Report

Current Biology

Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula

Highlights

- Iberian hunter-gatherers show dual Late Pleistocene genetic ancestry
- Dual hunter-gatherer ancestry is the result of admixture from different refugia
- This mixed Late Pleistocene ancestry can be traced in Iberian Neolithic farmers

Authors

Vanessa Villalba-Mouco, Marieke S. van de Loosdrecht, Cosimo Posth, ..., Pilar Utrilla, Johannes Krause, Wolfgang Haak

Correspondence

haak@shh.mpg.de

In Brief

Villalba-Mouco et al. generate genome-wide data from prehistoric Iberian hunter-gatherers and early farmers and show that two lineages of Late Pleistocene ancestry survived in Iberian foragers, likely as a result of admixture from different southern refugia. Newly arrived Neolithic farmers retain this genetic signature, suggesting local admixture.

Villalba-Mouco et al., 2019, Current Biology *29*, 1–9
April 1, 2019
https://doi.org/10.1016/j.cub.2019.02.006

CellPress

Please cite this article in press as: Villalba-Mouco et al., Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula, Current Biology (2019), https://doi.org/10.1016/j.cub.2019.02.006

Current Biology
Report

Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula

Vanessa Villalba-Mouco,[1,2] Marieke S. van de Loosdrecht,[1] Cosimo Posth,[1] Rafael Mora,[3] Jorge Martínez-Moreno,[3] Manuel Rojo-Guerra,[4] Domingo C. Salazar-García,[5] José I. Royo-Guillén,[6] Michael Kunst,[7] Hélène Rougier,[8] Isabelle Crevecoeur,[9] Héctor Arcusa-Magallón,[10] Cristina Tejedor-Rodríguez,[11] Iñigo García-Martínez de Lagrán,[12] Rafael Garrido-Pena,[13] Kurt W. Alt,[14,15] Choongwon Jeong,[1] Stephan Schiffels,[1] Pilar Utrilla,[2] Johannes Krause,[1] and Wolfgang Haak[1,16,*]

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Kahlaische Straße 10, 07745 Jena, Germany
[2]Departamento de Ciencias de la Antigüedad, Grupo Primeros Pobladores del Valle del Ebro (PPVE), Instituto de Investigación en Ciencias Ambientales (IUCA), Universidad de Zaragoza, Pedro Cerbuna, 50009 Zaragoza, Spain
[3]Centre d'Estudis del Patrimoni Arqueològic de la Prehistòria (CEPAP), Facultat de Lletres, Universitat Autònoma Barcelona, 08190 Bellaterra, Spain
[4]Department of Prehistory, University of Valladolid, Plaza del Campus, 47011 Valladolid, Spain
[5]Grupo de Investigación en Prehistoria IT-622-13 (UPV-EHU)/IKERBASQUE-Basque Foundation for Science, Euskal Herriko Unibertsitatea, Francisco Tomas y Valiente s/n., 01006 Vitoria, Spain
[6]Dirección General de Cultura y Patrimonio, Gobierno de Aragón, Avenida de Ranillas, 5 D., 50018 Zaragoza, Spain
[7]Instituto Arqueológico Alemán, Calle Serrano 159, E-28002 Madrid, Spain
[8]Department of Anthropology, California State University, Northridge, Northridge, California 91330, USA
[9]Université de Bordeaux, CNRS, UMR 5199-PACEA, 33615 Pessac Cedex, France
[10]Arcadia-FUNGE, Fundación General de la Universidad de Valladolid, 47002 Valladolid, Spain
[11]Juan de la Cierva-Formación Program, Institute of Heritage Sciences, Spanish National Research Council (Incipit, CSIC), Av. de Vigo, 15705 Santiago de Compostela, Spain
[12]Juan de la Cierva-Incorporación Program, Department of Prehistory, Valladolid University, Plaza del Campus, 47011 Valladolid, Spain
[13]Department of Prehistory, Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid, Spain
[14]Center of Natural and Cultural Human History, Danube Private University, Steiner Landstr. 124, A-3500 Krems, Austria
[15]Department of Biomedical Engineering, University of Basel, Gewerbestrasse 14-16, CH-4123 Allschwil, Switzerland
[16]Lead Contact
*Correspondence: haak@shh.mpg.de
https://doi.org/10.1016/j.cub.2019.02.006

SUMMARY

The Iberian Peninsula in southwestern Europe represents an important test case for the study of human population movements during prehistoric periods. During the Last Glacial Maximum (LGM), the peninsula formed a periglacial refugium [1] for hunter-gatherers (HGs) and thus served as a potential source for the re-peopling of northern latitudes [2]. The post-LGM genetic signature was previously described as a cline from Western HG (WHG) to Eastern HG (EHG), further shaped by later Holocene expansions from the Near East and the North Pontic steppes [3–9]. Western and central Europe were dominated by ancestry associated with the ~14,000-year-old individual from Villabruna, Italy, which had largely replaced earlier genetic ancestry, represented by 19,000–15,000-year-old individuals associated with the Magdalenian culture [2]. However, little is known about the genetic diversity in southern European refugia, the presence of distinct genetic clusters, and correspondence with geography. Here, we report new genome-wide data from 11 HGs and Neolithic individuals that highlight the late survival of Paleolithic ancestry in Iberia, reported previously in Magdalenian-associated individuals. We show that all Iberian HGs, including the oldest, a ~19,000-year-old individual from El Mirón in Spain, carry dual ancestry from both Villabruna and the Magdalenian-related individuals. Thus, our results suggest an early connection between two potential refugia, resulting in a genetic ancestry that survived in later Iberian HGs. Our new genomic data from Iberian Early and Middle Neolithic individuals show that the dual Iberian HG genomic legacy pertains in the peninsula, suggesting that expanding farmers mixed with local HGs.

RESULTS AND DISCUSSION

We successfully generated autosomal genome-wide data and mitochondrial genomes of 10 new individuals from key sites in the Iberian Peninsula ranging from ~13,000–6,000 calibrated years before present (years cal BP): Late Upper Paleolithic (n = 2), Mesolithic (n = 1), Early Neolithic (EN; n = 4), and Middle Neolithic (n = 3) (Figure 1, Data S1A, and STAR Methods). We furthermore improved the sequencing depth of one Upper Paleolithic individual from Troisième caverne of Goyet (Belgium) dated to ~15,000 years cal BP and associated with the Magdalenian culture [2]. For each individual, we generated multiple DNA libraries with unique index pairs [10, 11] for next-generation

Figure 1. Geo-chronological Location of Ancient Individuals from the Iberian Peninsula

(A) Map showing the geographical location of the new individuals and sites included in this study (black outlines) and relevant published data for HGs and EN/MN individuals from the Iberian Peninsula (no outlines).

(B) Radiocarbon dates of newly reported individuals in years cal BP (error bars indicate the 2-sigma range).

See also Data S1.

sequencing from ancient DNA (aDNA) extracted from teeth and bones [12] (Data S1B and STAR Methods). These were subsequently enriched using targeted in-solution capture for ~1240K informative nuclear SNPs [13], and independently for the complete mitogenome [14], and sequenced on Illumina platforms. All libraries contained short DNA fragments (46–65 bp length on average) with post mortem deamination patterns characteristic for aDNA (4%–16% for partial uracil-DNA glycosylase [UDG] and 19%–31% for non-UDG treated libraries at the first base pair position; Data S1C and S1D and STAR Methods). We estimated contamination rates from nuclear DNA in males to be 1.0%–3.4% for final merged libraries using ANGSD method 2 [15] and for mitogenomes in both sexes to be 0.14%–2.20% using ContamMix [16] (Data S1E and S1F and STAR Methods). After quality filtering, we obtained an endogenous DNA content of 1.3%–29.5% on the targeted 1240K SNPs (Data S1C and STAR Methods). We called pseudo-haploid genotypes for each individual (merged libraries) by randomly choosing a single base per site and intersected our data with a set of global modern populations genotyped for ~1240K nuclear SNP positions [17], including published ancient individuals from [2, 5, 7–9, 13, 14, 18–25]. The final dataset from the newly reported individuals contained 19,269–814,072 covered SNPs (Data S1G and STAR Methods). For principal-component analysis (PCA), we intersected our new data and published ancient individuals with a panel of worldwide modern populations genotyped on the Affymetrix Human Origins (HO) array [26].

Genetic Structure in Iberian Hunter-Gatherers

To characterize the genetic differentiation between HG individuals, we calculated genetic distances, defined as 1 – *f3,* where *f3* denotes the *f3*-outgroup statistics [26], for pairwise comparisons among all published and newly generated HG and visualized the results using multidimensional scaling (MDS) (STAR Methods, Figure 2A and STAR Methods). The hunter-gatherer (HG) individuals form distinguishable clusters on the MDS plot, supported by *f4*-statistics and clustering analysis (Figure S1 and STAR Methods), which we label in line with Fu et al. [2] as *Villabruna*, *Věstonice*, *Satsurblia*, and *Mal'ta* clusters, respectively. Henceforth, we present genetic clusters in italics and individuals in normal print. We introduce the *GoyetQ2* cluster (based on the highest genomic coverage) representing the Magdalenian-associated individuals Goyet Q-2, Hohle Fels 49, Rigney 1, El Mirón, and Burkhardtshöhle. With the newly generated data, we notice that Iberian HGs form a cline between the *GoyetQ2* and *Villabruna* clusters (Figures 2A and S2). This cline also includes El Mirón (the oldest individual from Iberia), which had previously been considered representing its own *El Mirón* cluster together with all individuals of the *GoyetQ2* cluster (yellow symbols in Figure 2A) [2]. Here, Canes 1 and La Braña 1 (Mesolithic individuals from the Cantabrian region in northern Iberia) are falling closer to the *Villabruna* cluster, while Chan (northwestern Iberia) and our newly reported individuals from Moita do Sebastião (Portuguese Atlantic coast) and Balma Guilanyà (Pre-Pyrenean region, northeastern Iberia) have more affinity with El Mirón, which is in turn closer to the Magdalenian *GoyetQ2* cluster.

This observation is confirmed by *f4*-statistics of the form *f4(GoyetQ2, Villabruna; test, Mbuti)*, which measures whether a test population shares more genetic drift with Goyet Q-2 than with the Villabruna individual. Three Iberian HGs (Chan, Moita do Sebastião, and El Mirón), as well as Hohle Fels 49 and the 35,000-year-old Goyet Q116-1, show significantly positive *f4*-values, indicating that these individuals shared more ancestry with Goyet Q-2 than with Villabruna (Figure 2B). This

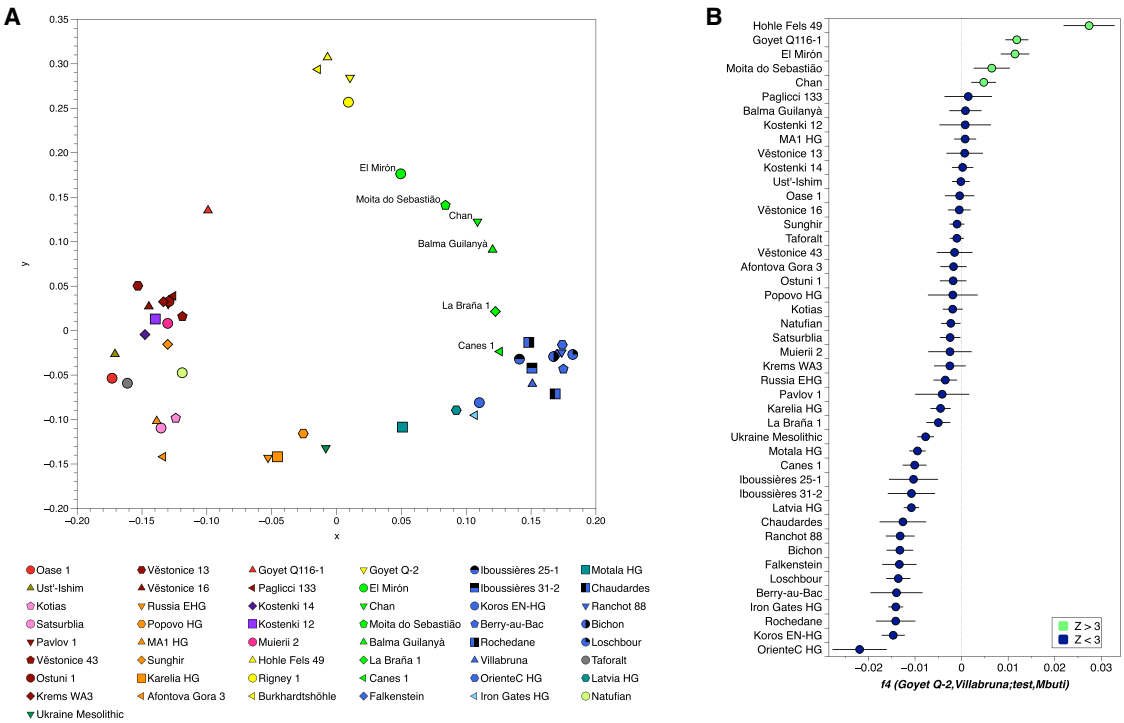

Figure 2. Genetic Distances between European HGs and Key *f4*-Statistics

(A) MDS plot of genetic distances between Eurasian HG individuals (>30,000 SNPs). The main genetic clusters defined previously [2] are: *Vĕstonice* (dark red), *Mal'ta* (orange: MA1 and Afontova Gora 3), *Satsurblia* (light pink: Kotias and Satsurblia), *Villabruna* (blue: Koros EN-HG, Berry-au-Bac 1, Rochedane, Villabruna, Chaudardes 1, Ranchot 88, La Braña 1, Loschbour), and *GoyetQ2* (yellow; newly defined). Iberian HGs are shown as green symbols.

