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

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

\textsuperscript{1} Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Kahlaische Straße 10, 07745 Jena, Germany.
\textsuperscript{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 Curbuna, 50009, Zaragoza, Spain.
\textsuperscript{3} Centre d’Estudis del Patrimoni Arqueològic de la Prehistòria (CEPAP), Facultat de Lletres, Universitat Autònoma Barcelona, 08190 Bellaterra, Spain.
\textsuperscript{4} Department of Prehistory, University of Valladolid, Plaza del campus, 47011 Valladolid, Spain.
\textsuperscript{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.
\textsuperscript{6} Dirección General de Cultura y Patrimonio. Gobierno de Aragón, Avenida de Aragón, 5 D. 50018 Zaragoza, Spain.
\textsuperscript{7} Deutsches Archäologisches Institut, Abteilung Madrid, E-28002 Madrid, Spain.
\textsuperscript{8} Department of Anthropology, California State University Northridge, Northridge, California 91330-8244, USA.
\textsuperscript{9} Université de Bordeaux, CNRS, UMR 5199-PACEA, 33615 Pessac Cedex, France.
\textsuperscript{10} Professional Archaeologist.
\textsuperscript{11} Postdoctoral researcher. Juan de la Cierva-Formación Program. Institute of Heritage Sciences. INCIPIT-CSIC, Av. de Vigo, 15705 Santiago de Compostela, Spain.
\textsuperscript{12} Postdoctoral researcher. Juan de la Cierva-Incorporación Program. Department of Prehistory. Valladolid University, Plaza del campus, 47011 Valladolid, Spain.
\textsuperscript{13} Department of Prehistory. Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049, Madrid, Spain.
\textsuperscript{14} Center of Natural and Cultural Human History, Danube Private University, Steiner Landstr. 124, A-3500 Krems, Austria.
\textsuperscript{15} Department of Biomedical Engineering, University of Basel, Switzerland, Gewerbestrasse 14-16, CH-4123 Allschwil
\textsuperscript{16} Corresponding author

Correspondence
haak@shh.mpg.de (W.H.)
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 (HG) 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 correspondance with geography. Here, we report new genome-wide data from eleven HG and Neolithic individuals that highlight the late survival of Paleolithic ancestry in Iberia, reported previously in Magdalenian-associated individuals. We show that all Iberian HG, including the oldest ~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 HG. 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 HG.

Keywords

Paleolithic, Mesolithic, Neolithic, ancient DNA, human, Last Glacial Maximum, Iberia, Europe, genome, ancestry
Results and Discussion

We successfully generated autosomal genome-wide data and mitochondrial genomes of ten new individuals from key sites in the Iberian Peninsula ranging from ~13,000–6,000 yrs cal BP: Late Upper Paleolithic (n=2), Mesolithic, (n=1), Early Neolithic (n=4) and Middle Neolithic (n=3) (STAR Methods, Data S1, Figure 1). We furthermore improved the sequencing depth of one Upper Paleolithic individual from the Troisíème caverne of Goyet cave (Belgium), dated to ~15,000 yrs cal BP and associated with the Magdalenian culture [2]. For each individual we generated multiple double-stranded DNA libraries with unique index pairs [10,11] for next generation sequencing from ancient DNA (aDNA) extracted from teeth and bones [12] (STAR Methods, Data S2). These were subsequently enriched using targeted in-solution capture for ~1240K informative nuclear single nucleotide polymorphisms (SNPs) [13], and independently for the complete mitogenome [14], and sequenced on an Illumina HiSeq4000 platform. All libraries contained short DNA fragments (46-65 base pair length on average) with post-mortem deamination patterns characteristic for aDNA (4-16% for UDG-half and 19-31% for non-UDG at the first base pair position; STAR Methods, Data S3 and S5). We estimate 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] (STAR Methods, Data S6). After quality filtering we obtained an endogenous DNA content of 1.3-29.5% on the targeted 1240K SNPs (STAR Methods, Data S3). We called pseudo-haplotype 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 [2,5,7–9,13,14,18,21–26]. The final dataset from the newly reported individuals contained 19,269-814,072 covered SNPs (STAR Methods, Data S3, S7). 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 [27]. This data set includes ~600K intersecting autosomal SNPs for which our individuals covered 10,602-447,287 SNPs.

