Neolithization of North Africa involved the migration of people from both the Levant and Europe

Rosa Fregel[1*], Fernando L. Méndez[1], Youssef Bokbot[2], Dimas Martín-Socas[3], María D. Camalich-Massieu[3], María C. Ávila-Arcos[1,4], Peter A. Underhill[1], Beth Shapiro[5], Genevieve Wojcik[1], Morten Rasmussen[1], Andre E. R. Soares[5§], Joshua Kapp[5], Alexandra Sockell[1], Francisco J. Rodríguez-Santos[6], Abdeslam Mikdad[2], Jonathan Santana[7], Aioze Trujillo-Mederos[3] & Carlos D. Bustamante[1¶]

[1]Department of Genetics, School of Medicine, Stanford University, Stanford, US. [2]Institut National des Sciences de l'Archéologie et du Patrimoine, Rabat, Morocco. [3]Department of Prehistory, Faculty of History, Universidad de La Laguna, San Cristóbal de La Laguna, Spain. [4]International Laboratory for Human Genome Research, National University of Mexico, Querétaro, Mexico. [5]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, US. [6]Instituto Internacional de Investigaciones Prehistóricas de Cantabria, Cantabria, Spain. [7]Department of Archaeology, Faculty of Arts and Humanities. Durham University, Durham, UK.

[¶]Present address: Department of Biomedical Data Science, School of Medicine, Stanford University, Stanford, US.

[§]Present address: Laboratório Nacional de Computação Científica, Petrópolis, Brazil

One of the greatest transitions in the human story was the change from hunter-gatherer to farmer. How farming traditions expanded from their birthplace in the Fertile Crescent has always been a matter of contention. Two models were proposed, one involving the movement of people and the other based on the transmission of ideas. Over the last decade, paleogenomics has been instrumental in settling long-disputed archaeological questions[1], including those surrounding the Neolithic revolution[2]. Compared to the extensive genetic work done on Europe and the Near East, the Neolithic transition in North Africa, including the Maghreb, remains largely uncharacterized. Archaeological evidence suggests this process may have happened through an *in situ* development from Epipaleolithic communities[3,4], or by demic diffusion from the Eastern Mediterranean shores[5] or Iberia[6]. In fact,

Neolithic pottery in North Africa strongly resembles that of European cultures like Cardial and Andalusian Early Neolithic, the southern-most early farmer culture from Iberia. Here, we present the first analysis of individuals' genome sequences from early and late Neolithic sites in Morocco, as well as Andalusian Early Neolithic individuals. We show that Early Neolithic Moroccans are distinct from any other reported ancient individuals and possess an endemic element retained in present-day Maghrebi populations, indicating long-term genetic continuity in the region. Among ancient populations, early Neolithic Moroccans share affinities with Levantine Natufian hunter-gatherers (~9,000 BCE) and Pre-Pottery Neolithic farmers (~6,500 BCE). Late Neolithic (~3,000 BCE) Moroccan remains, in comparison, share an Iberian component of a prominent European-wide demic expansion, supporting theories of trans-Gibraltar gene flow. Finally, the Andalusian Early Neolithic samples share the same genetic composition as the Cardial Mediterranean Neolithic culture that reached Iberia ~5,500 BCE. The cultural and genetic similarities of the Iberian Neolithic cultures with that of North African Neolithic sites further reinforce the model of an Iberian intrusion into the Maghreb.

Present genetic data from modern samples[7-9] suggests that North Africans' ancestry has contributions from four main sources: 1) an autochthonous Maghrebi component related to a back migration to Africa ~12,000 years ago from the Levant; 2) a Middle Eastern component probably associated with the Arab conquest; 3) a sub-Saharan component derived from trans-Saharan migrations; and 4) a European component that has been linked to recent historic movements. Paleogenomic studies have begun to provide insights into North African Prehistory[10,11]; however, no research to date has tested whether the Neolithic transition in the Maghreb was driven by an *in situ* development or the migration of people. Here, we perform genome-wide analysis of remains from the Early Neolithic site of Ifri n'Amr or Moussa (IAM; ~5,000 BCE, n=7) and the Late Neolithic site of Kelif el Boroud (KEB; ~3,000 BCE; n=8) (Supplementary Note 1). To test possible migrations through the Strait of Gibraltar, we also analyse samples from Early Neolithic sites in southern Iberia: El Toro (TOR; ~5,000 BCE; n=12) and Los Botijos (BOT; n=1) (Figure 1). This Andalusian Early Neolithic culture is thought to have arrived prior to Cardial technology, and bears similarities with early Maghrebi

61 farming traditions[12] (Supplementary Note 1). Including these southern samples in our
62 analysis enables a direct test of this hypothesis.

