

 Open access • Posted Content • DOI:10.1101/059311

## The genetic structure of the world's first farmers — Source link

Iosif Lazaridis, Dani Nadel, Gary O. Rollefson, Deborah C. Merrett ...+48 more authors

**Institutions:** Harvard University, University of Haifa, Whitman College, Simon Fraser University ...+18 more institutions

**Published on:** 16 Jun 2016 - bioRxiv (Cold Spring Harbor Labs Journals)

**Topics:** Southern Levant and Bronze Age

Related papers:

- [The genetic structure of the world's first farmers](#)
- [The Genomic History Of Southeastern Europe](#)
- [Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia](#)
- [The genetic history of Northern Europe](#)
- [Early farmers from across Europe directly descended from Neolithic Aegeans](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/the-genetic-structure-of-the-worlds-first-farmers-4qjlvctd08>

# 1 The genetic structure of the world's first farmers

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

Iosif Lazaridis<sup>1,2,†</sup>, Dani Nadel<sup>3</sup>, Gary Rollefson<sup>4</sup>, Deborah C. Merrett<sup>5</sup>, Nadin Rohland<sup>1</sup>,  
Swapan Mallick<sup>1,2,6</sup>, Daniel Fernandes<sup>7,8</sup>, Mario Novak<sup>7,9</sup>, Beatriz Gamarra<sup>7</sup>, Kendra Sirak<sup>7,10</sup>,  
Sarah Connell<sup>7</sup>, Kristin Stewardson<sup>1,6</sup>, Eadaoin Harney<sup>1,6,11</sup>, Qiaomei Fu<sup>1,12,13</sup>, Gloria  
Gonzalez-Fortes<sup>14</sup>, Songül Alpaslan Roodenberg<sup>15</sup>, György Lengyel<sup>16</sup>, Fanny Bocquentin<sup>17</sup>,  
Boris Gasparian<sup>18</sup>, Janet M. Monge<sup>19</sup>, Michael Gregg<sup>19</sup>, Vered Eshed<sup>20</sup>, Ahuva-Sivan  
Mizrahi<sup>20</sup>, Christopher Meiklejohn<sup>21</sup>, Fokke Gerritsen<sup>22</sup>, Luminita Bejenaru<sup>23</sup>, Matthias  
Blüher<sup>24</sup>, Archie Campbell<sup>25</sup>, Gianpiero Cavalleri<sup>26</sup>, David Comas<sup>27</sup>, Philippe Froguel<sup>28,29</sup>,  
Edmund Gilbert<sup>26</sup>, Shona M. Kerr<sup>25</sup>, Peter Kovacs<sup>30</sup>, Johannes Krause<sup>31,32,33</sup>, Darren  
McGettigan<sup>34</sup>, Michael Merrigan<sup>35</sup>, D. Andrew Merriwether<sup>36</sup>, Seamus O'Reilly<sup>35</sup>, Martin B.  
Richards<sup>37</sup>, Ornella Semino<sup>38</sup>, Michel Shamoon-Pour<sup>36</sup>, Gheorghe Stefanescu<sup>39</sup>, Michael  
Stumvoll<sup>24</sup>, Anke Tönjes<sup>24</sup>, Antonio Torroni<sup>38</sup>, James F. Wilson<sup>40,41</sup>, Loïc Yengo<sup>28,29</sup>, Nelli A.  
Hovhannisyian<sup>42</sup>, Nick Patterson<sup>2</sup>, Ron Pinhasi<sup>7,\*†</sup> and David Reich<sup>1,2,6,\*†</sup>

\*Co-senior authors;

†Correspondence and requests for materials should be addressed to:

I. L. ([lazaridis@genetics.med.harvard.edu](mailto:lazaridis@genetics.med.harvard.edu)), R. P. ([ron.pinhasi@ucd.ie](mailto:ron.pinhasi@ucd.ie)), or D. R.  
([reich@genetics.med.harvard.edu](mailto:reich@genetics.med.harvard.edu))

<sup>1</sup> Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

<sup>2</sup> Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA

<sup>3</sup> The Zinman Institute of Archaeology, University of Haifa, Haifa 3498838, Israel

<sup>4</sup> Dept. of Anthropology, Whitman College, Walla Walla, WA 99362, USA

<sup>5</sup> Dept. of Archaeology, Simon Fraser University, Burnaby, British Columbia V5A 1S6,  
Canada

<sup>6</sup> Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA

<sup>7</sup> School of Archaeology and Earth Institute, Belfield, University College Dublin, Dublin 4,  
Ireland

<sup>8</sup> CIAS, Department of Life Sciences, University of Coimbra, Coimbra 3000-456, Portugal

<sup>9</sup> Institute for Anthropological Research, Zagreb 10000, Croatia

<sup>10</sup> Dept. of Anthropology, Emory University, Atlanta, Georgia 30322, USA

<sup>11</sup> Dept. of Organismic and Evolutionary Biology, Harvard University, Cambridge 02138,  
USA

<sup>12</sup> Dept. of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology,  
Leipzig 04103, Germany

<sup>13</sup> Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of  
Sciences, IVPP, CAS, Beijing 100044, China

<sup>14</sup> Dept. of Biology and Evolution, University of Ferrara, Ferrara I-44121, Italy

<sup>15</sup> Independent Researcher, Santpoort-Noord, The Netherlands

<sup>16</sup> Department of Prehistory and Archaeology, University of Miskolc, 3515 Miskolc-  
Egyetemváros, Hungary

<sup>17</sup> French National Centre for Scientific Research, UMR 7041, 92023 Nanterre Cedex, France

<sup>18</sup> Institute of Archaeology and Ethnology, National Academy of Sciences of the Republic of  
Armenia, 0025 Yerevan, Republic of Armenia

<sup>19</sup> University of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, PA  
19104, USA

<sup>20</sup> Israel Antiquities Authority, Jerusalem 91710, Israel

<sup>21</sup> Dept. of Anthropology, University of Winnipeg, Winnipeg, Manitoba R3B 2E9, Canada

- 50 <sup>22</sup> Netherlands Institute in Turkey, Istiklal Caddesi, Nur-i Ziya Sokak 5, Beyoğlu, Istanbul,  
51 Turkey
- 52 <sup>23</sup> Alexandru Ioan Cuza University of Iasi, Romania 700505, Romania
- 53 <sup>24</sup> Department of Internal Medicine and Dermatology, Clinic of Endocrinology and  
54 Nephrology, 04103 Leipzig, Germany
- 55 <sup>25</sup> Generation Scotland, Centre for Genomic and Experimental Medicine, MRC Institute of  
56 Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital,  
57 Edinburgh EH4 2XU, Scotland
- 58 <sup>26</sup> RCSI Molecular & Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2,  
59 Ireland
- 60 <sup>27</sup> Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experimentals i de la  
61 Salut, Universitat Pompeu Fabra, 08003 Barcelona, Spain
- 62 <sup>28</sup> Univ. Lille, CNRS, Institut Pasteur de Lille, UMR 8199 - EGID, F-59000 Lille, France
- 63 <sup>29</sup> Department of Genomics of common disease, London Hammersmith Hospital, London  
64 W12 0HS, UK
- 65 <sup>30</sup> Leipzig University Medical Center, IFB AdiposityDiseases, 04103 Leipzig, Germany
- 66 <sup>31</sup> Max Planck Institute for the Science of Human History, 07745 Jena, Germany
- 67 <sup>32</sup> Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of  
68 Tübingen, Tübingen, Germany
- 69 <sup>33</sup> Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen,  
70 72072 Tübingen, Germany
- 71 <sup>34</sup> Independent research, County Wicklow, Ireland
- 72 <sup>35</sup> Genealogical Society of Ireland, Dún Laoghaire, County Dublin, Ireland
- 73 <sup>36</sup> Department of Anthropology, Binghamton University, New York 13902, USA
- 74 <sup>37</sup> Department of Biological Sciences, School of Applied Sciences, University of  
75 Huddersfield, Queensgate, Huddersfield HD1 3DH, UK
- 76 <sup>38</sup> Dipartimento di Biologia e Biotecnologie "L. Spallanzani", Università di Pavia, 27100  
77 Pavia, Italy
- 78 <sup>39</sup> Institutul de Cercetari Biologice, Iași 700505, Romania
- 79 <sup>40</sup> Usher Institute for Population Health Sciences and Informatics, University of Edinburgh,  
80 Edinburgh EH8 9AG, Scotland
- 81 <sup>41</sup> MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine,  
82 University of Edinburgh, Edinburgh EH4 2XU, Scotland
- 83 <sup>42</sup> Dept. of Ecology and Nature Protection, Yerevan State University, 0025 Yerevan, Republic  
84 of Armenia

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

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

109 In order to overcome the obstacle of poor DNA preservation, we took advantage of two  
110 methodological developments. First, we sampled from the inner ear region of the petrous  
111 bone<sup>3,4</sup> that can yield up to ~100 times more endogenous DNA than other skeletal elements<sup>4</sup>.  
112 Second, we used in-solution hybridization<sup>5</sup> to enrich extracted DNA for about 1.2 million  
113 single nucleotide polymorphism (SNP) targets<sup>6,7</sup>, making efficient sequencing practical by  
114 filtering out microbial and non-informative human DNA. We merged all sequences extracted  
115 from each individual, and randomly sampled a single sequence to represent each SNP,  
116 restricting to individuals with at least 9,000 SNPs covered at least once. We obtained  
117 genome-wide data passing quality control for 45 individuals on whom we had a median

118 coverage of 172,819 SNPs (Methods). We assembled radiocarbon dates for 26 individuals  
119 (22 new generated for this study) (Supplementary Data Table 1).

120 The newly reported ancient individuals date to ~12,000-1,400 BCE and come from the  
121 southern Caucasus (Armenia), northwestern Anatolia (Turkey), Iran, and the southern Levant  
122 (Israel and Jordan) (Supplementary Data Table 1, Fig. 1a). (One individual had a radiocarbon  
123 date that was not in agreement with the date of its archaeological context and was also a  
124 genetic outlier.) The samples include Epipaleolithic Natufian hunter-gatherers from Raqefet  
125 Cave in the Levant (12,000-9,800 BCE); a likely Mesolithic individual from Hotu Cave in the  
126 Alborz mountains of Iran (probable date of 9,100-8,600 BCE); Pre-Pottery Neolithic farmers  
127 from 'Ain Ghazal and Motza in the southern Levant (8,300-6,700 BCE); and early farmers  
128 from Ganj Dareh in the Zagros mountains of western Iran (8,200-7,600 BCE). The samples  
129 also include later Neolithic, Chalcolithic (~4,800-3,700 BCE), and Bronze Age (~3,350-  
130 1,400 BCE) individuals (Supplementary Information, section 1). We combined our data with  
131 previously published ancient data<sup>7,8,9,10,8,10-15</sup> to form a dataset of 281 ancient individuals. We  
132 then further merged with 2,583 present-day people genotyped on the Affymetrix Human  
133 Origins array<sup>13,16</sup> (238 new) (Supplementary Data Table 2; Supplementary Information,  
134 section 2). We grouped the ancient individuals based on archaeological culture and  
135 chronology (Fig. 1a; Supplementary Data Table 1). We refined the grouping based on  
136 patterns evident in Principal Components Analysis (PCA)<sup>17</sup> (Fig. 1b; Extended Data Fig. 1),  
137 ADMIXTURE model-based clustering<sup>18</sup> (Fig. 1c), and 'outgroup'  $f_3$ -analysis (Extended Data  
138 Fig. 2). We used  $f_4$ -statistics to identify outlier individuals and to cluster phylogenetically  
139 indistinguishable groups into 'Analysis Labels' (Supplementary Information, section 3).

140 We analyzed these data to address six questions. (1) Previous work has shown that the first  
141 European farmers harboured ancestry from a Basal Eurasian lineage that diverged from the  
142 ancestors of north Eurasian hunter-gatherers and East Asians before they separated from each  
143 other<sup>13</sup> What was the distribution of Basal Eurasian ancestry in the ancient Near East? (2)  
144 Were the first farmers of the Near East part of a single homogeneous population, or were they  
145 regionally differentiated? (3) Was there continuity between late pre-agricultural hunter-  
146 gatherers and early farming populations, or were the hunter-gatherers largely displaced by a  
147 single expansive population as in early Neolithic Europe?<sup>8</sup> (4) What is the genetic  
148 contribution of these early Near Eastern farmers to later populations of the Near East? (5)  
149 What is the genetic contribution of the early Near Eastern farmers to later populations of

150 mainland Europe, the Eurasian steppe, and to populations outside West Eurasia? (6) Do our  
151 data provide broader insights about population transformations in West Eurasia?