Genetic structure in Iberian hunter-gatherers

To characterise the genetic differentiation between HG individuals, we calculated genetic distances, defined as 1 – f3, where f3 denotes the f3-outgroup statistics [27], for pairwise comparisons among all published and newly generated HG, and visualized the results using Multidimensional Scaling (MDS) (STAR Methods, Figure 2A). The HG individuals form distinguishable clusters on the MDS plot, supported by f4-statistics and clustering analysis (Figure S1), 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 Italic 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, and Burkhardtshöhle. With more data available, we notice that Iberian HG form a cline between the GoyetQ2 and Villabruna clusters. This cline also includes El Mirón, 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 Cantabria in northern Iberia) are falling closer to the Villabruna cluster, whilst Chan (northwestern Iberia) and our newly reported individuals from Moita do Sebastião (Portuguese Atlantic coast) and Balma Guilanya (Pre-Pyrenean region, northeastern Iberia) are closer to El Mirón, which is in turn closer to the Magdalenian GoyetQ2 cluster (Supplemental Information 2.1).

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 HG (Chan, Moita do Sebastião, and El Mirón) as well as Hohle Fels 49 and Goyet Q-116-1 show significantly positive f4-values, indicating that these
individuals shared more ancestry with Goyet Q-2 than with Villabruna (Figure 2B). This heterogeneity in Iberian HG cannot be explained by genetic drift alone (to which this type of F-statistics is robust against), but only by admixture between two sources related to Goyet Q-2 and Villabruna, respectively. We visualise 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 towards 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 towards the x-axis, indicating excess genetic drift with Goyet Q-2. Iberian HG 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(Goyet-Q-2 cluster, ElMirón; 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 Burkhardshöhle. 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 S8). Based on this observation, we used the test f4(Goyet Q-2, Goyet Q-2 cluster; Villabruna, Mbuti), for which El Mirón is significantly negative (Figure 3B, Data S8), 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 Goyet Q-2 individuals we used the test f4(Goyet Q-2 cluster, Villabruna; El Mirón, Mbuti). Here, all individuals of the Goyet Q-2 cluster are significantly positive (El Mirón Z-score = 11.25), 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, Data S8). The mixed ancestry of El Mirón could also explain its reduced affinity to the 35,000-year-old Goyet Q116-1 individual when compared to the younger GoyetQ2 cluster in the test f4(‘GoyetQ2 cluster’, Goyet Q-2; Goyet Q116-1, Mbuti) (Data S8).

Having established two potential Paleolithic source populations surviving in Iberia from ~19,000 yrs BP onwards, we used the admixture modelling programs qpWave and qpAdm (ADMIXTOOLS; STAR Methods, Figure 3C) to explore the dual ancestry in 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 HG and is consistent with the cline between Villabruna and Goyet Q-2-like ancestries described above (Figure 3D, STAR methods). Here, Villabruna-like ancestry is the dominant component (69.8 ± 4.3 -100%) in individuals of the Villabruna cluster, and Goyet Q-2-like ancestry the dominant component (61.9 ± 6.3 - 94.3 ± 5.7 %) in GoyetQ2 cluster individuals (Figure 3C, Data S8). 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). Congruent with the pattern observed in MDS (Figure 1A), the F-statistic based tests (Figure 1B, 2A), 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 HG (ranging from 23.7 to 75.3%) than to contemporaneous WHG outside of Iberia. Balma Guilanyà, La Braña 1, and Canes 1 (Figure 3B, Data S9) show elevated Villabruna admixture proportions, but also higher Goyet Q-2 proportions than non-Iberian HG.

**Dual hunter-gatherer genetic legacy in Iberian Neolithic individuals**

During the Neolithic transition ~7600 years ago, human expansions reached the Iberian Peninsula relatively swiftly via expanding early farmers from western Anatolia [3–5]. The rapid expansion of Early Neolithic (EN) individuals associated with farming practises 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 [28,29]. 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 [30,31]. 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 HGs signatures [18].

