University of Huddersfield Repository

Lazaridis, Iosif et al.

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

Original Citation

Lazaridis, Iosif et al. (2016) Genomic insights into the origin of farming in the ancient Near East. Nature, 536 (7617). pp. 419-424. ISSN 00280836

The genetic structure of the world's first farmers

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], Songül Alpaslan Roodenberg[15], György Lengyel[16], Fanny Bocquentin[17], Boris Gasparian[18], Janet M. Monge[19], Michael Gregg[19], Vered Eshed[20], Ahuva-Sivan Mizrahi[20], Christopher Meiklejohn[21], Fokke Gerritsen[22], Luminita Bejenaru[23], Matthias Blüher[24], Archie Campbell[25], Gianpiero Cavalleri[26], David Comas[27], Philippe Froguel[28,29], Edmund Gilbert[26], Shona M. Kerr[25], Peter Kovacs[30], Johannes Krause[31,32,33], Darren McGettigan[34], Michael Merrigan[35], D. Andrew Merriwether[36], Seamus O'Reilly[35], Martin B. Richards[37], Ornella Semino[38], Michel Shamoon-Pour[36], Gheorghe Stefanescu[39], Michael Stumvoll[24], Anke Tönjes[24], Antonio Torroni[38], James F. Wilson[40,41], Loic Yengo[28,29], Nelli A. Hovhannisyan[42], Nick Patterson[2], Ron Pinhasi[7,*,†] and David Reich[1,2,6,*,†]

*Co-senior authors;
†Correspondence and requests for materials should be addressed to:
I. L. (lazaridis@genetics.med.harvard.edu), R. P. (ron.pinhasi@ucd.ie), or D. R. (reich@genetics.med.harvard.edu)

[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] Dept. of Anthropology, Whitman College, Walla Walla, WA 99362, USA
[5] Dept. of Archaeology, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada
[6] Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 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] Dept. of Anthropology, Emory University, Atlanta, Georgia 30322, USA
[11] Dept. of Organismic and Evolutionary Biology, Harvard University, Cambridge 02138, USA
[12] Dept. 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] Dept. of Biology and Evolution, University of Ferrara, Ferrara I-44121, Italy
[15] Independent Researcher, Santpoort-Noord, The Netherlands
[16] Department of Prehistory and Archaeology, University of Miskolc, 3515 Miskolc-Egyetemváros, Hungary
[17] French National Centre for Scientific Research, UMR 7041, 92023 Nanterre Cedex, France
[18] Institute of Archaeology and Ethnology,National Academy of Sciences of the Republic of Armenia, 0025 Yerevan, Republic of Armenia
[19] University of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, PA 19104, USA
[20] Israel Antiquities Authority, Jerusalem 91710, Israel
[21] Dept. of Anthropology, University of Winnipeg, Winnipeg, Manitoba R3B 2E9, Canada

[22] Netherlands Institute in Turkey, Istiklal Caddesi, Nur-i Ziya Sokak 5, Beyoğlu, Istanbul, Turkey
[23] Alexandru Ioan Cuza University of Iasi, Romania 700505, Romania
[24] Department of Internal Medicine and Dermatology, Clinic of Endocrinology and Nephrology, 04103 Leipzig, Germany
[25] Generation Scotland, Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, Scotland
[26] RCSI Molecular & Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland
[27] Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08003 Barcelona, Spain
[28] Univ. Lille, CNRS, Institut Pasteur de Lille, UMR 8199 - EGID, F-59000 Lille, France
[29] Department of Genomics of common disease, London Hammersmith Hospital, London W12 0HS, UK
[30] Leipzig University Medical Center, IFB AdiposityDiseases, 04103 Leipzig, Germany
[31] Max Planck Institute for the Science of Human History, 07745 Jena, Germany
[32] Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany
[33] Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, 72072 Tübingen, Germany
[34] Independent research, County Wicklow, Ireland
[35] Genealogical Society of Ireland, Dún Laoghaire, County Dublin, Ireland
[36] Department of Anthropology, Binghamton University, New York 13902, USA
[37] Department of Biological Sciences, School of Applied Sciences, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK
[38] Dipartimento di Biologia e Biotecnologie "L. Spallanzani", Università di Pavia, 27100 Pavia, Italy
[39] Institutul de Cercetari Biologice, Iaşi 700505, Romania
[40] Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh EH8 9AG, Scotland
[41] MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, Scotland
[42] Dept. of Ecology and Nature Protection, Yerevan State University, 0025 Yerevan, Republic of Armenia

We report genome-wide ancient DNA from 44 ancient Near Easterners ranging in time between ~12,000-1,400 BCE, 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 prior to 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 drastically 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 from 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 10,000-9,000 BCE, 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, due 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] that 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 to represent each SNP, restricting to individuals with at least 9,000 SNPs covered at least once. We obtained genome-wide data passing quality control for 45 individuals on whom we had a median coverage of 172,819 SNPs (Methods). We assembled radiocarbon dates for 26 individuals (22 new generated for this study) (Supplementary Data Table 1).

The newly reported ancient individuals date to ~12,000-1,400 BCE and come from the southern Caucasus (Armenia), northwestern Anatolia (Turkey), Iran, and the southern Levant (Israel and Jordan) (Supplementary Data Table 1, 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 Epipaleolithic Natufian hunter-gatherers from Raqefet Cave in the Levant (12,000-9,800 BCE); a likely Mesolithic individual from Hotu Cave in the Alborz mountains of Iran (probable date of 9,100-8,600 BCE); Pre-Pottery Neolithic farmers from ‘Ain Ghazal and Motza in the southern Levant (8,300-6,700 BCE); and early farmers from Ganj Dareh in the Zagros mountains of western Iran (8,200-7,600 BCE). The samples also include later Neolithic, Chalcolithic (~4,800-3,700 BCE), and Bronze Age (~3,350-1,400 BCE) individuals (Supplementary Information, section 1). We combined our data with previously published ancient data[7,8,9,10,8,10-15] to form a dataset of 281 ancient individuals. We then further merged with 2,583 present-day people genotyped on the Affymetrix Human Origins array[13,16] (238 new) (Supplementary Data Table 2; Supplementary Information, section 2). We grouped the ancient individuals based on archaeological culture and chronology (Fig. 1a; Supplementary Data Table 1). We refined the grouping based on patterns evident in Principal Components Analysis (PCA)[17] (Fig. 1b; Extended Data Fig. 1), ADMIXTURE model-based clustering[18] (Fig. 1c), and ‘outgroup’ f_3-analysis (Extended Data Fig. 2). We used f_4-statistics to identify outlier individuals and to cluster phylogenetically indistinguishable groups into ‘Analysis Labels’ (Supplementary Information, section 3).

We analyzed 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 ancestry was pervasive in the ancient Near East and associated with reduced Neanderthal ancestry

The 'Basal Eurasians' are a lineage hypothesized[13] to have split off prior to the differentiation of all other Eurasian lineages, including both eastern non-African populations like the Han Chinese, and even the early diverged lineage represented by the genome sequence of the ~45,000 year old Upper Paleolithic 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 their having ancestry from a deeply divergent Eurasian lineage that separated from the ancestors of most Eurasians prior to the separation of Han and Ust'-Ishim. We used *qpAdm*[7] to estimate Basal Eurasian ancestry in each *Test* population. We obtain 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 Epipaleolithic 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[19,20-23], which could be explained if West Eurasians (but not East Asians) have partial ancestry from a source diluting their Neandertal inheritance[21]. 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; Fig. 2). By extrapolation, we infer that the Basal Eurasian population had lower Neanderthal ancestry than non-Basal Eurasian populations and possibly none (ninety-five percent 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[19,24]. This could be explained if Basal Eurasians thoroughly admixed into the Near East before the time of the samples we analyzed 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 that the Natufians may have migrated from north or sub-Saharan Africa[25,26], a result that finds some support from Y chromosome analysis which shows that the Natufians and successor Levantine Neolithic populations carried haplogroup E, of likely ultimate African origin, which has not been detected in other ancient males from West Eurasia (Supplementary Information, section 6) [7,8]. 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[27,28].) 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 regional differentiation in the ancient Near East

PCA on present-day West Eurasian populations (Methods) (Extended Data Fig. 1) on which we projected the ancient individuals (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 (Extended Data Fig. 1)[7,8,13]. Ancient samples from the Levant project 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, and 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 the hypothesis 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*[7] (Supplementary Information, section 7).