152 **Basal Eurasian ancestry was pervasive in the ancient Near East and associated with**  
153 **reduced Neanderthal ancestry**

154 The ‘Basal Eurasians’ are a lineage hypothesized<sup>13</sup> to have split off prior to the differentiation  
155 of all other Eurasian lineages, including both eastern non-African populations like the Han  
156 Chinese, and even the early diverged lineage represented by the genome sequence of the  
157 ~45,000 year old Upper Paleolithic Siberian from Ust’-Ishim<sup>11</sup>. To test for Basal Eurasian  
158 ancestry, we computed the statistic  $f_d(\text{Test}, \text{Han}; \text{Ust’-Ishim}, \text{Chimp})$  (Supplementary  
159 Information, section 4), which measures the excess of allele sharing of Ust’-Ishim with a  
160 variety of *Test* populations compared to Han as a baseline. This statistic is significantly  
161 negative ( $Z < -3.7$ ) for all ancient Near Easterners as well as Neolithic and later Europeans,  
162 consistent with their having ancestry from a deeply divergent Eurasian lineage that separated  
163 from the ancestors of most Eurasians prior to the separation of Han and Ust’-Ishim. We used  
164  $qpAdm^7$  to estimate Basal Eurasian ancestry in each *Test* population. We obtain the highest  
165 estimates in the earliest populations from both Iran ( $66 \pm 13\%$  in the likely Mesolithic sample,  
166  $48 \pm 6\%$  in Neolithic samples), and the Levant ( $44 \pm 8\%$  in Epipaleolithic Natufians) (Fig. 2),  
167 showing that Basal Eurasian ancestry was widespread across the ancient Near East.

168 West Eurasians harbour significantly less Neanderthal ancestry than East Asians<sup>19,20-23</sup>, which  
169 could be explained if West Eurasians (but not East Asians) have partial ancestry from a  
170 source diluting their Neanderthal inheritance<sup>21</sup>. Supporting this theory, we observe a negative  
171 correlation between Basal Eurasian ancestry and the rate of shared alleles with Neanderthals<sup>19</sup>  
172 (Supplementary Information, section 5; Fig. 2). By extrapolation, we infer that the Basal  
173 Eurasian population had lower Neanderthal ancestry than non-Basal Eurasian populations and  
174 possibly none (ninety-five percent confidence interval truncated at zero of 0-60%; Fig. 2;  
175 Methods). The finding of little if any Neanderthal ancestry in Basal Eurasians could be  
176 explained if the Neanderthal admixture into modern humans 50,000-60,000 years ago<sup>11</sup>  
177 largely occurred after the splitting of the Basal Eurasians from other non-Africans.

178 It is striking that the highest estimates of Basal Eurasian ancestry are from the Near East,  
179 given the hypothesis that it was there that most admixture between Neanderthals and modern  
180 humans occurred<sup>19,24</sup>. This could be explained if Basal Eurasians thoroughly admixed into the  
181 Near East before the time of the samples we analyzed but after the Neanderthal admixture.

182 Alternatively, the ancestors of Basal Eurasians may have always lived in the Near East, but  
183 the lineage of which they were a part did not participate in the Neanderthal admixture.

184 A population without Neanderthal admixture, basal to other Eurasians, may have plausibly  
185 lived in Africa. Craniometric analyses have suggested that the Natufians may have migrated  
186 from north or sub-Saharan Africa<sup>25,26</sup>, a result that finds some support from Y chromosome  
187 analysis which shows that the Natufians and successor Levantine Neolithic populations  
188 carried haplogroup E, of likely ultimate African origin, which has not been detected in other  
189 ancient males from West Eurasia (Supplementary Information, section 6)<sup>7,8</sup>. However, no  
190 affinity of Natufians to sub-Saharan Africans is evident in our genome-wide analysis, as  
191 present-day sub-Saharan Africans do not share more alleles with Natufians than with other  
192 ancient Eurasians (Extended Data Table 1). (We could not test for a link to present-day North  
193 Africans, who owe most of their ancestry to back-migration from Eurasia<sup>27,28</sup>.) The idea of  
194 Natufians as a vector for the movement of Basal Eurasian ancestry into the Near East is also  
195 not supported by our data, as the Basal Eurasian ancestry in the Natufians (44±8%) is  
196 consistent with stemming from the same population as that in the Neolithic and Mesolithic  
197 populations of Iran, and is not greater than in those populations (Supplementary Information,  
198 section 4). Further insight into the origins and legacy of the Natufians could come from  
199 comparison to Natufians from additional sites, and to ancient DNA from north Africa.

## 200 **Extreme regional differentiation in the ancient Near East**

201 PCA on present-day West Eurasian populations (Methods) (Extended Data Fig. 1) on which  
202 we projected the ancient individuals (Fig. 1b) replicates previous findings of a Europe-Near  
203 East contrast along the horizontal Principal Component 1 (PC1) and parallel clines (PC2) in  
204 both Europe and the Near East (Extended Data Fig. 1)<sup>7,8,13</sup>. Ancient samples from the Levant  
205 project at one end of the Near Eastern cline, and ancient samples from Iran at the other. The  
206 two Caucasus Hunter Gatherers (CHG)<sup>9</sup> are less extreme along PC1 than the Mesolithic and  
207 Neolithic individuals from Iran, while individuals from Chalcolithic Anatolia, Iran, and  
208 Armenia, and Bronze Age Armenia occupy intermediate positions. Qualitatively, the PCA  
209 has the appearance of a quadrangle whose four corners are some of the oldest samples:  
210 bottom-left: Western Hunter Gatherers (WHG), top-left: Eastern Hunter Gatherers (EHG),  
211 bottom-right: Neolithic Levant and Natufians, top-right: Neolithic Iran. This suggests the  
212 hypothesis that diverse ancient West Eurasians can be modelled as mixtures of as few as four



213 streams of ancestry related to these populations, which we confirmed using *qpWave*<sup>7</sup>  
214 (Supplementary Information, section 7).

215 We computed squared allele frequency differentiation between all pairs of ancient West  
216 Eurasians<sup>29</sup> (Methods; Fig. 3; Extended Data Fig. 3), and found that the populations at the  
217 four corners of the quadrangle had differentiation of  $F_{ST}=0.08-0.15$ , comparable to the value  
218 of  $0.09-0.13$  seen between present-day West Eurasians and East Asians (Han)  
219 (Supplementary Data Table 3). In contrast, by the Bronze Age, genetic differentiation  
220 between pairs of West Eurasian populations had reached its present-day low levels (Fig. 3):  
221 today,  $F_{ST}$  is  $\leq 0.025$  for 95% of the pairs of West Eurasian populations and  $\leq 0.046$  for all  
222 pairs. These results point to a demographic process that established high differentiation  
223 across West Eurasia and then reduced this differentiation over time.

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

225 Our data document continuity across the hunter-gatherer / farming transition, separately in  
226 the southern Levant and in the southern Caucasus-Iran highlands. The qualitative evidence  
227 for this is that PCA, ADMIXTURE, and outgroup  $f_3$  analysis cluster Levantine hunter-  
228 gatherers (Natufians) with Levantine farmers, and Iranian and Caucasus Hunter Gatherers  
229 with Iranian farmers (Fig. 1b; Extended Data Fig. 1; Extended Data Fig. 2). We confirm this  
230 in the Levant by showing that its early farmers share significantly more alleles with Natufians  
231 than with the early farmers of Iran: the statistic  $f_4(\text{Levant\_N}, \text{Chimp}; \text{Natufian}, \text{Iran\_N})$  is  
232 significantly positive ( $Z=13.6$ ). The early farmers of the Caucasus-Iran highlands similarly  
233 share significantly more alleles with the hunter-gatherers of this region than with the early  
234 farmers from the Levant: the statistic  $f_4(\text{Iran\_N}, \text{Chimp}; \text{Caucasus or Iran highland hunter-}$   
235  $\text{gatherers}, \text{Levant\_N})$  is significantly positive ( $Z>6$ ).

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

237 Almost all ancient and present-day West Eurasians have evidence of significant admixture  
238 between two or more ancestral populations, as documented by statistics of the form  $f_3(\text{Test};$   
239  $\text{Reference}_1, \text{Reference}_2)$  which if negative, show that a *Test* population's allele frequencies  
240 tend to be intermediate between two *Reference* populations<sup>16</sup> (Extended Data Table 2). To  
241 better understand the admixture history beyond these patterns, we used *qpAdm*<sup>7</sup>, which can  
242 evaluate whether a particular *Test* population is consistent with being derived from a set of  
243 proposed source populations, and if so, infer mixture proportions (Methods). We used this



244 approach to carry out a systematic survey of ancient West Eurasian populations to explore  
245 their possible sources of admixture (Fig. 4; Supplementary Information, section 7).

246 Among first farmers, those of the Levant trace  $\sim 2/3$  of their ancestry to people related to  
247 Natufian hunter-gatherers and  $\sim 1/3$  to people related to Anatolian farmers (Supplementary  
248 Information, section 7). Western Iranian first farmers cluster with the likely Mesolithic  
249 HotuIIIb individual and more remotely with hunter-gatherers from the southern Caucasus  
250 (Fig. 1b), and share alleles at an equal rate with Anatolian and Levantine early farmers  
251 (Supplementary Information, section 7), highlighting the long-term isolation of western Iran.

252 During subsequent millennia, the early farmer populations of the Near East expanded in all  
253 directions and mixed, as we can only model populations of the Chalcolithic and subsequent  
254 Bronze Age as having ancestry from two or more sources. The Chalcolithic people of western  
255 Iran can be modelled as a mixture of the Neolithic people of western Iran, the Levant, and  
256 Caucasus Hunter Gatherers (CHG), consistent with their position in the PCA (Fig. 1b).  
257 Admixture from populations related to the Chalcolithic people of western Iran had a wide  
258 impact, consistent with contributing  $\sim 44\%$  of the ancestry of Levantine Bronze Age  
259 populations in the south and  $\sim 33\%$  of the ancestry of the Chalcolithic northwest Anatolians in  
260 the west. Our analysis show that the ancient populations of the Chalcolithic Iran, Chalcolithic  
261 Armenia, Bronze Age Armenia and Chalcolithic Anatolia were all composed of the same  
262 ancestral components, albeit in slightly different proportions (Fig. 4b; Supplementary  
263 Information, section 7).

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

265 Admixture did not only occur within the Near East but extended towards Europe. To the  
266 north, a population related to people of the Iran Chalcolithic contributed  $\sim 43\%$  of the  
267 ancestry of early Bronze Age populations of the steppe. The spread of Near Eastern ancestry  
268 into the Eurasian steppe was previously inferred<sup>7</sup> without access to ancient samples, by  
269 hypothesizing a population related to present-day Armenians as a source<sup>7,8</sup>. To the west, the  
270 early farmers of mainland Europe were descended from a population related to Neolithic  
271 northwestern Anatolians<sup>8</sup>. This is consistent with an Anatolian origin of farming in Europe,  
272 but does not reject other sources, since the spatial distribution of the Anatolian/European-like  
273 farmer populations is unknown. We can rule out the hypothesis that European farmers stem  
274 directly from a population related to the ancient farmers of the southern Levant<sup>30,31</sup>, however,

275 since they share more allele with Anatolian Neolithic farmers than with Levantine farmers as  
276 attested by the positive statistic  $f_4(\text{Europe\_EN}, \text{Chimp}; \text{Anatolia\_N}, \text{Levant\_N})$  ( $Z=15$ ).

277 Migrations from the Near East also occurred towards the southwest into East African  
278 populations which experienced West Eurasian admixture  $\sim 1,000$  BCE<sup>32,33</sup>. Previously, the  
279 West Eurasian population known to be the best proxy for this ancestry was present-day  
280 Sardinians<sup>33</sup>, who resemble Neolithic Europeans genetically<sup>13,34</sup>. However, our analysis  
281 shows that East African ancestry is significantly better modelled by Levantine early farmers  
282 than by Anatolian or early European farmers, implying that the spread of this ancestry to East  
283 Africa was not from the same group that spread Near Eastern ancestry into Europe (Extended  
284 Data Fig. 4; Supplementary Information, section 8).

