Paleolithic DNA from the Caucasus reveals core of West Eurasian ancestry

Iosif Lazaridis[1,2], Anna Belfer-Cohen[3], Swapan Mallick[1], Nick Patterson[2], Olivia Cheronet[4,5,6], Nadin Rohland[1], Guy Bar-Oz[7], Ofer Bar-Yosef[8], Nino Jakeli[9], Eliso Kvavadze[9], David Lordkipanidze[9], Zinovi Matzkevich[10], Tengiz Meshveliani[9], Brendan J. Culleton[11], Douglas J. Kennett[11], Ron Pinhasi[4] §, David Reich[1,2,12]§

[1] Department of Genetics, Harvard Medical School, New Research Building, 77 Ave. Louis Pasteur, Boston, MA 02115, USA
[2] Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02141, USA
[3] Institute of Archaeology, The Hebrew University of Jerusalem, Mt. Scopus, Jerusalem 9190501, Israel
[4] Department of Evolutionary Anthropology, University of Vienna, 1090 Vienna, Austria.
[5] Earth Institute, University College Dublin, Dublin 4, Ireland.
[6] School of Archaeology, University College Dublin, Dublin 4, Ireland.
[7] Zinman Institute of Archaeology, University of Haifa, Haifa 3498838, Israel
[8] Department of Anthropology, Peabody Museum, Harvard University, Cambridge, MA 02138, USA
[9] Georgian National Museum, 3 Purtseladze St., Tbilisi 0105, Georgia
[10] Israel Antiquities Authority, PO Box 586, Jerusalem 9100402, Israel
[11] Department of Anthropology, The Pennsylvania State University, University Park, PA 16802, USA
[12] Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA

§ These authors jointly supervised this work

The earliest ancient DNA data of modern humans from Europe dates to ~40 thousand years ago[1-4], but that from the Caucasus and the Near East to only ~14 thousand years ago[5,6], from populations who lived long after the Last Glacial Maximum (LGM) ~26.5-19 thousand years ago[7]. To address this imbalance and to better understand the relationship of Europeans and Near Easterners, we report genome-wide data from two ~26 thousand year old individuals from Dzudzuana Cave in Georgia in the Caucasus from around the beginning of the LGM. Surprisingly, the Dzudzuana population was more closely related to early agriculturalists from western Anatolia ~8 thousand years ago[8] than to the hunter-gatherers of the Caucasus from the same region of western Georgia of ~13-10 thousand years ago[5]. Most of the Dzudzuana population's ancestry was deeply related to the post-glacial western European hunter-gatherers of the 'Villabruna cluster'[3], but it also had ancestry from a lineage that had separated from the great majority of non-African populations before they separated from each other, proving that such 'Basal Eurasians'[6,9] were present in West Eurasia twice as early as

41 **previously recorded[5,6]. We document major population turnover in the Near East after**

42 **the time of Dzudzuana, showing that the highly differentiated Holocene populations of**

43 **the region[6] were formed by 'Ancient North Eurasian'[3,9,10] admixture into the Caucasus**

44 **and Iran and North African[11,12] admixture into the Natufians of the Levant. We finally**

45 **show that the Dzudzuana population contributed the majority of the ancestry of post-**

46 **Ice Age people in the Near East, North Africa, and even parts of Europe, thereby**

47 **becoming the largest single contributor of ancestry of all present-day West Eurasians.**

48 Ancient DNA has revealed more about the deep history of Europe than of any other

49 continent, with dozens of Paleolithic samples reported to date[1-5] (Fig. 1a). Genetic analyses

50 show that the first populations related to present-day West Eurasians arrived in Europe at

51 least ~36 thousand years ago (kya)[2]. A new group of populations (Věstonice cluster),

52 associated with the archaeologically defined Gravettian entity, appeared in the genetic record

53 of Europe by ~30kya, while another group, associated with the archaeologically defined

54 Magdalenian culture, appeared in Europe by ~20kya (El Mirón cluster) [3]. By ~14kya a third

55 group, the Villabruna cluster, appeared throughout mainland Europe, coinciding with the

56 Bølling-Allerød warming period[3]. Members of this cluster, which has also been called

57 western European hunter-gatherers (WHG), were found across Europe during Late Upper

58 Paleolithic-to-Mesolithic times, and were the main pre-agricultural Europeans prior to the

59 Neolithic ~8kya[9].

60 In contrast to this detailed knowledge about Europe during the Paleolithic, no Ice Age DNA

61 has been published from the Near East (including the Caucasus) whose post-glacial and

62 Holocene-era populations <15kya were highly differentiated from both those of Europe and

63 also from each other[5,6,8,13,14]. To address this deficit, we analyzed teeth from two individuals

64 recovered from Dzudzuana Cave[15], Southern Caucasus, (Fig. 1a; Supplementary Information

65 section 1) from an archaeological layer previously dated to ~27-24kya and whose age

66 determination was confirmed by a series of 8 new dates (Extended Data Figure 1;

67 Supplementary Information section 1), thus allowing us to probe a population on the cusp of

68 the beginning of the LGM. Of the two individuals, one yielded data at a total of 185,744

69 targeted single nucleotide polymorphisms (SNPs) on chromosomes 1-22 (I2949, Dzu2, Upper

70 Area square H16b), while the other yielded 90,916 SNPs (I2963, Dzu3, Lower Area square

71 18b). Both individuals had mitochondrial DNA sequences (U6 and N) that are consistent with

72 deriving from lineages that are rare in the Caucasus or Europe today. The two individuals

73 were genetically similar to each other, consistent with belonging to the same population

74 (Extended Data Fig. 2) and we thus analyze them jointly.

75 The Dzudzuana samples represent the earliest ancient modern human DNA outside of

76 Europe, Siberia, and China (Fig. 1b). In the local context, they help us answer the question of

77 the relationship of Ice Age populations of the region to their post-glacial successors[5]: was

78 there discontinuity in the Caucasus as in Europe? A broader issue that we wished to address

79 is the changing relationship between human populations from Europe and the Near East, as

80 the Caucasus is situated at the border between them. The Villabruna cluster has been modeled

81 as contributing to both the ~30kya Věstonice and ~20kya El Mirón-cluster populations[3],

82 suggesting that it must have existed somewhere in relatively unmixed form long before the

83 oldest genetic data we have from it at ~14kya[3,5]. However, it is unlikely that the Villabruna

84 cluster sojourned in mainland Europe, as members of the cluster have been attested there only

85 by ~14kya, marking an increased affinity of these European populations of the time to Near

86 Eastern ones[3]. Was there migration at the time from mainland Europe to the Near East or vice

87 versa, or, indeed from a geographically intermediate Ice Age refugium in southeast Europe,

88 Anatolia, or the circum-Pontic (Black Sea) region that might explain the affinity of post-

89 glacial Levantine and Anatolian populations to those of Europe[6]? It is also unknown how the

90 affinity between early populations in the eastern European-Caucasus-Iran zone[6] first arose.

91 Eastern European hunter-gatherers (EHG)[16] ~8kya can be modeled as a mixture of peoples of

92 WHG and Upper Paleolithic Siberians first known ~24kya[10] (also known as 'Ancient North

93 Eurasians' (ANE)). Caucasus hunter-gatherers (CHG)—sampled in Georgia in Satsurblia and

94 Kotias Klde caves <50km from Dzudzuana[5]—were genetically intermediate between EHG

95 and the first agriculturalists of Iran sampled from the Zagros mountains (Iran_N; ~10kya)[6,13].

96 We first estimated F_{ST}, a measure of population genetic differentiation, to assess the genetic

97 relationships between ancient West Eurasian populations (Extended Data Table 1; Methods).

98 Post-glacial Near Easterners and North Africans (PGNE) (CHG, Natufians, Taforalt[11] Ibero-

99 Maurusians from North Africa, and early Neolithic farmers from Anatolia[8], Iran[6], the

100 Levant[6], and the Maghreb[17]) are strongly differentiated from all European and Siberian

101 hunter-gatherers (ESHG) (F_{ST} = 0.078−0.267). By contrast, Dzudzuana is genetically closer

102 to both contemporaneous Gravettians from Europe (0.051±0.012) and also to the much later

103 Neolithic Anatolian farmers (0.039±0.005) who are genetically closest to them according to

104 this measure. Genetic drift inflates F_{ST} over time, so the affinity to the Gravettians may partly

105   be due to the great age of these samples. However, age cannot explain the affinity to much

106   later Neolithic Anatolians of ~8kya, a population closer to Dzudzuana than any other PGNE

107   (0.052−0.195).

108   Outgroup f_3-statistics[10] show that Dzudzuana clusters with Near Eastern populations

109   primarily from Anatolia and secondarily from the Levant, but not with the geographically

110   proximate CHG (Extended Data Fig. 3). A genetic relationship between Dzudzuana and

111   Neolithic Anatolians is also shown by principal components analysis (PCA) in the space of

112   'outgroup f_4-statistics'[16] of the form $f_4(Test, O_1; O_2, O_3)$ where $(O_1; O_2, O_3)$ is a triple of

113   outgroups (Fig. 1c; Methods); performing PCA on the space defined by these statistics has

114   the advantage of not being affected by genetic drift peculiar to the *Test* populations. It also

115   allows us to visualize genetic relationships between ancient populations alone, without

116   projecting onto the variation of present-day people. European hunter-gatherers in our analysis

117   form a cline with Villabruna/WHG samples on one end and ANE on the other. None of the

118   PGNE populations other than the Neolithic Anatolians cluster with the Ice Age Caucasus

119   population from Dzudzuana. As reported previously, present-day West Eurasians are much

120   more homogeneous than ancient ones, reflecting extensive post-Neolithic admixture[6].

121   However, they continue to be differentially related to ancient local populations in Europe and

122   the Near East (Extended Data Fig. 4).