We computed squared allele frequency differentiation between all pairs of ancient West Eurasians[29] (Methods; Fig. 3; Extended Data Fig. 3), 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 Data Table 3). In 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. These results point to a demographic process that established high differentiation across West Eurasia and then reduced this differentiation over time.

Continuity between pre-farming hunter-gatherers and early farmers of the Near East

Our data document continuity across the hunter-gatherer / farming transition, 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 Caucasus Hunter Gatherers with Iranian farmers (Fig. 1b; Extended Data Fig. 1; Extended Data Fig. 2). 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).

How diverse first farmers of the Near East mixed to form the region's later populations

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 intermediate between two *Reference* populations[16] (Extended Data Table 2). To better understand the admixture history beyond these patterns, we used *qpAdm*[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; Supplementary Information, section 7).

Among first farmers, those of the Levant trace ~2/3 of their ancestry to people related to Natufian hunter-gatherers and ~1/3 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 only model populations of the Chalcolithic and subsequent Bronze Age 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 Caucasus Hunter Gatherers (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 ~44% of the ancestry of Levantine Bronze Age populations in the south and ~33% of the ancestry of the Chalcolithic northwest Anatolians in the west. Our analysis show that the ancient populations of the 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; Supplementary Information, section 7).

The Near Eastern contribution to Europeans, East Africans and South Asians

Admixture did not only occur within the Near East but extended towards Europe. To the north, a population related to people of the Iran Chalcolithic contributed ~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, by hypothesizing a population related to present-day Armenians as a source[7,8]. To the west, the early farmers of mainland Europe were descended from a population related to Neolithic northwestern Anatolians[8]. This is consistent with an Anatolian origin of farming in Europe, but does not reject other sources, since 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[30,31], however, since they share more allele 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).

Migrations from the Near East also occurred towards the southwest into East African populations which experienced West Eurasian admixture ~1,000 BCE[32,33]. Previously, the West Eurasian population known to be the best proxy for this ancestry was present-day Sardinians[33], who resemble Neolithic Europeans genetically[13,34]. 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. 4; 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[35,36] (Supplementary Information, section 9) (Extended Data Fig. 4). 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 ancestry related to both early farmers of western Iran and to 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[35,36] is inferred to have ~18% steppe-related ancestry, while the Kalash of Pakistan are inferred to have ~50%, similar to present-day northern Europeans[7].

Broader insights into population transformations across 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. 5).

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; Extended Data Fig. 6). 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 while withholding major populations, we validated some of our key inferences, successfully inferring mixture proportions consistent with those obtained when the populations are included in the analysis. Application of this methods highlights the impact of Ancient North Eurasian (ANE) ancestry related to the ~22,000 BCE Mal'ta 1 and ~15,000 BCE Afontova Gora 2[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; Extended Data Fig. 7), ranging from ~40% ANE in Native Americans matching previous findings[13,15], to no less than ~5-10% ANE in diverse East Asian groups including Han Chinese (Extended Data Fig. 4; Extended Data Fig. 6f). We also document a cline of ANE ancestry across the east-west extent of Eurasia. Eastern Hunter Gatherers (EHG) derive ~3/4 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 the Upper Paleolithic 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[37] that contribute 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 of 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.

Acknowledgements

We thank the 238 human subjects who voluntarily donated the samples whose genome-wide data we newly report in this study. We thank D. Labuda for sharing the collection of DNA samples from Poland, and P. Zalloua for sharing the collection of DNA samples from Lebanon. 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 M. Ferry and M. Michel for their work on the in-solution enrichment experiments. S.C. was supported 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 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 from Generation Scotland received funding from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006], while the Scottish Donor DNA Databank (GS:3D) was funded by 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 grants from the Collaborative Research Center funded by the German Research Foundation (CRC 1052; B01, B03, C01). M.S.-P. was partially funded by a Wenner-Gren Foundation Dissertation Fieldwork Grant (#9005), and by the National Science Foundation DDRIG (BCS-1455744). P.K. was supported by the Federal Ministry of Education and Research (BMBF), Germany (FKZ: 01EO1501). J.F.W acknowledges the MRC “QTL in Health and Disease” programme grant. The Romanian contribution to this work was supported by the EC Commission, Directorate General XII, within the framework of the Cooperation in Science and Technology with Central and Eastern European Countries (Supplementary Agreement ERBCIPDCT 940038 to the Contract ERBCHRXCT 920032, coordinated by Prof. 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 strategic theme "Towards a governance model for international migration: an interdisciplinary and diachronic perspective" (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., 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 archaeological material. N.R., D.F., M.N., B.Gam., K.Si., S.C., K.St., E.H., Q.F., G.G.-F., 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.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. analyzed data. I.L., R.P. and D.R. wrote the manuscript and supplements. All authors read the manuscript and provided comments.

Author Information

The aligned sequences are available through the European Nucleotide Archive under accession number xxx. 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).

Online Methods

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[38], 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[39], of which we treated 167 with Uracil-DNA glycosylase (UDG) to remove the characteristic C-to-T errors of ancient DNA[40] . 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[39]. 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 1x Isothermal Amplification buffer (NEB) in a final volume of 25μL fill-in reaction, and second, we used the entire inactivated 25μL fill-in reaction in a total volume of 100μL PCR mix with 1 μM of each primer[41]. 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 lab 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[42], using the protocol described previously[7]. We sequenced the enriched libraries using an Illumina NextSeq 500 instrument using 2×76bp 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[43], using the Burrows Wheeler Aligner[44] (*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 (http://www.nature.com/nature/journal/v522/n7555/abs/nature14317.html#supplementary-information) and Supplementary Data 1 of ref. 6. (http://www.nature.com/nature/journal/v524/n7564/full/nature14558.html#supplementary-information). We sequenced the libraries on an Illumina NextSeq 500 using 2×76bp 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 2 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 10-fold 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 Data Table 1). We quantified the rate of C-to-T substitution in the final nucleotide of the sequences analyzed, relative to the human reference genome sequence, and found that all the libraries analyzed had rates of at least 3%[39], consistent with genuine ancient DNA. For the nuclear data from males, we used the ANGSD software[45] to estimate a conservative X chromosome estimate of contamination. We determined that all libraries passing 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 Data 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 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. De-identified 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 samples, researchers should a signed letter to D.R. containing the following text: "(a) I will not distribute the samples marked "signed letter" outside my collaboration; (b) I will not post data from the samples marked "signed letter" publicly; (c) I will make no attempt to connect the genetic data for the samples marked "signed letter" to personal identifiers; (d) I will not use the data for samples marked "signed letter" for commercial purposes." Supplementary Data 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 Data 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 1240k 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 use 235 individuals from Pagani et al.[32] on 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 of the *HO* dataset on 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[46,47] 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 Fig. 1c (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 jackknife[48]. 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[49]. We computed the dendrogram of Extended Data Fig. 2 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 jackknife[48], 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 modeling

We carried out Admixture Graph modeling 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*[35,50] 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[50] 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 related differently to the ‘Right’ set.

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 modeling) 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 jackknife for the difference between the statistics $f_4(u_0, \textit{Test}; v_0, v)$ and $f_4(u_0, \textit{Estimated Test}; v_0, v)$ 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.

Modeling 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).

Figures Legends

Figure 1: Genetic structure of ancient West Eurasia. (a) Sampling locations and times in six West Eurasian regions. Sample sizes for each population are given below each bar. Abbreviations used: 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 Paleolithic Ust_Ishim, Kostenki14, and MA1). To avoid visual clutter, population labels of present-day individuals are shown in Extended Data Fig. 1. (c) 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.