285 In South Asia, our dataset provides insight into the sources of Ancestral North Indians (ANI),  
286 a West Eurasian related population that no longer exists in unmixed form but contributes a  
287 variable amount of the ancestry of South Asians<sup>35,36</sup> (Supplementary Information, section 9)  
288 (Extended Data Fig. 4). We show that it is impossible to model the ANI as being derived  
289 from any single ancient population in our dataset. However, it can be modelled as a mix of  
290 ancestry related to both early farmers of western Iran and to people of the Bronze Age  
291 Eurasian steppe; all sampled South Asian groups are inferred to have significant amounts of  
292 both ancestral types. The demographic impact of steppe related populations on South Asia  
293 was substantial, as the Mala, a south Indian population with minimal ANI along the ‘Indian  
294 Cline’ of such ancestry<sup>35,36</sup> is inferred to have  $\sim 18\%$  steppe-related ancestry, while the Kalash  
295 of Pakistan are inferred to have  $\sim 50\%$ , similar to present-day northern Europeans<sup>7</sup>.

## 296 **Broader insights into population transformations across West Eurasia and beyond**

297 We were concerned that our conclusions might be biased by the particular populations we  
298 happened to sample, and that we would have obtained qualitatively different conclusions  
299 without data from some key populations. We tested our conclusions by plotting the inferred  
300 position of admixed populations in PCA against a weighted combination of their inferred  
301 source populations and obtained qualitatively consistent results (Extended Data Fig. 5).

302 To further assess the robustness of our inferences, we developed a method to infer the  
303 existence and genetic affinities of ancient populations from unobserved ‘ghost’ populations  
304 (Supplementary Information, section 10; Extended Data Fig. 6). This method takes advantage  
305 of the insight that if an unsampled ghost population admixes with differentiated ‘substratum’

306 populations, it is possible to extrapolate its identity by intersecting clines of populations with  
307 variable proportions of ‘ghost’ and ‘substratum’ ancestry. Applying this while withholding  
308 major populations, we validated some of our key inferences, successfully inferring mixture  
309 proportions consistent with those obtained when the populations are included in the analysis.  
310 Application of this methods highlights the impact of Ancient North Eurasian (ANE) ancestry  
311 related to the ~22,000 BCE Mal’ta 1 and ~15,000 BCE Afontova Gora 2<sup>15</sup> on populations  
312 living in Europe, the Americas, and Eastern Eurasia. Eastern Eurasians can be modelled as  
313 arrayed along a cline with different proportions of ANE ancestry (Supplementary  
314 Information, section 11; Extended Data Fig. 7), ranging from ~40% ANE in Native  
315 Americans matching previous findings<sup>13,15</sup>, to no less than ~5-10% ANE in diverse East  
316 Asian groups including Han Chinese (Extended Data Fig. 4; Extended Data Fig. 6f). We also  
317 document a cline of ANE ancestry across the east-west extent of Eurasia. Eastern Hunter  
318 Gatherers (EHG) derive ~3/4 of their ancestry from the ANE (Supplementary Information,  
319 section 11); Scandinavian hunter-gatherers<sup>7,8,13</sup> (SHG) are a mix of EHG and WHG; and  
320 WHG are a mix of EHG and the Upper Paleolithic Bichon from Switzerland (Supplementary  
321 Information, section 7). Northwest Anatolians—with ancestry from a population related to  
322 European hunter-gatherers (Supplementary Information, section 7)—are better modelled if  
323 this ancestry is taken as more extreme than Bichon (Supplementary Information, section 10).

324 The population structure of the ancient Near East was not independent of that of Europe  
325 (Supplementary Information, section 4), as evidenced by the highly significant ( $Z=-8.9$ )  
326 statistic  $f_4(\text{Iran\_N}, \text{Natufian}; \text{WHG}, \text{EHG})$  which suggests gene flow in ‘northeastern’  
327 (Neolithic Iran/EHG) and ‘southwestern’ (Levant/WHG) interaction spheres (Fig. 4d). This  
328 interdependence of the ancestry of Europe and the Near East may have been mediated by  
329 unsampled geographically intermediate populations<sup>37</sup> that contribute ancestry to both regions.

### 330 **Conclusions**

331 By analysing genome-wide ancient DNA data from ancient individuals from the Levant,  
332 Anatolia, the southern Caucasus and Iran, we have provided a first glimpse of the  
333 demographic structure of the human populations that transitioned to farming. We reject the  
334 hypothesis that the spread of agriculture in the Near East was achieved by the dispersal of a  
335 single farming population displacing the hunter-gatherers they encountered. Instead, the  
336 spread of ideas and farming technology moved faster than the spread of people, as we can  
337 determine from the fact that the population structure of the Near East was maintained

338 throughout the transition to agriculture. A priority for future ancient DNA studies should be  
339 to obtain data from older periods, which would reveal the deeper origins of the population  
340 structure in the Near East. It will also be important to obtain data from the ancient  
341 civilizations of the Near East to bridge the gap between the region's prehistoric inhabitants  
342 and those of the present.

## 343 **Acknowledgements**

344 We thank the 238 human subjects who voluntarily donated the samples whose genome-wide  
345 data we newly report in this study. We thank D. Labuda for sharing the collection of DNA  
346 samples from Poland, and P. Zalloua for sharing the collection of DNA samples from  
347 Lebanon. We thank O. Bar-Yosef, M. Bonogofsky, I. Hershkowitz, M. Lipson, I. Mathieson,  
348 H. May, R. Meadow, I. Olalde, S. Paabo, P. Skoglund, and N. Nakatsuka for comments and  
349 critiques, and M. Ferry and M. Michel for their work on the in-solution enrichment  
350 experiments. S.C. was supported by the Irish Research Council for Humanities and Social  
351 Sciences (IRCHSS) ERC Support Programme. Q.F. was funded by the Bureau of  
352 International Cooperation of Chinese Academy of Sciences, the National Natural Science  
353 Foundation of China (L1524016) and the Chinese Academy of Sciences Discipline  
354 Development Strategy Project (2015-DX-C-03). The Scottish diversity data from Generation  
355 Scotland received funding from the Chief Scientist Office of the Scottish Government Health  
356 Directorates [CZD/16/6] and the Scottish Funding Council [HR03006], while the Scottish  
357 Donor DNA Databank (GS:3D) was funded by a project grant from the Scottish Executive  
358 Health Department, Chief Scientist Office [CZB/4/285]. M.S., A.Tön., M.B. and P.K. were  
359 supported by grants from the Collaborative Research Center funded by the German Research  
360 Foundation (CRC 1052; B01, B03, C01). M.S.-P. was partially funded by a Wenner-Gren  
361 Foundation Dissertation Fieldwork Grant (#9005), and by the National Science Foundation  
362 DDRIG (BCS-1455744). P.K. was supported by the Federal Ministry of Education and  
363 Research (BMBF), Germany (FKZ: 01EO1501). J.F.W acknowledges the MRC “QTL in  
364 Health and Disease” programme grant. The Romanian contribution to this work was  
365 supported by the EC Commission, Directorate General XII, within the framework of the  
366 Cooperation in Science and Technology with Central and Eastern European Countries  
367 (Supplementary Agreement ERBCIPDCT 940038 to the Contract ERBCHRXCT 920032,  
368 coordinated by Prof. A. Piazza, Turin, Italy). M.R. received support from the Leverhulme  
369 Trust’s Doctoral Scholarship programme. O.S. and A.Tor. were supported by the University  
370 of Pavia strategic theme "Towards a governance model for international migration: an  
371 interdisciplinary and diachronic perspective" (MIGRAT-IN-G) and the Italian Ministry of  
372 Education, University and Research: Progetti Ricerca Interesse Nazionale 2012. The Raqefet  
373 Cave Natufian project was supported by funds from the National Geographic Society (Grant  
374 #8915-11), the Wenner-Gren Foundation (Grant #7481) and the Irene Levi-Sala CARE  
375 Foundation, while radiocarbon dating on the samples was funded by the Israel Science

376 Foundation (Grant 475/10; E. Boaretto). R.P. was supported by ERC starting grant  
377 ADNABIOARC (263441). D.R. was supported by NIH grant GM100233, by NSF  
378 HOMINID BCS-1032255, and is a Howard Hughes Medical Institute investigator.

### 379 **Author Contributions**

380 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.,  
381 B.Gas., J.M.M., M.G., V.E., A.M., C.M., F.G., N.A.H. and R.P. assembled archaeological  
382 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.  
383 performed or supervised ancient DNA wet laboratory work. L.B, M.B., A.C., G.C., D.C.,  
384 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.,  
385 A.Tön., A.Tor., J.F.W., L.Y. and D.R. assembled present-day samples for genotyping. I.L,  
386 N.P. and D.R. developed methods for data analysis. I.L., S.M., Q.F., N.P. and D.R. analyzed  
387 data. I.L., R.P. and D.R. wrote the manuscript and supplements. All authors read the  
388 manuscript and provided comments.

### 389 **Author Information**

390 The aligned sequences are available through the European Nucleotide Archive under  
391 accession number xxx. Fully public subsets of the analysis datasets are at  
392 ([http://genetics.med.harvard.edu/reichlab/Reich\\_Lab/Datasets.html](http://genetics.med.harvard.edu/reichlab/Reich_Lab/Datasets.html)). The complete dataset  
393 (including present-day humans for which the informed consent is not consistent with public  
394 posting of data) is available to researchers who send a signed letter to D.R. indicating that  
395 they will abide by specified usage conditions (Supplementary Information, section 2).

## 396 **Online Methods**

### 397 **Ancient DNA data**

398 In a dedicated ancient DNA laboratory at University College Dublin, we prepared powder  
399 from 132 ancient Near Eastern samples, either by dissecting the inner ear region of the  
400 petrous bone using a sandblaster (Renfert), or by drilling using a Dremel tool and single-use  
401 drill bits and selecting the best preserved bone fragments based on anatomical criteria. These  
402 fragments were then powdered using a mixer mill (Retsch Mixer Mill 400)<sup>4</sup>.

403 We performed all subsequent processing steps in a dedicated ancient DNA laboratory at  
404 Harvard Medical School, where we extracted DNA from the powder (usually 75 mg, range  
405 14-81 mg) using an optimized ancient DNA extraction protocol<sup>38</sup>, but replaced the assembly  
406 of Qiagen MinElute columns and extension reservoirs from Zymo Research with a High Pure  
407 Extender Assembly from the High Pure Viral Nucleic Acid Large Volume Kit (Roche  
408 Applied Science). We built a total of 170 barcoded double-stranded Illumina sequencing  
409 libraries for these samples<sup>39</sup>, of which we treated 167 with Uracil-DNA glycosylase (UDG) to  
410 remove the characteristic C-to-T errors of ancient DNA<sup>40</sup>. The UDG treatment strategy is  
411 (by-design) inefficient at removing terminal uracils, allowing the mismatch rate to the human  
412 genome at the terminal nucleotide to be used for authentication<sup>39</sup>. We updated this library  
413 preparation protocol in two ways compared to the original publication: first, we used 16U  
414 Bst2.0 Polymerase, Large Fragment (NEB) and 1x Isothermal Amplification buffer (NEB) in  
415 a final volume of 25 $\mu$ L fill-in reaction, and second, we used the entire inactivated 25 $\mu$ L fill-in  
416 reaction in a total volume of 100 $\mu$ L PCR mix with 1  $\mu$ M of each primer<sup>41</sup>. We included  
417 extraction negative controls (where no sample powder was used) and library negative  
418 controls (where extract was supplemented by water) in every batch of samples processed and  
419 carried them through the entire wet lab processing to test for reagent contamination.

420 We screened the libraries by hybridizing them in solution to a set of oligonucleotide probes  
421 tiling the mitochondrial genome<sup>42</sup>, using the protocol described previously<sup>7</sup>. We sequenced  
422 the enriched libraries using an Illumina NextSeq 500 instrument using 2 $\times$ 76bp reads,  
423 trimmed identifying sequences (seven base pair molecular barcodes at either end) and any  
424 trailing adapters, merged read pairs that overlapped by at least 15 base pairs, and mapped the  
425 merged sequences to the RSRS mitochondrial DNA reference genome<sup>43</sup>, using the Burrows  
426 Wheeler Aligner<sup>44</sup> (*bwa*) and the command *samse* (v0.6.1).



