	T2b
RISE586	F0597, gr. 6	48,8	17,02	F	Europe_	Unetice_EBA	AllentoftNa	minus	Shotgun	..	Moravska N	Czech Repu	0,176	195448	K1b1a
RISE00	grave	59,41	27,03	F	Europe_	Corded_Ware_Esto	AllentoftNa	minus	Shotgun	..	Sope	Estonia	0,91	656226	H5a1
RISE546	Kurgan 1, g	46,54	43,7	M	Steppe_	Yamnaya_Kalmykia	AllentoftNa	minus	Shotgun	..	Temrta IV	Russia	0,151	167261	U5a1d2b
RISE548	Kurgan 1, g	46,54	43,7	M	Steppe_	Yamnaya_Kalmykia	AllentoftNa	minus	Shotgun	..	Temrta IV	Russia	1,089	733271	U4
RISE412	#10	40,38	45,18	F	Armenia	Armenia_LBA	AllentoftNa	minus	Shotgun	1193-945 calBCE (2885±31 BP, UBA-27940)	Noratus	Armenia	0,105	120701	..
RISE413	#11	40,14	45,26	M	Armenia	Armenia_MBA	AllentoftNa	minus	Shotgun	1906-1698 calBCE (3493±34 BP, UBA-28941)	Nerquin Ge	Armenia	0,034	26038	..
RISE416	#14	40,14	45,26	M	Armenia	Armenia_MBA	AllentoftNa	minus	Shotgun	1643-1445 calBCE (3259±40 BP, UBA-27942)	Nerquin Ge	Armenia	0,025	29941	..
RISE423	#21	40,14	45,26	M	Armenia	Armenia_MBA	AllentoftNa	minus	Shotgun	1402-1211 calBCE (3038±32 BP, UBA-27944)	Nerquin Ge	Armenia	0,642	391520	..
RISE407	#5	40,15	45,86	F	Armenia	Armenia_LBA	AllentoftNa	minus	Shotgun	1115-895 calBCE (2827±40 BP, UBA-27938)	Norabak	Armenia	0,249	188600	..
RISE408	#6	40,15	45,86	M	Armenia	Armenia_LBA	AllentoftNa	minus	Shotgun	1209-1009 calBCE (2908±32 BP, UBA-27939)	Norabak	Armenia	0,106	119508	..
RISE396	tomb 6 ske	39,2	46,4	F	Armenia	Armenia_LBA	AllentoftNa	minus	Shotgun	1192-937 calBCE (2879±31 BP, OxA-31001)	Kapan	Armenia	0,056	66403	..
RISE397	tomb 6 ske	39,2	46,4	M	Armenia	Armenia_LBA	AllentoftNa	minus	Shotgun	1048-855 calBCE (2807±31 BP, OxA-31002)	Kapan	Armenia	0,241	247537	..
RISE247	ID 3437	47,33	18,96	M	Europe_	Vatya	AllentoftNa	minus	Shotgun	1746-1611 calBCE (3372±29 BP, OxA-29769)	Százhalomb	Hungary	0,273	237322	H11a
RISE487	T56	45,26	10,38	M	Europe_	Remedello_BA	AllentoftNa	minus	Shotgun	3483-3107 calBCE (4557±28 BP, OxA-X-2621)	Remedello	Italy	0,223	239530	H2a
RISE483	ID 106/159	47,34	18,9	F	Europe_	Vatya	AllentoftNa	minus	Shotgun	2000-1500 BCE	Erd 4	Hungary	0,121	134337	H2a1
RISE71	PMD 57, I	56,68	10,03	F	Europe_	Nordic_LN	AllentoftNa	minus	Shotgun	2196-2023 calBCE (3701±26 BP, OxA-28269)	Falshöj	Denmark	0,26	203604	H3b
RISE47	N 358 grave	56,97	9,552	M	Europe_	Nordic_BA	AllentoftNa	minus	Shotgun	1499-1324 calBCE (3153±26 BP, OxA-28258)	Sebber sko	Denmark	0,105	103513	I
RISE435	3/2	48,93	12,26	F	Europe_	Corded_Ware_Gern	AllentoftNa	minus	Shotgun	2863-2498 calBCE (4094±33 BP, UBA-27947)	Tiefbrunn	Germany	0,043	50680	J1b1a1
RISE486	T78	45,26	10,38	M	Europe_	Remedello_BA	AllentoftNa	minus	Shotgun	2134-1773 calBCE (3595±55 BP, ETH-12913)	Remedello	Italy	0,289	203575	J1c1b
RISE386	burial 4	52,45	55,16	M	Steppe_	Sintashta_MBA	AllentoftNa	minus	Shotgun	2298-2045 calBCE (3775±34 BP, OxA-30991)	Bulanovo	Russia	0,415	315156	J1c1b1a
RISE61	PMD 17, V,	55,7	11,86	M	Europe_	Nordic_MN_B	AllentoftNa	minus	Shotgun	2851-2492 calBCE (4071±27 BP, OxA-28296)	Kyndelöse	Denmark	0,318	293152	J1c4
RISE254	ID 4091	47,33	18,96	M	Europe_	Vatya	AllentoftNa	minus	Shotgun	2128-1909 calBCE (3631±29 BP, OxA-29842)	Százhalomb	Hungary	0,038	42943	J1c9
RISE511	6136-6	54,58	90,78	F	Steppe_	Afanasievo	AllentoftNa	minus	Shotgun	2909-2679 calBCE (4224±36 BP, OxA-31568)	Bateni	Russia	3,755	1102979	J2a2a
RISE510	6136-9	54,58	90,78	F	Steppe_	Afanasievo	AllentoftNa	minus	Shotgun	2851-2468 calBCE (4040±45 BP, OxA-31222)	Bateni	Russia	0,236	249386	J2a2a
RISE392	kurgan 4 bu	53,88	59,08	M	Steppe_	Sintashta_MBA	AllentoftNa	minus	Shotgun	2126-1896 calBCE (3626±33 BP, OxA-30999)	Stepnoe VI	Russia	0,598	526675	J2b1a2a
RISE373	Grave # 12	46,22	20,2	F	Europe_	Maros	AllentoftNa	minus	Shotgun	1886-1696 calBCE (3476±30 BP, OxA-31104)	Szöreg - C (	Hungary	0,282	259708	K1a2a
RISE94	grave 26:I	56,03	14,23	M	Europe_	BattleAxe_Sweden	AllentoftNa	minus	Shotgun	2621-2472 calBCE (4025±30 BP, OxA-29033)	Viby	Sweden	0,927	619914	K1a2a
RISE179	barrow I gr	55,4	13,6	M	Europe_	Nordic_LN	AllentoftNa	minus	Shotgun	2010-1776 calBCE (3556±28 BP, OxA-29193)	Abekås I	Sweden	0,045	50762	K1a3
RISE154	grave 3	50,95	16,94	F	Europe_	Unetice_EBA	AllentoftNa	minus	Shotgun	1925-1765 calBCE (3522±24 BP, Uba-16555)	Szczepanko	Poland	0,178	171391	K1a4a1

RISE98	grave 49, S	55,38	13,45	M	Europe_	Nordic_LN	AllentoftNa	minus	Shotgun	2275-2032 calBCE (3736±32 BP, OxA-28987)	L Beddinge	Sweden	6,125	1125643	K1b1a1
RISE97	grave 72(II)	55,56	13,06	F	Europe_	Nordic_LN	AllentoftNa	minus	Shotgun	2025-1885 calBCE (3590±29 BP, OxA-28986)	Fredriksber	Sweden	0,746	580367	K2a5
RISE555	CGG_2_011	48,72	44,5	M	Steppe_	Russia_EBA	AllentoftNa	minus	Shotgun	2857-2497 calBCE (4082±28 BP, AAR-20358)	Stalingrad (	Russia	0,218	232118	N1a1a-T152C
RISE391	kurgan 7 bu	50,59	56,83	F	Steppe_	Sintashta_MBA	AllentoftNa	minus	Shotgun	2120-1887 calBCE (3612±34 BP, OxA-30998)	Tanaberge	Kazakhstan	0,056	55088	N1a1a1a1
RISE175	barrow I gr	55,4	13,6	M	Europe_	Nordic_BA	AllentoftNa	minus	Shotgun	1395-1132 calBCE (3025±30 BP, OxA-28998)	Abekås I	Sweden	0,104	112552	T1a1
RISE484	ID 772/117	47,34	18,9	F	Europe_	Vatya	AllentoftNa	minus	Shotgun	2000-1500 BCE	Erd 4	Hungary	0,152	168871	T1a1
RISE210	Cranium VI	56	14,1	M	Europe_	Nordic_BA	AllentoftNa	minus	Shotgun	1432-1292 calBCE (3105±28 BP, OxA-29654)	Ängamölla	Sweden	0,046	51095	T2a1a
RISE547	Kurgan 1, g	46,54	43,7	M	Steppe_	Yamnaya_Kalmykia	AllentoftNa	minus	Shotgun	2887-2634 calBCE (4175±35 BP, GrA-58960)	Temrta IV	Russia	0,729	628581	T2a1a
RISE552	Kurgan 4, g	46,62	43,33	M	Steppe_	Yamnaya_Kalmykia	AllentoftNa	minus	Shotgun	2849-2143 calBCE (3940±90 BP, IGAN-4079)	Ulan IV	Russia	2,839	971022	T2a1a
RISE276	bog find 19	55,91	11,57	M	Europe_	Nordic_LBA	AllentoftNa	minus	Shotgun	794-547 calBCE (2525±25 BP, OxA-30485)	Trundholm	Denmark	0,083	92841	T2b
RISE374	Grave # 147	46,22	20,2	M	Europe_	Maros	AllentoftNa	minus	Shotgun	1866-1619 calBCE (3402±34 BP, OxA-30989)	Szöreg - C (	Hungary	0,149	146423	T2b
RISE479	ID 1129/17	47,34	18,9	M	Europe_	Vatya	AllentoftNa	minus	Shotgun	2000-1500 BCE	Erd 4	Hungary	1,4	884093	T2b
RISE349	Grave # 33	46,36	20,99	F	Europe_	Hungary_MBA	AllentoftNa	minus	Shotgun	2034-1784 calBCE (3588±34 BP, OxA-30987)	Battonya V	Hungary	0,082	96499	T2b3
RISE509	6136-5	54,58	90,78	F	Steppe_	Afanasievo	AllentoftNa	minus	Shotgun	2887-2677 calBCE (4186±27 BP, OxA-31221)	Bateni	Russia	0,974	703520	T2c1a2
RISE431	Barrow 4, s	52,14	16,54	M	Europe_	Corded_Ware_Proto	AllentoftNa	minus	Shotgun	2286-2048 calBCE (3762±27 BP, OxA-27967)	Leki Male	Poland	0,111	124341	T2e
RISE394	burial 6 ske	52,45	55,16	F	Steppe_	Sintashta_MBA	AllentoftNa	minus	Shotgun	1949-1754 calBCE (3532±34 BP, OxA-30993)	Bulanovo	Russia	0,363	336702	U2e1e
RISE395	kurgan 25 b	52,64	59,54	F	Steppe_	Sintashta_MBA	AllentoftNa	minus	Shotgun	1960-1756 calBCE (3540±33 BP, OxA-30996)	Bol'shekara	Russia	2,397	888393	U2e1h
RISE503	6652-41	53,46	85,45	F	Steppe_	Andronovo	AllentoftNa	minus	Shotgun	1727-1511 calBCE (3328±38 BP, OxA-31445)	Kytmanovo	Russia	0,879	678907	U2e2
RISE434	3/1	48,93	12,26	M	Europe_	Corded_Ware_Gern	AllentoftNa	minus	Shotgun	2880-2630 calBCE (4161±34 BP, UBA-27946)	Tiefbrunn	Germany	0,1	111108	U4
RISE109	grave 1044	50,98	17,07	F	Europe_	Unetice_EBA	AllentoftNa	minus	Shotgun	1954-1772 calBCE (3544±26 BP, UB-16557)	Wojkowice	Poland	0,215	211732	U4
RISE505	6652-42	53,46	85,45	F	Steppe_	Andronovo	AllentoftNa	minus	Shotgun	1746-1626 calBCE (3391±27 BP, OxA-31216)	Kytmanovo	Russia	5,277	1170027	U4a1b
RISE507	5910-2A	51,5	85,97	F	Steppe_	Afanasievo	AllentoftNa	minus	Shotgun	3322-2923 calBCE (4423±29 BP, OxA-31219)	River Kuyu	Russia	0,155	170131	U5a1a1
RISE508	5910-2B	51,5	85,97	F	Steppe_	Afanasievo	AllentoftNa	minus	Shotgun	3331-2935 calBCE (4442±29 BP, OxA-31220)	River Kuyu	Russia	0,339	322453	U5a1a1
RISE150	grave 02	50,92	16,96	F	Europe_	Unetice_EBA	AllentoftNa	minus	Shotgun	1885-1693 calBCE (3469±31 BP, Ua-42401)	Przeclawice	Poland	2,653	654921	U5a1b1
RISE240	kurgan 1, g	46,58	43,68	F	Steppe_	Yamnaya_Kalmykia	AllentoftNa	minus	Shotgun	2880-2632 calBCE (4160±30 BP, GrA-45038)	Sukhaya Te	Russia	0,201	212900	U5a1d1
RISE550	Kurgan 1, g	46,56	43,68	M	Steppe_	Yamnaya_Kalmykia	AllentoftNa	minus	Shotgun	3334-2635 calBCE (4312±94 BP, IGAN-2880)	Peshany V	Russia	0,521	476829	U5a1i
RISE480	ID 1039/15	47,34	18,9	F	Europe_	Vatya	AllentoftNa	minus	Shotgun	1700-1500 BCE	Erd 4	Hungary	0,148	158897	U5a2a
RISE371	Grave # 105	46,22	20,2	F	Europe_	Maros	AllentoftNa	minus	Shotgun	2136-1941 calBCE (3653±32 BP, OxA-30988)	Szöreg - C (	Hungary	0,045	53698	U5a2b
RISE436	3/3	48,93	12,26	M	Europe_	Corded_Ware_Gern	AllentoftNa	minus	Shotgun	2868-2580 calBCE (4124±31 BP, UBA-27948)	Tiefbrunn	Germany	0,076	63030	U5b1c2
RISE446	burial 13 m	50,01	10,18	M	Europe_	Corded_Ware_Gern	AllentoftNa	minus	Shotgun	2829-2465 calBCE (4015±38 BP, UBA-27950)	Bergrheinfe	Germany	0,172	187862	U5b1c2
RISE489	T65	45,26	10,38	M	Europe_	Remedello_BA	AllentoftNa	minus	Shotgun	2908-2578 calBCE (4185±70 BP, ETH-12188)	Remedello	Italy	0,548	510496	X2c1
Ust_Ishim	UstIshim	57,7	71,1	M	Ust_Ishi	Ust_Ishim_HG	FuNature2(	half	Shotgun	45530-40610 calBCE [46064-40920 calBCE (41400±13(	Ust'-Ishim,	Russia	40	1240000	..
ATP16	ATP16	42,35	-3,518	F	Europe_	Iberia_Chalcolithic	GuntherPN	minus	Shotgun	3261-2916 calBCE (4400±30, Beta-368289)	El Portalon	Spain	1,184	839051	..
ATP2	ATP2	42,35	-3,518	M	Europe_	Iberia_Chalcolithic	GuntherPN	minus	Shotgun	2899-2678 calBCE (4210±30, Beta-386394)	El Portalon	Spain	4,484	1205481	..
ATP9	ATP9	42,35	-3,518	F	Iberia_B	Iberia_Bronze_Age	GuntherPN	minus	Shotgun	1750-1618 calBCE (3390±30, Beta-386395)	El Portalon	Spain	0,461	449162	..
Matojo	Matojo	42,35	-3,518	M	Europe_	Iberia_Chalcolithic	GuntherPN	minus	Shotgun	3010-3879 calBCE (4300±30, Beta-368295)	El Portalon	Spain	0,087	102497	..
Bichon	Bichon	47,10	6,87	M	Switzerl	Switzerland_HG	JonesNatur	minus	Shotgun	11820-11610 calBCE (11855±50 BP, OxA-27763)	Grotte du E	Switzerlanc	8,443	1223510	..
KK1	Kotias	42,28	43,28	M	CHG	CHG	JonesNatur	minus	Shotgun	7940-7600 calBCE [7938-7580 calBCE (8665±65 BP, RT	Kotias Klde	Georgia	12,482	1230084	..
SATP	Satsurblia	42,38	42,59	M	CHG	CHG	JonesNatur	minus	Shotgun	11430-11180 calBCE (11415±50 BP, OxA-34632)	Satsurblia	Georgia	1,191	839553	..
Stuttgart	Stuttgart	48,78	9,18	F	Europe_	Stuttgart	LazaridisNa	half	Shotgun	5310-5070 calBCE (6246±30 BP, MAMS-24635)	Viesenhaeu	Germany	19	1240000	T2c1b
Loschbour	Loschbour	49,81	6,4	M	WHG	Loschbour_HG	LazaridisNa	Mixed:ha	Shotgun	6210-5990 calBCE (7205±50 BP, OxA-7738)	Echternach	Luxembour	22	1240000	U5b1a
I0172	ESP24	51,42	11,68	M	Europe_	Esperstedt_MN	MathiesonI	Mixed:ha	1240K.ca	3360-3086 calBCE (Er-8699)	Esperstedt	Germany	25,372	1158724	T2b
I0174	BAM25	46,2	18,7	M	Europe_	Starcevo_EN	MathiesonI	plus	1240K.ca	5702-5536 calBCE (6695±40BP, MAMS-11939)	Alsonyek-B	Hungary	0,224	231471	N1a1a1
I0118	ALB3	51,45	11,63	F	Europe_	Alberstedt_LN	MathiesonI	plus	1240K.ca	2471-2246 calBCE (3892±32 BP, MAMS-21492)	Alberstedt	Germany	11,371	1089011	HV
I0059	BZH6	51,82	10,91	F	Europe_	BenzigerodeHeimbu	MathiesonI	Mixed:ha	1240K.ca	2337-2138 calBCE (3796±30 BP, MAMS-21486)	Benzingero	Germany	1,303	752357	H1
I0104	ESP11	51,42	11,68	M	Europe_	Corded_Ware_Gern	MathiesonI	plus	1240K.ca	2559-2296 calBCE (3927±37 BP, MAMS-21487)	Esperstedt	Germany	4,184	990775	U4b1a1a1
I0103	ESP16	51,42	11,68	F	Europe_	Corded_Ware_Gern	MathiesonI	Mixed:ha	1240K.ca	2578-2468 calBCE (4000±31 BP, MAMS-21488)	Esperstedt	Germany	17,528	1109474	W6a
I0049	ESP22	51,42	11,68	F	Europe_	Corded_Ware_Gern	MathiesonI	plus	1240K.ca	2464-2210 calBCE (3867±35 BP, MAMS-21489)	Esperstedt	Germany	0,649	534318	X2b4
I0117	ESP29	51,42	11,68	F	Europe_	Unetice_EBA	MathiesonI	Mixed:ha	1240K.ca	2272-2039 calBCE (3743±25 BP, MAMS-21496)	Esperstedt	Germany	1,902	874673	I3a
I0409	Troc1	42,5	0,5	F	Europe_	Iberia_EN	MathiesonI	plus	1240K.ca	5310-5218 calBCE (6280±25 BP, MAMS-16159)	Els Trocs	Spain	0,686	528513	J1c3

I0412	Troc5	42,5	0,5	M	Europe_	Iberia_EN	MathiesonI	half	1240K.c	5308-5080 calBCE (6249±28 BP, MAMS-16164)	Els Trocs	Spain	13,785	1025420	N1a1a1
I0054	UWS4	51,66	11,53	F	Europe_	LBK_EN	MathiesonI	plus	1240K.c	5222-5022 calBCE (6180±34 BP, MAMS-21485)	Unterwied	Germany	18,61	1032320	J1c17
I0106	ESP26	51,42	11,68	F	Europe_	Corded_Ware_Gern	MathiesonI	plus	1240K.c	2464-2210 calBCE (3867±32 BP, MAMS-21490)	Esperstedt	Germany	0,224	230482	T2a1b1
I0115	ESP3	51,42	11,68	F	Europe_	Unetice_EBA	MathiesonI	plus	1240K.c	1954-1760 calBCE (3540±31 BP, MAMS-21494)	Esperstedt	Germany	0,409	365760	U5a1i
I0116	ESP4	51,42	11,68	M	Europe_	Unetice_EBA	MathiesonI	plus	1240K.c	2134-1939 calBCE (3650±32 BP, MAMS-21495)	Esperstedt	Germany	1,648	640413	W3a1
I0803	EUL41	51,17	11,85	F	Europe_	Unetice_EBA	MathiesonI	half	1240K.c	2132-1942 calBCE (3650±26 BP, MAMS-22822)	Eulau	Germany	0,228	224184	H4a1a1a
I0804	EUL57	51,17	11,85	M	Europe_	Unetice_EBA	MathiesonI	half	1240K.c	2137-1965 calBCE (3671±26 BP, MAMS-22821)	Eulau	Germany	0,054	62496	H3
I0047	HAL16	51,89	11,04	F	Europe_	Unetice_EBA	MathiesonI	half	1240K.c	2111-1891 calBCE (3612±30 BP, MAMS-21481)	Halberstad	Germany	1,655	836247	V9
I0048	HAL25	51,89	11,04	M	Europe_	LBK_EN	MathiesonI	half	1240K.c	5211-5009 calBCE (6153±33 BP, MAMS-21482)	Halberstad	Germany	0,456	422925	K1a
I0057	HAL34	51,89	11,04	F	Europe_	LBK_EN	MathiesonI	half	1240K.c	5218-5019 calBCE (6173±34 BP, MAMS-21483)	Halberstad	Germany	0,139	151700	N1a1a1
I0099	HAL36C	51,89	11,04	M	Europe_	Halberstadt_LBA	MathiesonI	half	1240K.c	1193-979 calBCE (2889±30 BP, MAMS-21484)	Halberstad	Germany	4,948	985957	H23
I0046	HAL5	51,89	11,04	F	Europe_	LBK_EN	MathiesonI	half	1240K.c	5212-4989 calBCE (6137±36 BP, MAMS-21479)	Halberstad	Germany	1,945	907019	T2c1
I0550	KAR22A	51,27	11,66	F	Europe_	Karsdorf_LN	MathiesonI	half	1240K.c	2570-2471 calBCE (3993±21 BP, MAMS-23344)	Karsdorf	Germany	0,15	166519	T1a1
I0795	KAR6	51,27	11,66	M	Europe_	LBK_EN	MathiesonI	half	1240K.c	5216-5036 calBCE (6174±29 BP, MAMS-22823)	Karsdorf	Germany	0,097	107480	H1 or H1au1b
I0559	QLB15D	51,79	11,14	M	Europe_	Baalberge_MN	MathiesonI	half	1240K.c	3652-3527 calBCE (4815±26 BP, MAMS-22818)	Quedlinbur	Germany	0,163	178578	HV
I0164	QUEVIII6	51,79	11,14	F	Europe_	Unetice_EBA	MathiesonI	plus	1240K.c	2023-1894 calBCE (3599±25 BP, MAMS-21497)	Quedlinbur	Germany	2,487	863754	U5b2a1b
I0176	SZEH4	46,4	18,74	F	Europe_	LBKT_EN	MathiesonI	plus	1240K.c	5207-4944 calBCE (6110±30 BP, Beta-310038)	Szemely-He	Hungary	0,066	71473	N1a1a1a3
I0410	Troc3	42,5	0,5	M	Europe_	Iberia_EN	MathiesonI	plus	1240K.c	5295-5066 calBCE (6217±25 BP, MAMS-16161)	Els Trocs	Spain	1,305	528396	T2c1d or T2c1d2
I0413	Troc7	42,5	0,5	F	Europe_	Iberia_EN	MathiesonI	plus	1240K.c	5302-5074 calBCE (6234±28 BP, MAMS-16166)	Els Trocs	Spain	1,572	657071	V
I0061	UzOO74	61,65	35,65	M	EHG	Karelia_HG	MathiesonI	plus	1240K.c	6850-6000 BCE [6270-6000 calBCE (7280±80 BP, OxA-	Yuzhnyy Ol	Russia	5,272	1057094	C1
I0246	SVP41	53,45	50,45	M	Steppe_	Potapovka	MathiesonI	half	1240K.c	2469-1928 calBCE (3760±100 BP, AA-12568)	Utyevka VI,	Russia	0,078	90021	C1
I0247	SVP56	52,43	51,16	M	Steppe_	Scythian_IA	MathiesonI	half	1240K.c	375-203 calBCE (2220±30 BP, Beta-392493)	Nadezhdinl	Russia	2,812	951695	G2a4
I1498	HUNG302,	47,52	21,59	F	Europe_	Hungary_MN	MathiesonI	half	1240K.c	5290-5060 calBCE (6207±30 BP, OxA-27858)	Debrecen T	Hungary	4,879	839760	H
I1497	HUNG353,	47,17	19,83	F	Europe_	Hungary_Late_Copp	MathiesonI	half	1240K.c	2900-2700 calBCE (4421±27 BP, MAMS-14825)	Apc-Bereka	Hungary	4,563	843595	H
I0726	M15, M15.	40,26	29,65	F	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6400-5600 BCE	Mentese	Turkey	0,221	236641	H or H5-C16192T
I0805	QLB26	51,79	11,15	M	Europe_	Bell_Beaker_Germa	MathiesonI	half	1240K.c	2467-2142 calBCE (3839±55 BP, Er-8558)	Quedlinbur	Germany	0,232	231622	H1
I0806	QLB28	51,79	11,14	M	Europe_	Bell_Beaker_Germa	MathiesonI	half	1240K.c	2431-2150 calBCE (3824±25 BP, MAMS-22820)	Quedlinbur	Germany	0,13	142638	H1
I0406	Mina4	41,25	-2,326	M	Europe_	Iberia_MN	MathiesonI	plus	1240K.c	3900-3600 BCE	La Mina	Spain	3,947	738288	H1
I0374	SVP16, NIK	49,97	44,67	M	Steppe_	Poltavka	MathiesonI	half	1240K.c	2800-2200 BCE	Nikolaevka	Russia	1,977	891602	H13a1a
I0370	SVP10	51,27	58,18	M	Steppe_	Yamnaya_Samara	MathiesonI	half	1240K.c	3300-2700 BCE	Ishkinovka	Russia	1,043	712941	H13a1a1
I0112	QUEXII6	51,79	11,14	F	Europe_	Bell_Beaker_Germa	MathiesonI	plus	1240K.c	2457-2142 calBCE (3820±42 BP, Er-7038)	Quedlinbur	Germany	3,768	985107	H13a1a2
I0807	ESP30	51,42	11,68	M	Europe_	Baalberge_MN	MathiesonI	half	1240K.c	3970-3710 calBCE (5061±62 BP, Er-7784)	Esperstedt	Germany	0,086	94260	H1e1a
I1281	MIR18	42,33	-3,5	F	Europe_	Iberia_Chalcolithic	MathiesonI	half	1240K.c	2865-2575 calBCE (4110±30 BP, Beta-416458)	El Mirador	Spain	1,575	646025	H1t
I0122	SVP35	52,22	48,1	M	Steppe_	Samara_Eneolithic	MathiesonI	Mixed:h	1240K.c	5200-4000 BCE	Khvalynsk I	Russia	0,723	582950	H2a1
I0441	SVP54	52,3	52,05	F	Steppe_	Yamnaya_Samara	MathiesonI	half	1240K.c	3010-2622 calBCE (4234±60 BP, AA-47805)	Kurmanaev	Russia	0,091	104430	H2b
I0431	SVP40	53,08	50,36	F	Steppe_	Srubnaya	MathiesonI	half	1240K.c	1850-1600 BCE	Spiridonovl	Russia	0,094	107284	H2b
I1277	MIR14	42,33	-3,5	M	Europe_	Iberia_Chalcolithic	MathiesonI	half	1240K.c	2568-2346 calBCE (3950±30 BP, Beta-416457)	El Mirador	Spain	0,941	566637	H3
I1282	MIR19	42,33	-3,5	M	Europe_	Iberia_Chalcolithic	MathiesonI	half	1240K.c	2900-2346 BCE [2568-2346 calBCE (3950±30 BP, Beta-	El Mirador	Spain	0,06	69288	H3
I1284	MIR21	42,33	-3,5	M	Europe_	Iberia_Chalcolithic	MathiesonI	half	1240K.c	2900-2346 BCE [2568-2346 calBCE (3950±30 BP, Beta-	El Mirador	Spain	0,114	125450	H3
I0111	ROT4	51,45	11,54	F	Europe_	Bell_Beaker_Germa	MathiesonI	Mixed:h	1240K.c	2475-2204 calBCE (3881±50 BP, Er-8712)	Rothenschi	Germany	0,731	558181	H3ao
I1276	MIR13	42,33	-3,5	F	Europe_	Iberia_Chalcolithic	MathiesonI	half	1240K.c	2900-2346 BCE [2568-2346 calBCE (3950±30 BP, Beta-	El Mirador	Spain	0,104	117488	H3c3
I0430	SVP39	53,08	50,36	M	Steppe_	Srubnaya	MathiesonI	half	1240K.c	1850-1600 BCE	Spiridonovl	Russia	3,21	881469	H3g
I0797	KAR16A	51,28	11,65	M	Europe_	LBK_EN	MathiesonI	half	1240K.c	5500-4775 BCE	Karsdorf	Germany	0,087	95833	H46b
I1580	L12-393	40,3	29,57	F	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	2,903	934855	H5
I0108	ROT6	51,45	11,54	F	Europe_	Bell_Beaker_Germa	MathiesonI	Mixed:h	1240K.c	2575-2299 calBCE (3953±47 BP, Er-8710)	Rothenschi	Germany	1,197	764099	H5a3
I0361	SVP9	53,03	50,39	M	Steppe_	Srubnaya	MathiesonI	half	1240K.c	1850-1200 BCE	Spiridonovl	Russia	0,448	415327	H5b
I0358	SVP6	53,03	50,39	F	Steppe_	Srubnaya	MathiesonI	half	1240K.c	1906-1631 calBCE (3455±56 BP, AA-47808)	Spiridonovl	Russia	0,264	273856	H6a1a
I0444	SVP58	53,1	51,13	M	Steppe_	Yamnaya_Samara	MathiesonI	half	1240K.c	3335-2882 calBCE (4370±75 BP, AA-12570)	Kutuluk I, K	Russia	0,653	551461	H6a1b
I0126	SVP51	53,31	51,15	M	Steppe_	Poltavka	MathiesonI	half	1240K.c	2867-2486 calBCE (4081±54 BP, AA-53803)	Kutuluk III,	Russia	0,347	339812	H6a2
I0234	SVP25	53,14	50,01	F	Steppe_	Srubnaya	MathiesonI	half	1240K.c	1850-1600 BCE	Rozhdestve	Russia	0,513	467183	I1a1

I0440	SVP53	53,74	49,38	M	Steppe_	Poltavka	MathiesonI	half	1240K.c	2887-2666 calBCE (4180±30 BP, Beta-392492)	Lopatino II,	Russia	4,953	1058753	I3a	
I1585	M11-59	40,3	29,57	F	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	3,098	852587	J1 or J1c	
I1500	HUNG372,	47,17	20,83	M	Europe_	ALPC_Hungary_MN	MathiesonI	half	1240K.c	5210-4990 calBCE (6164±64 BP,OxA-23763)	Kompolt-Ki	Hungary	4,246	854790	J1c1	
I1280	MIR17	42,33	-3,5	F	Europe_	Iberia_Chalcolithic	MathiesonI	half	1240K.c	2900-2346 BCE [2568-2346 calBCE (3950±30 BP, Beta-	El Mirador	Spain	0,22	208463	J1c1	
I0744	M10-275	40,3	29,57	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	2,39	907009	J1c11	
I1539	ESP25	51,42	11,68	F	Europe_	Corded_Ware_Gern	MathiesonI	half	1240K.c	2625-2291 calBCE (3967±57 BP, Er-7779)	Esperstedt	Germany	0,132	143190	J1c1b1a	
I1532	ESP8	51,42	11,68	M	Europe_	Corded_Ware_Gern	MathiesonI	half	1240K.c	2500-2050 BCE	Esperstedt	Germany	0,566	475207	J1c2e	
I0113	QUEXII4	51,79	11,14	F	Europe_	Bell_Beaker_Germa	MathiesonI	Mixed:h	1240K.c	2346-2033 calBCE (3773±47 BP, Er-7283)	Quedlinbur	Germany	0,777	609165	J1c5	
I1505	PF839/119	47,88	21,19	F	Europe_	Hungary_Neolithic	MathiesonI	half	1240K.c	5290-5050 calBCE (6153±33 BP, OxA-28020)	Polgar Fere	Hungary	3,494	797508	J1c5	
I1538	ESP20	51,42	11,68	M	Europe_	Corded_Ware_Gern	MathiesonI	half	1240K.c	2500-2050 BCE	Esperstedt	Germany	0,126	137950	J1c5	
I1540	ESP28	51,42	11,68	M	Europe_	Corded_Ware_Gern	MathiesonI	half	1240K.c	2500-2050 BCE	Esperstedt	Germany	0,298	291946	J1c5	
I1314	MIR26	42,33	-3,5	M	Europe_	Iberia_Chalcolithic	MathiesonI	half	1240K.c	2900-2346 BCE [2568-2346 calBCE (3950±30 BP, Beta-	El Mirador	Spain	0,268	224117	J2a1a1	
I1542	ESP33	51,42	11,68	M	Europe_	Corded_Ware_Gern	MathiesonI	half	1240K.c	2500-2050 BCE	Esperstedt	Germany	0,068	78085	J2a2a	
I0423	SVP31	52,54	50,5	M	Steppe_	Srubnaya_Outlier	MathiesonI	half	1240K.c	1850-1200 BCE	Barinovka I	Russia	0,485	418611	J2b1a2a	
I1508	HUNG276,	47,32	21,53	F	Europe_	Koros_Hungary_EN	MathiesonI	half	1240K.c	5710-5570 calBCE (6726±35 BP, OxA-28101)	Berettyóújf	Hungary	0,914	585392	K1a	
I1271	MIR1	42,33	-3,5	F	Europe_	Iberia_Chalcolithic	MathiesonI	half	1240K.c	2900-2346 BCE [2568-2346 calBCE (3950±30 BP, Beta-	El Mirador	Spain	0,241	239592	K1a	
I0746	L11-322	40,3	29,57	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	8,465	1033308	K1a or K1a1	
I1100	M11-351	40,3	29,57	F	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	0,331	304802	K1a or K1a6	
I1579	M13-72	40,3	29,57	F	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	2,507	847361	K1a-C150T	
I1504	HUNG381,	47,82	19,95	M	Europe_	Late_Bronze_Age	MathiesonI	half	1240K.c	1270-1110 calBCE (2769±24 BP, OxA-27859)	Ludas-Varju	Hungary	1,332	704178	K1a1a	
I0405	Mina3	41,25	-2,326	M	Europe_	Iberia_MN	MathiesonI	plus	1240K.c	3900-3600 BCE	La Mina	Spain	0,377	327044	K1a1b1	
I1534	ESP14	51,42	11,68	M	Europe_	Corded_Ware_Gern	MathiesonI	half	1240K.c	2500-2050 BCE	Esperstedt	Germany	0,158	167366	K1a1b2a	
I0727	M24 / UA J	40,26	29,65	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6400-5600 BCE	Mentese	Turkey	0,171	186798	K1a2	
I1550	HAL19	51,89	11,04	F	Europe_	LBK_EN	MathiesonI	half	1240K.c	5500-4775 BCE	Halberstad	Germany	0,745	514954	K1a2	
I1583	L14-200	40,3	29,57	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	9,396	1012449	K1a2	
I1300	MIR22	42,33	-3,5	F	Europe_	Iberia_Chalcolithic	MathiesonI	half	1240K.c	2900-2346 BCE [2568-2346 calBCE (3950±30 BP, Beta-	El Mirador	Spain	2,792	807280	K1a2a	
I0060	ROT3	51,45	11,54	F	Europe_	Bell_Beaker_Germa	MathiesonI	plus	1240K.c	2428-2149 calBCE (3822±25 BP, MAMS-22819)	Rothenschi	Germany	0,105	118605	K1a2c	
I1102	M11-354	40,3	29,57	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	0,55	423429	K1a3a	
I1496	HUNG352,	47,17	19,83	M	Europe_	LBK_Hungary_MN	MathiesonI	half	1240K.c	5300-4950 calBCE (6135±33 BP, MAMS-14821)	Apc-Bereka	Hungary	5,246	835911	K1a3a3	
I0724	T2 / UP	40,26	29,65	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6400-5600 BCE	Mentese	Turkey	0,231	239349	K1a4	
I0707	BAR2 / L11	40,3	29,57	F	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	9,431	1026916	K1a4	
I1272	MIR2	42,33	-3,5	F	Europe_	Iberia_Chalcolithic	MathiesonI	half	1240K.c	2857-2496 calBCE (4080±30 BP, Beta-416456)	El Mirador	Spain	0,977	599689	K1b1a1	
I0407	Mina6	41,25	-2,326	F	Europe_	Iberia_MN	MathiesonI	plus	1240K.c	3900-3600 BCE	La Mina	Spain	1,363	567058	K1b1a1	
I1103	M11-S-350	40,3	29,57	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	1,135	622029	K1b1b1	
I0235	SVP26	53,14	50,01	F	Steppe_	Srubnaya	MathiesonI	half	1240K.c	1850-1600 BCE	Rozhdestve	Russia	0,454	420407	K1b2a	
I1502	HUNG370,	47,17	20,83	F	Europe_	Hungary_Mako_EB/	MathiesonI	half	1240K.c	2190-1980 calBCE (3686±28 BP,OxA-23799)	Kompolt-Ki	Hungary	4,682	824420	K1c1	
I1544	ESP36	51,42	11,68	M	Europe_	Corded_Ware_Gern	MathiesonI	half	1240K.c	2500-2050 BCE	Esperstedt	Germany	0,104	117069	K2b2	
I1096	BAR26 / M1	40,3	29,57	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	2,865	780073	N1a1a1	
I0100	HAL4	51,89	11,04	F	Europe_	LBK_EN	MathiesonI	half	1240K.c	5202-4852 calBCE (6080±32 BP, KIA-40341)	Halberstad	Germany	6,727	1002375	N1a1a1a	
I1495	HUNG347,	47,17	19,83	M	Europe_	Hungary_MN	MathiesonI	half	1240K.c	4490-4360 calBCE (5598±32 BP, MAMS-14819)	Apc-Bereka	Hungary	4,853	854267	N1a1a1a	
I0736	L11-216	40,3	29,57	F	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	2,248	825825	N1a1a1a	
I0659	HAL2	51,89	11,04	M	Europe_	LBK_EN	MathiesonI	half	1240K.c	5211-4963 calBCE [5211-4963 calBCE (6130±39 BP, KI	Halberstad	Germany	0,703	582038	N1a1a1a2	
I0708	BAR6 / L11	40,3	29,57	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	6,948	1007876	N1b1a	
I1507	HUNG345a	47,93	21,2	M	WHG	Koros_Hungary_EN_	MathiesonI	half	1240K.c	5780-5640 calBCE (6835±34 BP, OxA-23757)	Tiszaszolos	Hungary	5,103	824434	R1b1	
I0418	SVP24	53,45	50,45	F	Steppe_	Potapovka	MathiesonI	half	1240K.c	2125-1769 calBCE (3583±52 BP, AA-53802)	Utyevka IV,	Russia	0,123	138537	T1a1	
I0422	SVP30	52,54	50,5	F	Steppe_	Srubnaya	MathiesonI	half	1240K.c	1850-1200 BCE	Barinovka I	Russia	1,372	750562	T1a1	
I1099	L11-S-488	40,3	29,57	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	0,827	556176	T2b	
I0025	LBK1992	48,78	9,18	F	Europe_	LBK_EN	MathiesonI	half	1240K.c	5500-4800 BCE	Viesenhaeu	Germany	2,655	866455	T2b	
I0026	LBK2155	48,78	9,18	F	Europe_	LBK_EN	MathiesonI	half	1240K.c	5500-4800 BCE	Viesenhaeu	Germany	3,625	926521	T2b	
I1101	M11-352a	40,3	29,57	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.c	6500-6200 BCE	Barcın	Turkey	1,56	680133	T2b	

I0424	SVP32	53,12	48,37	M	Steppe_	Srubnaya	MathiesonI	half	1240K.ca	1850-1600 BCE	Uvarovka I,	Russia	0,59	517489	T2b4
I0429	SVP38	53,38	50,39	M	Steppe_	Yamnaya_Samara	MathiesonI	half	1240K.ca	3339-2918 calBCE (4432±66 BP, AA-47804)	Lopatino I,	Russia	0,843	616123	T2c1a2
I0022	LBK1976	48,78	9,18	F	Europe_	LBK_EN	MathiesonI	half	1240K.ca	5500-4800 BCE	Viesenhaeu	Germany	0,441	379736	T2e
I0560	QLB18A	51,79	11,14	F	Europe_	Baalberge_MN	MathiesonI	half	1240K.ca	3640-3376 calBCE (4745±52 BP, Er-7856)	Quedlinbur	Germany	0,407	389067	T2e1
I0371	SVP11	53,34	50,34	M	Steppe_	Poltavka	MathiesonI	half	1240K.ca	2872-2583 calBCE (4130±30 BP, Beta-392488)	Grachevka	Russia	0,81	613900	U2d2
I0012	Motala2	58,54	15,05	M	SHG	Motala_HG	MathiesonI	plus	1240K.ca	5714-5575 calBCE (6734±30 BP, Ua-51722)	Motala, Ka	Sweden	1,756	805855	U2e1
I0017	Motala12	58,54	15,05	M	SHG	Motala_HG	MathiesonI	plus	1240K.ca	5721-5631 calBCE (6773±30 BP, Ua-51723)	Motala, Ka	Sweden	4,391	944923	U2e1
I0419	SVP27	53,45	50,45	M	Steppe_	Potapovka	MathiesonI	half	1240K.ca	2200-1900 BCE	Utyevka VI,	Russia	0,516	474599	U2e1h
I0709	BAR20/ M1	40,3	29,57	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.ca	6500-6200 BCE	Barcın	Turkey	9,765	1015958	U3
I1581	L12-502	40,3	29,57	F	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.ca	6500-6200 BCE	Barcın	Turkey	2,803	853125	U3
I1303	MIR25	42,33	-3,5	M	Europe_	Iberia_Chalcolithic	MathiesonI	half	1240K.ca	2900-2346 BCE [2568-2346 calBCE (3950±30 BP, Beta-	El Mirador	Spain	0,733	468021	U3a1
I0551	SALZ3B	51,52	11,85	M	Europe_	Salzmuende_MN	MathiesonI	half	1240K.ca	3400-3025 BCE	Salzmuend	Germany	0,091	104265	U3a1
I0211	UzOO40	61,65	35,65	M	EHG	Karelia_HG	MathiesonI	half	1240K.ca	5500-5000 BCE	Yuzhnyy Ol	Russia	0,136	149746	U4a
I0231	SVP3	52,42	48,24	M	Steppe_	Yamnaya_Samara	MathiesonI	half	1240K.ca	2921-2762 calBCE (4260±30 BP, Beta-392487)	Ekaterinovl	Russia	6,199	1099877	U4a1a or U4a1d
I0434	SVP47	52,22	48,1	M	Steppe_	Samara_Eneolithic	MathiesonI	half	1240K.ca	5200-4000 BCE	Khvalynsk I	Russia	0,053	62724	U4a2 or U4d
I0011	Motala1	58,54	15,05	F	SHG	Motala_HG	MathiesonI	plus	1240K.ca	5721-5516 calBCE (6701±64 BP, Ua-42116)	Motala, Ka	Sweden	1,03	643050	U5a1
I0013	Motala3	58,54	15,05	M	SHG	Motala_HG	MathiesonI	plus	1240K.ca	5964-5638 calBCE (6877±69 BP, Ua-42117)	Motala, Ka	Sweden	0,657	320332	U5a1
I0354	SVP1	53,03	50,39	F	Steppe_	Srubnaya_Outlier	MathiesonI	half	1240K.ca	2014-1692 calBCE (3517±56 BP, AA-47809)	Spiridonovl	Russia	0,39	385103	U5a1
I0360	SVP8	53,03	50,39	M	Steppe_	Srubnaya	MathiesonI	half	1240K.ca	1850-1200 BCE	Spiridonovl	Russia	0,049	58414	U5a1
I0438	SVP50	53,38	50,38	M	Steppe_	Yamnaya_Samara	MathiesonI	half	1240K.ca	3021-2635 calBCE (4254±61 BP, AA-47807)	Luzkhi I, Sa	Russia	0,736	579356	U5a1a1
I0439	SVP52	53,38	50,39	M	Steppe_	Yamnaya_Samara	MathiesonI	half	1240K.ca	3321-2921 calBCE (4420±30 BP, Beta-392491)	Lopatino I,	Russia	0,26	273652	U5a1a1
I0171	BZH12	51,82	10,91	F	Europe_	BenzigerodeHeimbu	MathiesonI	plus	1240K.ca	2287-2041 calBCE (3758±33 BP, KIA-27952)	Benzingero	Germany	0,143	144777	U5a1a2a
I1546	BZH2	51,82	10,91	F	Europe_	Bell_Beaker_Germa	MathiesonI	half	1240K.ca	2500-2050 BCE	Benzingero	Germany	0,143	155325	U5a1b1
I0432	SVP42	53,66	50,67	M	Steppe_	Poltavka_outlier	MathiesonI	half	1240K.ca	2925-2491 calBCE (4180±84 BP, AA-12569)	Potapovka	Russia	0,87	648053	U5a1c
I0124	SVP44	53,4	50,4	M	EHG	Samara_HG	MathiesonI	half	1240K.ca	5657-5541 calBCE (6680±30 BP, Beta-392490)	Lebyanzhin	Russia	0,598	486585	U5a1d
I0232	SVP12	48,1	54,44	M	Steppe_	Srubnaya	MathiesonI	half	1240K.ca	1850-1200 BCE	Novoselki,	Russia	8,202	1153002	U5a1f2
I1536	ESP17	51,42	11,68	M	Europe_	Corded_Ware_Gern	MathiesonI	half	1240K.ca	2500-2050 BCE	Esperstedt	Germany	0,105	116878	U5a1g
I0433	SVP46	52,22	48,1	M	Steppe_	Samara_Eneolithic	MathiesonI	half	1240K.ca	5200-4000 BCE	Khvalynsk I	Russia	0,476	440667	U5a1i
I0359	SVP7	53,03	50,39	F	Steppe_	Srubnaya	MathiesonI	half	1240K.ca	1850-1200 BCE	Spiridonovl	Russia	1,367	754432	U5a2a1
I0014	Motala4	58,54	15,05	F	SHG	Motala_HG	MathiesonI	plus	1240K.ca	5877-5629 calBCE (6842±68 BP, Ua-42118)	Motala, Ka	Sweden	2,709	763001	U5a2d
I0015	Motala6	58,54	15,05	M	SHG	Motala_HG	MathiesonI	plus	1240K.ca	5964-5629 calBCE (6863±75 BP, Ua-42120)	Motala, Ka	Sweden	1,562	611053	U5a2d
I0408	Mina18	41,25	-2,326	F	Europe_	Iberia_MN	MathiesonI	Mixed:ha	1240K.ca	3900-3600 BCE	La Mina	Spain	13,608	958102	U5b1
I0585	LB1	42,91	-5,378	M	WHG	Iberia_HG	MathiesonI	half	1240K.ca	5983-5747 calBCE (6980±50 BP, Beta-226472)	La Brana-A	Spain	19,536	1057168	U5b2c
I1506	PF325, NE1	47,88	21,19	F	Europe_	LBK_Hungary_EN	MathiesonI	half	1240K.ca	5310-5070 calBCE (6237±32 BP, OxA-27861)	Polgar Fere	Hungary	1,517	729538	U5b2c
I0745	M11-363	40,3	29,57	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.ca	6500-6200 BCE	Barcın	Turkey	7,785	1025390	U8b1b1
I1097	BAR271 / M	40,3	29,57	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.ca	6500-6200 BCE	Barcın	Turkey	2,125	776163	W1-T119C
I1549	BZH15	51,82	10,91	F	Europe_	Bell_Beaker_Germa	MathiesonI	half	1240K.ca	2500-2050 BCE	Benzingero	Germany	2,357	809566	W1c1
I0443	SVP57	53,38	50,38	M	Steppe_	Yamnaya_Samara	MathiesonI	half	1240K.ca	3300-2700 BCE	Lopatino II,	Russia	5,317	1025251	W3a1a
I0357	SVP5	53,38	50,39	F	Steppe_	Yamnaya_Samara	MathiesonI	half	1240K.ca	3090-2913 calBCE (4380±30 BP, Beta-392489)	Lopatino I,	Russia	0,625	544122	W6c
I1499	HUNG86, N	48,52	21,17	F	Europe_	Bu_kk_Culture_Hun	MathiesonI	half	1240K.ca	5210-5010 calBCE (6185±34 BP, OxA-27732)	Garadna	Hungary	3,319	847840	X2b-T226C
I0821	HAL24	51,9	11,05	M	Europe_	LBK_EN	MathiesonI	half	1240K.ca	5201-4850 calBCE (6076±34 BP, KIA-40348)	Halberstad	Germany	0,116	128759	X2d1
I1098	BAR99 / M1	40,3	29,57	F	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.ca	6500-6200 BCE	Barcın	Turkey	2,994	809388	X2d2
I0723	T1, M229 /	40,26	29,65	M	Anatolia	Anatolia_Neolithic	MathiesonI	half	1240K.ca	6400-5600 BCE	Mentese	Turkey	1,213	720119	X2m2
CB13	CB13	41,37	1,89	F	Europe_	Iberia_EN	OlaldeMBE	minus	Shotgun	5469-5327 calBCE (6410±30, Beta-384724)	Cova Bonic	Spain	0,979	746511	K1a2a
MA1	Malta1	52,9	103,5	M	MA1	MA1_HG	RaghavanN	minus	Shotgun	22570-22140 calBCE (20240±60 BP, UCIAMS-79666)	Mal'ta	Russia	1,206	841673	..
Kostenki14	Kostenki14	51,23	39,3	M	Kostenki	Kostenki14	SeguinOrla	minus	Shotgun	36730-34310 calBCE (33250±500 BP, OxA-X-2395-15)	Kostenki	Russia	2,806	1104397	..
I1949	GD37	34,45	48,12	M	Iran_N	Ganj_Dareh_Iran_N	This study (	half	1240K.ca	8000-7700 BCE	Ganj Dareh	Iran	0,045	52242	..
I1069	Nat5	32,65	35,07	M	Natufian	Israel_Natufian	This study (	half	1240K.ca	11840-9760 BCE	Raqefet Ca	Israel	0,009	10288	..
I1415	AG84/2	31,99	35,98	M	Levant_I	PPNB	This study (	half	1240K.ca	8197-7653 calBCE (8790±50 BP, Poz-81101)	'Ain Ghazal	Jordan	0,016	19435	..
I1690	NAT6	32,65	35,07	M	Natufian	Israel_Natufian	This study (	half	1240K.ca	11840-9760 BCE	Raqefet Ca	Israel	0,014	16866	..

I1687	NAT13	32,65	35,07	F	Natufian	Israel_Natufian	This study (half	1240K.ca	11520-11110 calBCE (11405±120 BP, RTK-6607)	Raqefet Ca	Israel	0,013	15422	..
I1416	AG83/1	31,99	35,98	M	Levant_N	PPNB	This study (half	1240K.ca	8300-7900 BCE	'Ain Ghazal	Jordan	0,009	11289	..
I1709	AG84_8	31,99	35,98	M	Levant_N	PPNB	This study (half	1240K.ca	8300-7900 BCE	'Ain Ghazal	Jordan	0,012	14816	..
I1679	AG037C	31,99	35,98	F	Levant_N	PPNC	This study (half	1240K.ca	6900-6800 BCE	'Ain Ghazal	Jordan	0,04	47421	I
I1944	GD14B	34,45	48,12	F	Iran_N	Ganj_Dareh_Iran_N	This study (half	1240K.ca	8000-7700 BCE	Ganj Dareh	Iran	0,017	19812	R2
I1685	NAT4	32,65	35,07	M	Natufian	Israel_Natufian	This study (half	1240K.ca	11840-9760 BCE	Raqefet Ca	Israel	0,02	23472	J2a2
I1700	AG88_1	31,99	35,98	M	Levant_N	PPNB	This study (half	1240K.ca	8300-7900 BCE	'Ain Ghazal	Jordan	0,019	22703	T1a2
I1414	AG84/1	31,99	35,98	M	Levant_N	PPNB	This study (half	1240K.ca	8300-7900 BCE	'Ain Ghazal	Jordan	0,024	28156	K1a18
I1951	GD39	34,45	48,12	F	Iran_N	Ganj_Dareh_Iran_N	This study (half	1240K.ca	8202-7681 calBCE (8800±50 BP, Poz-81109)	Ganj Dareh	Iran	0,008	9056	HV0f
I1701	AG83_3	31,99	35,98	F	Levant_N	PPNB	This study (half	1240K.ca	7750-7569 calBCE (8620±50 BP, Poz-81094)	'Ain Ghazal	Jordan	0,009	11347	K1a18
I1293	HotuIIIb	35,59	53,5	M	Iran_Ho	Iran_HotuIIIb	This study (Mixed:h	1240K.ca	9100-8600 BCE (questionable direct date of 7250 ± 40	Hotu Cave	Iran	0,139	150581	HV2
I1727	AG_83_308	31,99	35,98	M	Levant_N	PPNB	This study (half	1240K.ca	8300-7900 BCE	'Ain Ghazal	Jordan	0,041	47767	T1a2
I0861	Nat10	32,65	35,07	M	Natufian	Israel_Natufian	This study (half	1240K.ca	11840-9760 BCE	Raqefet Ca	Israel	0,107	118389	J2a2
I1407	ARE12.1	39,73	45,2	M	Armenia	Armenia_Chalcolith	This study (half	1240K.ca	4350-3700 BCE	Areni1	Armenia	2,164	814272	H
I1705	AG98_1	31,99	35,98	M	Levant_N	Jordan_EBA	This study (half	1240K.ca	2198-1966 calBCE (3690±35 BP, Poz-81096)	'Ain Ghazal	Jordan	2,037	720543	H14a
I1633	KA1/14	40,65	45,12	F	Armenia	Armenia_EBA	This study (half	1240K.ca	2619-2410 calBCE (3990±35 BP, Poz-22234)	Kalavan	Armenia	3,342	859724	H1u
I1661	SG16	34,5	47,96	F	Iran_ChL	Iran_Chalcolithic	This study (half	1240K.ca	4696-4491 calBCE (5740±40 BP, Poz-81104)	Seh Gabi	Iran	1,849	795427	H29
I1634	AR1/44	39,73	45,2	M	Armenia	Armenia_Chalcolith	This study (half	1240K.ca	4330-4060 calBCE (5366±31 BP, OxA-19331)	Areni1	Armenia	1,833	771623	H2a1
I1674	SG21	34,5	47,96	M	Iran_ChL	Iran_Chalcolithic	This study (half	1240K.ca	3972-3800 calBCE (5105±35 BP, Poz-81108)	Seh Gabi	Iran	0,915	640114	I1c
I1945	GD16	34,45	48,12	M	Iran_N	Ganj_Dareh_Iran_N	This study (half	1240K.ca	8000-7700 BCE	Ganj Dareh	Iran	0,112	126090	J1c10
I1662	SG7	34,5	47,96	M	Iran_ChL	Iran_Chalcolithic	This study (half	1240K.ca	4831-4612 calBCE (5860±40 BP, Poz-81105)	Seh Gabi	Iran	0,906	622642	K1a12a
I1671	SG2	34,5	47,96	M	Iran_LN	Iran_Late_Neolithic	This study (half	1240K.ca	5837-5659 calBCE (6850±50 BP, OxA-33168)	Seh Gabi	Iran	0,663	532043	K1a12a
I1584	M10-111	40,3	29,57	F	Anatolia	Anatolia_Chalcolithi	This study (half	1240K.ca	3943-3708 calBCE (5016±31 BP, OxA-32776)	Barcın	Turkey	0,966	597854	K1a17
I0867	Motz1	31,79	35,17	M	Levant_N	PPNB	This study (half	1240K.ca	7300-6750 BCE	Motza	Israel	2,087	772778	K1a4b
I1631	AR1/43C	39,73	45,2	F	Armenia	Armenia_Chalcolith	This study (half	1240K.ca	4250-4050 calBCE (5323±30 BP, OxA-19332)	Areni1	Armenia	3,202	847507	K1a8
I1632	AR1/46	39,73	45,2	M	Armenia	Armenia_Chalcolith	This study (half	1240K.ca	4230-4000 calBCE (5285±29 BP, OxA-18599)	Areni1	Armenia	2,963	835863	K1a8
I1072	NAT9	32,65	35,07	M	Natufian	Israel_Natufian	This study (half	1240K.ca	11840-9760 BCE	Raqefet Ca	Israel	0,526	442605	N1b
I1707	AG83_5	31,99	35,98	M	Levant_N	PPNB	This study (half	1240K.ca	7722-7541 calBCE (8590±50 BP, Poz-81097)	'Ain Ghazal	Jordan	0,142	152234	R0a
I1730	AG_84_308	31,99	35,98	M	Levant_N	Jordan_EBA	This study (half	1240K.ca	2489-2299 calBCE (3925±31 BP, OxA-32775)	'Ain Ghazal	Jordan	1,673	709491	R0a1a
I1699	AG84_5	31,99	35,98	F	Levant_N	PPNC	This study (half	1240K.ca	6800-6700 BCE	'Ain Ghazal	Jordan	0,309	303496	R0a2
I1704	AG89_1	31,99	35,98	F	Levant_N	PPNB	This study (half	1240K.ca	7446-7058 calBCE (8190±60 BP, Poz-81095)	'Ain Ghazal	Jordan	0,164	172819	T1a
I1656	Kat16	40,38	43,94	F	Armenia	Armenia_MBA	This study (half	1240K.ca	1501-1402 calBCE (3168±27 BP, OxA-31674)	Katnaghbiu	Armenia	3,507	861643	T1a1'3
I1710	AG83_6	31,99	35,98	M	Levant_N	PPNB	This study (half	1240K.ca	7733-7526 calBCE (8580±60 BP, Poz-81098)	'Ain Ghazal	Jordan	0,117	129614	T1a2
I1955	GD1150	34,45	48,12	F	Iran_rec	Ganj_Dareh_Iran_re	This study (half	1240K.ca	1430-1485 calCE (430±30 BP, Beta-432801)	Ganj Dareh	Iran	4,231	812079	U1a1
I1670	SG11	34,5	47,96	F	Iran_ChL	Iran_Chalcolithic	This study (half	1240K.ca	4839-4617 calBCE (5870±40 BP, Poz-81107)	Seh Gabi	Iran	0,132	147283	U3a'c
I1658	TA3/R8	40,38	43,87	F	Armenia	Armenia_EBA	This study (half	1240K.ca	3347-3092 calBCE (4492±29 BP, OxA-31874)	Talin	Armenia	3,408	845938	U3a2
I1409	ARE20	39,73	45,2	F	Armenia	Armenia_Chalcolith	This study (half	1240K.ca	4229-3985 calBCE (5260±30 BP, Poz-81110)	Areni1	Armenia	0,083	95691	U4a
I1665	SG19	34,5	47,96	F	Iran_ChL	Iran_Chalcolithic	This study (half	1240K.ca	3956-3796 calBCE (5070±30 BP, Poz-81106)	Seh Gabi	Iran	1,477	758386	U7a
I1290	GD13A	34,45	48,12	F	Iran_N	Ganj_Dareh_Iran_N	This study (half	1240K.ca	8179-7613 calBCE (8780±50 BP, Poz-81100)	Ganj Dareh	Iran	2,098	870535	X2
I1635	KA1/12	40,65	45,12	M	Armenia	Armenia_EBA	This study (half	1240K.ca	2619-2465 calBCE (4005±35 BP, Poz-81102)	Kalavan	Armenia	2,877	841677	X2f
I1706	AG98_2	31,99	35,98	F	Levant_N	Jordan_EBA	This study (half	1240K.ca	2490-2300 BCE	'Ain Ghazal	Jordan	3,749	718277	X2m

Sample ID	Sex	Population ID (terse)	Population ID (verbose)	Country	Town / province / language	Latitude	Longitude	DNA_Source	Access
A306	F	Romanian	Romanian	Romania	Apuseni mountains, Horea village	46,502	22,950	Genomic_from_blood	Fully Public
A325	F	Romanian	Romanian	Romania	Apuseni mountains, Horea village	46,502	22,950	Genomic_from_blood	Fully Public
A343	F	Romanian	Romanian	Romania	Apuseni mountains, Horea village	46,502	22,950	Genomic_from_blood	Fully Public
A362	F	Romanian	Romanian	Romania	Apuseni mountains, Horea village	46,502	22,950	Genomic_from_blood	Fully Public
A374	F	Romanian	Romanian	Romania	Apuseni mountains, Horea village	46,502	22,950	Genomic_from_blood	Fully Public
Assyrian151	F	Assyrian	Assyrian	Turkey	Cilo (Jilu), Hakkâri	37,47	43,94	Genomic_from_saliva	Fully Public
Assyrian152	M	Assyrian	Assyrian	Turkey	Çukurca, Hakkâri	37,28	43,54	Genomic_from_saliva	Fully Public
Assyrian153	M	Assyrian	Assyrian	Turkey	Yüksekova (Gawar), Hakkâri	37,56	44,28	Genomic_from_saliva	Fully Public
Assyrian155	M	Assyrian	Assyrian	Iraq	Mosul	36,34	43,13	Genomic_from_saliva	Fully Public
Assyrian159	F	Assyrian	Assyrian	Iraq	Bibad, Amadiya	37,10	43,45	Genomic_from_saliva	Fully Public
Assyrian160	M	Assyrian	Assyrian	Iran	Gug Tappeh, Urmia	37,51	45,14	Genomic_from_saliva	Fully Public
Assyrian161	M	Assyrian	Assyrian	Iran	Adeh, Urmia	37,70	45,18	Genomic_from_saliva	Fully Public
Assyrian162	M	Assyrian	Assyrian	Iraq	Shaqlawa, Erbil	36,40	44,32	Genomic_from_saliva	Fully Public
Assyrian163	F	Assyrian	Assyrian	Turkey	Yüksekova (Gawar), Hakkâri	37,56	44,28	Genomic_from_saliva	Fully Public
Assyrian164	M	Assyrian	Assyrian	Iran	Urmia	37,54	45,06	Genomic_from_saliva	Fully Public
Assyrian165	F	Assyrian	Assyrian	Turkey	Jilu, Hakkâri	37,47	43,94	Genomic_from_saliva	Fully Public
French23812	M	French	French_West	France	Corse-du-Sud	48,299	-4,123	Genomic_from_blood	Fully Public
French23814	M	French	French_Northwest	France	Nièvre	50,428	3,247	Genomic_from_blood	Fully Public
French23821	F	French	French_Northwest	France	Orne	50,462	2,333	Genomic_from_blood	Fully Public
French23830	M	French	French_West	France	Hérault	48,193	-1,628	Genomic_from_blood	Fully Public
French23833	F	French	French_West	France	Corse-du-Sud	48,299	-4,123	Genomic_from_blood	Fully Public
French23862	M	French	French_Northwest	France	Orne	50,462	2,333	Genomic_from_blood	Fully Public
French23915	M	French	French_Northwest	France	Paris	49,675	0,978	Genomic_from_blood	Fully Public
French23919	F	French	French_East	France	Vosges	47,848	3,593	Genomic_from_blood	Fully Public
French23989	F	French	French_West	France	Creuse	48,475	-2,839	Genomic_from_blood	Fully Public
French24061	M	French	French_West	France	Haute-Loire	47,343	-1,667	Genomic_from_blood	Fully Public
French24075	F	French	French_East	France	Morbihan	49,047	6,615	Genomic_from_blood	Fully Public
French24076	M	French	French_East	France	Morbihan	49,047	6,615	Genomic_from_blood	Fully Public
French24090	M	French	French_West	France	Creuse	48,475	-2,839	Genomic_from_blood	Fully Public
French24118	M	French	French_East	France	Côtes-d'Armor	47,373	4,794	Genomic_from_blood	Fully Public
French24120	F	French	French_West	France	Corse-du-Sud	48,299	-4,123	Genomic_from_blood	Fully Public
French24124	M	French	French_East	France	Morbihan	49,047	6,615	Genomic_from_blood	Fully Public
French24144	F	French	French_Northwest	France	Nièvre	50,428	3,247	Genomic_from_blood	Fully Public
French24148	M	French	French_East	France	Morbihan	49,047	6,615	Genomic_from_blood	Fully Public
French24178	F	French	French_East	France	Morbihan	49,047	6,615	Genomic_from_blood	Fully Public
French24247	F	French	French_West	France	Creuse	48,475	-2,839	Genomic_from_blood	Fully Public
French24381	F	French	French_West	France	Corse-du-Sud	48,299	-4,123	Genomic_from_blood	Fully Public
French24400	M	French	French_West	France	Corse-du-Sud	48,299	-4,123	Genomic_from_blood	Fully Public
French24408	F	French	French_Northwest	France	Paris	49,675	0,978	Genomic_from_blood	Fully Public
French24433	M	French	French_East	France	Morbihan	49,047	6,615	Genomic_from_blood	Fully Public
French24434	F	French	French_East	France	Morbihan	49,047	6,615	Genomic_from_blood	Fully Public