123   To better understand the relationship of Dzudzuana to other ancient West Eurasian

124   populations, we performed symmetry testing using *f*-statistics[18] (Extended Data Fig. 5). These

125   analyses show that ESHG share more alleles with Dzudzuana than with PGNE populations,

126   except Neolithic Anatolians who form a clade with Dzudzuana to the exclusion of ESHG

127   (Extended Data Fig. 5a). Thus, our results prove that the European affinity of Neolithic

128   Anatolians[6] does not necessarily reflect any admixture into the Near East from Europe, as an

129   Anatolian Neolithic-like population already existed in parts of the Near East by ~26kya.

130   Furthermore, Dzudzuana shares more alleles with Villabruna-cluster groups than with other

131   ESHG (Extended Data Fig. 5b), suggesting that this European affinity was specifically

132   related to the Villabruna cluster, and indicating that the Villabruna affinity of PGNE

133   populations from Anatolia and the Levant is not the result of a migration into the Near East

134   from Europe. Rather, ancestry deeply related to the Villabruna cluster was present not only in

135   Gravettian and Magdalenian-era Europeans[3] but also in the populations of the Caucasus, by

136   ~26kya. Neolithic Anatolians, while forming a clade with Dzudzuana with respect to ESHG

137   (Extended Data Fig. 5a), share more alleles with all other PGNE (Extended Data Fig. 5d),

138   suggesting that PGNE share at least partially common descent to the exclusion of the much

139   older samples from Dzudzuana.

140   All known ancient Near Eastern populations prior to this work were inferred to harbor 'Basal

141   Eurasian' ancestry[9], a branch that diverged from all other non-Africans (including ESHG and

142   present-day East Asians and Oceanians) before they split from each other. The CHG,

143   geographically intermediate between Europe and the Near East resembled Near Eastern

144   populations in the possession of Basal Eurasian ancestry[5]. The Dzudzuana population was not

145   identical to the WHG, as it shared fewer alleles with both an early Upper Paleolithic Siberian

146   (Ust'Ishim[19]) and an early Upper Paleolithic East Asian (Tianyuan[20]) (Extended Data Fig.

147   5c), thus, it too—like the PGNE populations—had Basal Eurasian ancestry[6,9]. The detection

148   of this type of ancestry, twice as early as previously documented[5,6] and at the northern edge

149   of the Near East, lends weight to the hypothesis that it represents a deep Near Eastern lineage

150   rather than a recent arrival from Africa[6].

151   We used qpGraph[18] to build an admixture graph model of the relationship between ESHG

152   and Dzudzuana, also including the earliest PGNE populations from North Africa (Taforalt)

153   and the Epipaleolithic Levant (Natufians) (Fig. 2). While potentially oversimplifying the

154   history of these populations by considering only discrete binary admixture events as opposed

155   to continuous gene flow, the model is useful for its insights into possible evolutionary

156   relationships between populations and for representing the minimum complexity that these

157   relationships had. According to this model, a common population contributed ancestry to

158   Gravettians (represented by Vestonice16) and to a "Common West Eurasian" population that

159   contributed all the ancestry of Villabruna and most of the ancestry of Dzudzuana which also

160   had 28.4±4.2% Basal Eurasian ancestry[21] (Supplementary Information section 2).

161   Our co-modeling of Epipaleolithic Natufians and Ibero-Maurusians from Taforalt confirms

162   that the Taforalt population was mixed[11], but instead of specifying gene flow from the

163   ancestors of Natufians into the ancestors of Taforalt as originally reported, we infer gene flow

164   in the reverse direction (into Natufians). The Neolithic population from Morocco, closely

165   related to Taforalt[17] is also consistent with being descended from the source of this gene flow,

166   and appears to have no admixture from the Levantine Neolithic (Supplementary Information

167   section 3). If our model is correct, Epipaleolithic Natufians trace part of their ancestry to

168   North Africa, consistent with morphological and archaeological studies that indicate a spread

169   of morphological features[22] and artifacts from North Africa into the Near East. Such a

170   scenario would also explain the presence of Y-chromosome haplogroup E in the Natufians

171   and Levantine farmers[6], a common link between the Levant and Africa. Moreover, our model

172   predicts that West Africans (represented by Yoruba) had 12.5±1.1% ancestry from a Taforalt-

173   related group rather than Taforalt having ancestry from an unknown Sub-Saharan African

174   source[11]; this may have mediated the limited Neanderthal admixture present in West

175   Africans[23]. An advantage of our model is that it allows for a local North African component

176   in the ancestry of Taforalt, rather than deriving them exclusively from Levantine and Sub-

177   Saharan sources.

178   We also used the qpWave/qpAdm framework[16] to model ancient populations without strong

179   phylogenetic assumptions (Supplementary Information section 3; Table 1). This analysis

180   shows that we cannot reject the hypothesis that Dzudzuana and the much later Neolithic

181   Anatolians form a clade with respect to ESHG (P=0.286), consistent with the latter being a

182   population largely descended from Dzudzuana-like pre-Neolithic populations whose

183   geographical extent spanned both Anatolia and the Caucasus. Dzudzuana itself can be

184   modeled as a 2-way mixture of Villabruna-related ancestry and a Basal Eurasian lineage.

185   Western PGNE populations, including Neolithic Anatolians, pre-pottery Neolithic farmers

186   from the Levant (PPNB), Natufians, and Taforalt, can all be modeled as a mixture of

187   Dzudzuana and additional 'Deep' ancestry that may represent an even earlier split than the

188   Basal Eurasians. Considering 2-way mixtures, we can model Karelia_HG as deriving

189   34±2.8% of its ancestry from a Villabruna-related source, with the remainder mainly from

190   ANE represented by the AfontovaGora3 (AG3) sample from Lake Baikal[3] ~17kya. Finally,

191   we can model CHG and samples from Neolithic Iran (Iran_N) as deriving their ancestry

192   largely (~58-64% using qpAdm and ~45-62% using qpGraph) from a Dzudzuana-like

193   population, but with ancestry from both 'Deep' and ANE sources, thus proving that ANE

194   ancestry had reached Western Eurasia long before the Bronze Age Eurasian steppe

195   migrations that carried further westward into mainland Europe.

196   In qpAdm modeling, a deeply divergent hunter-gatherer lineage that contributed in relatively

197   unmixed form to the much later hunter-gatherers of the Villabruna cluster is specified as

198   contributing to earlier hunter-gatherer groups (Gravettian Vestonice16: 35.7±11.3% and

199   Magdalenian ElMiron: 60.6±11.3%) and to populations of the Caucasus (Dzudzuana:

200   72.5±3.7%, virtually identical to that inferred using ADMIXTUREGRAPH). In Europe,

201   descendants of this lineage admixed with pre-existing hunter-gatherers related to Sunghir3

202   from Russia[4] for the Gravettians and GoyetQ116-1 from Belgium[3] for the Magdalenians,

203   while in the Near East it did so with Basal Eurasians. Later Europeans prior to the arrival of

204   agriculture were the product of re-settlement of this lineage after ~15kya in mainland Europe,

205   while in eastern Europe they admixed with Siberian hunter-gatherers forming the WHG-ANE

206   cline of ancestry (Fig. 1c). In the Near East, the Dzudzuana-related population admixed with

207   North African-related ancestry in the Levant and with Siberian hunter-gatherer and eastern

208   non-African-related ancestry in Iran and the Caucasus. Thus, the highly differentiated

209   populations at the dawn of the Neolithic[6] were primarily descended from Villabruna Cluster

210   and Dzudzuana-related ancestors, with varying degrees of additional input related to both

211   North Africa and Ancient North/East Eurasia whose proximate sources may be clarified by

212   future sampling of geographically and temporally intermediate populations.

213   The ancestry of present-day Europeans has been traced to the proximate sources of

214   Mesolithic hunter-gatherers, Early European/Anatolian farmers, and steppe pastoralists[16], but

215   the ancestry of Near Eastern and North African populations has not been investigated due to

216   lack of appropriate ancient sources. We present a unified analysis of diverse European, Near

217   Eastern, North African populations in terms of the deepest known sources of ancestry (Fig.

218   3), which suggests that Dzudzuana-related ancestry makes up ~46-88% of the ancestry of all

219   these populations, with Dzudzuana-related ancestry more strongly found in southern

220   populations across West Eurasia (Fig. 3; Extended Data Fig. 6). Dzudzuana-like ancestry

221   must have spread across West Eurasia with Neolithic migrations out of the Near East, but it

222   had not been previously completely absent from Europe as several hunter-gatherer

223   populations in southeastern Europe, eastern Europe, and Scandinavia can only be modeled

224   with some such ancestry (Extended Data Fig. 6; Supplementary Information section 4). Both

225   Europeans and Near Easterners also share in AG3-related ancestry of up to ~30% in eastern

226   Europe down to ~0% in parts of North Africa. Europeans are differentiated by an excess of

227   up to ~20% Villabruna-related ancestry relative to non-European populations and also by a

228   relative lack of extra 'Deep' ancestry compared to the Near East and North Africa, a type of

229   ancestry that may only partially be explained by the Basal Eurasian ancestry of ancient West

230   Eurasian populations and must also trace to Africa (Extended Data Fig. 7). 'Deep' ancestry,

231   including Basal Eurasian ancestry, is associated with reduced Neandertal ancestry

232   (Supplementary Information section 5, Extended Data Fig. 8), confirming that Neandertal

233   ancestry in West Eurasia[6] has been diluted by admixture.

234   Future studies must investigate when Dzudzuana-like populations first formed: does the

235   Basal Eurasian component represent the earliest Near Eastern population stratum or a later

accretion? What were the mechanisms and proximate sources of the Siberian- and North African-related ancestry that affected West Eurasia? We caution that the inference of Dzudzuana-related ancestry as the core component of ancient and present-day West Eurasia does not constitute proof for migrations specifically from the Caucasus: given that this is the only ancient DNA data from this time period and broad region, the geographical and temporal extent of this population and its relatives remains unknown. Both in its past (formed by admixture with Basal Eurasians), and in its future (admixing with populations from Africa, Europe, and Siberia in post-glacial, Neolithic, and later periods), Dzudzuana stands in the middle of an ongoing process of admixture of diverse elements from which West Eurasians, in all their diversity, eventually formed.

1. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216-219, (2015).
2. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118, (2014).
3. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200-205, (2016).
4. Sikora, M. *et al.* Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science*, (2017).
5. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912, (2015).
6. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419-424, (2016).
7. Clark, P. U. *et al.* The Last Glacial Maximum. *Science* **325**, 710, (2009).
8. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503, (2015).
9. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, (2014).
10. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87-91, (2014).
11. van de Loosdrecht, M. *et al.* Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science*, (2018).
12. Fregel, R. *et al.* Neolithization of North Africa involved the migration of people from both the Levant and Europe. *bioRxiv*, (2017).
13. Broushaki, F. *et al.* Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499-503, (2016).
14. Kılınç, Gülşah M. *et al.* The Demographic Development of the First Farmers in Anatolia. *Curr. Biol.* **26**, 2659–2666, (2016).
15. Bar-Yosef, O. *et al.* Dzudzuana: an Upper Palaeolithic cave site in the Caucasus foothills (Georgia). *Antiquity* **85**, 331-349, (2015).
16. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211, (2015).
17. Fregel, R. *et al.* Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proceedings of the National Academy of Sciences*, (2018).
18. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093, (2012).
19. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449, (2014).
20. Yang, M. A. *et al.* 40,000-Year-Old Individual from Asia Provides Insight into Early Population Structure in Eurasia. *Curr. Biol.* **27**, 3202-3208.e3209, (2017).