(B) *f4*-statistics highlighting the excess affinity to Goyet Q-2 in Iberian and European HGs (>20,000 SNPs; error bars indicate ± 3 SE; $Z > 3$ [green]).

See also Figures S1 and S2.

heterogeneity in Iberian HGs cannot be explained by genetic drift alone (against which this type of F-statistics is robust) but only by admixture between two sources related to Goyet Q-2 and Villabruna, respectively. We visualize this admixture cline using contrasting *f3*-outgroup statistics of the form *f3(GoyetQ2; test, Mbuti)* and *f3(Villabruna; test, Mbuti)* (Figure 3A). The individuals from the *Villabruna* cluster deviate from the symmetry line x = y toward the y axis, expectedly, indicating excess genetic drift shared with Villabruna. In contrast, individuals of the *GoyetQ2* cluster deviate from the symmetry line x = y toward the x axis, indicating excess genetic drift with Goyet Q-2. Iberian HGs fall between the two clusters, which is inconsistent with them forming a clade with either group, but can only be explained by admixture. Here, Iberian HG Canes 1 and La Braña 1 share more Villabruna-like ancestry while El Mirón, Moita do Sebastião, and Chan share more Goyet Q-2-like ancestry.

To further confirm the potential admixture of El Mirón, we used *f4(GoyetQ2 cluster, El Mirón; Villabruna, Mbuti)* to test if Magdalenian-associated individuals were cladal with El Mirón. Here, we obtained significantly negative *Z* scores for Hohle Fels 49, Goyet Q-2, and Burkhardtshöhle. Among these, Goyet Q-2 has the highest data quality and most negative *Z* score and thus represents the best proxy for the non-*Villabruna*-like ancestry proportion in individuals such as El Mirón ($Z = -6.82$) (Data S1H). Based on this observation, we used the test *f4(Goyet Q-2, GoyetQ2 cluster; Villabruna, Mbuti)*, for which El Mirón is significantly negative (Figure 3B and Data S1H), confirming shared ancestry with Villabruna.

To show that the affinity of El Mirón with the Villabruna individual cannot be explained by El Mirón representing a basal split from Villabruna and the other *GoyetQ2* individuals, we used the test *f4(GoyetQ2 cluster, Villabruna; El Mirón, Mbuti)*. Here, all individuals of the *GoyetQ2* cluster are significantly positive, indicating that this cluster does not represent a sister branch of Villabruna and that El Mirón is not an outgroup to both the *Villabruna* and *GoyetQ2* clusters (Figure 3C and Data S1H). The mixed ancestry of El Mirón could also explain its reduced affinity to Goyet Q116-1 when compared to the younger *GoyetQ2* cluster in the test *f4(GoyetQ2 cluster, Goyet Q-2; Goyet Q116-1, Mbuti)* (Data S1H).

Having established two potential Paleolithic source populations surviving in Iberia from ~19,000 years cal BP onward, we used the admixture modeling programs *qpWave* and *qpAdm* (Figure 3D and STAR Methods) to explore the ancestry of all Iberian HGs. We used Villabruna and Goyet Q-2 as ultimate sources to model the dual ancestry in European HGs relative to outgroups that can distinguish these two sources from shared deeper ancestries (STAR Methods). Our two-source admixture model provides a good fit for the genetic profiles of most European HGs and is consistent with the cline between Villabruna- and Goyet Q-2-like ancestries described above (Figure 3D and STAR Methods). Here, Villabruna-like ancestry is the dominant component (69.8% ± 4.3% – 100%) in individuals of the *Villabruna* cluster,

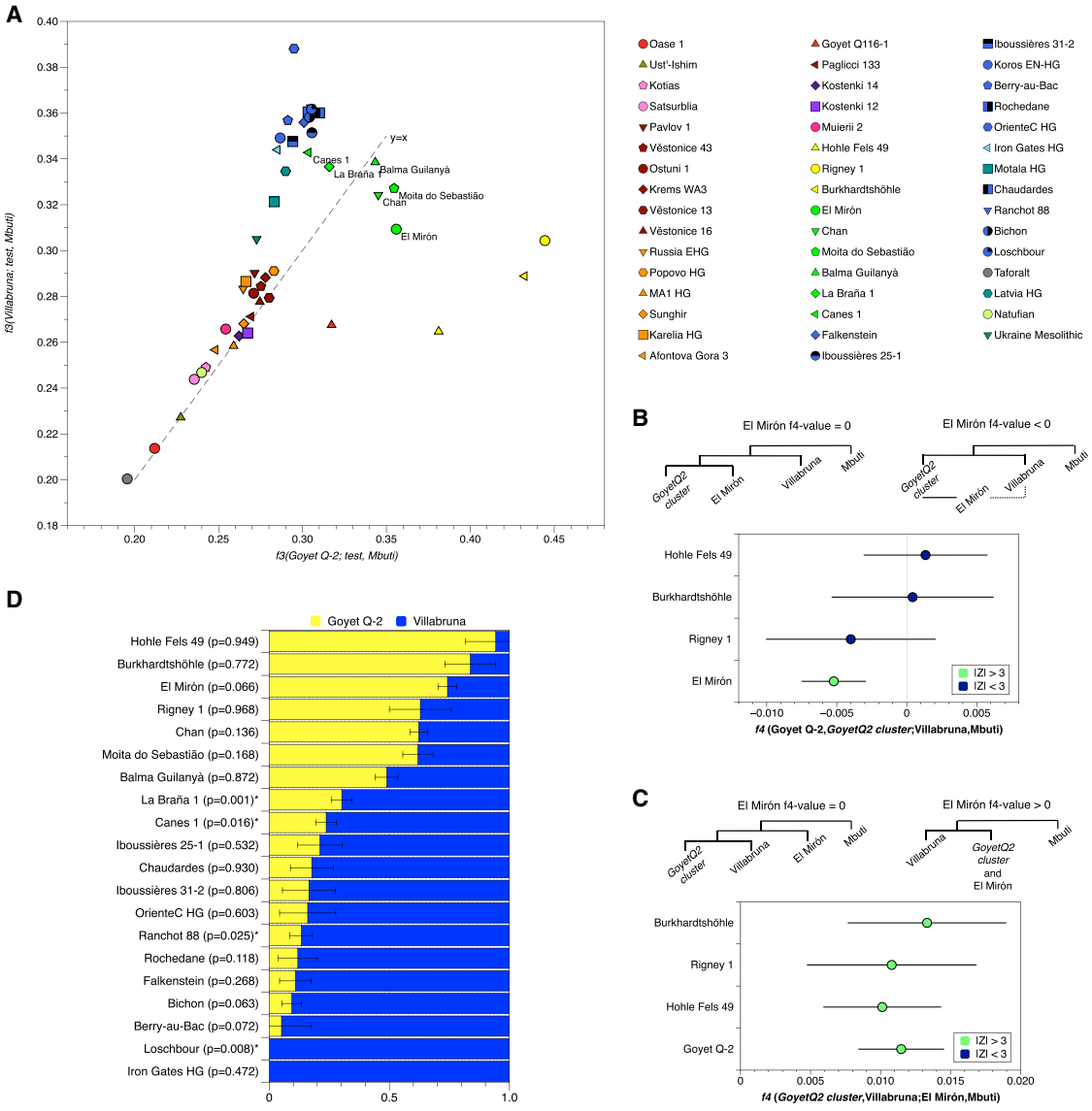

Figure 3. Key *f3*-Outgroup Tests, *f4*-Statistics, and *qpAdm* Results

(A) Biplot of *f3*-outgroup tests illustrating the Villabruna-like and Goyet Q-2-like ancestries in European HGs. The x = y axis marks full symmetry between Goyet Q-2- and Villabruna-like ancestries, and deviations mark excess ancestry shared with Goyet Q-2 (yellow) or Villabruna (blue).

(B) Results of *f4*-statistics highlighting the shared genetic drift between El Mirón and Villabruna individuals (≥20,000 SNPs; error bars indicated ±3 SE).

(C) Results of *f4*-statistics showing that El Mirón is not a sister clade of Villabruna (error bars indicated ±3 SE).

(D) Modeling European HGs as a two-way admixture of Villabruna- and Goyet Q-2-like ancestry (error bars indicated ±1 SE).

See also Figure 3B and Data S1.

and Goyet Q-2-like ancestry the dominant component (61.9% ± 6.3% – 94.3% ± 5.7%) in *GoyetQ2* cluster individuals (Figure 3D and Data S1I). These results underline the power of our outgroups and choice of proxies to differentiate Goyet Q-2- from Villabruna-like ancestry within our test individuals (STAR Methods; see Figure S3A for a replication with more proximal sources El Mirón and Loschbour). Congruent with the pattern observed in MDS (Figure 2A), clustering analysis (Figure S1), PCA (Figure S2), F-statistic-based tests (Figure 2B), and the biplot of *f3*-outgroup tests (Figure 3B), the two-source admixture model assigns a higher proportion of Goyet Q-2-like ancestry to Iberian HGs (ranging from 23.7% to 75.3%) than to contemporaneous Western HG (WHG) outside of Iberia. In fact, Balma Guilanyà, La Braña 1, and Canes 1 show elevated Villabruna admixture proportions but still higher Goyet Q-2 proportions than non-Iberian HGs (Figure 3D and Data S1I). We notice an additional contribution of Villabruna-like ancestry in the 12,000-year-old Balma Guilanyà individual from northeastern Iberia. Villabruna-like ancestry becomes even stronger during the Mesolithic in the Cantabrian region (La Braña 1 and Canes 1), suggesting extra HG flux into north/northeastern Iberia, which must have had a higher impact in this region. We were able to track a correlation between increasing Villabruna-like ancestry

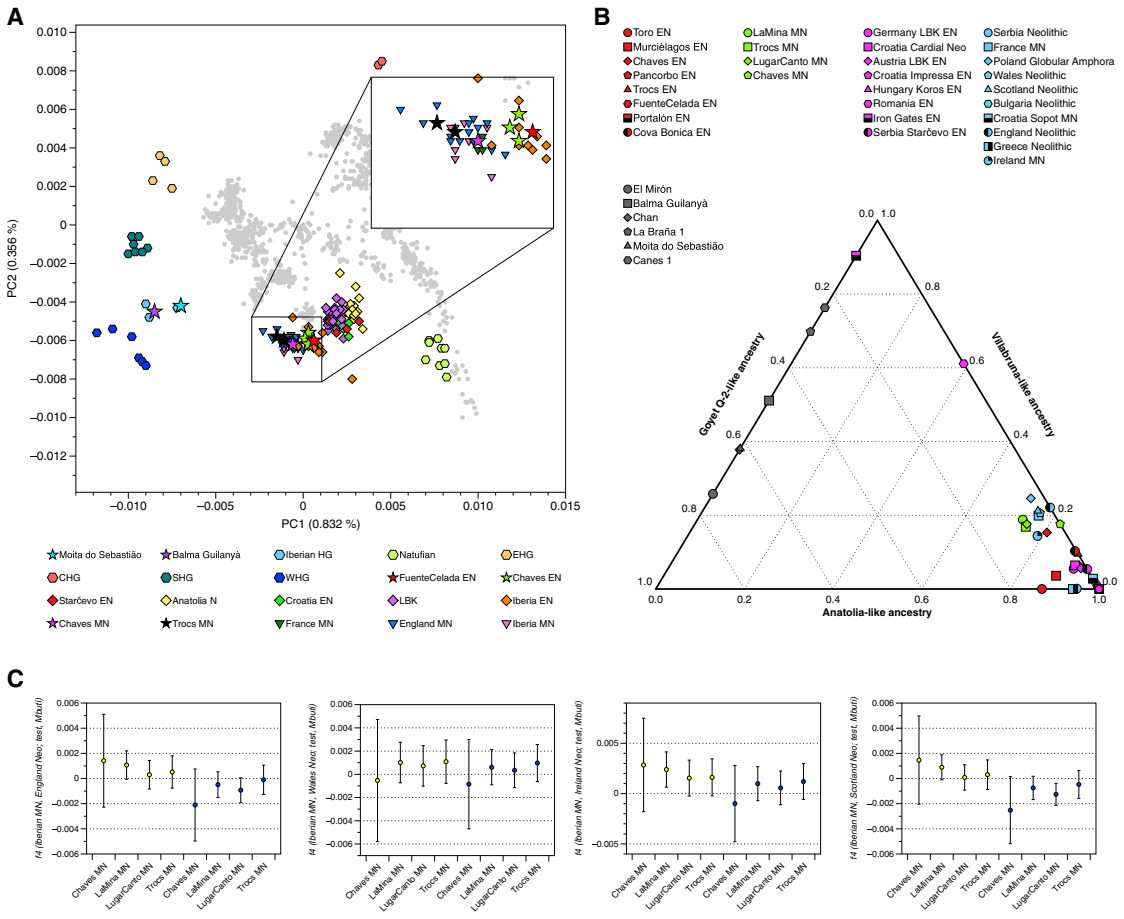

Figure 4. PCA Results and *qpAdm* Admixture Models

(A) Published ancient and newly reported individuals (stars) projected onto 777 present-day West Eurasians.