French24437	F	French	French_East	France	Morbihan	49,047	6,615	Genomic_from_blood	Fully Public
French24690	F	French	French_Northwest	France	Orne	50,462	2,333	Genomic_from_blood	Fully Public
French24817	M	French	French_Northwest	France	Aisne	49,548	3,529	Genomic_from_blood	Fully Public
French25068	F	French	French_Northwest	France	Orne	50,462	2,333	Genomic_from_blood	Fully Public
G408	F	Romanian	Romanian	Romania	Gorj county, Tismana village	45,051	22,949	Genomic_from_blood	Fully Public
G421	F	Romanian	Romanian	Romania	Gorj county, Tismana village	45,051	22,949	Genomic_from_blood	Fully Public
G428	F	Romanian	Romanian	Romania	Gorj county, Tismana village	45,051	22,949	Genomic_from_blood	Fully Public
G429	F	Romanian	Romanian	Romania	Gorj county, Tismana village	45,051	22,949	Genomic_from_blood	Fully Public
G434	F	Romanian	Romanian	Romania	Gorj county, Tismana village	45,051	22,949	Genomic_from_blood	Fully Public
ITS2	F	Italian_South	Italian_South	Italy	Naples	40,510	16,550	Genome_from_blood	Fully public
ITS4	F	Italian_South	Italian_South	Italy	Naples	40,680	14,770	Genome_from_blood	Fully public
ITS5	F	Italian_South	Italian_South	Italy	Salerno	40,390	16,550	Genome_from_blood	Fully public
ITS7	F	Italian_South	Italian_South	Italy	Crispiano	40,850	14,270	Genome_from_blood	Fully Public
Lebanese10AQ127	F	Lebanese_Muslim	Lebanese_Muslim	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese10AR37	M	Lebanese_Christian	Lebanese_Christian	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese11AS14	F	Lebanese_Muslim	Lebanese_Muslim	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese15AR37	M	Lebanese_Christian	Lebanese_Christian	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese1AQ127	F	Lebanese_Christian	Lebanese_Christian	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese1AQ170	F	Lebanese_Christian	Lebanese_Christian	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese20AR21	F	Lebanese_Muslim	Lebanese_Muslim	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese22BA23	F	Lebanese_Christian	Lebanese_Christian	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese24AR27	M	Lebanese_Muslim	Lebanese_Muslim	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese2AQ121	F	Lebanese_Muslim	Lebanese_Muslim	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese2AQ127	F	Lebanese_Muslim	Lebanese_Muslim	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese30AR21	F	Lebanese_Muslim	Lebanese_Muslim	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese4AQ115	M	Lebanese_Christian	Lebanese_Christian	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese6AQ115	M	Lebanese_Christian	Lebanese_Christian	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese6AQ170	F	Lebanese_Christian	Lebanese_Christian	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese6AS15	M	Lebanese_Muslim	Lebanese_Muslim	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese7AQ150	M	Lebanese_Muslim	Lebanese_Muslim	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese7AR20	M	Lebanese_Muslim	Lebanese_Muslim	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese7AR23	M	Lebanese_Muslim	Lebanese_Muslim	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
Lebanese8AS15	M	Lebanese_Christian	Lebanese_Christian	Lebanon	Beirut	33,890	35,500	Genome_from_blood	Fully Public
LIB13	F	Libyan	Libyan	Libya	Tripoli	32,880	13,190	Genome_from_blood	Fully Public
LIB18	F	Libyan	Libyan	Libya	Tripoli	32,880	13,190	Genome_from_blood	Fully Public
LIB27	F	Libyan	Libyan	Libya	Tripoli	32,880	13,190	Genome_from_blood	Fully Public
LIB30	F	Libyan	Libyan	Libya	Tripoli	32,880	13,190	Genome_from_blood	Fully Public
LIB7	F	Libyan	Libyan	Libya	Tripoli	32,880	13,190	Genome_from_blood	Fully Public
MCA14	F	Moroccan	Moroccan	Morocco	Casablanca	33,530	-7,580	Genome_from_blood	Fully Public
MCA16	F	Moroccan	Moroccan	Morocco	Casablanca	33,530	-7,580	Genome_from_blood	Fully Public
MCA19	F	Moroccan	Moroccan	Morocco	Casablanca	33,530	-7,580	Genome_from_blood	Fully Public
MCA24	F	Moroccan	Moroccan	Morocco	Casablanca	33,530	-7,580	Genome_from_blood	Fully Public
MCA37	F	Moroccan	Moroccan	Morocco	Casablanca	33,530	-7,580	Genome_from_blood	Fully Public
MCA38	F	Moroccan	Moroccan	Morocco	Casablanca	33,530	-7,580	Genome_from_blood	Fully Public

MCA39	F	Moroccan	Moroccan	Morocco	Casablanca	33,530	-7,580	Genome_from_blood	Fully Public
MCA7	F	Moroccan	Moroccan	Morocco	Casablanca	33,530	-7,580	Genome_from_blood	Fully Public
MCA8	F	Moroccan	Moroccan	Morocco	Casablanca	33,530	-7,580	Genome_from_blood	Fully Public
MCA9	F	Moroccan	Moroccan	Morocco	Casablanca	33,530	-7,580	Genome_from_blood	Fully Public
PV001	M	Iranian_Bandari	Iranian_Bandari	Iran	Bandar Abbas / Hormozgan	27,183	56,267	Genomic_from_blood	Fully Public
PV002	M	Iranian_Bandari	Iranian_Bandari	Iran	Bandar Abbas / Hormozgan	27,183	56,267	Genomic_from_blood	Fully Public
PV003	M	Iranian_Bandari	Iranian_Bandari	Iran	Bandar Abbas / Hormozgan	27,183	56,267	Genomic_from_blood	Fully Public
PV004	M	Iranian_Bandari	Iranian_Bandari	Iran	Bandar Abbas / Hormozgan	27,183	56,267	Genomic_from_blood	Fully Public
PV005	M	Iranian_Bandari	Iranian_Bandari	Iran	Bandar Abbas / Hormozgan	27,183	56,267	Genomic_from_blood	Fully Public
PV006	M	Iranian_Bandari	Iranian_Bandari	Iran	Bandar Abbas / Hormozgan	27,183	56,267	Genomic_from_blood	Fully Public
PV007	M	Iranian_Bandari	Iranian_Bandari	Iran	Bandar Abbas / Hormozgan	27,183	56,267	Genomic_from_blood	Fully Public
PV008	M	Iranian_Bandari	Iranian_Bandari	Iran	Bandar Abbas / Hormozgan	27,183	56,267	Genomic_from_blood	Fully Public
PV009	M	Iranian	Iranian_Lor	Iran	Khoramabad / Lorestan / Lori	33,465	48,339	Genomic_from_blood	Fully Public
PV010	M	Iranian	Iranian_Lor	Iran	Khoramabad / Lorestan / Lori	33,465	48,339	Genomic_from_blood	Fully Public
PV011	M	Iranian	Iranian_Lor	Iran	Khoramabad / Lorestan / Lori	33,465	48,339	Genomic_from_blood	Fully Public
PV012	M	Iranian	Iranian_Lor	Iran	Khoramabad / Lorestan / Lori	33,465	48,339	Genomic_from_blood	Fully Public
PV013	M	Iranian	Iranian_Lor	Iran	Khoramabad / Lorestan / Lori	33,465	48,339	Genomic_from_blood	Fully Public
PV014	M	Iranian	Iranian_Lor	Iran	Khoramabad / Lorestan / Lori	33,465	48,339	Genomic_from_blood	Fully Public
PV015	M	Iranian	Iranian_Lor	Iran	Khoramabad / Lorestan / Lori	33,465	48,339	Genomic_from_blood	Fully Public
PV016	M	Iranian	Iranian_Lor	Iran	Khoramabad / Lorestan / Lori	33,465	48,339	Genomic_from_blood	Fully Public
PV017	M	Iranian	Iranian_Lor	Iran	Khoramabad / Lorestan / Lori	33,465	48,339	Genomic_from_blood	Fully Public
PV018	M	Iranian	Iranian_Lor	Iran	Khoramabad / Lorestan / Lori	33,465	48,339	Genomic_from_blood	Fully Public
PV019	M	Iranian	Iranian_Mazandarani	Iran	Babol / Mazandaran / Mazandarani	36,529	52,671	Genomic_from_blood	Fully Public
PV020	M	Iranian	Iranian_Mazandarani	Iran	Babol / Mazandaran / Mazandarani	36,529	52,671	Genomic_from_blood	Fully Public
PV021	M	Iranian	Iranian_Mazandarani	Iran	Babol / Mazandaran / Mazandarani	36,529	52,671	Genomic_from_blood	Fully Public
PV022	M	Iranian	Iranian_Mazandarani	Iran	Babol / Mazandaran / Mazandarani	36,529	52,671	Genomic_from_blood	Fully Public
PV023	M	Iranian	Iranian_Mazandarani	Iran	Babol / Mazandaran / Mazandarani	36,529	52,671	Genomic_from_blood	Fully Public
PV024	M	Iranian	Iranian_Mazandarani	Iran	Babol / Mazandaran / Mazandarani	36,529	52,671	Genomic_from_blood	Fully Public
PV025	M	Iranian	Iranian_Mazandarani	Iran	Babol / Mazandaran / Mazandarani	36,529	52,671	Genomic_from_blood	Fully Public
PV026	M	Iranian	Iranian_Mazandarani	Iran	Babol / Mazandaran / Mazandarani	36,529	52,671	Genomic_from_blood	Fully Public
PV027	M	Iranian	Iranian_Mazandarani	Iran	Babol / Mazandaran / Mazandarani	36,529	52,671	Genomic_from_blood	Fully Public
PV028	M	Iranian	Iranian_Mazandarani	Iran	Babol / Mazandaran / Mazandarani	36,529	52,671	Genomic_from_blood	Fully Public
PV029	M	Iranian	Iranian_Persian	Iran	Shiraz / Fars / Persian	29,591	52,584	Genomic_from_blood	Fully Public
PV030	M	Iranian	Iranian_Persian	Iran	Shiraz / Fars / Persian	29,591	52,584	Genomic_from_blood	Fully Public
PV031	M	Iranian	Iranian_Persian	Iran	Shiraz / Fars / Persian	29,591	52,584	Genomic_from_blood	Fully Public
PV032	M	Iranian	Iranian_Persian	Iran	Shiraz / Fars / Persian	29,591	52,584	Genomic_from_blood	Fully Public
PV033	M	Iranian	Iranian_Persian	Iran	Shiraz / Fars / Persian	29,591	52,584	Genomic_from_blood	Fully Public
PV034	M	Iranian	Iranian_Persian	Iran	Shiraz / Fars / Persian	29,591	52,584	Genomic_from_blood	Fully Public
PV035	M	Iranian	Iranian_Persian	Iran	Shiraz / Fars / Persian	29,591	52,584	Genomic_from_blood	Fully Public
PV036	M	Iranian	Iranian_Persian	Iran	Shiraz / Fars / Persian	29,591	52,584	Genomic_from_blood	Fully Public
PV037	M	Iranian	Iranian_Persian	Iran	Shiraz / Fars / Persian	29,591	52,584	Genomic_from_blood	Fully Public
PV038	M	Iranian	Iranian_Persian	Iran	Shiraz / Fars / Persian	29,591	52,584	Genomic_from_blood	Fully Public
Irish0063	F	Irish_Ulster	Irish_Ulster	Ireland	Ulster	54,710	-6,950	Genomic_from_saliva	Letter
Irish0067	M	Irish_Ulster	Irish_Ulster	Ireland	Ulster	54,710	-6,950	Genomic_from_saliva	Letter

Irish0073	M	Irish	Irish_Leinster	Ireland	Leinster	53,090	-6,880	Genomic_from_saliva	Letter
Irish0074	M	Irish	Irish_Leinster	Ireland	Leinster	53,090	-6,880	Genomic_from_saliva	Letter
Irish0079	M	Irish_Ulster	Irish_Ulster	Ireland	Ulster	54,710	-6,950	Genomic_from_saliva	Letter
Irish0083	F	Irish	Irish_Cork_Kerry	Ireland	West Cork/South Kerry	51,840	-9,200	Genomic_from_saliva	Letter
Irish0085	M	Irish	Irish_Leinster	Ireland	Leinster	53,090	-6,880	Genomic_from_saliva	Letter
Irish0089	M	Irish	Irish_Cork_Kerry	Ireland	West Cork/South Kerry	51,840	-9,200	Genomic_from_saliva	Letter
Irish0091	M	Irish	Irish_Cork_Kerry	Ireland	West Cork/South Kerry	51,840	-9,200	Genomic_from_saliva	Letter
Irish0093	M	Irish	Irish_Connacht	Ireland	Connacht	53,720	-8,490	Genomic_from_saliva	Letter
Irish0094	M	Irish	Irish_Cork_Kerry	Ireland	West Cork/South Kerry	51,840	-9,200	Genomic_from_saliva	Letter
Irish0097	M	Irish	Irish_Connacht	Ireland	Connacht	53,720	-8,490	Genomic_from_saliva	Letter
Irish0129	M	Irish	Irish_Cork_Kerry	Ireland	West Cork/South Kerry	51,840	-9,200	Genomic_from_saliva	Letter
Irish0131	M	Irish_Ulster	Irish_Ulster	Ireland	Ulster	54,710	-6,950	Genomic_from_saliva	Letter
Irish0132	M	Irish	Irish_Leinster	Ireland	Leinster	53,090	-6,880	Genomic_from_saliva	Letter
Irish0135	F	Irish_Ulster	Irish_Ulster	Ireland	Ulster	54,710	-6,950	Genomic_from_saliva	Letter
Irish0137	F	Irish_Ulster	Irish_Ulster	Ireland	Ulster	54,710	-6,950	Genomic_from_saliva	Letter
Irish0139	M	Irish	Irish_Munster	Ireland	Munster	52,360	-8,490	Genomic_from_saliva	Letter
Irish0145	M	Irish	Irish_Munster	Ireland	Munster	52,360	-8,490	Genomic_from_saliva	Letter
Irish0151	M	Irish	Irish_Cork_Kerry	Ireland	West Cork/South Kerry	51,840	-9,200	Genomic_from_saliva	Letter
Irish0157	M	Irish	Irish_Munster	Ireland	Munster	52,360	-8,490	Genomic_from_saliva	Letter
Irish0158	F	Irish	Irish_Munster	Ireland	Munster	52,360	-8,490	Genomic_from_saliva	Letter
Irish0175	F	Irish	Irish_Connacht	Ireland	Connacht	53,720	-8,490	Genomic_from_saliva	Letter
Irish0177	M	Irish	Irish_Connacht	Ireland	Connacht	53,720	-8,490	Genomic_from_saliva	Letter
Irish0180	M	Irish	Irish_Leinster	Ireland	Leinster	53,090	-6,880	Genomic_from_saliva	Letter
Irish0181	M	Irish	Irish_Connacht	Ireland	Connacht	53,720	-8,490	Genomic_from_saliva	Letter
Irish0182	F	Irish	Irish_Connacht	Ireland	Connacht	53,720	-8,490	Genomic_from_saliva	Letter
Irish0185	F	Irish	Irish_Leinster	Ireland	Leinster	53,090	-6,880	Genomic_from_saliva	Letter
Irish0188	M	Irish	Irish_Munster	Ireland	Munster	52,360	-8,490	Genomic_from_saliva	Letter
Irish0189	F	Irish	Irish_Munster	Ireland	Munster	52,360	-8,490	Genomic_from_saliva	Letter
leipzig2659	M	German	German_Lipsian	Germany	Leipzig, German	51,340	12,373	Genomic_from_blood	Letter
leipzig2660	M	German	German_Lipsian	Germany	Leipzig, German	51,340	12,373	Genomic_from_blood	Letter
leipzig2661	M	German	German_Lipsian	Germany	Leipzig, German	51,340	12,373	Genomic_from_blood	Letter
leipzig2662	F	German	German_Lipsian	Germany	Leipzig, German	51,340	12,373	Genomic_from_blood	Letter
leipzig2705	M	German	German_Lipsian	Germany	Leipzig, German	51,340	12,373	Genomic_from_blood	Letter
leipzig2722	F	German	German_Lipsian	Germany	Leipzig, German	51,340	12,373	Genomic_from_blood	Letter
leipzig2785	F	German	German_Lipsian	Germany	Leipzig, German	51,340	12,373	Genomic_from_blood	Letter
leipzig2790	F	German	German_Lipsian	Germany	Leipzig, German	51,340	12,373	Genomic_from_blood	Letter
leipzig2823	F	German	German_Lipsian	Germany	Leipzig, German	51,340	12,373	Genomic_from_blood	Letter
leipzigBG80609	F	German	German_Lipsian	Germany	Leipzig, German	51,340	12,373	Genomic_from_blood	Letter
Polish101A	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish102B	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish103B	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish104A	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish106A	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish107A	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter

Polish108A	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish111A	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish113A	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish114B	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish115B	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish117B	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish119A	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish120A	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish121A	M	Polish	Polish	Poland	Poznan, Wielkopolskie	52,240	16,920	Genomic_from_blood	Letter
Polish201B	M	Polish	Polish	Poland	Lublin, Lubelskie	51,250	22,567	Genomic_from_blood	Letter
Polish202A	M	Polish	Polish	Poland	Lublin, Lubelskie	51,250	22,567	Genomic_from_blood	Letter
Polish206C	F	Polish	Polish	Poland	Lublin, Lubelskie	51,250	22,567	Genomic_from_blood	Letter
Polish207A	M	Polish	Polish	Poland	Lublin, Lubelskie	51,250	22,567	Genomic_from_blood	Letter
Polish209B	M	Polish	Polish	Poland	Lublin, Lubelskie	51,250	22,567	Genomic_from_blood	Letter
pos3	F	Polish	Polish	Poland	Lublin, Lubelskie	51,250	22,567	Genomic_from_blood	Letter
pos4	F	Polish	Polish	Poland	Lublin, Lubelskie	51,250	22,567	Genomic_from_blood	Letter
pos5	F	Polish	Polish	Poland	Lublin, Lubelskie	51,250	22,567	Genomic_from_blood	Letter
Scottish0111	M	Shetlandic	Scottish_Shetlandic	GreatBritain	Shetland	60,150	-1,150	Genomic_from_blood	Letter
Scottish0153	M	Shetlandic	Scottish_Shetlandic	GreatBritain	Shetland	60,150	-1,150	Genomic_from_blood	Letter
Scottish0236	M	Shetlandic	Scottish_Shetlandic	GreatBritain	Shetland	60,150	-1,150	Genomic_from_blood	Letter
Scottish0237	M	Shetlandic	Scottish_Shetlandic	GreatBritain	Shetland	60,150	-1,150	Genomic_from_blood	Letter
Scottish0347	M	Shetlandic	Scottish_Shetlandic	GreatBritain	Shetland	60,150	-1,150	Genomic_from_blood	Letter
Scottish0356	M	Shetlandic	Scottish_Shetlandic	GreatBritain	Shetland	60,150	-1,150	Genomic_from_blood	Letter
Scottish0462	M	Shetlandic	Scottish_Shetlandic	GreatBritain	Shetland	60,150	-1,150	Genomic_from_blood	Letter
Scottish0683	M	Shetlandic	Scottish_Shetlandic	GreatBritain	Shetland	60,150	-1,150	Genomic_from_blood	Letter
Scottish1045	M	Shetlandic	Scottish_Shetlandic	GreatBritain	Shetland	60,150	-1,150	Genomic_from_blood	Letter
Scottish1288	M	Shetlandic	Scottish_Shetlandic	GreatBritain	Shetland	60,150	-1,150	Genomic_from_blood	Letter
Scottish1334	F	Scottish	Scottish_Borders	GreatBritain	Borders	55,500	-2,500	Genomic_from_blood	Letter
Scottish1338	F	Scottish	Scottish_Borders	GreatBritain	Borders	55,500	-2,500	Genomic_from_blood	Letter
Scottish1340	F	Scottish	Scottish_Borders	GreatBritain	Borders	55,500	-2,500	Genomic_from_blood	Letter
Scottish1341	F	Scottish	Scottish_Borders	GreatBritain	Borders	55,500	-2,500	Genomic_from_blood	Letter
Scottish1343	F	Scottish	Scottish_Borders	GreatBritain	Borders	55,500	-2,500	Genomic_from_blood	Letter
Scottish1344	F	Scottish	Scottish_Borders	GreatBritain	Borders	55,500	-2,500	Genomic_from_blood	Letter
Scottish2401	F	Scottish	Scottish_Dumfries_Galloway	GreatBritain	Dumfries & Galloway	55,000	-4,000	Genomic_from_blood	Letter
Scottish2433	F	Scottish	Scottish_Dumfries_Galloway	GreatBritain	Dumfries & Galloway	55,000	-4,000	Genomic_from_blood	Letter
Scottish2436	F	Scottish	Scottish_Dumfries_Galloway	GreatBritain	Dumfries & Galloway	55,000	-4,000	Genomic_from_blood	Letter
Scottish2441	F	Scottish	Scottish_Dumfries_Galloway	GreatBritain	Dumfries & Galloway	55,000	-4,000	Genomic_from_blood	Letter
Scottish2443	F	Scottish	Scottish_Dumfries_Galloway	GreatBritain	Dumfries & Galloway	55,000	-4,000	Genomic_from_blood	Letter
Scottish3140	F	Scottish	Scottish_Fife	GreatBritain	Fife & Central	56,250	-3,000	Genomic_from_blood	Letter
Scottish3146	F	Scottish	Scottish_Fife	GreatBritain	Fife & Central	56,250	-3,000	Genomic_from_blood	Letter
Scottish3151	F	Scottish	Scottish_Fife	GreatBritain	Fife & Central	56,250	-3,000	Genomic_from_blood	Letter
Scottish3154	F	Scottish	Scottish_Fife	GreatBritain	Fife & Central	56,250	-3,000	Genomic_from_blood	Letter
Scottish3157	F	Scottish	Scottish_Fife	GreatBritain	Fife & Central	56,250	-3,000	Genomic_from_blood	Letter
Scottish3212	F	Scottish	Scottish_Fife	GreatBritain	Fife & Central	56,250	-3,000	Genomic_from_blood	Letter

Scottish3512	F	Scottish	Scottish_Highlands	GreatBritain	Scottish Highlands (mainland)	57,500	-4,250	Genomic_from_blood	Letter
Scottish3521	F	Scottish	Scottish_Highlands	GreatBritain	Scottish Highlands (mainland)	57,500	-4,250	Genomic_from_blood	Letter
Scottish3529	F	Scottish	Scottish_Highlands	GreatBritain	Scottish Highlands (mainland)	57,500	-4,250	Genomic_from_blood	Letter
Scottish3531	F	Scottish	Scottish_Highlands	GreatBritain	Scottish Highlands (mainland)	57,500	-4,250	Genomic_from_blood	Letter
Scottish3532	F	Scottish	Scottish_Highlands	GreatBritain	Scottish Highlands (mainland)	57,500	-4,250	Genomic_from_blood	Letter
Scottish3533	F	Scottish	Scottish_Highlands	GreatBritain	Scottish Highlands (mainland)	57,500	-4,250	Genomic_from_blood	Letter
Scottish4283	F	Scottish	Scottish_Grampian	GreatBritain	Grampian	57,250	-2,250	Genomic_from_blood	Letter
Scottish4290	F	Scottish	Scottish_Grampian	GreatBritain	Grampian	57,250	-2,250	Genomic_from_blood	Letter
Scottish4292	F	Scottish	Scottish_Grampian	GreatBritain	Grampian	57,250	-2,250	Genomic_from_blood	Letter
Scottish4293	F	Scottish	Scottish_Grampian	GreatBritain	Grampian	57,250	-2,250	Genomic_from_blood	Letter
Scottish4294	F	Scottish	Scottish_Grampian	GreatBritain	Grampian	57,250	-2,250	Genomic_from_blood	Letter
Scottish4300	F	Scottish	Scottish_Grampian	GreatBritain	Grampian	57,250	-2,250	Genomic_from_blood	Letter
sorb2908523	M	Sorb	German_Sorb	Germany	Kamenz, Sorbian	51,275	14,102	Genomic_from_blood	Letter
sorb2908543	F	Sorb	German_Sorb	Germany	Kamenz, Sorbian	51,275	14,102	Genomic_from_blood	Letter
sorb2908546	F	Sorb	German_Sorb	Germany	Kamenz, Sorbian	51,275	14,102	Genomic_from_blood	Letter
sorb2908672	F	Sorb	German_Sorb	Germany	Kamenz, Sorbian	51,275	14,102	Genomic_from_blood	Letter
sorb2908808	M	Sorb	German_Sorb	Germany	Kamenz, Sorbian	51,275	14,102	Genomic_from_blood	Letter
sorb2908809	M	Sorb	German_Sorb	Germany	Kamenz, Sorbian	51,275	14,102	Genomic_from_blood	Letter
sorb2908823	F	Sorb	German_Sorb	Germany	Kamenz, Sorbian	51,275	14,102	Genomic_from_blood	Letter
sorb2908843	F	Sorb	German_Sorb	Germany	Kamenz, Sorbian	51,275	14,102	Genomic_from_blood	Letter
sorb2908869	F	Sorb	German_Sorb	Germany	Kamenz, Sorbian	51,275	14,102	Genomic_from_blood	Letter

Fst computed on the HO dataset

Fst below and its standard error above the diagonal

Values above 0.02, 0.04, and 0.06 are highlighted in color

	Abkhasian	Adyge	Albanian	**Anatolia_N**	**Armenia_ChL**	**Armenia_EBA**	**Armenia_MLBA**	Armenian	Assyrian	Balkar	Basque	BedouinA	BedouinB	Belarusian	Bulgarian
Abkhasian		0,000	0,001	0,001	0,001	0,001	0,001	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,000
Adygei	0,003		0,000	0,000	0,001	0,001	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Albanian	0,008	0,008		0,001	0,001	0,001	0,001	0,000	0,001	0,001	0,000	0,000	0,001	0,001	0,000
Anatolia_N	0,028	0,029	0,017		0,001	0,001	0,001	0,000	0,000	0,001	0,000	0,000	0,001	0,001	0,000
Armenia_ChL	0,018	0,018	0,017	0,026		0,002	0,002	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
Armenia_EBA	0,007	0,008	0,011	0,022	0,012		0,002	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
Armenia_MLBA	0,003	0,005	0,007	0,023	0,015	0,002		0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
Armenian	0,004	0,005	0,005	0,021	0,015	0,007	0,003		0,000	0,000	0,000	0,000	0,001	0,000	0,000
Assyrian	0,007	0,008	0,008	0,023	0,018	0,008	0,006	0,003		0,000	0,000	0,000	0,000	0,001	0,000
Balkar	0,003	0,003	0,007	0,029	0,018	0,008	0,005	0,005	0,008		0,000	0,000	0,001	0,000	0,000
Basque	0,020	0,018	0,010	0,026	0,028	0,024	0,018	0,017	0,020	0,018		0,000	0,001	0,000	0,000
BedouinA	0,012	0,013	0,010	0,023	0,022	0,014	0,010	0,006	0,008	0,012	0,021		0,000	0,000	0,000
BedouinB	0,032	0,033	0,030	0,040	0,042	0,033	0,030	0,025	0,027	0,033	0,041	0,018		0,001	0,001
Belarusian	0,016	0,013	0,006	0,030	0,025	0,020	0,012	0,015	0,018	0,012	0,012	0,020	0,041		0,000
Bulgarian	0,008	0,008	0,000	0,018	0,017	0,012	0,005	0,006	0,009	0,007	0,009	0,011	0,031	0,004	
Canary_Islander	0,010	0,009	0,001	0,017	0,019	0,014	0,008	0,006	0,009	0,009	0,005	0,008	0,028	0,006	0,001
Chechen	0,006	0,006	0,009	0,032	0,020	0,010	0,006	0,007	0,010	0,006	0,020	0,014	0,035	0,014	0,009
CHG	0,033	0,036	0,049	0,073	0,050	0,030	0,033	0,040	0,042	0,037	0,063	0,050	0,072	0,054	0,049

doi:10.1038/nature19310

Genomic insights into the origin of farming in the ancient Near East

Table of Contents 1

Supplementary Information 1

Description of ancient samples and archaeological context

This note presents information on the archaeological context of 45 individuals from the ancient Near East for whom we report genome-wide data in this study. We give uncalibrated dates in radiocarbon years ("bp"). We give calibrated dates in years before the common era, converting from uncalibrated to calibrated dates using IntCal13[1] and OxCal4.2[2] (https://c14.arch.ox.ac.uk/oxcal/OxCal.html). We indicate dates obtained directly on the skeleton for which we obtained ancient DNA by the suffix "calBCE". We indicate indirect dates obtained based on archaeological context as "BCE".

Iran

Hotu Cave (Iran)

Hotu Cave is located in the foothills of the Alborz Mountains, near the modern town of Behshahr on the southern shore of the Caspian Sea in northern Iran. The site was excavated to a depth of 12 meters by the American physical anthropologist Carleton Coon in the spring of 1951, and provides evidence of human occupation at various times from the early Mesolithic to the post-Achaemenid periods[3-5]. Skeletal remains of three sets of individuals (I, II, III) were recovered from Pleistocene Gravel 4 of Trench D, the deepest sounding undertaken (Section 2)[3]. In Coon's preliminary excavation report, this gravel was initially identified as an aceramic Neolithic level, but was subsequently reclassified as a Mesolithic horizon.[1,2,3] In 1955, two charcoal samples recovered in association with the skeletal remains from depths of 9.5 and 10.15 meters yielded uncalibrated radiocarbon ages of 9190±590 bp and 9270±570 bp respectively[6]. In 2013, AMS analyses of collagen extracted from a tooth from Mesolithic skeleton I yielded an uncalibrated radiocarbon age of 9480±40 bp (9119-8637 calBCE)[7]. The set III of individuals includes one adult partially complete skeleton (HotuIIIa) and the remains of a young child estimated to be 1.5-2 years old (HotuIIIb). We sampled the right petrous bone of HotuIIIb for DNA analysis.

Subsequent AMS analysis of this specimen yielded an uncalibrated radiocarbon age of 7250 ± 40 bp (6218-6034 calBCE). The discrepancy in the direct dates of skeletal remains (Hotu I and IIIb) recovered from such close proximity to one another in the same archaeological horizon causes us to doubt the age of the Hotu IIIb specimen. There are no stratigraphic unconformities indicating that this specimen may have been introduced from overlying Neolithic cave deposits, and we suspect that the later than anticipated date may result from the use of modern-day organic materials such as the shellac or glue in post-excavation reconstruction of the Hotu IIIb crania. Genetic similarities between the CHG and Hotu IIIb specimens demonstrate that there are continuities between Iranian Neolithic

populations and local hunter-gatherer groups in northern Iran and the Caucasus, rather than with Epipalaeolithic or Neolithic populations originating in Anatolia or the Levant. The earlier date (9119-8637 BCE)[8] obtained from the Hotu I skeleton is not essential to support this observation, as the continuity of Iran Neolithic with hunter-gatherers of the southern Caucasus/Iran highlands rather than with those of the Levant is evident when either the HotuIIIb specimen or the securely dated pre-Neolithic Caucasus hunter-gatherers[8] from Georgia are used (main text). We show a conservatively wide range that includes both the earliest date from Hotu I and the later direct date from HotuIIIb in Fig. 1a, although based on the arguments made here we strongly favor the earlier date and Mesolithic attribution of the HotuIIIb individual.

• HotuIIIb (I1293): 1.5-2 year old individual.

Ganj Dareh (Iran)

Tepe Ganj Dareh is a mound on the floor of the Gamas-Ab Valley, and is situated at an altitude of ~1,400 meters in the High Zagros region of Kermanshah Province in western Iran. It is one of several mounds discovered during survey work in the area[9]. It measures about 40 meters in diameter and 7 to 8 meters in height and has five identified levels (A to E), with the lowest level E being the oldest. Permanent architecture is seen earliest at level D. Philip E. L. Smith led an excavation of roughly 20% of the mound in four seasons between 1967 and 1974. Zooarchaeological and archaeobotanical evidence show a population exploiting ovicaprids, with goats dominant, as well as evidence for use of wild barley but no plant domesticates. There is evidence of herding but no evidence for decreased size or changes in horn core morphology[10,11]. Current evidence places the occupation of the site at approximately 8000-7700 calBCE[11]. None of the human remains have been directly dated. The human remains are highly fragmentary. The current minimum number of individual is 116, of which 56 are catalogued as skeletons each represented by more than 4 elements[12]. The following six Ganj Dareh individuals (all petrous bone samples) are included in this study:

• GD13A (I1290): 30-50 year-old female from level C.
• GD14B (I1944): 3-4 year-old child from level B1.
• GD16 (I1945): 5.5-7.5 year-old child from level D.
• GD37 (I1949): 30-50 year-old male from levels D/E.
• GD39 (I1951): 1.5-2.5 year-old child from level D crypt.
• GD1150 (I1955): 18-30 year-old male from levels A/D. This individual is a clear genetic outlier and was analyzed separately from the other individuals pending radiocarbon dating, which confirmed that it was not from the Neolithic period but of recent origin (1430-1485 calCE (330±30 BP, Beta-432801); it is thus labeled Iran_recent.

Seh Gabi (Iran)

The site of Seh Gabi is a series of seven small mounds, identified as A through G, on a branch of the Gamas Ab River in the Kangavar Valley of the High Zagros region of Kermanshah Province, western Iran, about 6 kilometers from the much larger Chalcolithic and Bronze Age site of Godin Tepe. The mounds were scattered over an area of ~15 hectares with four excavated by Louis D. Levine and the Royal Ontario Museum between 1971 and 1973. Occupation lasted from the Late Neolithic through the Chalcolithic periods[13]. The human material, recovered in 1971 and 1973, comprises a minimum of 31 individuals that are almost entirely subadults: fetuses, infants and children <4 years of age; many with evidence of pathology. We report ancient DNA data on six samples (all petrous bones) from Mound A (n=2), Mound B (n=3), and Mound C (n=1). No radiocarbon dates have been obtained on Mound A, and so we use archaeological dates for the Chalcolithic of 4500-3500 BCE. Radiocarbon dates have been obtained from Mound B although not on the skeletons for which we report DNA. The dates obtained range from 5630±80 bp (SI-4915) to 5020±70 bp (SI-4910), and hence we represent the dates for samples from this bound by the union of the 95.4% calibrated confidence intervals (4680-3662 calBCE[14,15]). We directly dated the Mound C individual who is Late Neolithic:

• SG19 (I1665): fetal remains, ~26 weeks old from Mound A.
• SG21 (I1674): fetal remains, ~35 weeks old from Mound A.
• SG7 (I1662): fetal remains, ~34 weeks old from Mound B.
• SG11 (I1670): ~6 month-old infant from Mound B.
• SG16 (I1661): ~9 months old infant from Mound B.
• SG2 (I1671): 5837-5659 calBCE (OxA-33168, 6850±40 bp). ~6 month-old infant from Mound C.

Levant

Raqefet Cave (Israel)

Raqefet Cave is located within a southern extension of Mt. Carmel. It has a long prehistoric sequence spanning the Middle Palaeolithic through to the Neolithic periods[16]. The first excavation was led by Noy and Higgs (1970-1972), and the most recent excavation was led by Nadel and Lengyel (2004-2011). The first chamber was used extensively by the Natufians, who left on the terrace 100 bedrock mortars, cupmarks and cupules[17]. Some of the mortars are among the largest ever recorded for Natufian sites[18], and one of the deepest has a grid-like pattern incised inside the shaft[19]. The samples we analyze for ancient DNA (all of which are petrous bones) were found in single or double primary burials lying on their back or on their side in flexed positions. These are part of a cluster of Natufian graves created within a short time span in a dedicated area of the first chamber of the cave. The area

contains approximately 30 human skeletons, all but one of which are in an area of about 15 meters squared. Faunal analysis indicates feasting by the open graves and the burial of food remains within the graves[20]. Phytolith analysis indicates the use of plant materials in the graves and in the mortars[21]. In four graves, the pit was lined with greens and flowers prior to inhumation, as reflected by dozens of plant impressions in each[22]. Three skeletons in the series have direct dates, and one also produced ancient DNA data. The dated skeletons are: Homo19 at 11840-11340 calBCE (RTK-6480 11725±125 bp), Homo28 at 10600-9760 calBCE (RTK-6638, 10320±115 bp), and Homo18 (Nat13) at 11520-11110 calBCE (RTK-6607, 11405±120 bp). We use the union of the 95.4% confidence intervals for the three direct dates (11840-9760 BCE), to represent the five individuals without direct dates:

• Nat4/Homo14 (I1685): young adult, <30 years old at death.
• Nat5/Homo13 (I1069): pre-adolescent, 9-12 years old at death.
• Nat6/Homo10 (I1690): pre-adolescent, 9-12 years old at death.
• Nat9/Homo16 (I1072): young child, 2-3 years old at death.
• Nat10/Homo17 (I0861): young adult, <30 years old at death.
• Nat13/Homo18 (I1687): 11520-11110 calBCE (RTK-6607, 11405±120 bp). >30 years old at death.

Motza (Israel)

The site of Motza Tachtit is situated west of the main entrance to Jerusalem. It was uncovered during salvage excavations in the year 2012, and lies 900 meters south-east of the distinct archaeological site of Tel Motza. The excavations revealed a burial of an adult male buried in a flexed position with a fox mandible placed intentionally under its head (E. Vered, personal information). The skeleton (Locus 713) is dated to the end of the PPNB period, 7300-6750 BCE, based on its Level II context where other burials were also found[23].

• Motz1 (I0867): adult male (confirmed genetically).

'Ain Ghazal (Jordan)

'Ain Ghazal is one of the largest (approximately 14 hectare) Neolithic sites known in the Near East, and was continuously occupied from around 8300-5000 BCE, followed by sporadic occupation in the Chalcolithic and Byzantine periods[24-26]. The petrous bone samples of a total of 15 individuals gave ancient DNA data. Three of the early Middle PPNB skulls at 'Ain Ghazal were found in a cache with a range of ages (11, 20s, and >60), and all were at one time plastered. Another early Middle PPNB sample that provided DNA included painted and burnt cranial fragments, and was probably also plastered. Two of the Early Bronze Age samples analyzed in this study were found in a cave high

above the East Field of 'Ain Ghazal and are probably associated with an Early Bronze Age village about 450 meters south of the Neolithic town.

- AG_83_3082 (I1727): early MPPNB.
- AG83/1 (I1416): early MPPNB.
- AG84/1 (I1414): early MPPNB.
- AG83_3 (I1701): early MPPNB.
- AG84_8 (I1709): early MPPNB.
- AG88_1 (I1700): early MPPNB.
- AG83_6 (I1710): middle MPPNB.
- AG84/2 (I1415): late MPPNB.
- AG83_5 (I1707): late MPPNB.
- AG89_1 (I1704): early LPPNB.
- AG037C (I1679): early PPNC.
- AG84_5 (I1699): middle PPNC.
- AG98_1 (I1705): Early Bronze Age.
- AG98_2 (I1706): Early Bronze Age.
- AG_84_3083_116 (I1730): 2489-2299 calBCE (OxA-32775, 3925±31 bp). Early Bronze Age. AG_84_3083_116 is thought to be PPNB based on archaeological context, but the date and genetic clustering indicates unambiguously that the individual is from the Early Bronze Age and hence this is the label we use. The discrepancy with the archaeological assignment could be due either to an intrusive burial or to sample mislabeling.

For the individuals for whom we do not have direct radiocarbon dates, we give dates based on stratigraphic layers. The ages of these layers were determined based on dates of non-human archaeological samples (fauna, charcoal, seeds, etc.), results in the following calibrated date ranges:

8300-7900 BCE: early MPPNB (early Middle PPNB)
7900-7700 BCE: middle MPPNB (middle Middle PPNB)
7700-7500 BCE: late MPPNB (late Middle PPNB)
7500-7300 BCE: early LPPNB (early Late PPNB)
6900-6800 BCE: early PPNC
6800-6700 BCE: middle PPNC
2490-2300 BCE: Early Bronze Age (range based on dated individual AG_84_3083_116)

Anatolia

Barcın Höyük (Turkey)

Barcın Höyük is located in the Yenişehir Plain in Northwest Anatolia (Turkey). Excavations have demonstrated continuous occupation from around 6600-6000 BCE, producing about 4.5 meters of stratified Neolithic settlement deposits[27]. Late Chalcolithic deposits have been excavated at the centre of the mound, and these revealed fragmentary remains of two mud-brick houses, several large ovens, and a ditch that surrounded (part of) the settlement.[28] Two infant jar burials were found among the settlement remains. The burial of an adult male in flexed position in a simple pit is also thought to belong to the Late Chalcolithic[29]. The sample analyzed here is obtained from a petrous bone of a 12±4 month old infant[30] that was excavated in 2009 in Late Chalcolithic deposits of Trench M10, where it had been buried in a tall Hole Mouth jar. Although they were crushed, the cranial parts as well as the postcranial bones were almost complete. No pathology was observed.

• M10-111 (I1584): 3943-3708 calBCE (OxA-32776, 5016±31 bp). Chalcolithic.

Armenia

Areni-1 (Vayots-Dzor Province, Republic of Armenia)

Areni-1 (also known as Birds' Cave) is a three-chambered cave located on the left-hand side of the Arpa River basin, a tributary of the River Araxes, within the eastern portion of the modern village of Areni in the Vayots Dzor Province of southern Armenia. Excavations at the site began in 2007 and were directed by Boris Gasparyan (Institute of Archaeology and Ethnography, National Academy of Sciences, Armenia) and co-directed by Ron Pinhasi (School of Archaeology, University College Dublin, Ireland) and Gregory Areshian (Cotsen Institute of Archaeology at UCLA, USA). The significance of the site became clear during the initial excavations when very well preserved Chalcolithic (4300–3400 BCE) and Medieval (4th–18th centuries CE) occupations were exposed[31,32]. Owing to the cave's dry condition, the perishable organic materials are exceptionally well preserved. As a result, stratigraphic observations and a set of secure radiocarbon dates allow the reconstruction of the Late Chalcolithic sequence of Armenia (and the broader region) between 4300-3400 BCE with unparalleled precision. During this period, the cave was used for habitation, for keeping animals and storing plants, for the production of wine, and for ritual purposes. The data from the cave demonstrates evidence for early social complexity, and connections to contemporary Near Eastern and North Caucasian societies. Areni-1 yielded the world's earliest evidence of footwear[33], wine making[34], and a wealth of exceptionally well preserved organic material including seeds, wooden artefacts, reed matts, baskets, textiles and desiccated fruits. It also yielded secondary burials of subadult crania, each deposited in a clay pot[31,32]. The petrous bones of three subadult crania (jar burials) from the Early Late Chalcolithic (Horizon III) gave ancient DNA and are directly dated. In

addition, two remains from the Middle Late Chalcolithic (Horizon II) gave DNA (the only samples in the study not from petrous bones). For these two, we use a date of 4350-3700 BCE based on context.

- AR1/43c (I1631): 4250-4050 calBCE (OxA-19332, 5323±30 bp). Early Late Chalcolithic (Horizon III), Burial 1, age 8±2 years.
- AR1/44 (I1634): 4330-4060 calBCE (OxA-19331, 5366±31 bp). Early Late Chalcolithic (Horizon III), Burial 2, age 11±2.5 years.
- AR1/46 (I1632): 4230-4000 calBCE (OxA-18599, 5285±29 bp). Early Late Chalcolithic (Horizon III), Burial 3, age 15±2.5 years.
- ARE20 (I1409): Middle Late Chalcolithic (Horizon II), Trench 1, Unit 4, Square E19, Locus 81, Spit 6 - tooth sampled.
- ARE12 (I1407): Middle Late Chalcolithic (Horizon II), Trench 2A, Unit 7, Square S33/T33, Locus 9, Spit 23 - tooth sampled.

Talin necropolis (Aragatsotn Province, Republic of Armenia)

The necropolis is located at the limits of the city of Talin, and is distributed on both sides of the Talin-Gyumri Higway. The Early Iron Age remains are found in the northwestern part of the necropolis (north-western limits of Talin), in a cemetery occupying about 3 kilometers squared^2. The Early Bronze Age and Late Bronze Age cemeteries occupy around one kilometer squared^2. Systematic archaeological excavations at the site have been conducted since 1984, and over one hundred tombs dating from the last quarter of the 4th millennium BC through Hellenistic period have been excavated. The Early Bronze Age is represented by a ritualistic enclosure and four tombs. These are dated to the first phase of the Kura-Araxes culture, which overspread the region in the second half of the fourth millennium BCE to the early part of the third millennium BCE. The tombs are earth and stone tumuli, 0.4-0.6 meters high, but differ in their construction, with some having been built within pits, and others at ground level. Burial 115 was excavated as part of a group of 12 tombs in 2014, during rescue archaeology prior to road construction during the North-South Corridor Highway project [35,36].

- TA3/R8 (I1658): 3347-3092 calBCE (OxA-31874, 4492±29 bp). Early Bronze Age I, Burial 115, petrous bone from skull N1.

Kalavan-1 burial ground (Gegharkunik Province, Republic of Armenia)

Kalavan-1 is an open-air site 1,640 meters above sea level on the southwest slopes of the Aregunyats Range north of Lake Sevan, Northeast Armenia. Archaeological and geological investigations were conducted here between 2005 and 2009 as part of a collaborative Armenian and French project. The excavation revealed two main levels of occupation dated to the Terminal Palaeolithic, overlain by an

Early Bronze Age Kura-Araxes burial ground. The total excavated area approaches 70 meters squared. Five burial pits were uncovered, of which four, referred to as UF1, UF2, UF8 and UF9, contained single primary burials, while the fifth (UF5) is a multiple burial that held the remains of at least three individuals. Six consistent radiocarbon dates on human skeletal material from UF5, UF8 and UF9 span 2900-2400 BCE, during the later part of the Kura-Araxes cultural horizon, and this is the range we use for the undated sample. Stone heaps rising to approximately 0.7m in height marked the graves of the adults. These structures were oval-shaped with a major axis of 1 meter, reaching 1.7 meters above the multiple burial. The position of the body in the pits varied: sitting, tightly flexed, and flexed. Post-sepulchral recovery of skulls and long bones occurred. The adult burials were furnished with the same assemblage of black burnished pottery that has the strongest association with the Kura basin ceramics and UF9 also contained bronze ornaments: a ring and a bracelet found near the skull. The child burial was in flexed position on its right side and was adorned with a neck ornament composed of dog molars and two stone beads, one of which was made of carnelian[37,38]. The two human remains (petrous bones) used in ancient DNA analyses came from the Early Bronze Age III period burials UF1 and UF9:

• KA1/14 (I1633): 2619-2410 BCE (Poz-22234, 3990±35 bp). Early Bronze Age III, Burial UF9.
• KA1/12 (I1635): Early Bronze Age III, Burial UF1.

Katnaghbyur necropolis (Aragatsotn Province, Republic of Armenia)

This site is located on the outskirts of Katnaghbyur village, 1-2 kilometers from the Yerevan-Talin Highway. Most of the archaeological materials are from rescue excavations of 10 tombs (1952, 1987-1988) that were typologically dated to the Early Iron Age. Tombs 2 and 3, as well as tomb 2 of the 1952 excavation, are dated to the 1100-800 BCE, while the other tombs are dated to the 900-700 BCE. The funerary chambers are cists with basalt slabs as walls, which are oriented west-east or north-south. Tomb 3 does not have a defined contour and resembles a pit grave in which the burial was encircled by slabs. Most of the tombs contained individual inhumations, either secondary (tombs 2 and 7) or primary, with the skeleton in a flexed position lying on its side (tomb 3). Tomb 8 is a double inhumation of a man and a woman. The sample we analyzed is Burial 1 or Kurgan N1 dating to the Middle Bronze Age / Late Bronze Age. It was excavated in 2014 as part of rescue archaeology efforts during the North-South Corridor Highway project[39].

• KAT16 (I1656): 1501-1402 calBCE (OxA-31674, 3168±27 bp). Burial N1.

References

1 Reimer, P. J. *et al.* Intcal13 and Marine13 Radiocarbon Age Calibration Curves 0-50,000 Years cal BP. *Radiocarbon* **55**, 1869-1887 (2013).

2 Ramsey, C. B. & Lee, S. Recent and planned developments of the program OxCal. *Radiocarbon* **55**, 720-730 (2013).
3 Coon, C. S. Excavations at Hotu Cave, Iran, A Preliminary Report. *Proceedings of the American Philosophical Society* **96**, 231-249 (1952).
4 Coon, C. S. in *Catalogue des Hommes Fossiles V.* (eds H.V. Valois & H.L. Movius) 325-328 (1953).
5 Coon, C. S. *The Seven Caves*. (Alfred A. Knopf. New York, 1957).
6 Ralph, E. K. University of Pennsylvania Radiocarbon Dates I. *Science* **121**, 149-151 (1957).
7 Analytic, B. Report on Radiocarbon Analyses, Beta 344447, Hotu sample 532284 (2013).
8 Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun* **6** (2015).
9 Young Jr, T. C. & Smith, P. E. L. Research in the Prehistory of Central Western Iran. *Science* **153**, 386-391 (1966).
10 Hesse, B. C. Slaughter patterns and domestication: the beginnings of pastoralism in Western Iran. *Man ns* **17**, 403-417 (1982).
11 Zeder, M. A. & Hesse, B. C. The initial domestication of goats (Capra hircus) in the Zagros Mountains 10,000 years ago. *Science* **287**, 2254-2257 (2000).
12 Merrett, D. C. *Bioarchaeology in Early Neolithic Iran: assessment of health status and subsistence strategy* (Unpublished Ph.D. thesis, University of Manitoba 2004).
13 Young, T. C. J. & Levine, L. D. R. *Excavations of the Godin Project: second progress report.*, (1974).
14 Henrickson, E. *Ceramic Styles and Cultural Interaction in the Early and Middle Chalcolithic of the Central Zagros, Iran* PhD. thesis, (1983).
15 Voigt, M. M. & Dyson, R. H. T. in *Chronologies in Old World Archaeology, 3rd edition, vols.1+2* Vol. 2 (ed R.W. Ehrich) 122-178 (Chicago University Press, Chicago, 1992).
16 Lengyel, G., Bocquentin, F. & Nadel, D. in *Natufian Foragers in the Levant: Terminal Pleistocene Social Changes in Western Asia.* (eds O. Bar-Yosef & F. Valla) 478–504 (2013).
17 Nadel, D. & Lengyel, G. Human-made Bedrock Holes (mortars and cupmarks) as a Late Natufian social phenomenon. *Archaeology, Anthropology and Ethnology of Eurasia* **37**, 37–48 (2009).
18 Nadel, D., Filin, S., Rosenberg, D. & Miller, V. Prehistoric bedrock features: recent advances in 3D characterization and geometrical analyses. *Journal of Archaeological Science* **53**, 331-344 (2015).
19 Nadel, D. & Rosenberg, D. An incised pattern inside a Raqefet Cave bedrock mortar: New dimensions to Natufian stone carvings. *Journal of Lithic Studies* **In press** (2016).
20 Yeshurun, R., Bar-Oz, G. & Nadel, D. The social role of food in the Natufian cemetery of Raqefet Cave, Mount Carmel, Israel. *Journal of Anthropological Archaeology* **32**, 511-526 (2013).
21 Power, R. C., Rosen, A. M. & Nadel, D. The economic and ritual utilization of plants at the Raqefet Cave Natufian site: The evidence from phytoliths. *Journal of Anthropological Archaeology* **33**, 49-65 (2014).
22 Nadel, D. *et al.* Earliest floral grave lining from 13,700–11,700-y-old Natufian burials at Raqefet Cave, Mt. Carmel, Israel. *Proceedings of the National Academy of Sciences* **110**, 11774-11778 (2013).

23 Mizrahi, S. Moza Tahtit. *Hadashot Arkheologiyot* **127**, http://www.hadashot-esi.org.il/report_detail.aspx?id=24845&mag_id=24122 (2015).

24 Rollefson, G. & Kafafi, Z. in *Symbols at 'Ain Ghazal* (ed D. Schmandt-Besserat) 3-29 (Berlin: ex oriente, 2013).

25 Bonogofsky, M. *n Osteo-Archaeological Examination of the Ancestor Cult during the Pre-Pottery Neolithic B Period in the Levant* phD thesis, (2001).

26 Grindell, B. *Unmasked Equalities: An Examination of Mortuary Practices and Social Complexity in the Levantine Natufian and Pre-Pottery Neolithic*, (1998).

27 Gerritsen, F., Özbal, R. & Thissen, L. in *The Neolithic in Turkey. New Excavations and New Research. Vol. 5 Northwestern Turkey and Istanbul* (eds M. Özdoğan, N. Başgelen, & P. Kuniholm) 93-112 (Archaeology and Art Publications, 2013).

28 Gerritsen, F., Ozbal, R., Thissen, L., Ozbal, H. & Galik, A.-. The late Chalcolithic Settlements of barcin Hoyuk. Anatolica 36, p. 197-225. . *Anatolica* **36**, 197-225 (2010).

29 Roodenberg, J. J., As, A. V. & Alpaslan-Roodenberg, S. Barcin Hüyük in the Plain of Yenişehir (2005-2006). *Anatolica* **34**, 53-66 (2008).

30 Ubelaker, D. H. *Human skeletal remains: excavation, analysis, interpretation*. (Taraxacum Press, 1989).

31 Areshian, G. E. *et al.* The chalcolithic of the Near East and south-eastern Europe: discoveries and new perspectives from the cave complex Areni-1, Armenia. *Antiquity* **86**, 115-130 (2012).

32 Wilkinson, K. *et al.* The origins of the Bronze Age in the Caucasus: new discoveries at Birds' Cave, Vayots Dzor, Armenia. *Journal of Field Archaeology* **37**, 20-33 (2012).

33 Pinhasi, R. *et al.* First Direct Evidence of Chalcolithic Footwear from the Near Eastern Highlands. *PLoS ONE* **5**, e10984 (2010).

34 Barnard, H., Dooley, A. N., Areshian, G., Gasparyan, B. & Faull, K. F. Chemical evidence for wine production around 4000 BCE in the Late Chalcolithic Near Eastern highlands. *Journal of Archaeological Science* **38**, 977-984 (2011).

35 Kalantaryan, I. The Early Bronze Age complexes of Talin cemetery. , *Studii de Preistorie* **8**, 123-138 (2011).

36 Badalyan, R. S. & Avetisyan, P. S. in *Bronze and Iron Age Archaeological Sites in Armenia, I, Mt. Aragats and its Surrounding Region* (eds R. S. Badalyan & P.S. Avetisyan) 242-263 (BAR International Series 1697,, 2007).

37 Montoya, C. *et al.* The Upper Palaeolithic site of Kalavan 1 (Armenia): An Epigravettian settlement in the Lesser Caucasus. *Journal of Human Evolution* **65**, 621-640 (2013).

38 Chataigner, C. *et al.* in *Archaeology of Armenia in regional context* (eds P. Avetisyan & A. Bobokhyan) 37-43, 157=162 (: Gitutyun publishing house, 2012).

39 Badalyan, R. S. & Avetisyan, P. S. in *Bronze and Iron Age Archaeological Sites in Armenia, I, Mt. Aragats and its Surrounding Region* (eds R. S. Badalyan & P.S. Avetisyan) 152-155 (BAR International Series 1697,, 2007).

Supplementary Information 2

Human Origins dataset

For this study we genotyped 238 present-day individuals from West Eurasia on the Affymetrix Human Origins array[1], which we added to the 2,345 individuals of the same dataset previously analyzed by Lazaridis et al.[2]. A total of 2,918 individuals on the Human Origins array have been published to date, including 1,025 by Patterson et al.[1], 209 by Pickrell et al.[3], 1,458 by Lazaridis et al.[2], 178 by Qin and Stoneking[4], and 48 by Skoglund et al.[5] Of the total 621,799 SNPs, we analyze a subset of 592,146 autosomal SNPs in the *HO* dataset which includes autosomal SNPs used in the 1240k capture[6], intersected with the analysis set of 594,924 SNPs of ref.2.

The genotyped individuals are listed in Supplementary Data Table 2. Since our research permits for the collection of the samples analyzed in this dataset only cover studies of population prehistory, we have to stipulate some conditions before granting access to the data noted, as follows:

To access the dataset that includes individuals noted as "Letter" in the Access column of Supplementary Data Table 2, you need to send David Reich (reich@genetics.med.harvard.edu) a PDF of a signed letter containing the following language:

(a) I will not distribute the data outside my collaboration,
(b) I will not post the data publicly,
(c) I will make no attempt to connect the genetic data to personal identifiers for the samples,
(d) I will use the data only for studies of population history,
(e) I will not use the data for commercial purposes

References

1. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093, (2012).
2. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, (2014).
3. Pickrell, J. K. *et al.* The genetic prehistory of southern Africa. *Nat. Commun.* **3**, 1143, (2012).
4. Qin, P. & Stoneking, M. Denisovan Ancestry in East Eurasian and Native American Populations. *Molecular Biology and Evolution*, (2015).
5. Skoglund, P. *et al.* Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104-108, (2015).
6. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503, (2015).

Supplementary Information 3

Population groupings in the ancient Near East

We first assigned "detailed" population labels on the basis of provided information about the provenance, date, and cultural affiliation of sampled individuals and identified visible outliers in PCA and ADMIXTURE analysis (Fig. 1b, Extended Data Fig. 2a). Individuals from different population labels sometimes did not appear to be genetically distinct, so we sought to identify broader population groupings that would be used for analysis. Doing so is useful in order to reduce the number of distinct population names when there is no significant genetic differentiation between them, and also to enhance statistical power by including more individuals in each population used for analysis. Besides these technical benefits of lumping individuals in broader categories, doing so is also useful in determining whether groupings recognized by archaeologists correspond (or not) to genetically differentiated populations.

To arrive at the "analysis" population labels, we show in Table S3.1 f_4-statistics of the form f_4(Population$_1$, Population$_2$; *Test*, Chimp). If every *Test* population (including present-day and ancient populations in our dataset) is symmetrically related to the (Population$_1$, Population$_2$) pair, then we may join these in a single population. We generally use a $|Z|\leq 3$ threshold; this may be conservative, given the number of f_4-statistics considered, but we aimed to err on the side of caution and only group together populations that appear to form a strong clade to the limits of our resolution.

All pairwise comparisons in Iran suggest that none of the distinct groups can be safely joined into a broader category. The only exception is the pair (Iran_HotuIIIb, Iran_Late_Neolithic) with $|Z|\leq 2.8$. However, the low significance here may be due to the small number of shared SNPs (2,096 for the most significant Z=2.8 statistic) in these two singleton individuals. Ganj_Dareh_Iran_Neolithic does not form a clade with Iran_HotuIIIb (which is putatively Mesolithic, but whose inferred date is uncertain (Supplementary Information, section 1), and may either predate or postdate the Ganj_Dareh_Iran_Neolithic.), or with the succeeding Iran_Late_Neolithic. Ganj_Dareh_Iran_Neolithic shares fewer alleles with a Native American group (Ticuna) than does Iran_HotuIIIb, and it shares fewer alleles with the Anatolian Neolithic than does Iran_Late_Neolithic. Thus, we do not consider the non-significant statistic for the (Iran_HotuIIIb, Iran_Late_Neolithic) pair convincing, and with the additional consideration of the uncertainty regarding the age of Iran_HotuIIIb, we analyze them separately.

In Armenia, the Chalcolithic group differs from the Early Bronze Age group ($Z=4.1$) and so is studied separately. The Middle/Late Bronze Age groups appear to form a clade ($|Z|\leq 2$) and so are grouped into the Armenia_MLBA group. The Early Bronze Age population is analyzed separately as Armenia_EBA as it does not form a clade with the Middle Bronze Age ($|Z|=3$) and also forms a tight and distinctive cluster in PCA (Fig. 1b), shifted away from other ancient samples from Armenia in the direction of ancient Iran.

In the Levant, Natufians differ from all other later populations ($|Z|\geq 4.1$) and so are studied separately. The two Neolithic groups (PPNB, PPNC) do not differ significantly from each other ($|Z|\leq 1.7$), and so are grouped into a Levant_N population. The Bronze Age samples from Jordan differ significantly from the PPNB ($|Z|=5.0$) and form a tight cluster in PCA (Fig. 1b) and are thus studied separately as a Levant_BA population.

The correspondence between detailed and inclusive population labels is shown in Supplementary Data Table 1. With additional sampling, our power to distinguish between geographically or temporally distinct populations from the three Near Eastern regions may improve. We followed previous studies[1-3] for the groupings of non-Near Eastern populations. Some of these groupings (such as Europe_LNBA) are not genetically homogeneous (Fig. 1b) but represent several individual populations with varying ancestry (e.g., the Corded Ware group has more ancestry from the steppe than other Late Neolithic/Bronze Age Europeans[2]). We chose a coarser granularity for such groups as their relationships have been treated elsewhere and in the present paper we wish to focus the broader structure of West Eurasian populations, while exploring in greater detail the ancient Near Eastern populations described here for the first time.

Table S3.1: Statistics of the form f_4(Population$_1$, Population$_2$; *Test*, Chimp) that we use to motivate population groupings in the ancient Near East. The *Test* population producing the highest/lowest Z-score of this statistic is shown in this table.

Population1	Population2	Test	Z	# SNPs
Ganj_Dareh_Iran_Neolithic	Ganj_Dareh_Iran_recent	Sardinian	-10.0	17729
Ganj_Dareh_Iran_Neolithic	Ganj_Dareh_Iran_recent	Himba	-0.5	15142
Ganj_Dareh_Iran_Neolithic	Iran_Chalcolithic	Anatolia_N	-12.8	20581
Ganj_Dareh_Iran_Neolithic	Iran_Chalcolithic	Papuan	-2.0	20682
Ganj_Dareh_Iran_Neolithic	Iran_Late_Neolithic	Anatolia_N	-4.3	11078
Ganj_Dareh_Iran_Neolithic	Iran_Late_Neolithic	Wayuu	-0.5	11218
Ganj_Dareh_Iran_Neolithic	Iran_HotuIIIb	Ticuna	-3.4	3298
Ganj_Dareh_Iran_Neolithic	Iran_HotuIIIb	Anatolia_ChL	0.1	2471
Ganj_Dareh_Iran_recent	Iran_Chalcolithic	Dinka	-2.3	17015
Ganj_Dareh_Iran_recent	Iran_Chalcolithic	SHG	3.3	20400
Ganj_Dareh_Iran_recent	Iran_Late_Neolithic	Himba	-1.6	9243
Ganj_Dareh_Iran_recent	Iran_Late_Neolithic	Europe_MNChL	5.0	11896
Ganj_Dareh_Iran_recent	Iran_HotuIIIb	Himba	-1.2	2763
Ganj_Dareh_Iran_recent	Iran_HotuIIIb	Europe_EN	3.9	3564
Iran_Chalcolithic	Iran_Late_Neolithic	Himba	-1.3	10405
Iran_Chalcolithic	Iran_Late_Neolithic	Europe_EN	5.5	13253
Iran_Chalcolithic	Iran_HotuIIIb	Steppe_Eneolithic	-0.9	2907
Iran_Chalcolithic	Iran_HotuIIIb	Anatolia_N	5.5	3950
Iran_Late_Neolithic	Iran_HotuIIIb	Pima	-0.6	2163
Iran_Late_Neolithic	Iran_HotuIIIb	Levant_N	2.8	2096
Armenia_Chalcolithic	Armenia_EBA	Iran_N_outlier	-1.7	20000
Armenia_Chalcolithic	Armenia_EBA	EHG	4.8	25055
Armenia_Chalcolithic	Armenia_LBA	Tswana	-0.5	11035
Armenia_Chalcolithic	Armenia_LBA	Europe_EN	4.3	14149
Armenia_Chalcolithic	Armenia_MBA	Oromo	-0.9	20939
Armenia_Chalcolithic	Armenia_MBA	Anatolia_N	3.2	23875
Armenia_EBA	Armenia_LBA	Wayuu	-1.4	13209
Armenia_EBA	Armenia_LBA	Iran_N_outlier	2.8	11337
Armenia_EBA	Armenia_MBA	Cree	-3.0	22582
Armenia_EBA	Armenia_MBA	Anatolia_ChL	2.4	14061
Armenia_LBA	Armenia_MBA	Levant_N	-2.0	10522
Armenia_LBA	Armenia_MBA	Iran_HotuIIIb	1.0	2050
Israel_Natufian	Jordan_EBA	Kalash	-7.7	11373
Israel_Natufian	Jordan_EBA	Oromo	0.5	10833
Israel_Natufian	PPNB	Europe_MNChL	-5.8	10658
Israel_Natufian	PPNB	Oromo	0.9	9880
Israel_Natufian	PPNC	Abkhasian	-4.2	4462
Israel_Natufian	PPNC	Oromo	0.8	4191
Jordan_EBA	PPNB	Anatolia_N	-2.6	19127
Jordan_EBA	PPNB	Iran_N	5.0	17320
Jordan_EBA	PPNC	Anatolia_N	-3.0	7640
Jordan_EBA	PPNC	GujaratiD	2.0	7640
PPNB	PPNC	Wayuu	-1.1	6795
PPNB	PPNC	GujaratiD	1.7	6942

References

1. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, (2014).
2. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211, (2015).
3. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503, (2015).