Given the presence of two ancestral lineages in Iberian HG, we aimed to explore this potential genetic legacy in our newly generated Early and Middle Neolithic (MN) individuals. We first used principal component analysis (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 [5,7,8,13,19,26]. As shown before, MN individuals are shifted towards WHG individuals [3], including the newly reported MN individuals from Cova de Els Trocs and Cueva de Chaves.

Olalde et al. [8] have shown that Cova Bonica and other published EN individuals received either all or the majority of their HG ancestry en route from western Anatolia to Iberia, with only marginal additional ancestry from inside Iberia [8,9]. We thus tested our new Iberian HG individuals, who differ subtly from those previously reported. Using qpAdm models consistent with those above (Supplementary Figure S3), we aimed to trace and quantify the proportion of GoyetQ-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, GoyetQ-2 and Villabruna, respectively. We show that EN Iberians shared a higher proportion of GoyetQ-2-like ancestry than EN individuals from outside Iberia (Figure 4B, Data S10). GoyetQ-2-like ancestry is higher in EN from southern Iberia (Andalusia) suggesting additional admixture with local Iberian HG, 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-value 6.91e-05) in favour of two sources, despite being poorly supported (p-value 0.001) (Data S10). GoyetQ-2-like ancestry is highest in all Iberian MN (except Chaves MN) when compared with other MN populations that share a similar overall amount of HG ancestry (Figure 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 (2018) 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 S10), 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 Goyet Q-2-like ancestry in MN individuals outside Iberia also hints at Iberia as one possible 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 Upper Paleolithic/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 yrs 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 LGM refugia. It is possible that GoyetQ2 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-Q2 and Villabruna 
ancestries originated 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 HG in Iberia during the 
time of the Mesolithic-Neolithic transition.

Acknowledgements
We thank the members of the Archaeogenetics Department of the Max Planck Institute for the 
Science of Human History, especially Maite 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 and 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 (WH). VVM was funded by a 
predoctoral scholarship of the Gobierno de Aragón and the Fondo Social Europeo 
(BOA20150701025) and a 3-months research stay grant (CH 76/16) by Programa CAI-Ibercaja 
de Estancias de Investigación. VVM and PU 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 
(Gr. 7837 to HR), the College of Social and Behavioral Sciences of CSUN, and the CSUN 
Competition for Research, Scholarship and Creative Activity Awards.

Author Contributions
provided archaeological context, V.V.-M., M.v.d.L. and C.P. performed aDNA lab work and 
sequencing, V.V.-M., C.P., M.v.d.L., S.S., C.J., and W.H. analysed data, V.V.-M., C.P., 
M.v.d.L., and W.H. wrote the manuscript with input from all co-authors.

Declaration of Interests
The authors declare no competing interests.
References


Radiocarbon dating the beginning of the Neolithic in Iberia: new results, new
Martinez-Moreno, J., Mora, R. (2009). Balma Guilanyà (Prepiríneo de Lleida) y el Azilien se
Guilaine, pp. 83
https://doi.org/10.1038/s41598-017-15480-9.


Main text figure legends

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 HG and EN/MN Neolithic individuals from
the Iberian Peninsula (no outlines). (B) Radiocarbon dates of newly reported individuals in
calibrated years BP (error bars indicate the 2-sigma range).

Figure 2. Genetics distances between European HG 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: Vestonice (dark red), Mal’ta (orange: Mal’ta 1 [MA1]
and Afontova Gora 3), Satsurblia (light pink: Kotias and Satsurblia), Villabruna (blue: Koros
EN-HG. Berry-au-Bac 1, Rochedane, Villabruna, Chaurdardes 1, Ranchot 88, LaBraña 1 and
Loschbour), and GoyetQ2 (yellow; newly defined). Iberian HG are shown as green symbols.
(B) f4-statistics highlighting the excess affinity to Goyet Q-2 in Iberian and European HG
(>20,000 SNPs; error bars indicate ±3 standard errors; Z-scores >3 (green)).