284 21. Busing, F. T. A., Meijer, E. & Leeden, R. Delete-m Jackknife for Unequal m. *Statistics and*
285   *Computing* **9**, 3-8, (1999).
286 22. Ferembach, D. Squelettes du Natoufien d'Israel., etude anthropologique. . *L'Anthropologie* **65**,
287   46-66, (1961).
288 23. Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains.
289   *Nature* **505**, 43-49, (2014).
290

Figure 1. Ancient West Eurasian population structure. (a) Geographical distribution of key ancient West Eurasian populations. (b) Temporal distribution of key ancient West Eurasian populations (approximate date in ky BP). (c) PCA of key ancient West Eurasians, including additional populations (shown with grey shells), in the space of outgroup f_4-statistics (Methods).
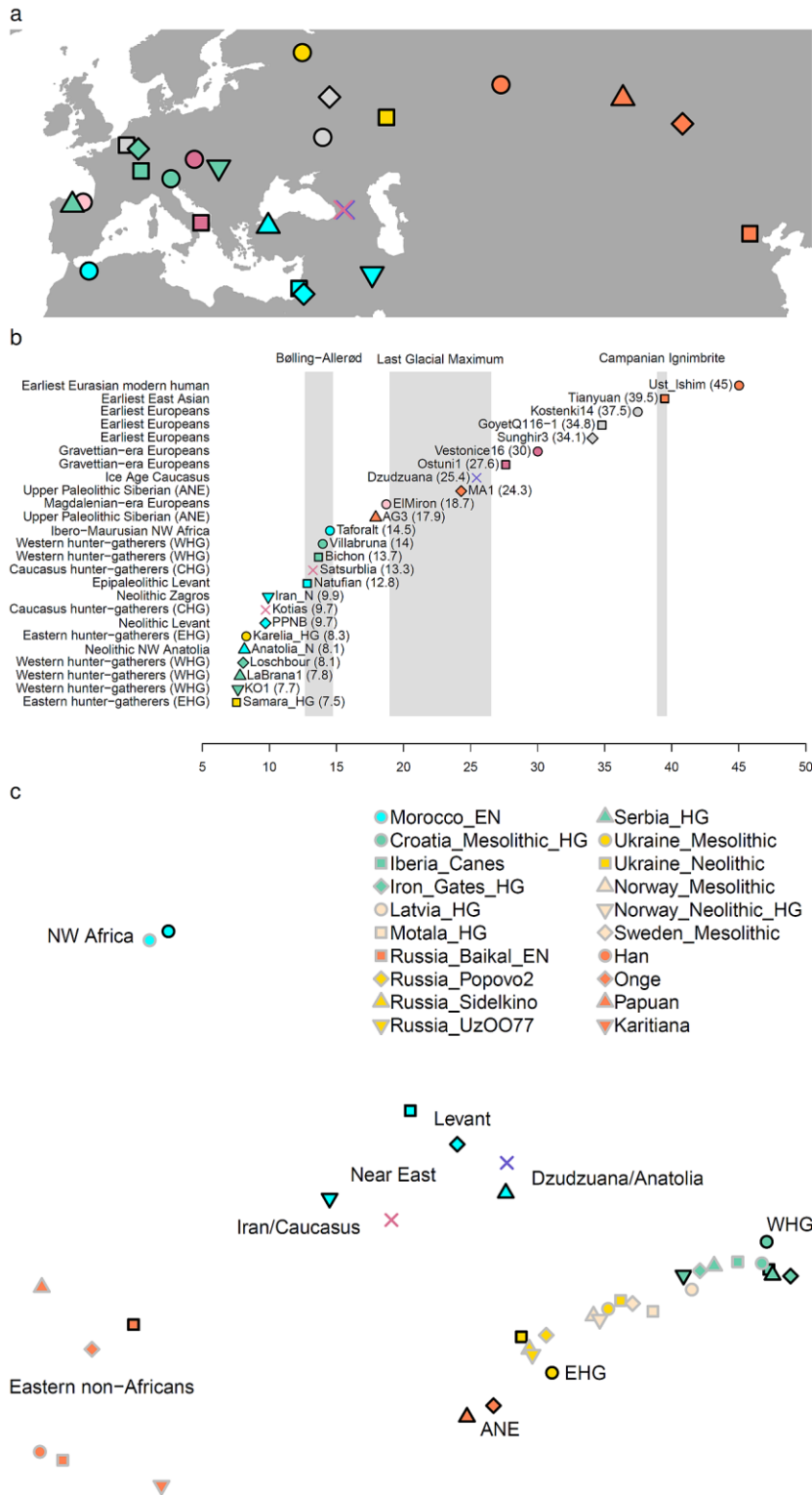
297 **Figure 2. An admixture graph model of Paleolithic West Eurasians.** An automatically
298 generated admixture graph models fits 14 populations (worst Z-score of the difference
299 between estimated and fitted *f*-statistics is 2.7) or 15 populations (also including
300 South_Africa_HG, worst Z-score is 3.5). This is a simplified model assuming binary
301 admixture events and is not a unique solution (Supplementary Information section 2).
302 Sampled populations are shown with ovals and select labeled internal nodes with rectangles.

303

304
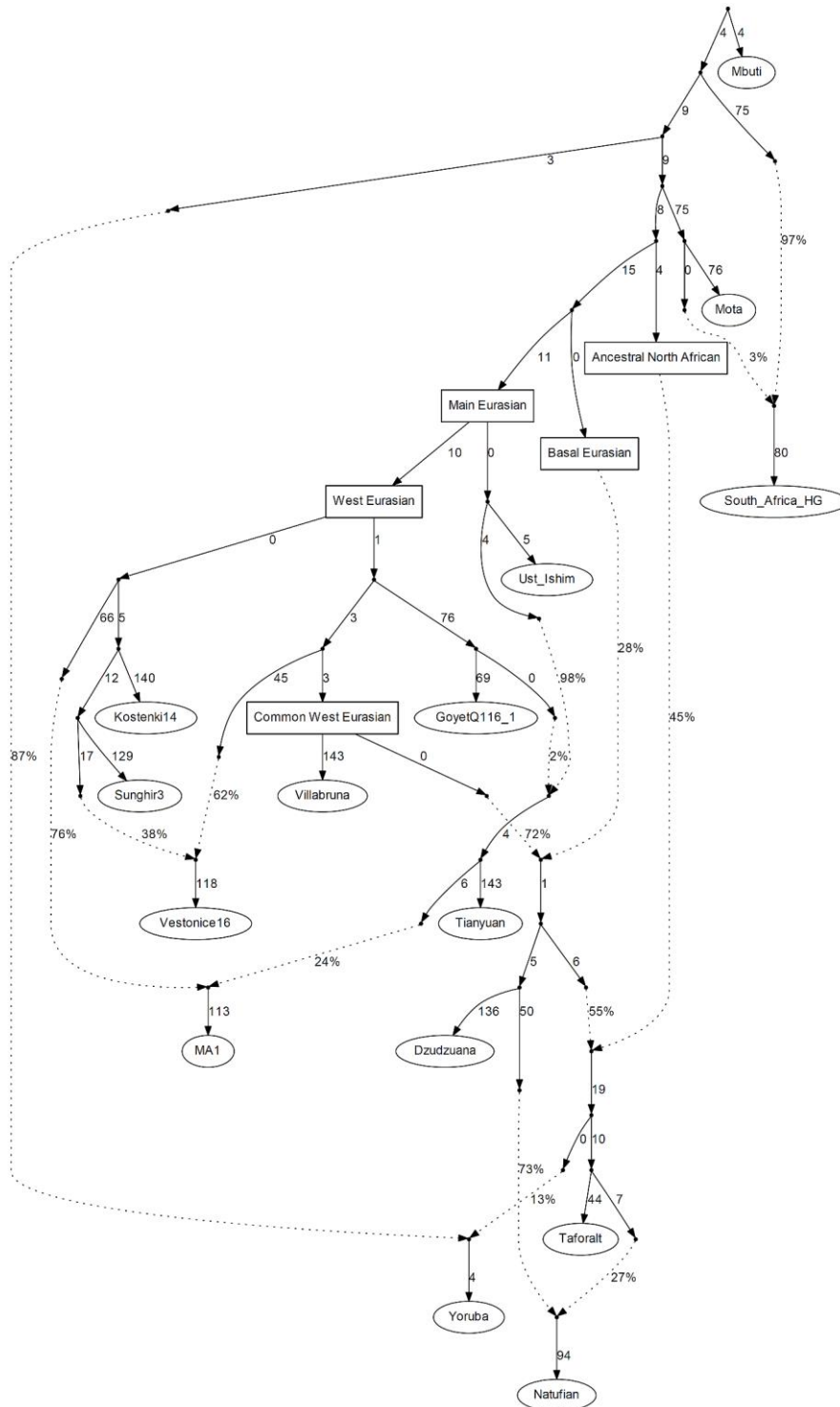
Figure 3: Modeling present-day West Eurasians and North Africans. Mixture proportions can be found in Extended Data Fig. 6 and Supplementary Information section 4. Ancestry with Mbuti as a source reflects all 'Deep' ancestry from lineages that split off prior to the 45kya Ust'Ishim. Jewish populations are shown with squares to visually distinguish them from geographically proximate populations.
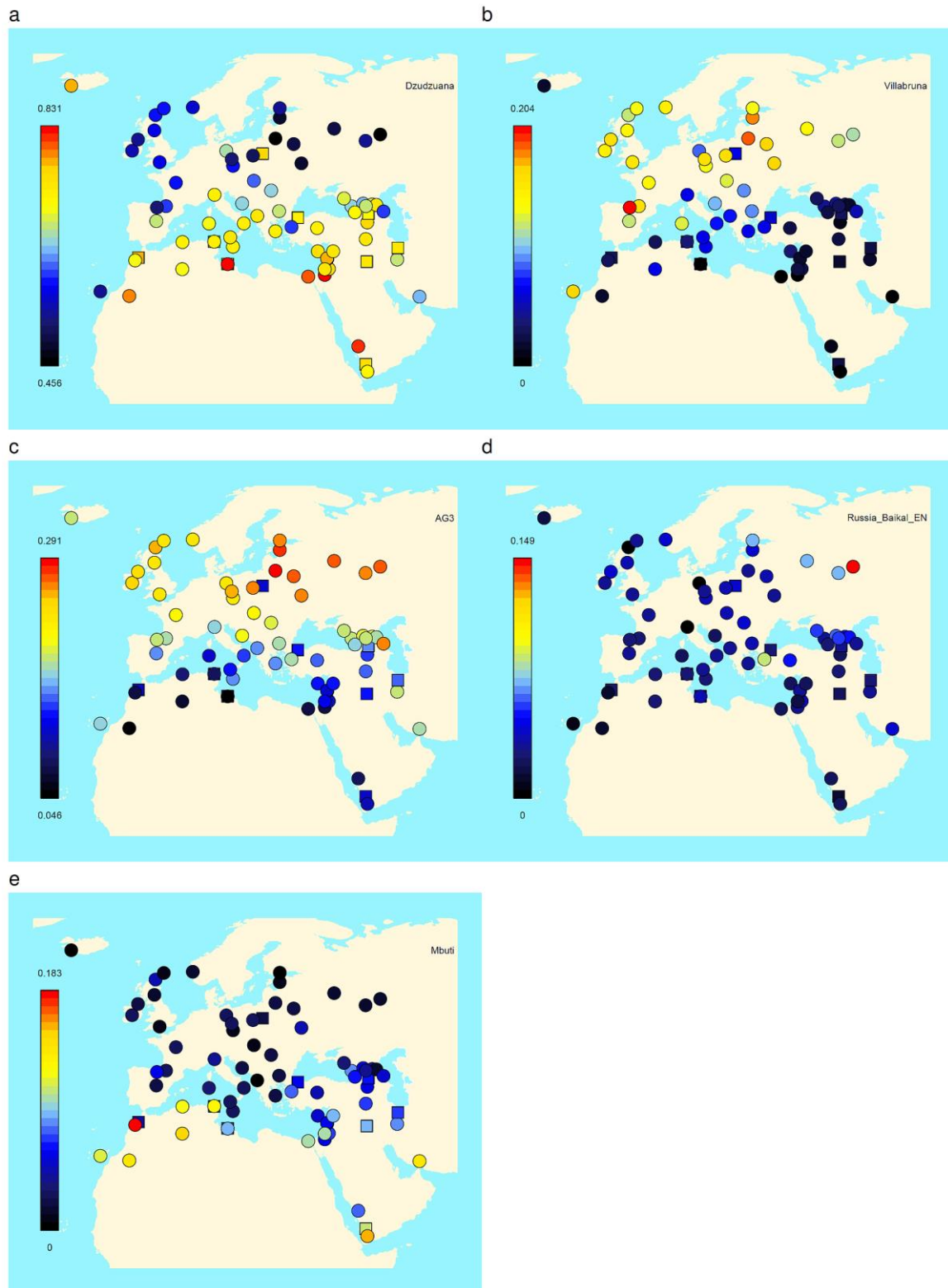
311 **Table 1: Mixture proportions of ancient populations.** The model with the fewest sources
312 for each population that fits the data is shown. Multiple models may fit some populations; we
313 show here the one with the highest P-value; others are shown in Supplementary Information
314 section 3. The P-value is (for *N* sources) for rank *N*-1 of (Test, Sources) with respect to a set
315 of diverse outgroups (Supplementary Information section 3). Neolithic Near Eastern
316 populations (four bottom rows) can also be fit as mixtures of Dzudzuana and Taforalt via
317 Natufians.