Supplementary Information 4

Pervasive Basal Eurasian ancestry in the ancient Near East

It has been proposed[1] that early European farmers had Basal Eurasian ancestry based on the observation that the Stuttgart[1] early farmer and Eurasian hunter-gatherers such as Loschbour[1], LaBrana[2], MA1[3], and Motala12[1] had significantly positive f_4-statistics of the form f_4(*Eurasian hunter-gatherer*, Stuttgart; *eastern non-African*, Chimp), for diverse eastern non-African groups from East Asia, the Andaman Islands, Papua New Guinea, North Asia and the Americas. A parsimonious explanation for these statistics that does not involve gene flow that affected all eastern non-African groups from west Eurasia or vice versa, is that the early farmers of Europe possessed ancestry from a deeply diverged lineage that split off from other Eurasians before the split of eastern non-Africans from west Eurasian hunter-gatherers.

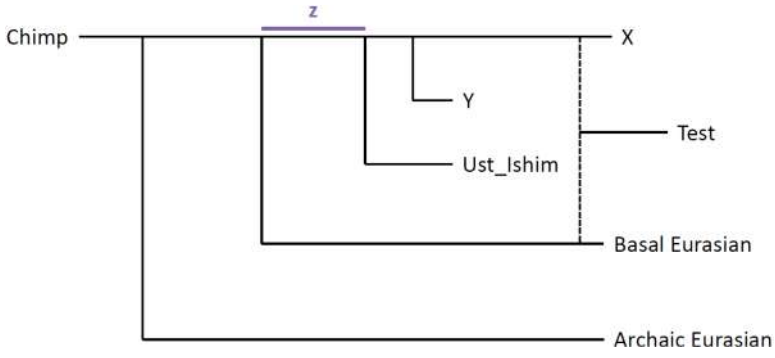
Subsequent research on the Ust_Ishim[4] and Oase1[5] individuals has revealed that these two individuals from Upper Paleolithic Siberia and Europe respectively share more alleles with present-day eastern non-Africans than with Europeans, but are symmetrically related to eastern non-Africans and ancient European hunter-gatherers. This finding has been interpreted as supporting the idea that recent Europeans have ancestry that was not present in ancient hunter-gatherers from Eurasia and which—because of its earlier split—is diluting the affinity of these Upper Paleolithic Eurasians to present-day Europeans. The genome of Kostenki14[6] was interpreted as having the same kind of Basal Eurasian ancestry as early European farmers, based on the observation that east Asians do *not* share more alleles with it than they do with early farmers[6]. However, this observation is also consistent with Kostenki14 having a different type of ancestry than the early farmers, or alternatively later gene flow between ancestors of present-day eastern non-Africans and Eurasian hunter-gatherers (see ref.7).

The prediction that the Basal Eurasian ancestry in present-day Europeans came from the Near East via early farmers[1] was not based on ancient genomes from the Near East (which were not available at the time), but rather on the observation that diverse Eurasian hunter-gatherers from Europe and Siberia do not differ significantly in their shared genetic drift with eastern non-Africans but systematically shared more than the early farmers did. As the ancient farmers could be modeled as a mixture of European hunter-gatherers and a Near Eastern source population[1], it followed that their Basal Eurasian ancestry was derived from the Near Eastern portion of their ancestry.

Basal Eurasian ancestry was prevalent in the ancient Near East

The current study confirms the presence of Basal Eurasian ancestry in the ancient Near East by the statistic f_4(*Ancient West Eurasian* Test, *Eastern non-African* or *European*/*Siberian hunter-gatherer* Y; Ust_Ishim, Chimp). This statistic tests whether the topology ((Test,Y), Ust_Ishim, Chimp) is valid (in which case it is not significantly different from zero), or the Test population has Basal Eurasian ancestry (in which case it has a negative value) (Fig. S4.1). It is also robust to both later gene flow within Eurasia as well as archaic admixture in an eastern non-African population such as Papuans who have more archaic admixture than other Eurasians[8] (Fig. S4.1).

Figure S4.1: Testing for Basal Eurasian ancestry in a population *X* with the statistic f_4(*Ancient Eurasian* Test, *Eastern non-African* or *European*/*Siberian hunter-gatherer* Y; Ust_Ishim, Chimp). If population Test traces its ancestry in part from a Basal Eurasian population, then the drift path *z* is traversed in the opposite direction in the paths Basal_Eurasian→Y and Ust_Ishim →Chimp. Gene flow between *X* and *Y* (i.e., not a clean split as shown here) does not affect the value of the statistic either positively or negatively. Archaic admixture in population *Y* affects the statistic in a *positive* direction as the drift paths X/Basal Eurasian→Archaic Eurasian and Ust_Ishim →Chimp are traversed in the same direction; thus, this will not result in a spurious negative f_4-statistic.



The Z-score of this statistic is shown in Table S4.1 for diverse ancient Eurasian populations on the *HO* dataset.

Table S4.1: Z-score of the statistic f_4(*Ancient Eurasian* Test, *Eastern non-African* or *European*/*Siberian hunter-gatherer* Y; Ust_Ishim, Chimp) on the *HO* dataset.

	Han	Onge	Papuan	Kostenki14	MA1	WHG	EHG	SHG	Switzerland_HG
Anatolia_ChL	-4.5	-4.2	-0.3	-2.1	-2.8	-3.6	-1.6	-2.9	-2.9
Anatolia_N	-6.8	-6.4	-0.2	-2.3	-3.6	-5.3	-2.7	-5.4	-3.5
Armenia_ChL	-4.7	-4.7	0.8	-1.3	-2.3	-3.4	-1.1	-3.2	-2.6
Armenia_EBA	-6.4	-6.0	-0.3	-2.3	-3.3	-4.6	-2.8	-4.6	-3.3
Armenia_MLBA	-5.9	-5.6	-0.2	-2.3	-3.0	-4.5	-1.8	-4.0	-3.5
CHG	-4.6	-4.5	0.4	-1.5	-2.9	-3.1	-1.4	-3.0	-2.4
EHG	-3.2	-3.2	1.6	-0.6	-1.4	-1.5		-1.5	-1.2
Europe_EN	-6.2	-5.8	0.3	-1.9	-3.3	-4.7	-1.9	-4.6	-3.0
Europe_LNBA	-5.1	-4.7	1.4	-0.9	-2.3	-3.2	-0.7	-3.0	-1.9
Europe_MNChL	-5.1	-5.0	0.9	-1.3	-2.6	-3.8	-1.2	-3.6	-2.3
Iberia_BA	-4.4	-4.1	0.0	-1.7	-2.6	-3.6	-1.5	-3.7	-2.7
Iran_ChL	-6.2	-6.1	-0.4	-2.3	-3.6	-4.9	-2.5	-4.7	-3.6
Iran_LN	-3.7	-3.7	0.6	-0.8	-2.1	-2.5	-1.0	-2.1	-2.2
Iran_HotuIIIb	-3.7	-4.1	-0.4	-1.6	-2.2	-3.3	-1.6	-2.4	-3.4
Iran_N	-7.1	-6.9	-1.7	-3.7	-4.7	-5.6	-3.6	-5.6	-4.4
Iran_recent	-6.0	-5.7	-0.8	-2.7	-3.4	-4.5	-3.2	-4.3	-3.8
Kostenki14	-2.1	-2.1	1.8		-1.0	-0.9	0.6	-0.5	-0.5
Levant_BA	-8.2	-8.0	-1.8	-4.0	-4.6	-7.3	-4.0	-6.7	-5.4
Levant_N	-8.8	-8.2	-2.9	-5.0	-5.6	-7.7	-5.5	-7.4	-5.7
MA1	-0.8	-0.9	2.8	1.0		0.2	1.4	0.7	0.2
Natufian	-7.0	-6.9	-2.1	-4.2	-4.3	-6.3	-3.7	-6.0	-4.8
SHG	-2.1	-2.3	2.6	0.5	-0.7	-0.7	1.5		-0.1
Steppe_EMBA	-4.8	-4.5	1.5	-0.9	-2.1	-2.7	-0.3	-2.5	-1.7
Steppe_Eneolithic	-2.5	-2.3	2.0	-0.1	-1.0	-1.0	0.8	-0.8	-0.5
Steppe_IA	-2.4	-2.1	1.9	-0.1	-1.0	-1.0	0.8	-0.5	-0.4
Steppe_MLBA	-4.6	-4.4	1.5	-0.8	-2.0	-2.9	-0.5	-2.7	-1.8
Switzerland_HG	-1.5	-1.7	2.3	0.5	-0.2	-0.3	1.2	0.1	
WHG	-1.6	-1.8	2.8	0.9	-0.2		1.5	0.7	0.3

Table S4.2: Z-score of the statistic f_4(*Ancient Eurasian* Test, *Eastern non-African* or *European*/*Siberian hunter-gatherer* Y; Ust_Ishim, Chimp) on the *HOIll* dataset.

	Kostenki14	MA1	WHG	EHG	SHG	Switzerland_HG
Anatolia_ChL	-2.6	-3.7	-4.0	-2.3	-3.8	-3.3
Anatolia_N	-2.6	-4.5	-5.5	-3.3	-6.1	-3.6
Armenia_ChL	-1.3	-3.1	-3.1	-1.5	-3.9	-2.2
Armenia_EBA	-2.1	-4.0	-4.0	-2.8	-4.9	-3.2
Armenia_MLBA	-2.6	-4.4	-4.9	-2.6	-5.0	-3.7
CHG	-2.5	-4.1	-4.2	-2.8	-4.8	-3.2
EHG	-0.5	-1.8	-1.3		-1.9	-1.1
Europe_EN	-2.1	-4.0	-4.9	-2.5	-5.6	-3.1
Europe_LNBA	-1.2	-3.1	-3.4	-1.4	-4.2	-1.9
Europe_MNChL	-1.7	-3.4	-4.2	-1.8	-4.7	-2.5
Iberia_BA	-1.4	-3.3	-3.2	-1.8	-3.5	-2.2
Iran_ChL	-2.6	-4.5	-4.9	-3.4	-5.5	-3.7
Iran_LN	-2.2	-3.6	-3.6	-2.4	-4.0	-3.3
Iran_HotuIIIb	-2.9	-2.8	-4.2	-3.3	-4.8	-4.3
Iran_N	-4.3	-6.0	-6.2	-4.7	-6.9	-5.3
Iran_recent	-2.8	-4.4	-4.7	-3.5	-4.8	-4.0
Kostenki14		-1.3	-0.8	0.5	-0.9	-0.4
Levant_BA	-3.9	-5.0	-6.8	-4.3	-7.2	-5.4
Levant_N	-5.8	-6.8	-8.4	-6.4	-8.7	-6.7
MA1	1.3		0.8	1.8	0.9	0.8
Natufian	-5.5	-6.1	-8.1	-6.0	-8.4	-6.6
SHG	0.9	-0.9	0.1	1.9		0.5
Steppe_EMBA	-0.8	-2.5	-2.4	-0.7	-3.2	-1.4
Steppe_Eneolithic	0.0	-1.8	-0.7	0.5	-1.2	-0.4
Steppe_IA	-0.1	-1.7	-0.9	0.5	-1.1	-0.4
Steppe_MLBA	-1.1	-2.9	-3.2	-1.2	-3.8	-1.8
Switzerland_HG	0.4	-0.8	-0.5	1.1	-0.5	
WHG	0.8	-0.8		1.3	-0.1	0.5

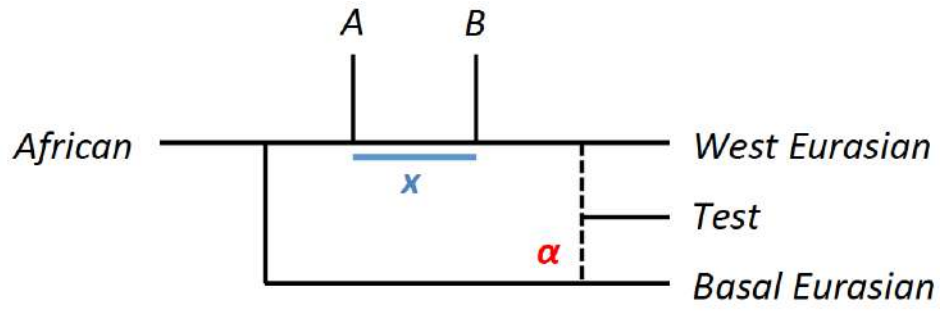
With the exception of the Y=Papuan column (where the statistic has a positive bias due to archaic admixture in this population; Fig. S4.1), the statistic is significantly negative (Z<-3) for all columns in four populations: Iran_N and all three populations from the Levant (Natufians, Levant_N, and Levant_BA). It is also significantly negative for all populations of partial Near Eastern ancestry when Y=Han, Onge, the present-day eastern non-African populations without substantial Denisovan admixture. To gain power, we repeated this analysis on the *HOIll* dataset, which has a larger number of SNPs, and show the results in Table S4.2. European hunter-gatherers (EHG, WHG, SHG, Kostenki14, Switzerland_HG show no evidence of Basal Eurasian ancestry, but populations of Near Eastern or partial Near Eastern ancestry do, reaching significance when Y=MA1, WHG, or SHG and being negative (although not always reaching significance at the |Z|>3 threshold) for the other hunter-gatherer groups.

We conclude that ancient Near Eastern populations and populations of partial Near Eastern ancestry have evidence of deriving part of their ancestry from a lineage that splits off prior to Ust_Ishim (and thus prior to ~45,000 years ago).

Estimating Basal Eurasian ancestry with a simple f_4-ratio

It is possible to estimate the proportion of ancestry derived from a Basal Eurasian population, assuming the topology of Fig. S4.2. The idea is to find a genetic drift path x equal to the statistic f_4(African, West Eurasian; A, B), where *West Eurasian* is a population without Basal Eurasian ancestry and A, B are two "ingroups" to the Basal Eurasian split node. The deepest split within the triple (A, B, West Eurasian) must occur after the split of the Basal Eurasians and *West Eurasian* must share a drift path length x with B that is not shared with A.

Figure S4.2: A topology for estimating the proportion α of Basal Eurasian ancestry. The ratio of the statistics f_4(Test, West Eurasian; A, B) / f_4(African, West Eurasian; A, B) = $\alpha x/x$ = α yields this proportion.



In Ref.1 we used A=Onge, B=MA1, West Eurasian=Loschbour (representative of WHG), and Test=Stuttgart (representative of Europe_EN).

We can now use A=Ust_Ishim instead of Onge as Ust_Ishim is symmetrically related to eastern non-Africans, MA1, and European hunter-gatherers (Table S4.1 and Table S4.2) and represents an even earlier split than that between Onge and European hunter-gatherers. Thus, we are estimating Basal Eurasian ancestry that stems from a clade not only prior to the split of European hunter-gatherers, MA1, and eastern non-Africans[1] but prior to the even earlier split of Ust_Ishim[4,7].

Using West_Eurasian=WHG as in Ref.1, we must ensure that the topology of Fig. S4.1 holds. We use MA1 and Kostenki14 (the two most ancient genomes sharing genetic drift with later West Eurasians) as candidates for B.

First, we examine the basic topology (without adding the Test population yet). In Table S4.3 we show f_4-statistics involving the set of four populations (Mbuti, A=Ust_Ishim, B=MA1 or Kostenki14, West Eurasian=WHG). These statistics show that (Kostenki14, WHG) and (MA1, WHG) form clades with respect to (Mbuti, Ust_Ishim) and that they mutually share more alleles with each other than with Ust_Ishim. Thus, the basic topology is validated.

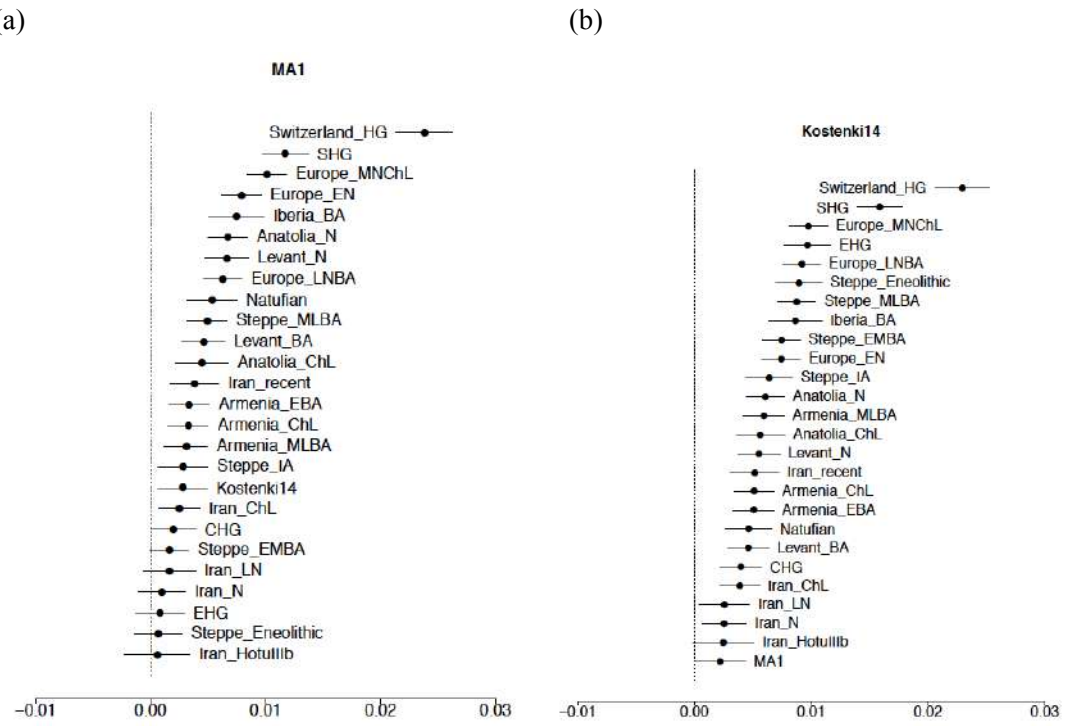
Table S4.3: Validating the topology of Fig. S4.2.

X	Y	Z	W	f_4(X,Y;W,Z)	Z
Mbuti	Ust_Ishim	Kostenki14	WHG	-0.00016	-0.3
Mbuti	WHG	Ust_Ishim	Kostenki14	0.00750	12.3
Mbuti	Kostenki14	Ust_Ishim	WHG	0.00734	11.4
Mbuti	Ust_Ishim	MA1	WHG	-0.00092	-1.7
Mbuti	WHG	Ust_Ishim	MA1	0.00777	13.0
Mbuti	MA1	Ust_Ishim	WHG	0.00686	10.5

Next, we examine whether the Test population shares more genetic drift with West_Eurasian=WHG than with B=MA1. In Fig. S4.3a we show the statistic f_4(WHG, MA1; Test, Ust_Ishim), showing that EHG and populations from early Iran are symmetrically related to (WHG, MA1), thus violating the topology of Fig. S4.2 in the sense that it is consistent with x=0. However, Kostenki14, which is the earliest known individual sharing West Eurasian-specific ancestry[6] does not violate the topology of Fig. S4.2 as the statistic f_4(WHG, Kostenki14; Test, Ust_Ishim) is significantly positive (Fig. S4.3b). Thus, we use B=Kostenki14. In the qpAdm section at the end of this note we will see additional evidence

that MA1 may not be a good outgroup for estimating Basal Eurasian ancestry and that Kostenki14 provides invaluable information to do so.

Figure S4.3: Statistic f_4(WHG, MA1 *or* Kostenki14; Test, Ust_Ishim). The value of the statistic ~3 standard errors is shown. (a) Populations from ancient Iran and EHG do not share significantly more alleles with WHG than with MA1 and are thus not consistent with the topology of Fig. S4.1. (b) All *Test* populations share significantly more alleles with WHG than with Kostenki14, and are thus consistent with the topology of Fig. S4.2.



Basal Eurasian ancestry can be inferred using the following f_4-ratio:

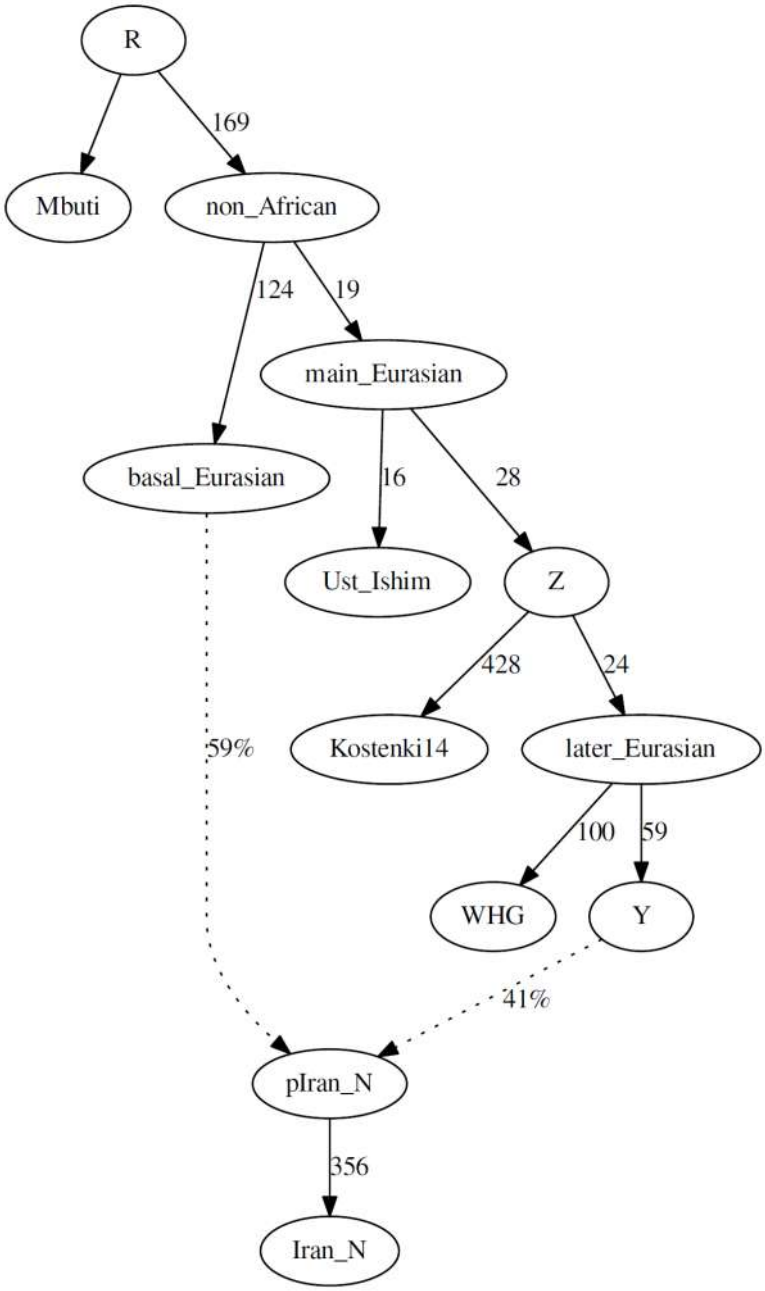
$$x = f_4(\text{Mbuti, WHG; Ust\_Ishim, Kostenki14})$$
$$\alpha x = f_4(\text{Test, WHG; Ust\_Ishim, Kostenki14})$$

The estimate of the mixture proportion α is shown in Table S4.4. We have also estimated this mixture proportion in a model-based way using the ADMIXTUREGRAPH software[9] which allows us to specify a phylogeny with added admixture edges and estimate its parameters (drift lengths and mixture proportions). We test the phylogeny of Fig. S4.2 and show a fitted model for the Iran_N population in Fig. S4.4.

Table S4.4: Mixture proportions estimated with the ratio f_4(Test, WHG; Ust_Ishim, Kostenki14)/f_4(Mbuti, WHG; Ust_Ishim, Kostenki14) and with ADMIXTUREGRAPH. The Z_{diff} column represents the number of standard errors of the f_4-ratio by which the ADMIXTUREGRAPH estimate differs.

Test	*f_4-ratio estimate*	*Standard Error*	*ADMIXTUREGRAPH estimate*	Z_{diff}
Anatolia_ChL	0.260	0.084	0.207	0.6
Anatolia_N	0.344	0.047	0.363	0.4
Armenia_ChL	0.365	0.054	0.366	0.0
Armenia_EBA	0.428	0.058	0.444	0.3
Armenia_MLBA	0.473	0.058	0.491	0.3
CHG	0.505	0.066	0.497	0.1
EHG	0.194	0.069	0.261	1.0
Europe_EN	0.323	0.047	0.342	0.4
Europe_LNBA	0.331	0.040	0.366	0.9
Europe_MNChL	0.293	0.043	0.310	0.4
Iberia_BA	0.277	0.087	0.428	1.7
Iran_ChL	0.446	0.056	0.464	0.3
Iran_LN	0.596	0.089	0.627	0.3
Iran_HotuIIIb	0.667	0.114	0.697	0.3
Iran_N	0.591	0.075	0.591	0.0
Iran_recent	0.394	0.082	0.394	0.0
Levant_BA	0.401	0.061	0.393	0.1
Levant_N	0.385	0.063	0.367	0.3
MA1	0.364	0.089	0.286	0.9
Natufian	0.460	0.082	0.438	0.3
SHG	0.170	0.050	0.200	0.6
Steppe_EMBA	0.367	0.044	0.402	0.8
Steppe_Eneolithic	0.314	0.066	0.302	0.2
Steppe_IA	0.280	0.076	0.269	0.1
Steppe_MLBA	0.327	0.043	0.377	1.2
Switzerland_HG	0.027	0.074	0.092	0.9

Figure S4.4: A fitted model for the topology of Fig. S4.3 for the *Test*=Iran_N population. The estimated mixture proportions for other *Test* populations are listed in Table S4.4. Drift lengths are multiplied by 1,000.



Alternative explanations for non-zero estimated Basal Eurasian ancestry in Ancient North Eurasian-related populations using the f_4-ratio

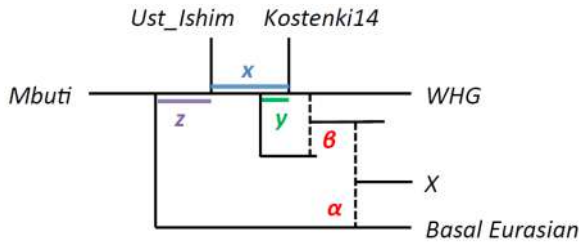
We were intrigued by the fact that X=MA1 *or* SHG *or* EHG, populations related to "Ancient North Eurasians", all have a positive estimate of "Basal Eurasian" ancestry according to the f_4-ratio estimate of Table S4.4, but do not have evidence of a significantly negative f_4(X, Kostenki14; Ust_Ishim, Chimp) statistic (Table S4.2) and thus it does not appear that Ust_Ishim shares more alleles with Kostenki14 than with X (which is expected if X has Basal

Eurasian ancestry). Thus, we do not think there is convincing evidence that these populations have Basal Eurasian ancestry and we consider here alternative explanations that would infer pseudo-Basal Eurasian for them.

A possible explanation, investigated further below, is that the "Basal Eurasian" ancestry in these populations branches off just prior to Ust_Ishim (minimizing the non_African→main_Eurasian genetic drift). Such a solution would preserve a statistic f_4(X, Kostenki14; Ust_Ishim, Chimp)~0, as (Basal Eurasian, Kostenki14, Ust_Ishim) would form an effective trifurcation.

A second possibility is illustrated in Fig. S4.5:

Figure S4.5: Bias in estimate of Basal Eurasian ancestry for populations tracing part of their ancestry from a split earlier than Kostenki14. If β=0 then the topology of Fig. S4.2 applies. If population X traces part of its ancestry to a split between Ust_Ishim and Kostenki14 then the value of the statistic f_4(X, WHG; Ust_Ishim, Kostenki14) = $\alpha x+(1-\alpha)\beta y$. This statistic is in the numerator of the f_4-ratio used in Table S4.4 to estimate Basal Eurasian admixture in a population, and can be positive even if α=0 (no Basal Eurasian ancestry).



Basal Eurasian ancestry is possible for many ancient populations

Fig. S4.5 shows how a positive estimate of Basal Eurasian ancestry is possible even in the absence of such ancestry. Such a bias may be present in a population X, as long as it traces part ($\beta(1-\alpha)$) of its ancestry from a population before the Kostenki14 but after the Ust_Ishim split (Fig. S4.5).

To investigate this bias we used ADMIXTUREGRAPH to write down the model of Fig. S4.5 and fit it for different populations. We show a fitted model in Fig. S4.6

Figure S4.6: A fitted model for the topology of Fig. S4.5 for *X*=Iran_N population. Key parameters of the fitted model are listed in Table S4.5. This model is a generalization of the model of Fig. S4.4; both fit statistically in the sense that they do not have any outliers |Z|>3, but the present model suggests that Iran_N may have some ancestry that splits off prior to the (Kostenki14, WHG) split. This model corresponds to the topology of Fig. S4.5 with *y* equal to the drift path Z→Z1.

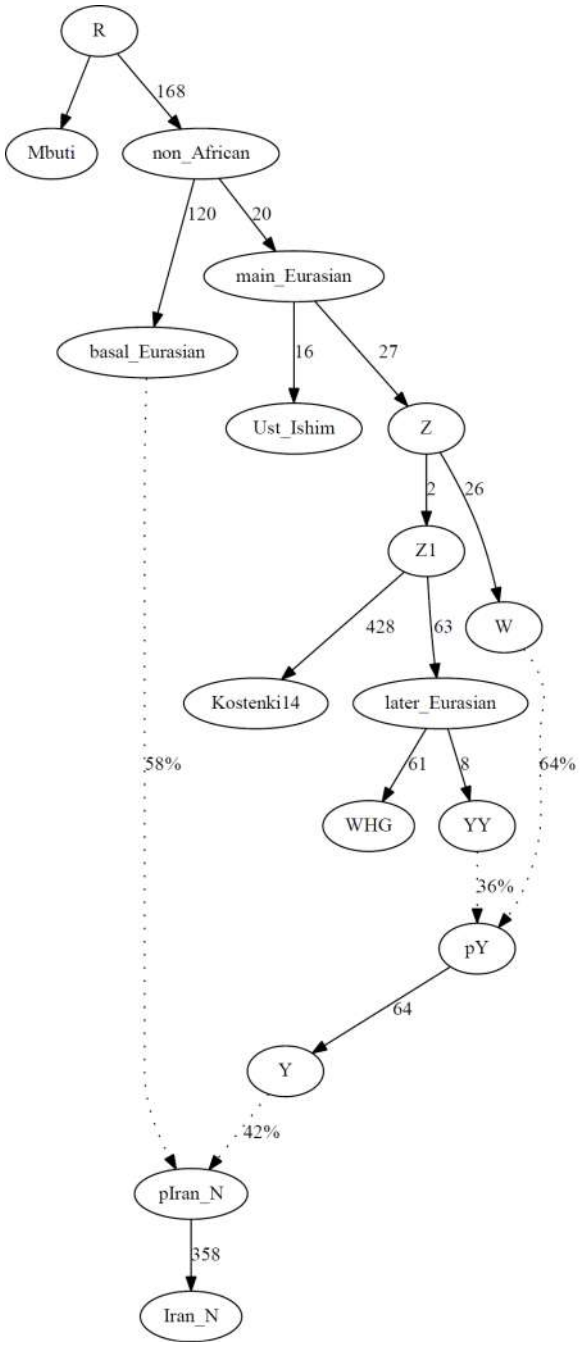


Table S4.5: Parameters of model of Figs. S4.5,6.

Test	*α*	*β*	z	Y	x-y
			non_African→main_Eurasian	Z→Z1	main_Eurasian→Z
Anatolia_ChL	0.2067	0.6583	0.0574	0.0000	0.0298
Anatolia_N	0.3629	0.5607	0.0218	0.0000	0.0283
Armenia_ChL	0.3663	0.6131	0.0205	0.0000	0.0276
Armenia_EBA	0.4439	0.5304	0.0207	0.0000	0.0281
Armenia_MLBA	0.4906	0.4026	0.0176	0.0000	0.0283
CHG	0.4965	0.5787	0.0153	0.0000	0.0283
EHG	0.2605	0.4953	0.0075	0.0000	0.0283
Europe_EN	0.3420	0.5283	0.0238	0.0000	0.0283
Europe_LNBA	0.3661	0.4462	0.0167	0.0000	0.0284
Europe_MNChL	0.3098	0.4709	0.0203	0.0000	0.0284
Iberia_BA	0.4228	0.3022	0.0167	0.0010	0.0275
Iran_ChL	0.4642	0.5947	0.0267	0.0000	0.0281
Iran_LN	0.5953	0.4819	0.0201	0.0046	0.0224
Iran_HotuIIIb	0.6792	0.6748	0.0142	0.0028	0.0282
Iran_N	0.5784	0.6365	0.0198	0.0015	0.0265
Iran_recent	0.3939	0.4882	0.0262	0.0000	0.0277
Levant_BA	0.3929	0.6185	0.0317	0.0000	0.0279
Levant_N	0.3673	0.5420	0.0394	0.0000	0.0279
MA1	0.2854	0.7655	0.0000	0.0000	0.0274
Natufian	0.4378	0.5032	0.0358	0.0000	0.0276
SHG	0.1994	0.4710	0.0087	0.0000	0.0285
Steppe_EMBA	0.4015	0.4798	0.0137	0.0000	0.0284
Steppe_Eneolithic	0.3021	0.4689	0.0180	0.0000	0.0288
Steppe_IA	0.2688	0.5818	0.0194	0.0000	0.0285
Steppe_MLBA	0.3766	0.4523	0.0159	0.0000	0.0284
Switzerland_HG	0.0892	0.3481	0.0068	0.0000	0.0290

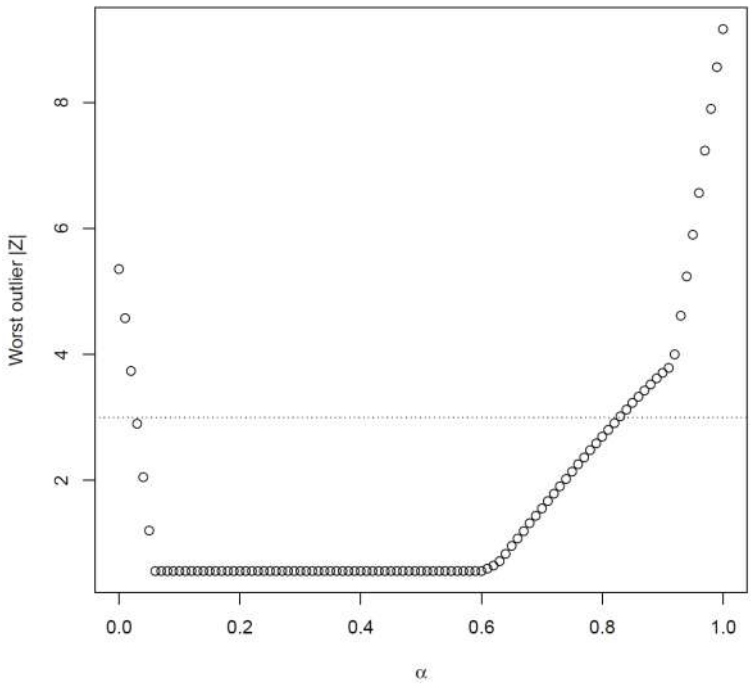
This procedure suggests that populations from ancient Iran may have ancestry from a population that splits off prior to Kostenki14 but after Ust_Ishim, as they have non-zero values (Table S4.5) of parameter *y* (Fig. S4.5). A different solution is arrived at for MA1 and EHG/SHG. While "Basal Eurasian" admixture is inferred for these populations, they also have low (or in the case of MA1 zero) drift *z* on the non_African→main_Eurasian branch. By forming, effectively, the trifurcation between ("Basal Eurasian", Ust_Ishim, and Kostenki14) previously mentioned.

Basal Eurasian ancestry is not necessary for European hunter-gatherer populations or MA1

The solutions of Table S4.5 optimize all graph parameters simultaneously. We next fixed the mixture proportion *α* of Basal Eurasian ancestry in the interval [0,1] in 0.01 increments and

re-fit the model. We show the Z-score of the worst outlier in Fig. S4.7 for the model of Fig. S4.6 and for the Iran_N population.

Figure S4.7: Worst *f*-statistic outlier Z-score for different amount of Basal Eurasian proportion *α* of the model of Fig. S4.6. The |Z|=3 line is shown. The worst outlier for α=0 is f_4(Mbuti, Ust_Ishim; Iran_N, WHG) whose estimated value is 5.3 standard errors higher than its fitted value. The worst outlier for α=1 is f_4(Mbuti, Iran_N; Ust_Ishim, WHG) whose estimated value is 9.2 standard errors higher than its fitted value. The twin constraints of (i) Ust_Ishim sharing more drift with WHG than with Iran_N, and of Iran_N sharing more drift with WHG than with Ust_Ishim make infeasible any models that derive all (or most) of the ancestry of Iran_N from either Basal Eurasians or a population related to WHG.



We list the feasible solution range (defined as the values of α for which the fitted model had no |Z|>3 outlier *f*-statistics) for each population in Table S4.6.

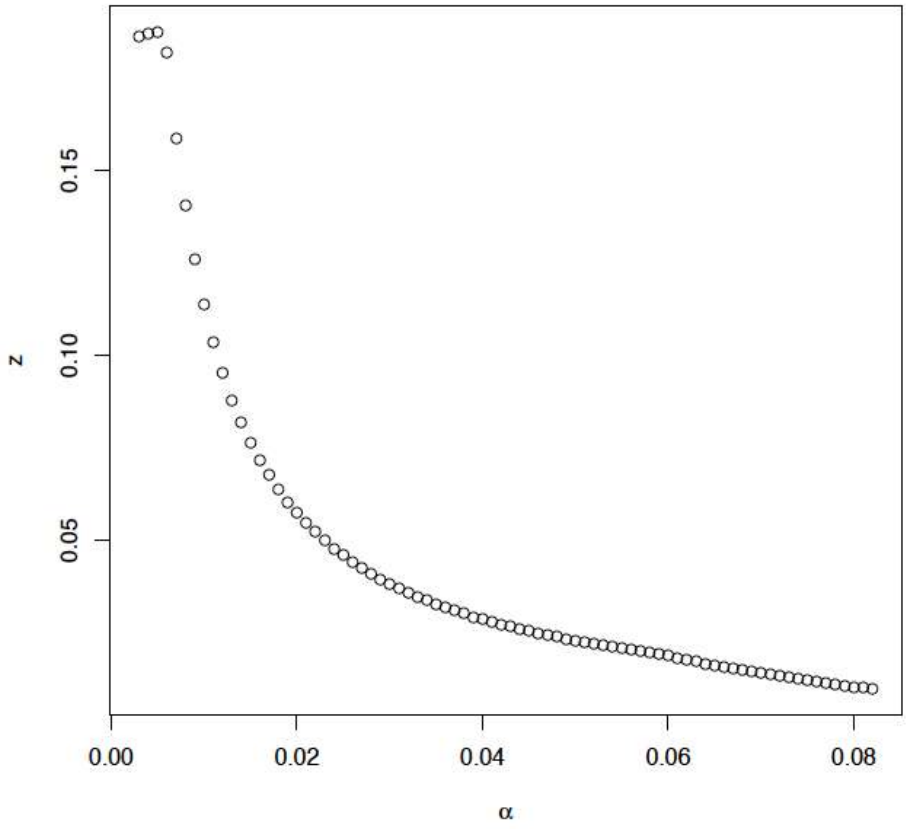
Table S4.6: Feasible ranges of Basal Eurasian admixture for different populations

Test	*Min*	*Max*
Anatolia_ChL	3	47
Anatolia_N	3	49
Armenia_ChL	2	52
Armenia_EBA	3	61
Armenia_MLBA	3	64
CHG	2	68
EHG	0	44
Europe_EN	3	46
Europe_LNBA	3	47
Europe_MNChL	2	42
Iberia_BA	0	70
Iran_ChL	5	61
Iran_LN	4	88
Iran_HotuIIIb	0	97
Iran_N	3	82
Iran_recent	3	62
Levant_BA	5	56
Levant_N	6	55
MA1	0	52
Natufian	5	68
SHG	0	33
Steppe_EMBA	2	52
Steppe_Eneolithic	1	49
Steppe_IA	1	47
Steppe_MLBA	2	49
Switzerland_HG	0	27

While no Basal Eurasian ancestry is feasible for MA1, EHG, SHG, and Switzerland_HG at least some such ancestry is necessary for the other populations. The range of this ancestry is wide. A useful illustration of the model behavior for different feasible values of α is shown in Fig. S4.8

A population with Basal Eurasian admixture has a positive f_4(Mbuti, Ust_Ishim; X, Kostenki14) = αz (Fig. S4.5). This is the equation of a hyperbola, and is the main explanation for the shape of the curve of Fig. S4.8. The first three feasible values of α (3-5%) are the only exception to this, where z plateaus to a maximum value constrained by the empirical genetic drift between Mbuti and non-Africans.

Figure S4.8: As the proportion of Basal Eurasian ancestry α decreases, it is derived from an ever more distant split as measured by the genetic drift z (Fig. S4.5). The plot for Iran_N is shown.

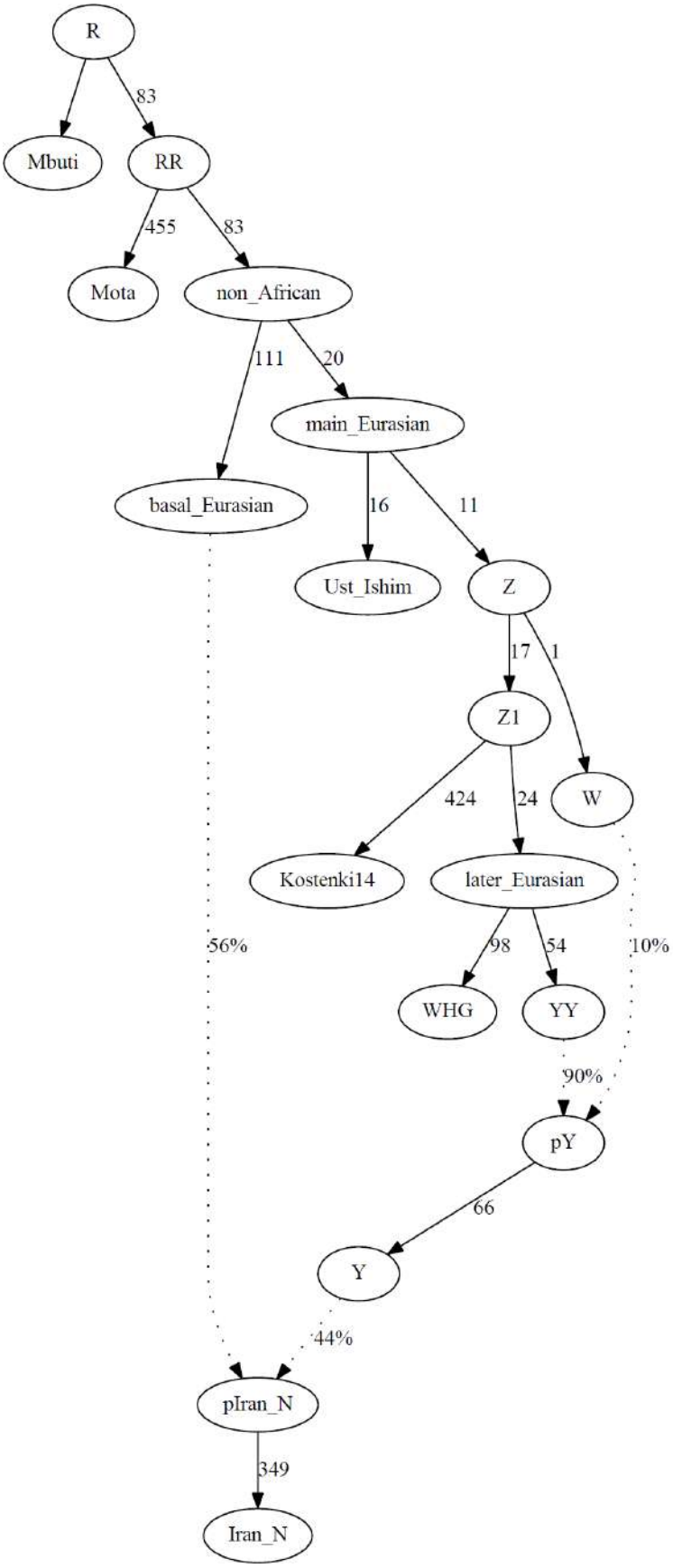


Adding Mota as an additional constraint

Fig. S4.8 suggests a way in which the proportion of Basal Eurasian can be constrained. We can find a population that is phylogenetically closer (than the Mbuti) to non-African populations, thus reducing the allowable range of z and placing a more stringent lower bound on α.

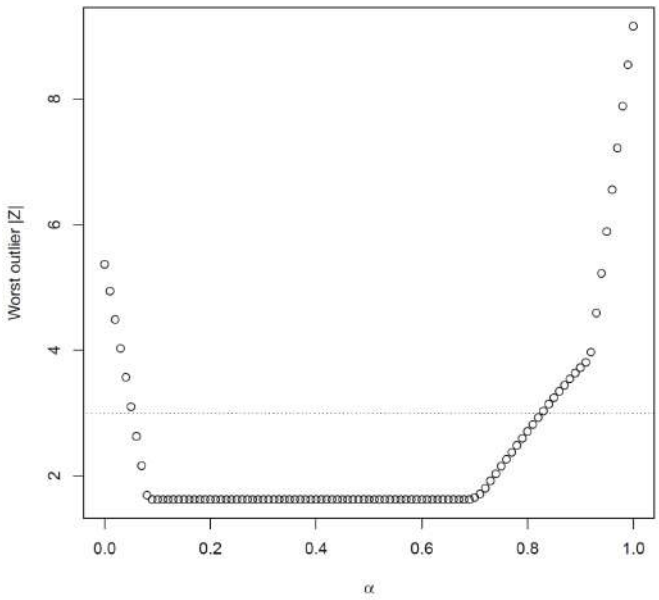
As a proof of concept, we add Mota[1010], a ~4,500 year old male from the Ethiopian highlands as an additional constraint (Fig. S4.9). This model fits, and the Out-of-Africa bottleneck drift RR→non_African becomes 0.083.

Figure S4.9: Adding Mota as an additional constraint to the model of Fig. S4.6



We next vary α in the [0,1] interval as before and show in Fig. S4.9 the Z-score of the worst *f*-statistic outlier. The feasible range is reduced from [3, 82] (Fig. S4.7) to [6, 82].

Figure S4.9: Worst *f*-statistic outlier Z-score for different amount of Basal Eurasian proportion α of the model of Fig. S4.8.



In Table S4.7 we list the feasible range of Basal Eurasian ancestry for the model structure of Fig. S4.8. Three populations did not have a feasible range according to the |Z|<3 criterion (Iberia_BA: |Z|=3.1, Steppe_EMBA: |Z|=3.2, Steppe_MLBA: |Z|=3.3).

Table S4.7: Feasible ranges of Basal Eurasian admixture for different populations (with Mota constraint).

Test	*Min*	*Max*
Anatolia_N	5	48
Armenia_ChL	4	52
Armenia_EBA	5	58
Armenia_MLBA	5	64
CHG	3	68
EHG	0	40
Europe_EN	5	46
Europe_LNBA	4	45
Europe_MNChL	4	42
Iran_ChL	8	61
Iran_LN	6	88
Iran_HotuIIIb	0	97
Iran_N	6	82
Iran_recent	5	62
Levant_BA	8	56

Levant_N	10	55
MA1	0	52
Natufian	9	68
SHG	0	33
Steppe_Eneolithic	1	49
Steppe_IA	1	46
Switzerland_HG	0	27

MA1, EHG, SHG, Switzerland_HG are consistent with having no Basal Eurasian ancestry, while at least some such ancestry is inferred for the remaining populations.

Neolithic Iran and Natufians could be derived from the same Basal Eurasian population but are genetically closer to EHG and WHG respectively

We take the model of Fig. S4.9 and attempt to fit Natufians as a mixture of the same Basal Eurasian population that contributes to Iran_N and any other population of the tree. Several solutions are feasible, and we show the best one (lowest ADMIXTUREGRAPH score) in Fig. S4.10. We can add both EHG and MA1 as simple branches to the model structure of Fig. S4.10 and show the results in Fig. S4.11. An interesting aspect of this model is that it derives both Natufians and Iran_N from Basal Eurasians but Natufians have ancestry from a population related to WHG, while Iran_N has ancestry related to EHG. Natufians and Iran_N may themselves reside on clines of WHG-related/EHG-related admixture. The fact that these two populations are differentially related to European hunter-gatherers can be directly seen from the following statistics:

X	Y	f_4(Iran_N, Natufian; X, Y)	Z
EHG	Mbuti	0.00204	3.4
WHG	Mbuti	-0.00241	-4.3
WHG	EHG	-0.00441	-8.9

The statistic f_4(Iran_N, Iran_HotuIIIb; EHG, Mbuti) = -0.00199 (Z=-2.4) suggests that the singleton individual from Hotu (Iran_HotuIIIb) was shifted towards EHG along the Iran_N/EHG cline, albeit it does not reach |Z|>3. There is uncertainty about the date of Iran_HotuIIIb, as it is not certain that it is of Mesolithic age and thus predates the Neolithic of Iran from Ganj Dareh. The fact that the Caucasus hunter-gatherers (CHG) (who are definitely pre-Neolithic) have extra EHG-related ancestry is also supportive of a substantial antiquity of this element in the Caucasus-Iran region. It is not clear whether the hunter-gatherers preceding the Neolithic in Ganj Dareh were similar to Iran_HotuIIIb or the CHG and their EHG affinity was diluted during Neolithization, or whether they are descended from an unsampled hunter-gatherer population that already had this reduced affinity to the EHG.

Figure S4.10: Co-modelling Iran_N and Natufians as a mixture of the same Basal Eurasian population.

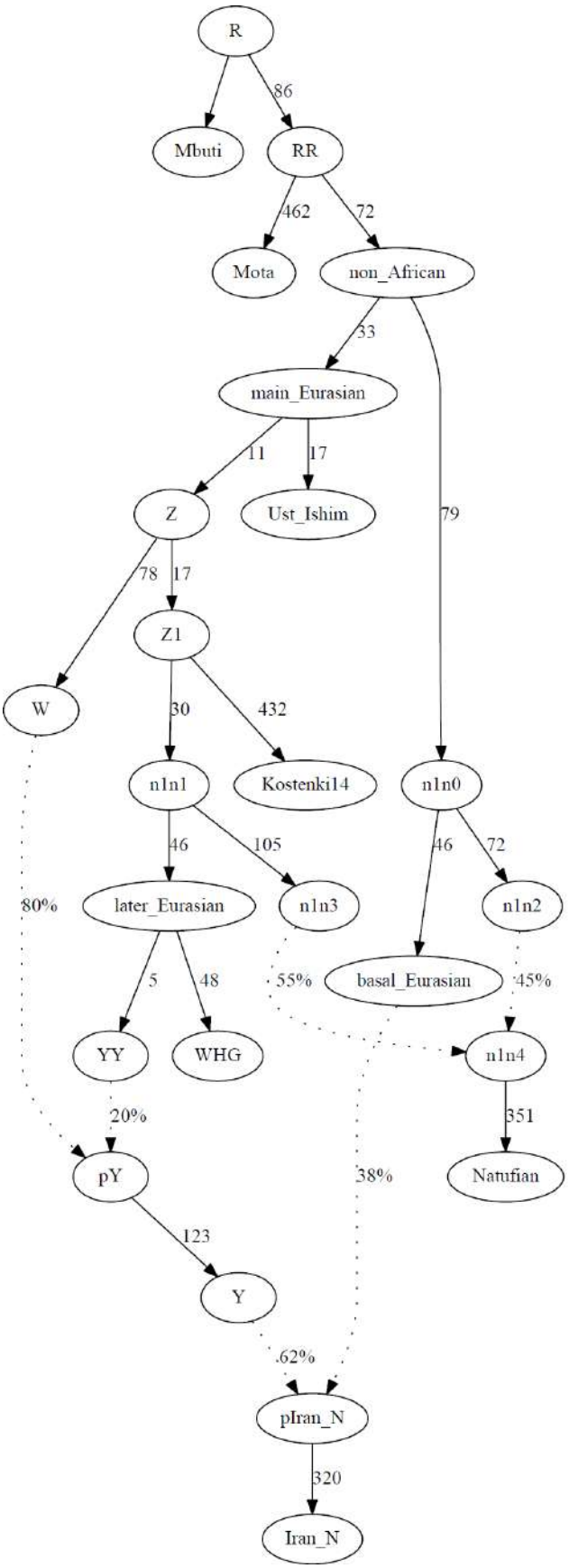
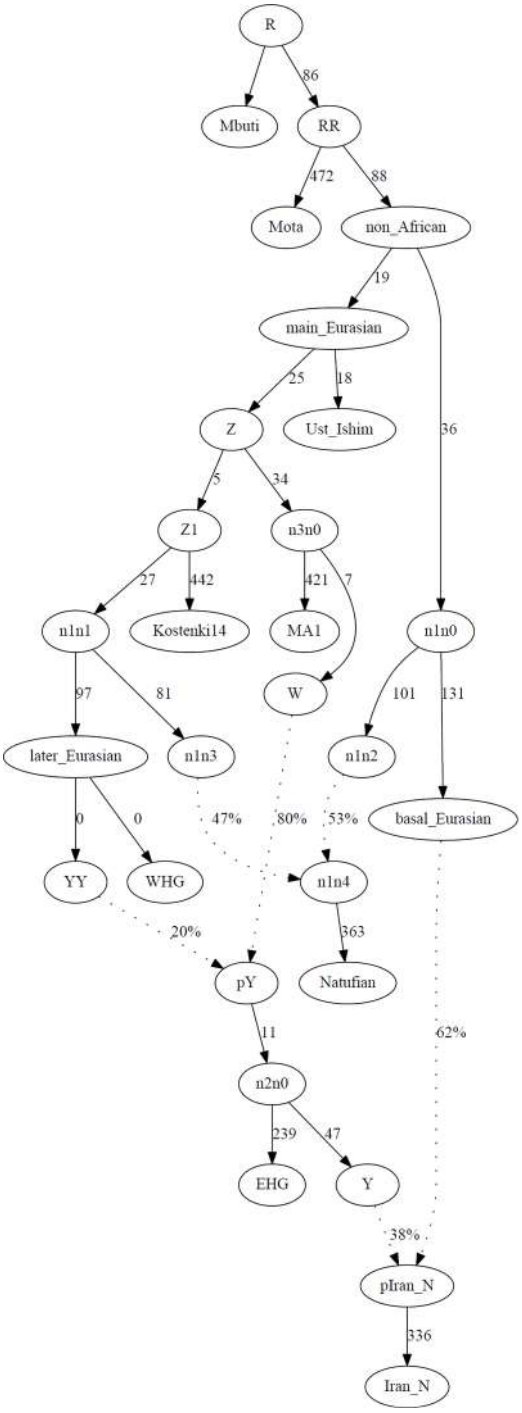


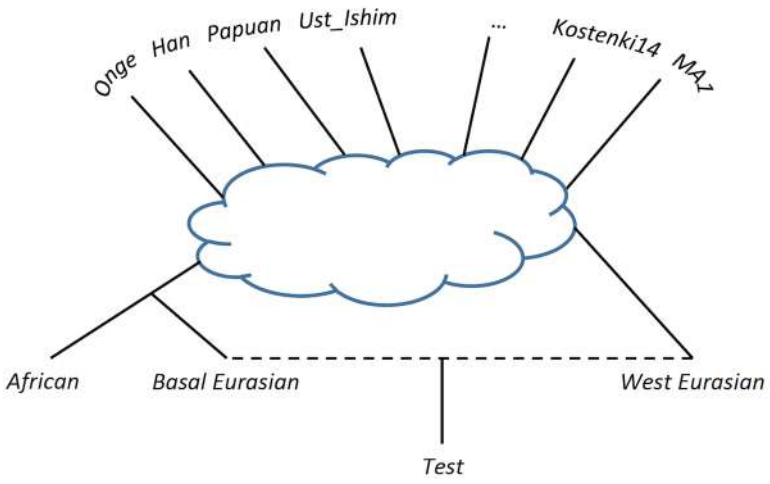
Figure S4.11: A model of western Eurasian history that incorporates Upper Paleolithic Eurasians (Ust_Ishim, Kostenki14, MA1) and the four most genetically differentiated Holocene populations (Natufians, Iran_N, WHG, EHG).



Estimating Basal Eurasian ancestry without an explicit model

An elegant solution to the problem of estimating α is not to model the phylogeny of diverse Eurasian populations directly, but rather to use the "qpAdm" methodology described in Supplementary Information, section 10 of ref. 7 to obtain mixture proportions without an explicit phylogeny. A schematic representation of this approach is seen in Fig. S4.12. We model the *Test* population as a mixture of a West Eurasian population with an African Outgroup. The qpAdm methodology also provides a formal test of whether the model of the *Test* population as a mixture of two source streams that are clades with West Eurasians and Africans – with all analyzed outgroups phylogenetically more distant – is a fit to the data.

Figure S4.12: Modeling Basal Eurasian admixture without an explicit phylogeny



We use the set of 6 outgroups shown in Fig. S4.12. This includes diverse eastern non-African populations (Onge from the Andaman Islands, Han from China, Papuans from New Guinea) and Upper Paleolithic Eurasians (Ust_Ishim[4], Kostenki14[6], MA1[3]). Modeling the relationships between all these populations is difficult and there are no ancient genomes from eastern Eurasia or Oceania that could constrain such models. Nonetheless, these populations are differentially related to ancient West Eurasian populations (Table S4.1) and thus potentially informative. We test three sets of populations for being derived from two streams of ancestry relative to the outgroups:

(*Test*, Mota, WHG)

(*Test*, Mota, EHG)

(*Test*, Mota, WHG, EHG)

The qpAdm method gives us both a P-value for whether the matrix is rank=1 or rank=2 (which if P>0.05 is consistent with two or three streams of ancestry respectively), and an estimate of the proportion α of ancestry that is a clade with Mota (that is, proportion of Basal Eurasian ancestry). Table S4.8 summarizes our results.

Table S4.8: Estimates of Basal Eurasian admixture with qpAdm methodology. The set of outgroups is (Ust_Ishim, Kostenki14, MA1, Han, Onge, Papuan). The sets of populations that are tested for being consistent with two streams of ancestry related to the outgroups are (*Test*, Mota, WHG), (*Test*, Mota, EHG), and with three streams of ancestry (*Test*, Mota, WHG, EHG). We highlight in red cases in which the P-value for rank=1 or rank=2 is >0.05, indicating that we cannot reject the model of descent from two or three streams of ancestry. The inferred mixture proportion α and its standard error are also presented.

	West Eurasian population(s) included among the references								
	WHG			EHG			WHG+EHG		
Test	P-value	α	s.e.	P-value	α	s.e.	P-value	α	s.e.
Anatolia_ChL	7.62E-03	0.167	0.057	8.98E-08	0.451	0.044	1.60E-02	0.213	0.061
Anatolia_N	8.18E-02	0.295	0.032	2.82E-18	0.566	0.030	4.67E-02	0.305	0.036
Armenia_ChL	3.06E-06	0.193	0.044	8.18E-07	0.458	0.030	3.32E-01	0.295	0.037
Armenia_EBA	9.10E-07	0.279	0.045	2.68E-06	0.513	0.031	1.01E-03	0.377	0.047
Armenia_MLBA	9.51E-15	0.267	0.066	5.89E-04	0.486	0.030	1.27E-03	0.420	0.045
CHG	2.10E-04	0.336	0.050	1.02E-03	0.532	0.034	6.91E-02	0.415	0.047
Europe_EN	2.06E-02	0.291	0.032	1.71E-17	0.570	0.029	1.13E-02	0.303	0.037
Europe_LNBA	5.21E-19	0.060	0.043	8.82E-08	0.387	0.024	1.26E-02	0.258	0.029
Europe_MNChL	5.35E-02	0.230	0.031	7.16E-21	0.524	0.032	6.95E-02	0.252	0.033
Iberia_BA	2.06E-02	0.116	0.069	1.17E-04	0.396	0.050	1.70E-01	0.189	0.067
Iran_ChL	1.96E-07	0.332	0.047	9.97E-06	0.540	0.030	2.14E-03	0.421	0.043
Iran_LN	4.37E-02	0.509	0.065	5.48E-02	0.671	0.044	1.42E-01	0.562	0.070
Iran_HotuIIIb	4.90E-02	0.646	0.108	6.44E-01	0.683	0.067	4.75E-01	0.684	0.098
Iran_N	1.53E-02	0.517	0.051	5.48E-02	0.648	0.036	2.45E-01	0.561	0.051
Iran_recent	5.10E-05	0.363	0.060	2.47E-04	0.552	0.041	1.11E-03	0.443	0.065
Levant_BA	1.62E-01	0.344	0.042	1.33E-08	0.584	0.033	2.31E-01	0.370	0.044
Levant_N	9.52E-02	0.334	0.043	5.38E-12	0.598	0.036	4.82E-02	0.334	0.048
Natufian	3.33E-01	0.399	0.054	1.34E-06	0.627	0.048	2.53E-01	0.384	0.058
SHG	2.27E-17	-0.210	0.059	2.11E-08	0.203	0.031	2.47E-01	0.051	0.033
Steppe_EMBA	1.57E-33	-0.208	0.077	3.32E-01	0.269	0.023	3.84E-01	0.239	0.032
Steppe_Eneolithic	2.83E-26	-0.327	0.092	8.04E-01	0.144	0.034	7.52E-01	0.163	0.047
Steppe_IA	3.94E-18	0.073	0.080	5.23E-07	0.339	0.040	1.55E-07	0.314	0.068
Steppe_MLBA	4.83E-23	-0.021	0.054	1.85E-05	0.340	0.024	1.71E-02	0.239	0.031
Switzerland_HG	2.49E-01	-0.009	0.051	1.25E-20	0.410	0.049	3.20E-01	-0.028	0.054

Disjoint sets of populations can be modeled as WHG+Mota or EHG+Mota mixtures. For WHG+Mota, the populations that can be modeled adequately are from the Levant, Anatolia, and Neolithic Europe. For EHG+Mota, the populations that can be modeled adequately are from Iran and the Eurasian steppe down to the Early/Middle Bronze Age (Steppe_EMBA) populations. The EHG+Mota modeling is adequate in the steppe continuing through the time of the Poltavka culture of the Middle Bronze Age[11], since as we show in Supplementary Information section 7, the Near Eastern migration into the steppe from Iran-related populations. The ~14,000 year old Upper Paleolithic hunter-gatherer from Switzerland[12] can also be modeled as WHG+Mota, but has no significant evidence of Basal Eurasian ancestry (α=-0.9±5.1%), consistent with its close relationship to WHG[12] (Fig. 1b). Scandinavian hunter-gatherers (SHG) from Motala[1,7,11] are an example of a population that cannot be modeled as either WHG+Mota (p=2.27E-17), or EHG+Mota (p=2.11E-08), but can be modeled as WHG+EHG+Mota (p=0.247). This population has both WHG and EHG ancestry[7]. By using both WHG and EHG as source populations, we are able to model it, and infer that its Basal Eurasian ancestry is 5.1±3.3% which is not significantly different than zero.

Several populations cannot be modeled even as WHG+EHG+Mota mixtures (Table S4.8), suggesting that their relationship to the outgroups may be more complex. By dropping some of the outgroups, we may remove populations that violate the topology of Fig. S4.12, that is, they interact with the Test population not via the reference populations. We can thus obtain an estimate of Basal Eurasian ancestry which is the parameter of interest here. The downside of dropping an outgroup is that we are removing potentially phylogenetic useful branching points within the "cloud" of Fig. S4.12 which provide leverage for inferring mixture proportions. We remove each of the six outgroups in turn and show the results in Table S4.9.

Three populations (Kostenki14, MA1, and Han), when dropped as outgroups, result in the quadruple (Test, WHG, EHG, Mota) being consistent with 3 streams of ancestry for all (or nearly all in the case of Han) Test populations. Removing Kostenki14 results in a blowup of the standard errors suggesting that it carries important phylogenetic information that is not present in the other outgroups. Removal of MA1 and Han suggests interactions between West Eurasia and Upper Paleolithic Siberia and East Asia which we explore in Supplementary Information, section 11. For our purpose of estimating Basal Eurasian ancestry, however (and unlike with Kostenki14), removing MA1 and Han from the set of outgroups does not result in a blowup of standard errors which remain modest (less than 10% for most Test populations). In Fig. 2 we plot graphically the Basal Eurasian estimate results when removing MA1, which results in successful modeling of all Test populations, and is thus the main estimate we use in the study.

Table S4.9: Estimates of Basal Eurasian ancestry with a single dropped outgroup population. The analysis here is the same as Table S4.8, with (Test, Mota, WHG, EHG) tested for consistency with being derived from three streams of ancestry (hence the P-value is for rank=2). We drop a single outgroup in each column of the table.