Figure 3. Key f3-outgroup tests, f4-statistics, and qpAdm results
(A) Biplot of f3-outgroup
tests illustrating the Villabruna-like and GoyetQ2-like genetic ancestries in European HG. The
x=y axis marks full symmetry between GoyetQ2-, and Villabruna-like ancestries and deviations
mark excess ancestry shared with GoyetQ2 (yellow) or Villabruna (blue). (B) Results of f4-
statistics highlighting the shared genetic drift between El Mirón and Villabruna individuals using
20,000 intersecting SNPs. (C) Results of f4-statistic showing that El Mirón is not a sister clade
of Villabruna. (D) Modelling European HG as a two-way admixture of Villabruna-, and
GoyetQ2-like ancestry using Villabruna and Goyet Q-2 as proxies, respectively.

Figure 4. PCA results and qpAdm admixture models
(A) PCA analysis calculated with 777
present-day West Eurasians on which published ancient individuals from western Europe and
newly reported individuals (stars) were projected. (B) Modelling EN and MN populations from
Iberian and western Europe asadmixture of three ancestral sources: Anatolian Neolithic,
Goyet Q-2 and Villabruna (Data S10). (C) f4(Iberian MN, Neolithic British Isles, test, Mbuti),
where test is Goyet-Q2 (yellow) or Villabruna (blue), highlighting the excess of Goyet Q-2
ancestry in Iberian MN compared to Neolithic England and Scotland. The affinity to Villabruna
is shown for comparison to avoid potential bias created by unspecified HG attraction.
**Star Methods:**

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

1. **KEY RESOURCES TABLE**
2. **CONTACT FOR REAGENT AND RESOURCE SHARING**
3. **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
4. **ARCHAEOLOGICAL SITES AND SAMPLE DESCRIPTION**
5. **ANCIENT DNA PROCESSING AND QUALITY CONTROL**
   - Sampling of ancient human remains
   - DNA extraction
   - Library preparation
   - Shotgun screening and in-solution enrichment of nuclear DNA (1240k capture) and mtDNA capture
6. **QUANTIFICATION AND STATISTICAL ANALYSIS**
   - Read processing and assessment of ancient DNA authenticity
   - Contamination tests
   - DNA damage
   - Contamination based on the match rate to the mtDNA dataset (ContamMix)
   - Sex determination and X-contamination
   - Genotyping and merging with data set.
   - Kinship relatedness and individual assessment.
   - Phenotypic traits analysis
   - Y and mtDNA haplogroups
   - Population genetic analysis
   - Labelling population groups
   - Principal Component Analysis
   - F-statistics
   - qpAdm and qpWave
   - Multi-Dimensional Scale analysis (MDS)

**DATA AVAILABILITY**

**Supplemental Information:**

Supplementary Data include a text file and four supplementary figures detailing additional population genomic analysis and archaeological implications of the main findings.
### REAGENT OR RESOURCE | SOURCE | IDENTIFIER
--- | --- | ---
Ancient individual | This study/ Troisième caverne of Goyet archaeological site | GoyetQ2
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