|						Mixture Proportions				Standard Errors			
Test	A	B	C	D	P-value	A	B	C	D	A	B	C	D
AG3	MA1				0.107	1.000							
Anatolia_N	Dzudzuana				0.286	1.000							
CHG	Mbuti	Tianyuan	AG3	Dzudzuana	0.685	0.054	0.081	0.222	0.643	0.040	0.028	0.031	0.041
Dzudzuana	Mbuti	Villabruna			0.274	0.275	0.725			0.037	0.037		
ElMiron	GoyetQ116-1	Villabruna			0.167	0.394	0.606			0.118	0.118		
Iran_N	Mbuti	Onge	AG3	Dzudzuana	0.129	0.097	0.109	0.218	0.577	0.047	0.025	0.026	0.041
Karelia_HG	Han	AG3	Villabruna		0.060	0.032	0.628	0.340		0.015	0.029	0.028	
Natufian	Mbuti	Dzudzuana			0.702	0.112	0.888			0.022	0.022		
PPNB	Mbuti	Dzudzuana			0.729	0.071	0.929			0.018	0.018		
Russia_Baikal_EN	Han	MA1			0.313	0.843	0.157			0.016	0.016		
Taforalt	Mbuti	Dzudzuana			0.556	0.272	0.728			0.024	0.024		
Morocco_EN	Taforalt				0.060	1.000							
Anatolia_N	Dzudzuana	Natufian			0.070	0.859	0.141			0.029	0.029		
Natufian	Dzudzuana	Taforalt			0.405	0.863	0.137			0.019	0.019		
PPNB	Dzudzuana	Natufian			0.910	0.409	0.591			0.042	0.042		

318

319 **Extended Data Figure 1: Dzudzuana Cave.** (a) Section drawing of Dzudzuana Cave.

320 Location of the two recovered teeth is indicated. (b) Calibrated dates for Layer C, upper pane:

321 new determinations (Supplementary Information section 1), lower pane: dates from ref.[24]

322 **a**
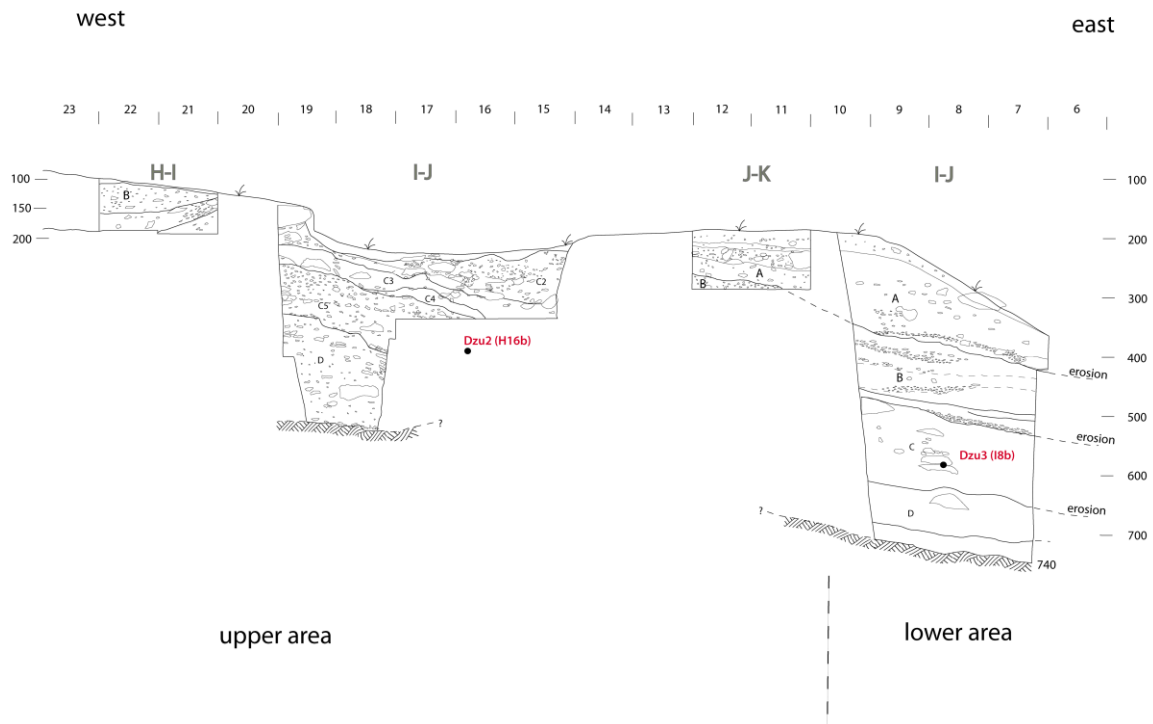
323
324 **b**

325
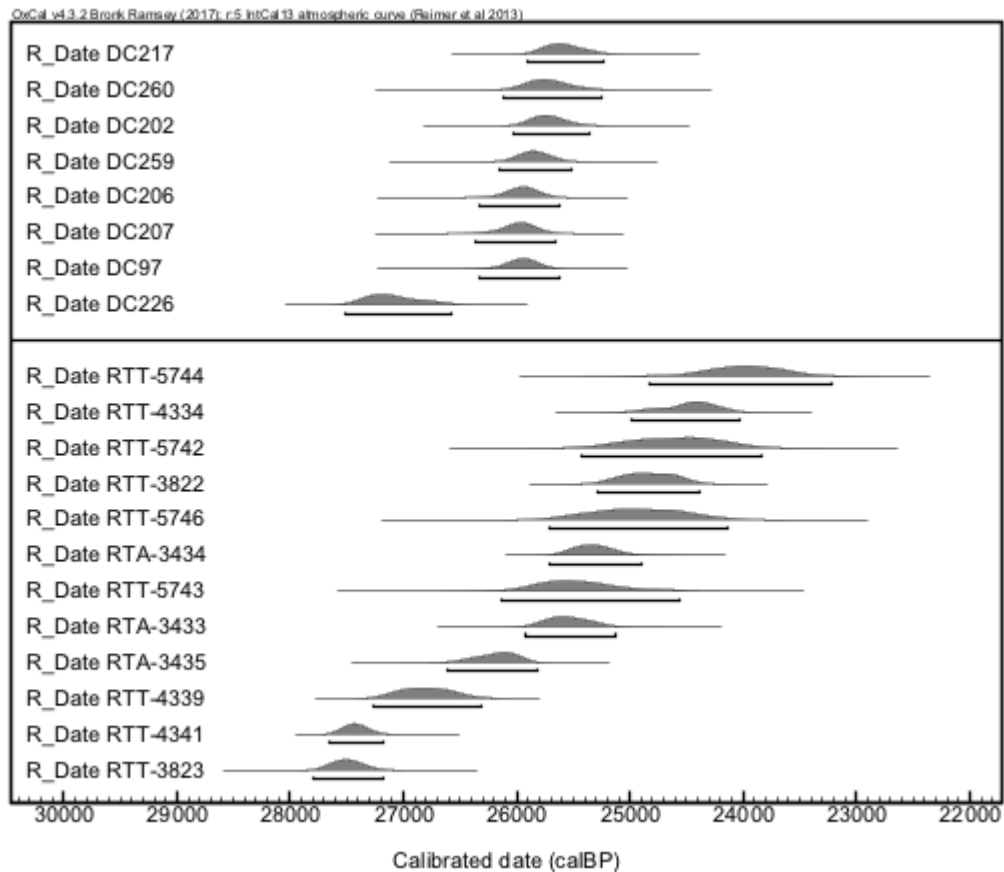
326 **Extended Data Figure 2: Symmetry testing and quality control.** (a) The two Dzudzuana
327 individuals are symmetrically related to others ($|Z| < 2.5$). (b) Individual I2949 shares more
328 alleles with I2963 than with others ($Z \in [-0.1, 11.6]$). (c) Individual I2963 shares more alleles
329 with I2949 than with others ($Z \in [0.1, 10.5]$). (d) Principal components analysis. We repeat the
330 analysis of Fig. 1c with the two Dzudzuana individuals shown separately. (e) Analysis of the
331 8 libraries separately. (f) Analysis of damage restricted sequences; point shapes and colors
332 not shown in legend are identical to Fig. 1.