	Dropped Outgroup																	
	Ust_Ishim			Kostenki14			MA1			Han			Papuan			Onge		
Test	P-value	α	s.e.	P-value	α	s.e.	P-value	α	s.e.	P-value	α	s.e.	P-value	α	s.e.	P-value	α	s.e.
Anatolia_ChL	6.08E-03	0.199	0.062	6.23E-01	1.051	2.212	7.08E-01	0.084	0.072	9.43E-01	0.228	0.055	1.26E-02	0.243	0.065	1.28E-02	0.229	0.061
Anatolia_N	2.27E-02	0.298	0.035	5.08E-01	0.134	0.619	2.52E-01	0.253	0.039	5.26E-01	0.309	0.034	1.30E-01	0.333	0.037	2.31E-02	0.309	0.036
Armenia_ChL	2.13E-01	0.295	0.037	4.53E-01	0.335	0.711	8.17E-01	0.248	0.045	8.49E-01	0.294	0.035	3.14E-01	0.309	0.039	1.99E-01	0.298	0.037
Armenia_EBA	5.40E-04	0.388	0.045	4.67E-01	0.859	1.832	5.88E-01	0.276	0.052	4.22E-01	0.375	0.041	4.83E-03	0.413	0.047	3.76E-04	0.380	0.047
Armenia_MLBA	1.22E-03	0.399	0.043	1.96E-01	105.830	26.253	2.73E-01	0.305	0.055	7.69E-02	0.410	0.040	3.46E-02	0.464	0.045	4.10E-04	0.417	0.045
CHG	1.67E-02	0.417	0.049	5.31E-01	0.212	0.832	2.63E-01	0.347	0.058	3.47E-01	0.414	0.046	3.56E-01	0.456	0.049	3.00E-02	0.417	0.048
Europe_EN	7.80E-03	0.306	0.036	3.48E-01	0.605	1.149	5.77E-01	0.239	0.038	6.59E-01	0.312	0.034	2.12E-02	0.328	0.038	5.23E-03	0.309	0.038
Europe_LNBA	1.44E-02	0.256	0.028	2.57E-01	1.473	1.418	7.60E-01	0.201	0.034	5.81E-01	0.256	0.026	2.36E-02	0.277	0.030	5.22E-03	0.260	0.029
Europe_MNChL	2.24E-02	0.252	0.033	3.88E-01	0.410	0.733	3.85E-01	0.210	0.035	8.66E-01	0.260	0.031	6.89E-02	0.268	0.034	6.76E-02	0.261	0.033
Iberia_BA	1.08E-01	0.161	0.067	5.57E-01	0.019	0.659	1.20E-01	0.162	0.074	1.70E-01	0.192	0.065	9.27E-01	0.229	0.065	9.07E-02	0.182	0.068
Iran_ChL	2.45E-03	0.431	0.042	1.91E-01	3.053	2.780	4.94E-01	0.317	0.051	1.02E-01	0.418	0.040	1.16E-02	0.451	0.044	6.78E-04	0.419	0.044
Iran_LN	1.56E-01	0.563	0.069	4.99E-01	0.916	1.276	9.31E-01	0.460	0.079	8.82E-01	0.567	0.065	1.20E-01	0.586	0.073	8.25E-02	0.574	0.072
Iran_HotuIIIb	3.63E-01	0.644	0.093	5.43E-01	-5.670	3.727	3.00E-01	0.665	0.125	3.09E-01	0.686	0.098	3.16E-01	0.698	0.106	7.84E-01	0.661	0.095
Iran_N	1.61E-01	0.577	0.052	2.33E-01	-378.404	674.748	9.27E-01	0.482	0.063	7.48E-01	0.559	0.050	1.37E-01	0.571	0.056	1.59E-01	0.568	0.052
Iran_recent	4.77E-02	0.450	0.057	4.33E-01	-86130.138	32292.937	2.32E-01	0.308	0.070	1.28E-02	0.432	0.061	1.07E-03	0.477	0.070	4.80E-04	0.436	0.064
Levant_BA	1.53E-01	0.368	0.044	4.01E-01	0.327	0.601	8.99E-01	0.313	0.049	8.18E-01	0.373	0.042	2.11E-01	0.389	0.047	1.23E-01	0.373	0.044
Levant_N	1.28E-02	0.326	0.049	4.39E-01	0.967	1.304	2.17E-01	0.277	0.050	6.46E-01	0.349	0.044	2.84E-02	0.353	0.052	1.40E-01	0.355	0.048
Natufian	5.87E-01	0.382	0.058	2.42E-01	0.997	0.788	3.91E-01	0.448	0.077	1.73E-01	0.385	0.058	1.64E-01	0.368	0.063	1.89E-01	0.393	0.059
SHG	2.01E-01	0.034	0.034	3.56E-01	-0.408	0.829	8.09E-01	0.097	0.039	4.72E-01	0.052	0.033	2.29E-01	0.038	0.035	1.29E-01	0.052	0.033
Steppe_EMBA	2.79E-01	0.235	0.032	3.33E-01	2.632	1.048	3.62E-01	0.216	0.042	2.56E-01	0.238	0.031	6.10E-01	0.254	0.033	4.24E-01	0.234	0.031
Steppe_Eneolithic	5.45E-01	0.146	0.047	9.55E-01	0.226	0.466	5.54E-01	0.171	0.061	5.99E-01	0.164	0.047	8.84E-01	0.145	0.049	5.48E-01	0.163	0.047
Steppe_IA	6.59E-06	0.329	0.062	2.58E-01	-7656.969	23037.913	8.96E-01	0.056	0.086	9.29E-02	0.274	0.055	6.83E-08	0.343	0.078	1.47E-07	0.322	0.068
Steppe_MLBA	1.77E-02	0.235	0.031	3.09E-01	0.453	0.986	2.54E-01	0.182	0.040	7.65E-02	0.237	0.030	1.20E-01	0.265	0.031	8.96E-03	0.236	0.031
Switzerland_HG	6.00E-01	-0.042	0.052	3.33E-01	1.171	1.396	4.26E-01	0.023	0.068	2.35E-01	-0.028	0.053	1.88E-01	-0.036	0.058	2.14E-01	-0.022	0.054

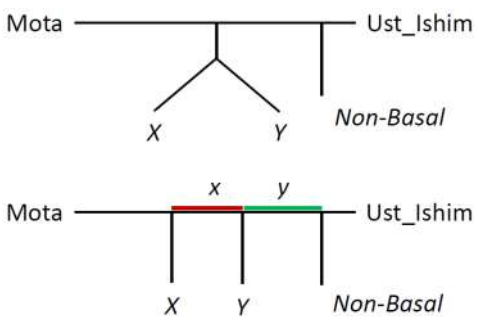
Basal Eurasian ancestry could stem from a single Out-of-Africa population

We have previously shown by ADMIXTUREGRAPH modeling (Fig. S4.10) that Natufians and Iran_N could derive their ancestry from the same Out-of-Africa Basal Eurasian population. Our method for inferring mixture proportions of Basal Eurasian ancestry without an explicit model (Fig. S4.12) relies on the fact that Basal Eurasians and an African reference (Mota) are both basal to other non-Africans and thus we can estimate mixture proportions of Basal Eurasian ancestry indirectly, using Mota as a reference population. However, this has the drawback that it estimates ancestry from any population that holds the same phylogenetic position and thus mixture proportions for the different tested populations (Table S4.9) do not necessarily reflect ancestry from a concrete 'Basal Eurasian' population, but (potentially) an aggregate of diverse kinds of ancestry with the common property that they derive from populations that split off prior to the differentiation of other non-Africans.

Populations as diverse as primates, archaic humans, Africans, and 'Basal Eurasians' all occupy a basal phylogenetic position to non-Africans and we would like some assurance that our estimates of 'Basal Eurasian' ancestry are not conflated by these other kinds of ancestry.

We can investigate this possibility by removing Mota from the set of reference populations and placing it into the right set of outgroups which now becomes: Mota, Ust_Ishim, Kostenki14, Han, Papuan, Onge. By doing so, we gain access to the phylogenetic branch Mota→Ust_Ishim, i.e., a branch that includes a sub-Saharan African population from East Africa and an 'undifferentiated' non-African population that is equally related to European hunter-gatherers and eastern non-Africans[4]. Two populations basal to non-Africans like Ust_Ishim may occupy the following phylogenetic positions vis a vis Mota, Ust_Ishim, and another *Non-Basal* population (Fig. S4.13).

Figure S4.13: Possible phylogenetic positions of two populations in relation to Ust_Ishim, *Non-Basal*, and Mota. In the top, *X* and *Y* form a clade vis a vis the other populations; in the bottom, the non-Africans (Ust_Ishim and *Non-Basal*) share drift length *x* with *Y* but not with *X* or Mota.



In the top phylogeny of Fig. S4.13, we have F_4(*X*, *Y*; Mota, Ust_Ishim or *Non-Basal*) = 0 and it is thus impossible (using Mota and Ust_Ishim as outgroups) to determine whether a population has ancestry from *X* or *Y*. These *X* and *Y* are descended from one stream of ancestry vis a vis (Mota, Ust_Ishim, *Non-Basal*). By contrast, in the bottom phylogeny of Fig. S4.13 we have F_4(*X*, *Y*; Mota, Ust_Ishim) = $x>0$ and *X* and *Y* are descended from two streams of ancestry vis a vis (Mota, Ust_Ishim, *Non-Basal*). Note that if Mota had not been introduced as an outgroup, the phylogenetic position of *X* and *Y* vis a vis the remaining outgroups (Ust_Ishim and *Non-Basal*) in the top and bottom phylogenies of Fig. S4.13 is equivalent.

We first set the Left set (WHG, EHG) and show that we reject rank=0 (p= 1.06e-14). We next use the Left set (Natufian, WHG, EHG) and show that we reject rank=1 (p=2.63e-11), hence Natufians have a type of ancestry that is different from what is found in the European hunter-gatherers WHG and EHG. Next, we use the Left set (*Test*, Natufian, WHG, EHG) for a variety of other *Test* populations and show the results in Table S4.10. If *Test* has yet another type of ancestry that is different from what is found in (Natufian, WHG, EHG) then by introducing *Test* to the Left set we will increase the rank.

However, for nearly all ancient populations (except for MA1 who interacts with Han in the set of outgroups; see Supplementary Information, section 11), rank=2 is not rejected, and hence each *Test* population does not have evidence for any type of ancestry other than what is found in the Natufians. This is supportive of the idea that the Basal Eurasian ancestry in the earliest population of the Levant (Natufians) and in virtually all other ancient West Eurasians (including ancient Iran) is from the same population. Note that the method can clearly reject rank=2 for other types of ancestry 'basal' to Eurasians, including Chimp, archaic humans (Altai and Denisovan), and Sub-Saharan Africans.

We should caution that our inability to reject different types of ancestry in Natufians and other ancient West Eurasian populations does not in itself prove that the Basal Eurasians were a single distinct population. For example, if there were multiple layers of Basal Eurasians with vanishingly small *x* (Fig. S4.13), then these would be effectively symmetrically related to the outgroups (the bottom case of Fig. S4.13 tends to the top one as $x\rightarrow 0$). Further advances in ancient DNA technology may allow future studies to study Upper and Middle Paleolithic populations directly and test whether early populations from the Levant such as the Skhul/Qafzeh hominins[13] and the makers of Middle Paleolithic stone industries from Arabia[14,15], or later ones from the cusp of the Upper Paleolithic transition[16], could correspond to the Basal Eurasians.

Table S4.10: Testing for rank=2 using Left=(*Test*, Natufian, WHG, EHG) and Right=(Mota, Ust_Ishim, Kostenki14, Han, Papuan, Onge) for different *Test* populations.

Test	p-value for rank=2
Ju_hoan_North	2.31E-04
MA1	3.32E-04
Mbuti	4.28E-04
Yoruba	4.64E-04
Dinka	8.87E-04
Chimp	3.38E-03
Denisovan	9.85E-03
Altai	2.10E-02
Levant_N	1.24E-01
Steppe_EMBA	1.67E-01
Iran_recent	1.81E-01
Armenia_MLBA	2.13E-01
Steppe_MLBA	2.14E-01
Iran_HotuIIIb	2.43E-01
Iran_ChL	2.68E-01
Anatolia_N	2.95E-01
Anatolia_ChL	3.04E-01
Steppe_Eneolithic	3.26E-01
CHG	3.43E-01
Iran_LN	3.71E-01
SHG	5.50E-01
Europe_EN	5.75E-01
Switzerland_HG	6.21E-01
Iran_N	7.03E-01
Iberia_BA	7.27E-01
Europe_LNBA	7.31E-01
Armenia_EBA	7.73E-01
Levant_BA	7.90E-01
Europe_MNChL	8.15E-01
Armenia_ChL	9.17E-01
Steppe_IA	9.66E-01

References

1 Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413 (2014).
2 Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225-228 (2014).
3 Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87-91 (2014).
4 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449 (2014).
5 Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216-219 (2015).
6 Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118 (2014).
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 Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060 (2010).
9 Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093 (2012).
10 Llorente, M. G. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* **350**, 820-822 (2015).
11 Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503 (2015).
12 Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun* **6** (2015).
13 Bar-Yosef, O. *The chronology of the Middle Paleolithic of the Levant*. 39-56 (New York: Plenum Press, 1998).
14 Armitage, S. J. *et al.* The southern route "Out of Africa": evidence for an early expansion of modern humans into Arabia. *Science* **331**, 453-456 (2011).
15 Rose, J. I. *et al.* The Nubian Complex of Dhofar, Oman: an African middle stone age industry in Southern Arabia. *PLoS One* **6**, e28239 (2011).
16 Hershkovitz, I. *et al.* Levantine cranium from Manot Cave (Israel) foreshadows the first European modern humans. *Nature* **520**, 216-219 (2015).

Supplementary Information 5

Ancient Near Easterners had less Neanderthal ancestry than ancient Europeans

It has been observed that West Eurasians have less Neanderthal admixture than Eastern non-African populations[1]. We wanted to investigate whether this could be due to West Eurasians having ancestry from a source that possessed less Neanderthal ancestry, and so studied statistics of the form f_4(Ancient West Eurasian$_1$, Ancient West Eurasian$_2$; Altai, Chimp) and f_4(Ancient West Eurasian$_1$, Ancient West Eurasian$_2$; Altai, Denisovan). The expected value of these statistics is zero if (Ancient West Eurasian$_1$, Ancient West Eurasian$_2$) both experienced a shared admixture event with Neanderthals prior to the differentiation and positive if Neanderthal admixture affected Ancient West Eurasian$_1$ more than Ancient West Eurasian$_2$.

Significantly positive statistics (Table S5.1) involve an individual from mainland Europe or the Eurasian steppe as Ancient West Eurasian$_1$ and a Near Eastern population as Ancient West Eurasian$_2$. The most convincing evidence is for the populations of Neolithic Iran where both sets of statistics (involving either Chimp or Denisovan) are significantly positive. These statistics suggest that ancient populations from the Near East shared fewer alleles with the Altai Neanderthal than those from Europe (including hunter-gatherers and later populations with European hunter-gatherer admixture).

We also combined ancient European and steppe samples into an "EuropeSteppe" meta-population and ancient Near Eastern samples into an "AncientNearEast" meta-population.

"EuropeSteppe": Anatolia_N, EHG, Europe_EN, Europe_LNBA, Europe_MNChL, Iberia_BA, SHG, Steppe_EMBA, Steppe_Eneolithic, Steppe_IA, Steppe_MLBA, Switzerland_HG, WHG

"AncientNearEast": Anatolia_ChL, Armenia_ChL, Armenia_EBA, Armenia_MLBA, CHG, Iran_ChL, Iran_LN, Iran_HotuIIIb, Iran_N, Iran_recent, Levant_BA, Levant_N, Natufian

(The Anatolian Neolithic and Chalcolithic samples are placed in EuropeSteppe and AncientNearEast respectively based on their overall placement in PCA; Fig. 1b).

To test the robustness of our inference using the high-coverage Altai Neanderthal[1], we also included additional Neanderthal genomes from Vindija (Vi_merge) and Mezmaiskaya (MezE) and report these statistics in Table S5.2. The signal is present for Altai and Vindija[2]

(3.6≤Z≤5.7) but not for Mezmaiskaya (Z≤1.3). The lower number of SNPs in this individual might weaken the signal; another possibility is this sample's provenance in the Caucasus which might interact with the Near East in two ways: first, contamination from present-day Caucasian populations into Mezmaiskaya, or second, Neanderthal admixture into ancient Near Eastern populations from a population more closely related to Mezmaiskaya than to the Neanderthals from Europe and the Altai.

The fact that Near Eastern populations have lower Neanderthal admixture than European ones led us to consider the possibility that Basal Eurasian admixture in the former is diluting their Neanderthal admixture. Indeed, when we plot estimates of Basal Eurasian ancestry (Fig. 2; Supplementary Information, section 4) against the statistic f_4(*Test*, Mbuti; Altai, Denisovan) (Fig. S5.1) which measures whether the Altai Neanderthal shares more alleles with a *Test* population than with the Mbuti Africans, we observe a striking negative correlation between the two (Fig. 2), suggesting that this is the case. Thus, allele sharing with Neanderthals is inversely correlated with the amount of Basal Eurasian ancestry. This provides direct evidence relevant to the long-standing question of the lower Neanderthal admixture in West Eurasian populations[3-6], in favor of the idea that ancient Near Eastern populations had lower Neanderthal admixture than those from Europe and through their later mixtures in the Neolithic diluted this ancestry throughout West Eurasia.

Our finding that at least Natufians had Y-chromosomes of African origin (Supplementary Information, section 6) suggested to us initially that gene flow from Africa (which did not experience Neanderthal admixture) may have contributed Basal Eurasian ancestry into the ancient Near East, although the alternative scenario of this ancestry stemming from earlier Out-of-Africa migrants >100 thousand years ago[7] cannot be rejected. However, Natufians do not have less Neanderthal ancestry (Fig. S5.1) or more Basal Eurasian ancestry (Supplementary Information, section 4) than the Neolithic of Iran, and so do not appear to be exceptional in either respect within the context of the ancient Near East. The presence of Basal Eurasian ancestry not only in the Epipaleolithic Natufians from the Levant, but also of the Upper Paleolithic/Mesolithic of Georgia[8] suggest that while this type of ancestry first appears in Europe with the Early Neolithic[7], it was already pervasive in the Near East before the advent of the Neolithic. Future studies of human remains from the Near East may determine (i) how much earlier the Basal Eurasians were present there, (ii) whether they represent a population of African origin or one that lived in Eurasia but did not experience the Neanderthal admixture of other Eurasians, and (iii) when the dilution of Neanderthal admixture first took place.

Table S5.1: Significant statistics (|Z|≥3) of the form f_4(Ancient West Eurasian$_1$, Ancient West Eurasian$_2$; Altai, Chimp) and f_4(Ancient West Eurasian$_1$, Ancient West Eurasian$_2$; Altai, Denisovan)

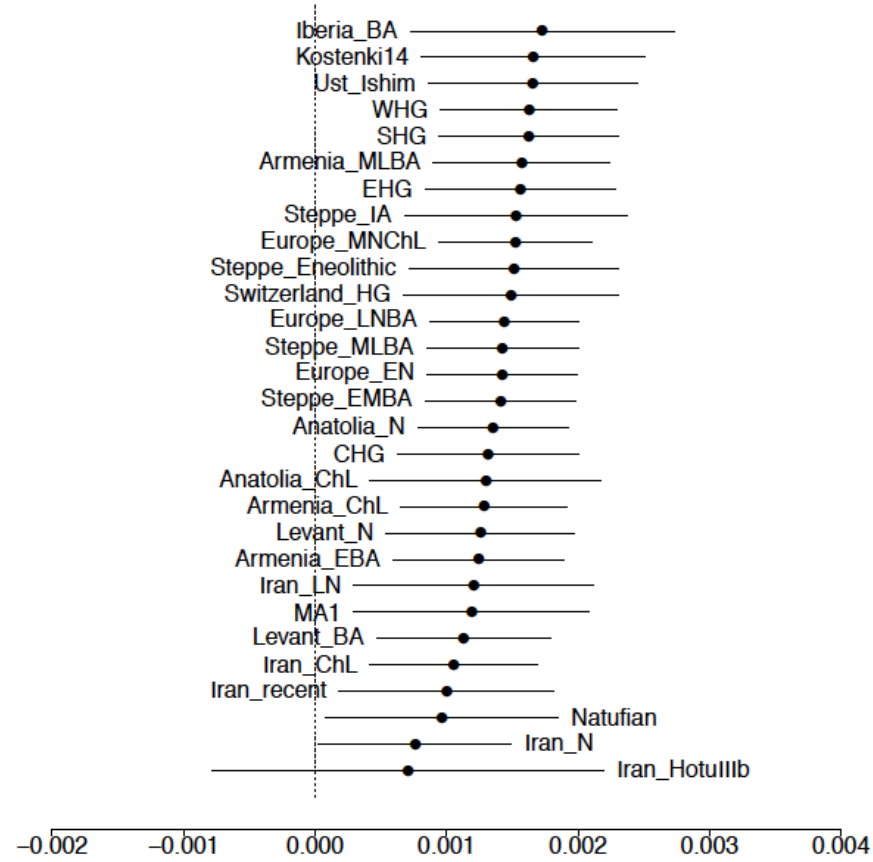
Ancient West Eurasian1	Ancient West Eurasiasn2	f_4(Ancient West Eurasian$_1$, Ancient West Eurasian$_2$, Altai, Chimp)	Z	Number of SNPs
Steppe_MLBA	Iran_N	0.00123	5.1	764177
Steppe_MLBA	Iran_ChL	0.00071	4.4	900462
Steppe_MLBA	Levant_N	0.00091	4.3	747971
Steppe_EMBA	Iran_N	0.00100	4.3	764161
Europe_MNChL	Iran_N	0.00102	4.2	764221
Europe_LNBA	Iran_N	0.00099	4.2	764219
Steppe_Eneolithic	Iran_N	0.00138	4.1	557406
WHG	Iran_N	0.00118	4.1	764094
Steppe_MLBA	Armenia_EBA	0.00063	4.0	896896
Steppe_MLBA	Anatolia_N	0.00048	4.0	1003388
Steppe_IA	Iran_N	0.00137	4.0	652653
EHG	Iran_N	0.00121	3.9	724059
SHG	Iran_N	0.00109	3.8	731428
Steppe_MLBA	Levant_BA	0.00067	3.7	783536
SHG	Iran_ChL	0.00078	3.7	851283
Steppe_MLBA	CHG	0.00070	3.6	1012914
Armenia_ChL	Iran_N	0.00090	3.6	736697
MA1	Iran_N	0.00135	3.6	550418
Steppe_MLBA	Iran_recent	0.00090	3.5	698462
Steppe_IA	Levant_N	0.00111	3.5	640122
Europe_EN	Iran_N	0.00084	3.5	764216
Steppe_MLBA	Europe_EN	0.00038	3.4	1013102
Armenia_MLBA	Iran_N	0.00095	3.4	711903
SHG	Levant_N	0.00089	3.4	717109
EHG	Iran_ChL	0.00083	3.3	843055
Steppe_MLBA	Iran_LN	0.00101	3.3	455766
MA1	Iran_LN	0.00159	3.2	330829
Steppe_Eneolithic	Levant_N	0.00098	3.2	548623
Europe_MNChL	Iran_ChL	0.00051	3.2	900612
Europe_LNBA	Levant_N	0.00063	3.2	747994
Europe_MNChL	Levant_N	0.00062	3.2	748010
Steppe_EMBA	Iran_ChL	0.00050	3.2	900423
Europe_LNBA	Iran_ChL	0.00048	3.2	900579
WHG	Levant_N	0.00083	3.1	747882
Anatolia_N	Iran_N	0.00075	3.1	763665
WHG	Iran_ChL	0.00071	3.1	900418
SHG	CHG	0.00074	3.1	921987
SHG	Armenia_EBA	0.00065	3.1	850497
MA1	Iran_recent	0.00126	3.1	506618
Iberia_BA	Iran_N	0.00135	3.1	293154
Armenia_ChL	Levant_N	0.00068	3.0	723619
EHG	Iran_recent	0.00097	3.0	670876

Ancient West Eurasian$_1$	Ancient West Eurasian$_2$	f_4(Ancient West Eurasian$_1$, Ancient West Eurasian$_2$, Altai, Denisovan)	Z	Number of SNPs
Europe_MNChL	Iran_N	0.00077	3.7	794839
Armenia_MLBA	Iran_N	0.00084	3.6	740676
WHG	Iran_N	0.00083	3.5	794703
Steppe_EMBA	Iran_N	0.00064	3.4	794777
SHG	Iran_N	0.00080	3.4	760506
Steppe_MLBA	Iran_N	0.00066	3.4	794793
Europe_LNBA	Iran_N	0.00064	3.3	794838
EHG	Iran_N	0.00083	3.3	752732
Armenia_MLBA	Iran_ChL	0.00054	3.2	851428
Europe_EN	Iran_N	0.00063	3.1	794835
Steppe_IA	Iran_N	0.00088	3.0	678032
Anatolia_N	Iran_N	0.00060	3.0	794257

Table S5.2: Statistics of the form f_4(AncientEurope, AncientNearEast; *Neanderthal*, *Outgroup*).

Neanderthal	*Outgroup*	f_4(AncientEurope, AncientNearEast; *Neanderthal*, *Outgroup*).	Z	Number of SNPs
All sites				
Altai	Chimp	0.00037	4.9	1014471
Altai	Denisovan	0.00024	3.8	1054004
MezE	Chimp	0.00026	1.6	111958
MezE	Denisovan	0.00011	0.8	116507
Vi_merge	Chimp	0.00047	5.6	677897
Vi_merge	Denisovan	0.00031	4.2	701178
Transversions				
Altai	Chimp	0.00021	1.5	189467
Altai	Denisovan	0.00025	2.2	197293
MezE	Chimp	0.00013	0.3	20124
MezE	Denisovan	0.00046	1.5	20995
Vi_merge	Chimp	0.00031	2.0	126395
Vi_merge	Denisovan	0.00039	2.8	130847

Figure S5.1: Symmetry test f_4(*Test*, Mbuti; Altai, Denisovan). Most ancient West Eurasian and Siberian populations share more alleles with the Altai Neanderthal[1] than with the Denisova individual[4], consistent with having more ancestry from Neanderthals than the Mbuti. The estimated value and ±3 standard errors is indicated.



References

1. Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, (2014).
2. Green, R. E. *et al.* A Draft Sequence of the Neanderthal Genome. *Science* **328**, 710-722, (2010).
3. Kim, Bernard Y. & Lohmueller, Kirk E. Selection and Reduced Population Size Cannot Explain Higher Amounts of Neanderthal Ancestry in East Asian than in European Human Populations. *The American Journal of Human Genetics* **96**, 454-461, (2015).
4. Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* **338**, 222-226, (2012).
5. Vernot, B. & Akey, Joshua M. Complex History of Admixture between Modern Humans and Neanderthals. *The American Journal of Human Genetics* **96**, 448-453, (2015).
6. Wall, J. D. *et al.* Higher Levels of Neanderthal Ancestry in East Asians Than in Europeans. *Genetics*, (2013).
7. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, (2014).
8. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, (2015).

Supplementary Information 6

Y-chromosome haplogroup variation in the ancient Near East

To determine the Y-chromosome haplogroup of male individuals, we used the nomenclature of the International Society of Genetic Genealogy (http://www.isogg.org) v. 9.129 (accessed Dec. 08, 2014). We called genotypes by requiring map quality and base quality ≥30, omitting 2 bases at the ends of reads, and taking the majority allele. For each individual, we determined the most derived SNP as well as downstream SNPs for which the individual was ancestral. The summary of Y-chromosome assignments is shown in Table S6.1, and the details are discussed below.

Table S6.1: Y-chromosome haplogroup assignments of ancient Near Easterners.

Region	ID	Haplogroup	Population
Armenia	I1407	L1a	Armenia_Chalcolithic
	I1632	L1a	Armenia_Chalcolithic
	I1634	L1a	Armenia_Chalcolithic
	I1635	R1b1x(R1b1a1, R1b1a2, R1b1c2, R1b1c3)	Armenia_EBA
Iran	I1293	J(xJ2a1b3,J2b2a1a1)	Iran_HotuIIIb
	I1945	P1(xQ,R1b1a2,R1a1a1b1a1b,R1a1a1b1a3a,R1a1a1b2a2a)	Ganj_Dareh_Iran_Neolithic
	I1949	CT	Ganj_Dareh_Iran_Neolithic
	I1671	G2a1(xG2a1a)	Iran_Late_Neolithic
	I1662	J(xJ1a,J2a1,J2b)	Iran_Chalcolithic
	I1674	G1a(xG1a1)	Iran_Chalcolithic
Levant	I0861	E1b1b1b2(xE1b1b1b2a,E1b1b1b2b)	Israel_Natufian
	I1069	E1b1(xE1b1a1,E1b1b1b1)	Israel_Natufian
	I1072	E1b1b1b2(xE1b1b1b2a,E1b1b1b2b)	Israel_Natufian
	I1685	CT	Israel_Natufian
	I1690	CT	Israel_Natufian
	I0867	H2	PPNB
	I1414	E(xE2,E1a,E1b1a1a1c2c3b1,E1b1b1b1a1,E1b1b1b2b)	PPNB
	I1415	E1b1b1	PPNB
	I1416	CT	PPNB
	I1707	T(xT1a1,T1a2a)	PPNB
	I1710	E1b1b1(xE1b1b1b1a1,E1b1b1a1b1,E1b1b1a1b2,E1b1b1b2a1c)	PPNB
	I1727	CT(xE,G,J,LT,R,Q1a,Q1b)	PPNB
	I1700	CT	PPNC
	I1705	J1(xJ1a)	Jordan_EBA
	I1730	J(xJ1,J2a,J2b2a)	Jordan_EBA

Armenia_ChL (Chalcolithic Armenia)

All three males from this population belong to Y-chromosome haplogroup L1a-M27/P329. The M27 mutation is common in South Asian haplogroup L Y-chromosomes[1,2], but was absent in a survey of Y-chromosomes from Anatolia[3]. Haplogroup L occurs at a very low ~2% frequency in present-day Armenians[4]. Our results indicate that it was present in Chalcolithic Armenia, but the fact that all three Chalcolithic Armenians belonged to it should not be necessarily interpreted as evidence that it was common there, as our samples are from a single location (Areni-1 cave) and may represent a local founder population.

I1407: L1a

This individual was derived for mutations P329, M27, M76 defining haplogroup L1a, as well as upstream mutations defining haplogroup L1 (L656, L1304) and L (L863, L878, M61). No calls were made for SNPs downstream of L1a, thus it is placed in haplogroup L1a.

I1632: L1a

This individual was derived for mutations P329, M27, M76 defining haplogroup L1a, as well as upstream mutations defining haplogroup L1 (L656, M22, L1304) and L (L878, M185, M61). No calls were made for SNPs downstream of L1a, thus it is placed in haplogroup L1a.

I1634: L1a

This individual was derived for mutations P329, M27 defining haplogroup L1a, as well as an upstream mutation defining haplogroup L1 (L656) and L (L878, M185, M11, M61, L855). No calls were made for SNPs downstream of L1a, thus it is placed in haplogroup L1a.

Armenia_EBA

I1635: R1b1

This individual was derived for mutations M415, L278 defining haplogroup R1b1, and upstream mutation defining haplogroup R1b (M343). It was ancestral for a downstream mutations defining haplogroup R1b1a1 (M478), several mutations defining haplogroup R1b1a2 (CTS12478, CTS3575, CTS8728, M269, PF6399, PF6430, PF6475, PF6482, PF6494, PF6497, PF6505), including the M269 mutation that is found in the great majority of present-day R1b Y-chromosomes[5]. It was also ancestral for haplogroup R1b1c2 (V35) and R1b1c3 (V69). Thus, it could be designated as R1b1x(R1b1a1, R1b1a2, R1b1c2, R1b1c3).

R1b1 not belonging to the dominant M269 subclade has been found in several early populations, including an Eastern European hunter-gatherer from Samara[6] on the Eurasian steppe, an Eneolithic individual from Samara[7], an Early Neolithic farmer from Spain[6] and the

~14,000 year old "Villabruna" hunter-gatherer from Italy[8]. All these samples are genetically differentiated based on analysis of their autosomal DNA: Villabruna belonged to the genetic cluster named after it which also includes Western European hunter-gatherers (WHG)[8], the samples from Samara were predominantly Eastern European hunter-gatherers (EHG)[6,7] with some admixture from the south during the Eneolithic, and the farmer from Spain belonged to the widespread Early European Farmer population[6]. Whatever the ultimate origin of R1b1, it became geographically widespread and associated with genetically differentiated autosomal profiles prior the steppe migration that coincided with the dispersal of M269 into mainland Europe[6,9]. The finding of R1b1 in Early Bronze Age Armenia may reflect either migration from the steppe or an earlier presence of this haplogroup in the area as part of the broader early geographic distribution in Eurasia outlined above. We note that unlike Europe[6], the appearance of R1b in Early Bronze Age Armenia does not correlate with steppe ancestry (f_4(Armenia_ChL, Armenia_EBA; Steppe_EMBA, Chimp)= 0.00059 (Z=2.3) compared to f_4(Europe_MNChL, Europe_LNBA; Steppe_EMBA, Chimp)= -0.00322 (Z=-25.1)). Thus, there is no evidence for steppe influx into Armenia between the time of Chalcolithic (~4,400-3,700 BCE) and the date of the I1635 individual (~2,600-2,500 BCE). The question of the timing and direction of gene flow may be resolved by obtaining earlier samples from Armenia and geographically intermediate areas between it and the Eurasian steppe.

Iran_HotuIIIb

I1293: J(xJ2a1b3, J2b2a1a1)

This individual belonged to haplogroup J, supported by mutations PF4519, FGC3271, PF4530, CTS5934, F2839, PF4619. It was found to be ancestral for L227 (J2a1b3) and Z639 (J2b2a1a1) and could be designated as J(xJ2a1b3, J2b2a1a1).

Iran_N

I1945: P1(xQ, R1b1a2, R1a1a1b1a1b, R1a1a1b1a3a, R1a1a1b2a2a)

This individual belonged to haplogroup P1 on the basis of mutation P282. It was ancestral for downstream haplogroups Q (F1237.1, FGC4603), R1b1a2 (CTS12478), R1a1a1b1a1b (CTS11962), R1a1a1b1a3a (L448), and R1a1a1b2a2a (Z2123). Thus, it could be designated P1(xQ, R1b1a2, R1a1a1b1a1b, R1a1a1b1a3a, R1a1a1b2a2a).

I1949: CT

This individual belonged to haplogroup CT, supported by mutations M5593, PF228, M5624, PF342, Z17710, CTS2842, CTS5532, M5730, M5751, M5765, CTS11358.

Iran_LN

I1671: G2a1(xG2a1a)

This individual belonged to haplogroup G2a1, supported by mutations FGC666, FGC587, FGC7537, FGC592, FGC7533, FGC593, FGC594, FGC7536, FGC600, FGC602, FGC606, FGC607, FGC610, FGC617, FGC618, FGC7543, FGC7547, FGC631, FGC7546, FGC635, FGC637, FGC639, FGC641. It was ancestral for mutations FGC703, FGC741, FGC762 defining haplogroup G2a1a. Thus, it could be designated G2a1(xG2a1a). Haplogroup G2a2 was common in Anatolian Neolithic[10] and Early European[6] farmers. Thus, the parent haplogroup G2a could be a genetic link between these Neolithic populations and those of the Iranian plateau, a link also supported by the statistic f_4(Iran_LN, Iran_N; Anatolia_N, Chimp) = 0.00213 (Z=4.3) which suggests that the Anatolian Neolithic shares more alleles with Late Neolithic Iran from Seh Gabi rather than with the Early Neolithic Iran from Ganj Dareh.

Iran_ChL (Chalcolithic Iran)

I1662: J(xJ1a, J2a1, J2b)

This individual was derived for mutations PF4505, CTS1250, PF4513, F1181, PF4519, PF4530, L60, F1744, CTS2769, F2116, CTS5628, CTS7229, PF4567, PF4572, F2746, F2769, CTS8974, CTS9877, CTS10446, F3119, PF4591, F3176, PF4594, FGC1599, L778, CTS11571, CTS11750 placing it in haplogroup J. It was ancestral for haplogroups J1a (CTS5368), J2a1 (L26), and J2b (M314). Thus, it is placed in haplogroup J(xJ1a, J2a1, J2b).

I1674: G1a(xG1a1)

This individual was derived for mutations F4113, F1761 placing it in haplogroup G1a, and upstream mutations Z17861, Z17868, Z17869, M285, M342 placing it in haplogroup G1. It was ancestral for downstream mutation L1324 (G1a1), and could be designated as G1a(xG1a1). A correspondence between haplogroup G1 and ancient Iranian speakers of both the Iranian plateau and steppe area has recently been proposed[11], with a suggestion based on STR variation of a likely homeland in the Iranian Plateau and the Armenian Highlands. While our data do not address this question directly, it does establish the presence of G1a in Iran as early as the Chalcolithic period (~4,500-3,500 BCE) prior to the formation and expansion of steppe pastoralist populations like the Yamnaya[12], making it less likely that it was introduced into the Iranian plateau by steppe migrants.

Natufians

I0861: E1b1b1b2(x E1b1b1b2a, E1b1b1b2b)

This individual was derived for CTS11781 placing it in haplogroup E1b1b1b2 mutations M5108, CTS3637, CTS7154, PF1755, L796 defining haplogroup E1b1b1 and L336 (E1b1b).

It was ancestral for L857 (E1b1b1b2a) and Z865 (E1b1b1b2b) and could be designated E1b1b1b2(x E1b1b1b2a, E1b1b1b2b).

I1069: E1b1(xE1b1a1, E1b1b1b1)

This individual was derived for mutation P179 defining haplogroup E1b1 and upstream mutation M5403 defining haplogroup E. It was ancestral for Z1116 (E1b1a1), CTS8649 (E1b1b1b1) and could be designated as E1b1(xE1b1a1, E1b1b1b1).

I1072: E1b1b1b2(xE1b1b1b2a, E1b1b1b2b)

This individual was derived for mutations CTS8182, CTS11781 defining haplogroup E1b1b1b2 and upstream mutations L117, PF1755, CTS9324, L796 defining haplogroup E1b1b1. It was ancestral for CTS1652 (E1b1b1b2a) and CTS11051, CTS11574 (E1b1b1b2b), and could be designated as E1b1b1b2(xE1b1b1b2a, E1b1b1b2b).

I1685: CT

This could only be assigned to haplogroup CT on the basis of mutations CTS9555, Y1580.

I1690: CT

This could only be assigned to haplogroup CT on the basis of mutations Y1462, M5723, L977.

Levant_N

I0867: H2 (PPNB)

This individual was derived for mutation P96 defining haplogroup H2 as well as upstream mutations M2713, M2896, M2936, M2942, M2992, M3070 defining haplogroup H. It was not **derived** for any downstream mutations so it could be designated simply as H2. Haplogroup H2 has also been found in Neolithic farmers from Anatolia and Europe[6,10], and thus represents a common genetic signature of several ancient Neolithic farmer groups.

I1414: E(xE2, E1a, E1b1a1a1c2c3b1, E1b1b1b1a1, E1b1b1b2b) (PPNB)

This individual was derived for mutation CTS2893 defining haplogroup E and also the upstream mutation P167 defining haplogroup DE. It was ancestral for haplogroup E1a (Z15455, Z912, CTS3507, CTS11248), E1b1a1a1c2c3b1 (Z16129, Z16130), E1b1b1b1a1 (CTS10196), E1b1b1b2b (M293), and E2 (CTS11446, CTS11447). Thus it could be designated E(xE2, E1a, E1b1a1a1c2c3b1, E1b1b1b1a1, E1b1b1b2b).

I1415: E1b1b1 (PPNB)

This individual was derived for mutation CTS2216 defining haplogroup E1b1b1.

I1416: CT (PPNB)

This individual could only be assigned to haplogroup CT on the basis of mutations CTS7933, M5786.

I1707: T(xT1a1, T1a2a) (PPNB)

This individual was derived for mutations PF7466, CTS7263, CTS10416 defining haplogroup T. It was ancestral for FGC3945.2 (T1a1) and P322 (T1a2a). Thus, it could be designated T(xT1a1, T1a2a). It has been suggested that haplogroup T first began to diversify in the Near East[13] and our results document that it was present there in some of the earliest Neolithic communities of the Near East, providing a plausible source for its appearance in the Early Neolithic of central Europe[6].

I1710: E1b1b1(x E1b1b1b1a1, E1b1b1a1b1, E1b1b1a1b2, E1b1b1b2a1c) (PPNB)

This individual was derived for mutation M5041 defining haplogroup E1b1b1, and upstream mutations CTS8479.1 (E1b1b), and E1 (CTS9753). It was ancestral for E1b1b1b1a1 (CTS5819), E1b1b1a1b1 (L618), E1b1b1a1b2 (CTS5479), and E1b1b1b2a1c (V23). Thus, it could be designated E1b1b1(x E1b1b1b1a1, E1b1b1a1b1, E1b1b1a1b2, E1b1b1b2a1c).

I1727: CT(xE, G, J, LT, R, Q1a, Q1b) (PPNB)

This individual was derived for mutations M5723, CTS7922, M5769, M5822, M5823 defining haplogroup CT. It was ancestral for haplogroup E (M5452), G (L770, M3479, PF2918, CTS2125, CTS2271), J (F1167, CTS7229, CTS9533), LT (L298), R (P285), Q1a (CTS5301), Q1b (Y1080, Y1996), and thus could be designated CT(xE, G, J, LT, R, Q1a, Q1b).

I1700: CT (PPNC)

This individual could only by assigned to haplogroup CT on the basis of mutations M5603, M5624, CTS3460, M5822.

Levant_BA

I1705: J1(xJ1a)

This individual belonged to haplogroup J1, supported by mutation M267. It was ancestral for haplogroup J1a (M365.1) and could be designated J1(xJ1a). Haplogroup J1 is common in present-day Levantine populations[14] and our result documents its earliest existence in the Bronze Age Levant.

I1730: J(xJ1, J2a, J2b2a).

This individual also belonged to haplogroup J, supported by mutations CTS687, CTS1250, PF4513, F1167, F1181, PF4519, PF4521, PF4524, FGC3271, PF4530, F1826, CTS4349, CTS5280, F2114, F2116, CTS5628, CTS5678, CTS7229, F2502, CTS7738, F2769, CTS8974, F2817, F2839, CTS10446, F3119, F4299, S22619, PF4591, F3176, FGC1599, YSC0000228, CTS10858, CTS11291, L778, CTS11571, CTS11750, PF4619, CTS12047. Unlike, I1705, it was ancestral for haplogroup J1 (M267). It was also ancestral for haplogroup J2a (L152, L212) and J2b2a (L283). Thus, it could be designated J(xJ1, J2a, J2b2a).

Discussion

We highlight two aspects of the Y-chromosome distribution in the ancient Near East.

First, the Mesolithic individual from Iran belonged to haplogroup J. This has also been detected in two hunter-gatherers from the Upper Paleolithic in Georgia[15], as well as in a hunter-gatherer from Karelia in northwest Russia[10], suggesting that it had a widespread early distribution prior to the spread of farming with which its current distribution was initially associated[16]. While the hunter-gatherers from Georgia resemble the one from Iran (Fig. 1b), their whole genome data shows very different patterns from the Eastern European hunter-gatherer who also possessed this haplogroup, as well as from a singleton Anatolian early farmer who belonged to haplogroup J2a[10]. This should serve as a note of caution against the idea that Y-chromosome lineages can be thought as markers of populations and population movements: the two sometimes coincide as with the sudden increase of haplogroups R1a and R1b in mainland Europe coinciding with an influx of steppe migrants around ~4,500 years ago[6], or the spread of migrants carrying haplogroup G2a2 during the early Neolithic settlement of Europe[10]. With this caveat in mind, it is nonetheless interesting that haplogroup J was present in EHG, CHG, and Mesolithic Iran, given the evidence for a relationship between these populations presented in Supplementary Information, sections 4, 7. Additional sampling of earlier individuals from eastern Europe and western Asia may elucidate the nature of this relationship. We also note one instance of a possible relationship between haplogroup J and the appearance of possible admixture event: two of the Bronze Age samples from the Levant belong to this haplogroup, while none of the earlier samples of Natufians and Pre-Pottery Neolithic individuals do. In Supplementary Information section 7, we infer that Levant_BA can be modeled as a mixture of Levant_N and Iran_ChL. The Y chromosome results suggest that haplogroup J migrants from the north of the Levant may have been involved in this mixture event.

Second, we observed that all three Natufian individuals that could be assigned to a specific haplogroup belonged to haplogroup E1b1. This is thought to have an East African origin, and a 4,500-year old individual from the Ethiopian highlands[17] belonged to it. An African origin of the Natufians has been proposed based on their 'Sub-Saharan' cranial morphology[18] in comparison to later West Eurasian samples. However, when we test whether present-day Sub-Saharan Africans share more alleles with Natufians than they do with other ancient West Eurasian individuals, we do not find the expected asymmetry if Natufians have ancestry related to present-day Sub-Saharan Africans (Extended Data Table 1). The existence of haplogroup E1b1 in ancient East Africa and the nearby Levant in the two earliest samples from both regions raises questions about its ultimate origin. This haplogroup continues to exist in the Pre-Pottery Neolithic period, but it is not documented in ancient Europe or the ancient Near East outside the Levant. This appears to be consistent with E1b1 being intrusive to the Near East from nearby Africa, rather than its having a long-term presence in the region. An association of E1b1 migrants from Africa with Basal Eurasian ancestry is possible (Basal Eurasians are so named because of their phylogenetic position basal to other Eurasians not for their geographical provenance; African migrants that did not participate in the initial Out-of-Africa expansion would occupy such a basal position to other Eurasians). However, such ancestry is not limited to the Levant, but also extends to the whole of Near East (where E1b1 chromosomes have not been detected). Thus, we think that both the late entry scenario of Basal Eurasian ancestry into the Near East (associated with gene flow from Africa), or its earlier presence[19] in anatomically modern-humans from the Levant and Arabia >100,000 years [20,21] are still plausible.

While some Y-chromosomal lineages (such as H2, T, and G2a) span more than one early Neolithic population in West Eurasia, none of them are found in all of them (Levant, Iran, and Northwestern Anatolia/Europe), in agreement with the conclusion based on the analysis of autosomal data that the Neolithic of West Eurasia either began (or was taken up soon after its beginning) by genetically diverse populations.

References

1 Thanseem, I. *et al.* Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. *BMC Genetics* **7**, 42 (2006).
2 Sengupta, S. *et al.* Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet* **78**, 202-221 (2006).
3 Cinnioğlu, C. *et al.* Excavating Y-chromosome haplotype strata in Anatolia. *Human Genetics* **114**, 127-148 (2004).
4 Herrera, K. J. *et al.* Neolithic patrilineal signals indicate that the Armenian plateau was repopulated by agriculturalists. *Eur J Hum Genet* **20**, 313-320 (2012).
5 Myres, N. M. *et al.* A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur J Hum Genet* **19**, 95-101 (2011).
6 Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211 (2015).
7 Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503 (2015).
8 Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205 (2016).
9 Batini, C. *et al.* Large-scale recent expansion of European patrilineages shown by population resequencing. *Nat Commun* **6** (2015).
10 Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **(in press)** (2015).
11 Balanovsky, O. *et al.* Deep Phylogenetic Analysis of Haplogroup G1 Provides Estimates of SNP and STR Mutation Rates on the Human Y-Chromosome and Reveals Migrations of Iranic Speakers. *PLoS ONE* **10**, e0122968 (2015).
12 Anthony, D. W. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. (Princeton University Press, 2007).
13 Mendez, F. L. *et al.* Increased Resolution of Y Chromosome Haplogroup T Defines Relationships among Populations of the Near East, Europe, and Africa. *Human Biology* **83**, 39-53 (2011).
14 Semino, O. *et al.* Origin, Diffusion, and Differentiation of Y-Chromosome Haplogroups E and J: Inferences on the Neolithization of Europe and Later Migratory Events in the Mediterranean Area. *American journal of human genetics* **74**, 1023-1034 (2004).
15 Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun* **6** (2015).
16 Rosser, Z. H. *et al.* Y-Chromosomal Diversity in Europe Is Clinal and Influenced Primarily by Geography, Rather than by Language. *American Journal of Human Genetics* **67**, 1526-1543 (2000).
17 Llorente, M. G. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* **350**, 820-822 (2015).
18 Brace, C. L. *et al.* The questionable contribution of the Neolithic and the Bronze Age to European craniofacial form. *Proc Natl Acad Sci U S A* **103**, 242-247 (2006).
19 Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413 (2014).
20 Bar-Yosef, O. *The chronology of the Middle Paleolithic of the Levant*. 39-56 (New York: Plenum Press, 1998).
21 Rose, J. I. *et al.* The Nubian Complex of Dhofar, Oman: an African middle stone age industry in Southern Arabia. *PLoS One* **6**, e28239 (2011).

Supplementary Information 7

Admixture history of ancient west Eurasians

In this section we study the admixture history of ancient west Eurasian populations using methods developed in ref. 1. The idea of these methods is to relate a *Test* population (whose history of admixture we are interested in studying) to a set of *Outgroup* populations (who are populations unlikely to have contributed to *Test* directly) via a set of *Reference* populations (who are clades – with respect to the *Outgroups* – of populations contributing ancestry to *Test*). Given this setup, we can write f_4-statistics of the form $f_4(Test, O_1; O_2, O_3)$ as a weighted sum of the statistics of the *Reference* populations:

$$f_4(Test, O_1; O_2, O_3) = \sum_{i=1}^{N} \alpha_i f_4(Ref_i, O_1; O_2, O_3)$$

Given *m Outgroups*, there are $m\binom{m-1}{2}$ equations of the above form, which allow us to fit the mixture coefficients by regression. A related formulation uses statistics of the form $f_4(Test, Ref_i; O_2, O_3)$. This is based on the observation that $f_4(Test, Ref_i; O_2, O_3) = f_4(Test, O_1; O_2, O_3) - f_4(Ref_i, O_1; O_2, O_3)$, which leads to a restatement of the admixture equation as:

$$\sum_{i=1}^{N} \alpha_i f_4(Test, Ref_i; O_2, O_3) = 0$$

There are $\binom{m}{2}$ equations of the above form that again allow us to fit the mixture coefficients. We do not repeat the theory here, except to say that if the *References* vary in their relationship to the *Outgroups*, it is possible to extract the mixture coefficients using this methodology. There are two main advantages: (i) First, these statistics are not influenced by genetic drift specific to either the *Test* or *Reference* populations, but only by genetic drift shared by the *Outgroups* and *References* (and through the *References* with *Test*). Thus, we do not need to have access to data from the exact population that participated in the mixing, and instead can use as a *Reference* a population that is genetically very drifted from it (albeit phylogenetically a clade with it). (ii) Second, the phylogeny relating the various *Outgroups* does not need to be known explicitly, and hence we can leverage genetic drift shared with populations whose relationships may be complicated without having to attempt to model those relationships explicitly. The fact that we do not need to determine the phylogenetic relationships among all the populations used in the analysis makes our inference procedure more robust.

This methodology has been implemented in the *qpWave* and *qpAdm* programs (https://github.com/DReichLab). *qpWave* tests whether a set of *Left* populations is consistent

with being descended from as few as *N* source populations (that is, the pattern of shared genetic drift with the *Right* populations can be entirely explained as due to *N* source populations). Typically, the *Right* population set is the set of *Outgroups*. The *Left* population set is either the set of *References*, allowing us to check whether *N References* can be distinguished by the set of *Outgroups*, or the combined set of the *Test* population with the *References*. This setup allows us to check whether the addition of the *Test* population increases the necessary number of source populations, which in turn indicates that the *Test* population cannot be well-modeled as an *N*-way mixture of populations that are clades with the *References*. In the discussion that follows, we will refer to the rank of the matrix:

$$X(u, v) = F_4(u_0, u; v_0, v)$$

Where *u* is a "*Left*" and *v* is a "*Right*" population and u_0,v_0 are basis populations from the *Left* and *Right* sets. It has been shown[2] that this matrix has rank *N*-1 if there are *N* waves of migration. Having verified that a population can be modeled as an *N*-way mixture, we can use *qpAdm* to compute its mixture coefficients.

We used a basic set of 9 *Outgroups*, including diverse non-west Eurasian present-day populations from Africa, East Asia, Oceania, the Andaman Islands, North Asia, and the Americas, as well as the Upper Paleolithic genomes of Ust_Ishim[3] from Siberia ~45,000 years ago, Kostenki14 from Russia ~37,000 years ago[4], and MA1 from Siberia ~24,000 years ago.[5] These samples reside in informative places of the phylogeny: Ust_Ishim is symmetrically related to European hunter-gatherers and eastern non-Africans[1,3]; Kostenki14 is the oldest sample that shares more alleles with recent west Eurasians than eastern non-Africans[4]; MA1 shares genetic drift with recent west Eurasians but also with native Americans as it is related to populations that admixed with both[5,6]. We do not use the Upper Paleolithic genome of Oase1 from Romania dating to ~40,000 years ago[7] due to its European contamination and high Neandertal admixture.

O9: Ust_Ishim, Kostenki14, MA1, Han, Papuan, Onge, Chukchi, Karitiana, Mbuti

Our goal in this section is to quantify admixture in ancient west Eurasian populations. Before we begin, two caveats are necessary:

First, we are dealing with samples from particular places and times, and our genetic models investigate relationships among them. As a result, when we derive a population's ancestry from Iran_N, we do not necessarily postulate a migration from present-day Iran, but rather admixture with a population related, perhaps anciently, to Iran_N. The geographic distribution of populations like Iran_N (or indeed any other) is currently unknown.

Second, we are likely missing pieces of the puzzle, and we make our best attempt to fit the existing pieces into a coherent picture. For example, previous work has modeled Europeans as mixtures of either farmers and hunter-gatherers from Sweden[8], or early farmers/present-day Sardinians and an 'Ancient North Eurasian' (ANE) population forming a clade with Native Americans[9,10]. This was amended by the publication of data from Upper Paleolithic Siberians[5] which showed that Native Americans were a mixture of Upper Paleolithic Siberians and eastern non-Africans[5], and of the western European hunter-gatherers[6] from Luxembourg[6] and Iberia[11,12] which allowed the inference that present-day Europeans had at least three ancestral populations[6] related to the early European farmers (EEF), Upper Paleolithic Siberians (ANE), and western European hunter-gatherers (WHG). The ANE component was absent in mainland Europe before and during the Early Neolithic[6] except in Scandinavian hunter-gatherers[6,13], suggesting that it became pervasive in Europe after the arrival of the early farmers[6]. A more proximate source and a better understanding of the arrival of ANE were made possible by the sampling of the eastern European hunter-gatherers (EHG)[1] and the Yamnaya steppe pastoralists[1,14]. Similarly, in a previous study we estimated Near Eastern admixture in early European farmers in a roundabout way using present-day Near Eastern populations[6], and in Yamnaya steppe pastoralists using present-day Armenians[1]. We are now in a position to study the history of admixture in west Eurasia using a more extensive collection of populations from both Europe and the Near East. Despite the far better population coverage, some pieces may still be missing, such as the early farmers from the northern Levant, Mesopotamia, Central and Eastern Anatolia, the Arabian Peninsula, and the southern and northern Caucasus which may differ from those in our study.

Levant_N, Iran_N, WHG, EHG (the "Four") are related by four streams of ancestry to the outgroups

The two Neolithic populations from the ancient Near East (Levant_N and Iran_N) and the two Mesolithic/Neolithic hunter-gatherer populations from Europe (WHG and EHG) are maximally differentiated in the PCA (Fig. 1b) We use Left=Four and Right=O9.

Table S7.1: The "Four" can be distinguished by the O9 outgroups

Rank	d.o.f.	χ^2	P
0	24	1343.399	6.16E-269
1	14	246.287	1.68E-44
2	6	27.132	1.37E-04

Rank=2 can be rejected (P=0.000137), thus there is at least rank=3 or 4 streams of ancestry relating the Four to the O9 outgroups. This is a useful check to show that the reference populations can be distinguished by the outgroups.

Ancient West Eurasian populations other than the "Four" do not require an additional stream of ancestry from the outgroups

We next add each ancient West Eurasian *Test* population to the set of "Four". If rank=3 can be excluded, this means that at least 5 streams of ancestry relate the quintuple (*Test*, Levant_N, Iran_N, WHG, EHG) to the O9. However, with the exception of the Steppe_IA population, we cannot exclude rank=3 (Table S7.2).

Table S7.3: Testing for rank=3 for the Left=(*Test*, Levant_N, Iran_N, WHG, EHG) and Right=O9.

		Proportions				Standard Errors			
Test	**P**	**Levant_N**	**Iran_N**	**WHG**	**EHG**	**Levant_N**	**Iran_N**	**WHG**	**EHG**
Anatolia_ChL	2.36E-01	0.649	-0.016	0.277	0.090	0.412	0.438	0.095	0.062
Anatolia_N	4.63E-01	0.331	0.331	0.407	-0.069	0.108	0.114	0.045	0.027
Armenia_ChL	9.54E-01	0.145	0.437	0.243	0.175	0.110	0.117	0.057	0.034
Armenia_EBA	5.16E-02	0.029	0.704	0.193	0.074	0.203	0.214	0.078	0.046
Armenia_MLBA	1.44E-01	-0.187	0.913	0.087	0.187	0.238	0.247	0.098	0.055
CHG	4.59E-01	0.037	0.700	0.159	0.105	0.189	0.204	0.082	0.045
Europe_EN	6.23E-01	0.176	0.455	0.431	-0.062	0.110	0.117	0.050	0.029
Europe_LNBA	2.85E-01	-0.089	0.532	0.317	0.240	0.121	0.130	0.052	0.031
Europe_MNChL	3.91E-01	0.168	0.365	0.520	-0.053	0.127	0.138	0.047	0.028
Iberia_BA	6.85E-01	0.474	0.066	0.326	0.135	0.272	0.297	0.097	0.052
Iran_ChL	5.11E-02	-0.011	0.841	0.114	0.055	0.218	0.229	0.082	0.047
Iran_LN	9.73E-01	0.095	0.880	0.105	-0.080	0.220	0.224	0.106	0.060
Iran_HotuIIIb	7.70E-01	-0.539	1.384	0.085	0.070	0.493	0.474	0.252	0.147
Iran_recent	3.91E-01	-0.258	1.031	0.201	0.026	0.269	0.284	0.129	0.071
Levant_BA	7.96E-01	0.583	0.301	0.134	-0.018	0.132	0.142	0.061	0.034
Natufian	1.35E-01	1.201	-0.094	-0.024	-0.083	0.493	0.482	0.153	0.096
SHG	2.74E-01	-0.008	0.090	0.479	0.439	0.163	0.167	0.057	0.037
Steppe_EMBA	3.97E-01	-0.236	0.597	0.131	0.508	0.139	0.145	0.061	0.038
Steppe_Eneolithic	5.74E-02	-0.430	0.635	0.069	0.727	0.327	0.338	0.104	0.072
Steppe_IA	2.16E-02	-1.232	1.588	0.374	0.270	0.806	0.876	0.199	0.117
Steppe_MLBA	1.96E-01	-0.115	0.530	0.251	0.334	0.136	0.144	0.055	0.034
Switzerland_HG	5.35E-01	0.311	-0.296	1.009	-0.023	0.237	0.254	0.087	0.055

This motivates us to attempt to derive ancient West Eurasian populations from the "Four". Note that while we can estimate mixture proportions (Table S7.3) in terms of the "Four" using the O9 set of outgroups, these are very "noisy" estimates, with high standard errors and infeasible negative estimated mixture proportions. Some of the *Test* populations may be

simpler than 4-way mixtures, and the mixture proportions of some may be clarified by introducing additional outgroups to the Right set that can tease out better differences between the reference populations.

The Natufians and the origin of the Neolithic in the Levant

When modeling different west Eurasian populations (Table S7.3) we infer that the Natufians have ancestry (>100%) from the Levant_N, suggesting a relationship between them and the Neolithic populations of the Pre-Pottery Neolithic that succeeded them. To better understand this relationship, we studied f_4-statistics of the form f_4(Natufian, Levant_N; *Test*, Chimp), f_4(*Test*, Natufian; Levant_N, Chimp), and f_4(*Test*, Levant_N; Natufian, Chimp) (Fig. S7.1-3).

Figure S7.1: Testing for symmetry between the Natufians and the Neolithic of the Levant with the statistic f_4(Natufian, Levant_N; *Test*, Chimp). In this and following figures, the estimate and ±3 standard errors is shown.

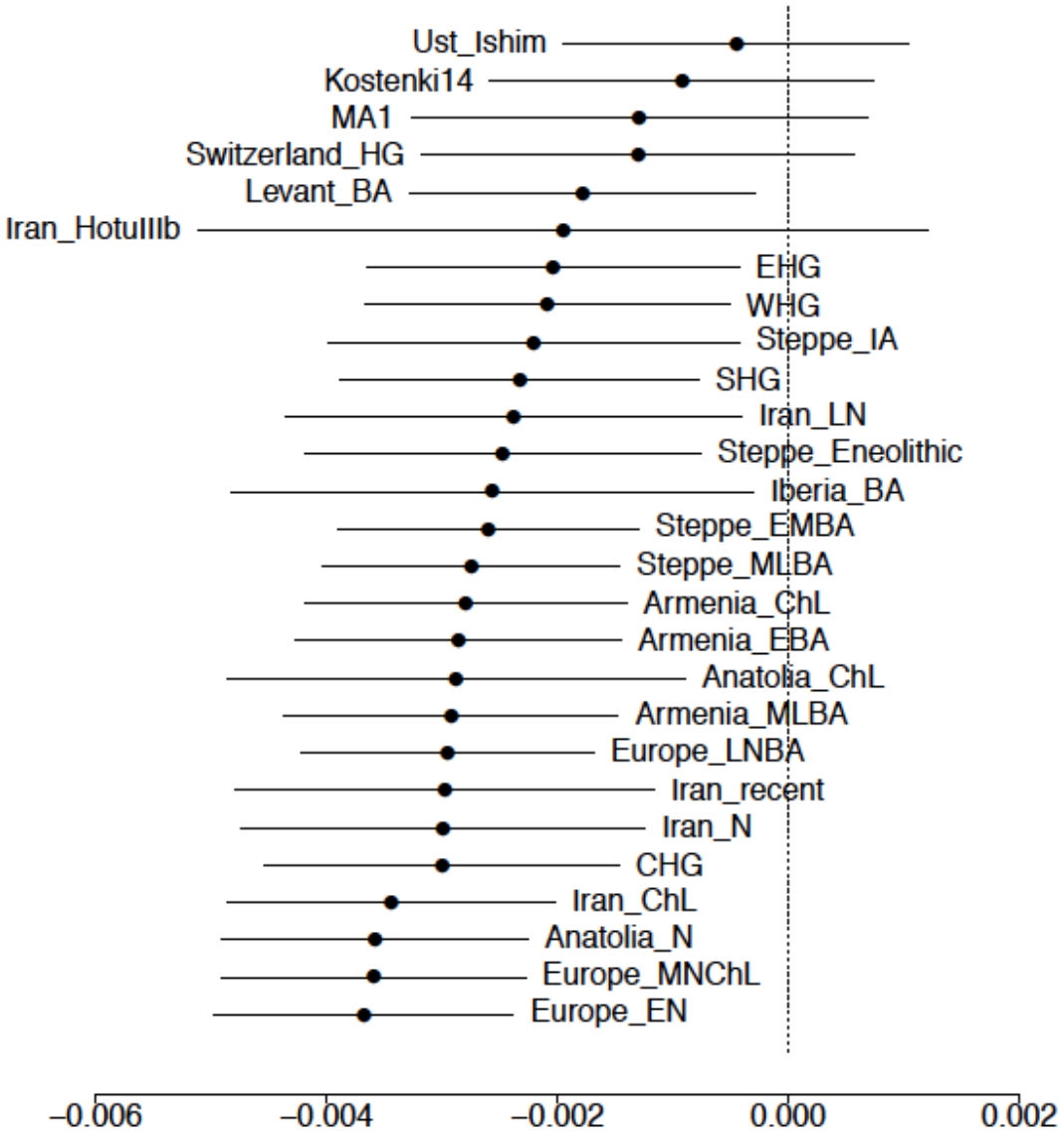


Fig. S7.1 shows that Natufians and the Neolithic of the Levant are asymmetric to ancient West Eurasian populations, with the Levantine Neolithic sharing more alleles with them than the Natufians did. Some of this asymmetry can be interpreted as gene flow from the Near East beginning in the Pre_Pottery Neolithic which begun ~3000 years prior to the Pottery Neolithic populations in Western Anatolia, which dispersed into Europe. The Pre-Pottery Neolithic is older than most other populations and migrations from the Near East and may have contributed ancestry from it to later populations of West Eurasia) However, this does not account for the fact that the three pre-Neolithic hunter-gatherer groups from Europe (WHG, EHG, SHG) and the one from the Caucasus (CHG) also share more alleles with the Neolithic of the Levant than with the Natufians (the value of the statistic is not significant for the HotuIIIb individual of likely Mesolithic date from Iran, however this is a singleton individual with fewer covered SNPs than the others).

A possible explanation for this phenomenon is an excess of Basal Eurasian ancestry in Natufians, as this would dilute their affinity to other populations, including the different pre-Neolithic Eurasians. However, several pieces of evidence argue against this interpretation. First, the statistic f_4(Natufian, Levant_N; Ust_Ishim, Chimp) = 0.00045 (Z=0.9) which does not provide evidence for asymmetry between Neolithic of the Levant and the Natufians with respect to Ust_Ishim as would be expected for a population with substantial extra Basal Eurasian ancestry. Second, if an excess of Basal Eurasian ancestry in Natufians were responsible for this phenomenon, the observed asymmetry would also be observed for older Upper Paleolithic Eurasians such as Ust_Ishim and Kostenki14, but it does not (Fig. S7.1). Third, the Levantine Neolithic does show evidence of admixture in terms of a negative f_3-statistic involving the Natufians and Neolithic Europeans and Anatolians (Table S7.4), suggesting that they are in fact admixed, and thus could share genetic drift with them via the non-Natufian portion of their ancestry.