63 [Figure 1]

64 We sequenced 39 Illumina pair-end libraries from 28 individuals, and selected best
65 conserved libraries for subsequent analyses. Endogenous DNA content was generally low
66 (2.06% on average) (Supplementary Note 2). Depth of coverage was consistently
67 improved when enriching using baits targeting specific sites (~100X), compared to
68 whole-genome capture (~15X) (Supplementary Note 2). Following enrichment, we
69 generated fifteen low-coverage genomes (5 from IAM, 4 from KEB and 5 from
70 TOR/BOT), with coverage ranging from 0.04X to 0.73X depth. All samples considered
71 in this study met the standard aDNA authentication criteria, including observation of
72 DNA fragmentation (∼46 bp average read length) and damage patterns due to cytosine
73 deamination toward the 5′ ends of molecules (Supplementary Note 3).

74 Mitochondrial DNA and Y-chromosome haplogroups obtained for IAM (Moroccan
75 Early Neolithic) and KEB (Moroccan Late Neolithic) indicate either a population
76 replacement or an important genetic influx into Morocco between 5,000−3,000 BCE.
77 IAM samples belong to the mtDNA haplogroups U6a and M1—both of which are
78 associated with back migration to Africa[13,14]—while KEB samples belong to haplogroups
79 K1, T2 and X2, prominently found in Anatolian and European Neolithic samples[2,15]
80 (Supplementary Note 4). Regarding the paternal lineages, IAM individuals carry Y
81 chromosomes distantly related to the typically North African E-M81 haplogroup, while
82 the Y chromosome from KEB belongs to the T-M184 haplogroup; though scarce and
83 broadly distributed today, this haplogroup has also been observed in European Neolithic
84 individuals[16] (Supplementary Note 5). Both mtDNA and Y-chromosome lineages (K1, J2
85 and T2 haplogroups, and G-M201 haplogroup, respectively) for samples from TOR/BOT
86 (Iberian Early Neolithic) are similar to those observed in Europe during Neolithic times[15].

87 West Eurasian populations can be modelled as admixture of four different ancestral
88 components[2]: Eastern and Western European hunter-gatherers, Iranian and Levant
89 Neolithic. We explored the placement of Moroccan and Southern Iberian Neolithic

samples in this context, and compared their genetic affinities to ancient and present-day West Eurasian and Levant populations. Interestingly, Principal Components Analysis (PCA) reveals that IAM individuals are different from any aDNA sample studied to date (Figure 2; Supplementary Note 6). When projected, IAM samples are close to modern North Africans, in the Levantine corner of the PCA space (Figure 2). Southern Iberian Neolithic individuals from TOR and BOT cluster with Sardinians and with other Anatolian and European Neolithic samples. Moreover, KEB samples are placed halfway between the IAM and Anatolian/European farmer clusters, in close proximity to Levant aDNA samples. We further explored the genetic structure of these samples using the program ADMIXTURE[16] (See Supplementary Note 7 for details), with values of K ranging between 2 and 8. At lower K values, IAM samples possess ~100% of a component partially shared by aDNA samples from the Middle East and Levant. At K=6, this IAM-like component is observed mainly in modern North Africa, following a west-to-east cline. TOR and other Early Neolithic samples from Iberia cluster together with farmers from Anatolia, the Aegean area and Europe. At K=8 the Early Neolithic individuals from Iberia differentiate from the Anatolian, Aegean and European Early Neolithic samples, and share their main component (purple) with Middle Neolithic/Chalcolitic samples (Figure 2). Finally, at low K values, KEB can be explained as having both IAM-like and European Neolithic components, suggesting an admixture process between IAM-like people and early farmers. Nevertheless, at K=8, the European component in KEB is predominantly "purple," with some "green" component. This "green" component is also present, at a low frequency, in Natufians and other ancient Levantine populations. The substantially larger contribution of the "purple" component, when compared with the "green", suggests a significant genetic contribution of ancient Iberians in Morocco (Figure 2).