333
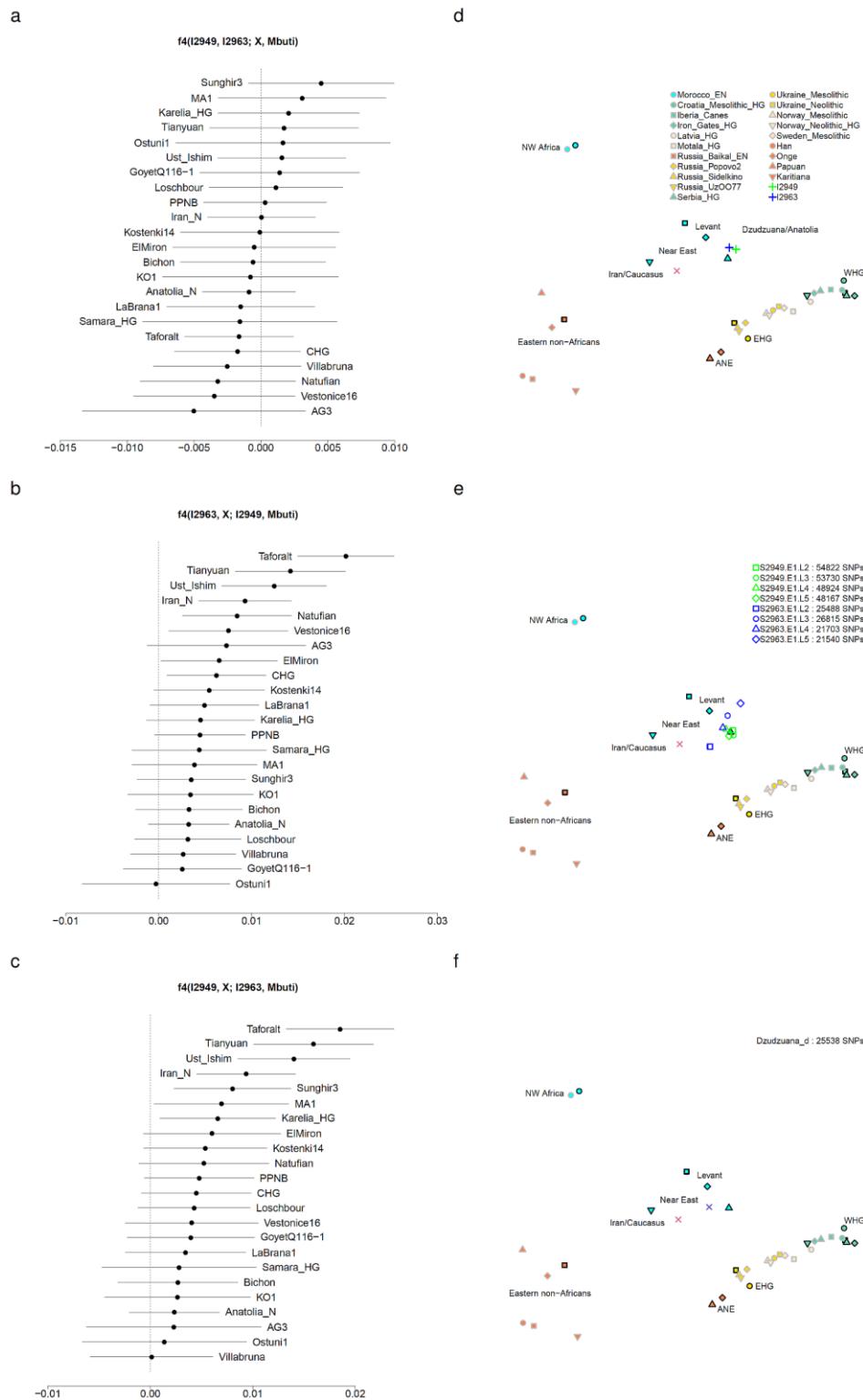
Extended Data Figure 3: Outgroup f_3-statistics of ancient West Eurasians. (a) We plot pairwise f_3(Mbuti; X, Y) for pairs of ancient West Eurasian populations, which measure the shared genetic drift between X and Y after their separation from an African outgroup (Mbuti pygmies). (b) The value of the statistic of panel (a) for X=Dzudzuana with ±3 standard errors. (c) We plot f_3(Ust'Ishim and Tianyuan; X, Y), which measures the shared genetic drift between X and Y after their separation from non-West Eurasians (made up of the Ust'Ishim and Tianyuan individuals of ~40-45kya age). The value of the statistic of panel (c) for X=Dzudzuana with ±3standard errors.
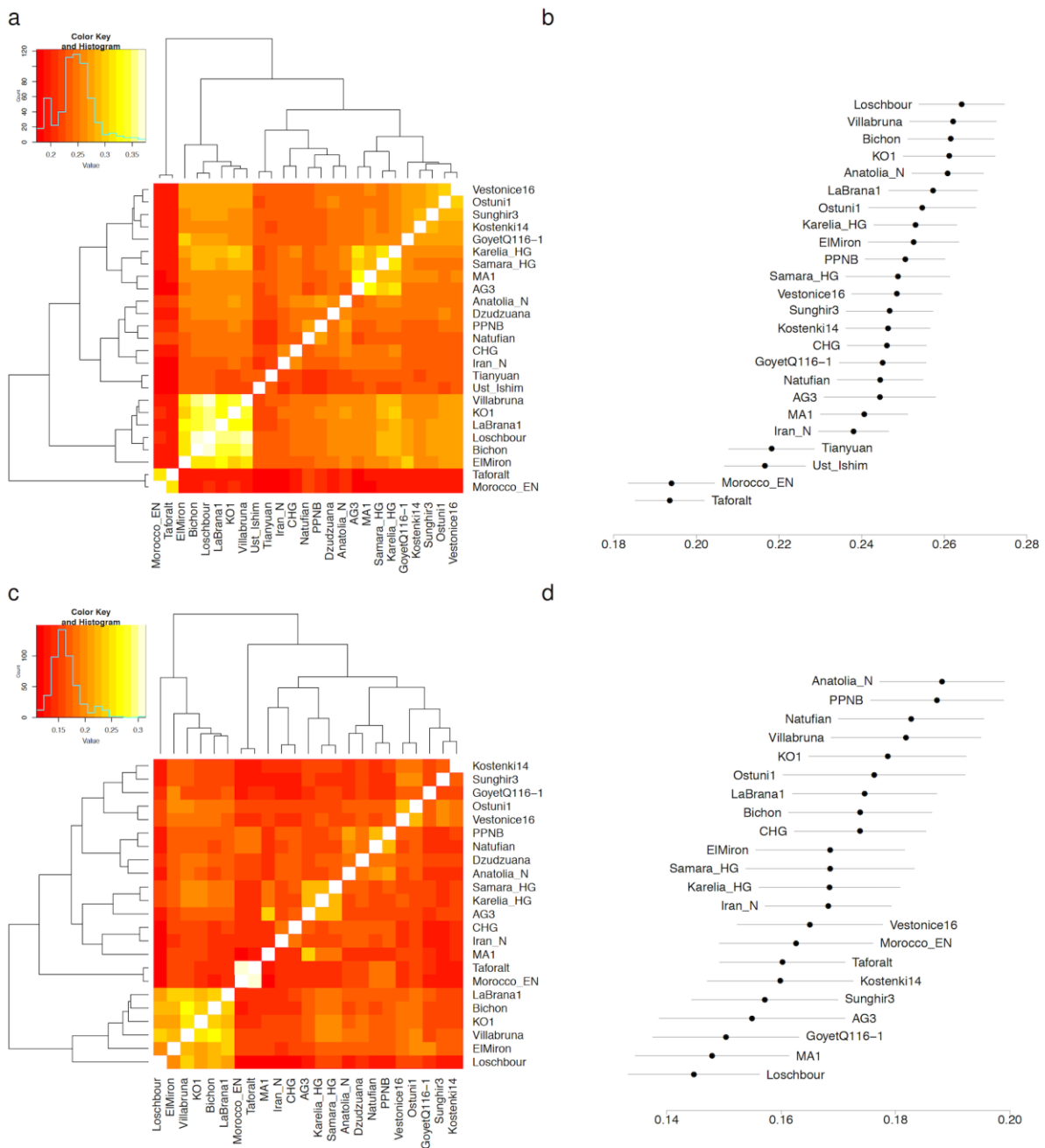
343 **Extended Data Figure 4: Principal components analysis of ancient and present-day**
344 **West Eurasians.** (a) View of all samples; present-day populations shown in white-filled
345 circles. This corresponds to Fig. 1c, but is computed on Human Origins data. (b) View of
346 present-day samples; the legend is split by the median PC1 value and shows a
347 correspondence between European (bottom-right) and Near Eastern (top-left) present-day
348 populations. Point shapes and colors not shown in legend are identical to Fig. 1.