(B) Modeling EN and MN populations from Iberian and western Europe as admixture of three ancestral sources: Anatolian Neolithic, Goyet Q-2, and Villabruna (Data S1J).

(C) *f4(Iberian MN, Neolithic British Isles, test, Mbuti)*, where test is Goyet Q-2 (yellow) or Villabruna (blue), highlighting the excess of Goyet Q-2 ancestry in Iberian MN compared to Neolithic England and Scotland (error bars indicated ±3 SE). The affinity to Villabruna is shown for comparison to avoid potential bias created by unspecified HG attraction.

See also Data S1.

and time in this region (Figure S3A), while Mesolithic HGs outside this region (Chan and Moita do Sebastiao) retain more *GoyetQ2* ancestry and do not fit this pattern (Figures S3A and S3C). Interestingly, we find no traces of African ancestry (Figure S3B and STAR Methods).

Dual Hunter-Gatherer Genetic Legacy in Iberian Neolithic Individuals

During the Neolithic transition ~7,600 years ago, human expansions reached the Iberian Peninsula relatively swiftly via expanding early farmers from western Anatolia [3–5]. The rapid expansion of EN individuals associated with farming practices across Europe resulted in a relatively low genetic variability in the reported Neolithic genomes, which makes it difficult to distinguish between the Mediterranean and Danubian routes of expansion of Neolithic lifeways [27, 28]. However, Olalde et al. [9] noted subtle regional differences between WHG individuals and used the proportion of HG ancestry from La Braña 1 in Neolithic Iberians to trace the expansion from southwestern Europe along the Atlantic coast to Britain. This movement corresponds well with the megalithic burial practices of these regions observed in the archaeological record [29, 30]. Under the assumption that these proportions reflect one, or potentially more, local admixture events along the routes of expansion, it is thus possible to distinguish Neolithic groups by their varying autochthonous HG signatures [18].

Given the presence of two ancestral lineages in Iberian HGs, we thus explored this potential genetic legacy in our newly generated EN and Middle Neolithic (MN) individuals. We first used PCA to assess the genetic affinities qualitatively (Figure 4A). Here, the new Neolithic Iberian individuals spread along a cline from Neolithic Anatolia to WHG, on which the new individuals cluster with contemporaneous Iberian Neolithic individuals [5, 7, 8, 13, 24, 25]. As shown before, MN individuals are shifted toward WHG individuals [3], including the newly reported MN individuals from Cova de Els Trocs and Cueva de Chaves.

Using *qpAdm* models consistent with those above (Figures 3D and S3), we aimed to trace and quantify the proportion of Goyet Q-2- and Villabruna-like HG ancestry in EN and MN groups from Iberia and western/central Europe as a mixture of three ancestral sources: Anatolian Neolithic, Goyet Q-2, and Villabruna, respectively. We show that EN Iberians shared a higher proportion of Goyet Q-2-like ancestry than EN individuals from outside Iberia (Figure 4B and Data S1J). Goyet Q-2-like ancestry is higher in EN from southern Iberia (Andalusia), suggesting additional admixture with local Iberian HGs, who carried mixed Upper Paleolithic ancestry.

Goyet Q-2 ancestry is continuously detectable in all Iberian MN individuals, including broadly contemporaneous individuals from Scotland, Wales, Ireland, and France, but not in Neolithic England, for which the *qpAdm* model with three sources fails ($p = 6.91e{-}05$) in favor of two sources, despite being poorly supported ($p = 0.001$) (Data S1J). Goyet Q-2-like ancestry is, however, highest in all Iberian MN (except Chaves MN) when compared with other MN populations that share a similar overall amount of HG ancestry (Figures 4B and 4C). We note the presence of Goyet Q-2 ancestry in MN Trocs, where this ancestry was not observed during the EN, but importantly also in MN individuals from France and Globular Amphora from Poland.

Olalde and colleagues [9] reported an elevated signal of La Braña 1 ancestry in Neolithic individuals from Wales and England (using KO1 HG from Hungary and Anatolian Neolithic as the other two sources) and thus argued for an Iberian contribution to the Neolithic in Britain [9]. We replicated these findings by using similar sources (El Mirón instead of Goyet Q-2; Data S1J), showing that these results are sensitive to the source populations used. However, our models with Goyet Q-2 as ultimate source highlight not only the admixed nature of La Braña 1 and El Mirón, but also that Goyet Q-2-like ancestry in MN individuals outside Iberia hints at Iberia as one possible source, but not the exclusive source, of the Neolithic in Britain. Further sampling from regions in today's France, the Netherlands, Belgium, Luxembourg, and Germany is needed to answer this question.

Conclusions

Our results highlight the unique genetic structure observed in Iberian HG individuals, which results from admixture of individuals related to the *GoyetQ2* and *Villabruna* clusters. This suggests a survival of two lineages of Late Pleistocene ancestry in Holocene western Europe, in particular the Iberian Peninsula, whereas HG ancestry in most other regions was largely replaced by Villabruna-like ancestry. With an age estimate of ~18,700 years cal BP for the El Mirón individual, the oldest representative of this mixed ancestry, the timing of this admixture suggests an early connection (*terminus ante quem*) between putative ancestries from different LGM refugia. It is possible that Goyet Q-2 ancestry could have existed in Iberia in unadmixed form, where it was complemented by Villabruna ancestry as early as ~18,700 years ago. Alternatively, both Magdalenian-associated Goyet Q-2 and Villabruna ancestries originated in regions outside Iberia and arrived in Iberia independently, where both lineages admixed, or had already existed in admixed form outside Iberia. Interestingly, the dual Upper Paleolithic ancestry was also found in EN individuals from the Iberian Peninsula, supporting the hypothesis of additional local admixture with resident HGs in Iberia during the time of the Mesolithic-Neolithic transition.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Archaeological sites and sample description
  - Ancient DNA processing and quality control
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Read processing and assessment of ancient DNA authenticity
  - Contamination tests
  - Genotyping and merging with dataset
  - Kinship relatedness and individual assessment
  - Phenotypic traits analysis
  - Mitochondrial and Y chromosome haplogroups
  - Population genetic analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found with this article online at https://doi.org/10.1016/j.cub.2019.02.006.

A video abstract is available at https://doi.org/10.1016/j.cub.2019.02.006#mmc4.

ACKNOWLEDGMENTS

We thank the members of the Archaeogenetics Department of the Max Planck Institute for the Science of Human History, especially Maïté Rivollat, Thiseas Lamnidis, Cody Parker, Rodrigo Barquera, Stephen Clayton, Aditya Kumar, and all technicians. We thank Iñigo Olalde for valuable comments on the manuscript. We are indebted to the Museo de Huesca, Patrick Semal, the Royal Belgian Institute of Natural Sciences, and all archaeologists involved in the excavations. The genetic research was funded by the Max Planck Society and the European Research Council ERC-CoG 771234 PALEoRIDER (W.H.). V.V.-M. was funded by a predoctoral scholarship of the Gobierno de Aragón and the Fondo Social Europeo (BOA20150701025) and a 3-month research stay grant (CH 76/16) by Programa CAI-Ibercaja de Estancias de Investigación. V.V.-M. and P.U. are members of the Spanish project HAR2014-59042-P (Transiciones climáticas y adaptaciones sociales en la prehistoria de la Cuenca del Ebro), and of the regional government of Aragon PPVE research group (H-07: Primeros Pobladores del Valle del Ebro). The Goyet project was funded by the Wenner-Gren Foundation (7837 to H.R.), the College of Social and Behavioral Sciences of CSUN, and the CSUN Competition for Research, Scholarship, and Creative Activity Awards.

AUTHOR CONTRIBUTIONS

V.V.-M., J.K., and W.H. conceived the study; R.M., J.M.-M., M.R.-G., D.C.S.-G., J.I.R.-G., M.K., H.R., I.C., H.A.-M., C.T.-R., I.G.-M.d.L., R.G.-P., K.W.A., and P.U. assembled samples and provided archaeological context; V.V.-M., M.S.v.d.L., and C.P. performed aDNA lab work and sequencing; V.V.-M., C.P., M.S.v.d.L., S.S., C.J., and W.H. analyzed data; and V.V.-M., C.P., M.S.v.d.L., and W.H. wrote the manuscript with input from all co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 16, 2018
Revised: January 4, 2019
Accepted: February 1, 2019
Published: March 14, 2019

REFERENCES

1. Stewart, J.R., and Stringer, C.B. (2012). Human evolution out of Africa: the role of refugia and climate change. Science *335*, 1317–1321.
2. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.
3. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.
4. Günther, T., and Jakobsson, M. (2016). Genes mirror migrations and cultures in prehistoric Europe-a population genomic perspective. Curr. Opin. Genet. Dev. *41*, 115–123.
5. Valdiosera, C., Günther, T., Vera-Rodríguez, J.C., Ureña, I., Iriarte, E., Rodríguez-Varela, R., Simões, L.G., Martínez-Sánchez, R.M., Svensson, E.M., Malmström, H., et al. (2018). Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. Proc. Natl. Acad. Sci. USA *115*, 3428–3433.
6. Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172.
7. Martiniano, R., Cassidy, L.M., Ó'Maoldúin, R., McLaughlin, R., Silva, N.M., Manco, L., Fidalgo, D., Pereira, T., Coelho, M.J., Serra, M., et al. (2017). The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. PLoS Genet. *13*, e1006852.
8. Olalde, I., Schroeder, H., Sandoval-Velasco, M., Vinner, L., Lobón, I., Ramirez, O., Civit, S., García Borja, P., Salazar-García, D.C., Talamo, S., et al. (2015). A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. Mol. Biol. Evol. *32*, 3132–3142.
9. Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. Nature *555*, 190–196.
10. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*.
11. Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. *40*, e3.
12. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. USA *110*, 15758–15763.
13. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.
14. Mittnik, A., Wang, C.-C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R., Khartanovich, V., Moiseyev, V., Tõrv, M., et al. (2018). The genetic prehistory of the Baltic Sea region. Nat. Commun. *9*, 442.
15. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*, 356.
16. Fu, Q., Mittnik, A., Johnson, P.L.F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. Curr. Biol. *23*, 553–559.
17. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.
18. Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., et al. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature *551*, 368–372.
19. Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., et al. (2018). The genomic history of southeastern Europe. Nature *555*, 197–203.
20. Cassidy, L.M., Martiniano, R., Murphy, E.M., Teasdale, M.D., Mallory, J., Hartwell, B., and Bradley, D.G. (2016). Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. Proc. Natl. Acad. Sci. USA *113*, 368–373.
21. Sikora, M., Seguin-Orlando, A., Sousa, V.C., Albrechtsen, A., Korneliussen, T., Ko, A., Rasmussen, S., Dupanloup, I., Nigst, P.R., Bosch, M.D., et al. (2017). Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. Science *358*, 659–662.
22. van de Loosdrecht, M., Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., Nickel, B., Nagel, S., Talbi, E.H., El Hajraoui, M.A., et al. (2018). Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. Science *360*, 548–552.
23. Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat. Commun. *6*, 8912.
24. González-Fortes, G., Jones, E.R., Lightfoot, E., Bonsall, C., Lazar, C., Grandal-d'Anglade, A., Garralda, M.D., Drak, L., Siska, V., Simalcsik, A., et al. (2017). Paleogenomic Evidence for Multi-generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin. Curr. Biol. *27*, 1801–1810.
25. Fregel, R., Méndez, F.L., Bokbot, Y., Martín-Socas, D., Camalich-Massieu, M.D., Santana, J., Morales, J., Ávila-Arcos, M.C., Underhill, P.A., Shapiro, B., et al. (2018). Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. Proc. Natl. Acad. Sci. USA *115*, 6774–6779.
26. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.
27. Manning, K., Timpson, A., Colledge, S., Crema, E., Edinborough, K., Kerig, T., and Shennan, S. (2014). The chronology of culture: a comparative assessment of European Neolithic dating approaches. Antiquity *88*, 1065–1080.
28. Perrin, T., Manen, C., Valdeyron, N., and Guilaine, J. (2018). Beyond the sea… The Neolithic transition in the southwest of France. Quat. Int. *470*, 318–332.
29. Sherratt, A. (1995). Instruments of conversion? The role of megaliths in the mesolithic/neolithic transition in Northwest Europe. Oxf. J. Archaeol. *14*, 245–260.
30. Masset, C. (1993). Les dolmens: Sociétés néolithiques, practiques funéraires (Paris: Ed. Errance).
31. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

32. Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biol. *17*, 60.

33. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.

34. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. *44* (W1), W58-63.

35. Poznik, G.D. (2016). Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. bioRxiv. https://doi.org/10.1101/088716.

36. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

37. Monroy Kuhn, J.M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. PLoS ONE *13*, e0195491.

38. Terradas, X., Pallarés, M., Mora, R., and Moreno, J.M. (1993). Estudi preliminar de les ocupacions humanes de la balma de Guilanyà (Navès, Solsonès). Rev. d'Arqueologia Ponent, 231–248.

39. Martínez-Moreno, J., Mora, R., and Casanova, J. (2006). Balma Guilanyà y la ocupación de la vertiente sur del Prepirineo del Noreste de la Península Ibérica durante el Tardiglaciar. In La cuenca mediterránea durante el paleolítico superior: 38.000-10.000 años, pp. 444–457.

40. Martínez-Moreno, J., Mora, R., and Casanova, J. (2007). El contexto cronométrico y tecno-tipológico durante el Tardiglaciar y Postglaciar de la viertiente sur de los Pirineos orientales. Rev. d'Arqueologia Ponent, 7–44.

41. Martínez-Moreno, J., and Mora, R. (2009). Balma Guilanyà (Prepirineo de Lleida) y el Aziliense en el noreste de la Península Ibérica. Trabajos de Prehistoria *66*, 45–60.

42. Garcia-Guixé, E., Martínez-Moreno, J., Mora, R., Núñez, M., and Richards, M.P. (2009). Stable isotope analysis of human and animal remains from the Late Upper Palaeolithic site of Balma Guilanyà, southeastern Pre-Pyrenees, Spain. J. Archaeol. Sci. *36*, 1018–1026.

43. Ruiz, J., García-Sívoli, C., Martínez-Moreno, J., and Subirá, M.E. (2006). Los restos humanos del Tardiglaciar de Balma Guilanyà. In La cuenca mediterránea durante el paleolítico superior: 38.000-10.000 años, pp. 458–467.

44. Martzluff, M., Martínez-Moreno, J., Guilaine, J., Mora, R., and Casanova, J. (2012). Transformaciones culturales y cambios climáticos en los Pirineos catalanes entre el Tardiglaciar y Holoceno antiguo: Aziliense y Sauveterriense en Balma de la Margineda y Balma Guilanyà. Cuaternario y Geomorfología *26*, 61–78.

45. Straus, L.G. (2015). Chronostratigraphy of the Pleistocene/Holocene boundary: the Azilian problem in the Franco-Cantabrian region. Palaeohistoria *27*, 89–122.

46. Szécsényi-Nagy, A., Roth, C., Brandt, G., Rihuete-Herrada, C., Tejedor-Rodríguez, C., Held, P., García-Martínez-de-Lagrán, Í., Arcusa Magallón, H., Zesch, S., Knipper, C., et al. (2017). The maternal genetic make-up of the Iberian Peninsula between the Neolithic and the Early Bronze Age. Sci. Rep. *7*, 15644.

47. Lubell, D., Jackes, M., Schwarcz, H., Knyf, M., and Meiklejohn, C. (1994). The Mesolithic-Neolithic transition in Portugal: isotopic and dental evidence of diet. J. Archaeol. Sci. *21*, 201–216.

48. Jackes, M., and Alvim, P. (1999). Reconstructing Moita do Sebastião, the first step. In Do Epipaleolítico ao Calcolítico na Península Iberica, Actas do IV Congresso de Arqueologia Peninsular, pp. 13–25.

49. Bicho, N.F. (1994). The End of the Paleolithic and the Mesolithic in Portugal. Curr. Anthropol. *35*, 664–674.

50. Gronenborn, D. (2017). Migrations before the Neolithic? The Late Mesolithic blade-and-trapeze horizon in Central Europe and beyond. Migration and Integration from Prehistory to the Middle Ages (Halle, Germany: LandesMuseum fur Vorgechichte), pp. 113–122.

51. Perrin, T., Marchand, G., Allard, P., Binder, D., Collina, C., Garcia Puchol, O., and Valdeyron, N. (2009). Le second Mesolithique d'Europe occidentale : origines et gradient chronologique. Annales de la Fondation Fyssen *24*, 160–176.

52. Utrilla Miranda, P., and Laborda Lorente, R. (2018). La cueva de Chaves (Bastaras, Huesca): 15 000 años de ocupación prehistórica. Trabajos de Prehistoria *75*, 248–269.

53. Castaños, P.M. (2004). Estudio arqueozoológico de los macromamíferos del Neolítico de la Cueva de Chaves (Huesca: Saldvie Estud. Prehist. y Arqueol), pp. 125–172.

54. Baldellou Martínez, V. (2011). La Cueva de Chaves (Bastarás-Casbas, Huesca). SAGVNTVM *12*, 141–144.

55. Utrilla, P., and Baldellou, V. (2001). Cantos pintados neolíticos de la Cueva de Chaves (Bastarás, Huesca: Saldvie Estud. Prehist. y Arqueol), pp. 45–126.

56. Utrilla, P., and Baldellou, V. (2007). Les galets peints de la Grotte de Chaves. Bull. la Société Préhistorique Ariège-Pyrénées *62*, 73–88.

57. Zilhão, J. (2001). Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. Proc. Natl. Acad. Sci. USA *98*, 14180–14185.

58. Isern, N., Zilhão, J., Fort, J., and Ammerman, A.J. (2017). Modeling the role of voyaging in the coastal spread of the Early Neolithic in the West Mediterranean. Proc. Natl. Acad. Sci. USA *114*, 897–902.

59. Bernabeu, J., Balaguer, L.M., Esquembre-Bebiá, M.A., Pérez, J.R.O., and Soler, J.d.D.B. (2009). La cerámica impresa mediterránea en el origen del Neolítico de la península Ibérica. In De Méditerranée et d'ailleurs...: mélanges offerts à Jean Guilaine, pp. 83–96.

60. Martins, H., Oms, F.X., Pereira, L., Pike, A.W.G., Rowsell, K., and Zilhão, J. (2015). Radiocarbon dating the beginning of the Neolithic in Iberia: new results, new problems. J. Mediterr. Archaeol. *28*, 105–131.

61. Villalba-Mouco, V., Utrilla, P., Laborda, R., Lorenzo, J.I., Martínez-Labarga, C., and Salazar-García, D.C. (2018). Reconstruction of human subsistence and husbandry strategies from the Iberian Early Neolithic: A stable isotope approach. Am. J. Phys. Anthropol. *167*, 257–271.

62. Cuenca-Romero, M.d.C.A., Carmona Ballestero, E., Pascual Blanco, S., Martínez, Díez, G., and Díez Pastor, C. (2011). El "campo de hoyos" calcolítico de Fuente Celada (Burgos): datos preliminares y perspectivas. Complutum *22*, 47–69.

63. Rojo-Guerra, M., Peña-Chocarro, L., Royo-Guillén, J.I., Tejedor, C., García-Martínez de Lagrán, I., Arcusa, H., Garrido Pena, R., Moreno, M., Mazzucco, N., Gibaja, J.F., et al. (2013). Pastores trashumantes del Neolítico Antiguo en un entorno de alta montaña: secuencia crono-cultural de la Cova de Els Trocs (San Feliú de Veri, Huesca: BSAA Arqueol), pp. 9–55.

64. Posth, C., Nägele, K., Colleran, H., Valentin, F., Bedford, S., Kami, K.W., Shing, R., Buckley, H., Kinaston, R., Walworth, M., et al. (2018). Language continuity despite population replacement in Remote Oceania. Nat. Ecol. Evol. *2*, 731–740.

65. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. Philos. Trans. R. Soc. B Biol. Sci. *370*.

66. Maricic, T., Whitten, M., and Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS ONE *5*, e14004.

67. Fu, Q., Hajdinjak, M., Moldovan, O.T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. Nature *524*, 216–219.

68. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.

69. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

70. Gamba, C., Fernández, E., Tirado, M., Deguilloux, M.F., Pemonge, M.H., Utrilla, P., Edo, M., Molist, M., Rasteiro, R., Chikhi, L., and Arroyo-Pardo, E. (2012). Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. Mol. Ecol. *21*, 45–56.

71. Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. Nat. Commun. *5*, 5257.

72. Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-Del-Molino, D., van Dorp, L., López, S., Kousathanas, A., Link, V., et al. (2016). Early farmers from across Europe directly descended from Neolithic Aegeans. Proc. Natl. Acad. Sci. USA *113*, 6886–6891.

73. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.

74. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Ancient individual	This study/ Troisième caverne of Goyet archaeological site	Goyet Q-2
Ancient individual	This study/ Balma Guilanyà archaeological site	BAL001/ E1206 Shown to be identical with BAL005
Ancient individual	This study/ Balma Guilanyà archaeological site	BAL005/ BG E 3214 Shown to be identical with BAL001
Ancient individual	This study/ Balma Guilanyà archaeological site	BAL003/ E9605
Ancient individual	This study/ Moita do Sebastião archaeological site	CMS001/ 22
Ancient individual	This study/ Cueva de Chaves archaeological site	CHA001/ 84C
Ancient individual	This study/ Cueva de Chaves archaeological site	CHA002/ CH.NIG.11559
Ancient individual	This study/ Cueva de Chaves archaeological site	CHA003/ CH.NIG.11558
Ancient individual	This study/ Cueva de Chaves archaeological site	CHA004/ Ch.Banda13
Ancient individual	This study/ Fuente Celada archaeological site	FUC003/ H62 UE 622
Ancient individual	This study/ Cova de Els Trocs archaeological site	ELT002/ UE 69 C: 589 S:7 No Inv: 14227
Ancient individual	This study/ Cova de Els Trocs archaeological site	ELT006/ UE:1 C: 650 S:1 No Inv: 22404
Chemicals, Peptides, and Recombinant Proteins		
0.5 M EDTA pH 8.0	Life Technologies	AM9261
1x Tris-EDTA pH 8.0	AppliChem	A8569,0500
Proteinase K	Sigma-Aldrich	P2308-100MG
Guanidine hydrochloride	Sigma-Aldrich	G3272-500 g
3M Sodium Acetate pH 5,2	Sigma-Aldrich	S7899-500ML
Tween 20	Sigma-Aldrich	P9416-50ML
Water Chromasolv Plus	Sigma-Aldrich	34877-2.5L
Ethanol	Merck	1009832511
Isopropanol	Merck	1070222511
Buffer Tango	Life Technologies	BY5
T4 DNA Polymerase	New England Biosciences	M0203 L
T4 Polynucleotide Kinase	New England Biosciences	M0201 L
User Enzyme	New England Biosciences	M5505 L
Uracil Glycosylase inhibitor (UGI)	New England Biosciences	M0281 S
Bst 2.0 DNA Polymerase	New England Biosciences	M0537 S
BSA 20mg/mL	New England Biosciences	B9000 S
ATP	New England Biosciences	P0756 S
dNTPs 25 mM	Thermo Scientific	R1121
D1000 ScreenTapes	Agilent Technologies	5067-5582
D1000 Reagents	Agilent Technologies	5067-5583
Pfu Turbo Cx Hotstart DNA Polymerase	Agilent Technologies	600412