[Figure 2 here]

To compare our samples directly to the genomes of ancient and modern populations, we calculated pair-wise F_{ST} distances, which, unlike PCA and global ancestry analyses, are insensitive to the inclusion of large numbers of individuals from modern populations. F_{ST} values indicate that the IAM samples are as differentiated from all other populations as Yoruba are from non-Africans (Supplementary Note 9), with the sole exception of KEB

121 and, to a lesser extent, modern North African populations. Given the relatively low
122 heterozygosity (~6.6 x 10^{-4}) and high identity-by-descent (~37%) proportions observed in
123 IAM (Supplementary Note 8), this differentiation could be related to isolation and genetic
124 drift. Although IAM is clearly more similar to KEB than any other population, the
125 converse is not true. KEB has lower F_{ST} distances with any Anatolian, European
126 (excluding European hunter gatherers), Levantine and Iranian population, rather than
127 with IAM. In the modern DNA reference panel, KEB is similar to North African,
128 European and Middle Eastern populations. Among the ancient populations, TOR is more
129 similar to Middle Neolithic/Chalcolithic Europeans, and, among modern populations, to
130 populations from Spain, North Italy and Sardinia.

131 To further investigate the genetic affinities of IAM, KEB and TOR samples, we
132 conducted outgroup f3-statistic analysis[17]. Results indicate that, when KEB is excluded,
133 IAM shares more drift with ancient Levantine populations, such as Natufians and Levant
134 Neolithic (Figure 3; Supplementary Note 10), than with any other ancient population.
135 This result is analogous to preliminary archaeological data pointing to links between
136 funerary practices in IAM and Pre-Pottery Neolithic B sites in Cyprus (Supplementary
137 Note 1).

138 Both F_{ST} and outgroup-f3 statistic analyses indicate that KEB shares ancestry with
139 IAM, but also more genetic drift with Neolithic and Chalcolithic populations from
140 Anatolia and Europe, with the highest shared genetic drift appearing in Iberian Early
141 Neolithic samples (Figure 3; Supplementary Note 10). This pattern and the result from
142 ADMIXTURE could be explained if the KEB population was a mixture between IAM-
143 related and European Neolithic groups. To formally test this hypothesis, we used an
144 admixture-f3 test[17], using KEB as the test population, IAM as a reference population and
145 one of the Anatolian and European Neolithic and Chalcolithic populations as the second
146 reference population. All comparisons produced negative values of the f3-statistic, which
147 suggests the KEB population can be modelled as a mixture of IAM and
148 Anatolian/European Neolithic. These results also parallel archaeological findings in the
149 region: Late Neolithic sites in North Africa contain pottery resembling that of the
150 Andalusian Early Neolithic and Cardial cultures, and ivory tools distinctly associated
151 with those of Iberian Neolithic sites (Supplementary Note 1).

152 TOR has more shared ancestry with Iberian Early Neolithic samples and other
153 Neolithic and Chalcolithic populations from Europe. Archaeological work in southern
154 Iberia, especially in the Nerja site, has pointed out that the Andalusian Early Neolithic
155 culture, previous to the Cardial expansion, may have had connections to farmer traditions
156 in the Maghreb[18]. However, we observe that TOR samples have a similar genetic
157 composition to that of Cardial individuals from Iberia, evidencing a common origin.

158 [Figure 3 here]

159 Finally, phenotypic predictions based on genetic variants of known effects agree with our
160 estimates of global ancestry. IAM people do not possess any of the European SNPs
161 associated with light pigmentation, and most likely had dark skin and eyes. IAM samples
162 present ancestral alleles for pigmentation-associated variants present in SLC24A5
163 (rs1426654), SLC45A2 (rs16891982) and OCA2 (rs16891982 and 12913832) genes. On
164 the other hand, KEB individuals exhibit some European- derived alleles that predispose
165 individuals to lighter skin and eye colour, including those on genes SLC24A5
166 (rs1426654) and OCA2 (rs16891982) (Supplementary Note 11).