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Source</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 M EDTA pH 8.0</td>
<td>Life Technologies</td>
<td>AM9261</td>
</tr>
<tr>
<td>1x Tris-EDTA pH 8.0</td>
<td>AppliChem</td>
<td>A8569,0500</td>
</tr>
<tr>
<td>Proteinase K</td>
<td>Sigma-Aldrich</td>
<td>P2308-100MG</td>
</tr>
<tr>
<td>Guanidine hydrochloride</td>
<td>Sigma-Aldrich</td>
<td>G3272-500g</td>
</tr>
<tr>
<td>3M Sodium Acetate pH 5.2</td>
<td>Sigma-Aldrich</td>
<td>S7899-500ML</td>
</tr>
<tr>
<td>Tween 20</td>
<td>Sigma-Aldrich</td>
<td>P9416-50ML</td>
</tr>
<tr>
<td>Water Chromasolv Plus</td>
<td>Sigma-Aldrich</td>
<td>34877-2.5L</td>
</tr>
<tr>
<td>Ethanol</td>
<td>Merck</td>
<td>1009832511</td>
</tr>
<tr>
<td>Isopropanol</td>
<td>Merck</td>
<td>1070222511</td>
</tr>
<tr>
<td>Buffer Tango</td>
<td>Life Technologies</td>
<td>BY5</td>
</tr>
<tr>
<td>T4 DNA Polymerase</td>
<td>NEB</td>
<td>M0203 L</td>
</tr>
<tr>
<td>T4 Polynucleotide Kinase</td>
<td>NEB</td>
<td>M0201 L</td>
</tr>
<tr>
<td>User Enzyme</td>
<td>NEB</td>
<td>M5505 L</td>
</tr>
<tr>
<td>Uracil Glycosylase inhibitor (UGI)</td>
<td>NEB</td>
<td>M0281 S</td>
</tr>
<tr>
<td>Bst 2.0 DNA Polymerase</td>
<td>NEB</td>
<td>M0537 S</td>
</tr>
<tr>
<td>Quick Ligation Kit</td>
<td>NEB</td>
<td>M2200 L</td>
</tr>
<tr>
<td>BSA 20mg/ml</td>
<td>NEB</td>
<td>B9000 S</td>
</tr>
<tr>
<td>ATP</td>
<td>NEB</td>
<td>P0756 S</td>
</tr>
<tr>
<td>Item</td>
<td>Supplier</td>
<td>Code</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>DyNAmo Flash SYBR Green qPCR Kit</td>
<td>Life Technologies</td>
<td>F-415L</td>
</tr>
<tr>
<td>dNTPs 25 mM</td>
<td>Thermo Scientific</td>
<td>R1121</td>
</tr>
<tr>
<td>Tween 20</td>
<td>Sigma-Aldrich</td>
<td>P9416-50ML</td>
</tr>
<tr>
<td>Ethanol</td>
<td>Merck</td>
<td>1009832511</td>
</tr>
<tr>
<td>D1000 ScreenTapes</td>
<td>Agilent Technologies</td>
<td>5067-5582</td>
</tr>
<tr>
<td>D1000 Reagents</td>
<td>Agilent Technologies</td>
<td>5067-5583</td>
</tr>
<tr>
<td>Pfu Turbo Cx Hotstart DNA Polymerase</td>
<td>Agilent Technologies</td>
<td>600412</td>
</tr>
<tr>
<td>Herculase II Fusion DNA Polymerase</td>
<td>Agilent Technologies</td>
<td>600679</td>
</tr>
<tr>
<td>Oligo aCGH/Chip-on-Chip Hybridization Kit</td>
<td>Agilent Technologies</td>
<td>5188-5220</td>
</tr>
<tr>
<td>1x TE-Puffer pH 8.0 low EDTA</td>
<td>AppliChem</td>
<td>A8569.0500</td>
</tr>
<tr>
<td>Sodiumhydroxide Pellets</td>
<td>Fisher Scientific</td>
<td>10308200</td>
</tr>
<tr>
<td>Sera-Mag Speed CM</td>
<td>GE Healthcare Lifescience</td>
<td>65152105050250</td>
</tr>
<tr>
<td>Dynabeads MyOne Streptavidin T1</td>
<td>Life Technologies</td>
<td>65601</td>
</tr>
<tr>
<td>DyNAmo Flash SYBR Green qPCR Kit</td>
<td>Life Technologies</td>
<td>F-415L</td>
</tr>
<tr>
<td>0.5 M EDTA pH 8.0</td>
<td>Life Technologies</td>
<td>AM9261</td>
</tr>
<tr>
<td>GeneRuler Ultra Low Range DNA Ladder</td>
<td>Life Technologies</td>
<td>SM1211</td>
</tr>
<tr>
<td>10x GeneAmp PCR Gold Buffer and MgCl2</td>
<td>Life Technologies</td>
<td>4379874</td>
</tr>
<tr>
<td>Human Cot-1 DNA</td>
<td>Life Technologies</td>
<td>15279011</td>
</tr>
<tr>
<td>1M Tris-HCl pH 8.0</td>
<td>Life Technologies</td>
<td>15568025</td>
</tr>
<tr>
<td>20x SOC Buffer</td>
<td>Life Technologies</td>
<td>AM9770</td>
</tr>
<tr>
<td>UltraPure™ Salmon Sperm DNA Solution</td>
<td>Life Technologies</td>
<td>15632011</td>
</tr>
<tr>
<td>Ethanol</td>
<td>Merck</td>
<td>1009831011</td>
</tr>
<tr>
<td>MiniElute PCR Purification Kit</td>
<td>Qiagen</td>
<td>28006</td>
</tr>
<tr>
<td>PEG 8000 Powder, Molecular Biology Grade</td>
<td>Promega</td>
<td>V3011</td>
</tr>
<tr>
<td>20% SDS Solution</td>
<td>Serva</td>
<td>39575.01</td>
</tr>
<tr>
<td>3M Sodium Acetate buffer solution pH 5.2</td>
<td>Sigma-Aldrich</td>
<td>S7899-500ML</td>
</tr>
<tr>
<td>5 M Sodium chloride solution</td>
<td>Sigma-Aldrich</td>
<td>S5150-1L</td>
</tr>
<tr>
<td>Denhardt's solution</td>
<td>Sigma-Aldrich</td>
<td>D9905-5MI</td>
</tr>
<tr>
<td>Tween 20</td>
<td>Sigma-Aldrich</td>
<td>P9416-50ML</td>
</tr>
</tbody>
</table>

