1   **CLASSIFICATION**

2   BIOLOGICAL SCIENCES: Population Biology

3   **TITLE (112; 135 CHARACTERS MAX.)**

4   Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from

5   both the Levant and Europe

6   **SHORT TITLE (48; 50 CHARACTERS MAX.)**

7   Paleogenomics of Moroccan Neolithic populations

8   **AUTHORS**

9   Rosa Fregel[a*], Fernando L. Méndez[a], Youssef Bokbot[b], Dimas Martín-Socas[c], María D.

10   Camalich-Massieu[c], Jonathan Santana[d], Jacob Morales[e], María C. Ávila-Arcos[a,f], Peter A.

11   Underhill[a], Beth Shapiro[g], Genevieve Wojcik[a], Morten Rasmussen[a], Andre E. R. Soares[g§],

12   Joshua Kapp[g], Alexandra Sockell[a], Francisco J. Rodríguez-Santos[h], Abdeslam Mikdad[b],

13   Aioze Trujillo-Mederos[c] & Carlos D. Bustamante[a¶]

14   **AFFILIATIONS**

15   [a]Department of Genetics, School of Medicine, Stanford University, 365 Lasuen Street, CA

16   94305, Stanford, US.

17   [b]Institut National des Sciences de l'Archéologie et du Patrimoine, Avenue Allal el Fassi,

18   6828, Rabat, Morocco.

19   [c]Departmento de Prehistoria, Universidad de La Laguna, Calle Prof. Jose Luis Moreno

20   Becerra, 38320, San Cristóbal de La Laguna, Spain.

21   [d]Department of Archaeology, Durham University, South Rd, Durham DH1 3LE, United

22   Kingdom.

23   [e]Departamento de Ciencias Históricas, Universidad de Las Palmas de Gran Canaria, Pérez

24   del Toro 1, 35003, Las Palmas de Gran Canaria, Spain.

25   [f]International Laboratory for Human Genome Research, National Autonomous University of

26   Mexico, Blvd. Juriquilla 3001, 76230, Querétaro, Mexico.

27   [g]Department of Ecology and Evolutionary Biology, University of California, 1156 High

28   Street, CA 95064, Santa Cruz, US.

29   [h]Instituto Internacional de Investigaciones Prehistóricas de Cantabria, Avda. de los Castros,

30   39005, Cantabria, Spain.

31

32   **CORRESPONDING AUTHOR**

33   Rosa Fregel

34   Departamento de Genética

35   Universidad de La Laguna

36   Avda. Astrofísico Fco. Sánchez, SN

37   San Cristóbal de La Laguna 38206

38   Spain

39   rfregel@ull.edu.es

40

43

ABSTRACT (249; 250 WORDS MAX.)

44

45   The extent to which prehistoric migrations of farmers influenced the genetic pool of western

46   North Africans remains unclear. Archaeological evidence suggests the Neolithization process

47   may have happened through the adoption of innovations by local Epipaleolithic

48   communities, or by demic diffusion from the Eastern Mediterranean shores or Iberia. Here,

49   we present the first analysis of individuals' genome sequences from early and late Neolithic

50   sites in Morocco, as well as Early Neolithic individuals from southern Iberia. We show that

51   Early Neolithic Moroccans are distinct from any other reported ancient individuals and

52   possess an endemic element retained in present-day Maghrebi populations, confirming a

53   long-term genetic continuity in the region. Among ancient populations, Early Neolithic

54   Moroccans are distantly related to Levantine Natufian hunter-gatherers (~9,000 BCE) and

55   Pre-Pottery Neolithic farmers (~6,500 BCE). Although an expansion in Early Neolithic times

56   is also plausible, the high divergence observed in Early Neolithic Moroccans suggests a

57   long-term isolation and an early arrival in North Africa for this population. This scenario is

58   consistent with early Neolithic traditions in North Africa deriving from Epipaleolithic

59   communities who adopted certain innovations from neighbouring populations. Late Neolithic

60   (~3,000 BCE) Moroccans, in contrast, share an Iberian component, supporting theories of

61   trans-Gibraltar gene flow. Finally, the southern Iberian Early Neolithic samples share the

62   same genetic composition as the Cardial Mediterranean Neolithic culture that reached Iberia

63   ~5,500 BCE. The cultural and genetic similarities of the Iberian Neolithic cultures with that

64   of North African Neolithic sites further reinforce the model of an Iberian migration into the

65   Maghreb.

66

SIGNIFICANCE STATEMENT (119; 120 WORDS MAX.)

67

68   The acquisition of agricultural techniques during the so-called Neolithic revolution has been

69   one of the major steps forward in human history. Using next-generation sequencing and

70   ancient DNA techniques, we directly test if Neolithization in North Africa occurred through

71   the transmission of ideas or by demic diffusion. We show that Early Neolithic Moroccans are

72   composed of an endemic Maghrebi element still retained in present-day North African

73   populations and distantly related to Epipaleolithic communities from the Levant. However,

74   late Neolithic individuals from North Africa are admixed, with a North African and a

75   European component. Our results support the idea that the Neolithization of North Africa

76   might have involved both the development of Epipaleolithic communities and the migration

77   of people from Europe.