167 Genetic analyses have revealed that the population history of modern North
168 Africans is quite complex[7]. Based on our aDNA analysis, we identify an Early Neolithic
169 Moroccan component that is restricted to North Africa in present-day populations[7], which
170 is the sole ancestry in IAM samples. We hypothesize that this component represents the
171 autochthonous Maghrebi ancestry associated with Berber populations. This Maghrebi
172 component was related to that of Epipaleolithic and Pre-Pottery Neolithic people from
173 the Levant. By 3,000 BCE, a European Neolithic expansion brought Mediterranean-like
174 ancestry to the Maghreb, most likely from Iberia. Our analyses demonstrate that at least
175 some of the European ancestry observed today in North Africa is related to prehistoric
176 migrations, and local Berber populations were already admixed with Europeans before
177 the Roman conquest. Furthermore, additional European/Iberian ancestry could have
178 reached the Maghreb after KEB people; this scenario is supported by the presence of
179 Iberian-like Bell-Beaker pottery in more recent stratigraphic layers of IAM and KEB
180 caves. Future palaeogenomic efforts in North Africa will help further disentangle the

181 complex history of migrations that forged the ancestry of many admixed populations we
182 observe today.

183 **Methods**

184 Measures to avoid and monitor contamination from modern DNA were applied, at all
185 times, during sample manipulation. Ancient DNA was extracted from teeth or bone, built
186 into double-stranded indexed libraries and sequenced on an Illumina NextSeq 500
187 (Supplementary Note 2). Due to the environmental conditions of the burial sites, we
188 expected to recover low proportions of endogenous DNA from these ancient remains. To
189 overcome limitations due to DNA degradation, we applied two different capture methods
190 to enrich for human reads (Supplementary Note 2): one targeting the whole genome[19] and
191 one targeting the variants of the Multiethnic Genotyping Array (MEGA) array (Illumina
192 Inc.).

193 Reads were trimmed and adapters removed using AdapterRemoval[20], and then
194 mapped to the human reference genome (hg19) using BWA[21]. Low quality (MAPQ<30)
195 and duplicate reads were removed using SAMtools[22]. MapDamage[23] was used to
196 visualize misincorporation and fragmentation patterns, and to rescale the quality of bases
197 likely affected by post-mortem damage. Confidence intervals of sex determination were
198 calculated following Skoglund et al.[24]. MtDNA haplogroups were determined using
199 HaploGrep[25]. Y-chromosome haplogroup inference was carried out as in Schroeder et
200 al.[26]. As the reference panel, we used both the Human Origins panel[2] and the HGDP
201 dataset genotyped with MEGA-ex (Illumina Inc.). For principal component analysis, we
202 projected the aDNA samples on the PCA space built with the modern dataset, using
203 smartpca[27] and LASER[28]. Admixture estimations were done using ADMIXTURE
204 software[16]. F_{ST} distances were calculated using smartpca[27]. Identity-by-descent
205 proportions were estimated using PLINK[29], and heterozygosity estimations using a newly
206 developed method for low-coverage genomes (Supplementary Note 8). f-statistics
207 estimates were calculated using admixtools software[17]. All plots were prepared using R
208 software[30]. Detailed information about methods is included in the Supplementary Notes.

209

Acknowledgements R.F. and C.D.B were funded by a grant from the National Science Foundation (1201234). R.F. was funded by Fundación Canaria Dr. Manuel Morales fellowship. A.E.R.S. was funded by Ciencia sem Fronteiras fellowship - CAPES, Brazil. BS and JK were funded by a grant from the Gordon and Betty Moore Foundation (GBMF-3804).

Author Contributions R.F. and C.D.B conceived the idea for the study, Y.B., D.M.S, M.D.C.M, F.J.R.S, A.M., J.S. and A.T.M assembled skeletal material, R.F., A.E.R.S., J.K. and A.S. performed wet laboratory work, F.M. developed methods for data analysis, M.R. developed methods for DNA capture, R.F., F.M., M.A.A, P.A.U, G.W., analysed data, C.D.B and B.S. supervised the study, R.F., F.M. and C.D.B. wrote the manuscript and supplements with input from all co-authors.