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Herculase II Fusion DNA Polymerase	Agilent Technologies	600679
1x TE-Puffer pH 8,0 low EDTA	AppliChem	A8569,0500
Sodiumhydroxide Pellets	Fisher Scientific	10306200
Sera-Mag Speed CM	GE Healthcare Lifescience	65152105050250
Dynabeads MyOne Streptavidin T1	Life Technologies	65601
GeneRuler Ultra Low Range DNA Ladder	Life Technologies	SM1211
10x GeneAmp PCR Gold Buffer and MgCl2	Life Technologies	4379874
Human Cot-1 DNA	Life Technologies	15279011
1M Tris-HCl pH 8.0	Life Technologies	15568025
20x SCC Buffer	Life Technologies	AM9770
UltraPure™ Salmon Sperm DNA Solution	Life Technologies	15632011
PEG 8000 Powder, Molecular Biology Grade	Promega	V3011
20% SDS Solution	Serva	39575.01
3M Sodium Acetate buffer solution pH 5,2	Sigma-Aldrich	S7899-500ML
5 M Sodium chloride solution	Sigma-Aldrich	S5150-1L
Denhardt's solution	Sigma-Aldrich	D9905-5Ml
Critical Commercial Assays		
High Pure Viral Nucleic Acid Large Volume Kit	Roche	5114403001
Quick Ligation Kit	New England Biosciences	M2200 L
MinElute PCR Purification Kit	QIAGEN	28006
DyNAmo Flash SYBR Green qPCR Kit	Life Technologies	F-415L
Oligo aCGH/Chip-on-Chip Hybridization Kit	Agilent Technologies	5188-5220
HighSeq 4000 SBS Kit	Ilumina	FC-410-1001/2
NextSeq 500/550 High Output Kit v2	Ilumina	FC-404-2002
Deposited Data		
Raw and analyzed data (European nucleotide archive)	This paper	ENA: PRJEB30985
Software and Algorithms		
Samtools	[31]	http://samtools.sourceforge.net/
EAGER	[32]	https://eager.readthedocs.io/en/latest/
ADMIXTOOLS	[25]	https://github.com/DReichLab/AdmixTools
smartpca	[33]	https://www.hsph.harvard.edu/alkes-price/software/
ANGSD	[15]	http://www.popgen.dk/angsd/index.php/Main_Page Contamination
Haplogrep 2	[34]	http://haplogrep.uibk.ac.at/
ContamMix	[16]	https://github.com/StanfordBioinformatics/DEFUNCT-env-modules/tree/master/contamMix
Yhaplo	[35]	https://github.com/23andMe/yhaplo
Geneious R8.1.974	[36]	https://www.geneious.com
READ	[37]	https://bitbucket.org/tguenther/read

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Wolfgang Haak (haak@shh.mpg.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The Iberian Peninsula in southwestern Europe is understood as a periglacial refugium for Pleistocene hunter-gatherers (HG) during the Last Glacial Maximum (LGM). The post-LGM genetic signature in western and central Europe was dominated by ancestry similar to the Villabruna individual, commonly described as WHG ancestry or '*Villabruna*' cluster [2]. This *Villabruna* cluster had largely

replaced the earlier *El Mirón* genetic cluster, comprised of 19,000-15,000-year-old individuals from central and western Europe associated with the Magdalenian culture [2].

By generating new genome-wide data from Belgian HG, Iberian HG and Neolithic individuals we aimed to further refine the HG genetic structure in the Iberian Peninsula during the Upper Paleolithic and Mesolithic, and to characterize the HG ancestry sources that contributed genetically to Neolithic groups. We hypothesize that (i) admixture events of different Upper Paleolithic HG ancestries resulted in genetic structure among various Iberian HG groups, which can be observed as asymmetric genetic affinities to the *Villabruna* and *El Mirón* cluster or another potential source, respectively, (ii) this structure during the Early Holocene is stronger in the Iberian Peninsula than in Central Europe (iii) the genetic structure was inherited by Neolithic individuals through mixture with local HG ancestry, especially in regions that initially were more affected by expanding farmers. All teeth and bone samples analyzed were obtained with relevant institutional permissions from the Gobierno de Aragón, Universitat Autònoma de Barcelona, Universidad de Valladolid, the German Archeological Institute Madrid.

Archaeological sites and sample description

Troisième caverne of Goyet (Upper Paleolithic)

The Troisième caverne of Goyet (Belgium) is a cave with an extensive Paleolithic record, from the Middle to the Upper Paleolithic periods (Aurignacian, Gravettian, and Magdalenian). The site was previously described and eight individuals were analyzed by Fu et al. [2]. For this study we have generated deeper sequencing data from individual Goyet Q-2 who is attributed to the Magdalenian period.

Goyet Q-2, juvenile individual (12,650 ± 50 BP [GrA-46168], 15,232–14,778 years cal BP [2-sigma value]) [2].

Balma Guilanyà (Late Upper Paleolithic)

Balma Guilanyà is a rock shelter located in Northeastern Iberia, at 1,150 m.a.s.l. (meters above sea level) in the Serra de Busa Pre-Pyrenean range (Navés, Lleida). After an initial test pit where Late Upper Paleolithic remains were recovered [38], the site was excavated between 2001 to 2008 [39, 40]. Two main chrono-cultural phases were defined. The oldest dates back to the Late Upper Paleolithic (15,000–11,000 years cal BP) and the youngest corresponds to the Early Mesolithic (11,000 – 9,500 years cal BP). The two chrono-cultural units are separated by a big fallen boulder which sealed the Late Upper Paleolithic levels [41]. A set of human skeletal remains were found under this big stone block without any evidence of funerary structures. Direct radiocarbon dates from two human remains (one human tooth and one human bone fragment) recovered from the same context dated to 13,380–12,660 years cal BP (Ua-34297) and 12,830–10,990 years cal BP (Ua- 34298) [42]. These dates fall inside the Bølling/Allerød interstadial and Younger Dryas stadial, which correspond to the Late Glacial. The Minimum Number of Individuals (MNI) was estimated to be three based on dental morphology: two adults and one immature individual [43]. The stable carbon and nitrogen isotope analyses performed on human bone collagen suggested a diet based on terrestrial herbivores, without any evidence of marine or freshwater resources [42]. The material cultural artifacts recovered from the same level as the human remains have been attributed to the Azilian techno-complex [41]. Balma Guilanyà shows clear technical parallels with the near Azilian site Balma Margineda [44]. However, in general the Azilian is considered to be more common in Vasco-Cantabrian northern Iberia and on the other side of the Pyrenees [45]. Here, we report the genome-wide data from two individuals from this site:

BAL0051, adult individual
BAL003, adult individual

Moita do Sebastião (Mesolithic)

This site was previously described in Szécsényi-Nagy et al. [46]. Moita do Sebastião is a Late Mesolithic shell midden site located in the Muge region (Salvaterra de Magos, Portugal) on the Atlantic coastline of Portugal. The Muge and Sado regions were very fertile estuaries and marshes during the Mesolithic, which were exploited by hunter-gatherers to obtain marine resources [47]. Although Mesolithic groups are not considered fully sedentary, Moita do Sebastião presents some cultural characteristics that suggest permanence at the site: the presence of post holes associated with hut building, and a big burial space [48]. These features have been interpreted as a systematic occupation of the estuarine areas, which is also reflected in the shell midden conformation. The Moita do Sebastião site was excavated by different archaeologists since the last century. The total minimum number of individuals (MNI) is unknown, but it could reach up to 100 individuals when summarizing the different campaigns [48]. The lithic assemblage of the Mesolithic phase is characterized by microburin technique and geometrics [49]. The Mesolithic geometric phenomenon is spread widely along eastern and western Europe and North Africa, but its origin is still debated [50]. Based on a chronological gradient an African origin was suggested, from where it spread into Europe through the South of Italy (Sicily) and from where it followed a Mediterranean expansion to Iberia [51]. In this study, we genetically analyze one Mesolithic individual from this site:

CMS001, adult individual, (7,240 ± 70 BP [To-131], 8,185–7,941 years cal BP [2-sigma value]) [46].

Cueva de Chaves (Early Neolithic)

Cueva de Chaves is located in Northeastern Iberia, at 663 m.a.s.l. in the Pre-Pyrenean mountain range of Sierra de Guara (Bastarás, Huesca). The site was excavated under the direction of Pilar Utrilla and Vicente Baldellou in between 1984 and 2007. Cueva de

Chaves was occupied during the Paleolithic, Neolithic, and sporadically during the Bronze Age and Late Roman periods. Neolithic deposits were divided in two archaeological levels and dated to the Early Neolithic period (Ia: 5,600–5,300 years cal BCE; Ib 5,300-5,000 years cal BCE) [52]. Both levels show a full Neolithic package consisting of domestic fauna [53], Cardial pottery [54] and schematic rock art painted on pebbles [55, 56]. The earliest Neolithic sites in the Iberian Peninsula are located in coastal areas [57, 58]. A long-standing hypothesis in archaeology to explain this is that the first arrival of the Neolithic in the Iberian Peninsula resulted from a Cardial expansion by a maritime route. In this context, Cueva de Chaves represents an interesting case study, because radiocarbon dates for the occupation of this cave overlap in time with other Cardial Early Neolithic sites in coastal Iberia [59, 60]. The pottery style, together with the radiocarbon dates, suggest an early expansion of the first farmers from coastal to the inland areas following the Ebro Basin [52]. An MNI of four individuals, directly radiocarbon dated, were recovered from this Early Neolithic context (although one radiocarbon date points back to the early Middle Neolithic). One of individuals was in a complete anatomical articulation. A human isotopic dietary study shows a high animal protein intake consumed by all individuals [61]. This was related to the existence of a specialized animal husbandry management community in which agriculture was not intensively developed. We included four Neolithic individuals for genetic analyses in this study:

CHA001, adult (6,230 ± 45 BP [GrA-26912], 7,257–7,006 years cal BP [2-sigma value]) [54].
CHA002, adult (6,227 ± 28 BP [MAMS 29127], 7,250–7,018 years cal BP [2-sigma value]) [61].
CHA003, infant (6,180 ± 54 BP [D-AMS 015821], 7,245–6,947 years cal BP [2-sigma value]) [61].
CHA004, adult (5,645 ± 31 BP [MAMS 28128], 6,494–6,321 years cal BP [2-sigma value]) [61].

Fuente Celada (Early Neolithic)

This site was described in Szécsényi-Nagy et al. [46]. Fuente Celada is an open-air settlement located in the northern Iberian Central Plateau (Quintanaduenas, Burgos). All the archaeological materials are from a rescue excavation carried out in 2008 by Alameda Cuenca-Romero et al. [62]. The site presents many negative structures, most of them from the Chalcolithic period, suggesting a habitat settlement. Some of these negative structures contain Chalcolithic humans remains. One of these burials gave an older date corresponding to the Early Neolithic. This burial contained an individual in a flexed position, with three bone rings close to the cervical vertebrae, which was interpreted as a necklace [62]. Here we include this Early Neolithic individual in our genetic analyses:

FUC003, adult (6,120 ± 30 BP [UGA-7565], 7,157–6,910 years cal BP [2-sigma value]) [62].

Cova de Els Trocs (Middle Neolithic)

This site was also described in Szécsényi-Nagy et al. [46]. Cova de Els Trocs is a cave located in Northeastern Iberia, at 1,564 m.a.s.l. in San Feliú de Veri (Bisaurri, Huesca) in the South of the Axial Pyrenees [63]. The excavation of the site is ongoing and lead by Manuel Rojo Guerra and José Ignacio Royo Guillén. Within the large stratigraphic sequence three different occupation phases are discerned that are supported by more than twenty radiocarbon dates [63]. The first phase corresponds to the Early Neolithic (ca. 5,300–4,800 years cal BCE.) for which genomic data has been published in Haak et al. [3] for seven individuals. The second phase dates back to the Middle Neolithic (ca. 4,500–4,300 years cal BCE.) when the cave was possibly used by animals and no human remains have been retrieved. During the third phase (ca. 4,000–3,700; 3,350–2,900 years cal BCE) the cave was again used as a burial place despite not being the only purpose. From the third phase, we have included two individuals in the present genomic study:

ELT002, adult (5,008 ± 23 BP [MAMS-16160], 5,882–5,658 years cal BP [2-sigma value]) [63].
ELT006, adult (5,035 ± 23 BP [MAMS-16165], 5,895–5,716 years cal BP [2-sigma value]) [63].