78

INTRODUCTION

One of the greatest transitions in human history was the transition from hunter-gatherer to farming lifestyle. How farming traditions expanded from their birthplace in the Fertile Crescent has 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 that some of the major innovations associated with the Neolithic, such as farming and pottery production, could have been introduced into northern Morocco through sea voyaging by people from Iberia or the central Mediterranean as early as *ca*. 5400 BCE[3,4]. In fact, some of the Neolithic pottery recorded in North Africa strongly resembles that of European cultures like the Cardial Early Neolithic, the Mediterranean early farmer culture located in Iberia[5]. However, other innovations such as some pottery traditions and bone and lithic technical customs could be the result of *in situ* development from Epipaleolithic communities, indicating a strong continuity in the local population since the Late Pleistocene[6-10].

Genetic data from present-day populations[11-13] suggests that North African ancestry has contributions from four main sources: 1) an autochthonous Maghrebi component related to a back migration to Africa ~12,000 years ago from Eurasia; 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[14-16]; however, no research to date has tested whether the Neolithic transition in the Maghreb was driven by local populations who adopted cultural and technological innovations 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; ~5,200 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 human remains from the

108   southern Iberian Early Neolithic site of El Toro (TOR; ~5,000 BCE; n=12) (Figure 1). This

109   Iberian Early Neolithic culture bears similarities with early Maghrebi pottery decoration, as

110   well as bone and lithic tool production traditions which suggest an African influence[17]

111   (Supplementary Note 1). Including these southern Iberian samples in our analysis enables a

112   direct test of this hypothesis.

113   [Figure 1 here]

114   **RESULTS AND DISCUSSION**

115   We sequenced 38 Illumina pair-end libraries from 27 individuals, and selected the best-

116   conserved libraries for subsequent analyses. Endogenous DNA content was generally low

117   (2.88% on average) (Supplementary Note 2). Depth of coverage was consistently improved

118   when enriching using baits targeting specific sites for the Multiethnic Genotyping Array

119   (MEGA) (~100X), compared to whole-genome capture (~15X) (Supplementary Note 2).

120   Following enrichment, we generated thirteen low-coverage genomes (five from IAM, four

121   from KEB and four from TOR), with MEGA coverage ranging from 0.04X to 1.72X depth,

122   and genome-wide coverage ranging from 0.01X to 0.74X depth (Table 1). All samples

123   considered in this study met the standard aDNA authentication criteria, including observation

124   of DNA fragmentation (~46 bp average read length) and damage patterns due to cytosine

125   deamination toward the 5′ ends of molecules (Supplementary Note 3).

126   Mitochondrial DNA and Y-chromosome haplogroups obtained for IAM (Moroccan Early

127   Neolithic) and KEB (Moroccan Late Neolithic) suggest either a population replacement or an

128   important genetic influx into Morocco between 5,000−3,000 BCE. IAM samples belong to

129   the mtDNA haplogroups U6a and M1—both of which are associated with the back migration

130   to Africa from Eurasia in Upper Palaeolithic times[18,19]—while KEB samples belong to

131   haplogroups K1, T2 and X2, prominently found in Anatolian and European Neolithic

132   samples[2,20] (Supplementary Note 4). Regarding the paternal lineages, IAM individuals carry

133   Y chromosomes distantly related to the typically North African E-M81 haplogroup, while

134   the Y chromosome from KEB belongs to the T-M184 haplogroup; though scarce and broadly

135   distributed today, this haplogroup has also been observed in European Neolithic individuals[16]

136   (Supplementary Note 5). Both mtDNA and Y-chromosome lineages (K1, J2 and T2

137   haplogroups, and G-M201 haplogroup, respectively) for samples from TOR (Iberian Early

138   Neolithic) are similar to those observed in Europe during Neolithic times[20].

139   [Table 1 here]

140   When projected on a Principal Components Analysis (PCA) space built using sub-Saharan

141   African, North African, European and Middle Eastern population of the Human Genome

142   Diversity Project (HGDP) dataset genotyped with MEGA, IAM samples are placed close to

143   Mozabites, while Iberian Neolithic samples fall close to southern European populations

144   (Supplementary Note 6). As suspected from the mtDNA and Y-chromosome data, KEB

145   samples do not cluster with IAM and are placed in an intermediate position between IAM

146   and TOR. We further explored the genetic structure of these samples using the program

147   ADMIXTURE[21] (Figure 2). At K=5, we observe sub-Saharan African (red), early European

148   Neolithic (green), North African (yellow), Middle Eastern (violet) and eastern European

149   components (orange). Congruently with PCA results, TOR is composed of the early

150   European component, clustering with Sardinian samples, and IAM is composed of the North

151   African component, clustering with Mozabites. Finally, KEB is placed in an intermediate

152   position, with ~50% of both early European Neolithic and North African ancestries. It is

153   worth mentioning that, compared to current North African samples, IAM and KEB do not

154   show any sub-Saharan African ancestry, suggesting that trans-Saharan migrations occurred

155   after Neolithic times. This is in agreement with the analysis of present-day genome-wide

156   data from Morocco, which estimated a migration of western African origin into Morocco

157   only ~1,200 years ago[11].