427 We enriched promising libraries for a targeted set of ~1.2 million SNPs<sup>8</sup> as in ref. 5, and  
428 adjusted the blocking oligonucleotide and primers to be appropriate for our libraries. The  
429 specific probe sequences are given in Supplementary Data 2 of ref. 7  
430 ([http://www.nature.com/nature/journal/v522/n7555/abs/nature14317.html#supplementary-](http://www.nature.com/nature/journal/v522/n7555/abs/nature14317.html#supplementary-information)  
431 [information](http://www.nature.com/nature/journal/v522/n7555/abs/nature14317.html#supplementary-information)) and Supplementary Data 1 of ref. 6.  
432 ([http://www.nature.com/nature/journal/v524/n7564/full/nature14558.html#supplementary-](http://www.nature.com/nature/journal/v524/n7564/full/nature14558.html#supplementary-information)  
433 [information](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.  
434 We trimmed identifying sequences (molecular barcodes) and any trailing adapters, merged  
435 pairs that overlapped by at least 15 base pairs (allowing up to one mismatch), and mapped the  
436 merged sequences to *hg19* using the single-ended aligner *samse* in *bwa* (v0.6.1). We  
437 removed duplicated sequences by identifying sets of sequences with the same orientation and  
438 start and end positions after alignment to *hg19*; we picked the highest quality sequence to  
439 represent each set. For each sample, we represented each SNP position by a randomly chosen  
440 sequence, restricting to sequences with a minimum mapping quality (MAPQ≥10), sites with a  
441 minimum sequencing quality (≥20), and removing 2 bases at the ends of reads. We sequenced  
442 the enriched products up to the point that we estimated that generating a hundred new  
443 sequences was expected to add data on less than about one new SNP<sup>8</sup>.

#### 444 **Testing for contamination and quality control**

445 For each ancient DNA library, we evaluated authenticity in several ways. First, we estimated  
446 the rate of matching to the consensus sequence for mitochondrial genomes sequenced to a  
447 coverage of at least 10-fold from the initial screening data. Of the 76 libraries that contributed  
448 to our dataset (coming from 45 samples), 70 had an estimated rate of sequencing matching to  
449 the consensus of >95% according to *contamMix*<sup>5</sup> (the remaining libraries had estimated  
450 match rates of 75-92%, but gave no sign of being outliers in principal component analysis or  
451 X chromosome contamination analysis so we retained them for analysis) (Supplementary  
452 Data Table 1). We quantified the rate of C-to-T substitution in the final nucleotide of the  
453 sequences analyzed, relative to the human reference genome sequence, and found that all the  
454 libraries analyzed had rates of at least 3%<sup>39</sup>, consistent with genuine ancient DNA. For the  
455 nuclear data from males, we used the *ANGSD* software<sup>45</sup> to estimate a conservative X  
456 chromosome estimate of contamination. We determined that all libraries passing our quality  
457 control and for which we had sufficient X chromosome data to make an assessment had  
458 contamination rates of 0-1.5%. Finally, we merged data for samples for which we had  
459 multiple libraries to produce an analysis dataset.

## 460 **Affymetrix Human Origins genotyping data**

461 We genotyped 238 present-day individuals from 17 diverse West Eurasian populations on the  
462 Affymetrix Human Origins array<sup>16</sup>, and applied quality control analyses as previously  
463 described<sup>13</sup> (Supplementary Data Table 2). We merged the newly generated data with data  
464 from 2,345 individuals previously genotyped on the same array<sup>13</sup>. All individuals that were  
465 genotyped provided informed consent consistent with studies of population history, following  
466 protocols approved by the ethical review committees of the institutions of the researchers  
467 who collected the samples. De-identified aliquots of DNA from all individuals were sent to  
468 the core facility of the Center for Applied Genomics at the Children's Hospital of  
469 Philadelphia for genotyping and data processing. For 127 of the individuals with newly  
470 reported data, the informed consent was consistent with public distribution of data, and the  
471 data can be downloaded at [http://genetics.med.harvard.edu/reich/Reich\\_Lab/Datasets.html](http://genetics.med.harvard.edu/reich/Reich_Lab/Datasets.html).  
472 To access data for the remaining 111 samples, researchers should a signed letter to D.R.  
473 containing the following text: "(a) I will not distribute the samples marked "signed letter"  
474 outside my collaboration; (b) I will not post data from the samples marked "signed letter"  
475 publicly; (c) I will make no attempt to connect the genetic data for the samples marked  
476 "signed letter" to personal identifiers; (d) I will not use the data for samples marked "signed  
477 letter" for commercial purposes." Supplementary Data Table 2 specifies which samples are  
478 consistent with which type of data distribution.

## 479 **Datasets**

480 We carried out population genetic analysis on two datasets: (i) *HO* includes 2,583 present-  
481 day humans genotyped on the Human Origins array<sup>13,16</sup> including 238 newly reported  
482 (Supplementary Data Table 2; Supplementary Information, section 2), and 281 ancient  
483 individuals on a total of 592,146 autosomal SNPs. (ii) *HOIII* includes the 281 ancient  
484 individuals on a total of 1,055,186 autosomal SNPs, including those present in both the  
485 Human Origins and Illumina genotyping platforms, but excluding SNPs on the sex  
486 chromosomes or additional SNPs of the 1240k capture array that were included because of  
487 their potential functional importance<sup>8</sup>. We used *HO* for analyses that involve both ancient and  
488 present-day individuals, and *HOIII* for analysis on ancient individuals alone. We also use 235  
489 individuals from Pagani et al.<sup>32</sup> on 418,700 autosomal SNPs to study admixture in East  
490 Africans (Supplementary Information, section 8). Ancient individuals are represented in  
491 'pseudo-haploid' form by randomly choosing one allele for each position of the array.

## 492 **Principal Components Analysis**

493 We carried out principal components analysis in the *smartpca* program of EIGENSOFT<sup>17</sup>,  
494 using default parameters and the `lsqproject: YES`<sup>13</sup> and `numoutlieriter: 0` options. We carried  
495 out PCA of the *HO* dataset on 991 present-day West Eurasians (Extended Data Fig. 1), and  
496 projected the 278 ancient individuals (Fig. 1b).

## 497 **ADMIXTURE Analysis**

498 We carried out ADMIXTURE analysis<sup>18</sup> of the *HO* dataset after pruning for linkage  
499 disequilibrium in PLINK<sup>46,47</sup> with parameters `--indep-pairwise 200 25 0.4` which retained  
500 296,309 SNPs. We performed analysis in 20 replicates with different random seeds, and  
501 retained the highest likelihood replicate for each value of K. We show the K=11 results for  
502 the 281 ancient samples in Fig. 1c (this is the lowest K for which components maximized in  
503 European hunter-gatherers, ancient Levant, and ancient Iran appear).

## 504 ***f*-statistics**

505 We carried out analysis of  $f_3$ -statistics,  $f_4$ -ratio, and  $f_4$ -statistics statistics using the  
506 ADMIXTOOLS<sup>16</sup> programs *qp3Pop*, *qpF4ratio* with default parameters, and *qpDstat* with  
507 `f4mode: YES`, and computed standard errors with a block jackknife<sup>48</sup>. For computing  $f_3$ -  
508 statistics with an ancient population as a target, we set the `inbreed: YES` parameter. We  
509 computed  $f$ -statistics on the *HOIII* dataset when no present-day humans were involved and on  
510 the *HO* dataset when they were. We computed the statistic  $f_4(\textit{Test}, \textit{Mbuti}; \textit{Altai}, \textit{Denisovan})$   
511 in Fig. 2 on the *HOIII* dataset after merging with whole genome data on 3 Mbuti individuals  
512 from Panel C of the Simons Genome Diversity Project<sup>49</sup>. We computed the dendrogram of  
513 Extended Data Fig. 2 showing hierarchical clustering of populations with outgroup  $f_3$ -  
514 statistics using the open source *heatmap.2* function of the *gplots* package in *R*.

## 515 **Negative correlation of Basal Eurasian ancestry with Neanderthal ancestry**

516 We used the *lm* function of R to fit a linear regression of the rate of allele sharing of a *Test*  
517 population with the Altai Neanderthal as measured by  $f_4(\textit{Test}, \textit{Mbuti}; \textit{Altai}, \textit{Denisovan})$  as  
518 the dependent variable, and the proportion of Basal Eurasian ancestry (Supplementary  
519 Information, section 4) as the predictor variable,. Extrapolating from the fitted line, we obtain  
520 the value of the statistic expected if *Test* is a population of 0% or 100% Basal Eurasian  
521 ancestry. We then compute the ratio of the Neanderthal ancestry estimate in Basal Eurasians

522 relative to non-Basal Eurasians as  $f_4(100\% \text{ Basal Eurasian, Mbuti; Altai, Denisovan})/f_4(0\%$   
523  $\text{Basal Eurasian, Mbuti; Altai, Denisovan})$ . We use a block jackknife<sup>48</sup>, dropping one of 100  
524 contiguous blocks of the genome at a time, to estimate the value and standard error of this  
525 quantity ( $9\pm 26\%$ ). We compute a 95% confidence interval based on the point estimate  $\pm 1.96$ -  
526 times the standard error: -42 to 60%. We truncated to 0-60% on the assumption that Basal  
527 Eurasians had no less Neanderthal admixture than Mbuti from sub-Saharan Africa.

### 528 **Estimation of $F_{ST}$ coefficients**

529 We estimated  $F_{ST}$  in *smartpca*<sup>17</sup> with default parameters, inbreed: YES, and fstonly: YES.

### 530 **Admixture Graph modeling**

531 We carried out Admixture Graph modeling with the *qpGraph* software<sup>16</sup> using Mbuti as an  
532 outgroup unless otherwise specified.

### 533 **Testing for the number of streams of ancestry**

534 We used the *qpWave*<sup>35,50</sup> software, described in Supplementary Information, section 10 of  
535 ref.<sup>7</sup>, to test whether a set of ‘Left’ populations is consistent with being related via as few as  
536  $N$  streams of ancestry to a set of ‘Right’ populations by studying statistics of the form  $X(u, v)$   
537  $= F_4(u_0, u; v_0, v)$  where  $u_0, v_0$  are basis populations chosen from the ‘Left’ and ‘Right’ sets  
538 and  $u, v$  are other populations from these sets. We use a Hotelling’s  $T^2$  test<sup>50</sup> to evaluate  
539 whether the matrix of size  $(L-1)*(R-1)$ , where  $L, R$  are the sizes of the ‘Left’ and ‘Right’ sets  
540 has rank  $m$ . If this is the case, we can conclude that the ‘Left’ set is related via at least  $N=m+1$   
541 streams of ancestry related differently to the ‘Right’ set.

### 542 **Inferring mixture proportions without an explicit phylogeny**

543 We used the *qpAdm* methodology described in Supplementary Information, section 10 of ref.  
544 <sup>7</sup> to estimate the proportions of ancestry in a *Test* population deriving from a mixture of  $N$   
545 ‘reference’ populations by exploiting (but not explicitly modeling) shared genetic drift with a  
546 set of ‘Outgroup’ populations (Supplementary Information, section 7). We set the details:  
547 YES parameter, which reports a normally distributed Z-score estimated with a block  
548 jackknife for the difference between the statistics  $f_4(u_0, \textit{Test}; v_0, v)$  and  $f_4(u_0, \textit{Estimated Test};$   
549  $v_0, v)$  where *Estimated Test* is  $\sum_{i=1}^N \alpha_i f_4(u_0, \textit{Ref}_i; v_0, v)$ , the average of these  $f_4$ -statistics  
550 weighed by the mixture proportions  $\alpha_i$  from the  $N$  reference populations.

551 **Modeling admixture from ghost populations**

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

## 559 **Figures Legends**

560 **Figure 1: Genetic structure of ancient West Eurasia.** (a) Sampling locations and times in six West  
561 Eurasian regions. Sample sizes for each population are given below each bar. Abbreviations used: E:  
562 Early, M: Middle, L: Late, HG: Hunter-Gatherer, N: Neolithic, ChL: Chalcolithic, BA: Bronze Age,  
563 IA: Iron Age. (b) Principal components analysis of 991 present-day West Eurasians (grey points) with  
564 278 projected ancient samples (excluding the Upper Paleolithic Ust\_Ishim, Kostenki14, and MA1).  
565 To avoid visual clutter, population labels of present-day individuals are shown in Extended Data Fig.  
566 1. (c) ADMIXTURE model-based clustering analysis of 2,583 present-day humans and 281 ancient  
567 samples; we show the results only for ancient samples for  $K=11$  clusters.

568 **Figure 2: Basal Eurasian ancestry explains the reduced Neanderthal admixture in West**  
569 **Eurasians.** Basal Eurasian ancestry estimates are negatively correlated to a statistic measuring  
570 Neanderthal ancestry  $f_4(\text{Test}, \text{Mbuti}; \text{Altai}, \text{Denisovan})$ .

571 **Figure 3: Genetic differentiation and its dramatic decrease over time in West Eurasia.** (a)  
572 Pairwise  $F_{ST}$  between 19 Ancient West Eurasian populations (arranged in approximate chronological  
573 order), and select present-day populations. (b) Pairwise  $F_{ST}$  distribution among populations belonging  
574 to four successive time slices in West Eurasia; the median (red) and range of  $F_{ST}$  is shown.