Table S7.4: Significantly negative statistics of the form f_3(Levant_N; Natufian, *Test*).

Test	f_3(Levant_N; Natufian, *Test*)	Z	Number of SNPs
Europe_EN	0.00140	-3.6	122590
Anatolia_N	0.00142	-3.3	122480
Europe_MNChL	0.00145	-3.0	121552

Figure S7.2: Testing whether the Neolithic of the Levant shares more alleles with the Natufians or with another *Test* population with the statistic f_4(*Test*, Natufian; Levant_N, Chimp).

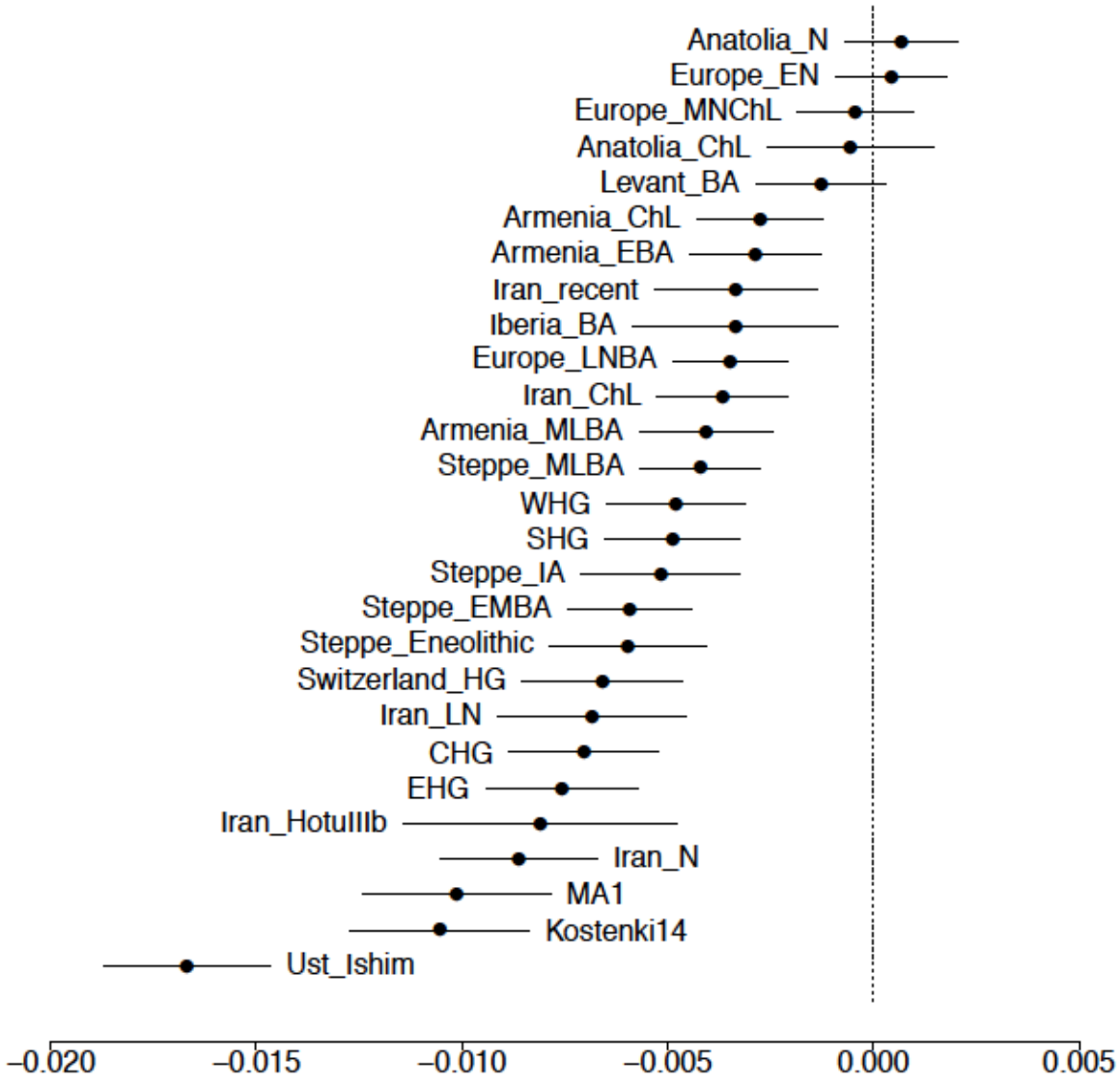


Figure S7.3: Testing whether the Natufians share more alleles with the Neolithic of the Levant or with another *Test* population with the statistic f_4(*Test*, Levant_N; Natufian, Chimp).

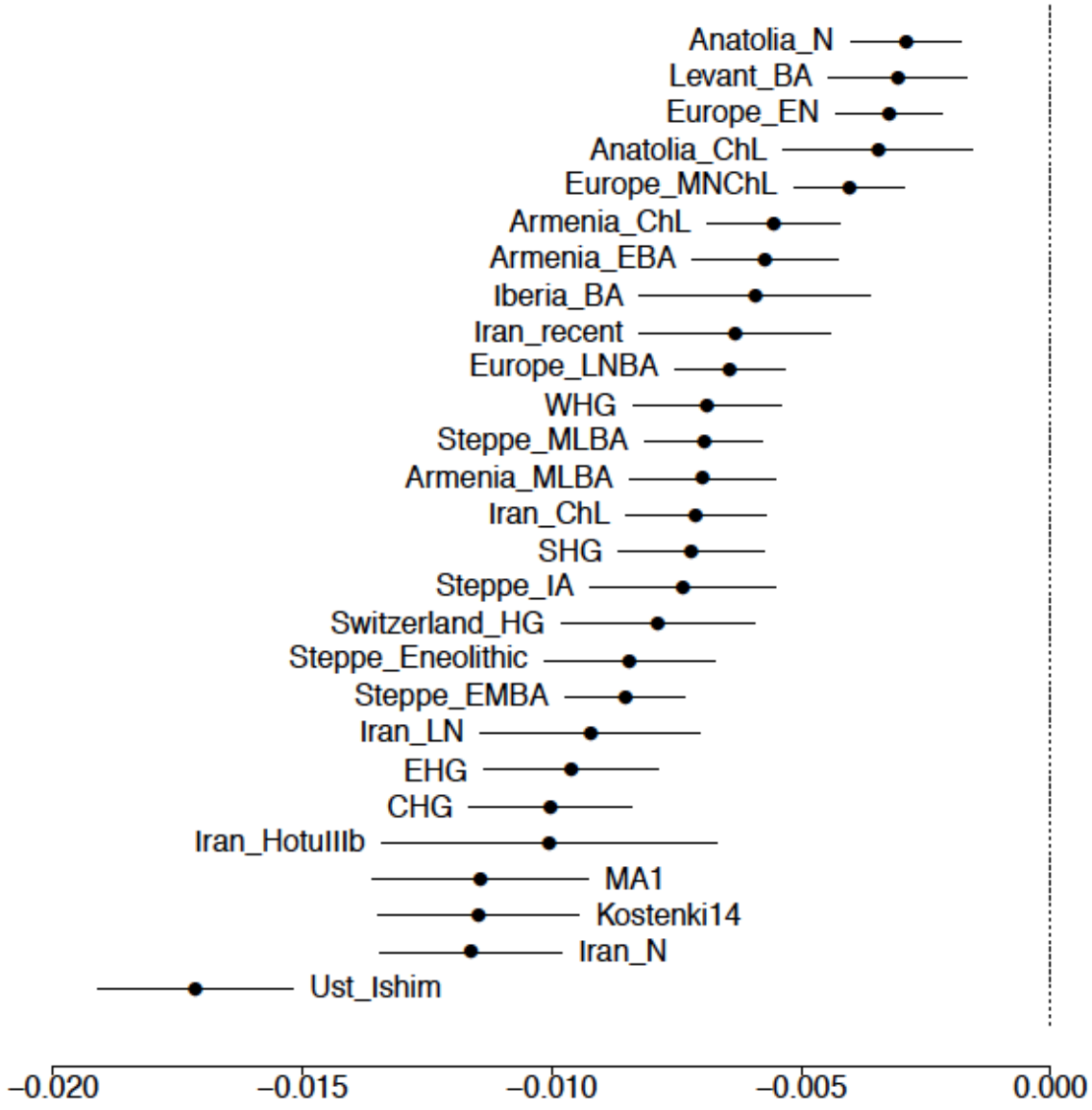


Fig. S7.2 shows that the Neolithic of the Levant shares more alleles with Natufians than all but five other ancient populations. These include the later Levantine Bronze Age population, but also populations from Anatolia and Neolithic Europe. The close relationship of the Anatolian and European Neolithic is established[14,15], and our results suggest that the Anatolian could be related to the Levantine Neolithic as shared drift between the two would result in the statistic being higher. Nonetheless, it does not become significantly positive for any ancient West Eurasian population, and the Levantine Neolithic shares equally or less with the Natufians than with any other ancient West Eurasian population. Fig. S7.3 shows that the Natufians share more alleles with the Levantine

Neolithic than with any other ancient West Eurasian population, consistent with its admixture estimate (Table S7.3) and their clustering (Fig. 1b) with that population.

To understand the origin of the Levantine Neolithic we pair it with Natufians and a population *A* and test Left=(Levant_N, Natufian, *A*). Before we proceed, we introduce a shorthand notation to denote additional populations added to the set of Right outgroups in Table S7.4. This will be used in the rest of this Supplementary section.

Table S7.4: Shorthand notation for adding outgroups to the Right set. For example, O9CEIW is the O9 set with the addition of CHG, EHG, Iran_N, and WHG.

Code	Population
A	Anatolia_N
C	CHG
E	EHG
I	Iran_N
L	Levant_N
N	Natufian
W	WHG
S	Switzerland_HG

In Table S7.5 we show the results of modeling Levant_N as a mix of Natufians and a *Test* population which we choose to be one of Anatolia_N, CHG, and Iran_N, i.e., the earliest population from the other three Near Eastern regions of Anatolia, Caucasus, and Iran (Fig. 1a). When using Right=O9 we observe that we cannot reject rank=1 for any of the chosen *A* populations; this likely reflects the fact that they all share more alleles with Levant_N than with Natufians (Fig. S7.1). By introducing additional outgroups to the Right set we are better able to distinguish between the different *Test* populations. In Supplementary Information, section 4 we observed that allele frequency differences between Neolithic Iran and Natufians are correlated with those of EHG and WHG, thus we introduce EHG and WHG into the Right set.

Table S7.5: Modeling Levant_N. While Levant_N can be modeled as a 2-way mixture of Natufians and various populations when the O9 outgroups are used, it can only be modeled as a mixture of Natufians and Anatolian Neolithic when introducing WHG and EHG as outgroups to the Right set.

			Mixture Proportions		
A	**Right**	**P for rank=1**	**Natufian**	***A***	**Std. Error**
Anatolia_N	O9	9.89E-02	0.670	0.330	0.098
Anatolia_N	**O9EW**	**2.05E-01**	**0.667**	**0.333**	**0.078**
CHG	O9	2.23E-01	0.715	0.285	0.078
CHG	O9EW	1.64E-02	0.854	0.146	0.085
Iran_N	O9	8.28E-02	0.633	0.367	0.146
Iran_N	O9EW	7.02E-03	1.045	-0.045	0.104

Using this approach we can reject every one of the candidates as sources for the Levant Neolithic *except* the Anatolian Neolithic (P=0.205). This should not be interpreted literally as evidence that people from Anatolia moved into the Levant, admixed with Natufians and became the Levantine Neolithic. First, because we do not know the geographical extent of populations similar to our sample from Northwestern Anatolia, and second because an even better candidate population may be sampled in the future (perhaps from eastern Anatolia or Mesopotamia, regions that have some of the earliest evidence from agriculture). Nonetheless, this result seems tentatively sensible in view of (i) our results from PCA and ADMIXTURE (Fig. 1; Extended Data Fig. 2) that show that Levant_N is intermediate between the Anatolian Neolithic and Natufians and shares with it an ancestral component, (ii) f_4-statistics which shows that the Anatolian Neolithic tops the statistic f_4(Test, Natufian; Levant_N, Chimp) (Fig. S7.2), and (iii) the negative f_3-statistic of the form f_3(Levant_N; Natufian, Anatolia_N) (Table S7.4).

Thus, the genetic evidence suggests a combination of continuity (Fig. S7.3) and admixture (Table S7.5) in the origin of the Levantine Neolithic, with a best estimate of ancestry of 66.7±7.8% from the Natufians and 33.3±7.8% from a population related to the Anatolian Neolithic.

Origin of the Neolithic of Iran and its relationship to the Mesolithic

In the Levant we have several samples of both the Neolithic and Natufian population, but from Iran, we possess a single individual of probable Mesolithic date (Iran_HotuIIIb; see Supplementary Information, section 1 for discussion regarding the uncertainty about its date) from Hotu Cave. This makes inferences about population continuity or change in this region more tentative. However, it is clear from PCA and ADMIXTURE analyses that the Neolithic of Iran from Ganj Dareh clusters closely with the likely Mesolithic individual (Fig. 1). Both are strikingly different from all Levantine individuals and cluster more remotely with later Chalcolithic samples from Iran, Chalcolithic and Bronze Age Armenians, and the recently described Caucasus Hunter-Gatherers (CHG) from Georgia.

In Fig. S7.4 we plot the statistic $f4$(*Test*, Iran_HotuIIIb; Iran_N, Chimp). Iran_N shares significantly more alleles with the Mesolithic of Iran than with most other populations in our dataset, consistent with the two populations clustering in PCA (Fig. 1b). For some populations there is no significant difference, including the pre-agricultural Caucasus hunter-gatherers (CHG) that as we will see below can be modeled as having most of their ancestry from a population like Iran_N, and later populations from Iran. This parallels Fig. S7.2 for the Levant in demonstrating a degree of continuity in the region.

In Fig. S7.5 we plot the statistic f_4(Iran_HotuIIIb, Iran_N; *Test*, Chimp). If Iran_N is derived from Iran_HotuIIIb by admixture with another population, then a negative statistic of the above form might help identify surrogates for that population. However, we do not observe any such statistics, but rather a couple of statistics that are borderline significantly positive (Z>3), involving the EHG and the related Eneolithic Steppe population also from eastern Europe and the related MA1 sample from Upper Paleolithic Siberia (Z=2.8). Thus it is rather the Mesolithic of Iran that shares more alleles with these eastern European groups than the Neolithic. Tentatively, this might suggest that the pre-Neolithic population of Iran had an affinity to the EHG/Ancient North Eurasians that was diluted during the Neolithic, although the lack of negative f_4-statistics does not allow us to discern what is the source of this dilution. Alternatively, there was no dilution, but the Neolithic of Iran was descended from an unsampled Mesolithic population.

We also searched more broadly by allowing *Test* to include any present-day population in the *HO* dataset, but do not find any significantly negative statistics (Z>-1.1). Moreover, we also do not find any negative f_3-statistics with the Iran Neolithic as a target for any reference pair of ancient populations in the *HOIll* dataset (Z>-1) or ancient/present-day populations in the *HO* dataset (Z>-1.1) Thus, it seems that the Neolithic of Iran represents an "extreme" population which cannot be modeled as a mixture of other populations in our dataset, consistent also with its position in the PCA (Fig. 1b) well outside the variation of any other populations.

The origin of the Neolithic of Iran does not appear to be related to either Anatolia or the Levant, as the Neolithic and Mesolithic of Iran are symmetrically related to either population (Fig. S7.5), providing no evidence for gene flow from either region into the Zagros, but hinting strongly that whatever role the exchange of ideas and technology may have played in the emergence of the Neolithic in the Zagros, this was not accompanied with any substantial gene flow from other ancient Near Eastern Neolithic centers of domestication.

We can model Iran_HotuIIIb as 90.6±4.4% Iran_N and 9.4±4.4% EHG, with O9ALNW as outgroups (P-value for rank=1 is 0.45). Additional and better data from pre-Neolithic individuals from the Zagros may clarify the population composition of its late inhabitants prior to the appearance of agriculture.

Figure S7.4: Testing whether the Neolithic of Iran shares more alleles with the Mesolithic of Iran or with another *Test* population with the statistic f_4(*Test*, Iran_HotuIIIb; Iran_N, Chimp).

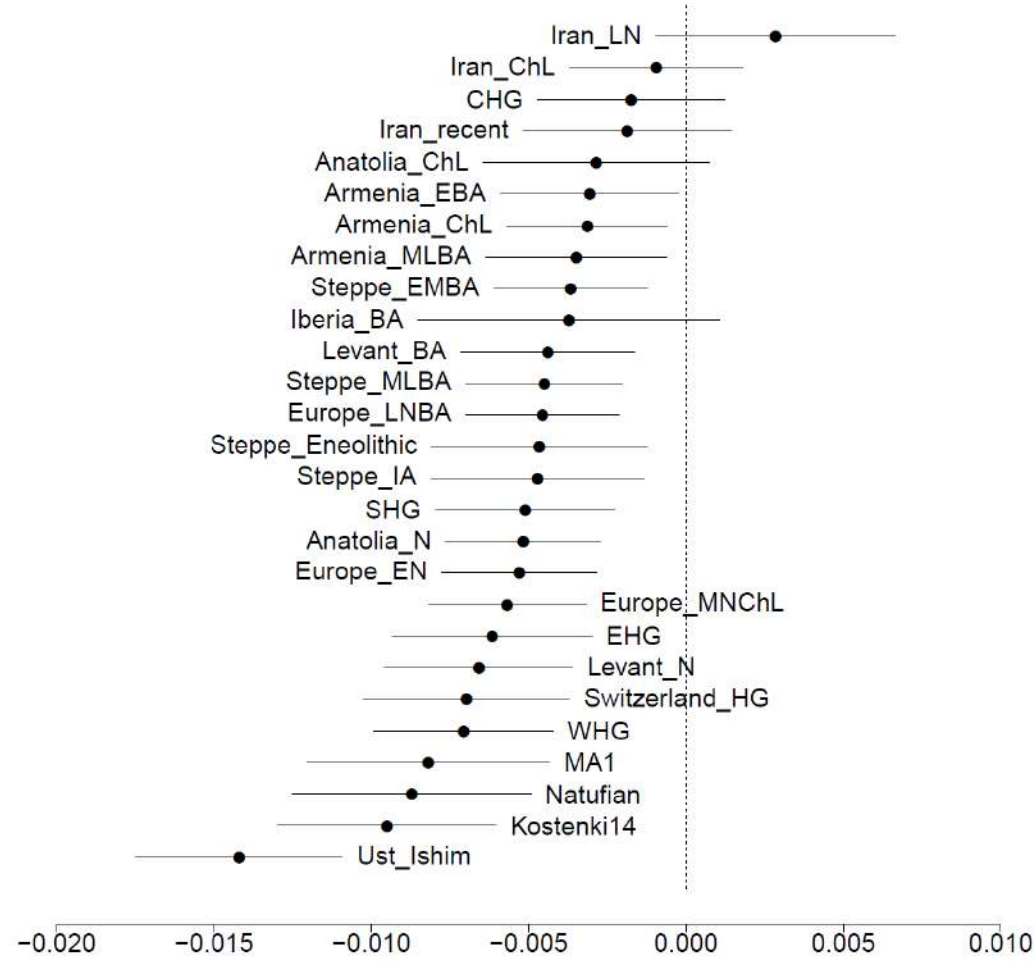
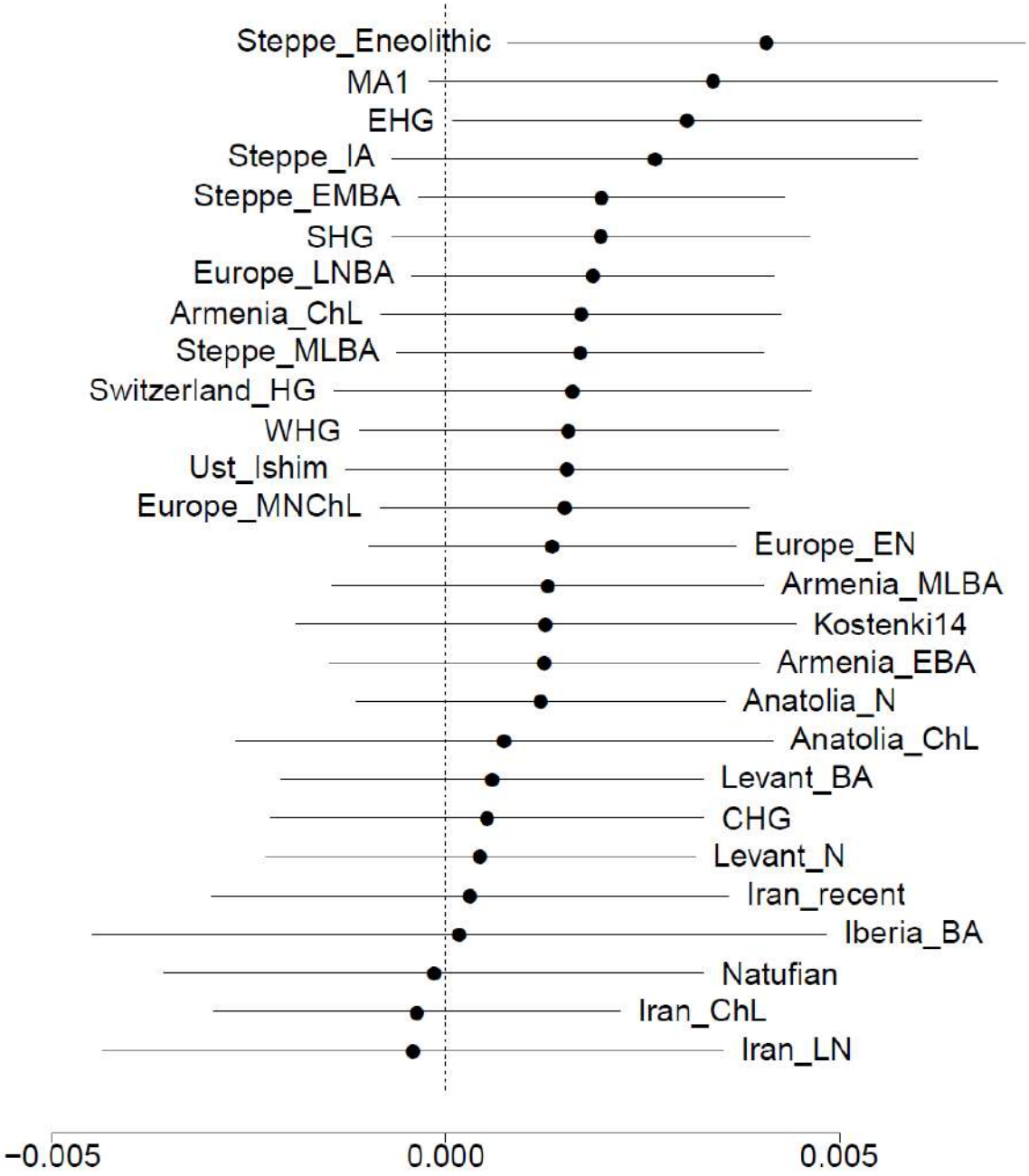


Figure S7.5: Testing for symmetry between the Mesolithic and Neolithic of Iran with the statistic f_4(Iran_HotuIIIb, Iran_N; *Test*, Chimp).



Caucasus hunter-gatherers can be modeled as a mixture of Neolithic Europe and European hunter-gatherers

The Caucasus hunter-gatherers cluster with the Iran_N in PCA (Fig. 1b), but are also shifted towards Europe. We verify with f_4-statistics that the CHG and Iran_N do not form a clade, but rather that most ancient West Eurasian populations share more alleles with the CHG than with the Iran_N (Fig. S7.6). Some of these statistics could be interpreted in terms of gene flow into Europe, associated with the migration of the Yamnaya that included a component from the Caucasus[1], for which the CHG could be a representative[16]. However, similar to the argument we previously made for the relationship of the Levantine Neolithic to Natufians, the positive f_4-statistics that involve European hunter-gatherers or even MA1 cannot be explained in this manner (as they predate the Yamnaya by thousands of years).

An alternative explanation is that the excess of genetic drift shared by the CHG and European hunter-gatherers is caused by the Iran_N having more Basal Eurasian ancestry than the CHG. The statistic f_4(CHG, Iran_N; Ust_Ishim, Chimp) = 0.00145 (Z=2.8; 762,778 SNPs) is suggestive that this is the case. The same could be the case for the Iran_HotuIIIb population, which is similar in magnitude, but on a much lower number of SNPs: f_4(Iran_HotuIIIb, Iran_N; Ust_Ishim, Chimp) = 0.00154 (Z=1.6; 116,123 SNPs).

Table S7.6: No strong evidence that allele frequency differences between CHG and Iran_N are strongly associated with a particular ancient West Eurasian hunter-gatherer population.

HG_1	HG_2	f_4(CHG, Iran_N; HG_1, HG_2)	Z	Number of SNPs
WHG	EHG	0.00068	1.5	448526
WHG	SHG	0.00101	3.0	451422
WHG	Switzerland_HG	0.00039	0.9	462807
WHG	MA1	0.00106	1.8	337136
WHG	Kostenki14	0.00157	3.0	434510
EHG	SHG	0.00032	0.8	440232
EHG	Switzerland_HG	-0.00037	-0.7	448141
EHG	MA1	0.00056	0.9	326727
EHG	Kostenki14	0.00110	1.9	420881
SHG	Switzerland_HG	-0.00067	-1.5	451050
SHG	MA1	-0.00004	-0.1	328854
SHG	Kostenki14	0.00064	1.2	423710
Switzerland_HG	MA1	0.00103	1.5	336886
Switzerland_HG	Kostenki14	0.00133	2.2	434174
MA1	Kostenki14	0.00085	1.2	316733

There does not appear to be any evidence that the difference between CHG and Iran_N is strongly correlated with any particular European hunter-gatherer population (Table S7.6), with the only two comparisons that reach significant (|Z|=3) hinting that CHG shares more alleles with WHG than with SHG or Kostenki14. We can model CHG as a mixture of Iran_N and different European hunter-gatherer populations (Table S7.7), with an estimate of 71.6±6.0% Iran_N, 7.0±3.8% WHG, 21.4±7.7% EHG. A sanity check of this estimate is that it predicts 0.716*.482=0.345 Basal Eurasian ancestry in the CHG, virtually identical to the 0.347±0.058 of Supplementary Information, section 4.

Figure S7.6: Testing for symmetry between the CHG and the Neolithic of Iran with the statistic f_4(CHG, Iran_N; *Test*, Chimp).

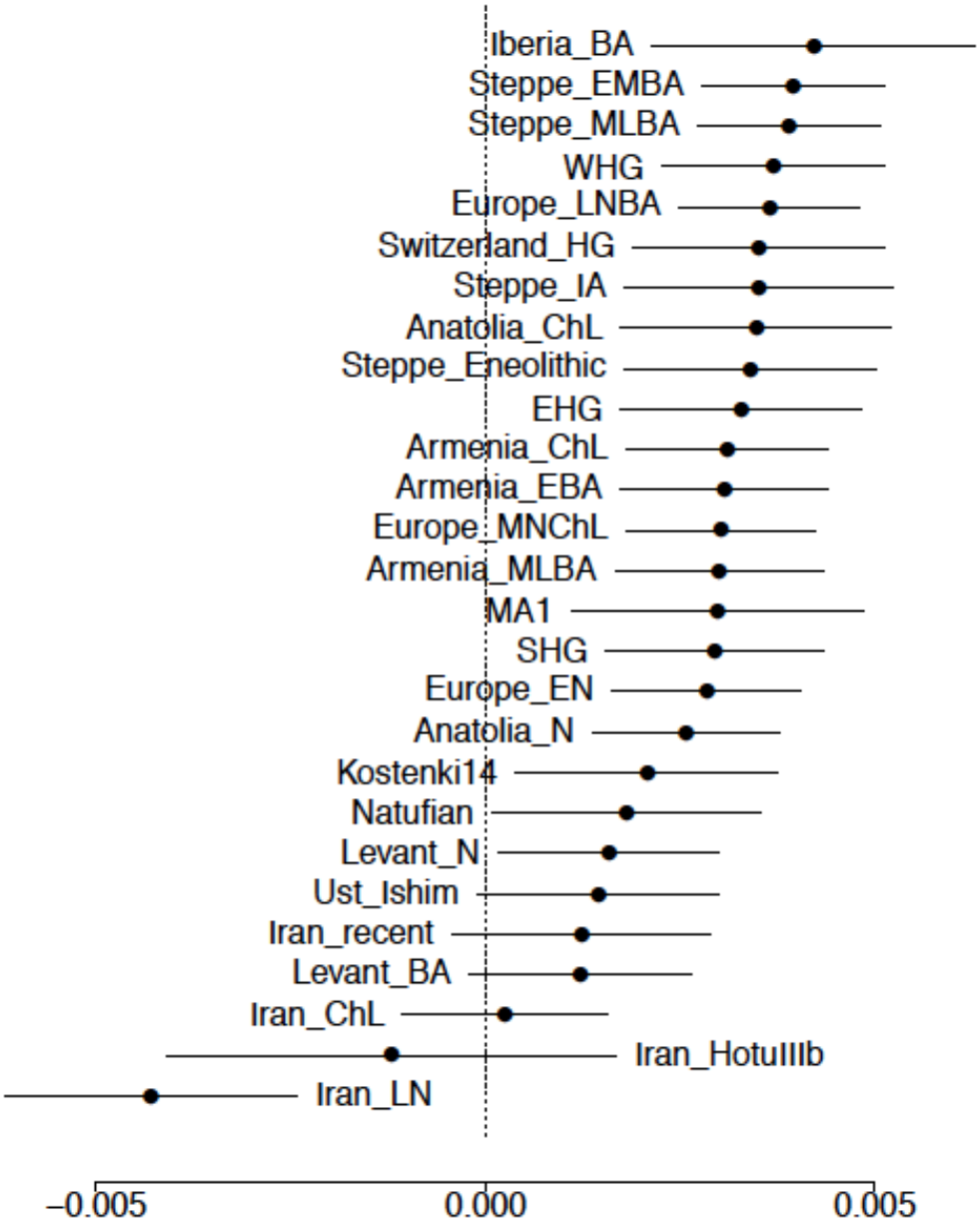


Table S7.7: Modeling CHG. CHG can be modeled as a mixture of Iran_N and European hunter-gatherers with the Neolithic Levant and Natufians as outgroups, but not with the Anatolian Neolithic as an outgroup.

				Mixture Proportions	
A	**Right**	P for rank=1	Iran_N	*A*	Std. Error
EHG	O9N	1.35E-01	0.852	0.148	0.033
EHG	O9LN	8.67E-02	0.867	0.133	0.032
EHG	O9ALN	7.62E-05	0.894	0.106	0.033
SHG	O9N	5.80E-01	0.731	0.269	0.048
SHG	O9LN	5.92E-01	0.734	0.266	0.048
SHG	O9ALN	3.89E-02	0.713	0.287	0.049
Switzerland_HG	O9N	9.87E-02	0.694	0.306	0.075
Switzerland_HG	O9LN	8.05E-02	0.713	0.287	0.076
Switzerland_HG	O9ALN	1.99E-03	0.678	0.322	0.083
WHG	O9N	1.97E-01	0.700	0.300	0.062
WHG	O9LN	2.50E-01	0.700	0.300	0.062
WHG	O9ALN	3.32E-02	0.655	0.345	0.061
			Mixture Proportions			Std. Errors	
		P for rank=2	Iran_N	WHG	EHG	Iran_N	WHG	EHG
WHG, EHG	**O9LN**	**4.52E-01**	**0.716**	**0.070**	**0.214**	**0.060**	**0.038**	**0.077**
WHG, EHG	O9ALN	2.42E-02	0.659	0.025	0.316	0.061	0.036	0.076

The Anatolian Neolithic

In the Levant and Iran we have Epipalaeolithic/Mesolithic samples and can thus compare the first Neolithic populations with the hunter-gatherers that preceded them. There are currently no samples of Epipalaeolithic Anatolians, but we observe that the Neolithic Anatolians are genetically shifted towards Europe in the PCA (Fig. 1b) and have ancestry from an ancestral population related to European hunter-gatherers according to ADMIXTURE analysis (Extended Data Fig. 2a). This should not be interpreted as evidence of ancestry from actual hunter-gatherers from Europe; while this is not implausible for our sample from Northwestern Anatolia[17], we have previously seen that populations of the ancient Near East are also differentially related to European hunter-gatherers. This suggests that populations related to European hunter-gatherers existed in the Near East and may be included in the Epipalaeolithic/Mesolithic ancestors of the Neolithic Anatolians without any need for a direct migration from Europe.

First, to understand how Neolithic Anatolians differ from the other two Neolithic Near Eastern populations we study f_4-statistics (Fig. S7.7, 8) of the form f_4(Anatolia_N, Levant_N or Iran_N; *Test*, Chimp) that demonstrate that Europeans share more alleles with the Anatolians than with the other two. This could be ascribed to the later Europeans having ancestry from Anatolia, but cannot account for the positive statistics of the above form when *Test* is a European hunter-gatherer population

(before any Neolithic migrations), including the ~13,000 year old Switzerland_HG[16], or the ~37,000 year old Kostenki14[4]. Thus, the greater affinity of Anatolia_N to European hunter-gatherer populations cannot be attributed to gene flow from Anatolia to Europe related to the Neolithization of the continent, or indeed to earlier possible gene flows that affected only post-glacial or Holocene-era European hunter-gatherers.

Figure S7.7: Statistics of the form f_4(Anatolia_N, Levant_N; *Test*, Chimp)

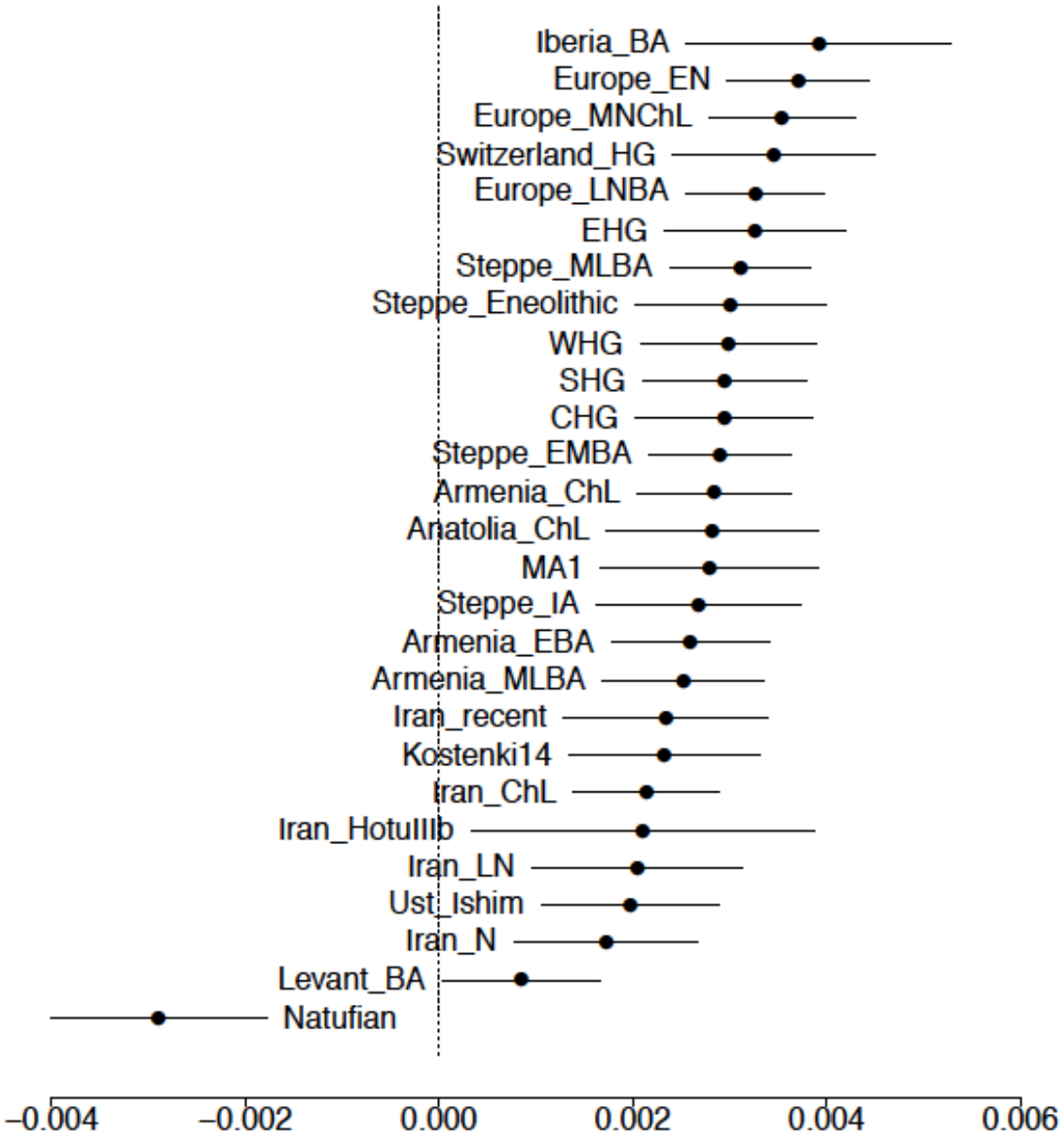
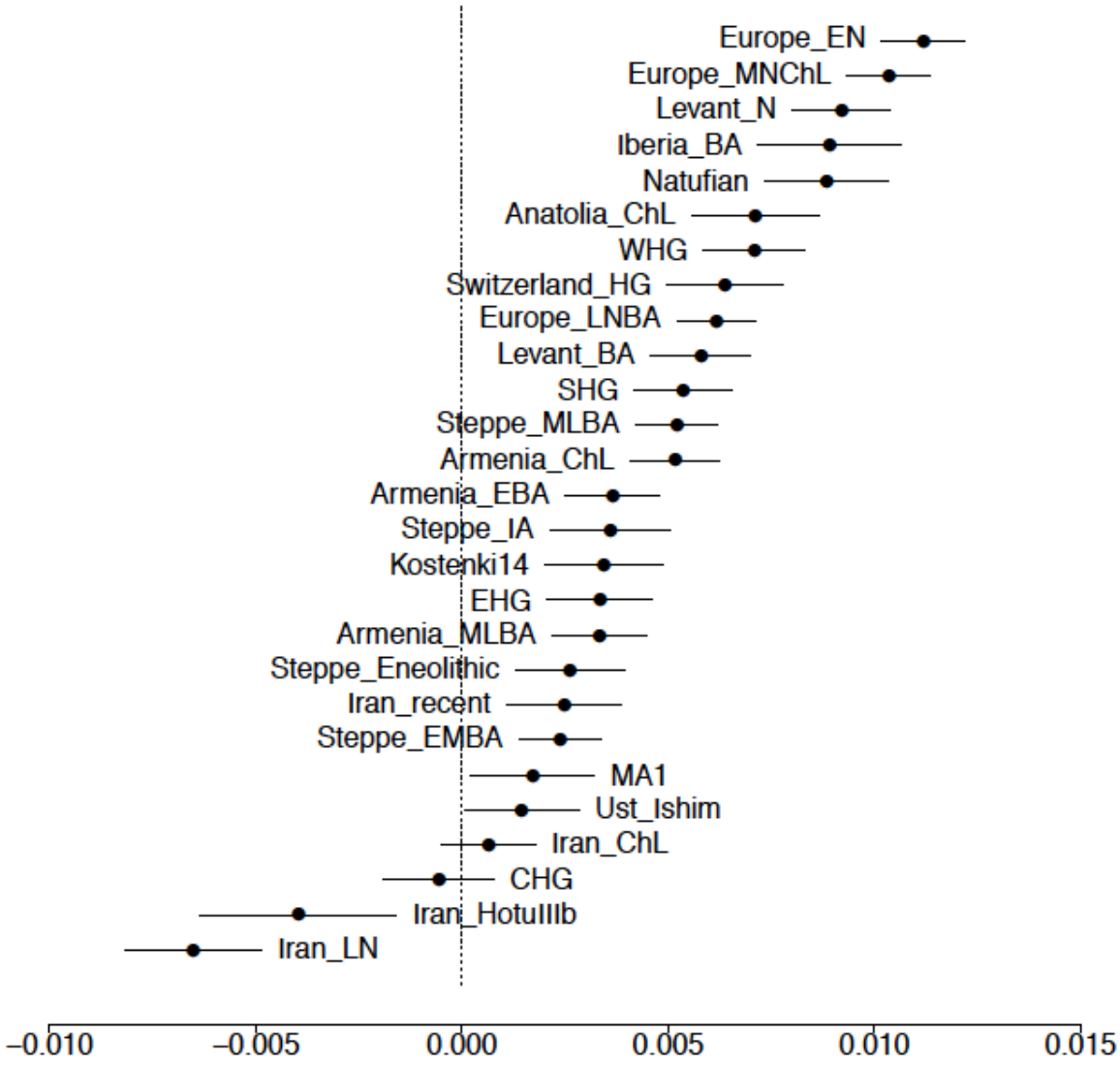


Figure S7.8: Statistics of the form f_4(Anatolia_N, Iran_N; *Test*, Chimp)



We modeled the Anatolian Neolithic as a 4-way mixture of the Four (Table S7.3), but with a negative inferred mixture proportion from EHG. We tried to model it as a simpler 3-way mixture (Table S7. 8), by dropping each of the Four in turn. All the simpler 3-way mixtures are rejected strongly (P<0.0001) except the one involving Neolithic Iran, Neolithic Levant, and WHG (P=0.0171), also suggesting that EHG is the population that does not contribute to the Anatolian Neolithic. In Supplementary Information section 10, we give an interpretation of the origin of the Anatolian Neolithic that interprets the negative EHG mixture proportion (Table S7.3) in terms of admixture (into the Anatolian Neolithic) of a population that is not exactly WHG, but rather residing on the EHG→WHG cline beyond WHG.

Table S7.8: Modeling Anatolian Neolithic as a 3-way mixture. We move each of the Four from the set of reference populations (Table S7.3) into the Right set and test rank=2 for Left=(Anatolia_N, A, B, C). Rank=2 is strongly rejected for all (A, B, C) triples (P<1E-07) except (Iran_N, Levant_N, WHG) (P=0.0171).

					Mixture Proportions			Standard Errors		
A	B	C	Right	P for rank=2	A	B	C	A	B	C
Iran_N	EHG	WHG	O9	9.03E-02	0.67	-0.113	0.444	0.048	0.033	0.065
Iran_N	EHG	WHG	O9L	2.64E-36	4.896	0.386	-4.282	8.096	2.412	10.473
Iran_N	Levant_N	EHG	O9	1.24E-04	-75.741	70.136	6.605	26.641	24.354	2.306
Iran_N	Levant_N	EHG	O9W	2.03E-10	-0.281	1.155	0.126	0.088	0.094	0.024
Iran_N	Levant_N	WHG	O9	4.14E-02	0.251	0.43	0.32	0.148	0.136	0.029
Iran_N	**Levant_N**	**WHG**	**O9E**	**1.71E-02**	**0.387**	**0.339**	**0.274**	**0.134**	**0.137**	**0.021**
Levant_N	EHG	WHG	O9	2.11E-02	0.638	-0.042	0.404	0.033	0.026	0.051
Levant_N	EHG	WHG	O9I	2.17E-08	0.742	-0.015	0.273	0.032	0.03	0.053

SHG, Europe_EN, Europe_MNChL, Europe_LNBA

The admixture history of mainland European Neolithic and Bronze Age populations has been recently discussed[1,6,17] and we confirm here the main findings of these earlier studies in the present context (Table S7.9). The Scandinavian hunter-gatherers (SHG) can be modeled as a mixture of WHG and EHG[1]. Early and Middle/Neolithic populations are modeled as a mixture of Anatolian Neolithic and WHG, with more hunter-gatherer ancestry in the Middle Neolithic/Chalcolithic populations[1,17]. Populations of the Late Neolithic/Bronze Age can be modeled as mixtures of the preceding Middle Neolithic/Chalcolithic populations and populations from the Early to Middle Bronze Age from the steppe[1,17].

Table S7.9: Modeling European populations with the O9 outgroups.

				Mixture Proportions		
Test	A	B	P for rank=1	A	B	Std. Error
SHG	EHG	WHG	1.38E-01	0.429	0.571	0.030
Europe_EN	Anatolia_N	WHG	2.50E-01	0.929	0.071	0.023
Europe_MNChL	Anatolia_N	WHG	1.15E-01	0.779	0.221	0.028
Europe_LNBA	Europe_MNChL	Steppe_EMBA	2.37E-01	0.469	0.531	0.016

Steppe_EMBA

It has been observed that while the Yamnaya populations from the Early Bronze Age steppe were carriers of Ancient North Eurasian ancestry into Europe, they had less of it than the preceding EHG[1]. The timeline of the "dilution of EHG ancestry" on the steppe was further clarified by the discovery that during the Eneolithic period the reduction of EHG ancestry had already begun[17]. The best surrogate for the population effecting this dilution was present-day Armenians[1,17], and a recent study has identified the CHG as an ancient population that could function as a source for the southern component in the ancestry of the Yamnaya. It has been observed that the Yamnaya[1,14], Afanasievo[14], and Middle Bronze Age Poltavka[17] culture formed a tight genetic cluster which we name here Steppe_EMBA.

In Fig. S7.9 we show that there are negative f_4(EHG, Steppe_EMBA; *A*, Chimp) statistics when *A* is a population of Near Eastern ancestry and positive ones when *A* is a European hunter-gatherer population, documenting the dilution of EHG ancestry from a Near Eastern source. Direct evidence that this is due to admixture is provided by the negative f_3(Steppe_EMBA; EHG, *A*) statistic (Fig. S7.10) when *A* is a Near Eastern source, with the CHG and populations of Iran providing the most negative statistics.

We verify that we can model the Steppe_EMBA as a mix of EHG and Near Eastern populations from the Caucasus, Armenia, and Iran (Table S7.10). We further test the plausible sources by adding outgroups to the Right set (Table S7.11). When we add Anatolia_N, Levant_N, Natufians, and WHG as additional outgroups, rank=1 is rejected for all populations (P<.001) except the Chalcolithic of Iran (P=0.057) and the recent individual from Ganj Dareh (P=0.205) that clusters with Chalcolithic Iran and ancient Armenia rather than the much older Neolithic individuals from Ganj Dareh (Iran_recent; Fig. 1b). We do not at present know the geographical distribution of populations like the Chalcolithic of Iran in the ancient Near East and the source of ancestry in the steppe could be from a geographically more proximate source in the Caucasus. This may be clarified with additional sampling, but it is important to note that regardless of the choice for the Near Eastern population, a relatively stable estimate of ~50/50% ancestry from the EHG and the Near East is inferred.

Figure S7.9: Testing for symmetry between EHG and Steppe_EMBA with the statistic f_4(EHG, Steppe_EMBA; *A*, Chimp)

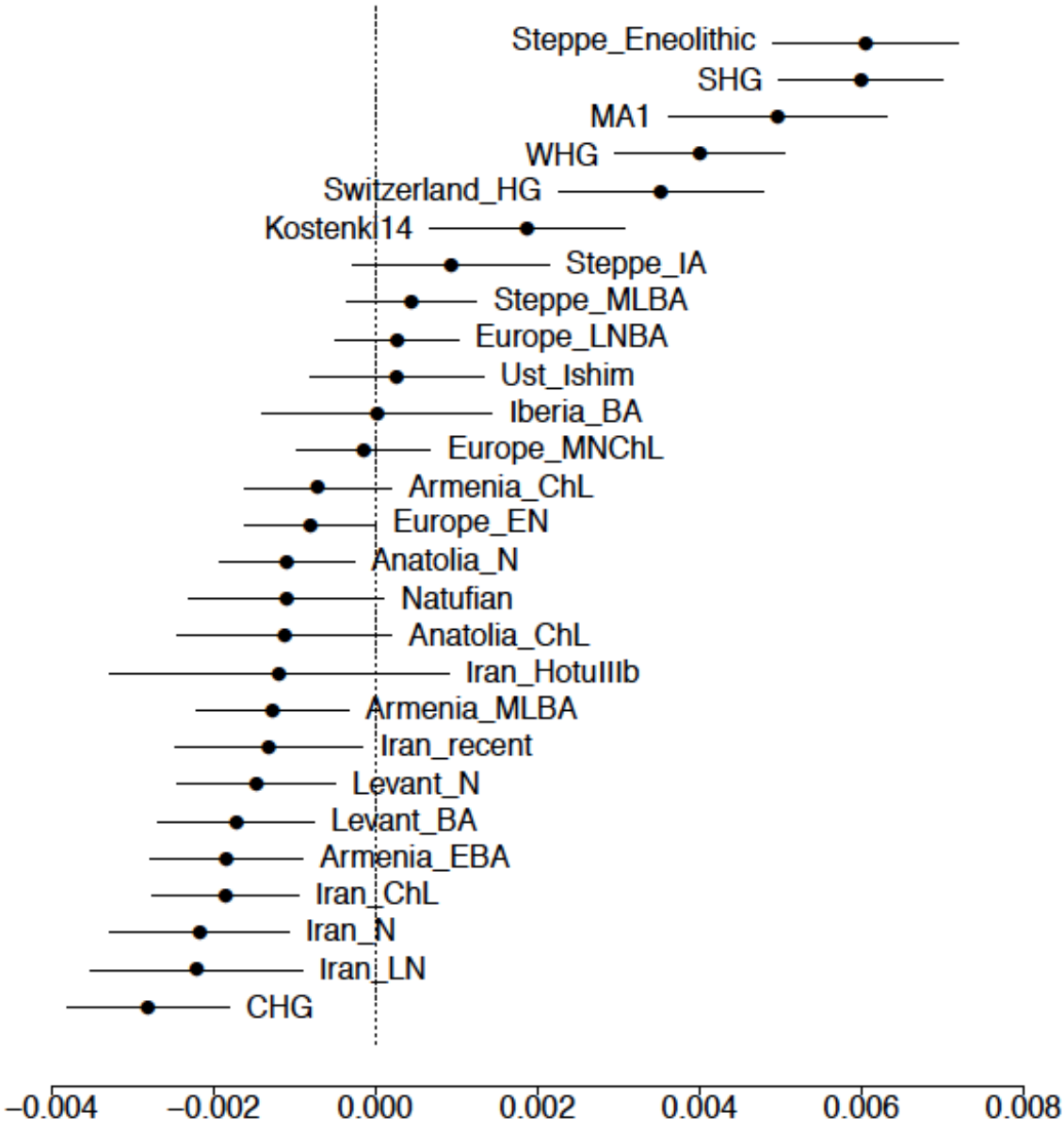


Figure S7.10: Testing for admixture in Steppe_EMBA with the statistic f_3(Steppe_EMBA; EHG, *A*)

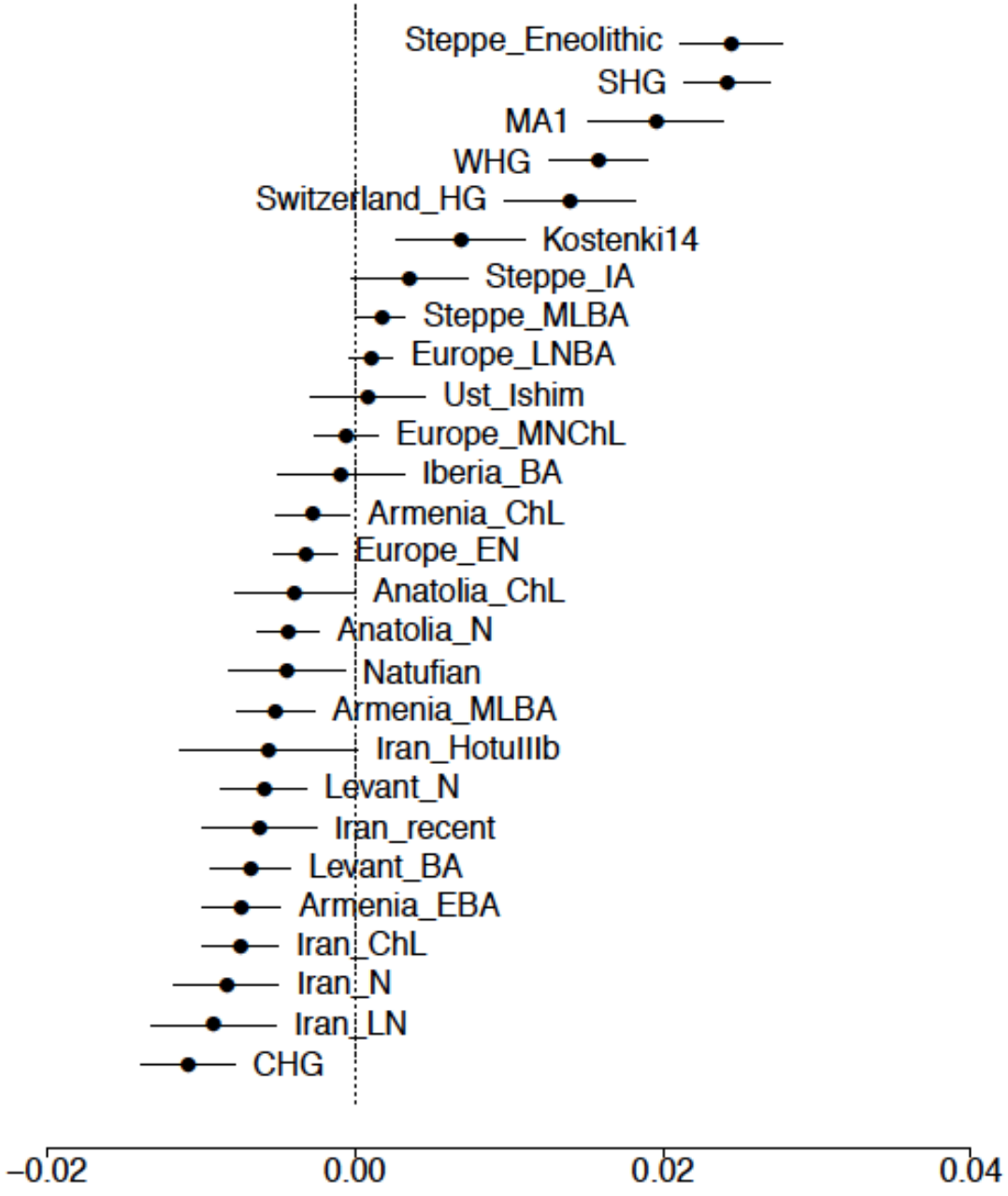


Table S7.10: Modeling Steppe_EMBA as a mix of EHG and a population *A*. When *A* is one of several populations from the Caucasus, Armenia, and Iran, the Left=(Steppe_EMBA, EHG, *A*) is related to O9 with 2 streams of ancestry.

			Mixture Proportions		
A	**Right**	**P for rank=1**	**EHG**	***A***	**Std. Error**
Anatolia_ChL	O9	1.76E-03	0.511	0.489	0.026
Anatolia_N	O9	5.08E-04	0.562	0.438	0.018
Armenia_ChL	O9	2.40E-02	0.439	0.561	0.022
Armenia_EBA	O9	1.35E-01	0.508	0.492	0.020
Armenia_MLBA	O9	3.50E-01	0.451	0.549	0.021
CHG	O9	3.82E-01	0.500	0.500	0.021
Europe_EN	O9	3.97E-04	0.550	0.450	0.018
Europe_LNBA	O9	9.21E-04	0.346	0.654	0.020
Europe_MNChL	O9	3.86E-05	0.532	0.468	0.019
Iberia_BA	O9	4.53E-02	0.441	0.559	0.031
Iran_ChL	O9	1.02E-01	0.535	0.465	0.019
Iran_LN	O9	2.95E-01	0.588	0.412	0.021
Iran_HotuIIIb	O9	2.22E-02	0.508	0.492	0.034
Iran_N	O9	3.65E-02	0.579	0.421	0.021
Iran_recent	O9	2.74E-01	0.513	0.487	0.023
Levant_BA	O9	8.95E-04	0.588	0.412	0.018
Levant_N	O9	6.30E-07	0.630	0.370	0.018
Natufian	O9	5.15E-05	0.657	0.343	0.019
SHG	O9	4.42E-13	-0.048	1.048	0.066
Steppe_Eneolithic	O9	4.46E-02	-0.955	1.955	0.269
Steppe_IA	O9	2.29E-08	0.183	0.817	0.055
Steppe_MLBA	O9	2.05E-02	0.253	0.747	0.021
Switzerland_HG	O9	5.71E-17	0.471	0.529	0.037
WHG	O9	4.66E-26	0.466	0.534	0.042

Table S7.11: Modeling Steppe_EMBA as EHG and a population *A*. We add Natufians, WHG, Levant_N, and Anatolia_N as outgroups to the Right set and test whether the triple (Steppe_EMBA, EHG, *A*) could be related via 2 streams of ancestry from the Right set.

			Mixture Proportions		
A	**Right**	**P for rank=1**	**EHG**	***A***	**Std. Error**
Armenia_EBA	O9N	6.59E-03	0.535	0.465	0.019
Armenia_EBA	O9W	4.03E-02	0.491	0.509	0.018
Armenia_EBA	O9NW	1.83E-03	0.518	0.482	0.017
Armenia_EBA	O9ANW	7.54E-06	0.563	0.437	0.012
Armenia_EBA	O9LNW	1.24E-04	0.542	0.458	0.015
Armenia_EBA	O9ALNW	1.37E-05	0.564	0.436	0.012
Armenia_MLBA	O9N	6.13E-03	0.477	0.523	0.021
Armenia_MLBA	O9W	1.99E-01	0.435	0.565	0.019
Armenia_MLBA	O9NW	3.61E-03	0.461	0.539	0.018
Armenia_MLBA	O9ANW	1.89E-04	0.496	0.504	0.014
Armenia_MLBA	O9LNW	6.14E-04	0.481	0.519	0.016
Armenia_MLBA	O9ALNW	2.73E-04	0.498	0.502	0.014
CHG	O9N	9.03E-02	0.488	0.512	0.022
CHG	O9W	3.52E-01	0.513	0.487	0.019
CHG	O9NW	8.48E-02	0.501	0.499	0.019
CHG	O9ANW	3.08E-05	0.459	0.541	0.017
CHG	O9LNW	9.93E-03	0.482	0.518	0.017
CHG	O9ALNW	6.21E-05	0.458	0.542	0.017
Iran_ChL	O9N	9.75E-02	0.543	0.457	0.018
Iran_ChL	O9W	1.39E-01	0.541	0.459	0.016
Iran_ChL	O9NW	1.33E-01	0.548	0.452	0.015
Iran_ChL	O9ANW	8.58E-02	0.565	0.435	0.012
Iran_ChL	O9LNW	3.78E-02	0.566	0.434	0.014
Iran_ChL	**O9ALNW**	**5.70E-02**	**0.568**	**0.432**	**0.012**
Iran_LN	O9N	2.31E-01	0.580	0.420	0.020
Iran_LN	O9W	2.67E-01	0.601	0.399	0.018
Iran_LN	O9NW	2.14E-01	0.593	0.407	0.017
Iran_LN	O9ANW	5.66E-06	0.548	0.452	0.016
Iran_LN	O9LNW	4.54E-02	0.578	0.422	0.016
Iran_LN	O9ALNW	9.59E-06	0.549	0.451	0.016
Iran_recent	O9N	2.89E-01	0.520	0.480	0.022
Iran_recent	O9W	2.65E-01	0.502	0.498	0.021
Iran_recent	O9NW	2.83E-01	0.508	0.492	0.020
Iran_recent	O9ANW	1.74E-01	0.529	0.471	0.015
Iran_recent	O9LNW	1.91E-01	0.523	0.477	0.017
Iran_recent	**O9ALNW**	**2.05E-01**	**0.531**	**0.469**	**0.015**

We were intrigued by the fact that populations from Iran were a better fit for the population admixing into the steppe than the CHG, so we examined f_4(Steppe_EMBA, Fitted Steppe_EMBA; Ust_Ishim, *Outgroup*) to see what interaction with the outgroups is causing better fit in Iran_ChL than in CHG when Anatolian and Levantine Neolithic are included in the outgroups (Table S7.11).

Table S7.12: Value of the statistic f_4(Steppe_EMBA, *Fitted Steppe_EMBA*; Ust_Ishim, *Outgroup*).

	Fitted Steppe_EMBA					
	.458EHG+.542CHG		.568EHG+.432Iran_ChL		.527EHG+.181CHG+.292Iran_ChL	
Outgroup	**Statistic**	**Z**	**Statistic**	**Z**	**Statistic**	**Z**
Kostenki14	0.00007	0.2	0.00050	1.8	0.00035	1.3
MA1	-0.00070	-2.1	0.00042	1.3	0.00002	0.1
Han	-0.00017	-0.7	0.00025	1.2	0.00010	0.5
Papuan	-0.00025	-1.1	0.00008	0.4	-0.00003	-0.1
Onge	-0.00033	-1.4	0.00002	0.1	-0.00009	-0.5
Chukchi	-0.00024	-1.0	0.00032	1.4	0.00012	0.6
Karitiana	-0.00021	-0.8	0.00066	2.6	0.00035	1.5
Mbuti	0.00011	0.4	0.00036	1.6	0.00029	1.4
Anatolia_N	-0.00077	-3.3	0.00044	2.0	0.00005	0.3
Levant_N	-0.00072	-2.6	0.00066	2.5	0.00022	0.9
Natufian	-0.00066	-2.0	0.00048	1.5	0.00011	0.4
WHG	-0.00064	-2.3	0.00033	1.3	-0.00002	-0.1

In Table S7.12 we see that when the southern population is CHG, several statistics are underestimated in the sense that the actual Steppe_EMBA shares more genetic drift with populations like Anatolia_N (Z=-3.3) than predicted by the fitted model of 0.458EHG+0.542CHG (Table S7.11). While no such extreme outlier is observed when the southern population is Iran_ChL, we nonetheless noticed that the opposite trend applies for Iran_ChL: differences between Fitted and actual Steppe_EMBA tend to be positive.

This led us to try one last model in which we model Steppe_EMBA as a 3-way mix of EHG, CHG, and Iran_ChL. The P-value for rank=2 is 0.241, so 3 streams of ancestry are consistent with the quadruple (Steppe_EMBA, EHG, CHG, Iran_ChL) and the fitted mixture proportions are 52.7±2.0% EHG, 18.1±7.4% CHG, 29.2±5.9% Iran_ChL. We do not want to overemphasize this 3-way mixture model as the 2-way one is more parsimonious and consistent with the data (Table S7.11). Nonetheless the 3-way model is also plausible as it suggests an explanation for the shared genetic drift between Steppe_EMBA and the Anatolian and Levantine Neolithic (underestimated when CHG alone is the southern population; Table S7.12), and makes geographical sense as admixture from the Near East

could have arrived on the steppe via the Caucasian isthmus where an addition of CHG ancestry could have occurred.

Steppe_MLBA

Beginning in the Middle Bronze age and in the Late Bronze Age a set of populations appear on the Eurasian steppe spanning both Europe, in the Srubnaya culture and in the trans-Ural steppe with the Andronovo and Sintashta cultures. These populations were descended from both the previous Yamnaya/Afanasievo/Poltavka ("Steppe_EMBA") population, but also Neolithic farmers from Europe. Table S7.13 shows our modeling of the Steppe_MLBA population as a mixture of Steppe_EMBA that preceded them on the Eurasian steppe and Europe_MNChL who occupied Europe prior to the steppe migration into mainland Europe. The Steppe_MLBA population thus resembles the Europe_LNBA population (Table S7.9), being composed of similar population elements; the two overlap in PCA (Fig. 1b) and represent a continuum of populations from mainland Europe to the eastern parts of the Eurasian steppe.

Table S7.13: Modeling Steppe_MLBA

		Mixture Proportions		
Right	**P for rank=1**	**Steppe_EMBA**	**Europe_MNChL**	**Std. Error**
O9	4.12E-01	0.684	0.315	0.021

Iran_ChL

The Chalcolithic population of Seh Gabi from Iran appears to be related to the Neolithic population from Ganj Dareh but is shifted towards Europe (Fig.1b) and occupies the northern end of the Near Eastern cline, close to the CHG, but shifted towards the Levant. Its relationship to Iran_N is shown by the statistic f_4(Iran_ChL, *A*; Iran_N, Chimp) (Fig. S7.11) and the fact that despite this relationship it does not form a strict clade with it is shown by the statistic f_4(Iran_N, Iran_ChL; *A*, Chimp) (Fig. S7.12). The two observations that Neolithic Iran shares more alleles with Chalcolithic Iran than with any other population outside Iran, but populations outside Iran share more alleles with Chalcolithic than Neolithic Iran, are suggestive that Iran_ChL is a population of mixed ancestry, having both a local (Iran_N) source but also at least one second source related to populations outside Iran. This is confirmed directly via statistics of the form f_3(Iran_ChL; Iran_N, *A*) (Fig. S7.13) which show that significantly negative statistics are produced for several populations *A*.