349
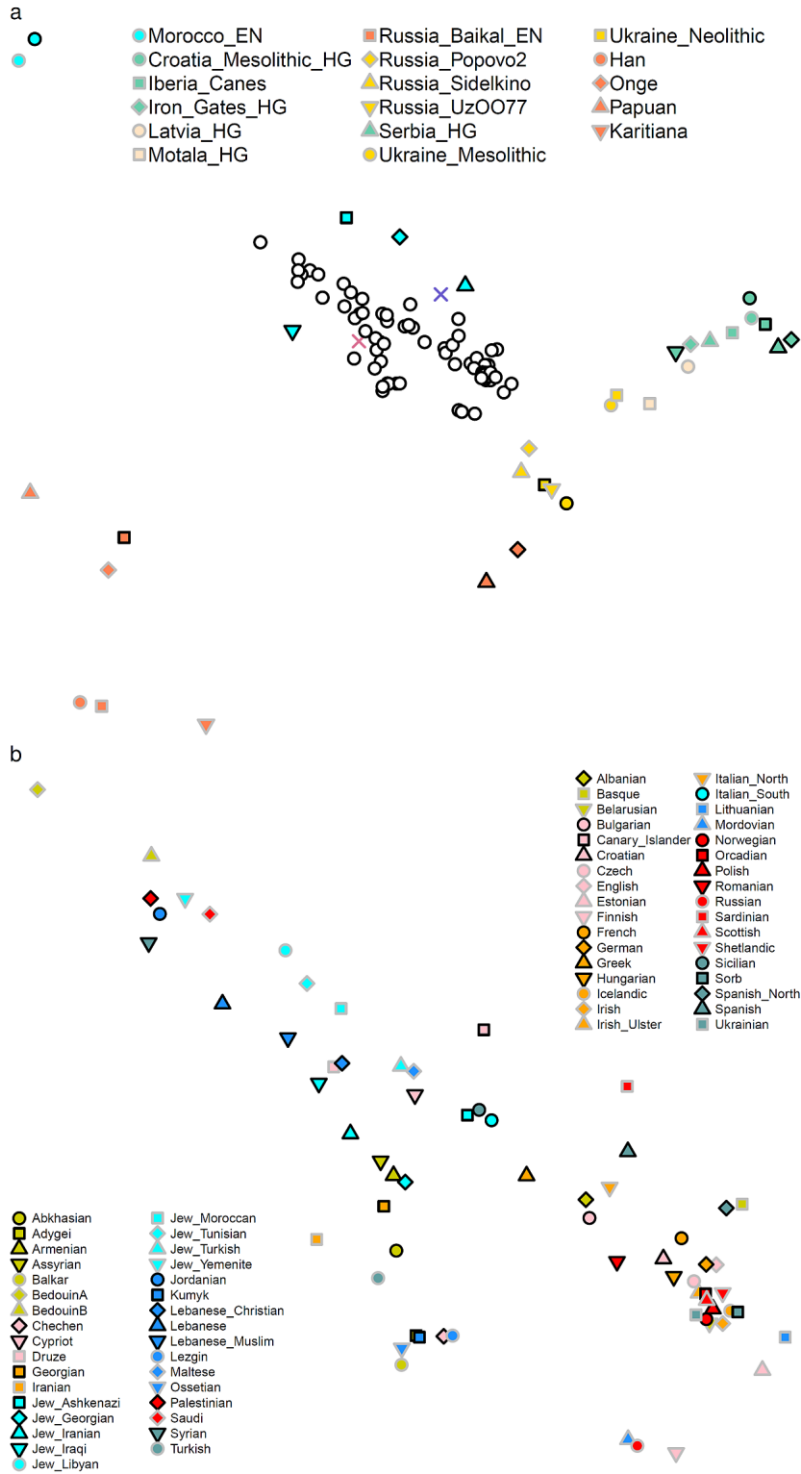
350 **Extended Data Figure 5: Symmetry testing.** In all panels we plot the pair (X, Y) as a label
351 for each statistic which is shown with ±3 standard errors (a) The statistic f_4(Dzudzuana, X, Y;
352 Mbuti) shows that diverse European hunter-gatherers Y share more alleles with Dzudzuana
353 than with other ancient Near Eastern populations except for Neolithic Anatolians. (b) The
354 statistic f_4(Dzudzuana, Mbuti; X, Y) shows that Dzudzuana shares more alleles with WHG
355 (both from the Upper Paleolithic and Mesolithic) than with other European hunter-gatherers.
356 (c) The statistic f_4(Dzudzuana, X; Y, Mbuti) shows that early Eurasians like Ust'Ishim,
357 Tianyuan and eastern non-Africans like Onge and Papuans share more alleles with European
358 hunter-gatherers than with Dzudzuana. (d) The statistic f_4(Dzudzuana, Anatolia_N; Y, Mbuti)
359 shows that Near Easterners Y share more alleles with Neolithic Anatolians than with
360 Dzudzuana.