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 dramatic decrease over time in West Eurasia. (a) Pairwise F_{ST} between 19 Ancient West Eurasian populations (arranged in approximate chronological order), and select present-day populations. (b) 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.

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. Clouds 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.

Extended Data Tables and Extended Data Figure Legends

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

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

Extended Data Table 2: Admixture *f*$_3$-statistics. We show the lowest Z-score of the statistic $f_3(Test; Reference_1, Refrence_2)$ for every ancient *Test* population 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.

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

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 color 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: 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 ones admixed with steppe ancestry; populations from the Levant from the Epipaleolithic (Natufians) to the Bronze Age; populations from Iran from the Mesolithic to the Late Neolithic.

Extended Data Figure 3: Reduction of genetic differentiation in West Eurasia over time. We measure differentiation by F_{ST}. Each column of the 5x5 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 4: 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 of Pagani et al. (2012); the remainder of the ancestry is a clade with Mota, a ~4,500 year old sample from Ethiopia. (c) EHG ancestry in Eastern Eurasians, or (d) Afontova Gora (AG2) 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 5: Inferred position of ancient populations in West Eurasian PCA according to the model of Fig. 4.

Extended Data Figure 6: Admixture from ghost populations using 'cline intersection'. We model each *Test* population (purple) in panels (a-f) 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 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 7: Admixture from a 'ghost' ANE population into both European and Eastern Eurasian ancestry. EHG, and Upper Paleolithic 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*).

References

1. Barker, G. & Goucher, C. *The Cambridge World History Volume II: A world with agriculture, 12,000 BCE-500 CE*. (Cambridge University Press, 2015).
2. Cavalli-Sforza, L. L. & Paolo, M. *The History and Geography of Human Genes*. (Princeton University 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-+, (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. *Proceedings of the National Academy of Sciences* **112**, 11917-11922, (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. Kim, Bernard Y. & Lohmueller, Kirk E. Selection and Reduced Population Size Cannot Explain Higher Amounts of Neandertal Ancestry in East Asian than in European Human Populations. *The American Journal of Human Genetics* **96**, 454-461, (2015).
21. Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* **338**, 222-226, (2012).

22. Vernot, B. & Akey, Joshua M. Complex History of Admixture between Modern Humans and Neandertals. *The American Journal of Human Genetics* **96**, 448-453, (2015).
23. Wall, J. D. *et al.* Higher Levels of Neanderthal Ancestry in East Asians Than in Europeans. *Genetics* **194**, 199-209, (2013).
24. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, (2010).
25. 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).
26. Ferembach, D. Squelettes du Natoufien d'Israel., etude anthropologique. . *L'Anthropologie* **65**, 46-66, (1961).
27. 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).
28. Henn, B. M. *et al.* Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS genetics* **8**, e1002397, (2012).
29. Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting F(ST): The impact of rare variants. *Genome Res.* **23**, 1514-1521, (2013).
30. Fernandez, 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 genetics* **10**, e1004401, (2014).
31. 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).
32. 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).
33. Pickrell, J. K. *et al.* Ancient west Eurasian ancestry in southern and eastern Africa. *Proc. Natl. Acad. Sci. USA* **111**, 2632–2637, (2014).
34. 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).
35. Moorjani, P. *et al.* Genetic evidence for recent population mixture in India. *Am. J. Hum. Genet.* **93**, 422-438, (2013).
36. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489-494, (2009).
37. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205, (2016).
38. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 15758-15763, (2013).
39. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **370**, DOI: 10.1098/rstb.2013.0624 (2014).
40. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic acids research* **38**, e87, (2010).
41. Korlevic, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *BioTechniques* **59**, 87-93, (2015).
42. Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403-406, (2014).
43. 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).
44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, (2009).
45. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 1-13, (2014).
46. 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).
47. Chang, C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7, (2015).

48. Busing, F. T. A., Meijer, E. & Leeden, R. Delete-m Jackknife for Unequal m. *Statistics and Computing* **9**, 3-8, (1999).
49. Sudmant, P. H. *et al.* Global diversity, population stratification, and selection of human copy-number variation. *Science* **349**, (2015).
50. Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370-374, (2012).