We first test Left=(Iran_ChL, Iran_N, *A*) to determine if Iran_ChL could be modeled as a simple 2-way mixture of Iran_N and another population *A* (Table S7.14). While many populations *A* are consistent with forming such a mixture when the O9 outgroups are used, only three are when we introduce Switzerland_HG and Natufians to the Right set of outgroups. These three populations are

Anatolia_ChL, Armenia_EBA, and Iran_recent. As we will see below, we can model Anatolia_ChL itself as a mix of Iran_ChL and Anatolia_N. The other two (Iran_recent and Armenia_EBA) cluster together (Fig. 1b) and Iran_recent may be part from an un-sampled populations of Iran or may date to a later period than the Neolithic. We do not have an earlier sample than the Chalcolithic in Armenia, which, however, is not from the sampling location as our Early Bronze Age samples; thus, it is in principle possible that the Early Bronze Age population there (which is later in time than the Chalcolithic of Iran) was formed earlier than the Bronze Age, and may indeed be a source for the Iran_ChL. Resolving these ambiguities will require dense spatio-temporal sampling of the Near East to determine how populations changed in each location. With our available data we can only show that populations can be modeled as mixtures of each other, without making strong claims about the archaeological consequences of these models.

Because of these uncertainties, we sought to gain further insight into the derivation of Iran_ChL in terms of unambiguously earlier populations from the different regions of West Eurasia. Thus, we attempt to model Iran_ChL as a 3-way mixture, fixing Iran_N as one of the sources, and allowing the other two to be any pair of (EHG, WHG, CHG, Anatolia_N, Levant_N). Only two quadruples are consistent with rank=2 (Table S7.15). The quadruple (Iran_ChL, Iran_N, CHG, EHG) (P=0.245), but Iran_ChL cannot be modeled as a 3-way mixture of these sources as can be seen by the negative mixture proportions when attempting to fit such a model (Table S7.15).

The only possible way (P=0.221) to model Iran_ChL as a 3-way mixture involves 10.1±14.6% Iran_N, 69.6±14.7% CHG, and 20.3±3.0% Levant_N ancestry, which is robust when adding both EHG and WHG as additional outgroups (P=0.267; 16.7±10.3% Iran_N, 63.1±10.8% CHG, 20.2±2.8% Levant_N). It might seem strange that the successful model infers that most of the ancestry of Iran_ChL is from CHG and only a small fraction from Iran_N, which is closer temporally, geographically, and genetically (Fig. S7.11) to Iran_ChL than to CHG. However, recall that our estimate of the ancestry of CHG involved 71.6% Iran_N ancestry, so our result is consistent with Iran_ChL having the majority of its ancestry from a population like Iran_N. Again, we emphasize that our analysis shows that Iran_ChL can be modeled in terms of other populations in our dataset and should not be interpreted literally as a confluence of people from the Caucasus and the Levant into Chalcolithic Iran.

The high proportion of CHG ancestry in Iran_ChL led us to also consider a simpler 2-way mixture model involving only CHG and Levant_N (Table S7.16). This model is successful when all European hunter-gatherer groups and Natufians are introduced as outgroups to the Right set with a predicted 79.9±3.2% CHG and 20.1±3.2% Levant_N ancestry. This should not be interpreted as evidence of a migration from the Caucasus to Iran, as the geographical extent of populations like CHG is unknown. This is demonstrated by the fact that when Iran_N is added to the Right set the rank increases (Table

S7.16), and the model of Iran_ChL=CHG+Levant_N underpredicts shared genetic drift with Iran_N as f_4(Iran_ChL, Fitted Iran_ChL; Ust_Ishim, Iran_N) = -0.00229 (Z=-5.1).

Finally, we did an additional sanity check of our "best estimate" from Table S7.15. Using the estimates of Basal Eurasian ancestry from Supplementary Information, section 4, we predict 0.167*0.482+0.631*0.347+0.202*0.277=35.5% Basal Eurasian ancestry for Iran_ChL, which is consistent with the 31.7±5.1% estimate of section 4.

Figure S7.11: A close relationship between Neolithic and Chalcolithic Iran shown by the statistic f_4(Iran_ChL, *A*; Iran_N, Chimp). Neolithic Iran shares more alleles with Chalcolithic Iran than with any populations outside Iran.

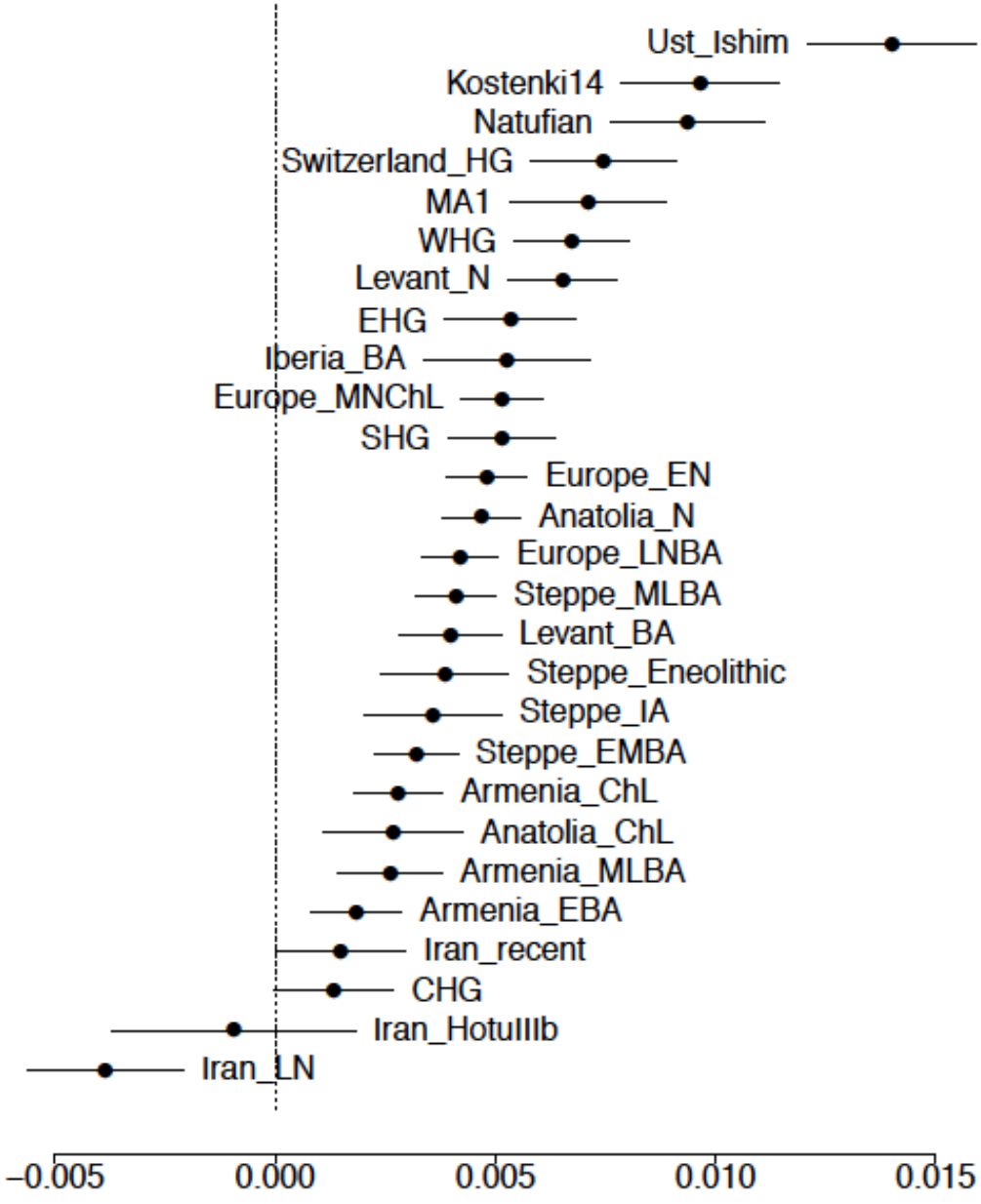


Figure S7.12: Testing for symmetry between Iran_N and Iran_ChL with the statistic f_4(Iran_N, Iran_ChL; *A*, Chimp). Chalcolithic Iran shares more alleles with all populations outside Iran.

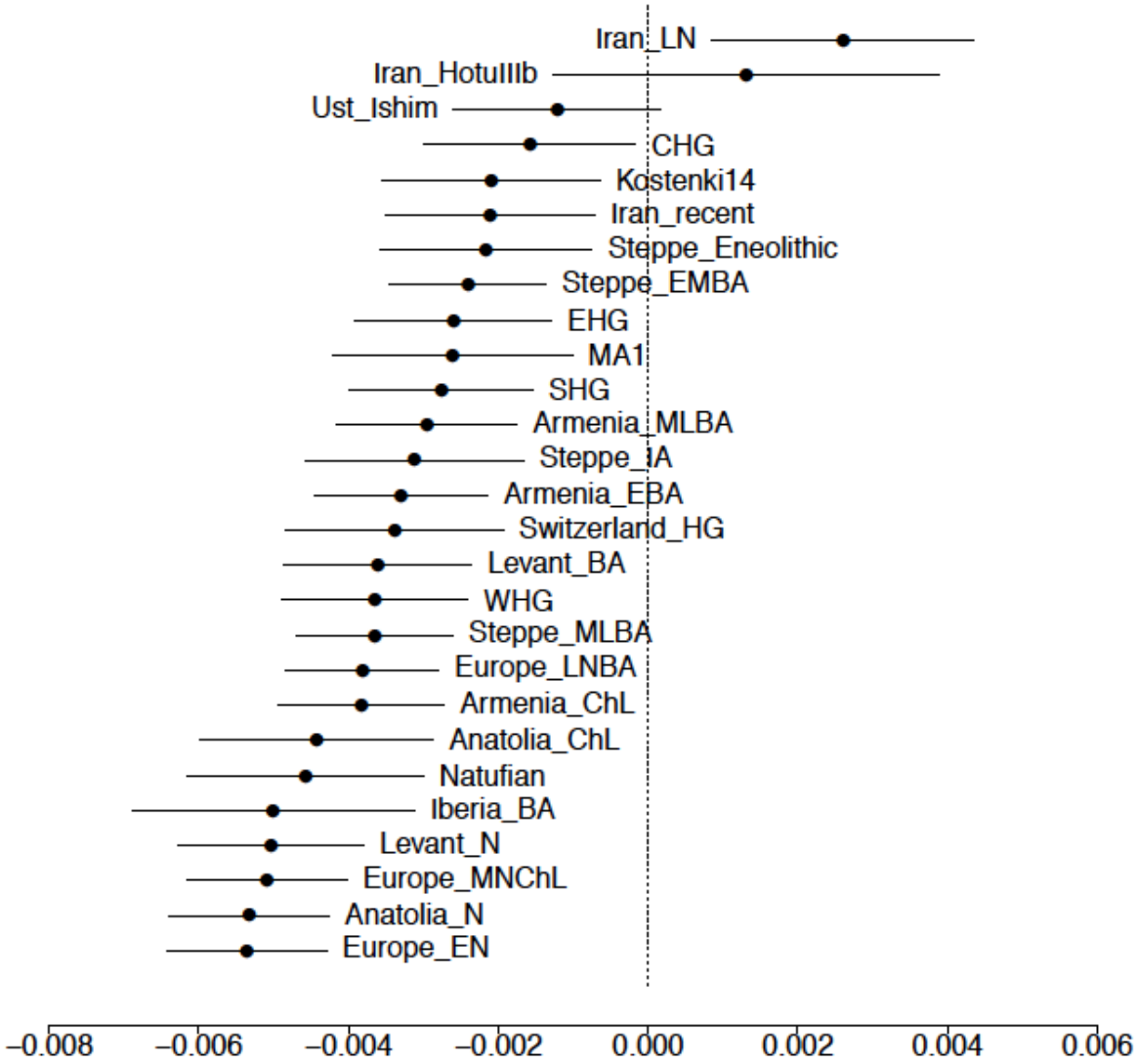


Figure S7.13: Testing for admixture in Iran_ChL with the statistic f_3(Iran_ChL; Iran_N, *A*)

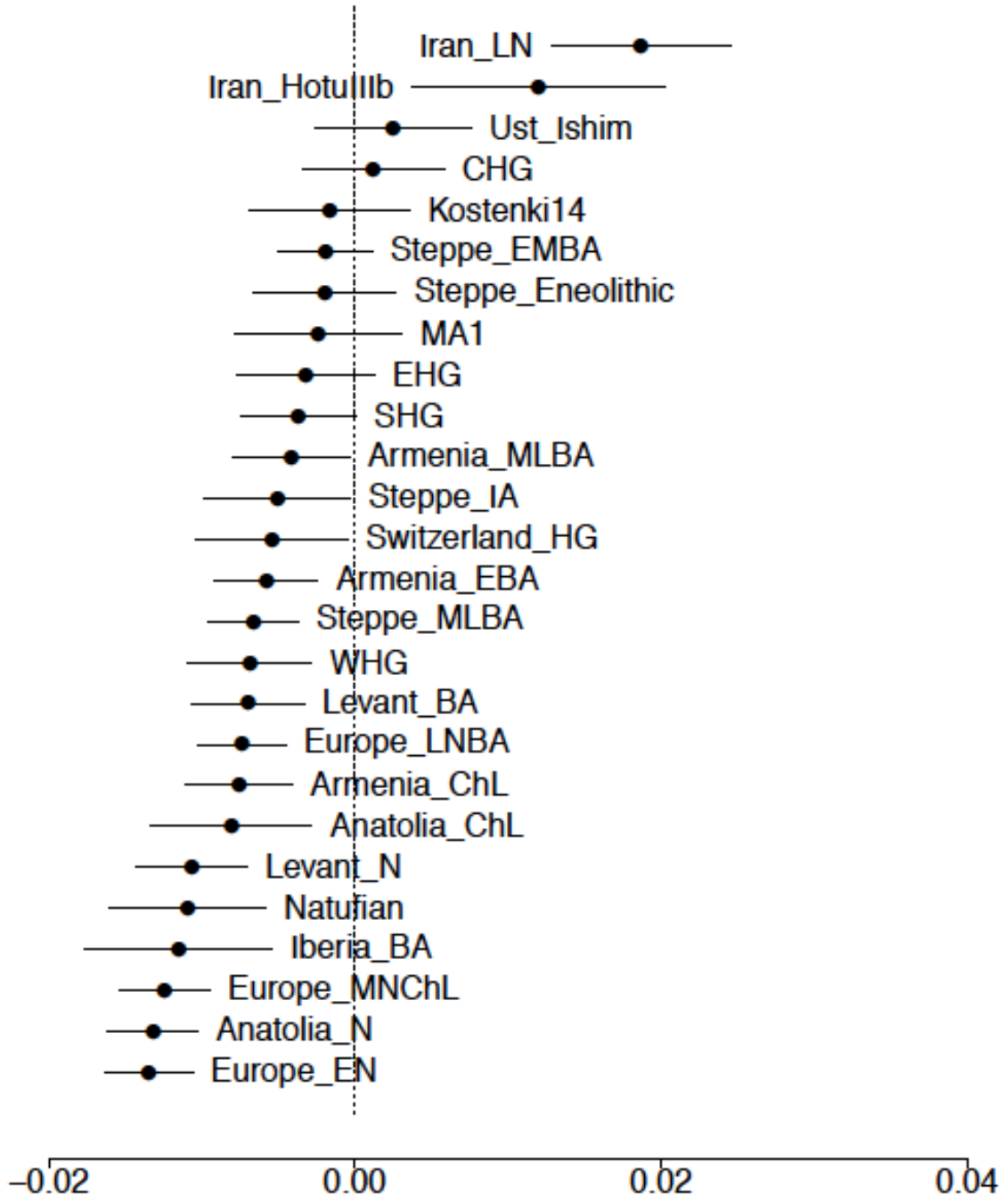


Table S7.14: Modeling Iran_ChL as a mixture of Iran_N and another population *A*.

			Mixture Proportions		
A	**Right**	**P-value for rank=1**	**Iran_N**	***A***	**Std. Error**
Anatolia_ChL	O9	3.22E-01	0.456	0.544	0.166
Anatolia_ChL	O9NS	4.53E-01	0.532	0.468	0.056
Anatolia_N	O9	3.66E-03	0.732	0.268	0.272
Anatolia_N	O9NS	4.66E-03	0.638	0.362	0.043
Armenia_ChL	O9	1.28E-01	0.618	0.382	0.094
Armenia_ChL	O9NS	7.81E-03	0.490	0.510	0.058
Armenia_EBA	O9	7.70E-01	0.293	0.707	0.129
Armenia_EBA	O9NS	4.37E-01	0.321	0.679	0.062
Armenia_MLBA	O9	1.41E-01	0.534	0.466	0.118
Armenia_MLBA	O9NS	1.15E-02	0.425	0.575	0.068
CHG	O9	3.31E-01	0.333	0.667	0.217
CHG	O9NS	5.42E-05	-0.418	1.418	0.585
EHG	O9	6.76E-02	0.906	0.094	0.030
EHG	O9NS	3.50E-10	0.932	0.068	0.029
Europe_EN	O9	5.27E-03	0.642	0.358	0.256
Europe_EN	O9NS	6.88E-04	0.665	0.335	0.044
Europe_LNBA	O9	1.22E-01	0.722	0.278	0.069
Europe_LNBA	O9NS	1.94E-07	0.775	0.225	0.044
Europe_MNChL	O9	1.79E-02	0.665	0.335	0.131
Europe_MNChL	O9NS	4.15E-06	0.779	0.221	0.037
Iberia_BA	O9	3.24E-01	0.593	0.407	0.113
Iberia_BA	O9NS	1.06E-05	0.729	0.271	0.052
Iran_LN	O9	9.30E-01	-105.971	106.971	34.482
Iran_LN	O9NS	6.10E-01	-1.032	2.032	1.959
Iran_HotuIIIb	O9	1.60E-01	-143.013	144.013	84.186
Iran_HotuIIIb	O9NS	3.06E-01	-68.720	69.720	25.978
Iran_recent	O9	3.28E-01	0.279	0.721	0.281
Iran_recent	O9NS	2.15E-01	0.217	0.783	0.114
Levant_BA	O9	3.36E-02	3052.349	-3051.349	26352.523
Levant_BA	O9NS	6.89E-05	0.622	0.378	0.058
Levant_N	O9	1.77E-02	1.327	-0.327	0.366
Levant_N	O9NS	1.39E-05	0.775	0.225	0.037
Natufian	O9	3.00E-02	1.286	-0.286	0.202
SHG	O9	1.05E-01	0.841	0.159	0.044
SHG	O9NS	1.04E-08	0.907	0.093	0.025
Steppe_EMBA	O9	7.44E-02	0.833	0.167	0.048
Steppe_EMBA	O9NS	5.61E-10	0.872	0.128	0.048
Steppe_Eneolithic	O9	5.11E-02	0.886	0.114	0.037
Steppe_Eneolithic	O9NS	2.97E-10	0.913	0.087	0.040
Steppe_IA	O9	1.59E-01	0.801	0.199	0.053
Steppe_IA	O9NS	5.16E-10	0.859	0.141	0.061
Steppe_MLBA	O9	1.44E-01	0.759	0.241	0.060
Steppe_MLBA	O9NS	1.04E-07	0.782	0.218	0.045
Switzerland_HG	O9	3.25E-02	0.841	0.159	0.061
WHG	O9	6.30E-02	0.820	0.180	0.056
WHG	O9NS	4.91E-09	0.941	0.059	0.017

Table S7.15: Modeling Iran_ChL as a mixture of Iran_N and a pair of populations from the set (WHG, EHG, CHG, Anatolia_N, Levant_N). We introduce Switzerland_HG (which is related to WHG) and Natufians (which is related to Levant_N) to better differentiate between the candidate source populations. Our "best estimate" is listed in bold.

				Mixture Proportions			Standard Errors		
A	***B***	**Outgroups**	**P-value for rank=2**	**Iran_N**	***A***	***B***	**Iran_N**	***A***	***B***
Anatolia_N	Levant_N	O9NS	2.64E-03	0.632	0.398	-0.030	0.047	0.104	0.076
CHG	Anatolia_N	O9NS	6.97E-03	0.381	0.329	0.290	0.286	0.342	0.069
CHG	Levant_N	O9NS	2.21E-01	0.101	0.696	0.203	0.146	0.147	0.030
CHG	**Levant_N**	**O9ENSW**	**2.67E-01**	**0.167**	**0.631**	**0.202**	**0.103**	**0.108**	**0.028**
EHG	Anatolia_N	O9NS	1.98E-02	0.602	0.052	0.347	0.043	0.022	0.042
EHG	CHG	O9NS	2.45E-01	-0.921	-0.282	2.203	0.845	0.187	1.018
EHG	Levant_N	O9NS	2.36E-02	0.635	0.110	0.255	0.041	0.022	0.032
WHG	Anatolia_N	O9NS	2.94E-03	0.627	-0.011	0.383	0.047	0.017	0.055
WHG	CHG	O9NS	1.85E-02	-0.713	-0.105	1.818	1.664	0.182	1.839
WHG	EHG	O9NS	2.20E-09	0.926	0.051	0.023	0.029	0.021	0.036
WHG	Levant_N	O9NS	1.79E-04	0.750	0.044	0.206	0.036	0.015	0.036

Table S7.16: Modeling Iran_ChL as a mixture of CHG and Levant_N. The simpler 2-way mixture CHG+Levant_N is consistent with the O9ENSW outgroups and parallels the position of Iran_ChL in the PCA (Fig. 1b) between CHG and the Levant. However, while CHG may encompass some ancestry related to Neolithic Iran, it cannot be the source of ancestry for Iran_ChL at the exclusion of Iran_N, as when Iran_N is added to the Outgroup set the CHG+Levant_N model fails.

		Mixture Proportions		
Outgroups	**P-value for rank=1**	**CHG**	**Levant_N**	**Std. Error**
O9	3.65E-01	0.825	0.175	0.092
O9ENSW	1.74E-01	0.799	0.201	0.032
O9EINSW	1.37E-05	0.870	0.130	0.033

Anatolia_ChL

A singleton individual from Chalcolithic Anatolia documents a population change that had apparently taken place since the Neolithic period samples also included in our study. Neolithic Anatolians cluster with Early European farmers and do not reside on the Near Eastern cline (Fig. 1b), but by the Chalcolithic period, it seems that they had shifted towards the northern end of the Near Eastern cline, occupied by present-day populations from western Asia (Extended Data Fig. 1). We carried out a similar analysis of Neolithic vs. Chalcolithic Anatolians as we did in Iran, studying statistics of the form f_4(Anatolia_ChL, *A*; Anatolia_N, Chimp) (Fig. S7.14) and f_4(Anatolia_N, Anatolia_ChL; *A*, Chimp) (Fig. S7.15), to understand population continuity or change in this far western part of Asia.

Neolithic Anatolia shares more alleles with Chalcolithic Anatolia than other populations (except European farmers descended from Anatolian ones), documenting a degree of continuity there. However, the two are not a clade, and Chalcolithic Anatolia differs from the Neolithic by sharing more alleles with "eastern" populations from the steppe, the Caucasus, and Iran. Thus, at the western end of western Asia, the population seems to become more "eastern" just as at the eastern end (Iran) it became more "western", confirming the visual impression from PCA (Fig. 1b) for highly differentiated Neolithic populations (Anatolia_N vs. Iran_N) but relatively similar Chalcolithic ones (Anatolia_ChL vs. Iran_ChL).

We first model Anatolia_ChL as a mix of Anatolia_N and a population *A* (Table S7.17). Only populations from Iran and Armenia work as sources of the input into Anatolia, confirming the visual impression from PCA. This input is quantified as at least 32.9±7.9% when Iran_ChL is used as a source population *A*.

Figure S7.14: The statistic f_4(Anatolia_ChL, *A*; Anatolia_N, Chimp) shows that Neolithic Anatolia shares more alleles with Chalcolithic Anatolia than with other ancient populations. The two exceptions are early farmers from Europe who share genetic drift with Anatolian Neolithic as most of their ancestry is from that population.

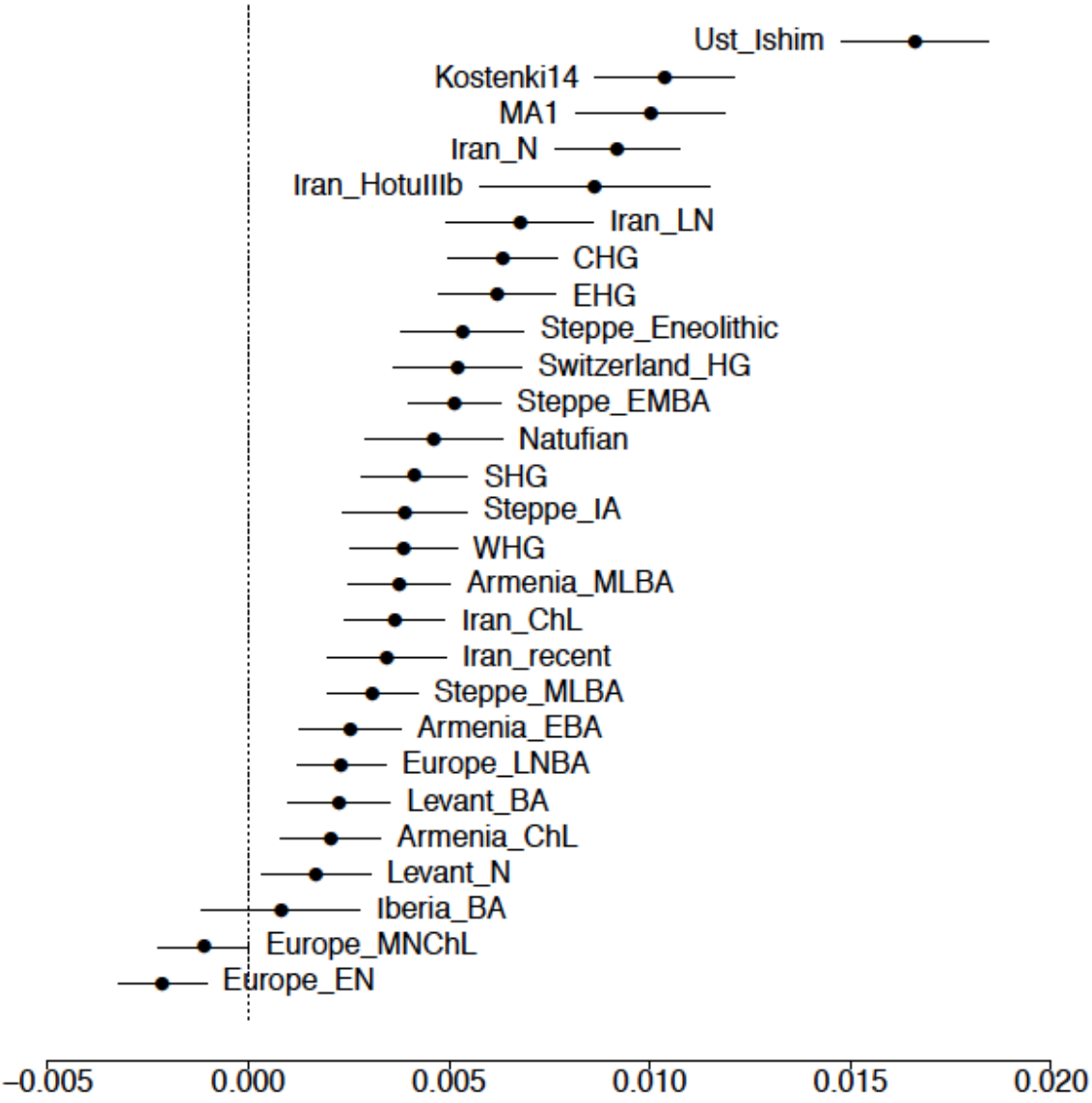


Figure S7.15: The statistic f_4(Anatolia_N, Anatolia_ChL; *A*, Chimp) becomes most negative for *A* being various populations east of Anatolia. Contrast with the analogous Fig. S7.12 for Iran that shows that Chalcolithic Iran is more "western" than Neolithic Iran.

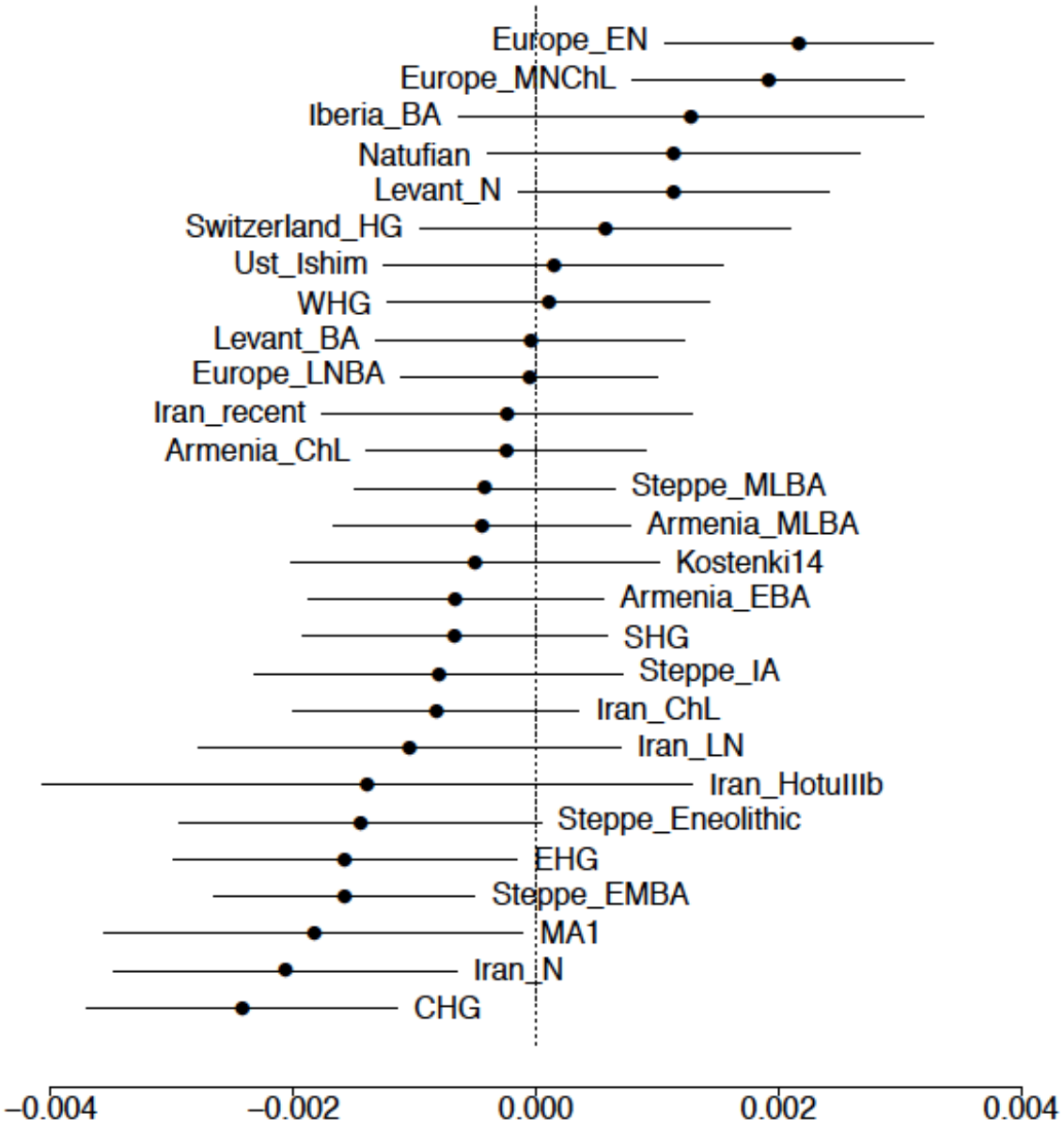


Table S7.17: Modeling Chalcolithic Anatolia as a mix of Neolithic Anatolia and a population *A*. This modeling suggests that Anatolia received genetic input from the east prior to ~3,950BCE. We use Iran_ChL as the "best" model as it has the lowest standard error.

			Mixture Proportions		
A	**Outgroups**	**P-value for rank=1**	**Anatolia_N**	***A***	**Std. Error**
Armenia_ChL	O9NW	1.36E-01	0.568	0.432	0.101
Armenia_EBA	O9NW	2.95E-01	0.543	0.457	0.097
Armenia_MLBA	O9NW	9.30E-02	0.624	0.376	0.093
CHG	O9NW	4.40E-02	0.751	0.249	0.064
EHG	O9NW	1.78E-03	0.926	0.074	0.028
Europe_EN	O9NW	1.27E-03	1.669	-0.669	0.345
Europe_LNBA	O9NW	2.33E-04	0.912	0.088	0.066
Europe_MNChL	O9NW	9.55E-04	1.189	-0.189	0.084
Iberia_BA	O9NW	1.29E-04	0.876	0.124	0.201
Iran_ChL	**O9NW**	**1.16E-01**	**0.671**	**0.329**	**0.075**
Iran_LN	O9NW	5.59E-03	0.815	0.185	0.065
Iran_HotuIIIb	O9NW	2.97E-02	0.831	0.169	0.054
Iran_N	O9NW	1.52E-02	0.826	0.174	0.050
Iran_recent	O9NW	1.10E-01	0.632	0.368	0.095
Levant_BA	O9NW	7.07E-03	0.536	0.464	0.169
Levant_N	O9NW	1.17E-04	1.036	-0.036	0.110
SHG	O9NW	1.62E-04	1.026	-0.026	0.028
Steppe_EMBA	O9NW	6.65E-03	0.853	0.147	0.045
Steppe_Eneolithic	O9NW	2.50E-03	0.902	0.098	0.036
Steppe_IA	O9NW	2.57E-02	0.826	0.174	0.048
Steppe_MLBA	O9NW	8.34E-04	0.866	0.134	0.060
Switzerland_HG	O9NW	2.66E-03	1.050	-0.050	0.018

Armenia_ChL

We do not have a pre-Chalcolithic sample from Armenia. We first model it as a 2-way mixture of any of WHG, EHG, CHG, Iran_N, Levant_N, Anatolia_N (Table S7.18), but we find no pair of these populations that could be ancestral to Armenia_ChL. We next model it as a 3-way mixture (Table S7.19), and determine that Armenia_ChL can be modeled as 18.3±1.5 EHG, 29.2±2.4% Iran_N, and 52.5±2.2% Anatolia_N. In the absence of a pre-Chalcolithic sample, we cannot be certain whether the Neolithic population of Armenia (which borders Anatolia from the east) was similar to that of Northwestern Anatolia and experienced gene flow from the east and north, or the reverse.

Table S7.18: Modeling Armenia_ChL as a 2-way mix of WHG, EHG, CHG, Anatolia_N, Iran_N, Levant_N. Only the triple (Armenia_ChL, CHG, Iran_N) is consistent with being derived from 2 streams. However, Armenia_ChL cannot be modeled as a mixture of CHG and Iran_N as it has a mixture proportion from CHG that is >1.

				Mixture Proportions		
A	***B***	**Outgroups**	**P-value for rank=1**	***A***	***B***	**Std. Error**
CHG	Anatolia_N	O9NS	1.06E-07	0.620	0.380	0.053
CHG	Iran_N	O9NS	9.04E-02	2.170	-1.170	0.363
CHG	Levant_N	O9NS	5.31E-09	0.792	0.208	0.036
EHG	Anatolia_N	O9NS	4.94E-14	0.233	0.767	0.017
EHG	CHG	O9NS	4.90E-14	0.059	0.941	0.035
EHG	Iran_N	O9NS	9.84E-28	0.224	0.776	0.034
EHG	Levant_N	O9NS	4.98E-20	0.389	0.611	0.016
Iran_N	Anatolia_N	O9NS	1.72E-21	0.471	0.529	0.054
Iran_N	Levant_N	O9NS	4.96E-31	0.761	0.239	0.046
Levant_N	Anatolia_N	O9NS	5.30E-23	-0.789	1.789	0.155
WHG	Anatolia_N	O9NS	8.60E-54	-0.006	1.006	0.015
WHG	CHG	O9NS	3.04E-13	0.048	0.952	0.018
WHG	EHG	O9NS	3.44E-173	-0.254	1.254	0.064
WHG	Iran_N	O9NS	8.07E-26	0.132	0.868	0.017
WHG	Levant_N	O9NS	4.60E-88	0.176	0.824	0.018

Table S7.19: Modeling Armenia_ChL as a 3-way mix of WHG, EHG, CHG, Anatolia_N, Iran_N, Levant_N. Several quadruples (Armenia_ChL, *A*, *B*, *C*) are consistent with rank=2; for some, such as (CHG, Iran_N, Anatolia_N) negative mixture proportions are inferred, hence they are not feasible. Three choices are feasible (EHG, CHG, Anatolia_N), (EHG, CHG, Levant_N), and (EHG, Iran_N, Anatolia_N). To decide between them we first add WHG to the outgroups which rejects (EHG, CHG, Anatolia_N). To decide between the other two that derive ancestry of Armenia_ChL from Levant_N or Anatolia_N, we add the Levant_N population to the outgroups for the model (EHG, Iran_N, Anatolia_N) or Anatolia_N for the model (EHG, CHG, Levant_N). Only the model (EHG, Iran_N, Anatolia_N) (marked in bold) is successful.

					Mixture Proportions			Standard Errors		
A	***B***	***C***	**Outgroups**	**P-value for rank=2**	***A***	***B***	***C***	***A***	***B***	***C***
CHG	Iran_N	Anatolia_N	O9NS	2.49E-01	1.363	-0.668	0.306	0.299	0.242	0.097
CHG	Iran_N	Levant_N	O9NS	7.88E-01	1.532	-0.766	0.234	0.218	0.208	0.059
CHG	Levant_N	Anatolia_N	O9NS	4.21E-08	0.600	-0.031	0.432	0.085	0.116	0.191
EHG	CHG	Anatolia_N	O9NS	6.90E-02	0.140	0.391	0.469	0.021	0.050	0.042
EHG	CHG	Anatolia_N	O9NSW	4.18E-02	0.124	0.441	0.435	0.019	0.042	0.038
EHG	CHG	Iran_N	O9NS	4.49E-01	-0.246	3.047	-1.801	0.206	1.136	0.954
EHG	CHG	Levant_N	O9NS	3.36E-01	0.195	0.483	0.322	0.024	0.045	0.030
EHG	CHG	Levant_N	O9NSW	4.24E-01	0.198	0.478	0.324	0.021	0.040	0.028
EHG	CHG	Levant_N	O9ANSW	1.51E-19	0.164	0.366	0.470	0.031	0.062	0.036
EHG	Iran_N	Anatolia_N	O9NS	1.21E-01	0.199	0.272	0.529	0.017	0.034	0.034
EHG	Iran_N	Anatolia_N	O9NSW	8.27E-02	0.191	0.304	0.505	0.017	0.027	0.030
EHG	**Iran_N**	**Anatolia_N**	**O9LNSW**	**8.91E-02**	**0.183**	**0.292**	**0.525**	**0.015**	**0.024**	**0.022**
EHG	Iran_N	Levant_N	O9NS	3.23E-03	0.287	0.344	0.369	0.019	0.035	0.028
EHG	Levant_N	Anatolia_N	O9NS	1.50E-14	0.226	-0.031	0.805	0.033	0.120	0.150
Iran_N	Levant_N	Anatolia_N	O9NS	1.28E-14	0.349	-0.420	1.071	0.058	0.111	0.145
WHG	CHG	Anatolia_N	O9NS	9.27E-08	0.020	0.623	0.357	0.015	0.053	0.057
WHG	CHG	Iran_N	O9NS	2.36E-01	-0.107	2.750	-1.643	0.097	0.892	0.810
WHG	CHG	Levant_N	O9NS	9.82E-07	0.060	0.719	0.221	0.015	0.037	0.033
WHG	EHG	Anatolia_N	O9NS	4.78E-05	-0.100	0.289	0.811	0.015	0.019	0.018
WHG	EHG	CHG	O9NS	1.06E-13	0.042	0.021	0.937	0.021	0.042	0.035
WHG	EHG	Iran_N	O9NS	9.07E-24	0.090	0.136	0.774	0.020	0.041	0.033
WHG	EHG	Levant_N	O9NS	4.85E-20	-0.029	0.413	0.616	0.019	0.023	0.017
WHG	Iran_N	Anatolia_N	O9NS	1.33E-18	0.062	0.516	0.423	0.017	0.056	0.065
WHG	Iran_N	Levant_N	O9NS	2.38E-19	0.121	0.645	0.234	0.015	0.040	0.038
WHG	Levant_N	Anatolia_N	O9NS	8.34E-08	-0.235	-1.283	2.518	0.045	0.210	0.248

Bronze Age Armenia: Armenia_EBA and Armenia_MLBA

The Bronze Age samples from Armenia cluster with the Chalcolithic ones, but do not overlap with them; during the Early Bronze Age there is an "eastward" shift away from Europe; and during the Middle/Late Bronze Age a partial "westward" counter-shift in the opposite direction. This is confirmed with f_4-statistics in Fig. S7.16, 17.

We can model Armenia_EBA as a 2-way mixture of 60.3±3.0% CHG and 39.7±3.0% Anatolia_N (Table S7.20). Recalling that our best model for Armenia_ChL involved 52.5±2.2% Anatolia_N ancestry, and this seems to correspond to the observed eastward shift in PCA. Recall that CHG itself can be modeled as a mixture of EHG and Iran_N, and thus our observation that Armenia_EBA can be modeled as an Anatolia_N+CHG mixture, while Armenia_ChL could be modeled as an Anatolia_N+EHG+Iran_N mix, does not indicate either the "disappearance" or the "appearance" of a wholly new population element in Armenia at the Bronze Age transition. It only indicates that while the CHG can account for the non-Anatolian element in the Early Bronze Age mix, it cannot do so in the Chalcolithic mix (Table S7.18).

We cannot model the Middle/Late Bronze Age population of Armenia as a 2-way mixture (Table S7.21). The 3-way mixture model with 10.5±2.0% EHG, 55.3±3.5% CHG, and 35.4±2.9% Anatolia_N fails marginally (P=0.0328) when several ancient outgroups are introduced to the Right set (Table S7.22). This suggests added complexity in this population, although it suggests an increase in European hunter-gatherer-related ancestry during the Middle/Late Bronze Age, consistent with the observed "westward" shift.

Figure S7.16: The statistic f_4(Armenia_ChL, Armenia_EBA; *A*, Chimp) becomes most positive for populations from Europe and most negative for populations from Iran and the Caucasus. This corresponds to the "eastward" shift in PCA (Fig. 1) of Early Bronze Age Armenia.

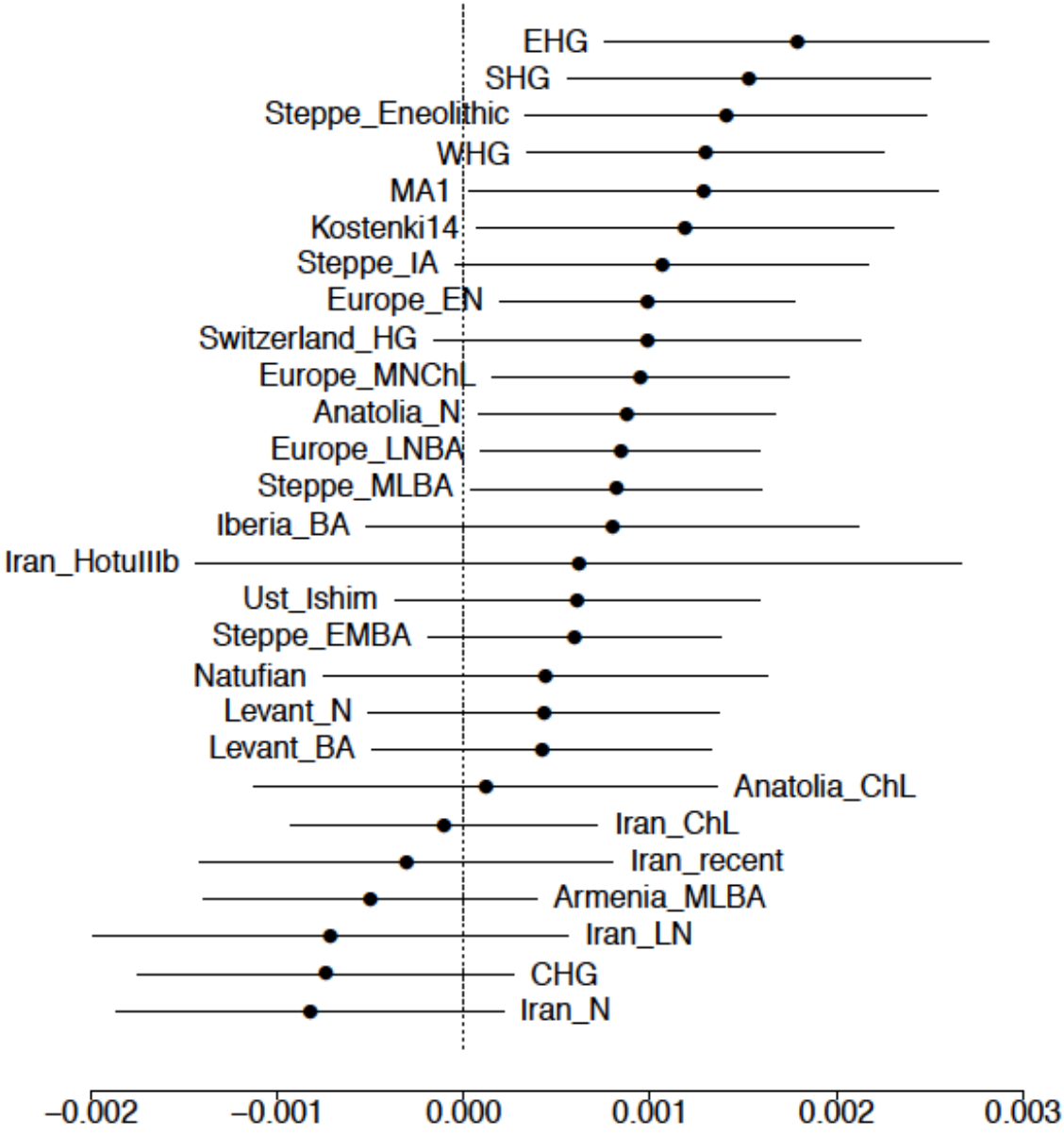


Figure S7.17: The statistic f_4(Armenia_EBA, Armenia_MLBA; *A*, Chimp) becomes most negative for populations from Europe and most positive for populations from the Near East. This corresponds to the counter-shift "westward" in PCA that reverses the Early Bronze Age eastward shift.

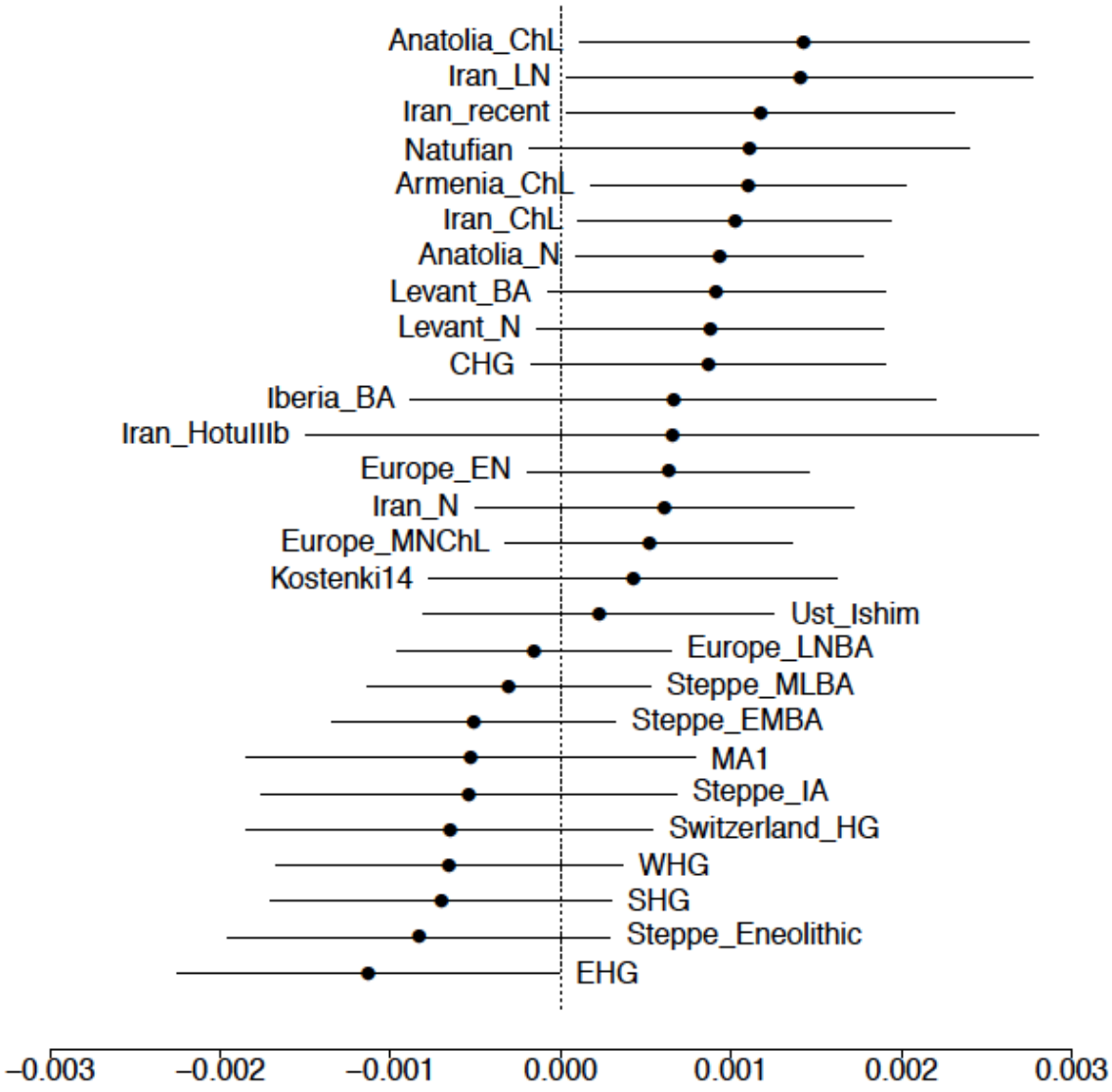


Table S7.20: Modeling Armenia_EBA as a 2-way mix of WHG, EHG, CHG, Anatolia_N, Iran_N, Levant_N. Only the pairs (CHG, Anatolia_N) and (CHG, Levant_N) can function as ancestral sources with the original set of outgroups (O9NS). Adding several other outgroups to the Right list, including Levant_N for the (CHG, Anatolia_N) model and Anatolia_N for the (CHG, Levant_N) model strongly rejects the latter. Thus, we can model Armenia_EBA only as a 2-way mix of CHG and Anatolia_N. The "best" model is marked in bold.

				Mixture Proportions		
A	***B***	**Outgroups**	**P-value for rank=1**	***A***	***B***	**Std. Error**
CHG	Anatolia_N	O9NS	0.228375543	0.587	0.413	0.048
CHG	**Anatolia_N**	**O9EILNSW**	**0.120301949**	**0.603**	**0.397**	**0.030**
CHG	Iran_N	O9NS	0.002686523	1.973	-0.973	0.434
CHG	Levant_N	O9NS	0.091447057	0.754	0.246	0.033
CHG	Levant_N	O9AEINSW	2.11E-09	0.629	0.371	0.021
EHG	Anatolia_N	O9NS	2.01E-25	0.155	0.845	0.021
EHG	CHG	O9NS	5.16E-08	-0.072	1.072	0.034
EHG	Iran_N	O9NS	3.11E-19	0.102	0.898	0.032
EHG	Levant_N	O9NS	6.01E-25	0.309	0.691	0.019
Iran_N	Anatolia_N	O9NS	1.54E-05	0.456	0.544	0.045
Iran_N	Levant_N	O9NS	8.31E-15	0.738	0.262	0.044
Levant_N	Anatolia_N	O9NS	6.94E-22	-0.651	1.651	0.133
WHG	Anatolia_N	O9NS	4.04E-37	-0.046	1.046	0.014
WHG	CHG	O9NS	6.07E-09	0.018	0.982	0.019
WHG	EHG	O9NS	1.03E-173	-0.288	1.288	0.062
WHG	Iran_N	O9NS	1.06E-14	0.107	0.893	0.018
WHG	Levant_N	O9NS	5.40E-69	0.121	0.879	0.017

Table S7.21: Modeling Armenia_MLBA as a 2-way mix of WHG, EHG, CHG, Anatolia_N, Iran_N, Levant_N. (Armenia_MLBA, *A*, *B*) is not consistent with only two ancestral populations for most choices of *A* and *B*. (Armenia_MLBA, CHG, Iran_N) is, but Armenia_MLBA cannot be modeled as a mixture of the other two populations.

				Mixture Proportions		
A	***B***	**Outgroups**	**P-value for rank=1**	***A***	***B***	**Std. Error**
CHG	Anatolia_N	O9NS	1.16E-05	0.735	0.265	0.062
CHG	Iran_N	O9NS	1.07E-01	1.895	-0.895	0.292
CHG	Levant_N	O9NS	3.29E-06	0.850	0.150	0.041
EHG	Anatolia_N	O9NS	8.07E-18	0.235	0.765	0.021
EHG	CHG	O9NS	7.74E-08	0.062	0.938	0.033
EHG	Iran_N	O9NS	2.50E-17	0.235	0.765	0.029
EHG	Levant_N	O9NS	2.30E-20	0.382	0.618	0.019
Iran_N	Anatolia_N	O9NS	6.48E-19	0.534	0.466	0.067
Iran_N	Levant_N	O9NS	8.32E-26	0.807	0.193	0.059
Levant_N	Anatolia_N	O9NS	2.98E-18	-0.920	1.920	0.161
WHG	Anatolia_N	O9NS	1.74E-47	0.002	0.998	0.016
WHG	CHG	O9NS	6.56E-07	0.055	0.945	0.018
WHG	EHG	O9NS	2.24E-153	-0.172	1.172	0.050
WHG	Iran_N	O9NS	1.55E-17	0.137	0.863	0.017
WHG	Levant_N	O9NS	2.78E-75	0.150	0.850	0.019

Table S7.22: Modeling Armenia_MLBA as a 3-way mix of WHG, EHG, CHG, Anatolia_N, Iran_N, Levant_N. The only plausible models with O9NS outgroups involve (EHG, CHG, Levant_N) (P=0.0503) and (EHG, CHG, Anatolia_N) (P=0.01). However, when we introduce Anatolia_N as an outgroup to the first of these models, it fails (P=9.42E-05), while the latter can withstand the addition of several outgroups (P=0.0328 for O9ILNSW). While this fails at the (P=0.05) level, it is the "best" one with 3 ancestral populations and is marked in bold.

					Mixture Proportions			*Standard Errors*		
A	***B***	***C***	**Outgroups**	**P-value for rank=2**	***A***	***B***	***C***	***A***	***B***	***C***
CHG	Iran_N	Anatolia_N	O9NS	1.99E-01	1.38	-0.608	0.228	0.277	0.226	0.096
CHG	Iran_N	Levant_N	O9NS	4.92E-01	1.491	-0.676	0.185	0.212	0.202	0.06
CHG	Levant_N	Anatolia_N	O9NS	5.22E-06	0.709	-0.047	0.338	0.108	0.162	0.258
EHG	CHG	Anatolia_N	O9NS	1.00E-02	0.118	0.526	0.356	0.025	0.066	0.055
EHG	CHG	Anatolia_N	O9LNS	1.72E-02	0.118	0.521	0.361	0.025	0.049	0.035
EHG	CHG	Anatolia_N	O9LNSW	2.34E-02	0.11	0.538	0.353	0.022	0.043	0.033
EHG	**CHG**	**Anatolia_N**	**O9ILNSW**	**3.28E-02**	**0.104**	**0.553**	**0.344**	**0.02**	**0.035**	**0.029**
EHG	CHG	Iran_N	O9NS	4.13E-01	-0.206	2.647	-1.441	0.183	0.952	0.787
EHG	CHG	Levant_N	O9NS	5.03E-02	0.161	0.576	0.263	0.028	0.058	0.039
EHG	CHG	Levant_N	O9ANS	9.42E-05	0.163	0.472	0.366	0.03	0.054	0.03
EHG	Iran_N	Anatolia_N	O9NS	4.70E-03	0.197	0.351	0.452	0.02	0.043	0.044
EHG	Iran_N	Levant_N	O9NS	5.13E-04	0.268	0.4	0.331	0.022	0.044	0.036
EHG	Levant_N	Anatolia_N	O9NS	1.91E-17	0.172	-0.272	1.1	0.091	0.382	0.471
Iran_N	Levant_N	Anatolia_N	O9NS	9.77E-11	0.388	-0.556	1.168	0.067	0.118	0.154
WHG	CHG	Anatolia_N	O9NS	2.27E-05	0.033	0.742	0.226	0.018	0.061	0.066
WHG	CHG	Iran_N	O9NS	1.25E-01	-0.055	2.2	-1.145	0.069	0.617	0.56
WHG	CHG	Levant_N	O9NS	1.31E-04	0.057	0.788	0.156	0.017	0.042	0.039
WHG	EHG	Anatolia_N	O9NS	4.18E-11	-0.097	0.294	0.803	0.017	0.024	0.021
WHG	EHG	CHG	O9NS	3.25E-07	0.048	0.026	0.926	0.02	0.037	0.032
WHG	EHG	Iran_N	O9NS	9.79E-14	0.086	0.16	0.754	0.019	0.033	0.028
WHG	EHG	Levant_N	O9NS	3.15E-20	-0.035	0.41	0.624	0.02	0.026	0.02
WHG	Iran_N	Anatolia_N	O9NS	9.84E-15	0.084	0.609	0.307	0.02	0.068	0.079
WHG	Iran_N	Levant_N	O9NS	3.32E-15	0.124	0.703	0.173	0.016	0.047	0.048
WHG	Levant_N	Anatolia_N	O9NS	1.30E-06	-0.249	-1.539	2.788	0.055	0.274	0.321

Levant_BA

Bronze Age Levant clusters with Neolithic Levant and Natufians (Fig. 1) but is shifted "northwards" along the Near Eastern cline. We confirm with f_4-statistics that this is a real shift and CHG and Neolithic/Mesolithic Iran shares more alleles with the Bronze Age than with the Neolithic Levant. We model Levant_BA as a mix of Levant_N and a population *A* (Table S7.23) and can model Levant_B as 55.7±2.8% Levant_N and 44.3±2.8% Iran_ChL. The appearance of ancestry related to Iran between the Neolithic and Bronze Age appears to parallel the evidence of Y-chromosomes where haplogroup J is absent in all Natufians and Neolithic male individuals but appears in the Bronze Age individuals and continues to be a major Y-chromosomal lineage of Near Eastern Levantine populations today. Again, we caution that the source of the Iran_ChL-related admixture need not be geographical present-day Iran, as we do not know the extent of populations similar to it in the ancient Near East. We also do not know whether the appearance of this admixture coincided with the beginning of the Bronze Age as there are intervening periods from the Levantine sequence between the Pre-Pottery Neolithic and the Bronze Age for which there are currently no data.

Figure S7.18: The statistic f_4(Levant_N, Levant_BA; *A*, Chimp) becomes most negative for populations from Iran and the Caucasus.

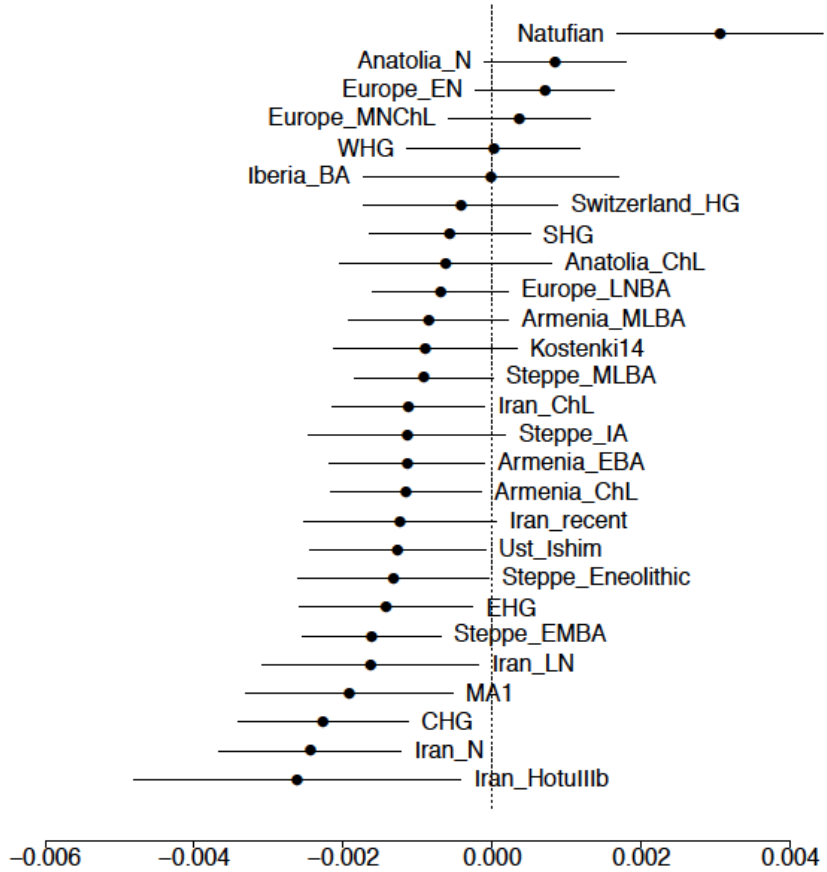


Table S7.23: Modeling Levant_BA as Levant_N and a population *A*. The Bronze Age Levant can be modeled as a mix of Neolithic Levant and all populations from Iran, as well as CHG and Armenia_EBA when O9NW are used as outgroups. Armenia_EBA is rejected as a source when Anatolia_N, EHG, and Switzerland_HG are added as outgroups. To distinguish between the remaining candidates, we add either Iran_N to the set of outgroups (rejecting CHG, P=0.0142), or CHG (rejecting all remaining populations except Iran_ChL and Iran_recent). Our "best" estimate in bold involves ancestry from Levant_N and Iran_ChL which is robust to a large number of included outgroups (P=0.806 for O9ACENSW).

			Mixture Proportions		
A	**Outgroups**	**P-value for rank=1**	**Levant_N**	***A***	**Std. Error**
Anatolia_ChL	O9NW	3.56E-04	0.213	0.787	0.117
Anatolia_N	O9NW	7.84E-16	0.580	0.420	0.109
Armenia_ChL	O9NW	1.54E-05	0.578	0.422	0.046
Armenia_EBA	O9NW	5.86E-02	0.489	0.511	0.045
Armenia_EBA	O9AENSW	3.54E-02	0.513	0.487	0.036
Armenia_MLBA	O9NW	3.55E-04	0.585	0.415	0.042
CHG	O9NW	6.16E-01	0.613	0.387	0.033
CHG	O9AENSW	5.88E-02	0.699	0.301	0.023
CHG	O9AEINSW	1.42E-02	0.681	0.319	0.022
EHG	O9NW	3.21E-13	0.891	0.109	0.019
Europe_EN	O9NW	2.71E-17	0.775	0.225	0.073
Europe_LNBA	O9NW	6.24E-14	0.817	0.183	0.033
Europe_MNChL	O9NW	8.52E-19	0.923	0.077	0.044
Iberia_BA	O9NW	1.28E-14	0.766	0.234	0.070
Iran_ChL	O9NW	7.96E-01	0.524	0.476	0.037
Iran_ChL	**O9ACENSW**	**8.06E-01**	**0.557**	**0.443**	**0.028**
Iran_LN	O9NW	2.71E-01	0.581	0.419	0.041
Iran_LN	O9ACENSW	2.19E-02	0.678	0.322	0.025
Iran_HotuIIIb	O9NW	6.13E-02	0.653	0.347	0.042
Iran_HotuIIIb	O9ACENSW	5.39E-03	0.729	0.271	0.022
Iran_N	O9NW	1.60E-01	0.657	0.343	0.031
Iran_N	O9ACENSW	2.98E-02	0.720	0.280	0.020
Iran_recent	O9NW	3.32E-01	0.531	0.469	0.044
Iran_recent	O9ACENSW	2.40E-01	0.552	0.448	0.033
SHG	O9NW	1.85E-18	0.957	0.043	0.019
Steppe_EMBA	O9NW	2.13E-09	0.799	0.201	0.026
Steppe_Eneolithic	O9NW	4.34E-12	0.862	0.138	0.023
Steppe_IA	O9NW	2.53E-05	0.756	0.244	0.028
Steppe_MLBA	O9NW	2.45E-12	0.802	0.198	0.031
Switzerland_HG	O9NW	1.69E-19	1.000	0.000	0.013

WHG

The Western European hunter-gatherers cluster with an Upper Paleolithic hunter-gatherer from Switzerland in PCA (Fig. 1) that is almost twice their age. This gives us the opportunity to study population change or continuity in Europe. Using f_4-statistics (Fig. S7.19) we can show that the ancient West Eurasians share more alleles with WHG than with the Switzerland_HG. For some of them, this can be explained by admixture with WHG-related populations. For example, early farmers of Europe admixed with WHG-related populations and are thus expected to share more genetic drift with them than with Switzerland_HG. However, significant statistics also exist for populations from the Near East, including both the Levant and Anatolia, and also from eastern Europe.