158   West Eurasian populations can be modelled as admixture of four different ancestral

159   components[2]: Eastern and Western European hunter-gatherers, and Iranian and Levantine

160   Neolithic. We explored the placement of Moroccan and Southern Iberian Neolithic samples

161   in this context, and compared their genetic affinities to ancient and present-day West

162   Eurasian and Levant populations in the Human Origins panel. Interestingly, PCA reveals that

163   IAM individuals are different from any aDNA sample studied to date (Figure 2;

164   Supplementary Note 6). When projected, IAM samples are close to modern North Africans,

165   in the Levantine corner of the PCA space (Figure 2). Southern Iberian Neolithic individuals

166   from TOR cluster with Sardinians and with other Anatolian and European Neolithic samples.

167   Moreover, KEB samples are placed halfway between the IAM and Anatolian/European

168   farmer clusters, in close proximity to Levant aDNA samples and also to Guanche samples[16],

169   the indigenous population of the Canary Islands known to have a Berber origin[22]. When

170   compared using ADMIXTURE (See Supplementary Note 7 for details), IAM samples

171   possess ~100% of a component partially shared by aDNA samples from the Middle East and

172   Levant at low K values. At K=6, this IAM-like component is observed mainly in modern

173   North Africa, following a west-to-east cline, and in the Guanches. TOR and other Early

174   Neolithic samples from Iberia cluster together with farmers from Anatolia, the Aegean area

175   and Europe. At K=8, the Early Neolithic individuals from Iberia differentiate from the

176   Anatolian, Aegean and European Early Neolithic samples, and share their main component

177   (purple) with Middle Neolithic/Chalcolitic samples (Figure 2). Finally, at low K values, KEB

178   can be explained as having both IAM-like and European Neolithic components, suggesting

179   an admixture process between IAM-like people and early farmers. Nevertheless, at K=8, the

180   European component in KEB is predominantly "purple," with some "green" component.

181   This "green" component is also present, at a low frequency, in Natufians and other ancient

182   Levantine populations. The substantially larger contribution of the "purple" component,

183   when compared with the "green", suggests a significant genetic contribution of ancient

184   Iberians in Morocco (Figure 2). The same admixture profile is observed in Guanches, but the

185   amount of IAM ancestry is consistently higher in all the samples. Given that the Guanches

186   could have had originated in a different area of the Maghreb, this result might suggest that

187   the European Neolithic impact in North Africa was heterogeneous.

188   [Figure 2 here]

189   To compare our samples directly to the genomes of ancient and modern populations, we

190   calculated pair-wise F_{ST} distances, which, unlike PCA and global ancestry analyses, are

191   insensitive to the inclusion of large numbers of individuals from modern populations. F_{ST}

192   values indicate that the IAM samples are as differentiated from all other populations as

193   Yoruba are from non-Africans (Supplementary Note 9), with the sole exception of KEB and,

194   to a lesser extent, the Guanches and modern North African populations. Given the relatively

195   low heterozygosity and high identity-by-descent proportions observed in IAM

196   (Supplementary Note 8), this differentiation could be driven by isolation and genetic drift.

197   IAM is divergent from the other populations, with the exception of populations that likely

198   received genetic influx from them. This raises the possibility that IAM was isolated in North

199   Africa since Palaeolithic times, when a back migration from Eurasia brought mtDNA

200   haplogroups M1 and U6 to the Maghreb[18]. Although IAM is clearly more similar to KEB

201   than to any other population, the converse is not true. KEB has lower F_{ST} distances with any

202   Anatolian, European (excluding European hunter gatherers), Levantine and Iranian

203   population, rather than with IAM. In the modern DNA reference panel, KEB is similar to

204   North African, European and Middle Eastern populations. Among the ancient populations,

205   TOR is more similar to Middle Neolithic/Chalcolithic Europeans, and, among modern

206   populations, to populations from Spain, North Italy and Sardinia.

207   To further investigate the genetic affinities of IAM, KEB and TOR samples, we conducted

208   outgroup f3-statistic analysis[23]. Results indicate that, when KEB and Guanches are excluded,

209   IAM shares more drift with ancient Levantine populations, such as Natufians (Epipaleolithic)

210   and Pre-Pottery Neolithic individuals (Figure 3; Supplementary Note 10), than with any

211   other ancient population. To explore further the connection between IAM and Levantine

212   populations, we performed an f4-statistic analysis to test whether IAM shares more alleles

213   with any other population in the Human Origins panel[2,24] than with ancient populations from

214   the Levant (Supplementary Note 10). Consistently, and also with the exception of KEB and

215   Guanches, all comparisons indicated higher similarity with Natufians and Levantine farmers.

216   This suggests that most of IAM ancestry originates from an out-of-Africa source, as IAM

217   shares more alleles with Levantines than with any sub-Saharan Africans, including the

218   4,500-year-old genome from Ethiopia[14]. To further test the hypothesis that IAM is more