Author Information Sequence data are available through the European Nucleotide Archive (PRJEB22699). Consensus mtDNA sequences are available at the National Center of Biotechnology Information (Accession Numbers XXXXXXX-XXXXXXX). The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.F. (e-mail: rfregel@stanford.edu).

References:

1 Nielsen, R. *et al.* Tracing the peopling of the world through genomics. *Nature* **541**, 302-310, doi:10.1038/nature21347 (2017).

2 Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419-424, doi:10.1038/nature19310 (2016).

3 Mulazzani, S. *et al.* The emergence of the Neolithic in North Africa: A new model for the Eastern Maghreb. *Quaternary International* **410**, 123 - 143, doi:http://dx.doi.org/10.1016/j.quaint.2015.11.089 (2016).

4 Barton, N. *et al.* Human Burial Evidence from Hattab II Cave and the Question of Continuity in Late Pleistocene–Holocene Mortuary Practices in Northwest Africa. *Cambridge Archaeological Journal* **18**, 195–214, doi: https://doi.org/10.1017/S0959774308000255 (2008).

5 Linstädter, J., Medved, I., Solich, M. & Weniger, G. C. Neolithisation process within the Alboran territory: Models and possible African impact. *Quaternary International* **274**, 219 - 232, doi:http://dx.doi.org/10.1016/j.quaint.2012.01.013 (2012).

6 Zilhão, J. Early prehistoric navigation in the Western Mediterranean: Implications for the Neolithic transition in Iberia and the Maghreb. *Eurasian Prehistory* **11**, 185-200 (2014).

7 Henn, B. M. *et al.* Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *Plos Genetics* **8**, e1002397-e1002397, doi:10.1371/journal.pgen.1002397 (2012).

8 Arauna, L. R. *et al.* Recent Historical Migrations Have Shaped the Gene Pool of Arabs and Berbers in North Africa. *Mol Biol Evol* **34**, 318-329, doi:10.1093/molbev/msw218 (2017).

9 Fadhlaoui-Zid, K. *et al.* Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa. *PLoS One* **8**, e80293, doi:10.1371/journal.pone.0080293 (2013).

10 Gallego Llorente, M. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**, 820-822, doi:10.1126/science.aad2879 (2015).

255 11 Schuenemann, V. J. *et al.* Ancient Egyptian mummy genomes suggest an increase
256 of Sub-Saharan African ancestry in post-Roman periods. *Nat Commun* **8**, 15694,
257 doi:10.1038/ncomms15694 (2017).
258 12 Garcia-Borja, P., Aura-Tortosa, J. E., Bernabeu-Auban, J. & Jorda-Pardo, J. F.
259 Nuevas perspectivas sobre la neolitizacion en la cueva de Nerja (Malaga-España):
260 La ceramica de la sala del vestibulo. *Zephyrus* **LXVI**, 109-132 (2010).
261 13 Pennarun, E. *et al.* Divorcing the Late Upper Palaeolithic demographic histories
262 of mtDNA haplogroups M1 and U6 in Africa. *BMC Evol Biol* **12**, 234,
263 doi:10.1186/1471-2148-12-234.; eng; ID: 4230 (2012).
264 14 Secher, B. *et al.* The history of the North African mitochondrial DNA haplogroup
265 U6 gene flow into the African, Eurasian and American continents. *BMC*
266 *evolutionary biology* **14**, 109, doi:1471-2148-14-109 [pii] (2014).
267 15 Haak, W. *et al.* Ancient DNA from European early neolithic farmers reveals their
268 near eastern affinities. *PLoS Biol* **8**, e1000536 (2010).
269 16 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of
270 ancestry in unrelated individuals. *Genome research* **19**, 1655-1664,
271 doi:10.1101/gr.094052.109 [doi] (2009).
272 17 Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192**, 1065-+,
273 doi:10.1534/genetics.112.145037 (2012).
274 18 Cortes Sanchez, M. *et al.* The Mesolithic-Neolithic transition in southern Iberia.
275 *Quaternary Research* **77**, 221-234 (2012).
276 19 Carpenter, M. L. *et al.* Pulling out the 1%: Whole-Genome Capture for the
277 Targeted Enrichment of Ancient DNA Sequencing Libraries. *American Journal of*
278 *Human Genetics* **93**, 852-864, doi:10.1016/j.ajhg.2013.10.002 (2013).
279 20 Lindgreen, S. AdapterRemoval: easy cleaning of next-generation sequencing
280 reads. *BMC research notes* **5**, 337, doi:10.1186/1756-0500-5-337 (2012).
281 21 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-
282 Wheeler transform. *Bioinformatics* **25**, 1754-1760,
283 doi:10.1093/bioinformatics/btp324 (2009).
284 22 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
285 **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