### Critical Commercial Assays
- High Pure Viral Nucleic Acid Large Volume Kit: Roche 5114403001
- MiMinElute PCR Purification Kit: Qiagen 28006
- NextSeq 500/550 High Output Kit v2: NextSeq 500/550 High Output Kit v2
- NextSeq 500/550 High Output Kit v2
- NextSeq 500/550 High Output Kit v2

### Deposited Data
- Raw and analyzed data: This paper
- ENA: XXXXXXXXXX

### Software and Algorithms
- ADMIXTOOLS: [https://github.com/DReichLab/AdmixTools](https://github.com/DReichLab/AdmixTools)
- smartpca: [https://www.hsph.harvard.edu/alkes-price/software/](https://www.hsph.harvard.edu/alkes-price/software/)
- ADIXMIXTURE: [http://www.genetics.ucla.edu/software/admixture/](http://www.genetics.ucla.edu/software/admixture/)
- haplogrep 2: [http://haplogrep.uibk.ac.at/](http://haplogrep.uibk.ac.at/)
- Contamination

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for resources and reagents should be directed to Wolfgang Haak (haak@shh.mpg.de).
Experimental model and subject details

We generated new genome-wide data from skeletal remains of eleven prehistoric individuals from the Belgian Upper Paleolithic and Iberian Late Upper Paleolithic, Mesolithic, Early and Middle Neolithic.

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 (Aurigniacian, Gravettian and Magdalenian). The site was previously described and eight individuals were analysed by Fu et al. 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 yrs 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. (metres above sea level) in the Serra de Busa Pre-Pyrenean range (Navès, Lleida). After an initial test pit where Late Upper Paleolithic remains were recovered [32], the site was excavated between 2001 to 2008 [33,34]. Two main chrono-cultural phases were defined. The oldest dates back to the Late Upper Paleolithic (15,000–11,000 yrs cal BP) and the youngest corresponds to the Early Mesolithic (11,000 – 9,500 yrs cal BP). The two chrono-cultural units are separated by a big fallen boulder which sealed the Late Upper Paleolithic levels [35]. 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 yrs cal BP (Ua-34297) and 12,830–10,990 yrs cal BP (Ua-34298) [36]. 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 [37]. 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 [36]. The material cultural artifacts recovered from the same level as the human remains have been attributed to the Azilian Culture [35]. However, in general this culture is considered to be more common in Vasco-Cantabrian northern Iberia and on the other side of the Pyrenees [38]. Balma Guilanyà shows clear technical parallels with the near Azilian site Balma Marguineda [39]. 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. [40]. 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 [41]. 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 [42]. These features have been interpreted as a systematic occupation of the estuarine areas, which is also
reflected in the shell midden conformation. The lithic assemblage is characterized by microburin technique and geometrics [44]. The Moita do Sebastião site was excavated by different archaeologists since the last century. The total MNI is unknown, but it could reach up to 100 individuals when summarizing the different campaigns [42]. In this study, we genetically analyse one Mesolithic individual from this site:

- CMS001, adult individual, (7,240 ± 70 BP [To-131], 8,185–7,941 yrs cal BP [2-sigma value]) [40].