Ancient DNA processing and quality control

Sampling of ancient human remains

For the ancient individuals analyzed in this study, we sampled various bones (a humerus, phalanges, metacarpals, mandibles and a cranial fragment) and teeth (molars) in the clean room of the Max Planck Institute for the Science of Human History (MPI-SHH) in Jena, Germany, and at the Institute of Anthropology, Johannes Gutenberg University, Mainz (Data S1B). Prior to sampling, samples were irradiated with UV-light for 30 min at all sides. Different sampling methods were used for different bones types, including sandblasting, grinding with mortar and pestle, and cutting and drilling in the denser regions (Data S1B). Teeth surfaces were cleaned with a low concentration bleach solution (3%). For the teeth sampled at the MPI-SHH, the crown was separated from the root by cutting with a hand saw along the cementum/enamel junction followed by drilling inside the pulp chamber [64]. For the teeth sampled in Mainz the complete tooth was ground using a mixer mill [46].

DNA extraction

DNA extraction was done following a modified version of the Dabney protocol [12], with an initial amount of 50-100 mg of bone or tooth powder. Samples were digested with extraction buffer (EDTA, UV H_2O and Proteinase K) during 16-24h in a rotator at 37°C. The suspension was centrifuged and the supernatant transferred into binding buffer (GuHCl, UV H_2O and Isopropanol) and then into silica columns (High Pure Viral Nucleic Acid Kit; Roche). The columns were first washed with wash buffer (High Pure Viral Nucleic Acid Kit; Roche) and then eluted in 100 μL TET (TE-buffer with 0.05% Tween). We included one or two extraction blanks in each extraction series to check for cross-contamination between samples and background contamination from the lab.

Library preparation

A total of 27 double-stranded (ds) libraries were created from 25 μL DNA template extract at the MPI-SHH, following a protocol by Meyer & Kircher [10] with unique index pairs [11]. We used a partial Uracil DNA Glycosylase treatment (UDG-half) that repairs damaged nucleotides by removing deaminated cytosines except for the final nucleotides at the 5′ and 3′ read ends to retain a damage pattern characteristic for ancient DNA [65]. The libraries generated from Goyet Q-2 were ds-non-UDG and ss-UDG-half treated. We repaired the terminal ends of the DNA fragments using T4 DNA Polymerase (NBE) and joined the Illumina adaptors using the Quick Ligation Kit (NBE). We also added one or two library blanks per batch. One aliquot of each library was used to quantify the DNA copy number with IS7/IS8 primers [10] outside the clean room using DyNAmo SYBP Green qPCR Kit (Thermo Fisher Scientific) on the LightCycler 480 (Roche). Libraries were double indexed with unique index combinations [11] before doing PCR amplifications outside the cleanroom with PfuTurbo DNA Polymerase (Agilent). After amplification, the indexed products were purified with MinElute columns (QIAGEN) and eluted in 50 μL TET buffer and quantified with IS5/IS6 primers using the DyNAmo SYBP Green qPCR Kit (Thermo Fisher Scientific) on the LightCycler 480 (Roche) [10]. We used Herculase II Fusion DNA Polymerase (Agilent) with the same IS5/IS6 primers for the further amplification of the indexed products up to a copy number of 10 e-13 molecules/μL. After another purification round, we quantified the indexed libraries on a TapeStation (TapeStation Nucleic Acid System, Agilent 4200) and made a 10nM equimolar pool. Data S1B shows an overview of the extracts and libraries generated for each ancient individual.

Shotgun screening and in-solution enrichment of nuclear DNA (1240k capture) and mtDNA (mitocapture)

The pooled double indexed libraries were sequenced on an Illumina HiSeq2500 for a depth of ~5 million read cycles, using either a single (1x75bp reads) or double end (2x50bp reads) configuration. Reads were analyzed with EAGER 1.92.32 [32] to check the quality and quantity of endogenous human DNA in each library. We selected samples for targeted in-solution capture enrichment that showed a damage pattern characteristic for ancient DNA and with > 0.2% endogenous DNA. We further amplified these libraries with the IS5/IS6 primer set to a concentration of 200-400 ng/μL. After that, the libraries were hybridized in-solution to different oligonucleotide probe sets synthetized by Agilent Technologies to enrich for the complete mitogenome (mtDNA capture [66]) and for 1,196,358 informative nuclear SNP markers (1240K capture [67]).

QUANTIFICATION AND STATISTICAL ANALYSIS

Read processing and assessment of ancient DNA authenticity

We demultiplexed the sequenced libraries according to expected read indexes, allowing for one mismatch. We clipped adapters with AdapterRemoval v2.2.0 [68]. For paired end reads, we restricted to merged fragments with an overlap of at least 30 bp. Single end reads shorter than 30 bp were discarded. We mapped fragments to the Human Reference Genome Hs37d5 using the Burrows-Wheeler Aligner (BWA, v0.7.12-r1039) *aln* and *samse* commands (-l 16500, -n0.01, -q30) [69] and removed duplicate reads using DeDup v0.12.1. We excluded reads with a mapping quality phred score < 30. A summary of quality statistics is given for 1240K SNP captured libraries in Data S1C and for mtDNA captured libraries in Data S1D.

Contamination tests

Prior to genotype calling we assessed the level of contamination in the mitochondrial and nuclear genome using several methods.

DNA damage

We determined and plotted the deamination rate pattern in our UDG-half libraries using MapDamage v.2.0.6 from EAGER 1.92.32 [32]. Although damage rates at the terminal read ends vary (5.3%–14.8%) in libraries for individuals from different sites, all libraries show deamination patterns expected for ancient DNA (Data S1C). Then we trimmed the reads for 2 bp at both terminal ends of the UDG-half libraries to reduce the bias of deamination from our genotype calls. Non-UDG libraries generated from Goyet Q-2 were trimmed for 10 bp.

Contamination based on the match rate to the mtDNA dataset (ContamMix)

We used ContamMix 1.0.10 to estimate the mitochondrial contamination levels in our mito-captured libraries taking a worldwide mitochondrial dataset to compare as a potential contamination source [16] (Data S1F). We find contamination rates below 2.2% for all libraries (Data S1F). We visualized the mitochondrial read alignment with Geneious R8.1.974 [36] and manually checked for heterozygous calls to confirm the ContamMix estimates. We found a substantial mitogenome heterozygosity level for BAL003_MT in the manual check, contrasting its respective ContamMix estimate of 2.2%, and therefore excluded this library from further genome analyses.

Sex determination and X-contamination

We determined genetic sex by calculating the X-ratio (targeted X-Chromosome SNPs/ targeted autosomal SNPs) and Y-ratio (targeted Y-Chromosome SNPs/ targeted autosomal SNPs) (Data S1F). For uncontaminated libraries, we expect an X ratio ~1 and Y ratio ~0 for females and X and Y ratio of 0.5 in males [2]. Potential individuals that fall in an intermediate position could indicate the presence of DNA contamination.

Method 2 of the ANGSD package was used on merged and unmerged libraries from male individuals to test the heterozygosity of polymorphic sites on the X chromosome [15]. For low coverage libraries from the same individual with < 200 SNPs on the X chromosome we merged them into a single BAM file using samtools v0.1.19 [31] to facilitate contamination estimation of the merged libraries (Data S1F). Finally, merged libraries with less than 3.3% contamination were selected for population genetic analysis.

Genotyping and merging with dataset

After trimming of potentially damaged terminal ends bamfiles were genotyped with pileupCaller (https://github.com/stschiff/sequenceTools/tree/master/src-pileupCaller), which call one SNP per position considering the human genome as pseudo-haploid genome. Genotyped data were merged with Human Origins panel (~600K SNPs) [26] and 1240K panel [17]. For Goyet Q-2 the genotyping was applied to clipped and unclipped bamfiles, calling only transversions in the latter to avoid residual ancient DNA damage and merging these extra SNPs in the final genotype. The number of SNPs covered per individual is shown in Data S1G.

Kinship relatedness and individual assessment

We first calculated the pairwise mismatch rate between bam files to rule out a potential duplication of individuals. We found the same low mismatch rate comparing different bamfiles combination from libraries of BAL001 and BAL005, respectively, suggesting that both samples come from the same individual. We consequently merged them as BAL051.

We then used Relationship Estimation from Ancient DNA (READ) to estimate the degree of genetic kinship relatedness among the final set of individuals [37]. This method can determine first and second-degree relatedness among individuals and can also be used to test for potential cross-contamination among libraries from the same batch of sampling processing. We calculated the proportion of non-matching alleles and normalized the results separating Neolithic from HG individuals taking into account the potential genetic diversity within each group to calculate the proportions of non-matching alleles (*P0*). After the normalization of both groups, *P0* of HG ranged between 0.964-1.029 and between 0.974-1.074 for Neolithic individuals, which in both cases is higher than 0.90625, the top value for second-degree related individuals. In sum, there are no first- or second-degree relatives among our newly reported ancient Iberian individuals.

Phenotypic traits analysis

We extracted a list of 36 SNPs of functional importance or related to known phenotypic traits [22] (e.g., lactase persistence, pigmentation, eye colors) (Figure S4) and calculated the genotype likelihood based on the number of reads (using a quality filter q30) for each specific position to determinate the presence of the ancestral or derived alleles [22].

We interrogated different SNPs positions on the gene *OCA2* related to light eye color. We obtain the ancestral allele (rs12913832, 3 reads) in individuals CHA001 and ELT002, and heterozygous allele calls for ELT006, suggesting dark color eyes for all of them. The SNP coverage was not sufficient to reliable type the remaining individuals. Another allele from the same gene (rs1800404) related to eye pigmentation could support darker pigmentation in ELT006 than in ELT002. We also checked different SNPs positions in the gene *SLC45A2.* We obtain the ancestral alleles (rs1426654, 2 reads) in BAL0051 and (rs16891982, 4 reads) in CHA001 which suggest a darker skin color than ELT002 and ELT006, who are heterozygous or homozygous for the derived allele. The coverage in the other individuals is very low to allow comparisons. The allele rs3827760 of the *EDAR* gene, related to straight and thick hair, is ancestral in all individuals, albeit with variable coverage (CHA001, 4 reads; CHA003, 3 reads; ELT002, 14 reads and; ELT006, 8 reads). Also, as reported before for pre-farming and Neolithic individuals [13], none of our newly typed individuals show evidence for Lactase persistence. Individuals from the Neolithic times (ELT002, ELT006 and FUC003) show different combinations of derived and ancestral alleles of the gene *rs174546*, which is related to the capacity of regulation of the production of long-chain polyunsaturated fatty acids (FADS1/FADS2). Results are shown in Figure S4.

Mitochondrial and Y chromosome haplogroups

Using an in-house mtDNA capture assay, we could recover complete mitochondrial genomes from individuals CHA002, CHA003, CHA004, ELT002, ELT006, and FUC003. The coverage of the mtDNA genome for the rest or the samples ranges from 88.86%–99.99% (Data S1D). We used samtools v1.3.1 to extract reads from mitocapture data [31] and mapped them to the rCRS and called the consensus sequences using Geneious R8.1.974 [36]. We downloaded these consensus sequences in *fasta* format and they were used to determinate mitochondrial haplotypes using Haplogrep 2 [34] (Data S1F).

Iberian HG individuals from Balma Guilanyà and Moita do Sebastião belong to haplogroup U, together with the two MN individuals CHA004 and ELT006 (Data S1F). Individual BAL003 could be assigned to U2′3′4′7′8′9′, also found in the Paglicci 108 (~27,000 years cal BP, Italy), Rigney 1 (~15,500 years cal BP, France) [2], and Grotta d'Oriente C HG (~14,000 years cal BP, Italy) [19]. Individual BAL0051 belongs to U5b2a, also found in Neolithic Scotland [9]. Moita do Sebastião (CMS001) carries haplogroup U5b1, which was reported from MN, Bell Beaker and Middle Bronze Age individuals from Portugal and Spain [7, 9], and in the Ranchot 88 HG (~10,000 years cal BP, France) [2].