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

## 582 Extended Data Tables and Extended Data Figure Legends

583 **Extended Data Table 1: No evidence for admixture related to sub-Saharan Africans in**  
 584 **Natufians.** We computed the statistic  $f_4(\text{Natufian}, \text{Other Ancient}, \text{African}, \text{Chimp})$  varying *African* to  
 585 be Mbuti, Yoruba, Ju\_hoan\_North, or the ancient Mota individual. Gene flow between Natufians and  
 586 African populations would be expected to bias these statistics positive. However, we find most of  
 587 them to be negative in sign and all of them to be non-significant ( $|Z| < 3$ ), providing no evidence that  
 588 Natufians differ from other ancient samples with respect to African populations.

<i>Other Ancient</i>	<i>African</i>	$f_4(\text{Natufian}, \text{Other Ancient}, \text{African}, \text{Chimp})$	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

589

590



591 **Extended Data Table 2: Admixture  $f_3$ -statistics.** We show the lowest Z-score of the statistic  $f_3(\text{Test};$   
 592  $\text{Reference}_1, \text{Reference}_2)$  for every ancient *Test* population with at least 2 individuals and every pair  
 593  $(\text{Reference}_1, \text{Reference}_2)$  of ancient or present-day source populations. Z-scores lower than -3 are  
 594 highlighted and indicate that the *Test* population is admixed from sources related to (but not identical  
 595 to) the reference populations. Z-scores greater than -3 are consistent with the population either being  
 596 admixed or not.

<i>Test</i>	<i>Reference</i> <sub>1</sub>	<i>Reference</i> <sub>2</sub>	$f_3(\text{Test}; \text{Reference}_1, \text{Reference}_2)$	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_HotulIb	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_HotulIb	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_HotulIb	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_HotulIb	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

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

601 **Extended Data Figure 2: Outgroup  $f_3(\text{Mbuti}; X, Y)$  for pairs of ancient populations.** The  
602 dendrogram is plotted for convenience and should not be interpreted as a phylogenetic tree. Areas of  
603 high shared genetic drift are ‘yellow’ and include from top-right to bottom-left along the diagonal:  
604 early Anatolian and European farmers; European hunter-gatherers, Steppe populations and ones  
605 admixed with steppe ancestry; populations from the Levant from the Epipaleolithic (Natufians) to the  
606 Bronze Age; populations from Iran from the Mesolithic to the Late Neolithic.

607 **Extended Data Figure 3: Reduction of genetic differentiation in West Eurasia over time.** We  
608 measure differentiation by  $F_{ST}$ . Each column of the 5x5 matrix of plots represents a major region and  
609 each row the earliest population with at least two individuals from each major region.

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

618 **Extended Data Figure 5: Inferred position of ancient populations in West Eurasian PCA**  
619 **according to the model of Fig. 4.**

620 **Extended Data Figure 6: Admixture from ghost populations using ‘cline intersection’.** We model  
621 each *Test* population (purple) in panels (a-f) as a mixture (pink) of a fixed reference population (blue)  
622 and a ghost population (orange) residing on the cline defined by two other populations (red and green)  
623 according to the visualization method of Supplementary Information, section 10. (a) Early/Middle  
624 Bronze Age steppe populations are a mixture of Iran\_ChL and a population on the WHG→SHG cline.  
625 (b) Scandinavian hunter-gatherers (SHG) are a mixture of WHG and a population on the  
626 Iran\_ChL→Steppe\_EMBA cline. (c) Caucasus hunter-gatherers (CHG) are a mixture of Iran\_N and  
627 both WHG and EHG. (d) Late Neolithic/Bronze Age Europeans are a mixture of the preceding  
628 Europe\_MNChL population and a population with both EHG and Iran\_ChL ancestry. (e) Somali are a  
629 mixture of Mota and a population on the Iran\_ChL→Levant\_BA cline. (f) Eastern European hunter-  
630 gatherers (EHG) are a mixture of WHG and a population on the Onge→Han cline.

631 **Extended Data Figure 7: Admixture from a ‘ghost’ ANE population into both European and**  
632 **Eastern Eurasian ancestry.** EHG, and Upper Paleolithic Siberians Mal’ta 1 (MA1) and Afontova  
633 Gora 2 (AG2) are positioned near the intersection of clines formed by European hunter-gatherers  
634 (WHG, SHG, EHG) and Eastern non-Africans in the space of outgroup  $f_3$ -statistics of the form  
635  $f_3(\text{Mbuti}; \text{Papuan}, \text{Test})$  and  $f_3(\text{Mbuti}; \text{Switzerland\_HG}, \text{Test})$ .

636

## 637 **References**

- 638 1. Barker, G. & Goucher, C. *The Cambridge World History Volume II: A world with*  
639 *agriculture, 12,000 BCE-500 CE.* (Cambridge University Press, 2015).
- 640 2. Cavalli-Sforza, L. L. & Paolo, M. *The History and Geography of Human Genes.* (Princeton  
641 University Press, 1994).
- 642 3. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory.  
643 *Nat. Commun.* **5**, 5257 (2014).
- 644 4. Pinhasi, R. *et al.* Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous  
645 Bone. *PLoS ONE* **10**, e0129102, (2015).
- 646 5. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc.*  
647 *Natl. Acad. Sci. USA* **110**, 2223–2227, (2013).
- 648 6. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor.  
649 *Nature* **524**, 216–+, (2015).
- 650 7. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages  
651 in Europe. *Nature* **522**, 207–211, (2015).
- 652 8. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**,  
653 499–503, (2015).
- 654 9. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat.*  
655 *Commun.* **6**, 8912, (2015).
- 656 10. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172,  
657 (2015).
- 658 11. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia.  
659 *Nature* **514**, 445–449, (2014).
- 660 12. Günther, T. *et al.* Ancient genomes link early farmers from Atapuerca in Spain to modern-day  
661 Basques. *Proceedings of the National Academy of Sciences* **112**, 11917–11922, (2015).
- 662 13. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-  
663 day Europeans. *Nature* **513**, 409–413, (2014).
- 664 14. Olalde, I. *et al.* A common genetic origin for early farmers from Mediterranean Cardial and  
665 Central European LBK cultures. *Mol. Biol. Evol.* **32**, 3132–3142, (2015).
- 666 15. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native  
667 Americans. *Nature* **505**, 87–91, (2014).
- 668 16. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093, (2012).
- 669 17. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**,  
670 e190, (2006).
- 671 18. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in  
672 unrelated individuals. *Genome Res.* **19**, 1655–1664, (2009).
- 673 19. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains.  
674 *Nature* **505**, 43–49, (2014).
- 675 20. Kim, Bernard Y. & Lohmueller, Kirk E. Selection and Reduced Population Size Cannot  
676 Explain Higher Amounts of Neanderthal Ancestry in East Asian than in European Human  
677 Populations. *The American Journal of Human Genetics* **96**, 454–461, (2015).
- 678 21. Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual.  
679 *Science* **338**, 222–226, (2012).

- 680 22. Vernot, B. & Akey, Joshua M. Complex History of Admixture between Modern Humans and  
681 Neandertals. *The American Journal of Human Genetics* **96**, 448-453, (2015).
- 682 23. Wall, J. D. *et al.* Higher Levels of Neanderthal Ancestry in East Asians Than in Europeans.  
683 *Genetics* **194**, 199-209, (2013).
- 684 24. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, (2010).
- 685 25. Brace, C. L. *et al.* The questionable contribution of the Neolithic and the Bronze Age to  
686 European craniofacial form. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 242-247, (2006).
- 687 26. Ferembach, D. Squelettes du Natoufien d'Israel., etude anthropologique. . *L'Anthropologie* **65**,  
688 46-66, (1961).
- 689 27. Fadhlouli-Zid, K. *et al.* Genome-Wide and Paternal Diversity Reveal a Recent Origin of  
690 Human Populations in North Africa. *PLoS ONE* **8**, e80293, (2013).
- 691 28. Henn, B. M. *et al.* Genomic ancestry of North Africans supports back-to-Africa migrations.  
692 *PLoS genetics* **8**, e1002397, (2012).
- 693 29. Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting F(ST):  
694 The impact of rare variants. *Genome Res.* **23**, 1514-1521, (2013).
- 695 30. Fernandez, E. *et al.* Ancient DNA analysis of 8000 B.C. near eastern farmers supports an  
696 early neolithic pioneer maritime colonization of Mainland Europe through Cyprus and the  
697 Aegean Islands. *PLoS genetics* **10**, e1004401, (2014).
- 698 31. Ammerman, A. J., Pinhasi, R. & Banffy, E. Comment on "Ancient DNA from the first  
699 European farmers in 7500-year-old Neolithic sites". *Science* **312**, 1875; author reply 1875,  
700 (2006).
- 701 32. Pagani, L. *et al.* Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex  
702 Influences on the Ethiopian Gene Pool. *Am. J. Hum. Genet.* **91**, 83-96, (2012).
- 703 33. Pickrell, J. K. *et al.* Ancient west Eurasian ancestry in southern and eastern Africa. *Proc.*  
704 *Natl. Acad. Sci. USA* **111**, 2632–2637, (2014).
- 705 34. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by  
706 whole-genome sequencing. *Nat. Commun.* **3**, 698, (2012).
- 707 35. Moorjani, P. *et al.* Genetic evidence for recent population mixture in India. *Am. J. Hum.*  
708 *Genet.* **93**, 422-438, (2013).
- 709 36. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian  
710 population history. *Nature* **461**, 489-494, (2009).
- 711 37. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205, (2016).
- 712 38. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear  
713 reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of*  
714 *Sciences of the United States of America* **110**, 15758-15763, (2013).
- 715 39. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil–DNA–  
716 glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal*  
717 *Society of London B: Biological Sciences* **370**, DOI: 10.1098/rstb.2013.0624 (2014).
- 718 40. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo methylation in  
719 ancient DNA. *Nucleic acids research* **38**, e87, (2010).
- 720 41. Korlevic, P. *et al.* Reducing microbial and human contamination in DNA extractions from  
721 ancient bones and teeth. *BioTechniques* **59**, 87-93, (2015).
- 722 42. Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos.  
723 *Nature* **505**, 403-406, (2014).
- 724 43. Behar, D. M. *et al.* A “Copernican” Reassessment of the Human Mitochondrial DNA Tree  
725 from its Root. *Am. J. Hum. Genet.* **90**, 675-684, (2012).
- 726 44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.  
727 *Bioinformatics* **25**, 1754-1760, (2009).
- 728 45. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation  
729 Sequencing Data. *BMC Bioinformatics* **15**, 1-13, (2014).
- 730 46. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based  
731 linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575, (2007).
- 732 47. Chang, C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer  
733 datasets. *GigaScience* **4**, 7, (2015).

- 734 48. Busing, F. T. A., Meijer, E. & Leeden, R. Delete-m Jackknife for Unequal m. *Statistics and*  
735 *Computing* **9**, 3-8, (1999).
- 736 49. Sudmant, P. H. *et al.* Global diversity, population stratification, and selection of human copy-  
737 number variation. *Science* **349**, (2015).
- 738 50. Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370-374,  
739 (2012).