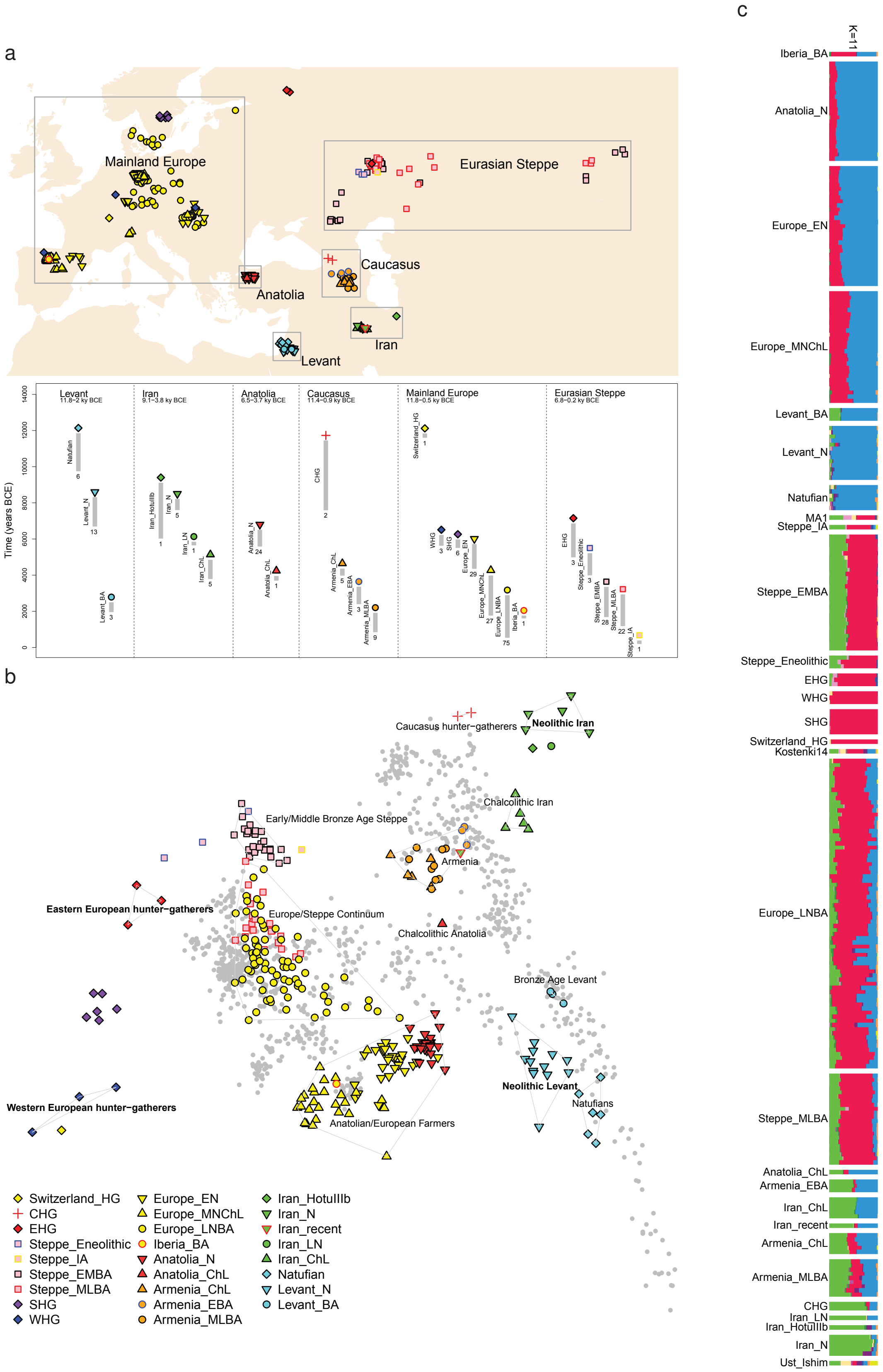

a
Mainland Europe
Eurasian Steppe
Caucasus
Anatolia
Iran
Levant
Time (years BCE)
Levant 11.8–2 ky BCE
Iran 9.1–3.8 ky BCE
Anatolia 6.5–3.7 ky BCE
Caucasus 11.4–0.9 ky BCE
Mainland Europe 11.8–0.5 ky BCE
Eurasian Steppe 6.8–0.2 ky BCE
Natufian
Levant_N
Levant_BA
Iran_HotuIIIb
Iran_N
Iran_LN
Iran_ChL
Anatolia_N
Anatolia_ChL
CHG
Armenia_ChL
Armenia_EBA
Armenia_MLBA
Switzerland_HG
WHG
SHG
Europe_EN
Europe_MNChL
Europe_LNBA
Iberia_BA
EHG
Steppe_Eneolithic
Steppe_EMBA
Steppe_MLBA
Steppe_IA
b
Caucasus hunter−gatherers
Neolithic Iran
Chalcolithic Iran
Early/Middle Bronze Age Steppe
Armenia
Eastern European hunter−gatherers
Europe/Steppe Continuum
Chalcolithic Anatolia
Bronze Age Levant
Neolithic Levant
Natufians
Western European hunter−gatherers
Anatolian/European Farmers
Switzerland_HG
CHG
EHG
Steppe_Eneolithic
Steppe_IA
Steppe_EMBA
Steppe_MLBA
SHG
WHG
Europe_EN
Europe_MNChL
Europe_LNBA
Iberia_BA
Anatolia_N
Anatolia_ChL
Armenia_ChL
Armenia_EBA
Armenia_MLBA
Iran_HotuIIIb
Iran_N
Iran_recent
Iran_LN
Iran_ChL
Natufian
Levant_N
Levant_BA
c
K=11
Iberia_BA
Anatolia_N
Europe_EN
Europe_MNChL
Levant_BA
Levant_N
Natufian
MA1
Steppe_IA
Steppe_EMBA
Steppe_Eneolithic
EHG
WHG
SHG
Switzerland_HG
Kostenki14
Europe_LNBA
Steppe_MLBA
Anatolia_ChL
Armenia_EBA
Iran_ChL
Iran_recent
Armenia_ChL
Armenia_MLBA
CHG
Iran_LN
Iran_HotuIIIb
Iran_N
Ust_Ishim

Figure 1

Figure 2

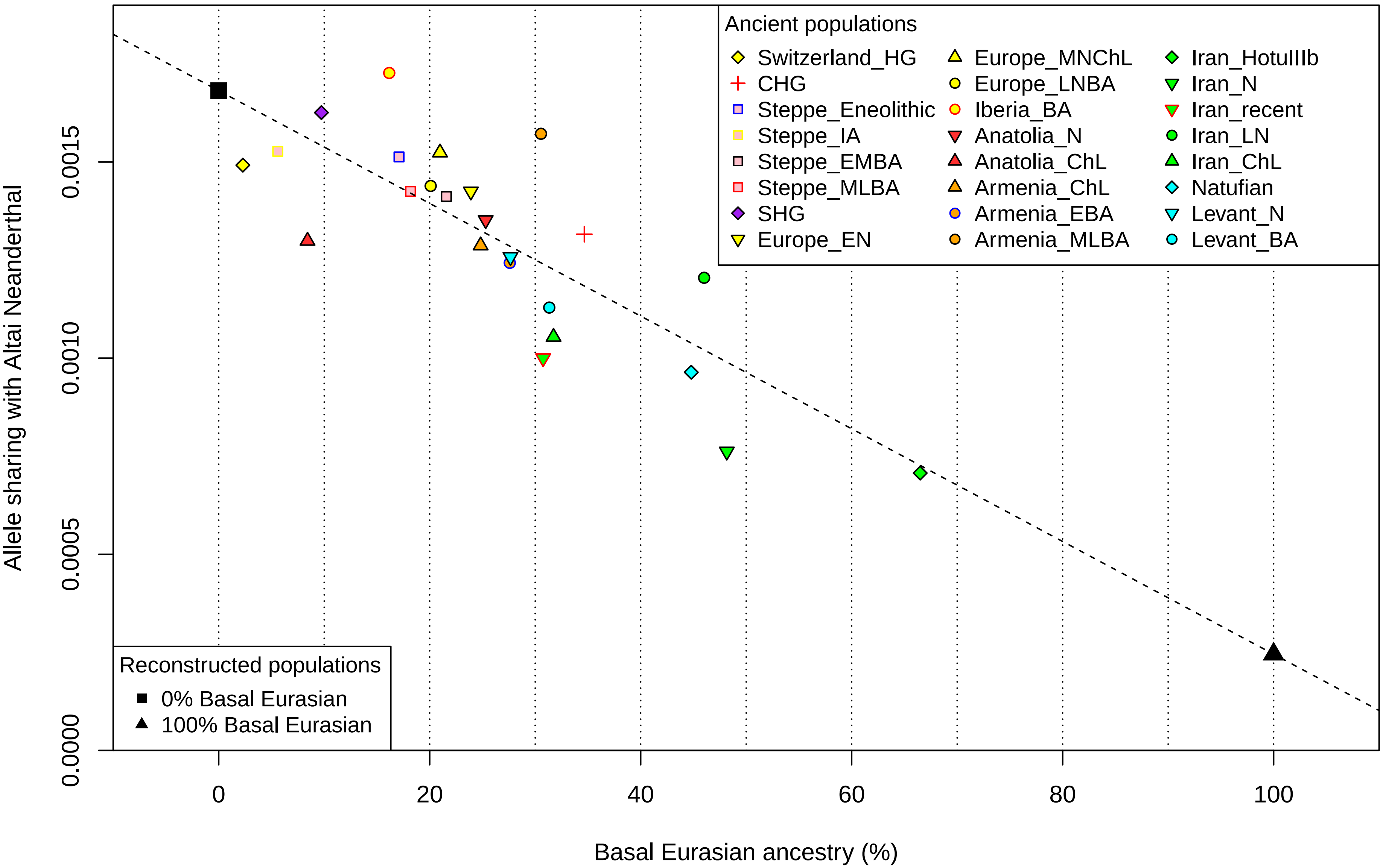
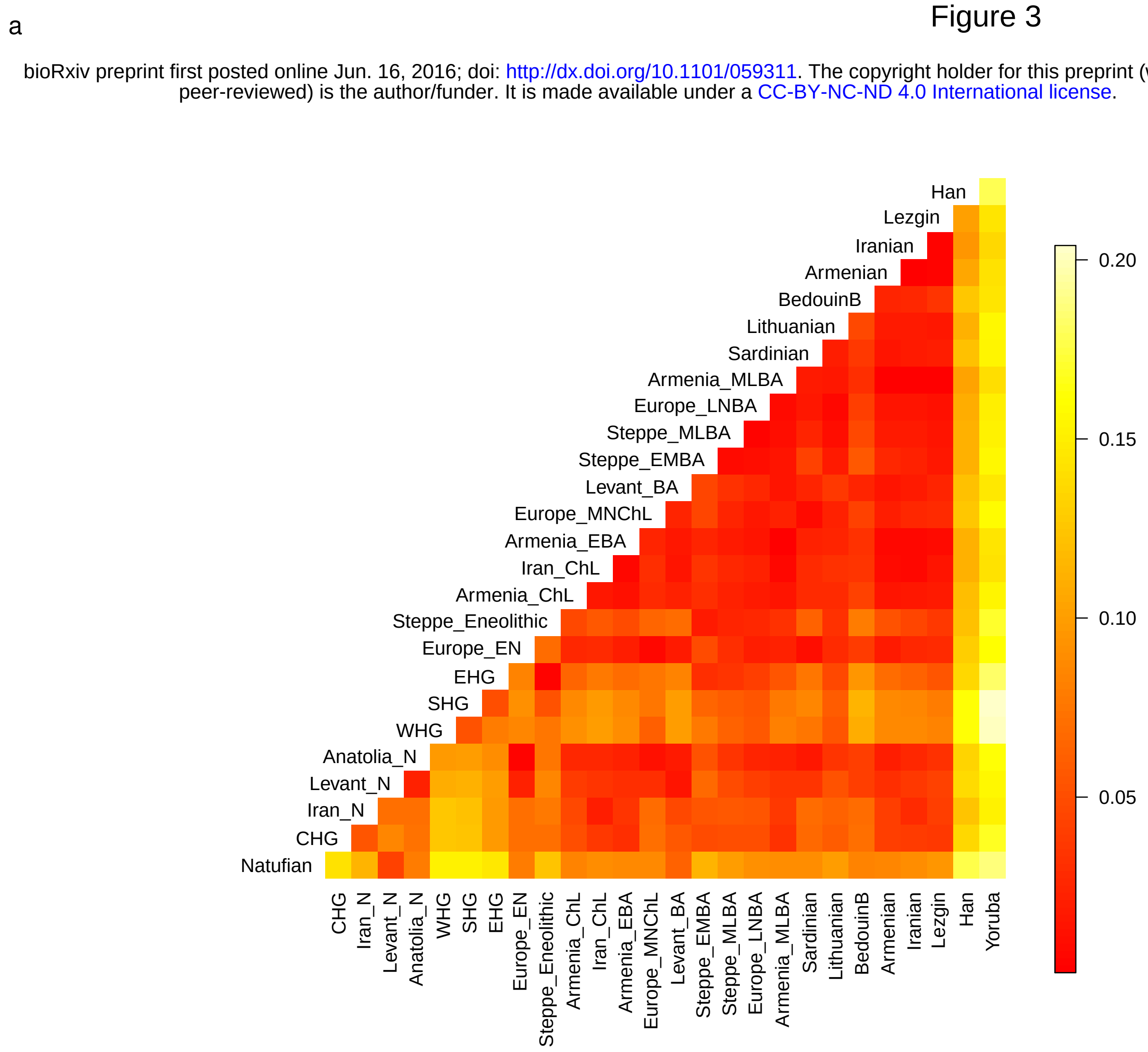
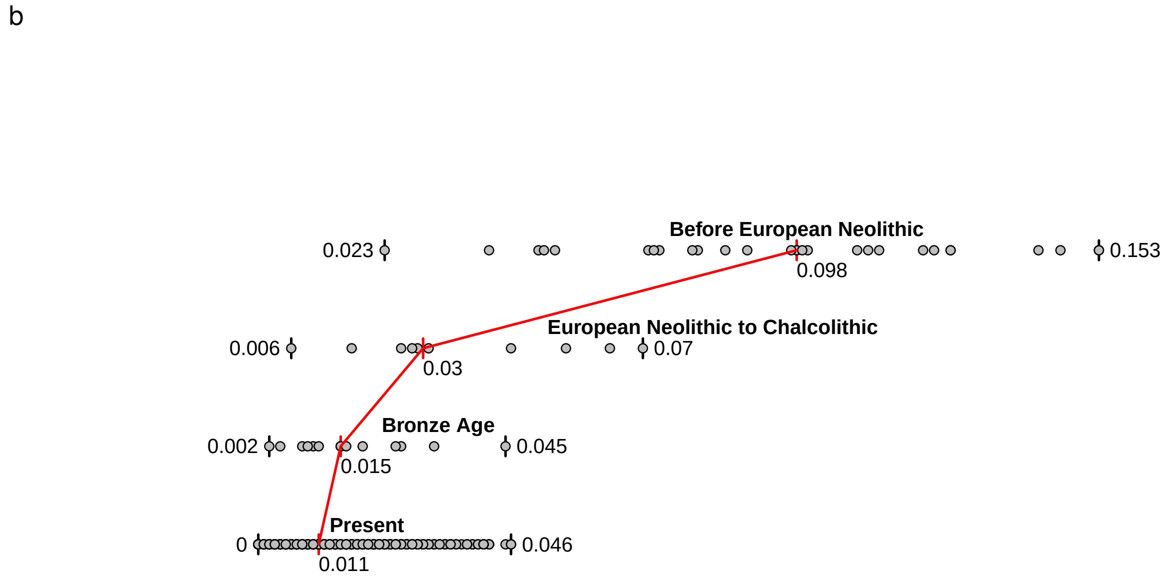