**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 yrs cal BCE; Ib 5,300–5,000 yrs cal BCE) [44]. Both levels show a full Neolithic package consisting of domestic fauna [45], Cardial pottery [46] and schematic rock art painted on pebbles [47,48]. The earliest Neolithic sites in the Iberian Peninsula are located in coastal areas [49,50]. 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 [51,52]. 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 [44]. 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 [53]. 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 yrs cal BP [2-sigma value]) [46].
- CHA002, adult (6,227 ± 28 BP [MAMS 29127], 7,250–7,018 yrs cal BP [2-sigma value]) [53].
- CHA003, infant (6,180 ± 54 BP [D-AMS 015821], 7,245–6,947 yrs cal BP [2-sigma value]) [53].
- CHA004, adult (5,645 ± 31 BP [MAMS 28128], 6,494–6,321 yrs cal BP [2-sigma value]) [53].

**Fuente Celada (Early Neolithic)**

This site was described in Szécsényi-Nagy et al. [40]. Fuente Celada is an open-air settlement located in the northern Iberian Central Plateau (Quintanaduena, Burgos). All the archaeological materials are from a rescue excavation carried out in 2008 by Alameda Cuenca-Romero et al. [54]. 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 [54]. Here we include this Early Neolithic individual in our genetic analyses:

- FUC003, adult (6,120 ± 30 BP [UGA-7565], 7,157–6,910 yrs cal BP [2-sigma value]) [54].
Cova de Els Trocs (Middle Neolithic)

This site was also described in Szécsényi-Nagy et al. [40]. Cova de Els Trocs is a cave located in Northeastern Iberia, at 1,564 m.a.s.l. in San Feliú de Verí (Bisauri, Huesca) in the South of the Axial Pyrenees [55]. 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 [55]. The first phase corresponds to the Early Neolithic (ca. 5,300–4,800 yrs 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 yrs 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 yrs cal BCE) the cave was used as a burial site. 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 yrs cal BP [2-sigma value])
- ELT006, adult (5,035 ± 23 BP [MAMS-16165], 5,895–5,716 yrs cal BP [2-sigma value])

Ancient DNA processing and quality control

Sampling of ancient human remains

For the ancient individuals analysed 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 in the Institute of Anthropology, Johannes Gutenberg University, Mainz (Data S2). 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 S2). Teeth surfaces were cleaned with a low concentration bleach solution (10%). 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 [56]. For the teeth sampled in Mainz the complete tooth was grinded down using a mixer mill [40].

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 H₂O 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 H₂O 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 strand (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 [57]. 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 S2 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 (mito-capture)**

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 [58] 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, [59]) and for 1,196,358 informative nuclear SNP markers (1240K capture, [60]).

**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 [61]. 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) [62] 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 S3 and for mtDNA captured libraries in Data S4.

**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 [58]. 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 S3). 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] (Table S3). We find contamination rates below 2.2% for all libraries (Data S6). We visualized the mitochondrial read alignment with Geneious R8.1.974 [63] 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 S3). 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 [64] to facilitate contamination estimation of the merged libraries (Data S3). Finally, merged libraries with less than 3.3% contamination were selected for population genetic analysis.

Genotyping and merging with data set.

After trimming of potentially damaged terminal ends bamfiles were genotyped with pileupCaller (https://github.com/stschiff/sequenceTools/tree/master/srcpileupCaller), which call one SNP per position considering the human genome as pseudo-haploid genome. Genotyped data were merged with Human Origins panel (~600K SNPs) [27] 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 S4.

Kinship relatedness and individual assessment.

We used Relationship Estimation from Ancient DNA (READ) to estimate the degree of genetic kinship relatedness among individuals [65]. 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. We found that individuals BAL001 and BAL005 were identical and consequently merged. All downstream analyses were performed with the merged version BAL051.

Phenotypic traits analysis

We called specifically some SNPs positions associated with some phenotypic traits from the captured SNPs (e.g. lactase persistence, pigmentation, eye colours) (Figure S4). We calculated the genotype likelihood based on number of reads of each specific position to determinate the presence of the ancestral or derived alleles in homozygous or heterozygous fashion [24]. Results are shown in Figure S4.
Y-chromosomal and mitochondrial haplogroups

We used samtools v1.3.1 to extract reads from mitocapture data [64] and mapped them to the rCRS and called the consensus sequences using Geneious R8.1.974 [63]. We downloaded these consensus sequences in fasta format and they were used to determine mitochondrial haplotypes using Haplogrep 2 [66] (Data S6).