740

741

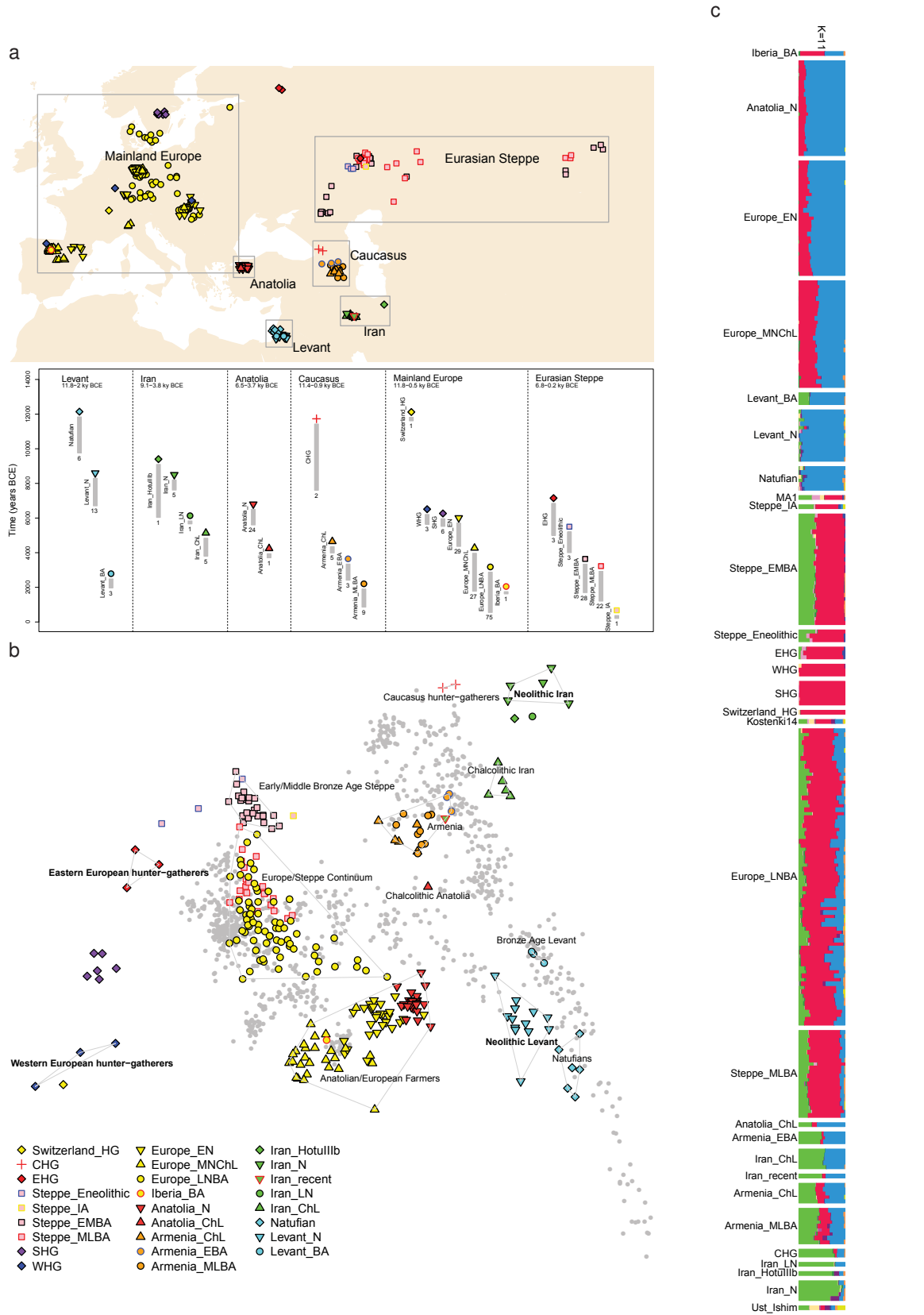


Figure 2

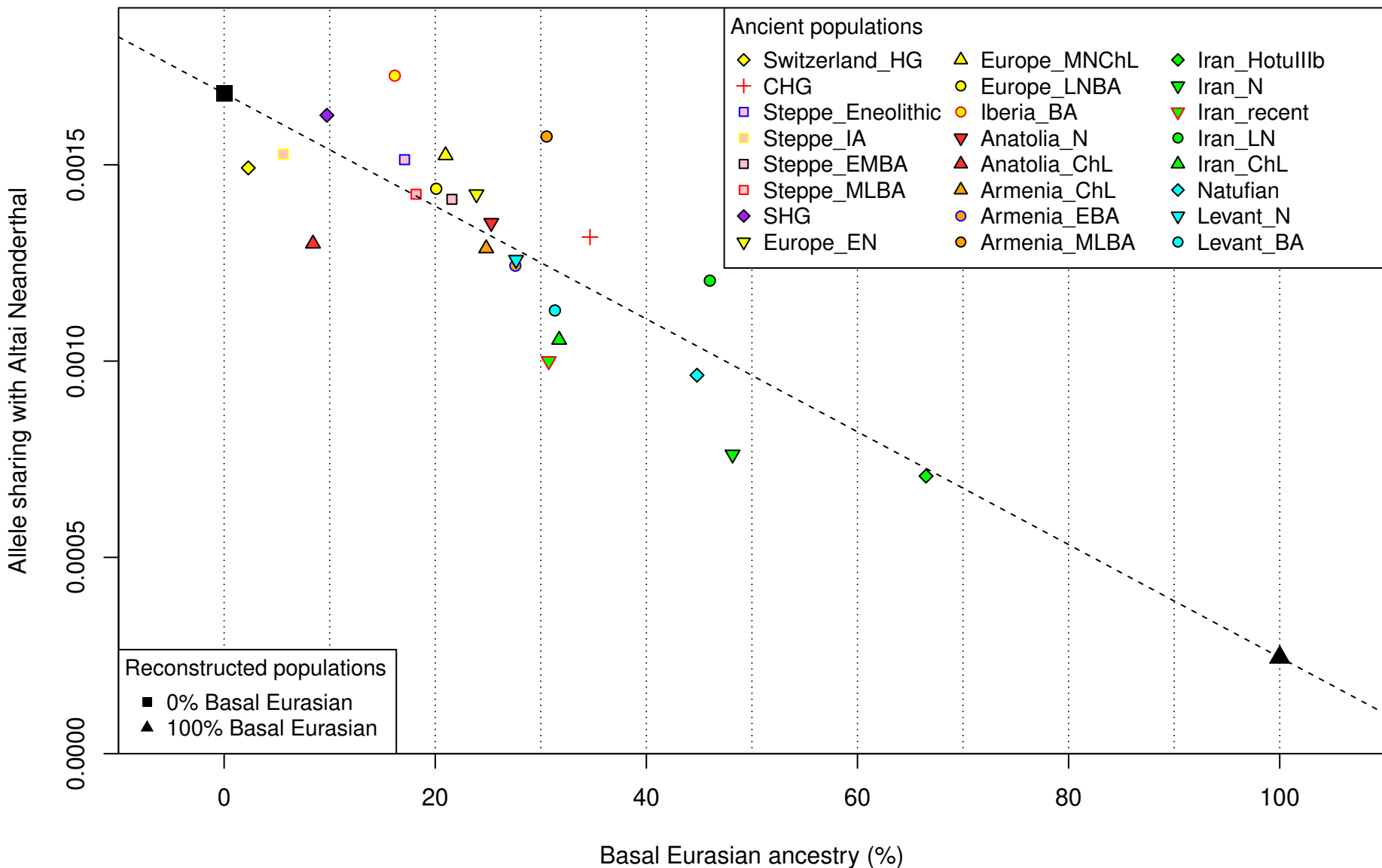
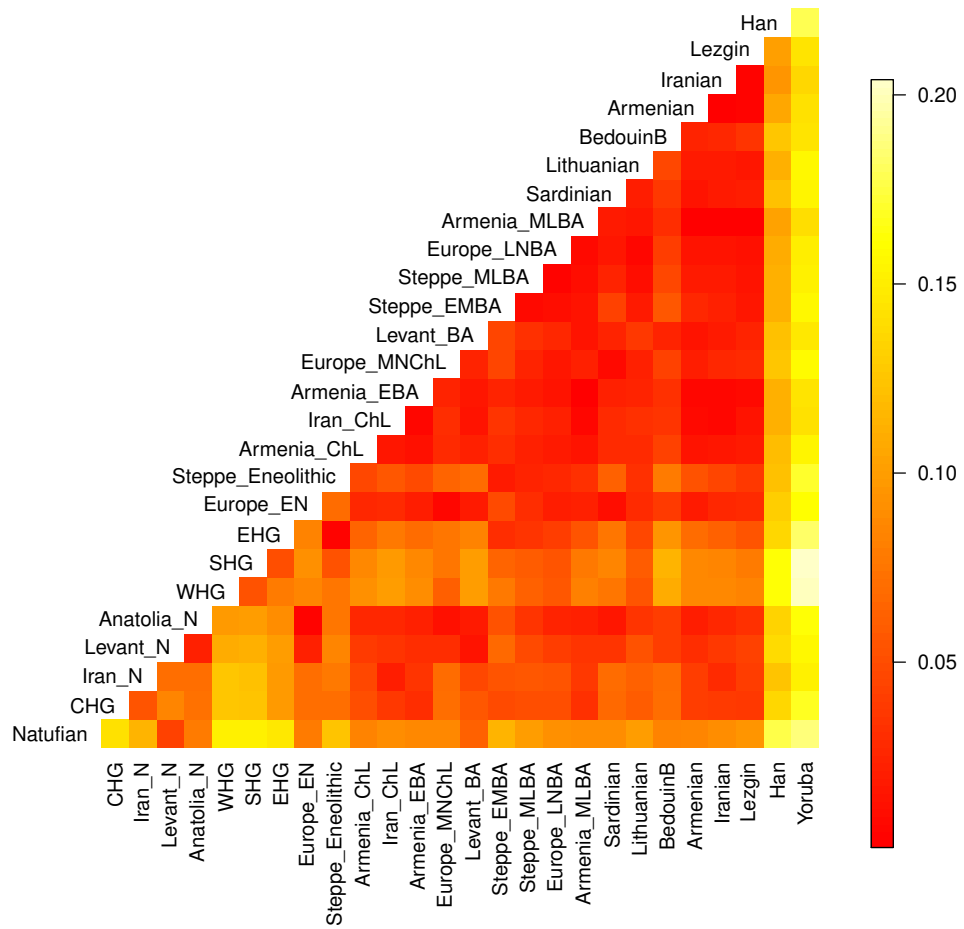




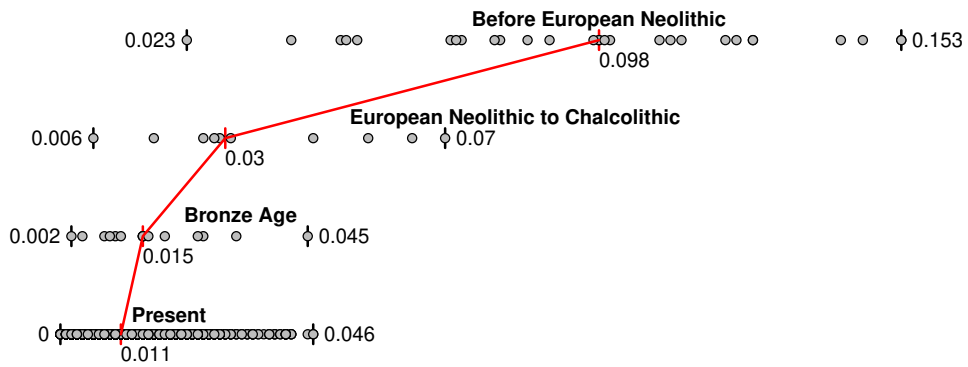
Figure 3

a

bioRxiv preprint doi: <https://doi.org/10.1101/059311>; this version posted June 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