219   closely related to out-of-Africa populations, we determined if we could detect Neanderthal

220   ancestry in IAM, which is typical of non-African populations. A signal of Neanderthal

221   ancestry has been detected in modern North African populations[25]. A lack of Neanderthal

222   ancestry in IAM would imply that the signal observed today is a product of more recent

223   migration into North Africa from the Middle East and Europe in historical times. When

224   compared to the Neanderthal high coverage genome sequence from Altai[26] and the low-

225   coverage sequence from Vindija Cave[27], and using the S-statistic[23], we detected a

226   Neanderthal introgression signal into IAM, suggesting derivation from the same event shared

227   by non-African populations. All these results together indicate that the origin of IAM was

228   outside Africa, most probably from the Levant. However, it is important to take into account

229   that the number of ancient genomes for comparison is low and future sampling can provide

230   further refinement in the origin of IAM.

231   Both F_{ST} and outgroup-f3 statistic analyses indicate that KEB shares ancestry with IAM, but

232   also more genetic drift with Neolithic and Chalcolithic populations from Anatolia and

233   Europe, with the highest shared genetic drift appearing in Iberian Early Neolithic samples

234   (Figure 3; Supplementary Note 10). This pattern and the result from ADMIXTURE could be

235   explained if the KEB population was a mixture between IAM-related and European

236   Neolithic groups. To formally test this hypothesis, we used an admixture-f3 test[23], using

237   KEB as the test population, IAM as a reference population and one of the Anatolian and

238   European Neolithic and Chalcolithic populations as the second reference population. All

239   comparisons produced negative values of the f3-statistic, which suggests the KEB population

240   can be modelled as a mixture of IAM and Anatolian/European Neolithic.

241   TOR has more shared ancestry with Iberian Early Neolithic samples and other Neolithic and

242   Chalcolithic populations from Europe. Archaeological studies have suggested that there was

243   an Andalusian Early Neolithic culture with North African influences before the Cardial

244   expansion into the Western Mediterranean basin[28]. However, we observe that TOR samples

245   have a similar genetic composition to that of Cardial individuals from Iberia, evidencing a

246   common origin, and ruling out an Andalusian Early Neolithic distinct from Cardial Culture.

247   [Figure 3 here]

248   Finally, although limited by low coverage, phenotypic predictions based on genetic variants

249   of known effects agree with our estimates of global ancestry. IAM people do not possess any

250   of the European SNPs associated with light pigmentation, and most likely had dark skin and

251   eyes. IAM samples present ancestral alleles for pigmentation-associated variants present in

252   SLC24A5 (rs1426654), SLC45A2 (rs16891982) and OCA2 (rs1800401 and 12913832)

253   genes. On the other hand, KEB individuals exhibit some European- derived alleles that

254   predispose individuals to lighter skin and eye colour, including those on genes SLC24A5

255   (rs1426654) and OCA2 (rs1800401) (Supplementary Note 11).

256   **CONCLUSION**

257   Genetic analyses have revealed that the population history of modern North Africans is quite

258   complex[11]. Based on our aDNA analysis, we identify an Early Neolithic Moroccan

259   component that is restricted to North Africa in present-day populations[11], and that is the sole

260   ancestry in IAM samples. We hypothesize that this component represents the autochthonous

261   Maghrebi ancestry associated with Berber populations. This Maghrebi component is

262   different from those of any ancient samples studied so far and is distantly related to that of

263   Epipaleolithic people from the Levant. Our data suggests that the IAM population was

264   isolated in the Maghreb since the Upper Palaeolithic back migration, although it is

265   impossible to be certain without paleogenomic data from North African Palaeolithic samples.

266   An expansion in Early Neolithic times followed by strong genetic drift might also be

267   plausible.

268   Our hypothesis is in agreement with archaeological research pointing to the first stage of the

269   Neolithic expansion in Morocco as the result of a local population who adopted some

270   technological innovations, such as pottery production or farming, from neighbouring areas.

271   By 3,000 BCE, a continuity in the Neolithic spread brought Mediterranean-like ancestry to

272   the Maghreb, most likely from Iberia. Other archaeological remains, such as African

273   elephant ivory and ostrich eggs found in Iberian sites, confirm the existence of contacts and

274   exchange networks through both sides of the Gibraltar strait at this time. Our analyses

275   strongly support that at least some of the European ancestry observed today in North Africa

276   is related to prehistoric migrations, and local Berber populations were already admixed with

277   Europeans before the Roman conquest. Furthermore, additional European/Iberian ancestry

278   could have reached the Maghreb after KEB people; this scenario is supported by the

279   presence of Iberian-like Bell-Beaker pottery in more recent stratigraphic layers of IAM and

280   KEB caves. Future palaeogenomic efforts in North Africa will further disentangle the

281   complex history of migrations that forged the ancestry of the admixed populations we

282   observe today.