361

362
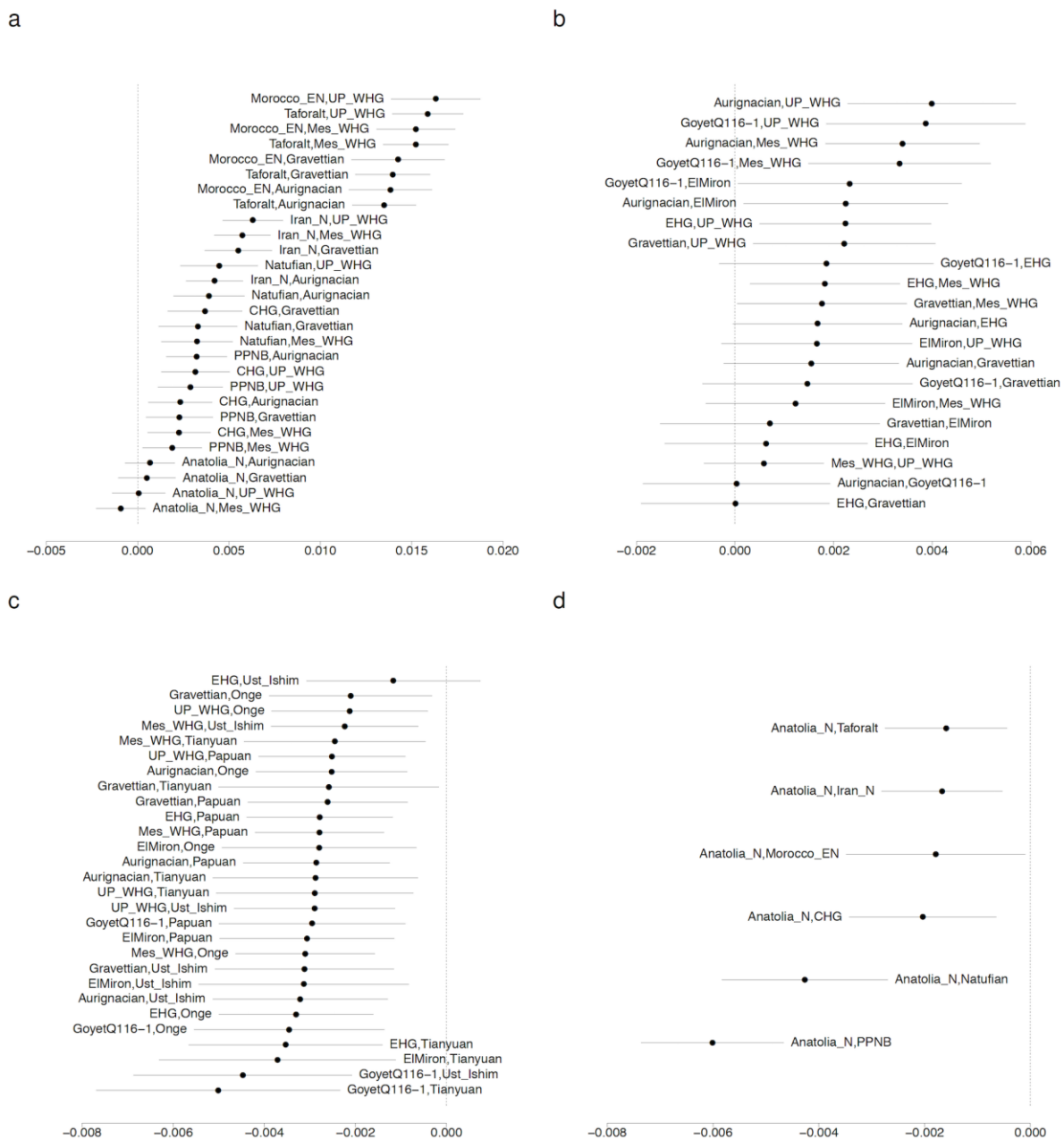
Extended Data Figure 6: Modeling present-day and ancient West-Eurasians. Mixture proportions computed with qpAdm (Supplementary Information section 4). The proportion of 'Mbuti' ancestry represents the total of 'Deep' ancestry from lineages that split prior to the split of Ust'Ishim, Tianyuan, and West Eurasians and can include both 'Basal Eurasian' and other (e.g., Sub-Saharan African) ancestry. (a) 'Conservative' estimates. Each population cannot be modeled with fewer admixture events than shown. (b) 'Speculative' estimates. The highest number of sources (≤5) with admixture estimates within [0,1] are shown for each population. Some of the admixture proportions are not significantly different from 0 (Supplementary Information section 4).
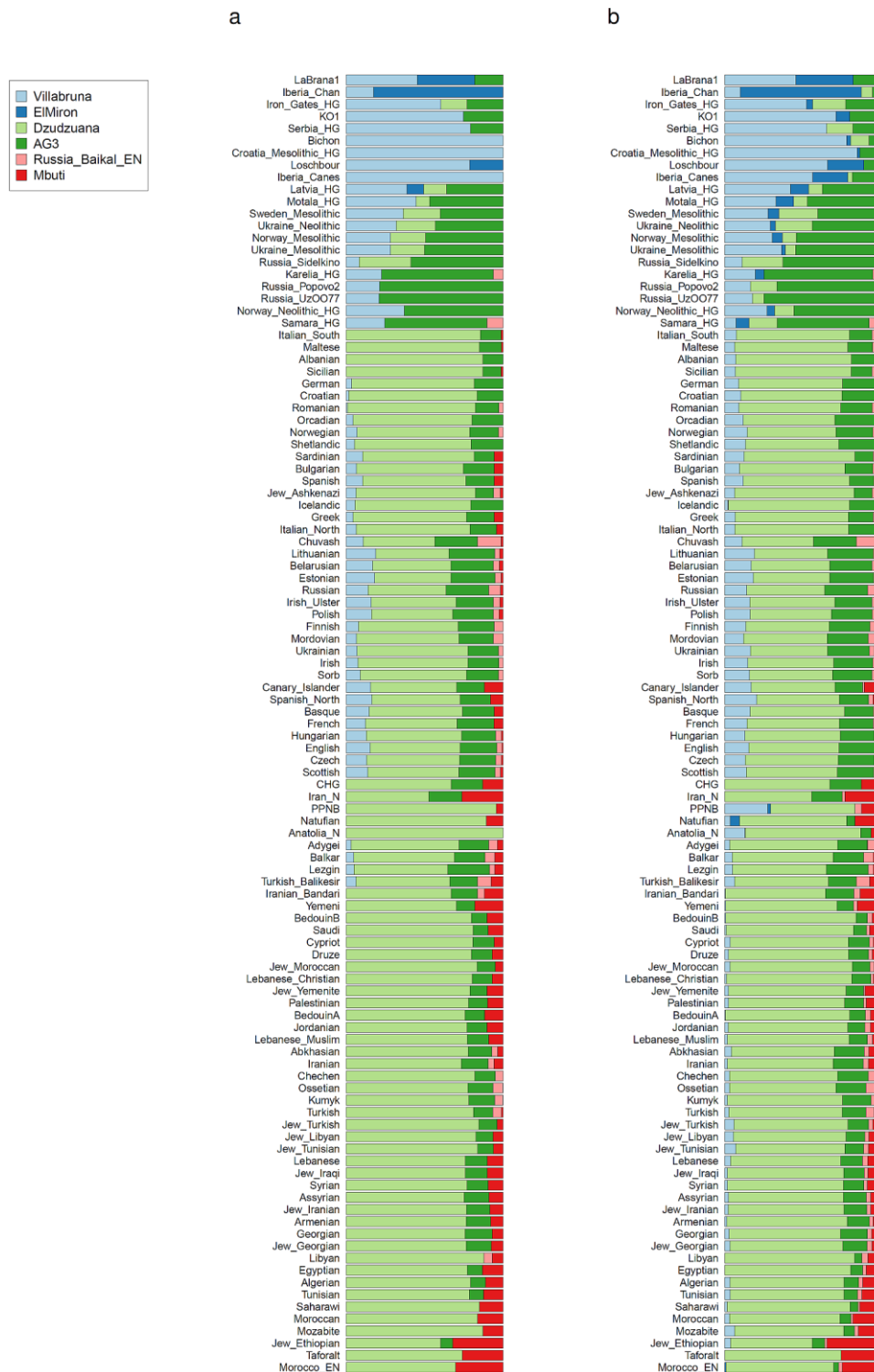
Extended Data Figure 7: Differential relationship of Basal ancestry to Africa. Basal ancestry (conservative estimate) is negatively correlated with the statistic f_4(X, Kostenki14, Ust'Ishim, Yoruba) which quantifies allele sharing between X and Ust'Ishim, consistent with this type of ancestry diluting the affinity of populations to this 45kya Siberian (earliest known modern human for which there are genomic data). For Taforalt and some populations from the Near East and North Africa this statistic is more negative, suggesting that they have North or Sub-Saharan-related ancestry that cannot be accounted for by any combination of the ancient West Eurasian sources whose convex hull is shown.
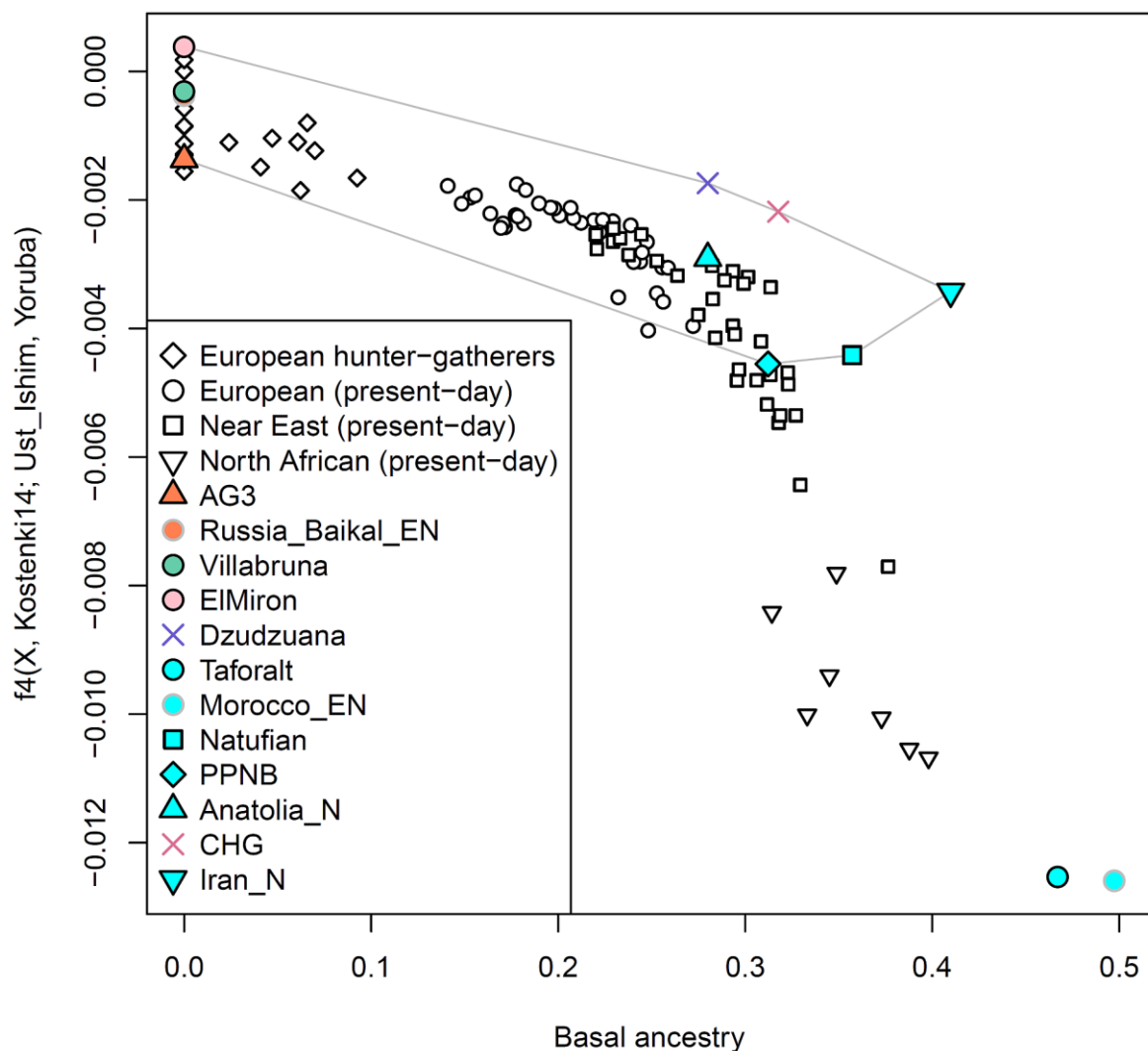
383 **Extended Data Figure 8: Deep ancestry of West Eurasians is associated with reduced**

384 **Neandertal ancestry.** Symbols used as in Fig. 1. Small grey-filled circles represent present-

385 day West Eurasian individuals sequenced in the Simons Genome Diversity Project. The three

386 left-most ancient samples and all the present-day populations are not used in the regression.

387 Standard errors of the linear regression line are computed with a block jackknife

388 (Supplementary Information section 5).

389

390
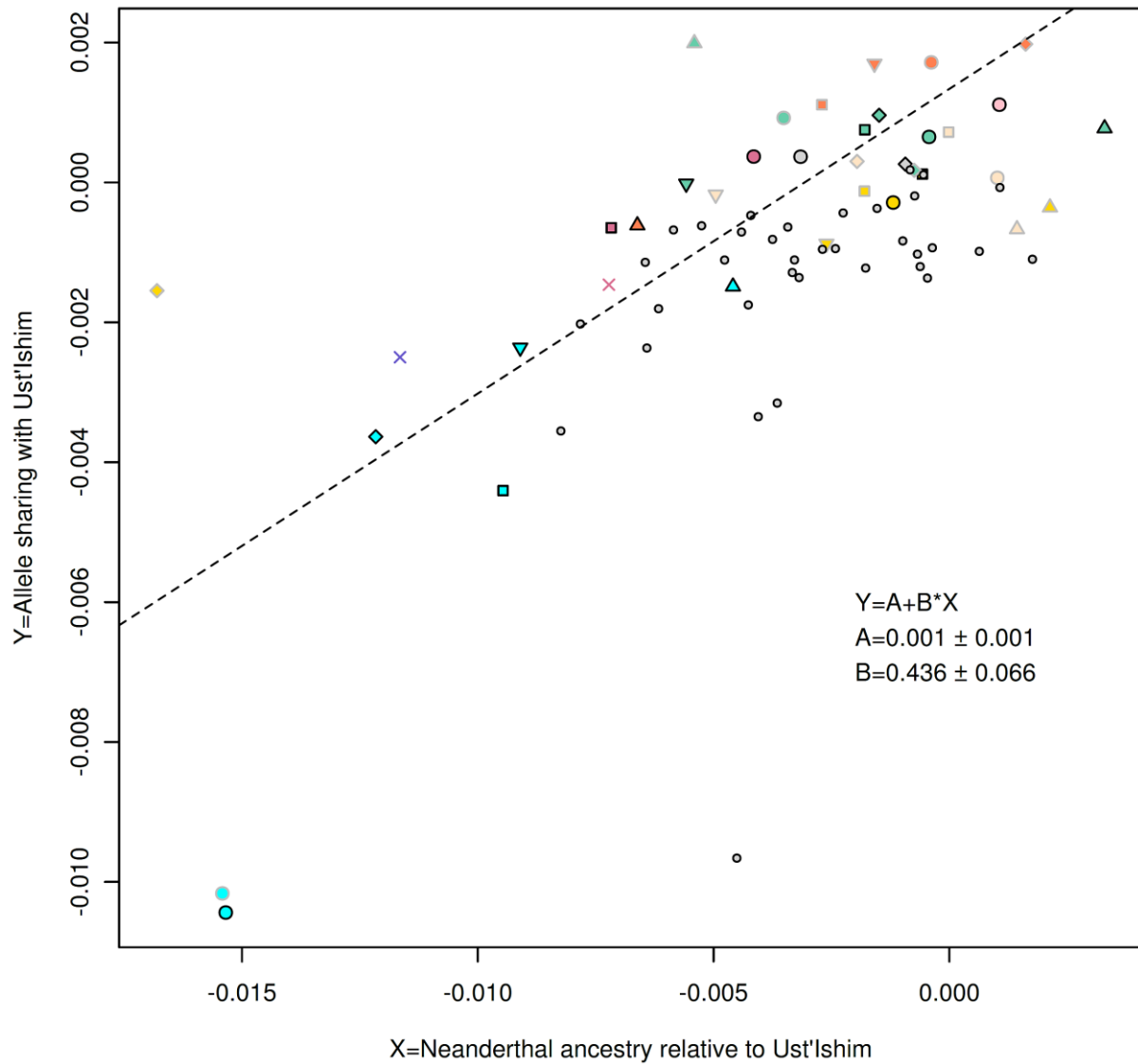
Extended Data Table 1. F_{ST} of ancient West Eurasian populations. The following groups were used as F_{ST} estimation requires at least two individuals per population. ANE=MA1[10] and AG3[3]. Aurignacian-era: Kostenki14 and Sunghir3[4]. CHG: Kotias and Satsurblia[5]. EHG: Karelia_HG and Samara_HG[8,16]. Gravettian-era: Vestonice16 and Ostuni1[3]. Mes_WHG: Loschbour[9], KO1[25], and LaBrana1[26]. UP_WHG: Villabruna[3] and Bichon[5]. The value of the statistic is shown below the diagonal and its standard error above it.