23 Ginolhac, A., Rasmussen, M., Gilbert, M. T., Willerslev, E. & Orlando, L. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* **27**, 2153-2155, doi:10.1093/bioinformatics/btr347 (2011).

24 Skoglund, P., Stora, J., Gotherstrom, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *Journal of Archaeological Science* **40**, 4477-4482 (2013).

25 Kloss-Brandstatter, A. *et al.* HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat* **32**, 25-32, doi:10.1002/humu.21382 (2011).

26 Schroeder, H. *et al.* Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proc Natl Acad Sci U S A* **112**, 3669-3673, doi:10.1073/pnas.1421784112 (2015).

27 Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909, doi:10.1038/ng1847 (2006).

28 Wang, C. *et al.* Ancestry estimation and control of population stratification for sequence-based association studies. *Nat Genet* **46**, 409-415, doi:10.1038/ng.2924 (2014).

29 Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559-575, doi:10.1086/519795 (2007).

30 A language and environment for statistical computing (R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org, 2008).

Figure 1. Geographical location (A) and calibrated radiocarbon date (B) of the samples included in this study, as well as other ancient DNA samples from the literature. *BOT sample was not radiocarbon dated, but was assigned to the Early Andalusian period on the basis of the associated material culture.

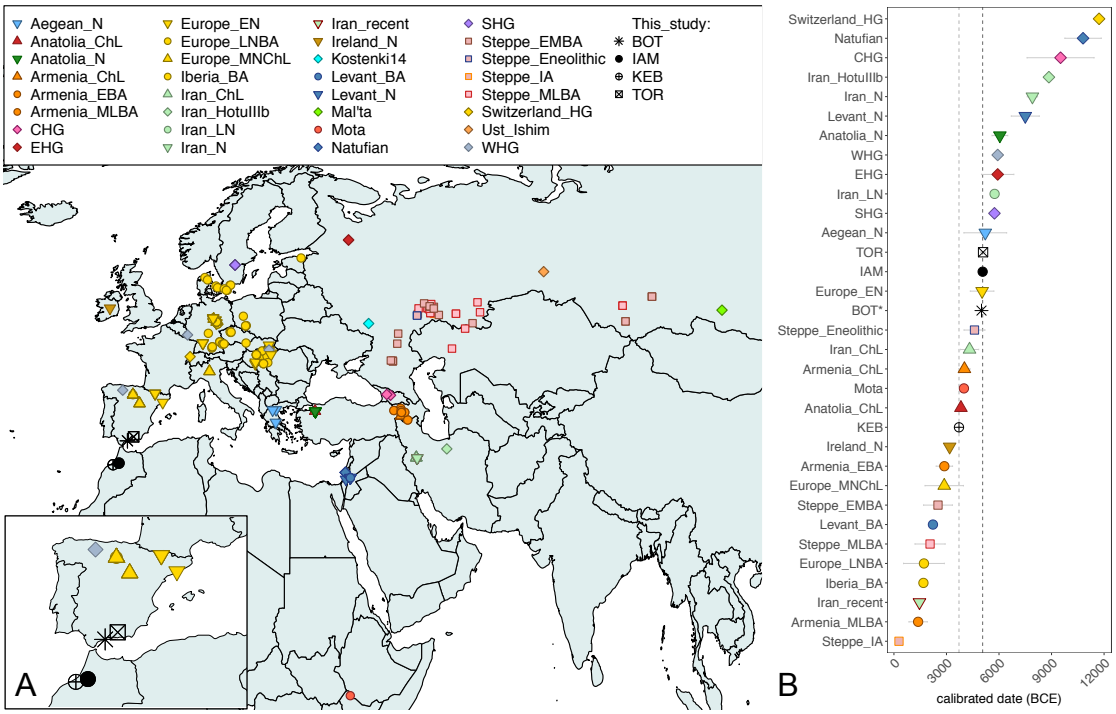

Figure 2. Ancestry inference in ancient samples from North Africa and the Iberian Peninsula. (a) PCA analysis, (b) ADMIXTURE analysis (K=2 - K=8), and (c) detailed ADMIXTURE analysis for European Neolithic samples (K=8).