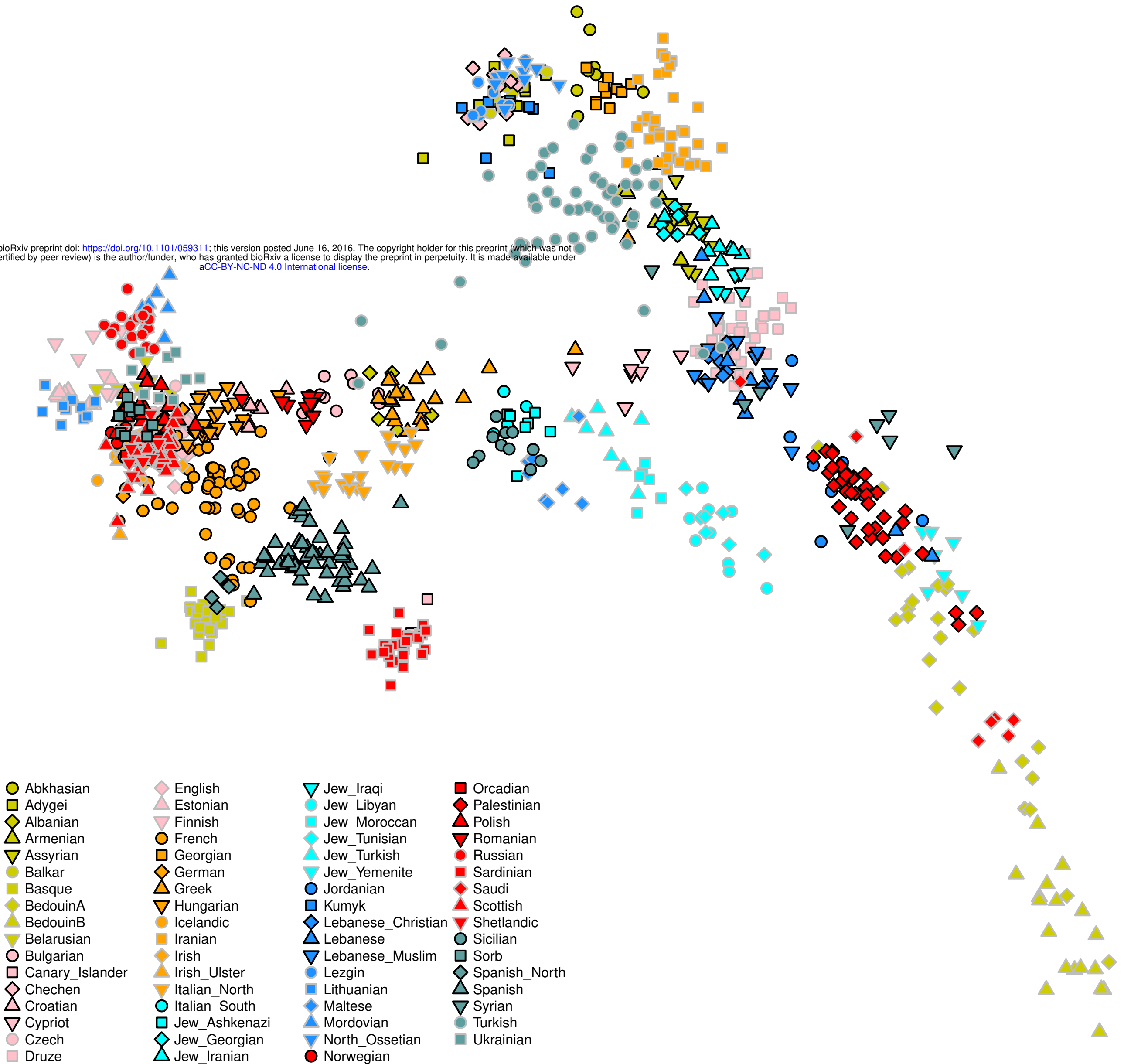
b



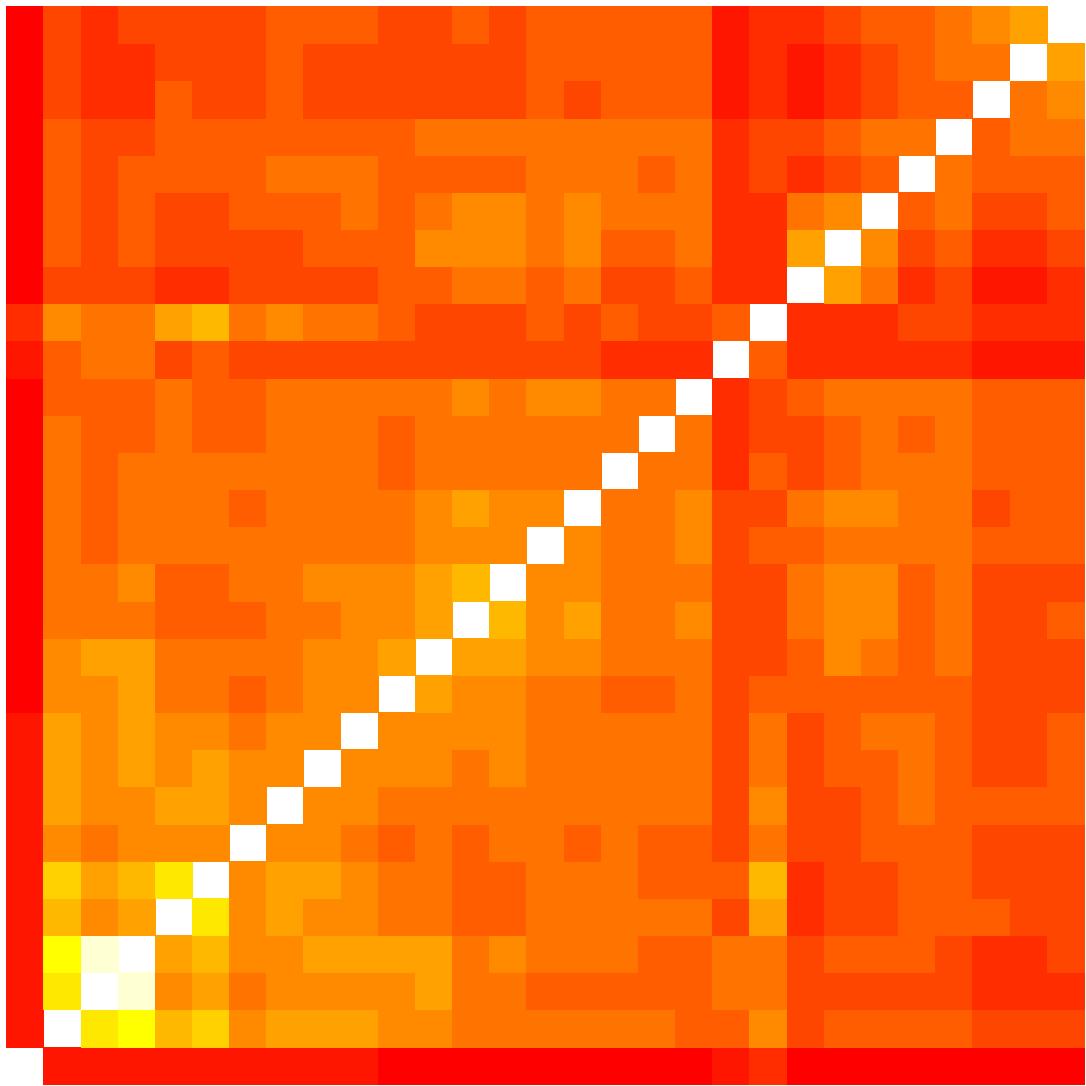
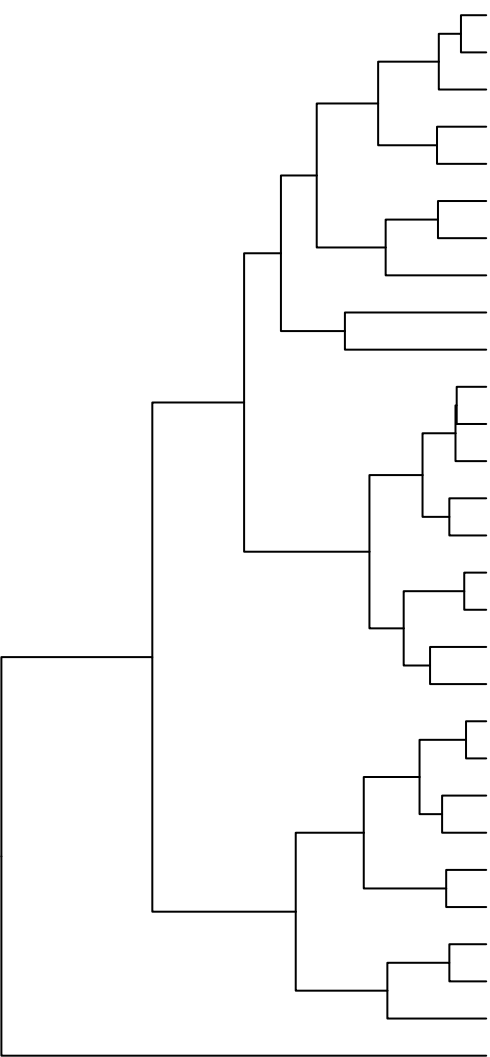
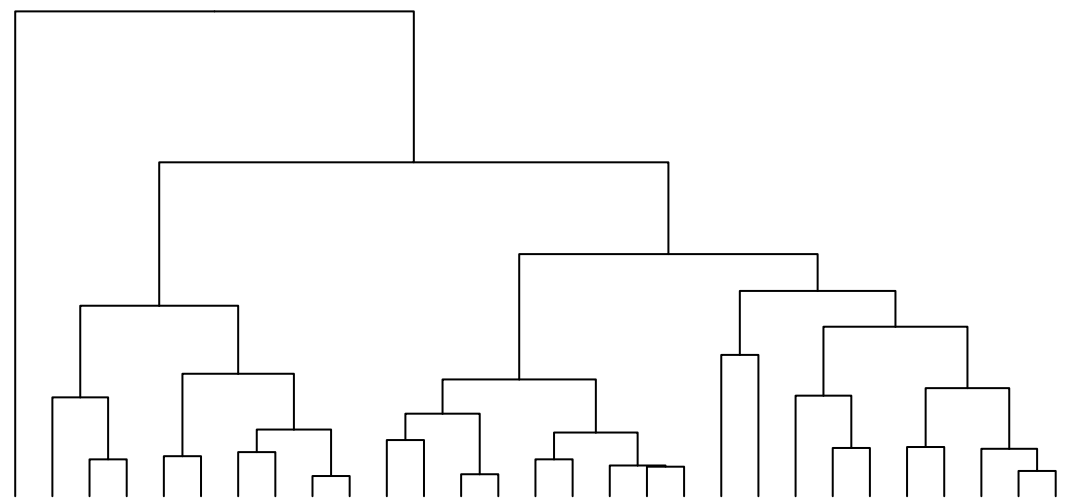
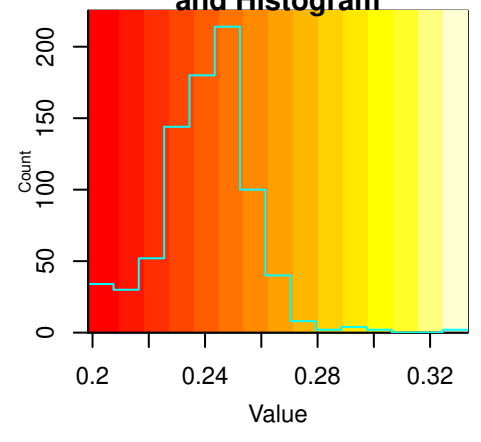


Extended Data Figure 1

bioRxiv preprint doi: <https://doi.org/10.1101/059311>; this version posted June 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