283   **MATERIAL AND METHODS**

284   Measures to avoid and monitor contamination from modern DNA were applied, at all times,

285   during sample manipulation. Ancient DNA was extracted from teeth or bone, built into

286   double-stranded indexed libraries and sequenced on an Illumina NextSeq 500

287   (Supplementary Note 2). Due to the environmental conditions of the burial sites, we

288   expected to recover low proportions of endogenous DNA from these ancient remains. To

289   overcome limitations due to DNA degradation, we applied two different capture methods to

290   enrich for human reads (Supplementary Note 2): one targeting the whole genome[29] and one

291   targeting the variants of the MEGA array (Illumina Inc.).

292   Reads were trimmed and adapters removed using AdapterRemoval[30], and then mapped to the

293   human reference genome (hg19) using BWA[31]. Low quality (MAPQ<30) and duplicate reads

294   were removed using SAMtools[32]. MapDamage[33] was used to visualize misincorporation and

295   fragmentation patterns, and to rescale the quality of bases likely affected by post-mortem

296   damage. Confidence intervals of sex determination were calculated following Skoglund et

297   al.[34]. MtDNA haplogroups were determined using HaploGrep[35]. Y-chromosome haplogroup

298   inference was carried out as in Schroeder et al.[36]. As the reference panel, we used both the

299   Human Origins panel[2] and the HGDP dataset genotyped with MEGA-ex (Illumina Inc.). For

300   principal component analysis, we projected the aDNA samples on the PCA space built with

301   the modern dataset, using smartpca[37] and LASER[38]. Admixture estimations were done using

302   ADMIXTURE software[21]. F_{ST} distances were calculated using smartpca[37]. Identity-by-

303   descent proportions were estimated using PLINK[39], and heterozygosity estimations using a

304   newly developed method for low-coverage genomes (Supplementary Note 8). f-statistics

305   estimates were calculated using admixtools software[23]. All plots were prepared using R

306   software[40]. Detailed information about methods is included in the Supplementary Notes.

307

312   **Author Contributions** R.F. and C.D.B conceived the idea for the study, D.M.S, M.D.C.M, J.S., J.M, Y.B.,

313   F.J.R.S, A.M. and A.T.M assembled skeletal material and provided archaeological background for the samples,

314   R.F., A.E.R.S., J.K. and A.S. performed work in the wet laboratory, F.M. developed methods for data analysis,

315   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.

316   supervised the study, R.F., F.M. and C.D.B. wrote the manuscript and supplements with input from all co-

317   authors.

318   **Author Information** Sequence data are available through the European Nucleotide Archive (PRJEB22699).

319   Consensus mtDNA sequences are available at the National Center of Biotechnology Information (Accession

320   Numbers MF991431-MF991448). The authors declare no competing financial interests. Correspondence and

321   requests for materials should be addressed to R.F. (e-mail: rfregel@stanford.edu).

REFERENCES

322

323   1   Nielsen, R. *et al.* Tracing the peopling of the world through genomics. *Nature* **541**,
324   302-310, doi:10.1038/nature21347 (2017).

325   2   Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near
326   East. *Nature* **536**, 419-424, doi:10.1038/nature19310 (2016).

327   3   Linstädter, J., Medved, I., Solich, M. & Weniger, G. C. Neolithisation process within
328   the Alboran territory: Models and possible African impact. *Quaternary International*
329   **274**, 219 - 232, doi:10.1016/j.quaint.2012.01.013 (2012).

330   4   Zilhão, J. Early prehistoric navigation in the Western Mediterranean: Implications for
331   the Neolithic transition in Iberia and the Maghreb. **Eurasian Prehistory**, 185-200
332   (2014).

333   5   Martínez-Sánchez, R. M., Vera-Rodríguez, J. C., Pérez-Jordà, G., Peña-Chocarro, L.
334   & Bokbot, Y. The beginning of the Neolithic in northwestern Morocco. *Quaternary*
335   *International* **In press** (2017).

336   6   Mulazzani, S. *et al.* The emergence of the Neolithic in North Africa: A new model
337   for the Eastern Maghreb. *Quaternary International* **410**, 123 - 143,
338   doi:10.1016/j.quaint.2015.11.089 (2016).

339   7   Barton, N. *et al.* Human Burial Evidence from Hattab II Cave and the Question of
340   Continuity in Late Pleistocene–Holocene Mortuary Practices in Northwest Africa.
341   *Cambridge Archaeological Journal* **18**, 195–214, doi:10.1017/S0959774308000255
342   (2008).

343   8   Lubell, D., Sheppard, P. & Jackes, M. in *Advances in World Archaeology* (eds F.
344   Wendorf & A. Close) 146-191 (Academic Press, 1984).

345   9   Linstädter, J. Epipalaeolithic-Neolithic-Transition in the Mediterranean region of
346   Northwest-Africa. *Quartär* **55**, 41-62 (2008).

347   10   de Groote, I. & Humphrey, L. T. Characterizing evulsion in the Later Stone Age
348   Maghreb: Age, sex and effects on mastication. *Quaternary International* **413**, 50 - 61,
349   doi:https://doi.org/10.1016/j.quaint.2015.08.082 (2016).