Ancient populations
Switzerland_HG
CHG
Steppe_Eneolithic
Steppe_IA
Steppe_EMBA
Steppe_MLBA
SHG
Europe_EN
Europe_MNChL
Europe_LNBA
Iberia_BA
Anatolia_N
Anatolia_ChL
Armenia_ChL
Armenia_EBA
Armenia_MLBA
Iran_HotuIIIb
Iran_N
Iran_recent
Iran_LN
Iran_ChL
Natufian
Levant_N
Levant_BA
Reconstructed populations
0% Basal Eurasian
100% Basal Eurasian
Allele sharing with Altai Neanderthal
0.0000
0.0005
0.0010
0.0015
0
20
40
60
80
100
Basal Eurasian ancestry (%)

Figure 3


a
Han
Lezgin
Iranian
Armenian
BedouinB
Lithuanian
Sardinian
Armenia_MLBA
Europe_LNBA
Steppe_MLBA
Steppe_EMBA
Levant_BA
Europe_MNChL
Armenia_EBA
Iran_ChL
Armenia_ChL
Steppe_Eneolithic
Europe_EN
EHG
SHG
WHG
Anatolia_N
Levant_N
Iran_N
CHG
Natufian
CHG
Iran_N
Levant_N
Anatolia_N
WHG
SHG
EHG
Europe_EN
Steppe_Eneolithic
Armenia_ChL
Iran_ChL
Armenia_EBA
Europe_MNChL
Levant_BA
Steppe_EMBA
Steppe_MLBA
Europe_LNBA
Armenia_MLBA
Sardinian
Lithuanian
BedouinB
Armenian
Iranian
Lezgin
Han
Yoruba
0.20
0.15
0.10
0.05


b
Before European Neolithic
0.023
0.098
0.153
European Neolithic to Chalcolithic
0.006
0.03
0.07
Bronze Age
0.002
0.015
0.045
Present
0
0.011
0.046