Early Neolithic individuals from Cueva de Chaves do not carry U haplogroups. Individual CHA001 could be assigned to haplogroup HV0+195, but was previously reported as K based on PCR-based results [70]. This haplogroup has been reported from MN Ireland [20], as well as MN Scotland and Bell Beaker individuals from England [9]. Individual CHA002 was assigned to K1a2a, which is common in Early Neolithic Iberia, e.g., Cova Bonica [8], Cova de Els Trocs [3] and Cueva del Toro [30], but also in Chalcolithic and Bell Beaker individuals from Iberia and Italy [9, 13]. CHA003 was assigned to K1a3a, so far reported from Neolithic Anatolia [13], Neolithic and Chalcolithic Scotland, and Bell Beaker individuals from Sicily [9]. Middle Neolithic individual CHA004 carried haplogroup U4a2f, found in HG from the Iron Gates, Romania, and Lithuania [14, 19]. The MN individual ELT002 carries haplogroup J1c1b, present in the Körös Early Neolithic [18], Neolithic from Scotland [9], Iberian Late Neolithic-Chalcolithic [7], and Bronze Age from Italy and Germany [6]. ELT006 was assigned to haplogroup U3a1, which has been reported from MN France and Germany [9, 19], and Chalcolithic Iberia [13]. Early Neolithic Fuente Celada carries haplogroup X2b+226, found in MN Hungary [71] and Middle Bronze Age Iberia [7]. X2b was

found in Iberian Late Neolithic [5], Chalcolithic [18] and Bell-Beaker individuals [9], Neolithic England [9] and Greece [72], and EN/LN Morocco [25].

For Y haplogroup determination, we first called the Y chromosome SNPs of the 1240K SNP panel from all male individuals using pileupCaller with MajorityCalling mode, (https://github.com/stschiff/sequenceTools), and mapping quality $\geq$30 and base quality $\geq$30 (Data S1E). Y chromosome haplogroups were called from the list of Y-SNPs included in the 1240K capture assay using the script *yhaplo* [35].

BAL0051 could be assigned to haplogroup I1, while BAL003 carries the C1a1a haplogroup. To the limits of our typing resolution, EN/MN individuals CHA001, CHA003, ELT002 and ELT006 share haplogroup I2a1b, which was also reported for Loschbour [73] and Motala HG [13], and other LN and Chalcolithic individuals from Iberia [7, 9], as well as Neolithic Scotland, France, England [9], and Lithuania [14]. Both C1 and I1/ I2 are considered typical European HG lineages prior to the arrival of farming. Interestingly, CHA002 was assigned to haplogroup R1b-M343, which together with an EN individual from Cova de Els Trocs (R1b1a) confirms the presence of R1b in Western Europe prior to the expansion of steppe pastoralists that established a related male lineage in Bronze Age Europe [3, 6, 9, 13, 19]. The geographical vicinity and contemporaneity of these two sites led us to run genomic kinship analysis in order to rule out any first or second degree of relatedness. Early Neolithic individual FUC003 carries the Y haplogroup G2a2a1, commonly found in other EN males from Neolithic Anatolia [13], Starçevo, LBK Hungary [18], *Impressa* from Croatia and Serbia Neolithic [19] and Czech Neolithic [9], but also in MN Croatia [19] and Chalcolithic Iberia [9].

Population genetic analysis

Labeling population groups

For the Paleolithic individuals, we adopted the labels from Fu et al. [2] and used the improved genotype calls from Mathieson et al. [19]. We included the HG with more than 15,000 SNPs covered in the PCA (Figure S1). If the HG individuals from the same site or chrono-cultural context clustered in the PCA analysis, we grouped them using the same label for the following population genetic analysis [14, 19, 21]. Neolithic individuals from Iberia were grouped by sites and by chrono-cultural context. For Neolithic individuals from outside of Iberia, we kept the label names from the respective initial publications [9, 13, 19, 20, 24, 74].

Principal Component Analysis

PCA analysis was run with the Human Origins dataset using *smartpca* v10210 (EIGENSOFT) with the option *SHRINKMODE* [33] using 777 modern populations to calculate eigenvectors on which aDNA samples were projected [74]. PC1 was multiplied by -1 (–PC1) in order to mirror geography.

F-statistics

D-statistics and F-statistics were calculated with qpDstat from ADMIXTOOLS (https://github.com/DReichLab). We used the 1240K panel to increase the number of SNPs covered by the ancient individuals and get more resolution in the statistic tests. Standard errors were calculated with the default block jacknife. We report and plot three standard errors in all F-statistics.

qpAdm and qpWave

We used *qpWave* and *qpAdm* from the ADMIXTOOLS package (https://github.com/DReichLab) to estimate admixture proportions. We used this framework to model and quantify the ancestry proportions of HG individuals in- and outside of Iberia (we only use HG or groups of HG with more than 30,000 SNPs). First, we tested whether Goyet Q-2 and Villabruna formed a clade with respect to the following set outgroups: Mota, Ust'-Ishim, Mal'ta 1 (MA1), Koros EN-HG, Goyet Q116-1, Mbuti, Papuan, Onge, Han, Karitiana and Natufian extending the set used by Olalde et al. [9]. The resulting *qpWave* model showed an extremely poor fit (p value = 3.07705115e-91), which means that our set of outgroups can be used to differentiate between Goyet Q-2 and Villabruna-related ancestry. We then modeled the ancestry in the HG groups as a mixture of Goyet Q-2 and Villabruna (Figure 3D and Data S1I). Alternatively, we also used El Mirón and Loschbour as potential source populations and the same ten outgroups, after checking that El Mirón and Loschbour are not equally related to the outgroups (p value = 5.41365181e-68) (Figure S3A and Data S1I).

We also used *qpWave* and *qpAdm* to explore the HG admixture in the Neolithic populations. In this case, the sources (left populations) were Anatolia Neolithic, Goyet Q-2 and Villabruna. We chose the same set of outgroups, and first checked that Anatolia Neolithic, Goyet Q-2 and Villabruna were not equally related to the outgroups (p value = 8.63552043e-92) (Figure 4B and Data S1J. We repeated the model with Anatolia Neolithic, El Mirón and Villabruna as sources using the same set of outgroups. However, the resulting *qpWave* model also resulted in a poor fit (p value = 9.5033482e-59) (Data S1J).

The results of all *qpWave* and *qpAdm* models are reported in Data S1I and S1J. The criteria to report these values were as follows: If the resulting p values were higher than 0.05 we report the three-sources model. If the p values were lower than 0.05 we show the best p value obtained from the three- or two-sources model (i.e., the nested model). In case of negative values for some of the sources, we report the two-sources model (nested model) and the respective p value of these models. We applied the same criteria to two-sources models.

Multi-Dimensional Scaling analysis (MDS)

We computed Multi-Dimensional Scaling (MDS) analysis using the R package *cmdscale* to measure the genetic dissimilarity among hunter-gatherers (HG), and then used the inverted [*1-f3(HG1; HG2, Mbuti)*] pairwise values among all the combinations [2].

DATA AND SOFTWARE AVAILABILITY

Data is available at ENA under study accession number PRJEB30985.

Current Biology, Volume *29*

Supplemental Information

Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula

Vanessa Villalba-Mouco, Marieke S. van de Loosdrecht, Cosimo Posth, Rafael Mora, Jorge Martínez-Moreno, Manuel Rojo-Guerra, Domingo C. Salazar-García, José I. Royo-Guillén, Michael Kunst, Hélène Rougier, Isabelle Crevecoeur, Héctor Arcusa-Magallón, Cristina Tejedor-Rodríguez, Iñigo García-Martínez de Lagrán, Rafael Garrido-Pena, Kurt W. Alt, Choongwon Jeong, Stephan Schiffels, Pilar Utrilla, Johannes Krause, and Wolfgang Haak

Supplemental Figures

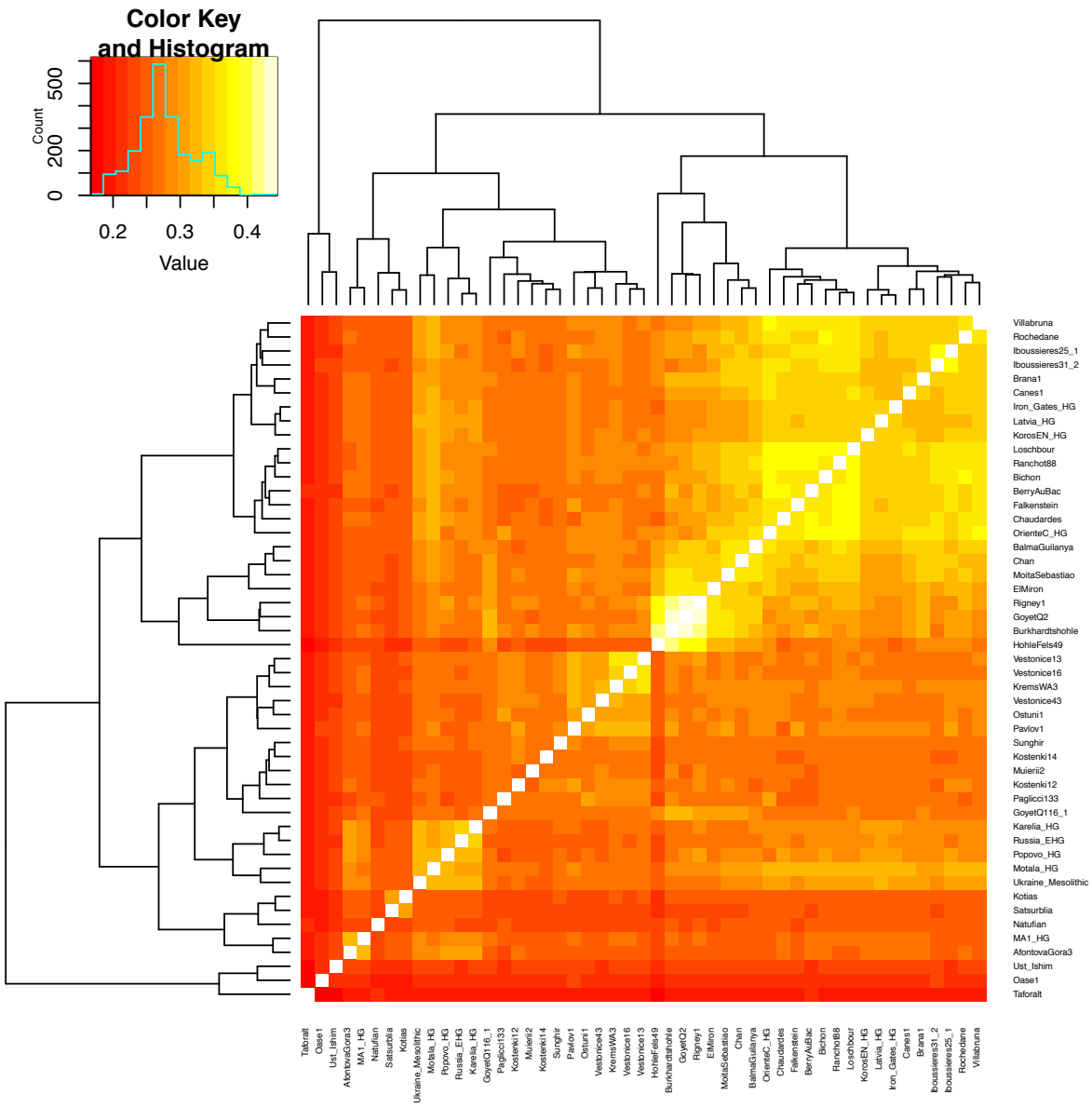

Figure S1. Heat plot showing the genetic distances between Eurasian HG, Related to Figure 2A. Genetic distances were calculated using *f3-outgroup* statistics of the form *f3(X;Y, Mbuti)*, with X and Y being Eurasian hunter-gatherers in all possible pairwise comparisons. The analysis has been restricted to samples with more than 30,000 SNPs, following Fu et al. [S1]. The clustering pattern is similar to the MDS plot (STAR methods, Figure 2A): the newly reported Moita do Sebastião [~8ky cal BP], Balma Guilanyà [~12ky cal BP], and Chan [~9ky cal BP] cluster with El Mirón and Goyet Q-2, whereas La Braña 1 and Canes 1 cluster with Villabruna.