350   11   Henn, B. M. *et al.* Genomic Ancestry of North Africans Supports Back-to-Africa
351   Migrations. *Plos Genetics* **8**, e1002397-e1002397, doi:10.1371/journal.pgen.1002397
352   (2012).
353   12   Arauna, L. R. *et al.* Recent Historical Migrations Have Shaped the Gene Pool of
354   Arabs and Berbers in North Africa. *Mol Biol Evol* **34**, 318-329,
355   doi:10.1093/molbev/msw218 (2017).
356   13   Fadhlaoui-Zid, K. *et al.* Genome-wide and paternal diversity reveal a recent origin of
357   human populations in North Africa. *PLoS One* **8**, e80293,
358   doi:10.1371/journal.pone.0080293 (2013).
359   14   Gallego Llorente, M. *et al.* Ancient Ethiopian genome reveals extensive Eurasian
360   admixture throughout the African continent. *Science* **350**, 820-822,
361   doi:10.1126/science.aad2879 (2015).
362   15   Schuenemann, V. J. *et al.* Ancient Egyptian mummy genomes suggest an increase of
363   Sub-Saharan African ancestry in post-Roman periods. *Nat Commun* **8**, 15694,
364   doi:10.1038/ncomms15694 (2017).
365   16   Rodriguez-Varela, R. *et al.* Genomic Analyses of Pre-European Conquest Human
366   Remains from the Canary Islands Reveal Close Affinity to Modern North Africans.
367   *Curr Biol* **27**, 3396-3402 e3395, doi:10.1016/j.cub.2017.09.059 (2017).
368   17   Garcia-Borja, P., Aura-Tortosa, J. E., Bernabeu-Auban, J. & Jorda-Pardo, J. F.
369   Nuevas perspectivas sobre la neolitizacion en la cueva de Nerja (Malaga-España): La
370   ceramica de la sala del vestibulo. *Zephyrus* **LXVI**, 109-132 (2010).
371   18   Pennarun, E. *et al.* Divorcing the Late Upper Palaeolithic demographic histories of
372   mtDNA haplogroups M1 and U6 in Africa. *BMC Evol Biol* **12**, 234,
373   doi:10.1186/1471-2148-12-234.; eng; ID: 4230 (2012).
374   19   Secher, B. *et al.* The history of the North African mitochondrial DNA haplogroup U6
375   gene flow into the African, Eurasian and American continents. *BMC evolutionary*
376   *biology* **14**, 109, doi:1471-2148-14-109 [pii] (2014).
377   20   Haak, W. *et al.* Ancient DNA from European early neolithic farmers reveals their
378   near eastern affinities. *PLoS Biol* **8**, e1000536 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/191569; this version posted February 20, 2018. 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 a CC-BY-NC-ND 4.0 International license.

16

379   21   Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry
380   in unrelated individuals. *Genome research* **19**, 1655-1664,
381   doi:10.1101/gr.094052.109 [doi] (2009).
382   22   Maca-Meyer, N. *et al.* Ancient mtDNA analysis and the origin of the Guanches. *Eur*
383   *J Hum Genet* **12**, 155-162 (2004).
384   23   Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192**, 1065-+,
385   doi:10.1534/genetics.112.145037 (2012).
386   24   Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for
387   present-day Europeans. *Nature* **513**, 409-413, doi:10.1038/nature13673 (2014).
388   25   Sanchez-Quinto, F. *et al.* North African populations carry the signature of admixture
389   with Neandertals. *PLoS One* **7**, e47765, doi:10.1371/journal.pone.0047765 (2012).
390   26   Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai
391   Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 [doi] (2014).
392   27   Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722
393   (2010).
394   28   Cortes Sanchez, M. *et al.* The Mesolithic-Neolithic transition in southern Iberia.
395   *Quaternary Research* **77**, 221-234 (2012).
396   29   Carpenter, M. L. *et al.* Pulling out the 1%: Whole-Genome Capture for the Targeted
397   Enrichment of Ancient DNA Sequencing Libraries. *American Journal of Human*
398   *Genetics* **93**, 852-864, doi:10.1016/j.ajhg.2013.10.002 (2013).
399   30   Lindgreen, S. AdapterRemoval: easy cleaning of next-generation sequencing reads.
400   *BMC research notes* **5**, 337, doi:10.1186/1756-0500-5-337 (2012).
401   31   Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
402   transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
403   32   Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
404   2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
405   33   Ginolhac, A., Rasmussen, M., Gilbert, M. T., Willerslev, E. & Orlando, L.
406   mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics*
407   **27**, 2153-2155, doi:10.1093/bioinformatics/btr347 (2011).