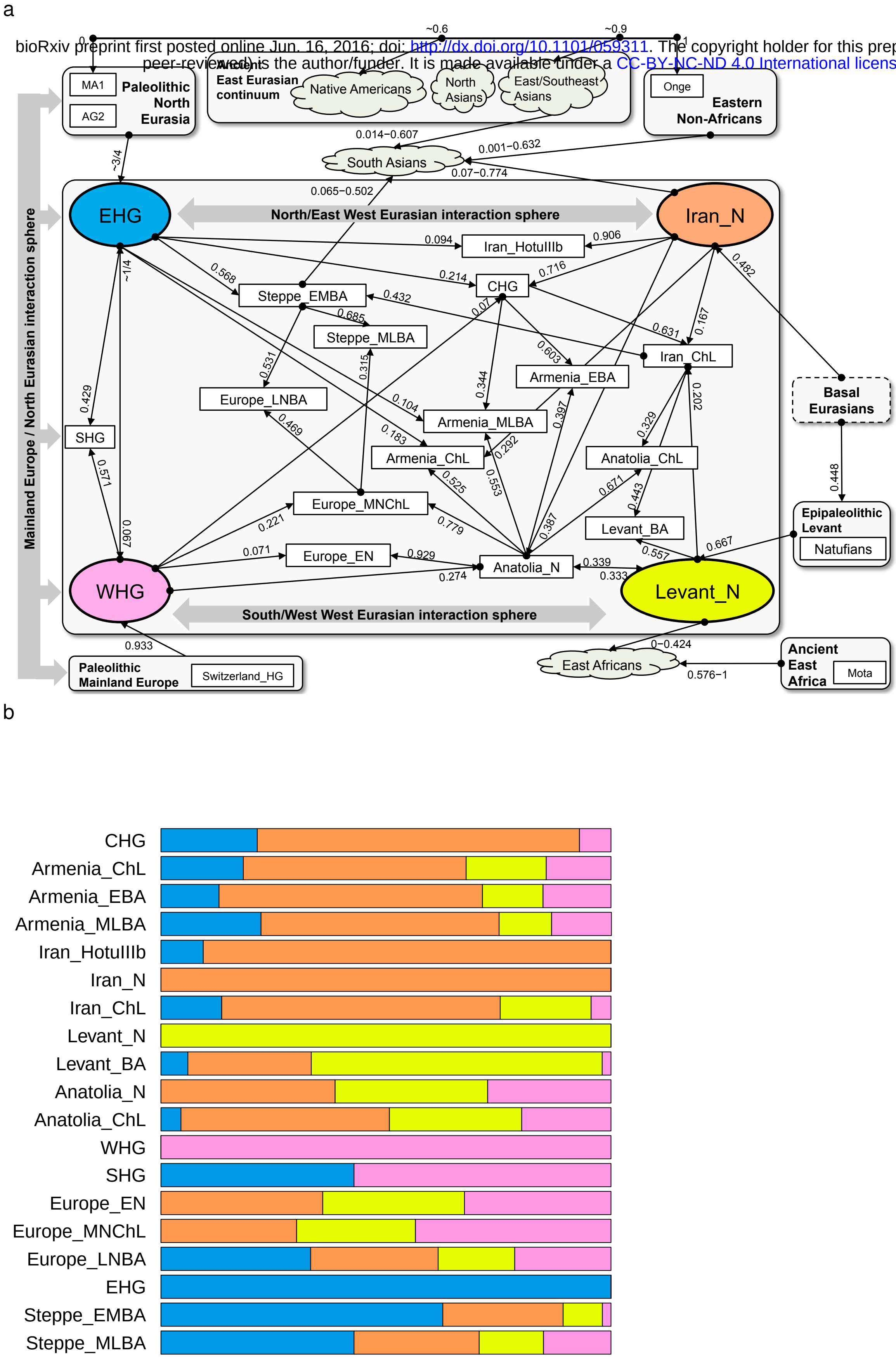
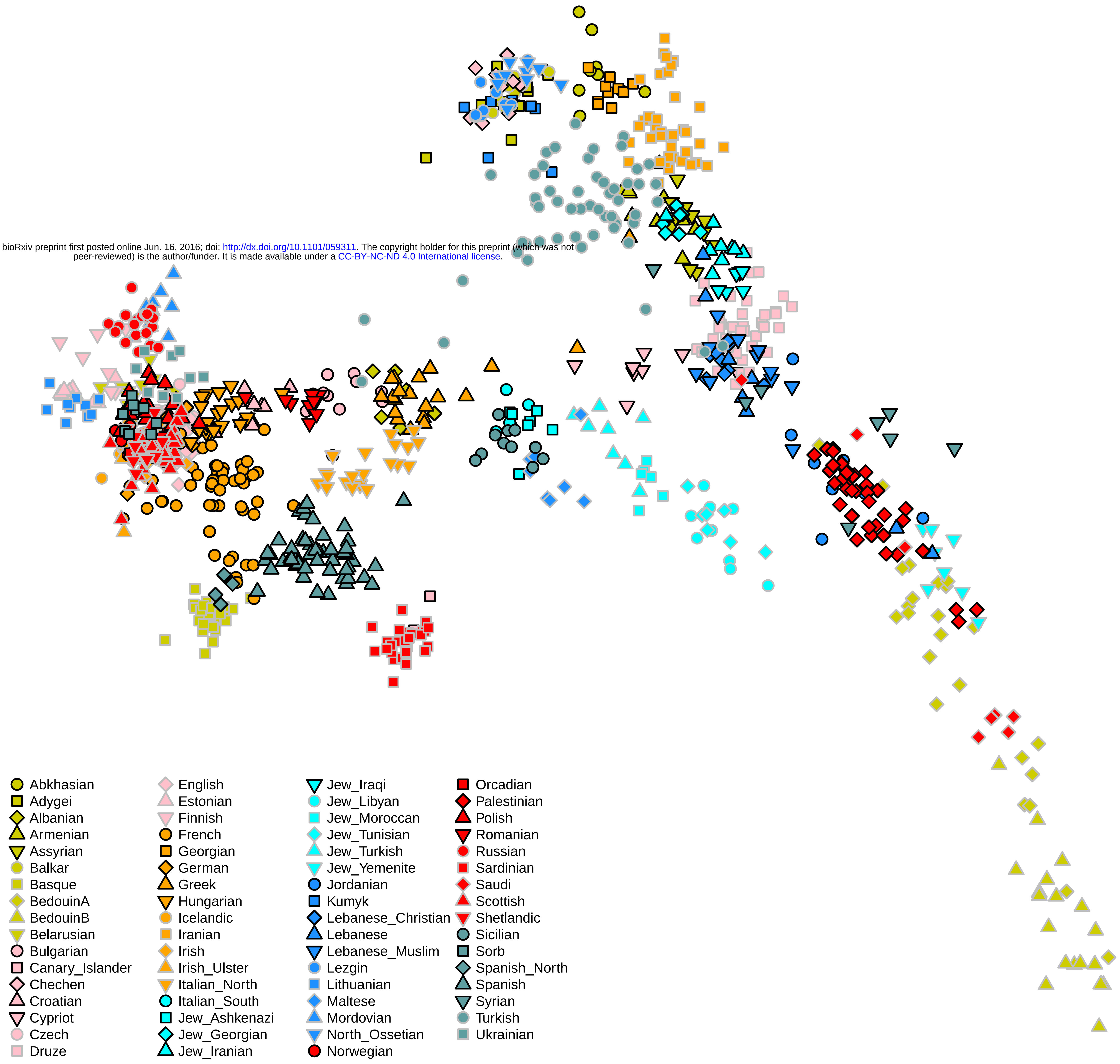
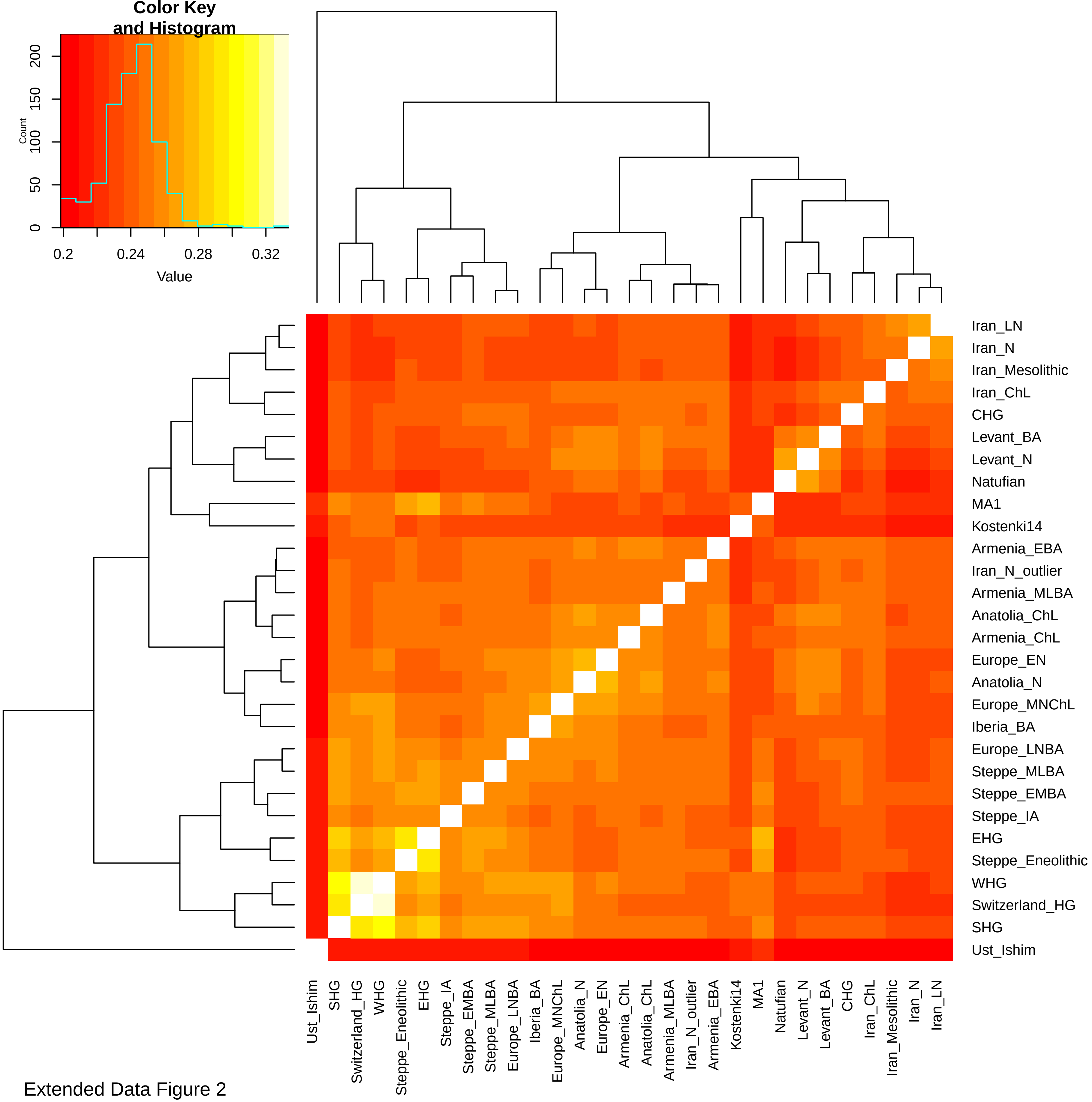