	Anatolia_N	ANE	Aurignacian	CHG	Dzudzuana	EHG	Gravettian	Iran_N	Mes_WHG	Natufian	PPNB	Taforalt	UP_WHG	Morocco_EN
Anatolia_N		0.003	0.002	0.002	0.005	0.002	0.002	0.001	0.002	0.002	0.001	0.002	0.003	0.003
ANE	0.119		0.004	0.004	0.015	0.004	0.005	0.003	0.003	0.005	0.004	0.003	0.004	0.007
Aurignacian	0.078	0.096		0.003	0.009	0.003	0.004	0.003	0.003	0.003	0.003	0.003	0.004	0.004
CHG	0.074	0.115	0.094		0.010	0.003	0.004	0.002	0.003	0.004	0.003	0.003	0.003	0.004
Dzudzuana	0.039	0.100	0.060	0.068		0.009	0.012	0.007	0.007	0.013	0.008	0.006	0.009	0.021
EHG	0.089	0.047	0.089	0.100	0.094		0.004	0.002	0.003	0.004	0.003	0.003	0.003	0.005
Gravettian	0.089	0.116	0.028	0.117	0.051	0.100		0.003	0.003	0.005	0.003	0.003	0.004	0.006
Iran_N	0.074	0.110	0.093	0.059	0.070	0.099	0.114		0.002	0.002	0.002	0.002	0.003	0.003
Mes_WHG	0.098	0.132	0.087	0.126	0.084	0.081	0.092	0.132		0.003	0.002	0.002	0.003	0.003
Natufian	0.076	0.164	0.121	0.131	0.106	0.142	0.131	0.126	0.148		0.003	0.003	0.004	0.007
PPNB	0.022	0.122	0.087	0.085	0.052	0.100	0.090	0.080	0.109	0.043		0.002	0.003	0.004
Taforalt	0.195	0.242	0.204	0.219	0.192	0.228	0.218	0.214	0.242	0.199	0.179		0.003	0.003
UP_WHG	0.112	0.149	0.100	0.139	0.097	0.096	0.100	0.144	0.001	0.160	0.125	0.252		0.004
Morocco_EN	0.206	0.254	0.219	0.230	0.195	0.236	0.234	0.225	0.252	0.218	0.190	0.101	0.267	

Methods

Ancient DNA

In a dedicated ancient DNA clean room facility at University College Dublin, we prepared powder from the roots of teeth, and then sent these powders to Harvard Medical School where in another clean room we extracted DNA[27,28] and prepared double-stranded libraries[29,30] which we enriched for sequenced overlapping the mitochondrial genome and 1.24 million SNPs in the nuclear genome[1,16,31] (Methods). We obtained usable data for two individuals after quality control and merging of data from 4 libraries for each of the two samples (treated with the enzyme UDG to reduce ancient DNA errors). The sequences from all libraries displayed characteristic damage in the terminal nucleotides consistent with partial UDG treatment.

Contamination testing

We assessed contamination by examining heterozygosity on mitochondrial DNA using contamMix[19] and schmutzi[32] (Supplementary Data Table 1).

Datasets

Our main analysis dataset included 1,233,013 SNPs, of which the 1,150,639 ones on the autosomes were analyzed. Present-day populations from the Simons Genome Diversity Panel[33] were included in this dataset. Analyses of Supplementary Information section 4 that included present-day populations[6,9,13,18] genotyped on the Affymetrix Human Origins array were performed on a dataset of 597,573 SNPs (593,124 on the autosomes).

Estimation of F_{ST}

We estimated F_{ST} in smartpca[34] using the default parameters and inbreed: YES[35] and fstonly: YES.

f-statistics

All *f*-statistics were computed in ADMIXTOOLS[18] using the programs qpDstat (with parameter f4mode: YES) and qp3Pop and default parameters.

Principal component analysis on outgroup f_4-statistics

426 We computed for n populations $Test_1$, $Test_2$,…,$Test_n$ and n_o outgroups, a matrix of $n \times d$

427 dimensionality where $d = n_o \binom{n_o - 1}{2}$ of outgroup f_4-statistics[16] of the form $f_4(Test_i, O_1; O_2,$

428 $O_3)$. The outgroup set was: Vindija[36], Altai[23], Denisova[37], Mbuti[33], South_Africa_HG[38],

429 Mota[39], Yoruba[33], Han[33], Onge[33], Papuan[33], Karitiana[33], Ust_Ishim[19], Tianyuan[20], MA1[10],

430 Kostenki14[3], GoyetQ116-1[3], Sunghir3[4], Vestonice16[3], Ostuni1[3], ElMiron[3], Dzudzuana. PCA

431 on the matrix was performed using the R package ppca[40,41] which allows for missing data,

432 thus allowing us to also plot populations that are included in the outgroup set. ppca was run

433 with parameters nPcs=$\lfloor\sqrt{n}\rfloor$ (the number of principal components used to fill in missing

434 values), and seed=123. Results are shown in Fig. 1c and Extended Data Fig. 4.

Admixture graph fitting

436 We used ADMIXTUREGRAPH program qpGraph with parameters outpop: NULL,

437 blgsize: 0.05, forcezmode: YES, lsqmode: YES, diag: .0001, bigiter: 6,

438 hires: YES, allsnps: YES, lambdascale: 1. We used the hash value computed by

439 qpreroot to avoid evaluating equivalent graphs. We fit graphs in an automated way described

440 in Supplementary Information section 2.

qpWave/qpAdm analyses

442 We used qpWave[16,35,42] and qpAdm[16] to test for the number of streams of ancestry from a set

443 of 'right' populations to a set of 'left' ones which includes the population whose history of

444 admixture we study, and to also estimate mixture proportions (Supplementary Information

445 section 3).

Acknowledgments

Author Contributions

451 I.L. analyzed data with input from D.R. I.L., A.B-C., R.P., D.R. wrote the manuscript with

452 input from other co-authors. A.B.-C., G.B.-O., O.B.-Y., N.J., E.K., D.L., Z.M. T.M.

453 undertook archaeological work in Dzudzuana cave; G.B.-O. studied the archaeozoology of

454 the cave and E.K. its palynology. S.M. performed bioinformatics. N.P. aided in the admixture

455   graph analysis. O.C. and N.R. performed ancient DNA work. B.J.C. and D.J.K. performed

456   [14]C dating. R.P. and D.R. co-ordinated the work.

Author Information

458   Correspondence and requests for materials should be addressed to I.L.

459   (lazaridis@genetics.med.harvard.edu) or R.P. (ron.pinhasi@univie.ac.at), or D.R.

460   (reich@genetics.med.harvard.edu)

Data and Code availability

462   The aligned sequences are available through the European Nucleotide Archive under

463   accession number xxx. Genotype datasets used in analysis are available at

464   https://reich.hms.harvard.edu/datasets. Code used to implement automated qpGraph fitting

465   can be obtained from I.L. by request. All other data are available from the corresponding

466   authors upon reasonable request.

References

469   24.   Bar-Yosef, O. *et al.* Dzudzuana: an Upper Palaeolithic cave site in the Caucasus foothills
470   (Georgia). *Antiquity* **85**, 331-349, (2011).
471   25.   Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory.
472   *Nat. Commun.* **5**, 5257 (2014).
473   26.   Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old
474   Mesolithic European. *Nature* **507**, 225-228, (2014).
475   27.   Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear
476   reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of*
477   *Sciences* **110**, 15758–15763, (2013).
478   28.   Korlevic, P. *et al.* Reducing microbial and human contamination in DNA extractions from
479   ancient bones and teeth. *BioTechniques* **59**, 87-93, (2015).
480   29.   Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed
481   target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb prot5448, (2010).
482   30.   Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil–DNA–
483   glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal*
484   *Society B: Biological Sciences* **370**, (2015).
485   31.   Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc.*
486   *Natl. Acad. Sci. USA* **110**, 2223–2227, (2013).
487   32.   Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and
488   endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224, (2015).
489   33.   Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse
490   populations. *Nature* **538**, 201-206, (2016).
491   34.   Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**,
492   e190, (2006).
493   35.   Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian
494   population history. *Nature* **461**, 489-494, (2009).
495   36.   Prüfer, K. *et al.* A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*,
496   (2017).

497   37.   Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan Individual.
498   *Science* **338**, 222-226, (2012).
499   38.   Skoglund, P. *et al.* Reconstructing Prehistoric African Population Structure. *Cell* **171**, 59-
500   71.e21, (2017).
501   39.   Llorente, M. G. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admixture in
502   Eastern Africa. *Science* **350**, 820-822, (2015).
503   40.   Porta, J. M., Verbeek, J. J. & Kröse, B. J. A. Active Appearance-Based Robot Localization
504   Using Stereo Vision. *Autonomous Robots* **18**, 59-80, (2005).
505   41.   Roweis, S. T. in *Advances in neural information processing systems.* 626-632.
506   42.   Moorjani, P. *et al.* Genetic evidence for recent population mixture in India. *Am. J. Hum.*
507   *Genet.* **93**, 422-438, (2013).

508

509