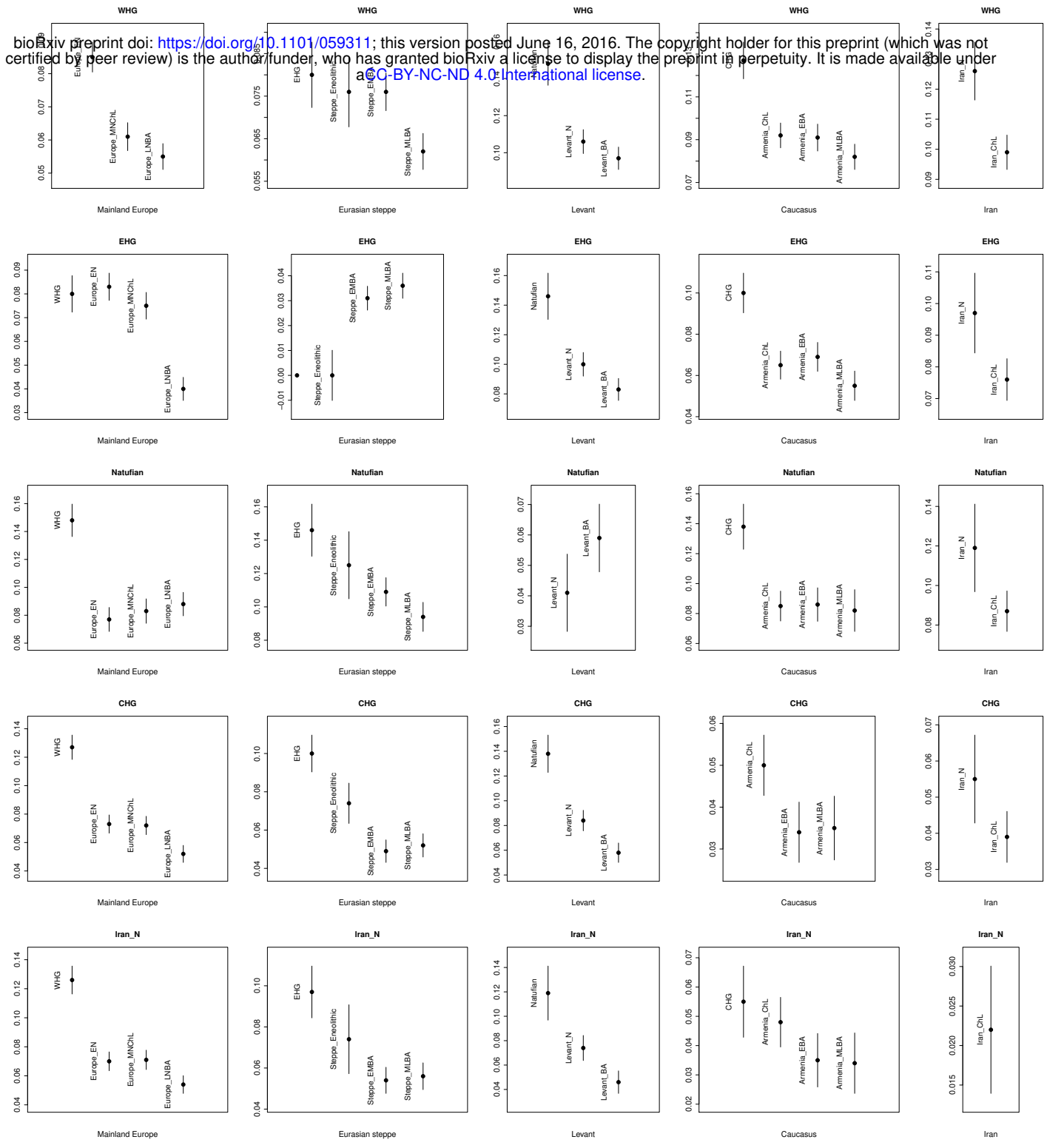
**Color Key  
and Histogram**



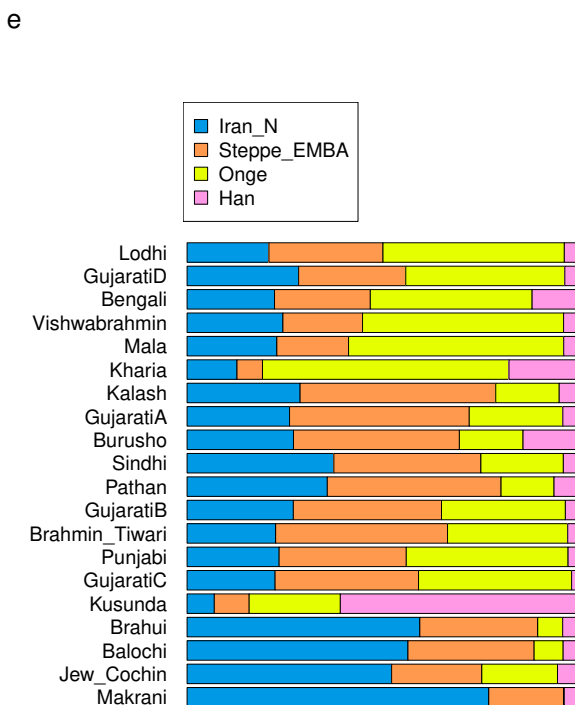
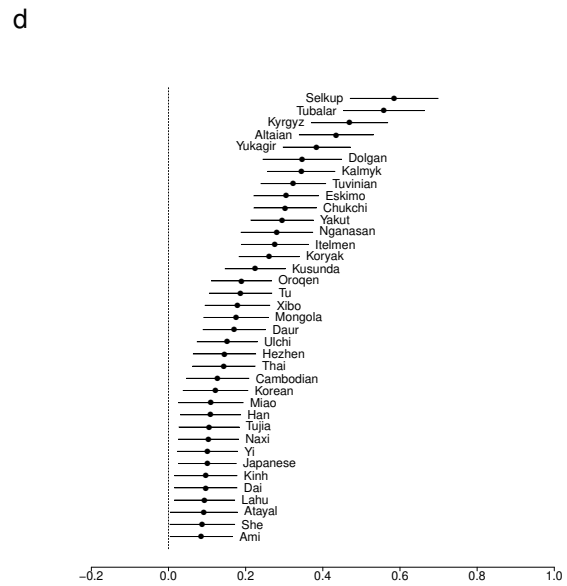
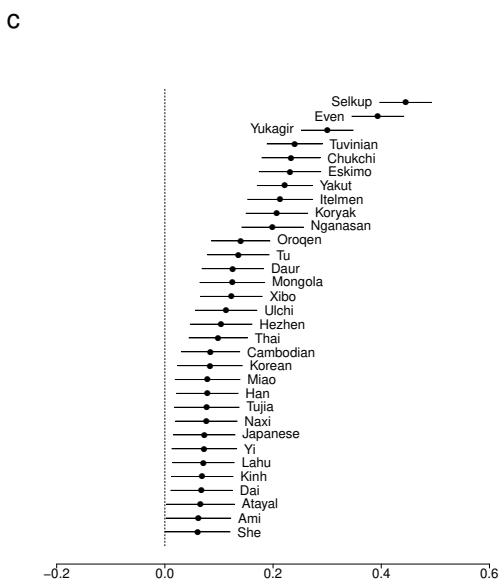
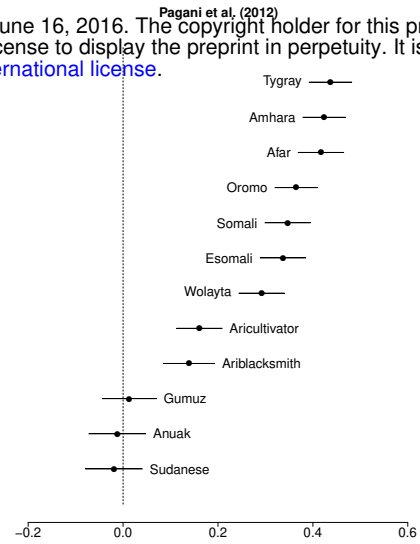
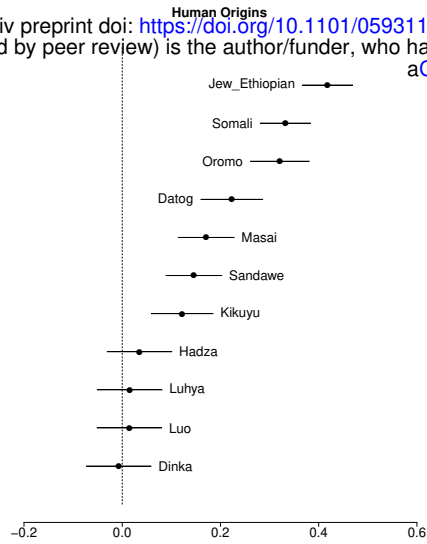
- Iran\_LN
- Iran\_N
- Iran\_Mesolithic
- Iran\_ChL
- CHG
- Levant\_BA
- Levant\_N
- Natufian
- MA1
- Kostenki14
- Armenia\_EBA
- Iran\_N\_outlier
- Armenia\_MLBA
- Anatolia\_ChL
- Armenia\_ChL
- Europe\_EN
- Anatolia\_N
- Europe\_MNChL
- Iberia\_BA
- Europe\_LNBA
- Steppe\_MLBA
- Steppe\_EMBA
- Steppe\_IA
- EHG
- Steppe\_Eneolithic
- WHG
- Switzerland\_HG
- SHG
- Ust\_Ishim

- Ust\_Ishim
- SHG
- Switzerland\_HG
- WHG
- Steppe\_Eneolithic
- EHG
- Steppe\_IA
- Steppe\_EMBA
- Steppe\_MLBA
- Europe\_LNBA
- Iberia\_BA
- Europe\_MNChL
- Anatolia\_N
- Europe\_EN
- Armenia\_ChL
- Anatolia\_ChL
- Armenia\_MLBA
- Iran\_N\_outlier
- Armenia\_EBA
- Kostenki14
- MA1
- Natufian
- Levant\_N
- Levant\_BA
- CHG
- Iran\_ChL
- Iran\_Mesolithic
- Iran\_N
- Iran\_LN

Extended Data Figure 2



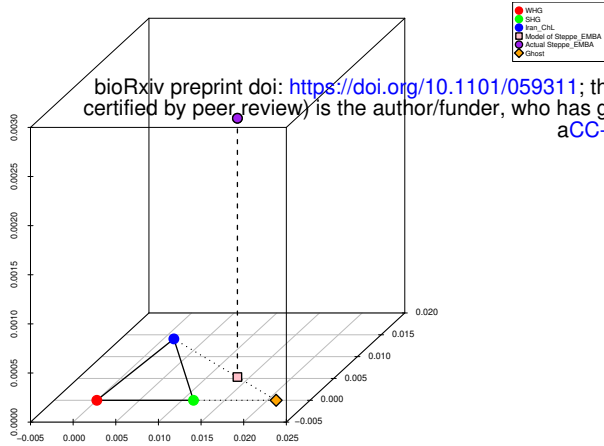
Extended Data Figure 3



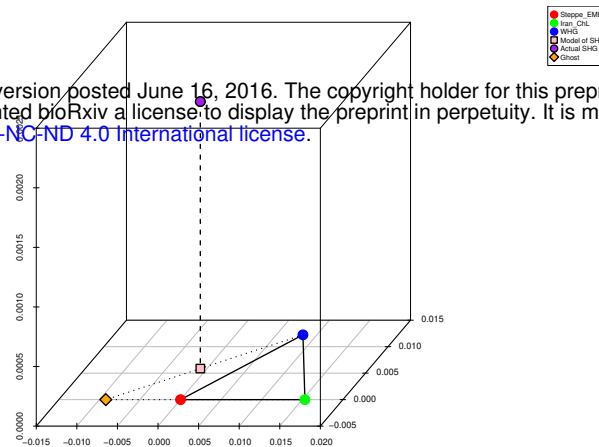
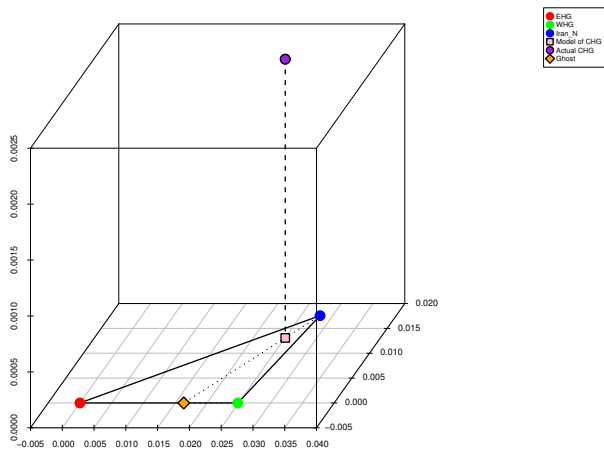
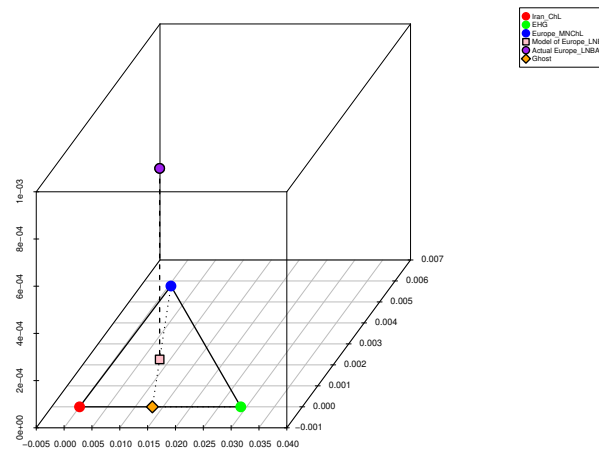
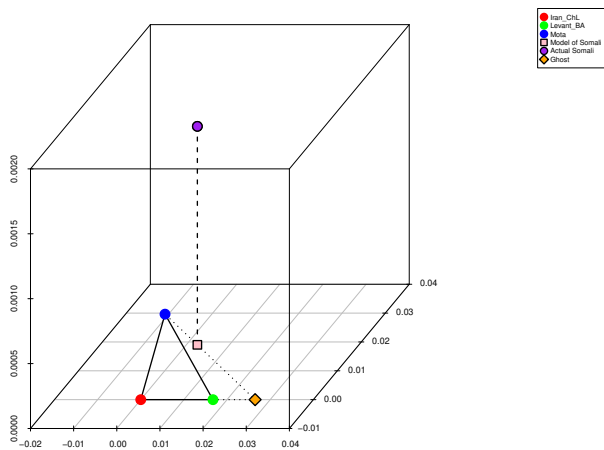
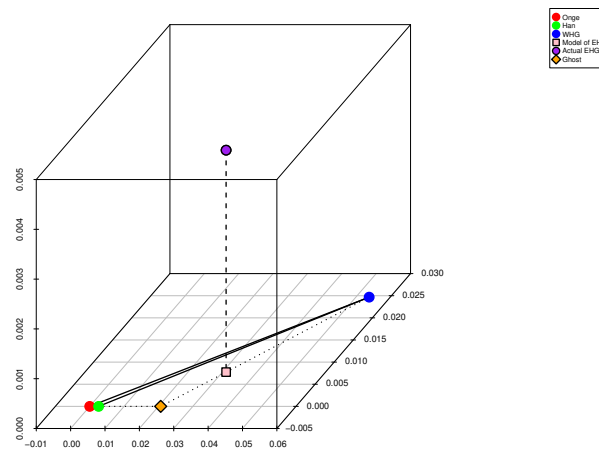


Extended Data Figure 5



**a**

bioRxiv preprint doi: <https://doi.org/10.1101/059311>; this version posted June 16, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**b****c****d****e****f**

Extended Data Figure 6

Extended Data Figure 7