a
0
~0.6
~0.9
1
MA1
AG2
Paleolithic North Eurasia
Ancient East Eurasian continuum
Native Americans
North Asians
East/Southeast Asians
Onge
Eastern Non-Africans
0.014−0.607
0.001−0.632
South Asians
0.07−0.774
0.065−0.502
~3/4
EHG
North/East West Eurasian interaction sphere
Iran_N
0.094
Iran_HotuIIIb
0.906
0.568
0.214
CHG
0.716
0.482
Steppe_EMBA
0.432
0.685
Stepppe_MLBA
0.07
0.631
0.167
Iran_ChL
0.603
Armenia_EBA
0.202
0.531
0.315
0.344
Europe_LNBA
0.104
Armenia_MLBA
0.397
0.329
Basal Eurasians
0.429
SHG
0.469
0.183
Armenia_ChL
0.292
Anatolia_ChL
0.448
0.571
0.525
0.553
0.671
Europe_MNChL
0.443
Epipaleolithic Levant
Natufians
0.221
0.779
Levant_BA
0.387
0.067
0.071
Europe_EN
0.929
0.557
0.667
Anatolia_N
0.339
0.274
0.333
~1/4
WHG
Levant_N
South/West West Eurasian interaction sphere
Mainland Europe / North Eurasian interaction sphere
0.933
0−0.424
East Africans
0.576−1
Paleolithic Mainland Europe
Switzerland_HG
Ancient East Africa
Mota
b
CHG
Armenia_ChL
Armenia_EBA
Armenia_MLBA
Iran_HotuIIIb
Iran_N
Iran_ChL
Levant_N
Levant_BA
Anatolia_N
Anatolia_ChL
WHG
SHG
Europe_EN
Europe_MNChL
Europe_LNBA
EHG
Steppe_EMBA
Steppe_MLBA