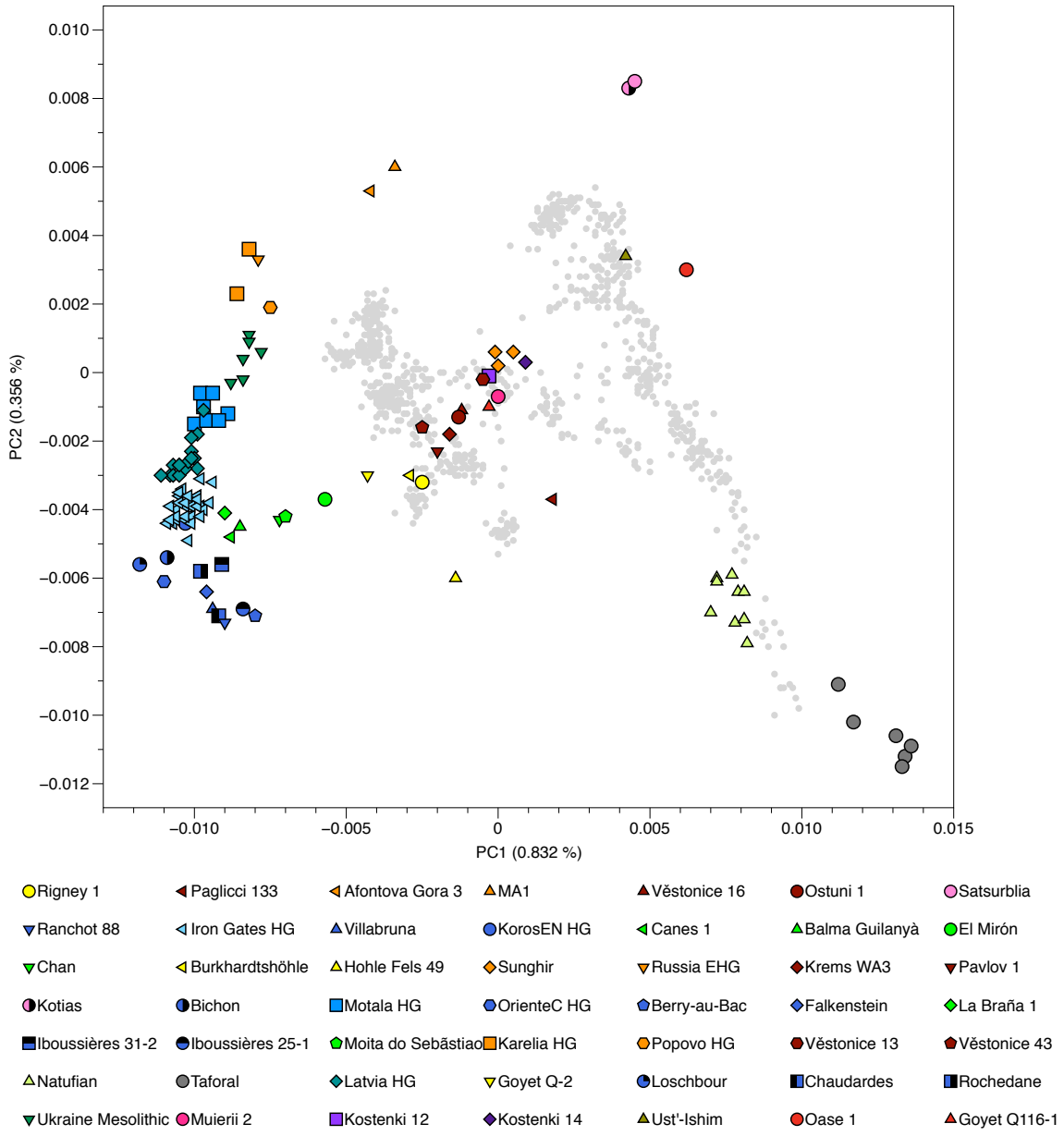

Figure S2. Principal Component Analysis of Hunter-gatherer individuals, Related to Figure 2A. PCA analysis calculated with 777 present day West Eurasians [S2] with option shrinkmode:YES on which HG individuals were projected.

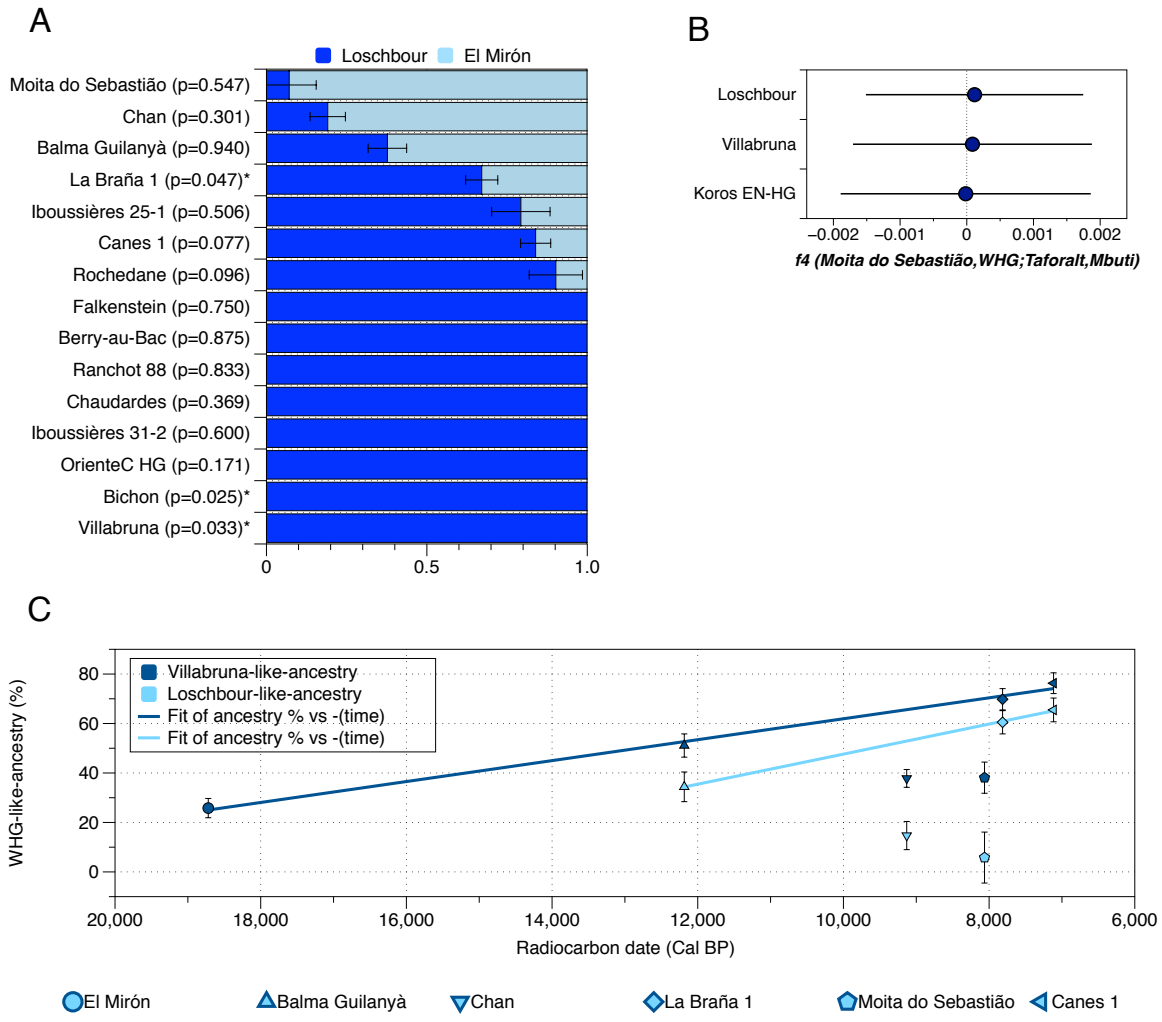

Figure S3. Hunter-gatherer ancestry and geographical correlation, **Related to Figure 3D**
A) Modelling European HG as admixture of Villabruna-like and El Mirón-like ancestry using *Loschbour* and *El Mirón* as proximal sources, respectively (error bars indicate ±1 standard error). **B)** f4-statistics showing **no** affinity between Geometric Mesolithic Moita do Sebastião from Portugal and Iberomaurusian HG from Taforalt, Morocco, North Africa. Taforalt individuals are a good proxy to test the African-Iberian connections due to the genetic continuity attested in North Africa from the Late Pleistocene to the Holocene (Early Neolithic) despite their chronologically older age [S3]; errors bars indicate ±3 standard errors **C)** Correlation between Villabruna-like ancestry (dark blue; Figure 3D) and Loschbour-like ancestry (light blue; Figure S3A) and time (error bars indicate the radiocarbon 2-sigma range). Both models result in a fit of R = 0.99 for individuals from north and northeast of Iberia Peninsula (to the exclusion of Chan and Moita do Sebastião), where we observe an increase of WHG-like ancestry similar to other parts of Europe.

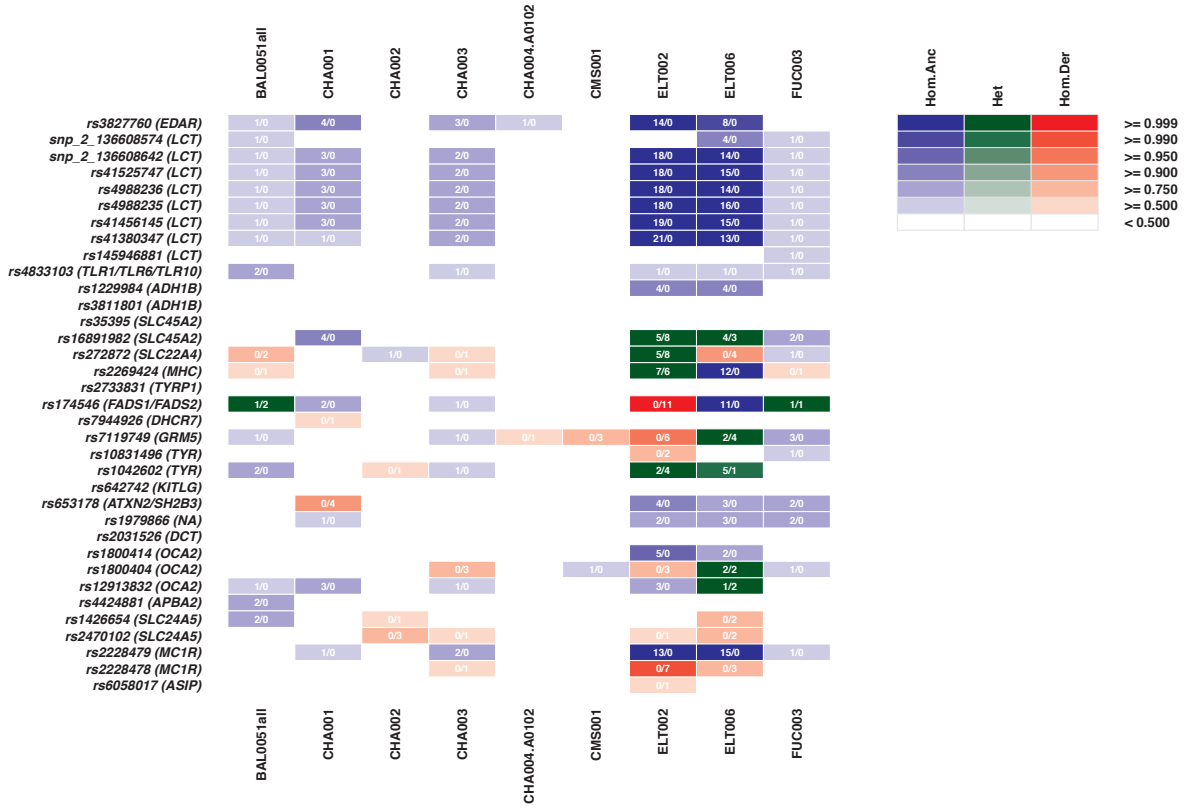

Figure S4. Summary of genotypes of phenotypic and functional SNPs, related to STAR methods. Colours indicate the homozygous ancestral/derived or heterozygous state of the SNPs reported in the left-hand column. Numbers in cells indicate the number of reads matching the ancestral/derived allele.

Supplemental References

S1. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., *et al.* (2016). The genetic history of Ice Age Europe. Nature *534*, 200-205.

S2. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., *et al.* (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419-424.

S3. Fregel, R., Méndez, F.L., Bokbot, Y., Martín-Socas, D., Camalich-Massieu, M.D., Santana, J., Morales, J., Ávila-Arcos, M.C., Underhill, P.A., Shapiro, B., *et al.* (2018). Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.1800851115

S4. Posth, C., Renaud, G., Mittnik, A., Drucker, D.G., Rougier, H., Cupillard, C., Valentin, F., Thevenet, C., Furtwängler, A., Wißing, C., et al. (2016). Current Biology *26*, 827-833.