408   34   Skoglund, P., Stora, J., Gotherstrom, A. & Jakobsson, M. Accurate sex identification
409   of ancient human remains using DNA shotgun sequencing. *Journal of*
410   *Archaeological Science* **40**, 4477-4482 (2013).
411   35   Kloss-Brandstatter, A. *et al.* HaploGrep: a fast and reliable algorithm for automatic
412   classification of mitochondrial DNA haplogroups. *Hum Mutat* **32**, 25-32,
413   doi:10.1002/humu.21382 (2011).
414   36   Schroeder, H. *et al.* Genome-wide ancestry of 17th-century enslaved Africans from
415   the Caribbean. *Proc Natl Acad Sci U S A* **112**, 3669-3673,
416   doi:10.1073/pnas.1421784112 (2015).
417   37   Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-
418   wide association studies. *Nat Genet* **38**, 904-909, doi:10.1038/ng1847 (2006).
419   38   Wang, C. *et al.* Ancestry estimation and control of population stratification for
420   sequence-based association studies. *Nat Genet* **46**, 409-415, doi:10.1038/ng.2924
421   (2014).
422   39   Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-
423   based linkage analyses. *American Journal of Human Genetics* **81**, 559-575,
424   doi:10.1086/519795 (2007).
425   40   A language and environment for statistical computing (R Foundation for Statistical
426   Computing. Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org,
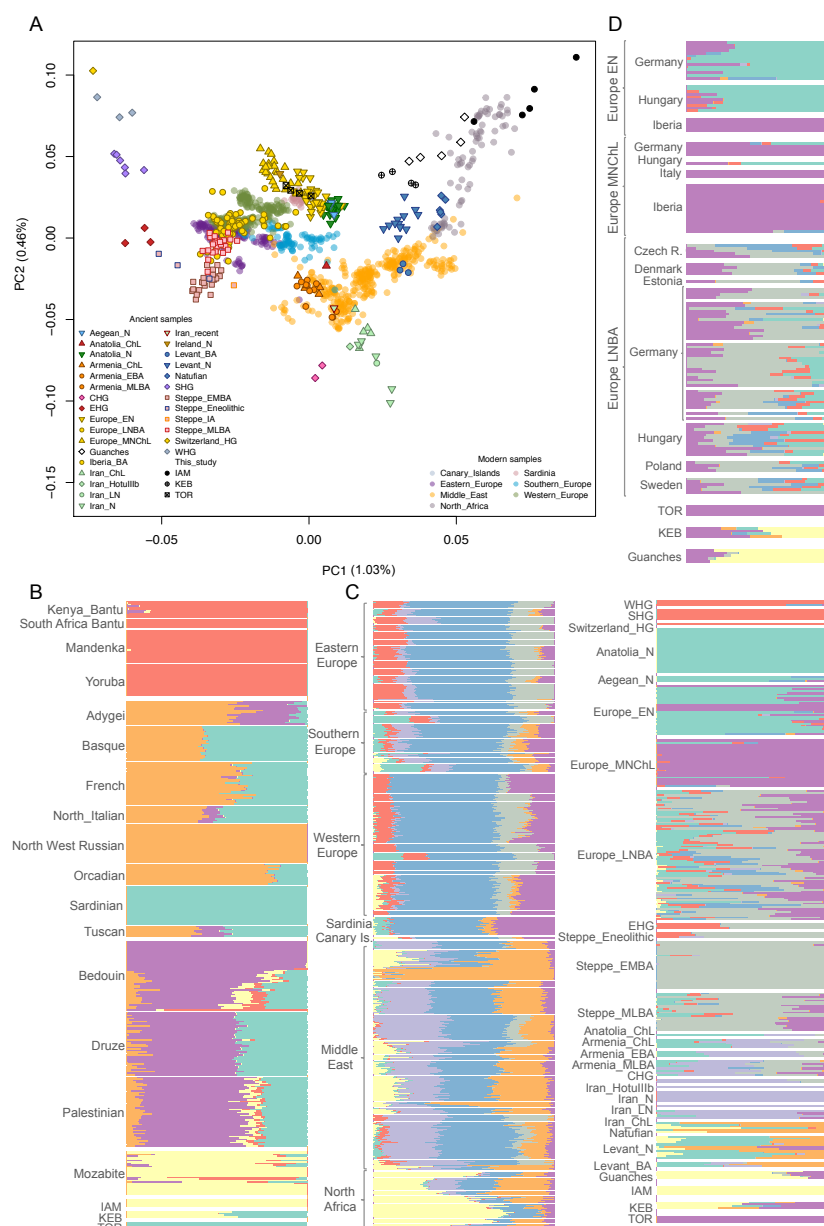427   2008).
428

432 **FIGURES**

433 Figure 1. Geographical location (A) and calibrated radiocarbon date (B) of the

434 samples included in this study, as well as other ancient DNA samples from the