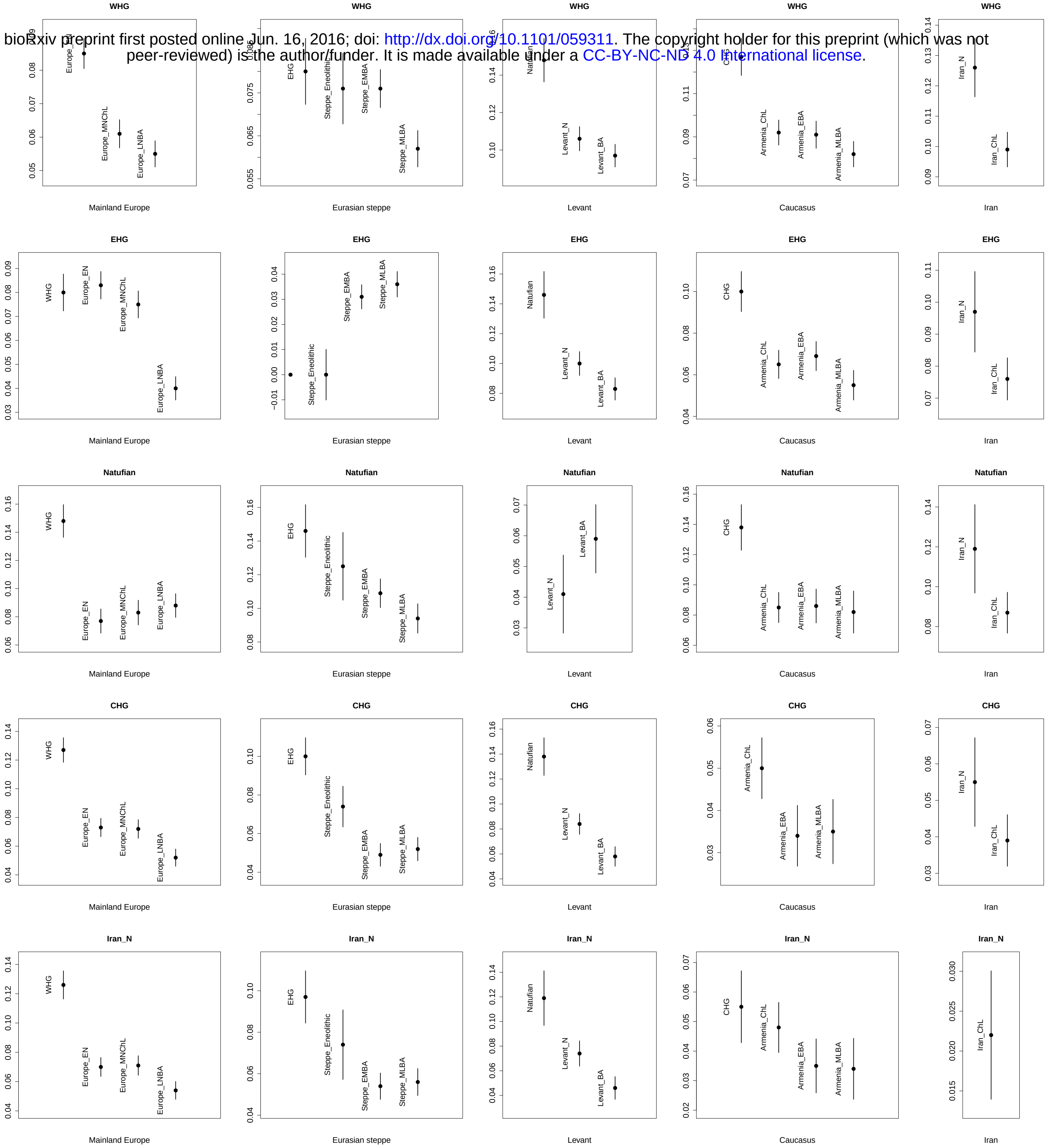
Figure 4

Extended Data Figure 1

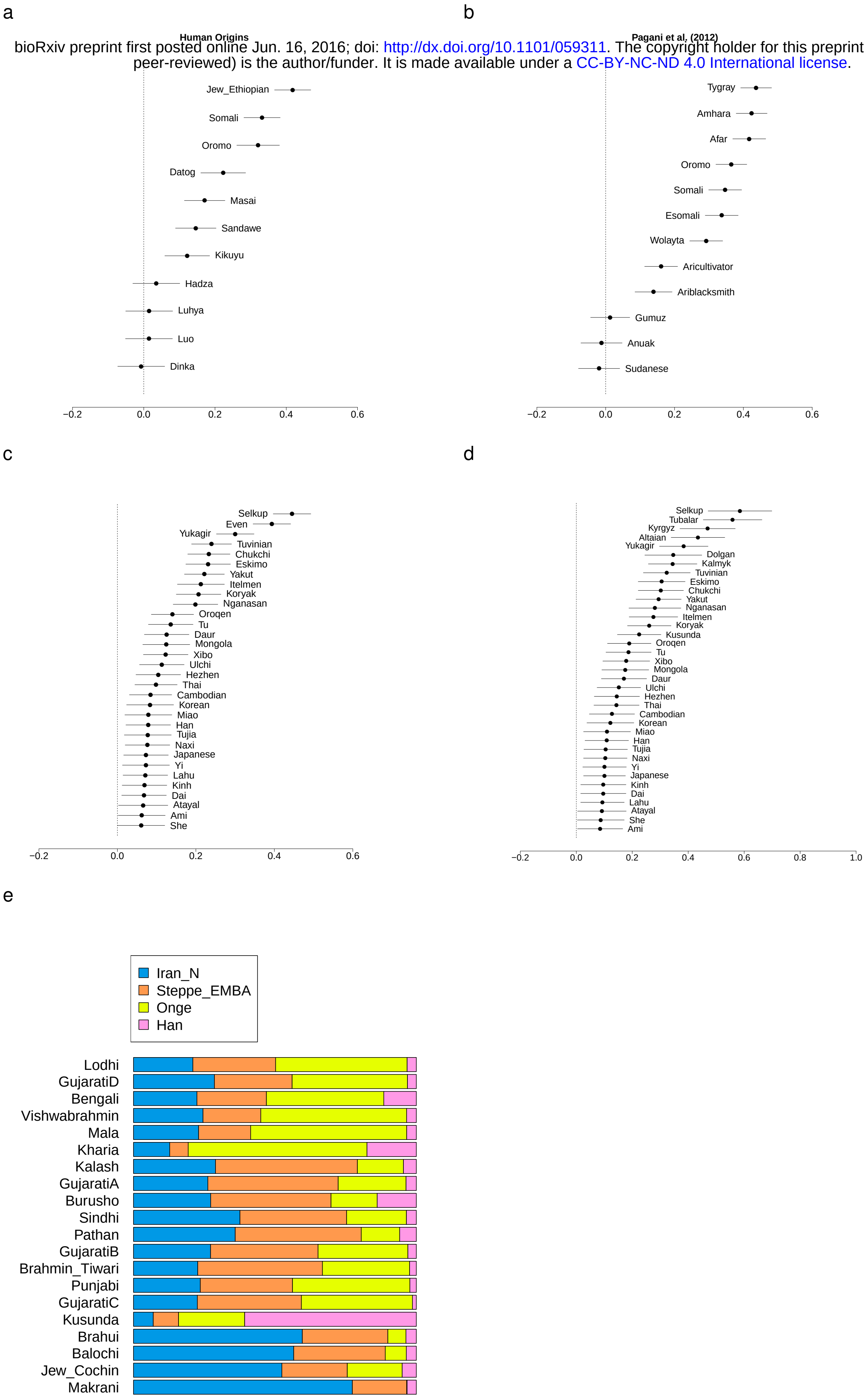




Extended Data Figure 2

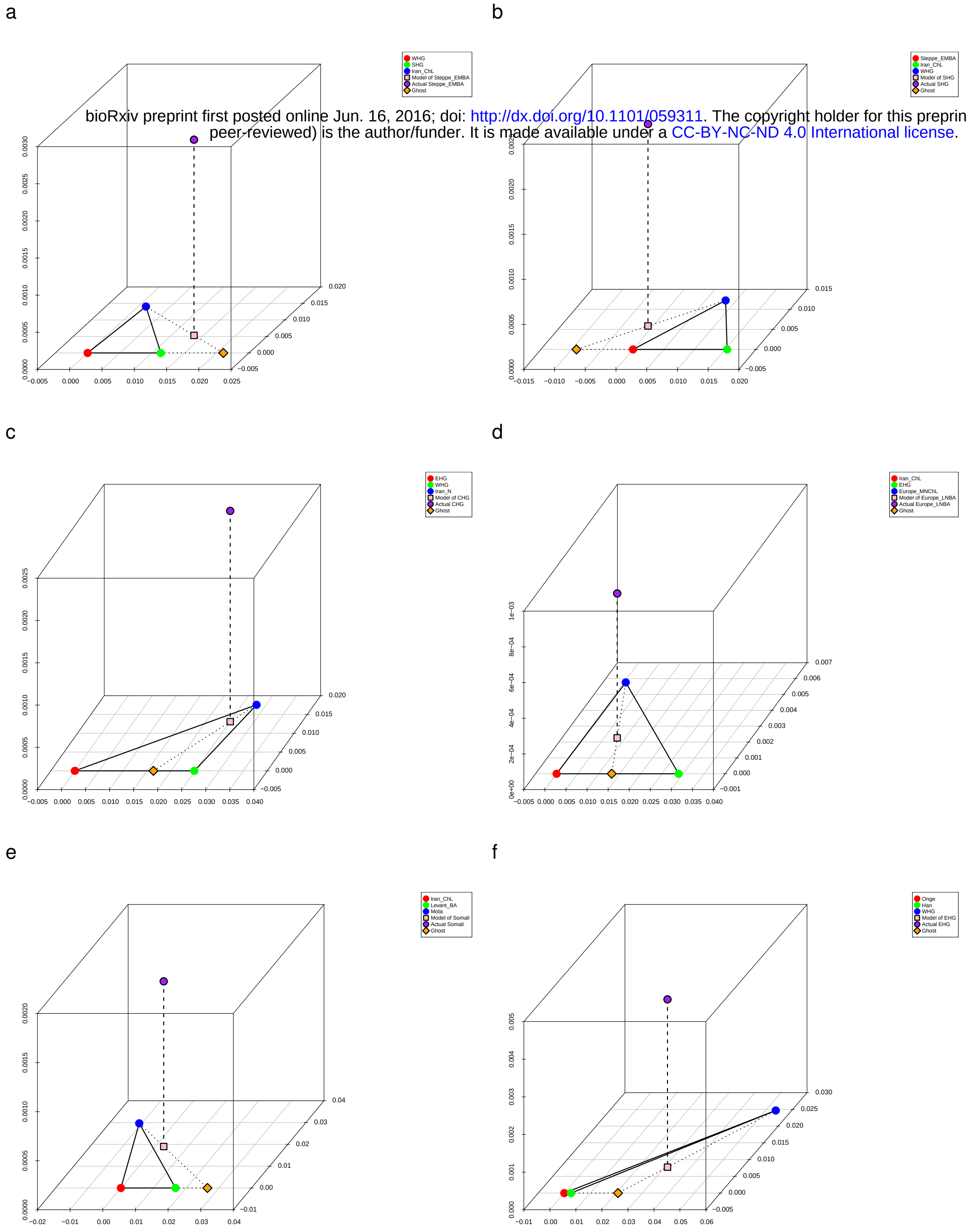


Extended Data Figure 3



Extended Data Figure 4

Extended Data Figure 5



Extended Data Figure 6

Extended Data Figure 7