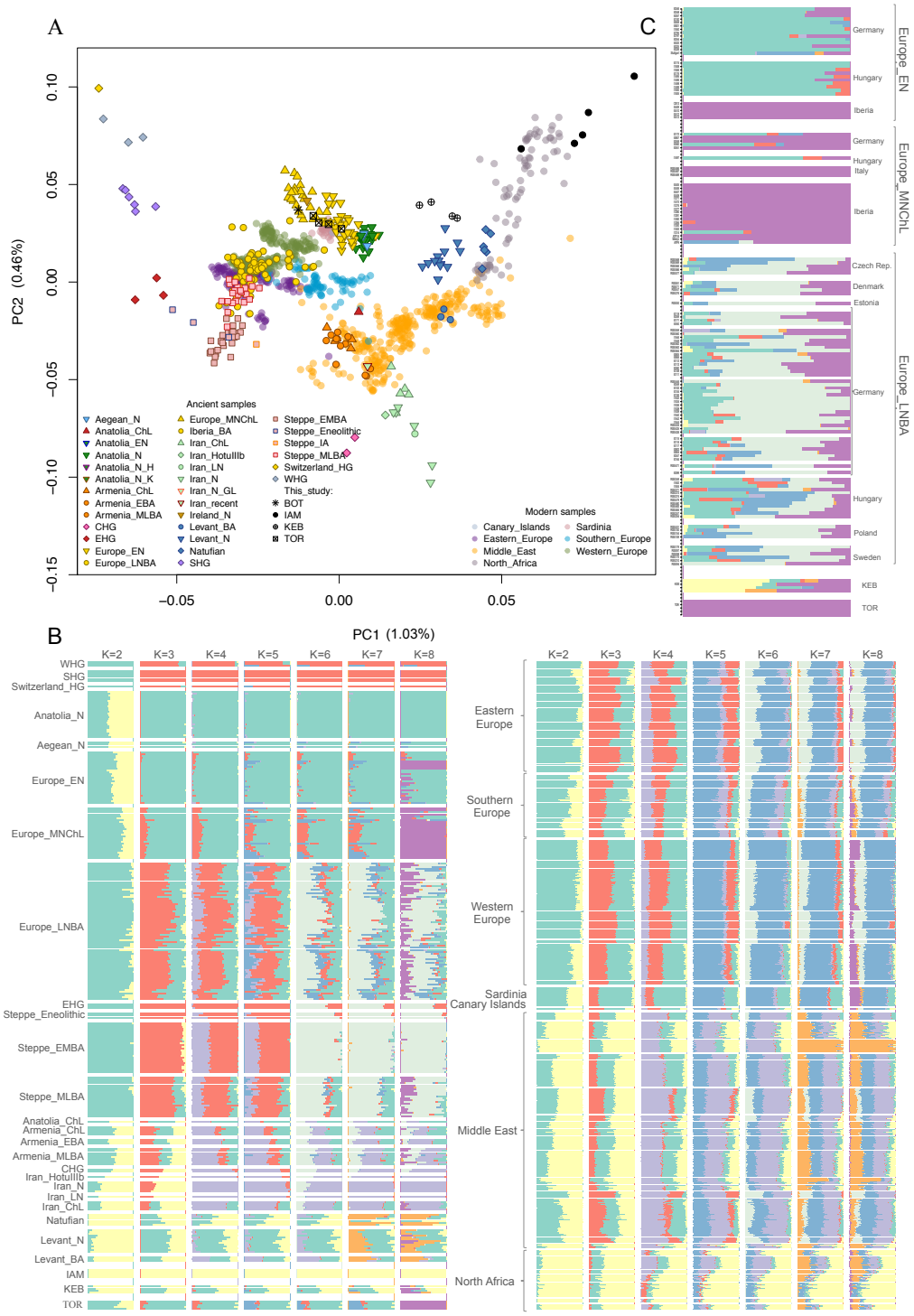
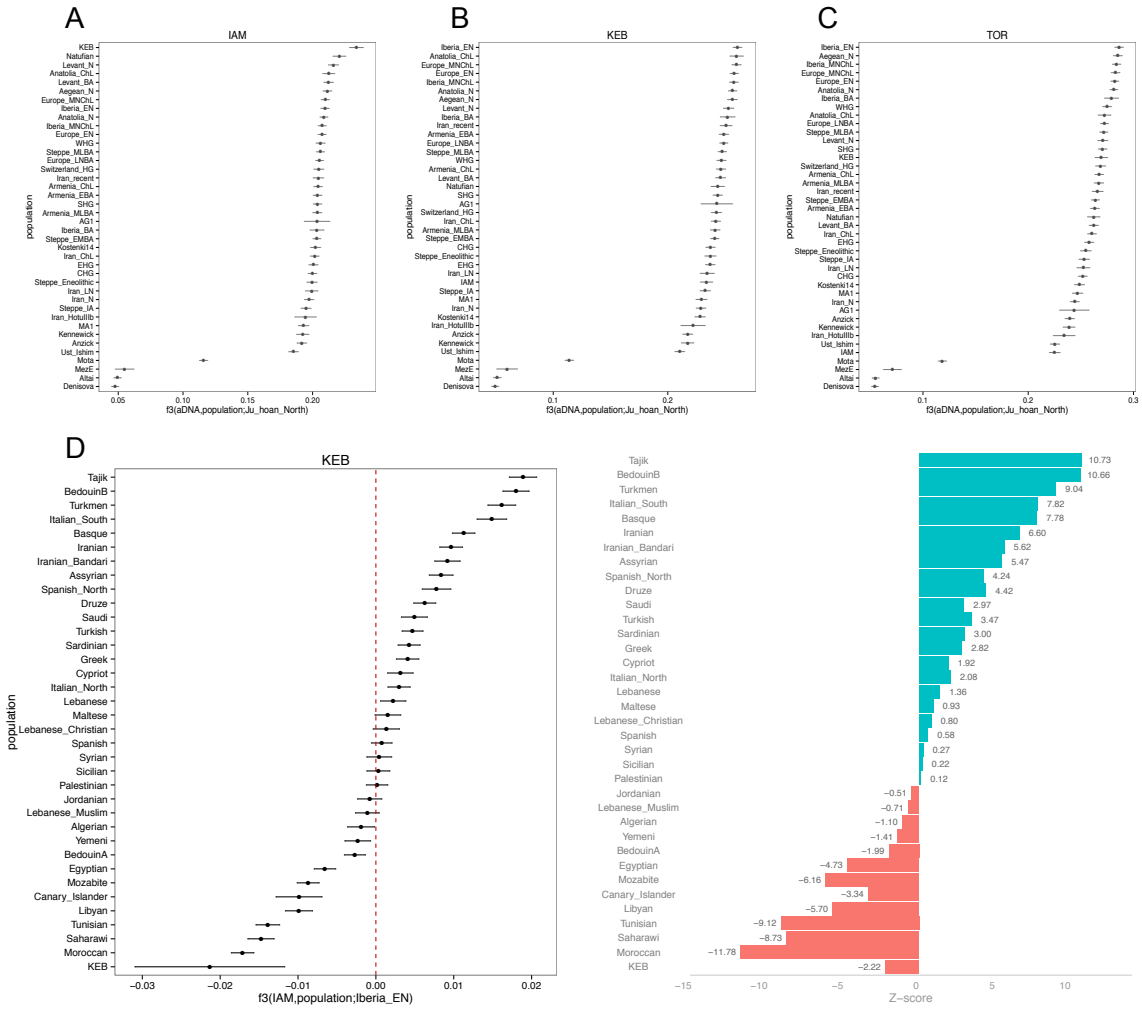
Figure 3. Outgroup f3-statistic for IAM (A), KEB (B) and TOR (C), and admixture f3-statistic for KEB (D).