435 literature.
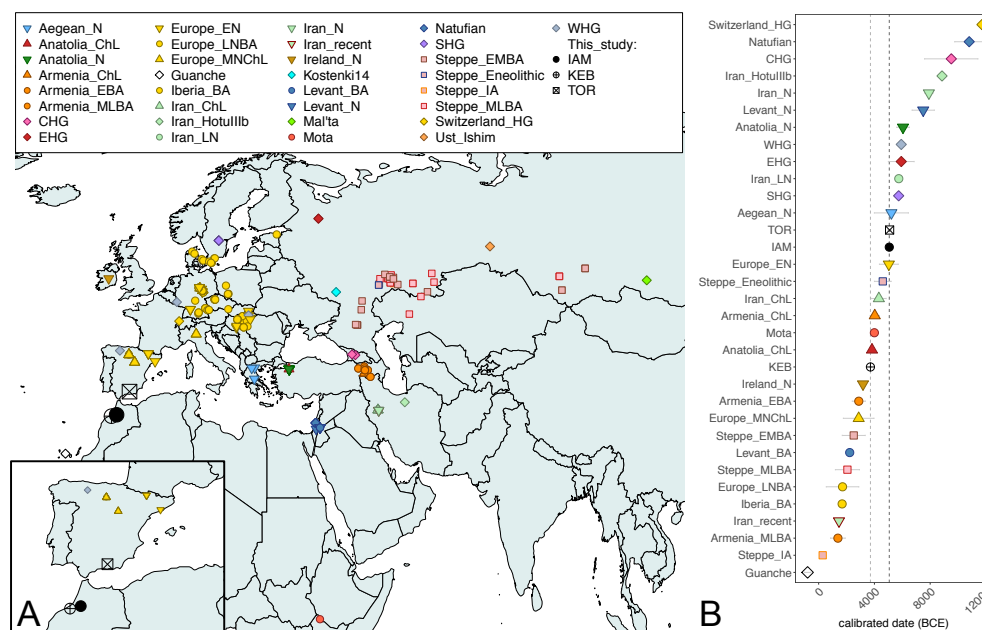
436

437

438 Figure 2. Ancestry inference in ancient samples from North Africa and the Iberian

439 Peninsula. (A) PCA analysis using the Human Origins panel, (B) ADMIXTURE

440 analysis using the HGDP-MEGA dataset (K=5), (C) ADMIXTURE analysis using

441 the Human Origins dataset for modern and ancient populations (K=8), and (D)

442 detailed ADMIXTURE analysis for European Neolithic samples (K=8).

443

444

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

446   f3-statistic for KEB (D).

447

448

449

Table 1 - Summary statistics for North African and Iberian samples.

Sample	Origin	2 sigma calibrated radiocarbon date (BCE)	MtDNA coverage	MtDNA haplogroup	Contamination estimation	Molecular sex	Y-chromosome haplogroup	SNPs in MEGA dataset	SNPs in Human Origins dataset	MEGA coverage	Genome-wide coverage
IAM.3	Ifri n'Amr or Moussa	5325 - 5210	5X	M1b1	6.42% (21.49% - 1.90%)	XX	-	26,798	5,353	0.04X	0.01X
IAM.4	Ifri n'Amr or Moussa	5215 - 5005	52X	U6a1b	4.66% (6.64% - 3.26%)	XY	E-L19*	62,729	17,093	0.12X	0.02X
IAM.5	Ifri n'Amr or Moussa	5199 - 5176, 5066 - 4942	82X	U6a1b	3.68% (5.26% - 2.62%)	XY	E-L19*	408,744	208,661	1.72X	0.74X
IAM.6	Ifri n'Amr or Moussa	5290 - 5265, 5230 - 5195, 5180 - 5060	21X	U6a7b2	5.28% (8.48% - 3.02%)	XX	-	90,226	18,285	0.17X	0.02X
IAM.7	Ifri n'Amr or Moussa	5000 - 4840	129X	U6a3	2.35% (3.39% - 1.67%)	XY	-	177,041	45,134	0.38X	0.06X
KEB.1	Kelif el Boroud		18X	X2b	3.03% (6.31% - 1.23%)	XX	-	96,946	39,312	0.16X	0.08X
KEB.3	Kelif el Boroud		20X	K1a1b1	3.16% (5.73% - 1.72%)	-	-	-	-	-	-
KEB.4	Kelif el Boroud	3780 - 3650	135X	K1a1b1	2.49% (3.27% - 1.89%)	XX	-	304,607	77,964	0.64X	0.12X
KEB.6	Kelif el Boroud		14X	K1a4a1	3.77% (6.72% - 1.98%)	XY	T-M184	95,152	52,694	0.16X	0.14X
KEB.7	Kelif el Boroud		12X	T2b3	1.44% (4.31% - 0.31%)	XY	-	-	-	-	-
KEB.8	Kelif el Boroud		23X	X2b	3.03% (6.23% - 1.27%)	XX	-	34,385	18,974	0.06X	0.05X
TOR.1	El Toro	5280 - 4780	170X	T2c1d	1.42% (2.07% - 0.98%)	-	-	-	-	-	-
TOR.5	El Toro	5040 - 4850	18X	J2b1a	4.47% (8.76% - 1.96%)	XY	G-M201	-	-	-	-
TOR.6	El Toro		53X	T2b3	4.15% (5.89% - 2.94%)	XX	-	318,110	101,698	0.67X	0.17X
TOR.7	El Toro		126X	T2b3	7.06% (8.36% - 5.88%)	XX	-	115,652	24,719	0.18X	0.03X
TOR.8	El Toro	5280 - 4780	234X	K1a1	3.95% (4.80% - 3.26%)	XX	-	212,370	46,364	0.38X	0.06X
TOR.11	El Toro		124X	K1a2a	3.88% (5.00% - 2.98%)	XX	-	80,687	17,306	0.15X	0.03X
TOR.12	El Toro		174X	J2b1a	1.50% (2.30% - 0.93%)	XY	-	-	-	-	-