We model WHG as a mix of Switzerland_HG and a population *A* (Table S7.24). Remarkably while all triples (WHG, Switzerland_HG, *A*) are consistent with only two streams of ancestry, only for a subset of them is WHG modeled as a mix of Switzerland_HG and *A* with a feasible (≤1) proportion of ancestry from Switzerland_HG. This subset includes only populations with EHG ancestry. We verify visually that WHG can be modeled as EHG+Switzerland_HG by plotting statistics of the form f_4(WHG, EHG; O_2, O_3) vs. f_4(WHG, Switzerland_HG; O_2, O_3) for all pairs (O2, O3) of the outgroups. Such statistics (see ref. 1) are anti-correlated and form a line through the origin when WHG is a mix of EHG and Switzerland_HG (Fig. S7.20). Direct evidence for admixture in WHG is provided by the negative f_3(WHG; Switzerland_HG, EHG) = -0.01635 (Z=-7.5).

The modeling of ref. 1 made it unclear whether WHG was an admixed population or not (it was determined that at least one of WHG, EHG, MA1 must be). By making use of Switzerland_HG we can now determine that at least WHG is admixed. Our study has shown that most ancient West Eurasian populations can be modeled as 2- or 3-way mixtures of earlier populations of Late Upper Paleolithic, Mesolithic, or Early Neolithic provenance, and it appears that this process of admixture was taking place even earlier in both Europe (in the case of WHG) and the Near East (where admixture with Basal Eurasians is evident; Supplementary Information, section 4). As more ancient samples from the Upper Paleolithic become available for study, we anticipate that the earlier history of admixture of West Eurasian populations will be further clarified.

Table S7.24: Modeling WHG as a mix of Switzerland_HG and a population *A* with the O9 outgroups. We choose EHG as the "best" model as it has the lowest standard error.

		Mixture Proportions		
A	**P-value for rank=1**	**Switzerland_HG**	***A***	**Std. Error**
Anatolia_ChL	2.31E-01	1.048	-0.048	0.094
Anatolia_N	2.72E-01	1.078	-0.078	0.090
Armenia_ChL	2.01E-01	1.002	-0.002	0.104
Armenia_EBA	2.09E-01	1.023	-0.023	0.085
Armenia_MLBA	2.03E-01	0.977	0.023	0.082
CHG	2.11E-01	1.025	-0.025	0.090
EHG	**4.35E-01**	**0.933**	**0.067**	**0.037**
Europe_EN	2.51E-01	1.069	-0.069	0.097
Europe_LNBA	2.37E-01	0.913	0.087	0.102
Europe_MNChL	2.59E-01	1.088	-0.088	0.121
Iberia_BA	2.16E-01	1.039	-0.039	0.125
Iran_ChL	2.10E-01	1.022	-0.022	0.072
Iran_LN	2.34E-01	1.038	-0.038	0.063
Iran_HotuIIIb	2.00E-01	0.993	0.007	0.062
Iran_N	2.29E-01	1.036	-0.036	0.062
Iran_recent	2.01E-01	0.996	0.004	0.083
Levant_BA	2.67E-01	1.058	-0.058	0.067
Levant_N	3.21E-01	1.065	-0.065	0.057
Natufian	4.66E-01	1.082	-0.082	0.053
SHG	3.71E-01	0.868	0.132	0.081
Steppe_EMBA	3.69E-01	0.895	0.105	0.065
Steppe_Eneolithic	3.60E-01	0.924	0.076	0.049
Steppe_IA	6.18E-01	0.847	0.153	0.067
Steppe_MLBA	3.01E-01	0.885	0.115	0.087

Figure S7.19: Ancient West Eurasians share more alleles with WHG than with a ~14,000 year old Upper Paleolithic individual from Switzerland, as measured by the statistic f_4(Switzerland_HG, WHG; *A*, Chimp).

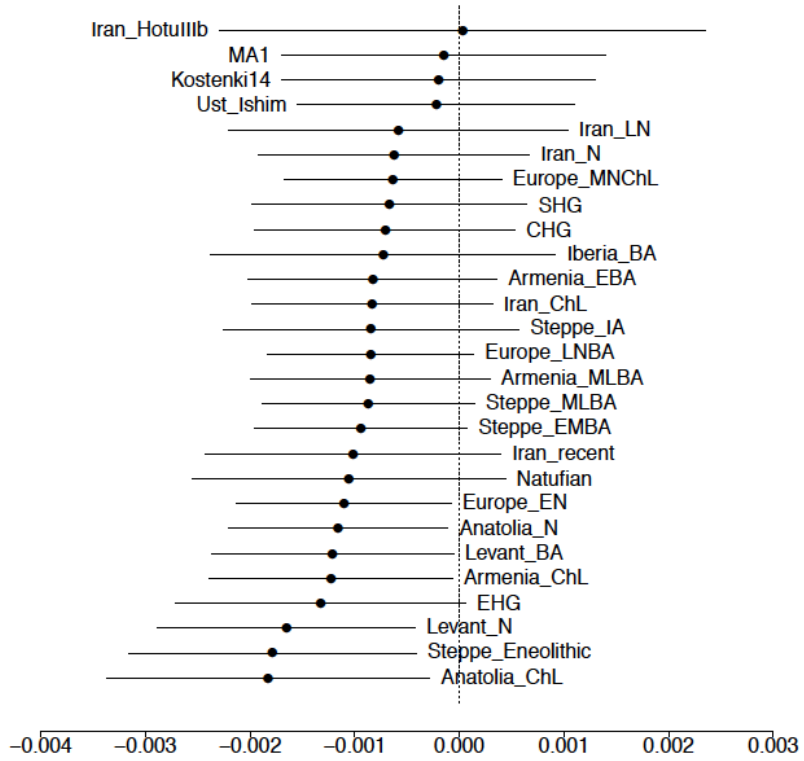
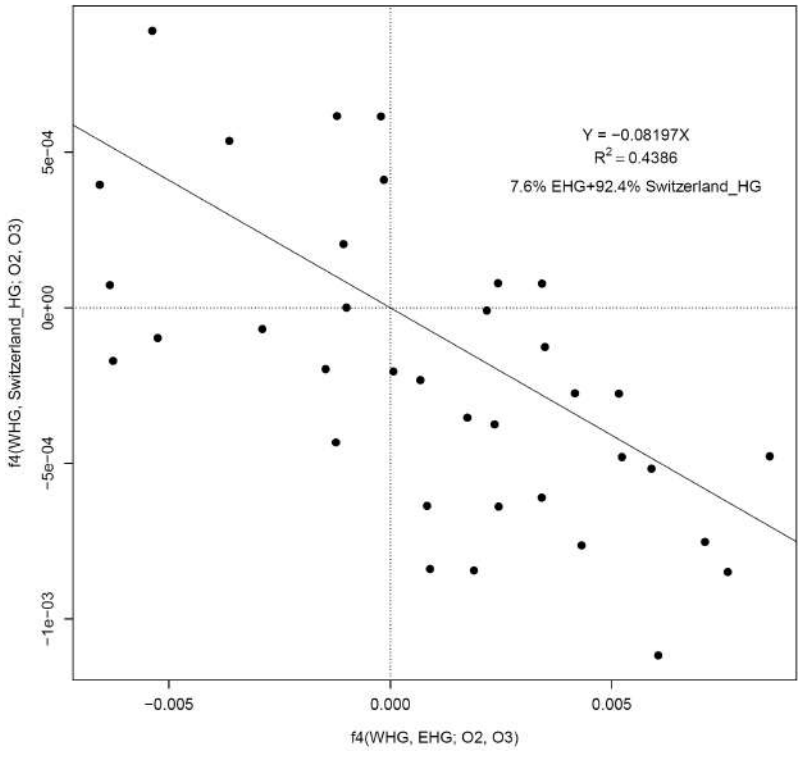


Figure S7.20: Anti-correlation between f_4(WHG, EHG; O_2, O_3) vs. f_4(WHG, Switzerland_HG; O_2, O_3) supports admixture in WHG from sources related to EHG and Switzerland_HG.



Summary of findings

Table S7.25 summarizes our best estimates of admixture in ancient West Eurasian populations. These are displayed graphically in Fig. 4.

Table S7.25: Mixture models for Ancient West Eurasian populations. This is a compilation of the best identified mixture models in this section (supported at the P>0.05 level). Mixture models supported at only a 0.05>P>0.01 level are listed in italics.

	Ancestral Source			Mixture Proportions			Standard Errors		
	A	B	C	A	B	C	A	B	C
Levant_N	Natufian	Anatolia_N		0.667	0.333		0.078	0.078	
Iran_HotuIIIb	Iran_N	EHG		0.906	0.094		0.044	0.044	
CHG	Iran_N	WHG	EHG	0.716	0.07	0.214	0.06	0.038	0.077
Anatolia_N	*Iran_N*	*Levant_N*	*WHG*	*0.387*	*0.339*	*0.274*	*0.134*	*0.137*	*0.021*
SHG	EHG	WHG		0.429	0.571		0.03	0.03	
Europe_EN	Anatolia_N	WHG		0.929	0.071		0.023	0.023	
Europe_MNChL	Anatolia_N	WHG		0.779	0.221		0.028	0.028	
Europe_LNBA	Europe_MNChL	Steppe_EMBA		0.469	0.531		0.016	0.016	
Steppe_EMBA	EHG	Iran_ChL		0.568	0.432		0.012	0.012	
Steppe_EMLBA	Steppe_EMBA	Europe_MNChL		0.684	0.315		0.021	0.021	
Iran_ChL	Iran_N	CHG	Levant_N	0.167	0.631	0.202	0.103	0.108	0.028
Anatolia_ChL	Anatolia_N	Iran_ChL		0.671	0.329		0.075	0.075	
Armenia_ChL	EHG	Iran_N	Anatolia_N	0.183	0.292	0.525	0.015	0.024	0.022
Armenia_EBA	CHG	Anatolia_N		0.603	0.397		0.03	0.03	
Armenia_MLBA	*EHG*	*CHG*	*Anatolia_N*	*0.104*	*0.553*	*0.344*	*0.02*	*0.035*	*0.029*
Levant_BA	Levant_N	Iran_ChL		0.557	0.443		0.028	0.028	
WHG	Switzerland_HG	EHG		0.933	0.067		0.037	0.037	

We tested the robustness of these models in two ways. First, we can use the mixture proportions to infer the amount of Basal Eurasian ancestry in different populations and compare it against our estimate from Supplementary Information, section 4. We find good agreement between the two (all are within |Z|<3 of the standard error of the Basal Eurasian estimate, and most are within |Z|<1) (Table S7.26). Second, we plot the PCA (Fig. 1) population means of each population and its ancestral sources together with the inferred mean when taking the weighted average of the sources according to the proportions of Table S7.25 (Extended Data Fig. 6). While the PCA (Fig. 1) is computed using present-day populations and it is not mathematically guaranteed that an ancient population will project near the weighted average of its ancestral populations, we nonetheless see that with some exceptions (such as Neolithic Anatolia, which is not well-modeled with existing populations, Table S7.25), actual and modeled means have good correspondence in PCA space.

Table S7.26: Correspondence of Basal Eurasian estimates from Supplementary Information, section 4 with those obtained by substituting a population with its ancestral components from Table S7.25.

Population	Basal Eurasian	Std. Error	Basal Eurasian from Mixture Model	Z
Levant_N	0.277	0.050	0.383	2.1
Iran_HotuIIIb	0.665	0.125	0.436	-1.8
CHG	0.347	0.058	0.345	0.0
Anatolia_N	0.253	0.039	0.280	0.7
Europe_EN	0.239	0.038	0.235	-0.1
Europe_MNChL	0.210	0.035	0.197	-0.4
Europe_LNBA	0.201	0.034	0.213	0.4
Steppe_EMBA	0.216	0.042	0.137	-1.9
Steppe_MLBA	0.182	0.040	0.214	0.8
Iran_ChL	0.317	0.051	0.355	0.7
Anatolia_ChL	0.084	0.072	0.274	2.6
Armenia_ChL	0.248	0.045	0.273	0.6
Armenia_EBA	0.276	0.052	0.309	0.6
Armenia_MLBA	0.305	0.055	0.279	-0.5
Levant_BA	0.313	0.049	0.295	-0.4

Conclusions

In this section we undertook the task of modeling the population history of ancient West Eurasians by means of the available samples. While it is important to remember that these samples are spatio-temporal snapshots of a genetic landscape that is only beginning to be explored, it is nonetheless encouraging that we could derive diverse ancient West Eurasian populations from each other in a master admixture graph (Table S7.25; Fig. 4) whose most interesting feature is the spread of ancestry from initially highly differentiated sources throughout the region of West Eurasia. Our results also hint that earlier admixture events contributed to the ancestry of the major Holocene ancestral sources (WHG, EHG, Iran_N, Levant_N). As the archaeogenetic record becomes more complete for earlier periods and regions outside West Eurasia, it may be possible to fully recreate the web of gene flows that began with the settlement of Eurasia by anatomically present humans.

References

1. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211, (2015).
2. Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370-374, (2012).
3. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449, (2014).
4. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118, (2014).
5. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87-91, (2014).
6. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, (2014).
7. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216-219, (2015).
8. Skoglund, P. *et al.* Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466-469, (2012).
9. Lipson, M. *et al.* Efficient moment-based inference of admixture parameters and sources of gene flow. *Mol. Biol. Evol.* **30**, 1788-1802, (2013).
10. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093, (2012).
11. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225-228, (2014).
12. Sánchez-Quinto, F. *et al.* Genomic Affinities of Two 7,000-Year-Old Iberian Hunter-Gatherers. *Curr. Biol.* **22**, 1494-1499, (2012).
13. Skoglund, P. *et al.* Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. *Science* **344**, 747-750, (2014).
14. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167-172, (2015).
15. Hofmanová, Z. *et al.* Early farmers from across Europe directly descended from Neolithic Aegeans. *bioRxiv*, (2015).
16. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912, (2015).
17. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503, (2015).

Supplementary Information 8

Population admixture in East Africa from the Levantine Neolithic

East Africans are admixed with West Eurasians[1,2], but the origin of the West Eurasian component in their ancestry has been obscure, as the strongest evidence of admixture among present-day East Africans is from Sardinians[2]. It has been proposed that an unsampled population from the Near East is the bearer of West Eurasian admixture into East Africa[2]. Here we systematically search for the source of this admixture. First, we verify that they do have such admixture by studying the statistic f_3(*East African*; Mota, *A*) where we iterate *A* to be any ancient or present-day population other than *East African*. This tests whether *East African* has admixture related to Mota[3], a ~4,500-year old sample from Ethiopia predating inferred dates of admixture in East Africa and present-day populations. For each East African population in our dataset, we present the most negative statistic in Table S8.1. It is important to recognize that the population *A* that contributes to the most negative f_3-statistic is not necessarily the one that is mostly closely related to the admixing population in a phylogenetic sense. A negative f_3-statistic proves a history of admixture related (perhaps deeply) to population *A*, but it is not always the case that the statistic is most negative when we use as population *A* the correct mixing population[4].

Table S8.1: East African populations are admixed

East African	*A*	f_3(*East African*; Mota, *A*)	Z
Luhya	Yoruba	-0.00236	-5.5
Luo	Yoruba	-0.00178	-4.2
Kikuyu	Saudi	-0.00526	-5.3
Jew_Ethiopian	Sardinian	-0.02488	-34.7
Somali	Saudi	-0.01776	-25.9
Oromo	Greek	-0.02348	-27.9
Masai	Saudi	-0.00951	-12.4
Dinka	Iran_HotuIIIb	0.02438	9.0
Datog	Sardinian	-0.01580	-15.1
Sandawe	Saudi	-0.00488	-6.6
Hadza	Iran_HotuIIIb	0.05270	17.4

Most populations have significantly negative f_3-statistics, proving that they are admixed. For some (like the Luhya), the source of the admixture seems to be related to West Africa, some of them (like the Dinka and the Hadza) do not have a negative f_3-statistic, but for most of the others, the minimizing population *A* is a West Eurasian one.

A way to identify the source of admixture is to use the qpAdm framework[5]. We use the following set of "Right" outgroups, using the notation introduced in Supplementary Information, section 7:

O8ENSW: Ust_Ishim, Kostenki14, MA1, Han, Papuan, Onge, Chukchi, Karitiana, EHG, Natufian, Switzerland_HG, WHG

This set includes the same populations used in Supplementary Information, section 7, but we exclude Africans (as we want to have an African reference population), and add EHG, Natufians, Switzerland_HG, and WHG as outgroups as they can help distinguish between different West Eurasian populations. The European hunter-gatherers are geographically improbable sources of admixture into East Africa, and the Natufians are temporally improbable as they predate by many thousands of years both the date of Mota and the inferred West Eurasian admixture into East Africa[1,2], and our data in Supplementary Information, section 7 suggests that they had been replaced in the Levant as early as the Pre-Pottery Neolithic period.

We show the P-value for rank=1 for the triple (*East African,* Mota, *A*) varying *A* to be any ancient population not included in the Right set (Table S8.2). We notice that rank=1 is not rejected regardless of the West Eurasian population *A* for four populations: Luhya, Luo, Dinka, and Hadza. We inspect the inferred mixture proportions for these four populations (Table S8.3) and see that these populations have very low estimates (or even negative) estimates of West Eurasian ancestry that are not statistically significantly different from zero. Thus, for example, the triple (Dinka, Mota, Levant_N) is consistent with 2 streams of ancestry not because Dinka is a mix of Mota and Levant_N but because (Dinka, Mota) and Levant_N represent two different ancestral populations. Because these four East African populations have negligible West Eurasian ancestry, it does not matter which West Eurasian population is chosen as a source of their ancestry (Table S8.2).

Considering the populations with non-negligible West Eurasian admixture, we observe that the Neolithic of the Levant is a good source for all but two of them, and the Bronze Age of the Levant for all but four of them (Table S8.2). This provides evidence that the source of the West Eurasian ancestry in East Africa is derived from the Levant and not from Europe or other parts of the Near East.

Table S8.2: Modeling East Africans in the *HO* dataset as a mix of Mota and an ancient population *A*. The P-value for rank=1 is shown for the triple Left=(*East African*, Mota, *A*) and Right=O8ENSW. P-values greater than 0.05 are highlighted in red.

	Anatolia_ChL	Anatolia_N	Armenia_ChL	Armenia_EBA	Armenia_MLBA	CHG	Europe_EN	Europe_LNBA	Europe_MNChL	Iberia_BA	Iran_ChL
Luhya	2.18E-01	1.95E-01	2.11E-01	2.12E-01	2.14E-01	2.00E-01	1.91E-01	1.92E-01	1.90E-01	1.89E-01	2.13E-01
Luo	1.83E-01	1.68E-01	1.77E-01	1.79E-01	1.81E-01	1.70E-01	1.66E-01	1.67E-01	1.70E-01	1.68E-01	1.80E-01
Kikuyu	1.18E-01	3.77E-02	2.71E-02	3.99E-02	3.03E-02	3.79E-03	1.81E-02	4.73E-03	2.62E-03	3.12E-03	3.99E-02
Jew_Ethiopian	8.09E-07	1.47E-10	2.71E-27	2.85E-23	2.56E-25	1.38E-39	4.42E-16	1.12E-33	8.79E-28	9.85E-23	8.07E-26
Somali	8.71E-05	9.11E-07	1.74E-15	4.60E-14	9.94E-16	6.10E-25	1.77E-10	1.22E-20	3.58E-18	6.06E-17	5.28E-16
Oromo	4.68E-05	1.56E-07	4.86E-14	2.26E-12	1.56E-14	3.11E-21	1.89E-10	9.15E-19	7.96E-17	4.14E-16	2.37E-14
Masai	2.13E-02	2.20E-03	1.17E-04	5.20E-04	1.52E-04	1.42E-06	3.00E-04	2.82E-06	4.60E-06	5.14E-06	3.99E-04
Dinka	1.42E-01	1.59E-01	1.45E-01	1.44E-01	1.43E-01	1.47E-01	1.65E-01	1.55E-01	1.84E-01	1.72E-01	1.44E-01
Datog	9.78E-03	6.32E-03	8.36E-06	6.37E-05	1.23E-05	3.49E-09	8.27E-04	2.29E-07	3.03E-06	1.74E-06	3.54E-05
Sandawe	1.08E-01	3.83E-02	4.41E-03	8.75E-03	3.72E-03	1.16E-04	9.35E-03	2.92E-04	4.32E-04	3.72E-04	7.40E-03
Hadza	5.37E-01	4.59E-01	4.87E-01	5.03E-01	5.03E-01	4.41E-01	4.25E-01	4.06E-01	3.70E-01	3.80E-01	5.21E-01

	Iran_LN	Iran_HotuIIb	Iran_N	Iran_recent	Levant_BA	Levant_N	SHG	Steppe_EMBA	Steppe_Eneolithic	Steppe_IA	Steppe_MLBA
Luhya	2.21E-01	2.55E-01	2.23E-01	2.09E-01	2.10E-01	2.08E-01	1.88E-01	2.09E-01	2.40E-01	2.19E-01	1.97E-01
Luo	1.76E-01	1.97E-01	1.79E-01	1.73E-01	1.77E-01	1.81E-01	1.67E-01	1.75E-01	1.93E-01	1.85E-01	1.69E-01
Kikuyu	1.92E-02	3.92E-03	3.76E-03	3.74E-02	1.34E-01	1.33E-01	8.04E-04	2.15E-03	2.08E-03	3.58E-03	4.93E-03
Jew_Ethiopian	2.27E-09	1.00E-15	2.32E-25	2.69E-18	4.20E-04	9.76E-04	5.95E-37	7.82E-46	7.41E-46	1.15E-42	9.97E-36
Somali	2.62E-11	3.66E-15	8.91E-26	8.80E-14	2.04E-02	2.18E-01	8.41E-24	2.57E-27	1.33E-27	1.68E-26	1.43E-21
Oromo	2.65E-09	4.47E-15	5.13E-21	1.51E-11	7.57E-04	1.86E-03	5.55E-22	1.23E-24	5.46E-25	4.17E-23	1.13E-19
Masai	4.17E-05	1.43E-07	1.96E-07	1.69E-04	3.19E-02	5.65E-02	8.67E-08	1.33E-07	7.63E-08	4.81E-07	1.95E-06
Dinka	1.44E-01	1.42E-01	1.42E-01	1.43E-01	1.46E-01	1.46E-01	1.71E-01	1.44E-01	1.44E-01	1.45E-01	1.50E-01
Datog	2.05E-06	1.15E-09	6.20E-10	4.08E-05	6.58E-02	1.07E-01	4.67E-09	3.47E-10	6.01E-11	1.53E-09	6.98E-08
Sandawe	1.24E-03	4.62E-05	4.81E-05	4.17E-03	2.40E-01	4.74E-01	2.82E-05	3.82E-05	2.65E-05	5.50E-05	2.30E-04
Hadza	4.95E-01	5.37E-01	4.84E-01	5.01E-01	5.38E-01	5.46E-01	3.61E-01	4.34E-01	4.65E-01	4.59E-01	4.20E-01

Table S8.3: Mixture Proportion of West Eurasian ancestry for East African populations in the *HO* dataset. Mixture proportions less than 3 standard errors from zero are highlighted.

		Luhya	Luo	Kikuyu	Jew_Ethiopian	Somali	Oromo	Masai	Dinka	Datog	Sandawe	Hadza
Mixture Proportions	Anatolia_ChL	0.019	0.016	0.125	0.437	0.328	0.329	0.173	-0.003	0.222	0.141	0.036
	Anatolia_N	0.009	0.006	0.114	0.408	0.313	0.310	0.159	-0.014	0.219	0.132	0.027
	Armenia_ChL	0.017	0.013	0.118	0.370	0.282	0.288	0.153	-0.006	0.202	0.126	0.032
	Armenia_EBA	0.019	0.015	0.133	0.438	0.326	0.336	0.177	-0.004	0.233	0.144	0.037
	Armenia_MLBA	0.018	0.015	0.122	0.389	0.294	0.294	0.158	-0.003	0.207	0.129	0.034
	CHG	0.015	0.010	0.131	0.419	0.304	0.317	0.168	-0.008	0.211	0.133	0.034
	Europe_EN	0.005	0.002	0.102	0.364	0.276	0.274	0.140	-0.015	0.197	0.116	0.022
	Europe_LNBA	0.005	0.003	0.079	0.246	0.192	0.193	0.102	-0.009	0.139	0.083	0.016
	Europe_MNChL	-0.003	-0.005	0.075	0.269	0.204	0.202	0.103	-0.017	0.149	0.084	0.011
	Iberia_BA	-0.001	-0.003	0.078	0.311	0.219	0.221	0.106	-0.015	0.150	0.086	0.013
	Iran_ChL	0.022	0.018	0.162	0.530	0.389	0.398	0.212	-0.004	0.278	0.173	0.046
	Iran_LN	0.037	0.024	0.208	0.938	0.699	0.707	0.291	-0.008	0.389	0.220	0.060
	Iran_HotuIIIb	0.041	0.032	0.166	0.974	0.619	0.569	0.222	0.007	0.287	0.164	0.056
	Iran_N	0.036	0.026	0.185	1.120	0.444	0.608	0.229	-0.001	0.314	0.179	0.056
	Iran_recent	0.021	0.015	0.146	0.534	0.365	0.385	0.189	-0.005	0.252	0.152	0.040
	Levant_BA	0.020	0.016	0.149	0.515	0.397	0.390	0.205	-0.008	0.275	0.172	0.042
	Levant_N	0.015	0.015	0.122	0.418	0.332	0.321	0.171	-0.007	0.223	0.146	0.035
	SHG	0.000	-0.002	0.043	0.140	0.109	0.107	0.055	-0.009	0.078	0.044	0.005
	Steppe_EMBA	0.011	0.008	0.074	0.182	0.156	0.156	0.089	-0.002	0.113	0.072	0.019
	Steppe_Eneolithic	0.014	0.010	0.057	0.140	0.120	0.120	0.068	0.001	0.081	0.054	0.017
	Steppe_IA	0.017	0.014	0.092	0.257	0.197	0.202	0.113	0.004	0.144	0.089	0.026
	Steppe_MLBA	0.008	0.005	0.079	0.235	0.184	0.185	0.099	-0.007	0.133	0.081	0.018
Standard Errors	Anatolia_ChL	0.023	0.023	0.022	0.024	0.021	0.024	0.021	0.023	0.023	0.020	0.023
	Anatolia_N	0.022	0.022	0.021	0.018	0.018	0.020	0.020	0.023	0.022	0.019	0.023
	Armenia_ChL	0.023	0.023	0.022	0.025	0.020	0.023	0.021	0.024	0.024	0.021	0.023
	Armenia_EBA	0.026	0.026	0.024	0.028	0.023	0.027	0.023	0.026	0.026	0.023	0.026
	Armenia_MLBA	0.024	0.024	0.023	0.024	0.021	0.024	0.021	0.024	0.024	0.021	0.024
	CHG	0.030	0.030	0.029	0.045	0.031	0.036	0.028	0.030	0.032	0.027	0.030
	Europe_EN	0.021	0.021	0.020	0.017	0.017	0.019	0.018	0.021	0.021	0.018	0.021
	Europe_LNBA	0.018	0.018	0.017	0.017	0.016	0.018	0.016	0.018	0.018	0.016	0.018
	Europe_MNChL	0.018	0.018	0.017	0.017	0.016	0.018	0.016	0.018	0.018	0.016	0.018
	Iberia_BA	0.018	0.018	0.018	0.023	0.019	0.022	0.017	0.018	0.020	0.017	0.018
	Iran_ChL	0.031	0.031	0.030	0.035	0.028	0.033	0.027	0.032	0.031	0.027	0.031
	Iran_LN	0.044	0.044	0.044	0.082	0.088	0.082	0.052	0.044	0.066	0.043	0.044
	Iran_HotuIIIb	0.038	0.040	0.044	0.169	0.117	0.105	0.062	0.039	0.084	0.047	0.038
	Iran_N	0.041	0.041	0.042	0.153	0.233	0.146	0.044	0.042	0.061	0.040	0.042
	Iran_recent	0.029	0.028	0.027	0.039	0.030	0.033	0.027	0.028	0.030	0.025	0.028
	Levant_BA	0.027	0.027	0.025	0.021	0.020	0.023	0.023	0.027	0.025	0.022	0.027
	Levant_N	0.022	0.022	0.021	0.017	0.017	0.020	0.019	0.022	0.021	0.019	0.022
	SHG	0.011	0.011	0.011	0.012	0.011	0.012	0.011	0.011	0.012	0.011	0.011
	Steppe_EMBA	0.018	0.017	0.017	0.018	0.016	0.018	0.016	0.017	0.018	0.016	0.018
	Steppe_Eneolithic	0.014	0.013	0.014	0.016	0.014	0.015	0.013	0.014	0.015	0.013	0.014
	Steppe_IA	0.021	0.021	0.021	0.029	0.022	0.024	0.020	0.021	0.023	0.020	0.021
	Steppe_MLBA	0.018	0.017	0.017	0.017	0.016	0.018	0.016	0.018	0.018	0.016	0.018

We also used the more varied Illumina genotype dataset of East African populations from Pagani et al.[1] All sources outside the Levant are strongly rejected (Table S8.4) except for the Levant Neolithic and Bronze Age. Inferred mixture proportions are shown in Table S8.5; all populations have non-zero West Eurasian admixture except the Anuak, Gumuz, and Sudanese where the mixture proportion is not more than three standard errors higher than zero; these populations may have negligible or no West Eurasian admixture. In Extended Data Figure 4 we show mixture proportions for populations of both the Human Origins and Pagani data sets with the Neolithic Levant as a source.

Table S8.4: Modeling East Africans in the Pagani dataset as a mix of Mota and an ancient population *A*. The P-value for rank=1 is shown for the triple Left=(*East African*, Mota, *A*) and Right=O8ENSW. P-values >0.05 are highlighted in red and >0.001 in yellow.

	Anatolia_ChL	Anatolia_N	Armenia_ChL	Armenia_EBA	Armenia_MLBA	CHG	Europe_EN	Europe_LNBA	Europe_MNChL	Iberia_BA	Iran_ChL
Afar	2.49E-10	6.74E-10	7.16E-28	1.74E-24	8.40E-29	2.09E-47	1.65E-17	7.46E-38	1.01E-31	1.17E-21	1.56E-24
Amhara	1.55E-11	1.02E-10	1.69E-31	7.96E-29	2.45E-33	2.58E-54	1.71E-19	3.98E-43	4.28E-36	1.09E-23	4.11E-29
Anuak	6.75E-01	7.07E-01	6.70E-01	6.67E-01	6.71E-01	6.75E-01	7.43E-01	7.37E-01	8.11E-01	7.83E-01	6.61E-01
Ariblacksmith	5.56E-02	9.16E-02	1.24E-02	1.25E-02	4.69E-03	1.14E-04	1.77E-02	3.22E-04	5.83E-04	7.65E-04	1.99E-02
Aricultivator	3.19E-02	7.99E-02	2.65E-03	4.39E-03	1.04E-03	2.96E-06	1.08E-02	3.31E-05	7.76E-05	1.42E-04	6.92E-03
Esomali	3.41E-09	2.60E-07	4.49E-18	5.27E-17	5.00E-20	1.23E-33	1.23E-12	5.90E-26	6.98E-23	3.78E-18	6.34E-17
Gumuz	7.26E-01	7.14E-01	7.26E-01	7.28E-01	7.25E-01	7.12E-01	7.07E-01	7.05E-01	7.08E-01	7.07E-01	7.37E-01
Oromo	1.06E-10	2.36E-09	7.45E-24	8.61E-23	2.71E-26	9.40E-43	3.61E-16	1.54E-33	2.49E-29	3.90E-21	9.18E-23
Somali	3.07E-09	1.53E-07	3.03E-20	8.29E-19	1.35E-21	1.21E-35	2.11E-12	5.75E-27	1.57E-22	3.75E-18	1.12E-18
Sudanese	7.48E-01	7.92E-01	7.41E-01	7.39E-01	7.43E-01	7.53E-01	8.28E-01	8.17E-01	8.91E-01	8.75E-01	7.29E-01
Tygray	1.80E-09	1.80E-09	1.60E-29	1.26E-27	7.06E-32	3.08E-53	1.69E-18	5.43E-42	5.68E-36	5.25E-23	8.56E-27
Wolayta	2.42E-06	1.08E-05	5.24E-12	5.02E-11	2.35E-13	2.29E-22	1.39E-09	3.70E-19	5.06E-17	1.34E-13	6.61E-11

	Iran_LN	Iran_HotuIIIb	Iran_N	Iran_recent	Levant_BA	Levant_N	SHG	Steppe_EMBA	Steppe_Eneolithic	Steppe_IA	Steppe_MLBA
Afar	1.02E-14	2.82E-14	8.32E-33	2.83E-17	2.68E-02	2.38E-03	3.33E-45	7.42E-52	5.22E-56	2.16E-46	2.99E-41
Amhara	1.88E-16	4.51E-16	4.29E-36	2.32E-20	1.89E-02	5.14E-03	4.08E-51	1.20E-59	4.32E-64	6.39E-53	4.86E-47
Anuak	6.49E-01	6.42E-01	6.49E-01	6.70E-01	6.72E-01	6.78E-01	8.07E-01	6.72E-01	6.54E-01	6.80E-01	7.12E-01
Ariblacksmith	2.81E-04	7.51E-05	2.29E-04	1.12E-02	4.80E-01	7.01E-01	1.00E-05	4.10E-05	7.77E-06	4.25E-05	2.16E-04
Aricultivator	1.17E-05	3.39E-06	5.87E-06	6.25E-03	5.07E-01	5.20E-01	1.35E-07	6.11E-07	2.90E-08	7.12E-07	1.48E-05
Esomali	8.60E-17	3.11E-14	3.57E-29	1.23E-14	4.39E-02	4.72E-02	9.41E-33	2.00E-34	8.10E-38	4.52E-33	4.09E-28
Gumuz	7.25E-01	7.67E-01	7.25E-01	7.28E-01	7.39E-01	7.40E-01	7.12E-01	7.14E-01	7.18E-01	7.13E-01	7.07E-01
Oromo	1.52E-16	2.34E-15	3.73E-34	8.52E-18	7.29E-03	4.89E-03	1.07E-41	2.96E-45	1.91E-49	2.83E-42	2.41E-36
Somali	5.66E-18	6.54E-15	1.03E-31	1.01E-14	9.30E-03	1.69E-02	9.31E-34	2.62E-37	3.63E-41	9.34E-35	1.93E-29
Sudanese	7.11E-01	6.60E-01	7.06E-01	7.39E-01	7.46E-01	7.47E-01	8.73E-01	7.33E-01	6.95E-01	7.51E-01	7.88E-01
Tygray	1.06E-14	4.62E-14	1.18E-33	1.71E-18	1.55E-01	1.66E-03	3.58E-50	6.60E-58	1.44E-61	2.35E-51	1.08E-45
Wolayta	2.16E-14	2.85E-12	2.22E-20	1.97E-09	5.67E-02	1.37E-02	1.23E-25	1.60E-25	2.83E-29	4.15E-24	1.43E-20

Table S8.5: Mixture proportions of West Eurasian ancestry for East African populations in the Pagani dataset. Mixture proportions less than 3 standard errors from zero are highlighted. Mixture proportions are valid for those combinations of *East African* populations (column) and source populations *A* where rank=1 is supported for the triple (*East African*, Mota, *A*) according to Table S8.4.

		Afar	Amhara	Anuak	Ariblacksmith	Aricultivator	Esomali	Gumuz	Oromo	Somali	Sudanese	Tygray	Wolayta
Mixture Proportions	Anatolia_ChL	0.432	0.431	-0.012	0.129	0.154	0.333	0.011	0.369	0.347	-0.020	0.449	0.297
	Anatolia_N	0.400	0.407	-0.016	0.127	0.153	0.323	0.007	0.349	0.331	-0.024	0.424	0.280
	Armenia_ChL	0.381	0.389	-0.011	0.124	0.148	0.298	0.011	0.328	0.304	-0.020	0.402	0.273
	Armenia_EBA	0.430	0.445	-0.012	0.138	0.166	0.339	0.013	0.373	0.344	-0.022	0.457	0.308
	Armenia_MLBA	0.385	0.395	-0.012	0.122	0.148	0.301	0.011	0.333	0.309	-0.020	0.410	0.276
	CHG	0.449	0.477	-0.015	0.128	0.160	0.330	0.008	0.378	0.347	-0.027	0.493	0.310
	Europe_EN	0.351	0.359	-0.019	0.111	0.137	0.283	0.003	0.308	0.293	-0.025	0.375	0.247
	Europe_LNBA	0.242	0.251	-0.016	0.079	0.100	0.195	0.001	0.215	0.205	-0.021	0.259	0.177
	Europe_MNChL	0.259	0.266	-0.021	0.081	0.103	0.207	-0.003	0.226	0.219	-0.026	0.276	0.183
	Iberia_BA	0.310	0.315	-0.019	0.083	0.106	0.237	-0.002	0.260	0.245	-0.025	0.325	0.202
	Iran_ChL	0.551	0.560	-0.012	0.170	0.204	0.427	0.018	0.470	0.434	-0.024	0.582	0.382
	Iran_LN	0.956	0.978	-0.009	0.192	0.235	0.739	0.020	0.846	0.788	-0.027	1.010	0.574
	Iran_HotuIIIb	0.865	0.882	0.005	0.152	0.199	0.645	0.026	0.745	0.724	-0.007	0.934	0.536
	Iran_N	0.868	0.943	-0.010	0.169	0.209	0.472	0.019	0.623	0.478	-0.025	0.968	0.416
	Iran_recent	0.482	0.483	-0.013	0.139	0.170	0.367	0.013	0.404	0.383	-0.021	0.509	0.326
	Levant_BA	0.504	0.511	-0.013	0.162	0.190	0.404	0.015	0.440	0.414	-0.023	0.527	0.353
	Levant_N	0.417	0.424	-0.012	0.139	0.161	0.337	0.013	0.365	0.347	-0.019	0.437	0.292
	SHG	0.132	0.137	-0.014	0.039	0.053	0.103	-0.003	0.114	0.111	-0.016	0.143	0.092
	Steppe_EMBA	0.185	0.193	-0.009	0.070	0.088	0.158	0.005	0.171	0.161	-0.014	0.203	0.150
	Steppe_Eneolithic	0.127	0.133	-0.005	0.048	0.061	0.108	0.005	0.118	0.110	-0.008	0.145	0.101
	Steppe_IA	0.266	0.283	-0.012	0.084	0.104	0.200	0.006	0.229	0.209	-0.019	0.291	0.193
	Steppe_MLBA	0.230	0.240	-0.014	0.078	0.098	0.187	0.003	0.206	0.196	-0.019	0.248	0.173
Standard Errors	Anatolia_ChL	0.023	0.021	0.021	0.019	0.018	0.021	0.020	0.020	0.021	0.021	0.021	0.020
	Anatolia_N	0.015	0.013	0.020	0.018	0.016	0.016	0.020	0.014	0.015	0.020	0.013	0.016
	Armenia_ChL	0.022	0.021	0.021	0.020	0.018	0.020	0.021	0.019	0.020	0.022	0.020	0.019
	Armenia_EBA	0.023	0.023	0.024	0.022	0.019	0.022	0.023	0.021	0.021	0.024	0.023	0.021
	Armenia_MLBA	0.023	0.021	0.022	0.020	0.018	0.021	0.021	0.020	0.021	0.022	0.022	0.020
	CHG	0.046	0.043	0.027	0.027	0.024	0.040	0.027	0.038	0.037	0.028	0.045	0.031
	Europe_EN	0.015	0.014	0.019	0.017	0.016	0.016	0.019	0.014	0.015	0.019	0.014	0.016
	Europe_LNBA	0.015	0.015	0.016	0.015	0.014	0.015	0.016	0.014	0.015	0.016	0.014	0.015
	Europe_MNChL	0.015	0.014	0.016	0.015	0.014	0.015	0.016	0.014	0.014	0.016	0.014	0.015
	Iberia_BA	0.020	0.018	0.016	0.016	0.015	0.019	0.016	0.017	0.018	0.016	0.019	0.017
	Iran_ChL	0.029	0.026	0.029	0.025	0.023	0.026	0.028	0.025	0.025	0.029	0.027	0.025
	Iran_LN	0.103	0.105	0.038	0.044	0.042	0.113	0.038	0.101	0.128	0.039	0.106	0.083
	Iran_HotuIIIb	0.123	0.129	0.032	0.044	0.042	0.105	0.032	0.116	0.118	0.033	0.129	0.087
	Iran_N	0.146	0.174	0.035	0.035	0.032	0.090	0.035	0.185	0.087	0.036	0.143	0.053
	Iran_recent	0.027	0.026	0.024	0.022	0.020	0.025	0.023	0.025	0.025	0.024	0.026	0.023
	Levant_BA	0.019	0.017	0.024	0.021	0.019	0.019	0.023	0.017	0.019	0.024	0.017	0.019
	Levant_N	0.016	0.015	0.020	0.018	0.016	0.016	0.019	0.015	0.016	0.020	0.015	0.016
	SHG	0.011	0.011	0.010	0.010	0.009	0.010	0.010	0.010	0.010	0.010	0.011	0.010
	Steppe_EMBA	0.017	0.017	0.016	0.016	0.014	0.016	0.016	0.016	0.016	0.016	0.016	0.016
	Steppe_Eneolithic	0.015	0.016	0.013	0.013	0.012	0.014	0.013	0.014	0.014	0.013	0.015	0.014
	Steppe_IA	0.028	0.029	0.019	0.019	0.017	0.023	0.019	0.024	0.023	0.020	0.028	0.022
	Steppe_MLBA	0.016	0.015	0.016	0.016	0.014	0.015	0.016	0.015	0.015	0.016	0.015	0.015

References

1 Pagani, L. *et al.* Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool. *American journal of human genetics* **91**, 83-96 (2012).

2 Pickrell, J. K. *et al.* Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl. Acad. Sci. USA* **111**, 2632–2637 (2014).

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

4 Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093 (2012).

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

Supplementary Information 9

Constraints on the origin of Ancestral North Indians

Previous studies[1,2] have uncovered evidence of admixture in South Asian populations from an "Ancestral North Indian" (ANI) source that is related to West Eurasian populations. It has been proposed that populations of the Caucasus such as Georgians were related to the ANI[1,3], a claim that has found additional support by the analysis of the Caucasus hunter-gatherers (CHG) from Georgia[4] which appear to be a source of ancestry for South Asian populations. However, South Asia is also linked to the Eurasian steppe[5] by the analysis of Y-chromosomes which detected the presence of Y-chromosome haplogroup R1a1a1b2-Z93 as a common element in ancient Bronze Age populations of eastern Europe[5], Mongolia[6], and Central[7]/South Asia[8] and which may mark spread of Indo-European languages eastward as suggested by the steppe origin theory of Indo-European languages[9].

The admixture history of ANI into the Indian subcontinent is likely to be complex, as there is evidence of more than one layer of admixture within the last 4,000 years[1]. Moreover, unlike Europe where a substantial number of pre-agricultural hunter-gatherers is available for study[4,5,10-16], the earliest population substratum of the "Ancestral South Indians" (ASI) is only indirectly known by its distant relationship[2] to the Onge hunter-gatherers from the Andaman Islands[17], a population that may be an imperfect proxy for the ASI. There is also evidence that Indian populations have ancestry related to Austroasiatic and Tibeto-Burman groups[2,18], although many of them can be modeled as a simpler mixture involving only the ANI-ASI ancestral populations[1,2].

In this section as in Supplementary Information, section 11 which deals with East Asians, for which there is also a lack of ancient DNA, we do not seek to solve the problem of their population history completely. Rather, we investigate whether the ANI could correspond to any of the numerous ancient populations of West Eurasia or to their combinations, using the qpAdm/qpWave methodology[12]. We use the following set of outgroups:

S7: Ust_Ishim, Kostenki14, MA1, Papuan, Chukchi, Karitiana, Mbuti

Note that this set corresponds to the O9 set introduced in Supplementary Information, section 7, but we have removed from it the Onge and the Han, as we wish to be able to use these populations to model admixture from ASI and/or East and Southeast Asian populations into South Asia.

We first set Left=(*South Asian*, *West Eurasian*, Onge, Han) and Right=S7, iterating between all *West Eurasian* and *South Asian* populations. If *South Asian* can be modeled as a 3-way mixture of *West Eurasian*, Onge, and Han, then the Left set will be consistent with 3 streams of ancestry. In particular, if the ANI portion of *South Asian* does not form a clade with *West Eurasian* and the

ASI/Southeast/East Asian portion of *South Asian* includes any ancestry that interacts with the S7 outgroups not via populations like the Onge, Han, then including *South Asian* into the set of outgroups will not be consistent with 3 streams of ancestry. Our goal is to reject possible sources of ancestry for the ANI, without making a strong claim about its identity. In Table S9.1 we show the p-value for rank=2 for different choices of *South Asian* and *West Eurasian*; this can be rejected in the great majority of cases, suggesting that no single population can account for ANI across South Asia.

We next tried all possible $\binom{20}{2} = 190$ pairs of the 20 *West Eurasian* populations of Table S9.1, thus Left=(*South Asian*, *West Eurasian*$_1$, *West Eurasian*$_2$, Onge, Han). South Asian populations could not be modeled as mixtures of most of the 190 quadruples (pairs of West Eurasian populations plus Onge and Han). In Table S9.2 we list the 49 pairs which resulted in successful modeling of at least 4 South Asian populations (p-value for rank=3 greater or equal to 0.05 and estimated mixture proportions in [0, 1] interval). A clear pattern emerges with the most successful pairs involving a population from Iran/the Near East and one from the Steppe/Eastern Europe. The fact that the Eastern European population is either EHG, Steppe_Eneolithic, or Steppe_EMBA is not surprising, as Steppe_Eneolithic and Steppe_EMBA groups are themselves mixtures of EHG and Iran-related populations as we have seen in Supplementary Information, section 7.

There is, however, uncertainty about the source of the Near Eastern-related component, as numerous South Asians can be modeled as having either Iran_N or Levant_N ancestry. We can introduce additional outgroups to the Right set to better distinguish between these choices. We add Natufians to the set of outgroups as it shares more alleles with Levant_N than with any other population (Supplementary Information, section 7) and can thus help differentiate between Levantine and other ancestry. We also add Switzerland_HG to help distinguish between the populations of Iran and those of Armenia and the Caucasus, all of which share more alleles with European hunter-gatherers than the Neolithic of Iran did (Supplementary Information, section 7). The new set of Right populations is:

S7NS: Ust_Ishim, Kostenki14, MA1, Papuan, Chukchi, Karitiana, Mbuti, Natufian, Switzerland_HG

In Table S9.3 we see that only 12 pairs of West Eurasian populations can be used to model at least 4 South Asian populations. All these pairs involve a steppe population and a population from Iran or CHG. Only steppe/Iran combinations can model all or nearly all South Asian populations successfully. In Table S9.4 we list the inferred mixture proportions for the top 3 pairs that can be used to model nearly all South Asian populations in our dataset. Ancestry from both Iran and Steppe-related sources is pervasive across South Asia. Estimated proportions of EHG ancestry are lower than Steppe_Eneolithic ancestry which are lower than Steppe_EMBA ancestry, consistent with the dilution

of EHG ancestry from a southern source[5] related to populations of Iran (Supplementary Information, section 7), thus more Steppe ancestry is required when the Steppe source has less EHG ancestry.

In Extended Data Fig. 5 we plot mixture proportions using Iran_N and Steppe_EMBA as sources as these source populations have the greatest sample sizes. Nonetheless, we do not claim that any particular pair of populations is directly ancestry to South Asian populations. This section places constraints on the possible sources of ANI ancestry in India by showing that single populations of ancient West Eurasia are not feasible sources (Table S9.1) while pairs of populations involving Steppe/Iran ancestry are (Tables S9.2 and S9.3). The geographical interpretation of this finding is unclear; mixtures of Steppe/Iran-related ancestry (such as Steppe_EMBA and Steppe_Eneolithic) were formed in West Eurasia and it is possible that a currently unsampled such mixed population is responsible for the ANI ancestry in South Asia. Alternatively, admixture may have taken place by combinations of Steppe/Iran-related groups in central or south Asia.

The analysis in this section reconciles the evidence presented in the first paragraph regarding the origin of the ANI by showing that is may be related both to "southern" populations related to Iran and the Caucasus and to "northern" steppe populations. Our results do not resolve the relationship between ANI and the origin of Indo-European speakers in South Asia, in the sense that they reveal that South Asian populations have ancestry both from regions related to the Eurasian steppe and ancient Iran, which is compatible with alternative homeland solutions[12]. West Eurasian-related ANI ancestry in South Asia may pre-date, coincide with, or postdate Indo-European dispersals, although a partial link between the two is suggested by the evidence for Bronze Age admixture in India[1] that contributed a large portion of ancestry especially in Indo-European speakers[1,2] whose magnitude would be compatible with major linguistic change. However, ANI ancestry related to both ancient Iran and the steppe is found across South Asia (Table S9.4) making it difficult to associate it strongly with any particular language family (Indo-European or otherwise). Nonetheless, the fact that we can reject West Eurasian population sources from Anatolia, mainland Europe, and the Levant diminishes the likelihood that these areas were sources of Indo-European (or other) languages in South Asia. While the Early/Middle Bronze Age 'Yamnaya'-related group (Steppe_EMBA) is a good genetic match (together with Neolithic Iran) for ANI, the later Middle/Late Bronze Age steppe population (Steppe_MLBA) is not. Steppe_MLBA includes Sintashta and Andronovo populations who have been proposed as identical to or related to ancestral Indo-Iranians[9,19], as well as the Srubnaya from eastern Europe which are related to South Asians by their possession of Y-chromosome haplogroup R1a1a1b2-Z93[5]. A useful direction of future research is a more comprehensive sampling of ancient DNA from steppe populations, as well as populations of central Asia (east of Iran and south of the steppe), which may reveal more proximate sources of the ANI than the ones considered here, and of South Asia to determine the trajectory of population change in the area directly.

Table S9.1: Modeling *South Asian* populations as a mix of *West Eurasian* plus Onge and Han using Right=S7. We show the p-value for rank=2 for Left=(*South Asian*, *West Eurasian*, Onge, Han). No *West Eurasian* populations appear to be the single source of ANI ancestry across South Asia.

	Anatolia_ChL	Anatolia_N	Armenia_ChL	Armenia_EBA	Armenia_MLBA	CHG	EHG	Europe_EN	Europe_LNBA	Europe_MNChL	Iran_ChL	Iran_LN	Iran_N	Levant_BA	Levant_N	SHG	Steppe_EMBA	Steppe_Eneolithic	Steppe_MLBA	WHG
Balochi	2.8E-03	1.4E-33	9.1E-06	7.6E-04	3.4E-01	3.3E-03	5.2E-64	5.7E-37	1.3E-32	1.3E-34	2.3E-05	1.8E-06	1.7E-07	1.7E-16	4.0E-26	2.7E-51	1.7E-56	1.5E-44	1.3E-33	1.9E-57
Bengali	2.0E-04	2.9E-14	8.7E-03	4.0E-05	2.7E-02	1.3E-04	3.7E-16	3.3E-14	6.1E-03	4.6E-12	8.6E-08	1.1E-08	9.1E-11	1.8E-13	1.1E-19	2.0E-10	2.0E-05	2.0E-10	1.4E-02	1.0E-17
Brahmin_Tiwari	1.1E-09	1.2E-53	3.6E-10	1.8E-17	3.4E-10	8.9E-13	1.4E-26	1.8E-54	3.3E-09	1.2E-44	1.7E-27	4.8E-19	3.9E-22	3.1E-43	6.8E-56	4.8E-15	1.6E-07	4.1E-14	5.4E-04	5.1E-35
Brahui	2.3E-03	3.3E-33	5.7E-08	8.6E-05	5.1E-02	1.3E-03	6.0E-66	1.1E-33	1.4E-39	2.4E-36	3.2E-05	1.2E-05	8.4E-07	2.0E-14	1.6E-23	3.0E-53	2.2E-59	1.5E-49	2.6E-40	1.5E-58
Burusho	5.9E-09	9.5E-60	3.0E-09	2.1E-16	1.7E-09	1.9E-11	8.3E-37	8.8E-61	1.1E-11	1.0E-49	1.5E-28	1.2E-16	3.9E-21	5.3E-43	2.8E-53	2.2E-22	4.7E-14	5.8E-20	7.9E-07	8.2E-41
GujaratiA	2.3E-06	5.8E-32	7.8E-05	8.9E-12	5.6E-06	1.4E-08	1.1E-29	4.4E-29	3.8E-03	3.9E-24	9.6E-18	4.2E-14	1.6E-18	1.1E-29	4.0E-41	2.4E-15	3.2E-10	3.7E-17	1.1E-02	1.1E-26
GujaratiB	2.1E-06	2.3E-23	4.0E-04	1.9E-08	4.3E-04	8.3E-07	1.8E-24	4.3E-24	2.0E-03	6.2E-19	9.2E-13	2.9E-13	2.7E-14	1.7E-22	2.2E-30	1.5E-13	7.8E-08	6.2E-15	1.6E-02	5.3E-24
GujaratiC	9.7E-07	3.2E-24	7.7E-06	2.2E-09	2.3E-05	3.6E-08	8.0E-18	4.9E-24	2.6E-04	9.4E-20	4.5E-13	4.1E-14	1.3E-15	3.0E-22	3.7E-30	9.7E-11	8.3E-05	1.5E-10	1.0E-02	3.7E-23
GujaratiD	4.4E-04	9.3E-15	3.1E-02	2.6E-05	3.1E-02	1.4E-04	3.2E-17	1.7E-14	3.5E-02	2.8E-11	1.8E-07	6.9E-10	7.2E-11	1.1E-14	1.6E-20	3.1E-09	1.1E-05	3.0E-11	5.0E-02	1.0E-15
Jew_Cochin	3.7E-02	1.4E-09	4.5E-03	1.9E-01	6.6E-01	3.1E-01	4.1E-42	1.1E-10	1.3E-12	1.7E-12	1.1E-01	2.2E-03	4.7E-04	6.9E-06	6.2E-12	3.6E-29	9.9E-27	6.6E-33	2.0E-15	1.7E-30
Kalash	2.2E-08	1.1E-51	1.3E-08	2.1E-15	3.4E-09	3.4E-12	1.4E-30	5.1E-50	5.9E-09	2.8E-39	1.0E-24	1.5E-16	4.0E-20	1.1E-37	5.2E-49	1.1E-17	1.1E-11	2.8E-18	3.7E-05	2.3E-36
Kharia	2.1E-01	4.2E-02	4.0E-01	4.6E-01	7.1E-01	3.9E-01	6.9E-04	3.5E-02	2.2E-01	3.6E-02	4.8E-01	7.7E-02	1.6E-01	9.7E-02	2.3E-02	6.3E-03	8.8E-02	6.0E-03	2.3E-01	3.4E-04
Kusunda	2.1E-01	1.9E-02	2.6E-01	2.7E-01	3.9E-01	1.4E-01	5.6E-05	2.0E-02	1.6E-01	2.7E-02	2.2E-01	1.1E-02	2.2E-02	3.1E-02	5.2E-03	1.9E-03	3.6E-02	1.3E-03	1.4E-01	6.8E-05
Lodhi	9.6E-07	8.8E-23	2.2E-05	2.2E-09	1.2E-04	1.1E-07	2.3E-16	8.2E-23	6.5E-05	3.9E-19	2.5E-13	5.6E-13	2.3E-15	6.2E-21	2.2E-27	2.2E-11	3.2E-05	3.2E-11	2.6E-03	2.7E-22
Makrani	1.3E-04	2.1E-46	1.2E-19	4.5E-07	2.1E-07	2.6E-05	3.5E-88	8.8E-49	4.7E-80	9.2E-58	6.3E-04	2.8E-03	7.2E-03	4.3E-10	3.3E-15	2.5E-83	5.9E-90	2.1E-68	3.8E-76	6.6E-81
Mala	1.4E-03	1.9E-09	4.8E-02	1.4E-03	1.6E-01	2.2E-03	2.4E-12	2.2E-09	3.2E-02	5.6E-08	1.7E-04	6.5E-08	4.1E-08	4.0E-09	1.7E-13	1.2E-07	2.5E-04	2.7E-08	4.1E-02	6.8E-13
Pathan	7.9E-07	3.1E-56	1.3E-06	9.1E-13	4.7E-06	4.0E-09	3.5E-44	5.6E-54	1.3E-13	6.6E-45	1.2E-21	1.6E-13	4.5E-18	6.6E-36	3.2E-50	3.0E-28	4.4E-24	1.3E-25	2.7E-10	1.3E-45
Punjabi	8.6E-06	3.9E-20	4.1E-04	8.8E-07	2.1E-03	2.4E-05	2.8E-22	7.9E-20	2.0E-04	7.2E-17	1.8E-10	1.9E-11	2.3E-13	3.4E-18	6.7E-26	4.8E-14	2.8E-08	6.3E-14	4.2E-04	8.9E-23
Sindhi	1.3E-05	1.5E-44	1.2E-05	1.4E-09	1.9E-04	4.4E-07	6.4E-42	1.5E-43	9.4E-15	3.8E-40	1.9E-16	1.2E-12	9.9E-15	1.2E-27	3.3E-39	5.4E-30	2.0E-25	7.4E-28	4.1E-12	2.4E-44
Vishwabrahmin	1.5E-05	1.0E-15	2.4E-04	1.9E-05	9.2E-03	3.6E-05	4.4E-16	7.4E-16	1.9E-05	1.2E-14	5.7E-07	6.7E-09	3.8E-09	9.6E-13	1.6E-17	9.2E-12	1.2E-06	6.6E-11	1.1E-04	1.2E-20

Table S9.2: Pairs of West Eurasian populations (out of $\binom{20}{2} = 190$ choices) and the number of South Asian populations (out of a total of 20) that could be modeled as 4-way mixtures involving each pair plus Onge and Han using the S7 outgroups. We show only population pairs that can be used to model at least 4 South Asian populations (for Left=(*South Asian*, *West Eurasian*$_1$, *West Eurasian*$_2$, Onge, Han) we have p-value for rank=3 greater or equal to 0.05 and estimated mixture proportions in [0, 1] interval).

Number of South Asian populations modeled as mixtures of ***Pair*** , Onge, and Han	***Pair***
20	Iran_N:Steppe_Eneolithic
20	Iran_N:Steppe_EMBA
20	Iran_LN:Steppe_Eneolithic
20	EHG:Iran_LN
19	Iran_LN:Steppe_EMBA
19	EHG:Iran_N
18	Iran_ChL:Steppe_Eneolithic
17	Iran_ChL:Steppe_EMBA
16	Levant_N:Steppe_Eneolithic
16	Levant_BA:Steppe_Eneolithic
16	EHG:Iran_ChL
15	Levant_N:Steppe_EMBA
14	Armenia_EBA:EHG
13	EHG:Levant_N
13	Armenia_EBA:Steppe_Eneolithic
12	Levant_BA:Steppe_EMBA
12	CHG:Steppe_Eneolithic
12	CHG:Steppe_EMBA
12	Armenia_EBA:Steppe_EMBA
11	EHG:Levant_BA
11	CHG:EHG
11	Armenia_MLBA:Steppe_EMBA
10	Armenia_MLBA:Steppe_Eneolithic
9	Iran_N:Steppe_MLBA
8	Levant_N:Steppe_MLBA
8	Iran_LN:Steppe_MLBA
8	Iran_ChL:Steppe_MLBA
8	CHG:Steppe_MLBA
8	Armenia_MLBA:EHG
8	Armenia_EBA:Steppe_MLBA
7	Levant_BA:Steppe_MLBA
7	Iran_N:SHG
7	Armenia_ChL:Steppe_Eneolithic
7	Armenia_ChL:Steppe_EMBA
7	Armenia_ChL:EHG
7	Anatolia_ChL:Steppe_Eneolithic
7	Anatolia_ChL:Steppe_EMBA
6	Armenia_MLBA:Steppe_MLBA
6	Armenia_MLBA:SHG
6	Anatolia_N:Steppe_EMBA
6	Anatolia_ChL:EHG
5	Iran_ChL:SHG
4	Iran_LN:SHG
4	Europe_LNBA:Iran_N
4	Europe_LNBA:Iran_ChL
4	Europe_EN:Steppe_EMBA
4	Armenia_ChL:Steppe_MLBA
4	Anatolia_N:Steppe_Eneolithic
4	Anatolia_ChL:Steppe_MLBA

Table S9.3: Pairs of West Eurasian populations (out of $\binom{20}{2} = 190$ choices) and the number of South Asian populations (out of a total of 20) that could be modeled as 4-way mixtures involving each pair plus Onge and Han using the S7NS outgroups. We show only population pairs that can be used to model at least 4 South Asian populations (for Left=(*South Asian*, *West Eurasian*$_1$, *West Eurasian*$_2$, Onge, Han) we have p-value for rank=3 greater or equal to 0.05 and estimated mixture proportions in [0, 1] interval).

Number of South Asian populations modeled as mixtures of *Pair* , Onge, and Han	*Pair*
20	Iran_LN:Steppe_Eneolithic
20	EHG:Iran_LN
19	Iran_N:Steppe_EMBA
18	Iran_N:Steppe_Eneolithic
16	Iran_LN:Steppe_EMBA
15	EHG:Iran_N
12	Iran_ChL:Steppe_Eneolithic
12	CHG:Steppe_Eneolithic
10	EHG:Iran_ChL
9	CHG:EHG
8	Iran_ChL:Steppe_EMBA
8	CHG:Steppe_EMBA

Table S9.4: Mixture proportions for South Asian populations.

		Mixture Proportions				Standard Errors					Mixture Proportions				Standard Errors					Mixture Proportions				Standard Errors			
South Asian	P-value for rank=3	Iran_LN	Steppe_Eneolithic	Onge	Han	Iran_LN	Steppe_Eneolithic	Onge	Han	P-value for rank=3	Iran_LN	EHG	Onge	Han	Iran_LN	EHG	Onge	Han	P-value for rank=3	Iran_N	Steppe_EMBA	Onge	Han	Iran_N	Steppe_EMBA	Onge	Han
Balochi	8.47E-01	0.605	0.269	0.091	0.035	0.036	0.031	0.034	0.025	7.43E-01	0.662	0.204	0.094	0.040	0.033	0.026	0.037	0.026	3.69E-01	0.566	0.324	0.074	0.035	0.037	0.030	0.028	0.018
Bengali	8.87E-01	0.268	0.197	0.419	0.116	0.027	0.022	0.028	0.019	6.25E-01	0.310	0.150	0.421	0.119	0.024	0.018	0.029	0.019	5.00E-01	0.224	0.246	0.415	0.115	0.032	0.025	0.027	0.017
Brahmin_Tiwari	8.68E-01	0.313	0.359	0.307	0.021	0.027	0.022	0.026	0.018	3.04E-01	0.386	0.276	0.316	0.023	0.024	0.019	0.027	0.019	1.12E-01	0.227	0.441	0.308	0.024	0.029	0.022	0.022	0.013
Brahui	4.23E-01	0.632	0.248	0.081	0.039	0.038	0.033	0.035	0.025	6.13E-01	0.678	0.193	0.089	0.040	0.034	0.026	0.038	0.027	5.32E-01	0.597	0.302	0.064	0.037	0.040	0.033	0.030	0.019
Burusho	6.50E-01	0.355	0.345	0.166	0.134	0.027	0.023	0.025	0.018	1.89E-01	0.424	0.264	0.175	0.138	0.025	0.019	0.027	0.019	1.70E-02	0.273	0.425	0.163	0.138	0.028	0.022	0.020	0.013
GujaratiA	7.96E-01	0.361	0.372	0.232	0.034	0.032	0.026	0.033	0.022	3.91E-01	0.434	0.287	0.241	0.039	0.030	0.022	0.035	0.023	2.99E-01	0.263	0.461	0.240	0.036	0.034	0.027	0.027	0.017
GujaratiB	8.92E-01	0.345	0.308	0.320	0.027	0.030	0.025	0.031	0.021	7.80E-01	0.405	0.239	0.327	0.029	0.027	0.021	0.032	0.022	6.12E-01	0.273	0.380	0.318	0.030	0.035	0.028	0.027	0.018
GujaratiC	6.66E-01	0.296	0.301	0.395	0.009	0.031	0.026	0.031	0.021	2.45E-01	0.356	0.230	0.403	0.012	0.027	0.021	0.032	0.022	1.38E-01	0.226	0.368	0.393	0.014	0.036	0.028	0.030	0.018
GujaratiD	4.77E-01	0.325	0.231	0.416	0.028	0.033	0.027	0.032	0.022	2.92E-01	0.373	0.176	0.422	0.029	0.029	0.022	0.034	0.023	1.52E-01	0.286	0.275	0.408	0.031	0.039	0.031	0.032	0.020
Jew_Cochin	7.41E-01	0.554	0.181	0.213	0.052	0.041	0.034	0.038	0.026	8.31E-01	0.587	0.141	0.220	0.052	0.036	0.027	0.039	0.027	6.82E-02	0.525	0.231	0.194	0.050	0.043	0.034	0.034	0.021
Kalash	4.76E-01	0.383	0.412	0.169	0.036	0.032	0.026	0.031	0.022	1.50E-01	0.468	0.317	0.173	0.042	0.030	0.022	0.035	0.024	5.65E-02	0.290	0.502	0.163	0.046	0.034	0.027	0.026	0.017
Kharia	8.39E-01	0.130	0.056	0.638	0.175	0.027	0.022	0.029	0.019	7.73E-01	0.142	0.043	0.639	0.176	0.024	0.017	0.029	0.019	8.51E-01	0.128	0.065	0.632	0.174	0.036	0.027	0.031	0.019
Kusunda	3.90E-01	0.082	0.076	0.239	0.602	0.023	0.018	0.023	0.016	3.45E-01	0.095	0.060	0.242	0.603	0.021	0.014	0.024	0.016	4.05E-01	0.070	0.089	0.234	0.607	0.030	0.023	0.025	0.016
Lodhi	2.94E-01	0.265	0.234	0.472	0.029	0.027	0.023	0.027	0.018	3.19E-01	0.310	0.183	0.476	0.031	0.024	0.018	0.027	0.018	1.07E-01	0.210	0.293	0.465	0.032	0.033	0.025	0.027	0.017
Makrani	1.02E-01	0.768	0.166	0.028	0.038	0.045	0.038	0.043	0.03	1.13E-01	0.800	0.128	0.032	0.039	0.040	0.031	0.044	0.031	4.00E-01	0.774	0.192	0.001	0.033	0.048	0.039	0.036	0.023
Mala	3.90E-01	0.250	0.155	0.564	0.031	0.027	0.023	0.029	0.02	2.18E-01	0.281	0.119	0.567	0.033	0.024	0.018	0.029	0.020	2.14E-01	0.230	0.184	0.552	0.034	0.034	0.026	0.029	0.019
Pathan	9.12E-01	0.441	0.363	0.140	0.057	0.030	0.026	0.028	0.02	4.85E-01	0.518	0.277	0.142	0.064	0.028	0.022	0.031	0.022	6.63E-02	0.360	0.446	0.136	0.059	0.028	0.023	0.021	0.013
Punjabi	8.09E-01	0.303	0.261	0.410	0.026	0.028	0.023	0.029	0.02	3.75E-01	0.354	0.200	0.417	0.030	0.025	0.019	0.030	0.021	5.47E-01	0.236	0.326	0.415	0.023	0.032	0.026	0.026	0.017
Sindhi	6.19E-01	0.432	0.311	0.227	0.029	0.029	0.025	0.028	0.02	6.74E-01	0.492	0.241	0.236	0.031	0.027	0.020	0.030	0.021	2.88E-01	0.376	0.377	0.211	0.035	0.030	0.024	0.024	0.015
Vishwabrahmin	1.96E-01	0.270	0.172	0.528	0.030	0.028	0.023	0.027	0.018	9.14E-02	0.306	0.131	0.530	0.033	0.025	0.019	0.027	0.019	9.25E-02	0.246	0.204	0.516	0.034	0.035	0.027	0.028	0.017

References

1. Moorjani, P. *et al.* Genetic evidence for recent population mixture in India. *Am. J. Hum. Genet.* **93**, 422-438, (2013).
2. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489-494, (2009).
3. Metspalu, M. *et al.* Shared and Unique Components of Human Population Structure and Genome-Wide Signals of Positive Selection in South Asia. *The American Journal of Human Genetics* **89**, 731-744, (2011).
4. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912, (2015).
5. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503, (2015).
6. Hollard, C. *et al.* Strong genetic admixture in the Altai at the Middle Bronze Age revealed by uniparental and ancestry informative markers. *Forensic Science International: Genetics* **12**, 199-207, (2014).
7. Pamjav, H., Fehér, T., Németh, E. & Pádár, Z. Brief communication: New Y-chromosome binary markers improve phylogenetic resolution within haplogroup R1a1. *Am. J. Phys. Anthropol.* **149**, 611-615, (2012).
8. Underhill, P. A. *et al.* The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur. J. Hum. Genet.* **23**, 124–131, (2014).
9. Anthony, D. W. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. (Princeton University Press, 2007).
10. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216-219, (2015).
11. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
12. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211, (2015).
13. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, (2014).
14. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225-228, (2014).
15. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118, (2014).
16. Skoglund, P. *et al.* Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. *Science* **344**, 747-750, (2014).
17. Thangaraj, K. *et al.* Reconstructing the origin of Andaman Islanders. *Science* **308**, 996-996, (2005).
18. Basu, A., Sarkar-Roy, N. & Majumder, P. P. Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. *Proceedings of the National Academy of Sciences*, (2016).
19. Kuz'mina, E. E. *The Origin of the Indo-Iranians* (Brill, 2007).

Supplementary Information 10

Modeling admixture from ghost populations

In this section we present a generalization of the framework of inferring admixture for a *Test* population from *N Reference* populations using a set of *Outgroup* ones[1,2] that we use in Supplementary Information, section 7. The basic idea is to relax the requirement that *Test* is descended from the *N* sources and allow it to be descended from an unobserved "ghost" population that resides on a line defined by two of the sources. The utility of this generalization will become clearer in its application, but intuitively this frees us to consider admixture with a population that may not correspond exactly to our available samples but is composed by the same elements as them in different proportions.

In the present we will mainly use it to show that we can recover mixture proportions from pseudo-"ghost" populations whose existence we actually know; this is a validation of the method, and also a test of the robustness of key population inferences as they can be shown to not be dependent on the available samples from the pseudo-"ghost" populations.

Inference framework

To simplify notation, we will use the following shorthand for statistics involving a *Test* and *Reference* population:

$$\langle t, r_i \rangle = F_4(Test, Ref_i; O_2, O_3)$$

Note that $\langle t, r_i \rangle$ is a $\binom{m}{2}$ long column vector for all pairs of m different outgroups.

If *Test* is descended from *N* sources with proportions α_i then:

$$\sum_{i=1}^{N} \alpha_i \langle t, r_i \rangle = 0_{\binom{m}{2}\times 1} \qquad \text{(Eq. 1)}$$

We remove the first two sources from this sum and replace them with a "ghost" population X that has the property (for a real λ):

$$x = r_1 + \lambda(r_2 - r_1) \qquad \text{(Eq. 2)}$$

X resides on the line formed by the two references. For λ=0 it is identical to Ref_1. For λ=1 it is identical to Ref_2. For 0<λ<1 it is a mixture of the two references, while for λ>1 it is more extreme than Ref_2 and for λ<0 it is more extreme than Ref_1. When the parameter λ is outside the [0, 1] interval it controls the amount of extrapolation along the cline defined by the two source populations.

If these two source populations are composed of the same two population elements (in different proportions), then they define a cline in the $\binom{m}{2}$-dimensional space of f_4-statistics.

The population X we are seeking also resides along this cline. To see why this is the case, consider that r_1, r_2 have respectively γ, 1-γ and δ, 1-δ proportions of ancestry from two unobserved non-admixed populations p and q. Then we may write:

$$r_1 = \gamma p + (1-\gamma)q$$
$$r_2 = \delta p + (1-\delta)q$$

Thus, our expression for the ghost population (Eq. 2) can be re-written as:

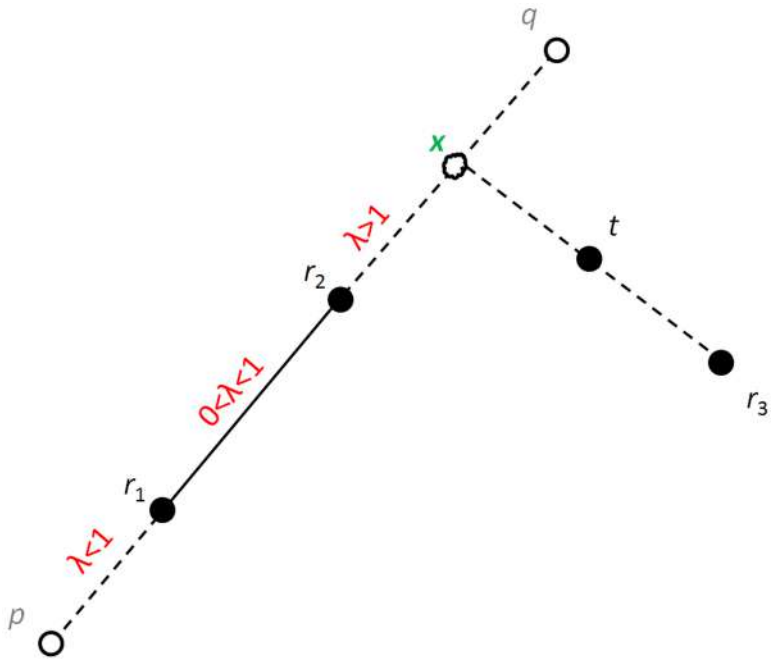
$$x = \gamma p + (1-\gamma)q + \lambda(\delta p + (1-\delta)q - \gamma p - (1-\gamma)q)$$

Or:

$$x = (\gamma + \lambda\delta - \lambda\gamma)p + (1-\gamma-\lambda\delta+\gamma\lambda)q$$

Thus population x is also modeled as a 2-way mix of $(\gamma + \lambda\delta - \lambda\gamma)$ of p and $1-(\gamma + \lambda\delta - \lambda\gamma)$ of q. The key point is that we do not need to sample either the unobserved populations p and q or the ghost population x. By having only r_1 and r_2, we may model any population x on the cline defined by r_1 and r_2 (Fig. S9.1)

Figure S9.1: Illustration of ghost admixture. Test population t cannot be modelled as a mix of the observed source populations (r_1, r_2, r_3) but can be modelled as a mix of r_3 and unobserved ghost population x. The ghost population is composed of the same unobserved elements p and q as reference populations r_1 and r_2 and resides on the cline defined by them.



Assuming β fraction of ancestry from the ghost population, we may re-write Eq. 1 as:

$$\beta\langle t,x\rangle + \sum_{i=3}^{N} \alpha_i \langle t,r_i\rangle = 0_{\binom{m}{2}\times 1}$$

Or:

$$\beta(1-\lambda)\langle t,r_1\rangle + \beta\lambda\langle t,r_2\rangle + \sum_{i=3}^{N} \alpha_i \langle t,r_i\rangle = 0_{\binom{m}{2}\times 1}$$

In practice (substituting theoretical F_4 statistics with empirically estimated f_4 statistics), the above equation holds approximately and we seek estimates of the coefficients $\beta(1-\lambda), \beta\lambda, \alpha_i$ that minimize the sum of squares:

$$\left\| \beta(1-\lambda)\langle t,r_1\rangle + \beta\lambda\langle t,r_2\rangle + \sum_{i=3}^{N} \alpha_i \langle t,r_i\rangle \right\|_2^2$$

Subject to the constraints $0 \le \beta, a_i \le 1$ and $\beta + \sum_{i=3}^{N} a_i = 1$.

Simplifying, we set $k_1 = \beta(1-\lambda), k_2 = \beta\lambda$, and $k_i = \alpha_i$ for $i > 2$. Then the optimization problem is written as:

$$min\left\|\sum_{i=1}^{N} k_i \langle t,r_i\rangle\right\|_2^2, \text{ subject to } \sum_{i=1}^{N} k_i = 1 \text{ and } 0 \le k_i \le 1 \text{ for } i > 2.$$

We use the *lsqlin* function of Matlab (http://www.mathworks.com/help/optim/ug/lsqlin.html) to solve this problem. This function can solve the problem:

$$min\frac{1}{2}\|C \cdot x - d\|_2^2 \text{ subject to } Aeq \cdot x = beq \text{ and } lb \leq x \leq ub$$

In our application:

- $x = \begin{bmatrix} k_1 \\ k_2 \\ \dots \\ k_N \end{bmatrix}$
- $C = [\langle t, r_1 \rangle \quad \langle t, r_2 \rangle \quad \dots \quad \langle t, r_N \rangle]$
- $d = 0_{\binom{m}{2}\times 1}$
- $Aeq = 1_{1\times N}$
- $beq = 1$
- $lb = \begin{bmatrix} -\infty \\ -\infty \\ 0 \\ \dots \\ 0 \end{bmatrix}$
- $ub = \begin{bmatrix} \infty \\ \infty \\ 1 \\ \dots \\ 1 \end{bmatrix}$

We practically set $\infty = 100$. This allows λ to stretch between [-100, 100], i.e., for the ghost population to be a 100-fold times outside the segment of a cline defined by the two reference populations.

Application

As we mentioned in the beginning of this note, we validate the new method using cases with known admixture events; these are mainly taken from Supplementary Information, section 7. We will use the O9 as the set of outgroups.

WHG

Scandinavian hunter-gatherers (SHG) can be modeled as a 2-way mix of EHG and WHG and the Middle Neolithic/Chalcolithic Europeans (Europe_MNChL) can be modeled as a 2-way mix of Anatolian Neolithic and WHG. These 2-way mixtures can be expressed in the framework discussed in this section as follows.

Test	Ref_1	Ref_2	α_{WHG}
Europe_MNChL	Anatolia_N	WHG	0.211
SHG	EHG	WHG	0.553

We withhold WHG and re-estimate mixture proportions. We also estimate the mixture proportion of Europe_MNChL by making use of the fact that WHG can be modeled as a mixture of Switzerland_HG and EHG.

Test	Ref_1	Ref_2	Ref_3	$\alpha_{Ghost\ WHG}$	λ
Europe_MNChL	EHG	SHG	Anatolia_N	0.224	1.64
Europe_MNChL	EHG	Switzerland_HG	Anatolia_N	0.188	0.81
SHG	Anatolia_N	Europe_MNChL	EHG	0.578	3.86

Thus, mixture proportions from "Ghost WHG" are within ~3% of those from actual WHG. Note the amount of extrapolation gauged by the λ parameter: "Ghost WHG" is inferred to be beyond the EHG→SHG cline segment, in-between (EHG, Switzerland_HG) and substantially beyond the Anatolia_N→Europe_MNChL cline segment.

EHG

SHG can be modeled as a 2-way mix of EHG and WHG, Early/Middle Bronze steppe populations (Steppe_EMBA) can be modeled as a 2-way mix of EHG and Iran_ChL, and WHG can be modeled as a 2-way mix of Switzerland_HG and EHG.

Test	Ref_1	Ref_2	α_{EHG}
SHG	WHG	EHG	0.447
Steppe_EMBA	Iran_ChL	EHG	0.528
WHG	Switzerland_HG	EHG	0.083

We withhold EHG and can estimate mixture proportions as follows.

Test	Ref_1	Ref_2	Ref_3	$\alpha_{Ghost\ EHG}$	λ
SHG	Iran_ChL	Steppe_EMBA	WHG	0.510	1.62
Steppe_EMBA	WHG	SHG	Iran_ChL	0.612	1.92
Steppe_EMBA	Switzerland_HG	WHG	Iran_ChL	0.580	5.94
WHG	Iran_ChL	Steppe_EMBA	Switzerland_HG	0.113	1.32

Thus, mixture proportions from "Ghost EHG" are within ~8% of those from actual EHG. Note the successful modeling of Steppe_EMBA when using the Switzerland_HG→WHG cline to infer the "Ghost EHG": since WHG has mostly Switzerland_HG ancestry, this involves projection well beyond the (Switzerland_HG, WHG) segment.

Steppe_EMBA

Europe_LNBA and Steppe_MLBA can both be modeled as 2-way mixtures of Europe_MNChL and Steppe_EMBA.

Test	Ref_1	Ref_3	α_{Steppe_EMBA}
Europe_LNBA	Europe_MNChL	Steppe_EMBA	0.524
Steppe_MLBA	Europe_MNChL	Steppe_EMBA	0.694

But, Steppe_EMBA can be modeled as a 2-way mixture of EHG and Iran_ChL. We thus withhold Steppe_EMBA and model Europe_LNBA and Steppe_MLBA as 3-way mixtures:

Test	Ref_1	Ref_2	Ref_3	$\alpha_{Ghost\ Steppe\_EMBA}$	λ
Europe_LNBA	EHG	Iran_ChL	Europe_MNChL	0.575	0.52
Steppe_MLBA	EHG	Iran_ChL	Europe_MNChL	0.716	0.49

These are within ~5% of those from the actual Steppe_EMBA. Thus, we could infer admixture from a population that is an approximately even mix of EHG and Iran_ChL into Late Neolithic/Bronze Age Europe and the Middle/Late Bronze Age Eurasian steppe even if we did not have any samples from the actual Early/Middle Bronze Age Yamnaya-related pastoralists that were the actual mediators of this admixture.

Iran_ChL

Steppe_EMBA can be modeled as a 2-way mix of EHG and Iran_ChL and Levant_BA can be modeled as a 2-way mix of Iran_ChL and Levant_N. The very strong differentiation between the two substratum populations (Levant_N and EHG) makes this an ideal case for applying our method. First, we model the admixed populations using Iran_ChL.

Test	Ref_1	Ref_2	α_{Iran_ChL}
Steppe_EMBA	EHG	Iran_ChL	0.472
Levant_BA	Levant_N	Iran_ChL	0.399

Next, we withhold Iran_ChL and model them as follows, using the EHG→Steppe_EMBA and Levant_N→Levant_BA clines:

Test	Ref_1	Ref_2	Ref_3	$\alpha_{Ghost\ Iran\_ChL}$	λ
Steppe_EMBA	Levant_N	Levant_BA	EHG	0.485	2.90
Levant_BA	EHG	Steppe_EMBA	Levant_N	0.252	1.86

The inference of Steppe_EMBA is robust (within ~2%), for Levant_BA less so (~15% off). When we model Levant_BA using qpAdm and the same O9 outgroups, we obtain an estimate of 0.43±0.102 of Iran_ChL ancestry, therefore our estimate above is off by 1.7 standard errors.

Levant_N

In Supplementary Information, section 8 we show that East African populations can be modeled as a mixture of Mota and Levant_N. Levant_BA can also be modeled as Iran_ChL and Levant_N. We estimate, using O8ENSW as outgroups:

Test	Ref_1	Ref_2	α_{Levant_N}
Levant_BA	Levant_N	Iran_ChL	0.527
Somali	Levant_N	Mota	0.330
Kikuyu	Levant_N	Mota	0.134
Masai	Levant_N	Mota	0.181
Datog	Levant_N	Mota	0.228
Sandawe	Levant_N	Mota	0.156

Withholding Levant_N we obtain:

Test	Ref_1	Ref_2	Ref_3	$\alpha_{Ghost\ Levant\_N}$	λ
Levant_BA	Mota	Somali	Iran_ChL	0.605	2.75
Levant_BA	Mota	Kikuyu	Iran_ChL	0.320	7.79
Levant_BA	Mota	Masai	Iran_ChL	0.463	5.24
Levant_BA	Mota	Datog	Iran_ChL	0.537	3.95
Levant_BA	Mota	Sandawe	Iran_ChL	0.426	6.46
Somali	Iran_ChL	Levant_BA	Mota	0.365	1.53
Kikuyu	Iran_ChL	Levant_BA	Mota	0.156	1.34
Masai	Iran_ChL	Levant_BA	Mota	0.200	1.52
Datog	Iran_ChL	Levant_BA	Mota	0.264	1.36
Sandawe	Iran_ChL	Levant_BA	Mota	0.159	1.81

We observe that mixture proportions with "Ghost Levant_N" for East African populations match quite closely (within ~4%) those with actual Levant_N, but mixture proportions with "Ghost Levant_N" for Levant_BA vary widely. The parameter λ which gauges the amount of extrapolation beyond the observed cline segment is useful to understand this discrepancy: when the "Test" population is East African, extrapolation along the Iran_ChL→Levant_BA involves only 34-81% projection beyong Levant_BA, but when the "Test" population is Levant_BA, extrapolation along the Mota→*East African* cline involves extrapolation of 175-679% beyond *East African*. For example, using Kikuyu as *East African* we are attempting to recreate the admixing population by using only Mota and a population that has its ancestry from ~6/7 Mota and only ~1/7 from West Eurasia. Assuming perfect estimation of *f*-statistics and perfect identification of ancestral sources any cline

segment would suffice to reproduce mixture proportions using our method, as any line segment of a line uniquely defines its direction. As these assumptions will not hold in practice, our results suggest that a high degree of extrapolation may lead to inaccurate estimates of mixture proportions.

Anatolia_N

In Supplementary Information, section 7 we inferred that Anatolia_N cannot be modeled as a mixture of Levant_N, Iran_N, WHG, EHG. Rank=3 is not rejected (P=0.463) for the Left quintuple (Anatolia_N, Levant_N, Iran_N, WHG, EHG) in relation to the O9 outgroups, but a negative proportion (-0.069±0.027) of EHG is inferred. This suggests that these five populations are descended from four streams of ancestry but Neolithic Anatolia cannot be reduced to a weighted combination of the other four populations.

We can re-formulate this problem within the framework introduced in this section by modeling Anatolia_N as a mix of Levant_N, Iran_N and a population on the EHG→WHG cline or EHG→Switzerland_HG or WHG→Switzerland_HG cline (as WHG itself is modeled as a mixture of EHG and WHG):

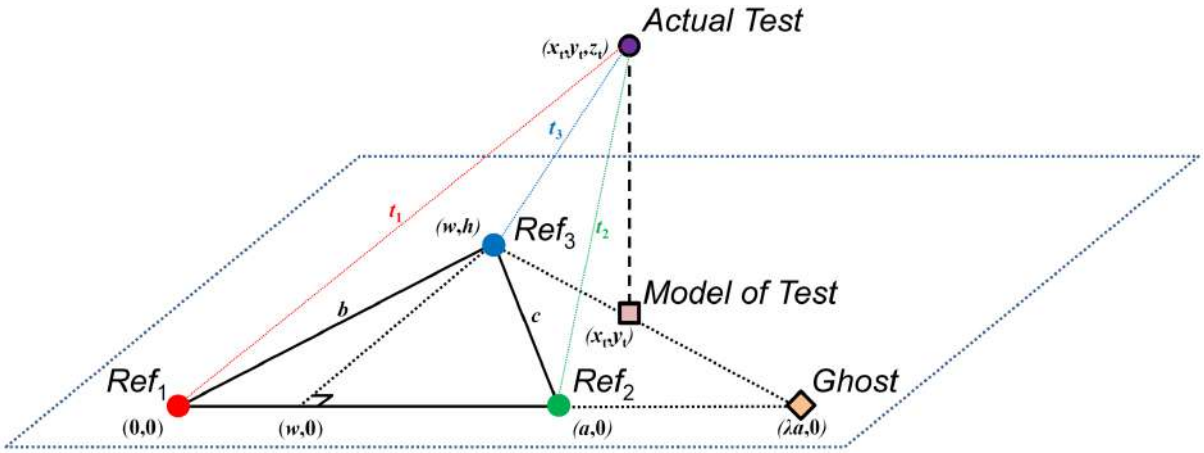
Test	Ref_1	Ref_2	Ref_3	Ref_4	α_{Ghost}	λ
Anatolia_N	EHG	WHG	Levant_N	Iran_N	0.3293	1.1364
Anatolia_N	EHG	Switzerland_HG	Levant_N	Iran_N	0.3339	1.0793
Anatolia_N	WHG	Switzerland_HG	Levant_N	Iran_N	0.3213	0.9297

Notice that the proportions from the "Ghost" population are fairly stable at around ~1/3, and this population is inferred to be (based on the λ parameter) beyond both WHG and Switzerland_HG on the respective EHG→WHG and EHG→Switzerland_HG clines and closer to Switzerland_HG than to WHG. Future sampling may reveal whether such a population, whose existence is hypothesized here as a way to better model the Anatolian Neolithic, did actually exist.

Visualization of cline intersection

We used the idea of intersecting clines (Fig. S9.1) to illustrate how our method works, but since it relies on statistics in $\binom{m}{2}$-dimensional space, it is not possible to visualize it directly. We introduce a method for visualizing cline intersection in the special case of N=3, i.e., when the admixed population is formed of a fixed reference population Ref_3 and a ghost population residing on the Ref_1, Ref_2 cline. Most of the applications of our method above involve N=3, although our formulation allows for arbitrary N.

Figure S9.2: Visualization of cline intersection. A *Test* population is projected on the *(x, y)* plane and is modeled as a mix of Ref_3 and a population on the Ref_1, Ref_2 cline.



The basic idea of the visualization technique is to embed populations in a 3D space in a way that preserves their Euclidean distances in the $\binom{m}{2}$-dimensional space of f_4-statistics (Fig. S9.2). Given a vector $\langle a, b\rangle = F_4(A, B; O_2, O_3)$ we define as the distance between populations A, B the 2-norm:

$$d_{A,B} = \|\langle a, b\rangle\|_2$$

There are three such distances between the three reference populations that can be thought of as three sides of a triangle $a = d_{Ref_1,Ref_2}$, $b = d_{Ref_1,Ref_3}$, $c = d_{Ref_2,Ref_3}$. We place this triangle on a 2D Cartesian plane (*x*, *y*) (Fig. S9.2) with vertex coordinates:

- Ref_1: (0,0)
- Ref_2: (0,*a*)
- Ref_3: $(w = \frac{a^2+b^2-c^2}{2a}, h = \sqrt{b^2 - w^2})$

Next, consider the *Test* population which has distances to the three reference populations: $t_1 = d_{Test,Ref_1}, t_2 = d_{Test,Ref_2}, t_3 = d_{Test,Ref_3}$. This can be placed in 3D space (*x*, *y*, *z*) at an intersection point (x_t, y_t, z_t) of three spheres centered on Ref_1, Ref_2, Ref_3 with radii t_1, t_2, t_3. An intersection of a pair of spheres is not always defined. We can find it by noting that:

$$t_1^2 = x_t^2 + y_t^2 + z_t^2$$
$$t_2^2 = (x_t - a)^2 + y_t^2 + z_t^2$$
$$t_3^2 = (x_t - w)^2 + (y_t - h)^2 + z_t^2$$

These equations yield:

$$x_t = \frac{a^2 + t_1^2 - t_2^2}{2a}$$

$$y_t = \frac{t_2^2 - t_3^2 - (x_t - a)^2 + (x_t - w)^2 + h^2}{2h}$$

$$z_t = \pm\sqrt{t_1^2 - x_t^2 - y_t^2}$$

The root is defined if $t_1^2 \geq x_t^2 + y_t^2$, that is if the projection of the *Test* population on the (x, y) plane (x_t, y_t) is enclosed in a sphere of radius $t_1 = d_{Test,Ref_1}$. We can plot either the positive or negative root that are symmetrically positioned about the (x, y) plane at distance $\sqrt{t_1^2 - x_t^2 - y_t^2}$.

Since the ghost population is allowed to be on the cline defined by Ref_1=(0,0) and Ref_2=(0,a), and the ghost population is placed on (λa, 0), where λ is a real-valued parameter. It is also placed on a line which passes through points (x_t, y_t) and (w, h). We obtain:

$$\lambda = \frac{hx_t - wy_t}{a(h - y_t)}$$

The method we just described is useful for exploratory data analysis as it allows one to gain insight by visualizing relationships between populations. We show examples of its application on several ghost populations discussed in this section and in Supplementary Information, sections 8 and 11, in Extended Data Fig. 7.

Conclusion

In this section we used the framework of inferring admixture from ghost populations to (i) show its feasibility, and (ii) demonstrate the coherence of the inferences presented in this paper: by removing populations and re-estimating mixture proportions from pseudo-ghost populations we could, nonetheless, estimate mixture proportions fairly accurately. A useful area of future work is to use this framework in an exploratory manner to uncover previously undetected ghost populations in the manner we attempted to do for the Anatolian Neolithic and to provide a way of gauging confidence in the existence of such populations.

References

1. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211, (2015).
2. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, (2014).

Supplementary Information 11

Admixture in East Asians and Eastern European hunter-gatherers

In Supplementary Information, section 8 we discussed the admixture between West Eurasian populations (likely from the Levant) and East Africans. In this section we show that there is previously unknown admixture from a population related to "Ancient North Eurasians"[1,2] and "Eastern European hunter-gatherers"[3] into East Asians.

Western Eurasians are not clade with respect to Eastern non-Africans and vice versa

If separate migrations settled Western Eurasia and Eastern Eurasia during the Upper Paleolithic, with subsequent genetic isolation, then statistics of the form f_4(West Eurasian$_1$, West Eurasian$_2$; Eastern non-African$_1$, Eastern non-African$_2$) have an expected value of 0. Table S11.1 shows that there are statistics that strongly deviate from 0, disproving the genetic isolation hypothesis of West Eurasians and eastern non-Africans. In particular, the Han share more alleles with the EHG than with other West Eurasians, with WHG than with Kostenki14, Switzerland_HG, Natufians, and with CHG than with Kostenki14, and Natufians.

Table S11.1: Asymmetry between West Eurasians and (Han, Papuan or Onge). The EHG share significantly more alleles than other West Eurasians with the Han. Significant statistics are shown for West Eurasians in the set: Switzerland_HG[4], WHG, EHG, Iran_N, Natufian, CHG[4], Kostenki14[5], MA1[2] and Eastern non-Africans in the set: Han, Papuan, Onge.

A	*B*	*C*	*D*	f_4(*A*, *B*; *C*, *D*)	*Z*
EHG	Kostenki14	Han	Papuan	0.00233	5.7
EHG	Natufian	Han	Papuan	0.00222	5.4
EHG	Iran_N	Han	Papuan	0.00168	4.6
EHG	Kostenki14	Han	Onge	0.00140	3.9
EHG	Natufian	Han	Onge	0.00133	3.7
EHG	WHG	Han	Papuan	0.00106	3.6
EHG	WHG	Han	Onge	0.00099	3.6
WHG	Kostenki14	Han	Papuan	0.00133	3.6
CHG	Kostenki14	Han	Papuan	0.00145	3.5
EHG	MA1	Han	Onge	0.00121	3.5
CHG	Natufian	Han	Papuan	0.00131	3.4
EHG	Switzerland_HG	Han	Papuan	0.00126	3.4
WHG	Natufian	Han	Papuan	0.00123	3.3
EHG	CHG	Han	Onge	0.00093	3.1

Eastern Eurasians have admixture from a population related to the EHG

In Supplementary Information, section 7 we showed that the WHG could be modeled as a mixture of EHG and Switzerland_HG. Discovering this admixture was possible by the availability of the Bichon (Switzerland_HG) individual. It is possible that EHG too is admixed, with Switzerland_HG as one source and a ghost population "X" as another. If this population "X" also contributed to the Han, then this would explain why the Han share more genetic drift with the EHG than with other West Eurasians. The Han would also share more genetic drift with populations with EHG admixture (such as WHG and CHG) than with other West Eurasians.

At this stage we are agnostic as to the identity of "X". A special case is that X=EHG, i.e., EHG is unadmixed, gene flow was exclusively from EHG→Han; a different special case is that X=Han, i.e., Han is unadmixed, gene flow was exclusively from Han→EHG. In the general case X is an unobserved population that contributed to both EHG and to Han.

As our problem involves both East and West Eurasian populations, we must withdraw populations from the set of outgroups (e.g., we want to treat Han as a potentially admixed population, so it cannot be in the set of outgroups). We use the following set that includes an African group, Upper Paleolithic Eurasians and an eastern non-African population (Papuan) that is unlikely to have been affected by recent gene flow with West Eurasians.

M5: Mbuti, Ust_Ishim, Papuan, Kostenki14, Switzerland_HG

We first test whether the triple Left=(EHG, WHG, Han) may be related to Right=M5 only via 2 streams of ancestry. We strongly reject rank=1 (P=6.11e-27), and thus EHG cannot be modeled as a mix of WHG and Han; this rejects the simple case where X=Han. We then test Left=(Han, Onge, EHG) which is not rejected (P=0.59). Thus, the simple case where X=EHG cannot be rejected. We estimate that Han can be modeled as having 92.1±1.9% Onge and 7.9±1.9% EHG ancestry.

It is useful to assess this visually, following the method of ref.[3] (Fig. S11.1) which makes use of the fact that when a *Test* population is a mix of populations related to Ref_1 and Ref_2 in proportions α and 1-α, then over pairs of outgroups (O_2, O_3), the following equation holds:

$$f_4(Test, Ref_2; O_2, O_3) \approx \frac{\alpha}{\alpha-1} f_4(Test, Ref_1; O_2, O_3)$$

This is an equation of a line with a negative slope ($\frac{\alpha}{1-\alpha}$) through the origin. Fig. S11.1 shows such a line when modeling Han as a mix of EHG and Onge but not when modeling EHG as a mix of Han and WHG. Several other East Asian populations can be modeled as an EHG+Onge mix. We show these in Table S11.2 and plot mixture proportions in Extended Data Fig. 5.

Two notes of caution are necessary here. First, the admixing populations need not be necessarily "close" (either geographically or genetically) to the EHG and Onge, but they are in some sense related to them so that present-day Eastern Eurasian populations have intermediate allele frequencies between them. This is plausible given that both EHG and Onge show genetic affinities that stretch well beyond eastern Europe and the Andaman Islands, and may represent in some sense more widely dispersed populations. The EHG share more alleles with a ~24,000-year old Upper Paleolithic individual (MA1) from Siberia[6] than any other ancient or present-day population, and the Onge are representative of both "Ancestral South Indians" contributing to populations of the Indian subcontinent[7,8] but also of a minor stream of ancestry present in Amazonians[9]. Second, in contrast to western Eurasia there is currently no genome-wide ancient DNA data from eastern Eurasia. As an analogy, prior to the availability of genomes other than the Tyrolean Iceman, Europeans were modeled as a 2-way mix of "Sardinians/Iceman" and an "Ancient North Eurasian" ancestry from which the Native Americans were also descended. However, subsequent work with ancient DNA provided a much more detailed picture, involving a three-way mixture and more proximate source populations[1-3,10]. Thus, while our results demonstrate widespread ancient admixture in eastern Eurasia, the story of eastern Eurasian origins will doubtlessly be more complex than our model.

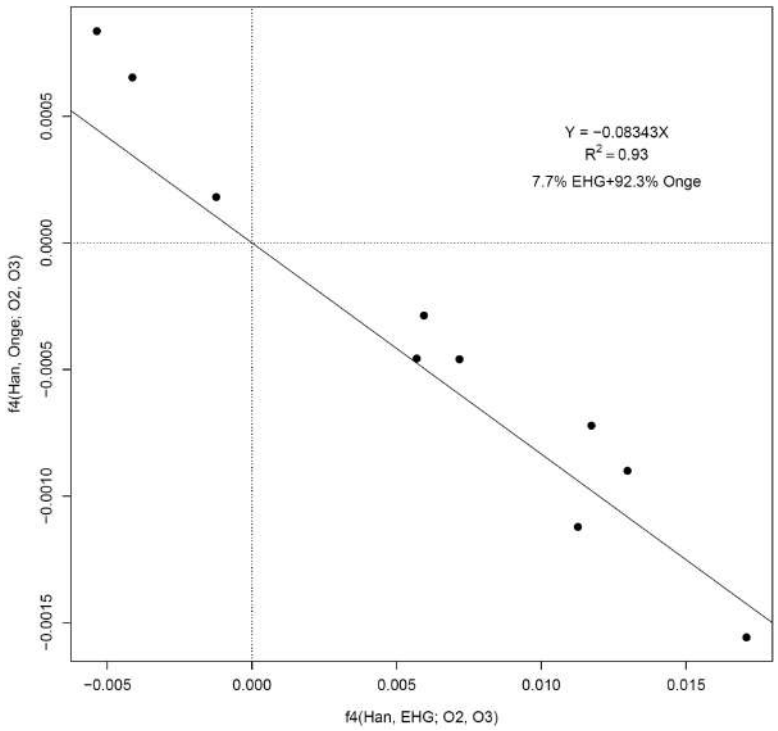
Table S11.2 assigns minimal EHG ancestry to the She people (6.1±2.0%). An alternative formulation is to use the She as an ancestral source. This allows us to place the Onge in the set of outgroups (M6 = M5 + Onge) and estimate whether eastern Eurasian populations can be modeled as a mix of She+EHG. Table S11.3 shows mixture proportions from this model. Many mainland East Asian populations have EHG proportions close to zero in this analysis (i.e., not significantly higher than the She), while others have substantially more such ancestry. The inferred mixture proportions of EHG ancestry in this analysis are consistent with those obtained using Onge as a reference.

Replacing EHG with a "Ghost" population

We modeled Eastern Eurasian populations as EHG+Onge/She mixtures, but it is possible that EHG may not correspond exactly to the admixing population. Recall that WHG can be modeled as a mix of Switzerland_HG and EHG, and SHG (Scandinavian hunter-gatherers) as a mix of WHG and EHG (Supplementary Information, section 7). As we are using Switzerland_HG as an outgroup, we can define three "clines" of ancestry for the remaining three populations (WHG→EHG, WHG→SHG, or SHG→EHG) and treat the population admixing into East Eurasians as a "ghost" population residing on these clines as in Supplementary Information, section 8. This analysis (Table S11.4) suggests that the admixing population may be beyond EHG on the WHG→EHG and SHG→EHG clines. Conversely, we can model EHG as a mix of WHG and a "Ghost" population on the Onge→*Eastern Eurasian* cline (Table S11.5). This suggests that EHG can be modeled as a mix of ~1/4 WHG and ~3/4 from the Ghost population.

Figure S11.1: Modeling Han as a mix of EHG and Onge (a) or EHG as a mix of Han and WHG. Consistent with the formal analysis, (a) is consistent with two streams of ancestry, with Han being "intermediate" between EHG and Onge, while (b) is inconsistent with two streams of ancestry, with EHG not clearly "intermediate" between WHG and Han.

(a)



(b)

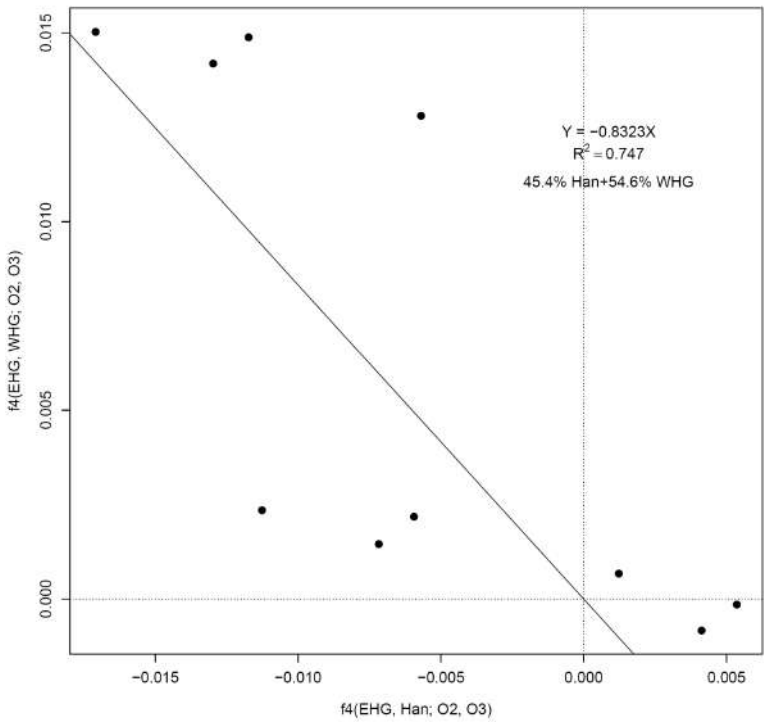


Table S11.2: Modeling Eastern Eurasian populations as a mix of EHG and Onge. We test whether (*Test*, EHG, Onge) is consistent with 2 streams of ancestry in relation to the M5 outgroups and list mixture proportions. P-values >0.05 are highlighted. The label "Eastern Eurasian" was applied to those populations that possess >50% of the "Eastern Eurasian" K=6 ADMIXTURE component; these do not necessarily reside in Eastern Eurasia

		Mixture Proportions		
Test	**P-value for rank=1**	**Onge**	**EHG**	**Std. Error**
Selkup	1.32E-01	0.555	0.445	0.016
Tubalar	8.25E-06	0.579	0.421	0.017
Even	6.00E-02	0.606	0.394	0.016
Kyrgyz	1.86E-04	0.656	0.344	0.017
Altaian	7.65E-04	0.673	0.327	0.018
Yukagir	3.70E-01	0.700	0.300	0.016
Dolgan	1.86E-02	0.747	0.253	0.020
Kalmyk	3.64E-03	0.749	0.251	0.018
Tuvinian	9.03E-02	0.760	0.240	0.017
Chukchi	1.65E-01	0.767	0.233	0.018
Eskimo	1.71E-01	0.769	0.231	0.019
Yakut	2.23E-01	0.778	0.222	0.017
Itelmen	1.27E-01	0.787	0.213	0.020
Koryak	1.48E-01	0.793	0.207	0.019
Nganasan	8.93E-02	0.801	0.199	0.019
Kusunda	7.59E-03	0.834	0.166	0.018
Oroqen	6.88E-01	0.860	0.140	0.018
Tu	1.58E-01	0.864	0.136	0.019
Daur	4.98E-01	0.874	0.126	0.019
Mongola	1.29E-01	0.875	0.125	0.020
Xibo	1.75E-01	0.877	0.123	0.019
Ulchi	4.87E-01	0.887	0.113	0.019
Hezhen	4.18E-01	0.896	0.104	0.019
Thai	1.58E-01	0.902	0.098	0.018
Cambodian	1.74E-01	0.915	0.085	0.018
Korean	2.18E-01	0.917	0.083	0.020
Miao	7.76E-01	0.921	0.079	0.020
Han	5.88E-01	0.921	0.079	0.019
Tujia	6.55E-01	0.923	0.077	0.020
Naxi	6.84E-01	0.923	0.077	0.019
Japanese	6.15E-01	0.927	0.073	0.019
Yi	6.44E-01	0.927	0.073	0.020
Lahu	5.73E-01	0.929	0.071	0.019
Kinh	7.55E-01	0.931	0.069	0.019
Dai	7.39E-01	0.932	0.068	0.019
Atayal	8.64E-01	0.934	0.066	0.021
Ami	7.82E-01	0.938	0.062	0.020
She	5.77E-01	0.939	0.061	0.020

Table S11.3: Modeling Eastern Eurasian populations as a mix of EHG and She. We test whether (*Test*, EHG, She) is consistent with 2 streams of ancestry in relation to the M6 (M5+Onge) outgroups and list mixture proportions. P-values >0.05 are highlighted. The right column lists "EHG-equivalent" proportions by summing the EHG column with 0.061 times the She column, as She are inferred to have 6.1% EHG ancestry in Table S11.2. For the set of populations that fit the EHG+She and EHG+Onge models, "EHG" proportions from Table S11.2 and "EHG equivalent" proportions from Table S11.3 differ by ≤1.3%.

		Mixture Proportions			
Test	**P-value for rank=1**	**She**	**EHG**	**Std. Error**	**EHG+0.061*She**
Selkup	1.70E-01	0.583	0.417	0.012	0.453
Tubalar	2.43E-07	0.592	0.408	0.012	0.444
Even	7.61E-02	0.645	0.355	0.011	0.395
Kyrgyz	1.64E-05	0.681	0.319	0.012	0.361
Altaian	1.02E-03	0.707	0.293	0.011	0.336
Yukagir	3.24E-01	0.740	0.260	0.010	0.305
Dolgan	1.20E-02	0.778	0.222	0.014	0.269
Kalmyk	9.33E-05	0.787	0.213	0.010	0.261
Tuvinian	5.77E-02	0.798	0.202	0.011	0.251
Eskimo	2.04E-01	0.805	0.195	0.012	0.244
Chukchi	7.26E-01	0.814	0.186	0.012	0.236
Yakut	3.66E-01	0.824	0.176	0.010	0.226
Itelmen	5.94E-01	0.835	0.165	0.014	0.216
Koryak	5.59E-01	0.845	0.155	0.013	0.206
Nganasan	1.11E-01	0.853	0.147	0.013	0.199
Kusunda	8.11E-04	0.883	0.117	0.011	0.170
Tu	1.71E-02	0.914	0.086	0.009	0.142
Oroqen	3.49E-01	0.918	0.082	0.010	0.138
Daur	3.35E-02	0.928	0.072	0.010	0.129
Mongola	2.96E-01	0.928	0.072	0.010	0.129
Xibo	3.12E-01	0.932	0.068	0.010	0.125
Ulchi	3.40E-01	0.944	0.056	0.010	0.114
Hezhen	3.89E-01	0.949	0.051	0.010	0.109
Thai	1.64E-02	0.961	0.039	0.010	0.097
Cambodian	1.13E-01	0.983	0.017	0.010	0.077
Korean	5.06E-02	0.984	0.016	0.010	0.076
Miao	6.06E-01	0.985	0.015	0.009	0.075
Han	1.30E-01	0.985	0.015	0.007	0.075
Naxi	1.27E-01	0.988	0.012	0.010	0.072
Tujia	3.09E-01	0.989	0.011	0.009	0.071
Yi	6.99E-01	0.989	0.011	0.009	0.071
Japanese	3.86E-01	0.993	0.007	0.008	0.068
Kinh	3.34E-01	0.996	0.004	0.009	0.065
Lahu	4.61E-02	1.000	0.000	0.010	0.061
Ami	7.59E-01	1.001	-0.001	0.010	0.060
Atayal	3.80E-01	1.004	-0.004	0.012	0.057
Dai	8.15E-02	1.006	-0.006	0.009	0.056

Table S11.4: Modeling Eastern Eurasian populations as a mixture of Onge and a ghost population. We list West Eurasian mixture proportions (α) estimated with the method of Supplementary Information, section 8 for mixtures of Onge + EHG (left column) and a population residing on the WHG→EHG, WHG→SHG, or SHG→EHG clines. Recall that the parameter λ for a cline $A \rightarrow B$ is defined as "Ghost"$=A+\lambda(B-A)$.

	EHG	WHG→EHG		WHG→SHG		SHG→EHG	
	α	λ	α	λ	α	λ	α
Ami	0.061	1.116	0.067	2.230	0.076	1.158	0.065
Atayal	0.065	1.224	0.078	2.381	0.086	1.420	0.076
Cambodian	0.083	1.354	0.114	2.634	0.129	1.672	0.108
Chukchi	0.229	1.186	0.268	2.283	0.289	1.358	0.262
Dai	0.067	1.245	0.083	2.429	0.092	1.455	0.080
Daur	0.124	1.152	0.140	2.225	0.151	1.289	0.138
Eskimo	0.230	1.260	0.288	2.385	0.305	1.535	0.282
Even	0.388	1.020	0.394	1.998	0.422	1.007	0.389
Han	0.077	1.226	0.094	2.401	0.105	1.411	0.090
Hezhen	0.104	1.300	0.136	2.498	0.148	1.594	0.131
Itelmen	0.212	1.212	0.254	2.321	0.273	1.419	0.248
Japanese	0.071	1.325	0.094	2.552	0.105	1.637	0.091
Kinh	0.068	1.354	0.094	2.573	0.102	1.724	0.091
Korean	0.081	1.488	0.131	2.818	0.145	2.002	0.124
Koryak	0.202	1.111	0.220	2.190	0.242	1.171	0.214
Lahu	0.067	0.575	0.051	1.490	0.060	-0.103	0.049
Miao	0.077	1.202	0.091	2.340	0.101	1.370	0.088
Mongola	0.122	1.302	0.159	2.520	0.177	1.584	0.153
Naxi	0.075	1.179	0.087	2.310	0.096	1.318	0.084
Nganasan	0.198	1.291	0.255	2.456	0.274	1.587	0.248
Oroqen	0.140	1.252	0.173	2.375	0.185	1.523	0.170
Selkup	0.441	1.120	0.486	2.136	0.514	1.241	0.481
She	0.058	1.353	0.080	2.618	0.090	1.685	0.076
Thai	0.098	1.327	0.132	2.549	0.145	1.649	0.127
Tu	0.134	1.249	0.166	2.419	0.182	1.477	0.160
Tujia	0.075	1.232	0.092	2.409	0.102	1.427	0.088
Tuvinian	0.240	1.154	0.272	2.223	0.294	1.292	0.267
Ulchi	0.111	1.240	0.136	2.392	0.149	1.464	0.132
Xibo	0.121	1.332	0.163	2.557	0.179	1.660	0.156
Yakut	0.222	1.202	0.263	2.308	0.282	1.402	0.257
Yi	0.072	1.334	0.097	2.560	0.107	1.664	0.093
Yukagir	0.298	1.098	0.322	2.103	0.341	1.187	0.319

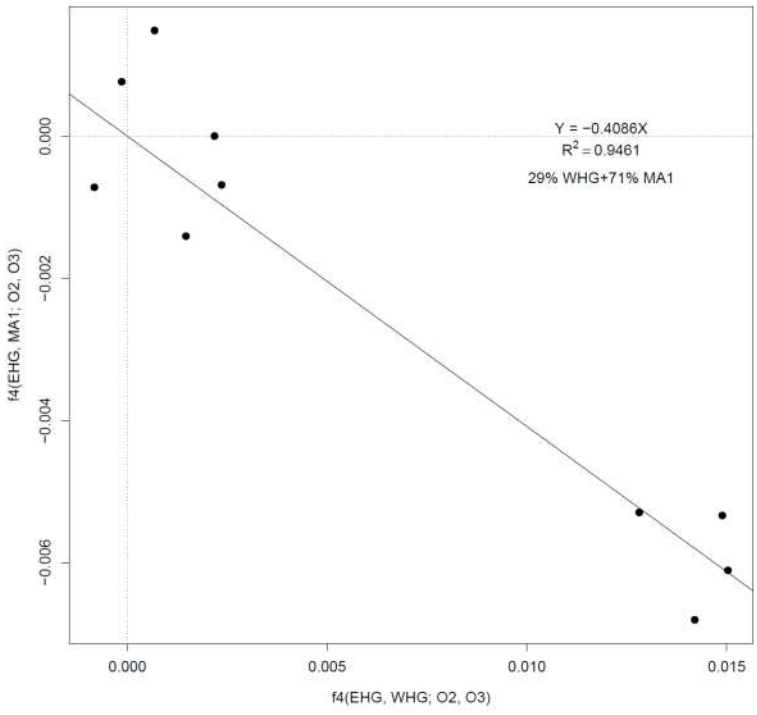
Table S11.5: Modeling EHG as a mixture of WHG and a ghost population on the Onge→*Eastern Eurasian* cline. Projection beyond the (Onge, *Eastern Eurasian*) segment varies from λ~2 to 12 depending on which *Eastern Eurasian* population is used, being higher for populations with less West Eurasian ancestry. However, the proportion from the "Ghost" population is fairly stable (average: 0.755, standard deviation: 0.076, range: 0.571-0.936).

Eastern Eurasian	λ	α_{GHOST}
Ami	9.390	0.661
Atayal	12.148	0.767
Cambodian	7.014	0.663
Chukchi	3.519	0.802
Dai	10.286	0.718
Daur	6.605	0.813
Eskimo	3.430	0.782
Even	2.437	0.936
Han	9.272	0.740
Hezhen	7.037	0.741
Itelmen	3.526	0.762
Japanese	9.732	0.713
Kinh	10.701	0.728
Korean	7.385	0.658
Koryak	3.825	0.782
Lahu	4.943	0.571
Miao	9.932	0.765
Mongola	5.564	0.715
Naxi	10.045	0.766
Nganasan	3.871	0.763
Oroqen	5.874	0.800
Selkup	2.051	0.888
She	9.067	0.638
Thai	6.958	0.713
Tu	5.546	0.749
Tujia	9.880	0.754
Tuvinian	3.549	0.838
Ulchi	7.030	0.771
Xibo	5.770	0.722
Yakut	3.787	0.823
Yi	9.836	0.725
Yukagir	3.087	0.902

The "Ghost" population may correspond to Ancient North Eurasians

We were curious to see if the population represented by the MA1 Upper Paleolithic Siberian[2] could be the "Ghost" population, as it shares more alleles with EHG than with other European hunter-gatherers[3] and there is a cline of shared genetic drift with it in East Asia[2]. Testing Left=(EHG, WHG, MA1), we weakly reject rank=1 (P=0.041). Fig. S11.2 shows a modeling of EHG as a mix of 71% MA1 and 29% WHG, although the regression is driven wholly by statistics involving the Switzerland_HG outgroup. The mixture proportions resemble the ~3/4 "Ghost" and ~1/4 "WHG" inferred for EHG (Table S11.5) and we think it is plausible that an "Ancient North Eurasian"-related population (which, however may not be exactly represented by MA1) was the "Ghost" population.

Figure S11.2: Modeling EHG as a WHG+MA1 mixture. There is a cluster of points about the origin (top-left) and the signal of negative slope is driven by the cluster of four points at the bottom-right that involve statistics of the form *f*4(EHG, MA1; O_2, Switzerland_HG) (that are negative), and *f*4(WHG, EHG; O_2, Switzerland_HG) (that are positive), suggesting that EHG are interemediate between WHG and Switzerland_HG. A previous analysis (Supplementary Information, section 9 of ref.3) prior to the availability of the Switzerland_HG individual did not discover this signal, highlighting the usefulness of including this sample to the set of outgroups as a way to differentiate between European hunter-gatherers like WHG and "Ancient North Eurasians" like MA1.



Another candidate for the source population might be the AG2[2] sample which is similar to MA1 but is more recent (it postdates the Last Glacial Maximum); we interpret this very cautiously as it is of poorer quality has a higher level of contamination, estimated to be ~30% using polymorphism on the X-chromosome for this male individual[2]. Nonetheless, we cannot reject a mixture model (P-value for rank=1 is 0.66) with EHG having 26.6±4.1% WHG and 73.4±4.1% AG2 ancestry. We obtain similar proportions when modeling EHG as a mix of WHG and a "Ghost" population defined wholly by the cline of Eastern Eurasians and when modeling it as a mix of WHG and MA1/AG2, suggesting that these two "Ancient North Eurasian" individuals from Siberia may be genetically close to the sought-after "Ghost".

We repeat the qpAdm admixture analysis of Eastern Eurasians using MA1 and AG2 (rather than EHG) as sources (Table S11.6). AG2 appears to be a valid source for populations across the Eastern Eurasian cline, but MA1 is not, especially for populations with the lowest levels of Onge ancestry. We list mixture proportions using either EHG or AG2 in Extended Data Fig. 5. We also fit a simple model that encapsulates admixture from an AG2-related population into both European hunter-gatherers and Eastern Eurasians (Fig. S11.3), and list mixture proportions for different Eastern Eurasian populations (Table S11.6); most of these are within 3 standard errors of those derived using qpAdm.

Figure S11.3: A simple mixture model deriving ancestry into European hunter-gatherers and East Asians from a common AG2-related "Ancient North Eurasian" source. Fit for Han is shown; mixture proportions for other populations in Table S11.6

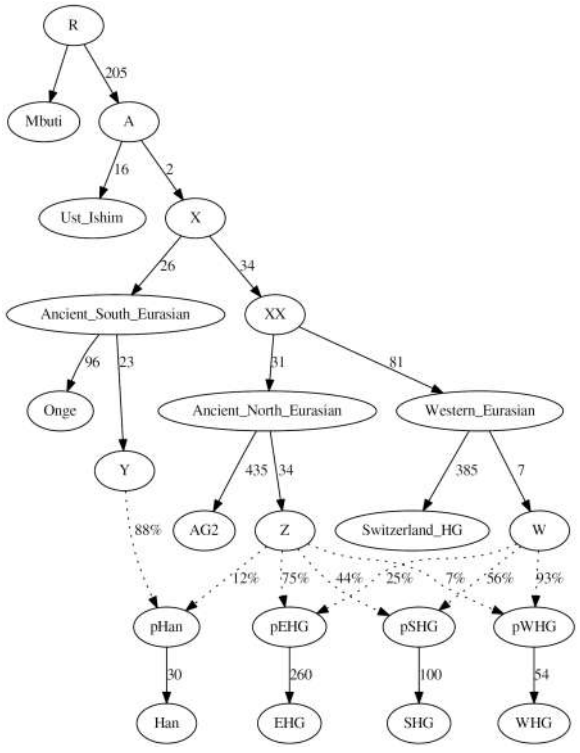


Table S11.6: Modeling Eastern Eurasian populations as a mix of MA1 or AG2 and Onge. We test whether (*Test*, MA1 or AG2, Onge) is consistent with 2 streams of ancestry in relation to the M5 outgroups and list mixture proportions. P-values for rank=1 that are >0.05 are highlighted. Mixture proportions from AG2 according to the model of Fig. S11.3 are shown in the penultimate column and the number of standard errors it differs from the qpAdm model in the last column.

	qpAdm				qpAdm				qpGraph	qpAdm -qpGraph
Test	P-value	Onge	MA1	Std. Err.	P-value	Onge	AG2	Std. Err.	AG2	Z
Altaian	1.17E-04	0.558	0.442	0.027	3.59E-01	0.565	0.435	0.032	0.344	2.9
Ami	7.91E-01	0.913	0.087	0.028	8.67E-01	0.915	0.085	0.027	0.112	-1
Atayal	7.59E-01	0.911	0.089	0.029	9.56E-01	0.908	0.092	0.029	0.143	-1.8
Cambodian	3.17E-01	0.88	0.12	0.025	5.93E-01	0.873	0.127	0.027	0.117	0.4
Chukchi	1.04E-01	0.678	0.322	0.026	4.53E-01	0.698	0.302	0.027	0.357	-2
Dai	7.85E-01	0.905	0.095	0.027	9.27E-01	0.903	0.097	0.027	0.13	-1.2
Daur	7.42E-02	0.835	0.165	0.026	5.32E-01	0.83	0.17	0.027	0.173	-0.1
Dolgan	3.37E-03	0.653	0.347	0.03	2.94E-01	0.653	0.347	0.034	0.256	2.7
Eskimo	2.97E-01	0.679	0.321	0.027	9.49E-01	0.695	0.305	0.028	0.365	-2.1
Even	3.16E-09	0.465	0.535	0.031	2.75E-02	0.491	0.509	0.034	0.419	2.6
Han	4.69E-01	0.892	0.108	0.026	7.50E-01	0.891	0.109	0.026	0.12	-0.4
Hezhen	4.15E-01	0.855	0.145	0.027	7.86E-01	0.855	0.145	0.027	0.166	-0.8
Itelmen	2.56E-01	0.697	0.303	0.028	5.33E-01	0.724	0.276	0.029	0.332	-1.9
Japanese	7.43E-01	0.896	0.104	0.026	9.25E-01	0.899	0.101	0.025	0.131	-1.2
Kalmyk	1.64E-04	0.661	0.339	0.026	2.90E-01	0.656	0.344	0.029	0.266	2.7
Kinh	8.08E-01	0.904	0.096	0.027	9.56E-01	0.903	0.097	0.027	0.114	-0.6
Korean	4.81E-01	0.878	0.122	0.027	7.49E-01	0.878	0.122	0.028	0.114	0.3
Koryak	2.31E-02	0.717	0.283	0.027	2.23E-01	0.739	0.261	0.026	0.337	-2.9
Kusunda	1.17E-03	0.78	0.22	0.025	1.26E-01	0.775	0.225	0.026	0.13	3.7
Kyrgyz	1.83E-05	0.53	0.47	0.027	6.57E-01	0.531	0.469	0.033	0.351	3.6
Lahu	2.38E-01	0.91	0.09	0.027	4.59E-01	0.906	0.094	0.026	0.082	0.5
Miao	6.88E-01	0.892	0.108	0.028	9.40E-01	0.89	0.11	0.028	0.128	-0.6
Mongola	2.22E-01	0.826	0.174	0.027	6.67E-01	0.825	0.175	0.028	0.151	0.9
Naxi	4.06E-01	0.899	0.101	0.026	7.25E-01	0.896	0.104	0.026	0.108	-0.2
Nganasan	6.08E-02	0.725	0.275	0.027	6.78E-01	0.719	0.281	0.031	0.275	0.2
Oroqen	4.93E-01	0.807	0.193	0.026	9.19E-01	0.811	0.189	0.026	0.191	-0.1
Selkup	5.85E-07	0.401	0.599	0.031	1.39E-01	0.415	0.585	0.038	0.476	2.9
She	8.08E-01	0.912	0.088	0.027	8.14E-01	0.912	0.088	0.028	0.1	-0.4
Thai	1.29E-01	0.865	0.135	0.025	5.37E-01	0.856	0.144	0.027	0.134	0.4
Tu	6.80E-02	0.817	0.183	0.026	4.66E-01	0.813	0.187	0.027	0.124	2.3
Tubalar	5.80E-07	0.423	0.577	0.031	2.69E-01	0.441	0.559	0.035	0.444	3.3
Tujia	5.76E-01	0.894	0.106	0.027	8.67E-01	0.895	0.105	0.026	0.107	-0.1
Tuvinian	1.25E-03	0.673	0.327	0.025	3.34E-01	0.676	0.324	0.028	0.278	1.6
Ulchi	3.08E-01	0.844	0.156	0.026	7.15E-01	0.848	0.152	0.026	0.185	-1.2
Xibo	3.46E-01	0.826	0.174	0.026	8.12E-01	0.821	0.179	0.028	0.137	1.5
Yakut	2.44E-02	0.696	0.304	0.025	7.06E-01	0.706	0.294	0.027	0.266	1.1
Yi	7.72E-01	0.896	0.104	0.027	9.22E-01	0.899	0.101	0.026	0.114	-0.5
Yukagir	3.31E-05	0.598	0.402	0.026	1.03E-01	0.616	0.384	0.029	0.33	1.9

The fact that MA1, AG2, and EHG could be related to a population contributing to both European hunter-gatherers and Eastern Eurasians is shown by plotting outgroup f_3-statistics of the form f_3(Mbuti; Papuan, *Test*) and f_3(Mbuti; Switzerland_HG, *Test*) (Extended Data Fig. 8). European hunter-gatherer populations are arrayed on a cline WHG→SHG→EHG and Eastern Eurasian populations form their own cline with Onge just beyond its southern end. Both MA1 and AG2 appear to be placed on the opposite end of the Eastern Eurasian cline. EHG, AG2, and MA1 surround the intersection point of the two clines, consistent with them being plausible sources for many East Asian populations (Tables S11.2, 6)

Future studies of ancient DNA from eastern Europe (where there is currently a ~30,000-year gap between Kostenki14[5] and the EHG[3,10]) and Siberia may identify the sources of this admixture. The fact that the WHG who have been sampled from throughout mainland Europe[1,11,12] as well as the SHG[1,3] possess EHG admixture (Supplementary Information, section 7) suggests that the "Ghost" population admixed into European hunter-gatherers some time before the ~8,000-year old time frame of the available EHG/SHG/WHG samples.

The timing of the inferred admixture in East Asians is unknown, although the fact that Native Americans also have ancestry[6] from "Ancient North Eurasians" (ANE), related to MA1 and AG2, suggests that mixtures between ANE and an eastern non-African populations distantly related to the Onge may predate the settlement of the Americas. We model Native American populations as a mix of Onge and MA1, AG2, or EHG (Table S11.7). Both MA1 and AG2 are valid sources for Native American populations. We estimate that Amazonian populations like the Karitiana and Surui have ~40% of their ancestry from ANE, consistent with previous estimates[1,2,9].

We also include two ancient genomes in Table S11.7, 'Clovis': the ~13,000-year old Anzick-1 from the Clovis culture in Montana[13], and 'Kennewick': the ~8,500-year old Kennewick Man from Washington State[14], whose mixture proportions are similar to present-day Native Americans, showing that the mixture between 'Ancient North Eurasians' and an Onge-like population must have happened at least by ~13,000 years ago. It is important to sample Siberia and East Asia more extensively in order to better understand the history of these two sources of the ancestry of East Eurasians and Native Americans. Mixture proportions for Native Americans (Table S11.7) are at the 'northern' end of those for east Eurasian populations (Tables S11.2, S11.6), consistent with a Siberian origin of Native Americans. However, the ancient geography of the ANE/Onge-related populations is likely to be complex, as present-day Amazonians not only have the 'general' eastern non-African admixture related to the Onge, but also a more specific stream of ancestry related to them and to Australasian populations[9].

Table S11.7: Modeling Native American populations. We model a *Native American* population as a mix of Onge and a population *X*, showing the P-value for rank=1 for Left=(*Native American*, Onge, *X*)

			Mixture Proportions		
X	***Native American***	**P-value for rank=1**	**Onge**	***X***	**Std. Error**
EHG	Karitiana	6.49E-03	0.721	0.279	0.023
MA1	Karitiana	7.04E-01	0.605	0.395	0.031
AG2	Karitiana	1.66E-01	0.613	0.387	0.037
EHG	Surui	1.24E-01	0.721	0.279	0.023
MA1	Surui	4.00E-01	0.610	0.390	0.033
AG2	Surui	9.63E-02	0.631	0.369	0.037
EHG	Mixe	1.92E-02	0.693	0.307	0.022
MA1	Mixe	2.89E-01	0.567	0.433	0.031
AG2	Mixe	7.85E-01	0.587	0.413	0.034
EHG	Pima	1.84E-02	0.702	0.298	0.022
MA1	Pima	1.17E-01	0.584	0.416	0.031
AG2	Pima	7.86E-01	0.599	0.401	0.035
EHG	Clovis	2.60E-04	0.711	0.289	0.029
MA1	Clovis	2.85E-01	0.580	0.420	0.038
AG2	Clovis	1.24E-01	0.580	0.420	0.049
EHG	Kennewick	7.53E-03	0.703	0.297	0.030
MA1	Kennewick	3.49E-01	0.553	0.447	0.046
AG2	Kennewick	8.41E-01	0.581	0.419	0.060

References

1. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, (2014).
2. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87-91, (2014).
3. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211, (2015).
4. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912, (2015).
5. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118, (2014).
6. Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345**, (2014).
7. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489-494, (2009).
8. Moorjani, P. *et al.* Genetic evidence for recent population mixture in India. *Am. J. Hum. Genet.* **93**, 422-438, (2013).
9. Skoglund, P. *et al.* Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104-108, (2015).
10. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503, (2015).
11. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
12. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225-228, (2014).
13. Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225-229, (2014).
14. Rasmussen, M. *et al.* The ancestry and affiliations of Kennewick Man. *Nature* **523**, 455-458, (2015).