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/tree/master/srcpileupCaller), and mapping quality ≥30 and base quality ≥30 (Data S5). Y haplogroups were determined using the program Yhaplo [67].

Population genetic analysis

Labelling 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. [21]. We included the HG with more than 15,000 SNPs covered in the PCA (Supplemental Item 1). If the hunter-gatherer 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 [14,21,23]. 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 [9,13,20–22,26].

Principal Component Analysis

PCA analysis was run with the Human Origins data set using smartpca v10210 (EIGENSOFT) with the option SHRINKMODE [68] using 777 modern populations to calculate eigenvectors on which aDNA samples were projected [20]. 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 Q-116-1, 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 GoyetQ2 and Villabruna-related ancestry. We then modelled the ancestry in the HG groups as a mixture of GoyetQ2 and Villabruna. Alternatively, we also used El Miron 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).

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 of 8.63552043e-92).
The results of all qpWave and qpAdm models are reported in tables S8 and S9. 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-value 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 \([1-f3(X; X, Mbuti)]\) pairwise values among all the combinations [2].

**Data Availability**

will be released on ENA at the time of publication.
Legends of supplemental figures and data files

**Figure S1.** Heat plot showing the genetic distances between Eurasian HG. Genetic distances were calculated using $f_3$-outgroup statistics of the form $f_3(X;Y, Mbuti)$, 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. [1].

**Figure S2.** PCA analysis calculated with 777 present day West Eurasians [3] with option shrinkmode:YES on which HG individuals were projected.

**Figure S3.** A) Modelling European HG as admixture of Villabruna-like and El Mirón-like ancestry using *Loschbour* and *El Mirón* as proximal sources, respectively. B) Map showing the distribution of the main sites related to the Azilian techno-complex, adapted from Strauss [4]. Note that Iberian HG restricted to the Cantabrian and the Pre-Pyrenean regions carry more Loschbour-like ancestry (e.g. Balma Guilanyà, La Braña 1, and Canes 1) than Iberian HG from outside this area (Chan and Moita do Sebastião individuals). The increase of Villabruna ancestry is first attested with the Balma Guilanyà individual, recovered from an Azilian archaeological level. C) $f_4$-statistics showing no affinity between Geometric Mesolithic Moita do Sebastião from Portugal and Iberomaurusian HG from Taforalt, Morocco, North Africa. D) Correlation between Villabruna-like ancestry (dark blue; Figure 3D) and Loschbour-like ancestry (light blue; Figure S3A) and time. 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. 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 Data**

Excel spreadsheet containing

Data S1. Chrono-cultural information
Data S2. Sampling method and library treatment
Data S3. Ancient DNA information for 1240K captured libraries UDG-half treated. DNA Capture (%) represents the percentage of overlapped reads with the targeted 1240K capture SNPs.
Data S4. Ancient DNA information for mtDNA captured in UDG-half treated libraries. DNA Capture (%) represents the percentage of overlapped reads with the targeted mitochondrial capture SNPs.
Data S5. Sex determination and nuclear contamination estimation on X Chromosome in male samples.
Data S6. Mitochondrial contamination and haplogroup determination
Data S7. Number of SNPs called and % SNPs covered in 1240K and HO panel.
Data S8. $F_4$-statistics
Data S9. qpAdm models for HG individuals
Data S10. qpAdm models for Neolithic individuals
Figure 1

A

B

Years cal BP

5,000
6,000
7,000
8,000
9,000
10,000
11,000
12,000
13,000
14,000

Balma Guilanyà (Late Upper Paleolithic)
Moita do Sebastião (Mesolithic)
Cueva de Chaves (Early Neolithic)
Fuente Celada (Early Neolithic)
Cueva de Chaves (Middle Neolithic)
Cova de Els Trocs (Middle Neolithic)

1 100 Km.
S1. Exploring the genetic structure of Iberian hunter-gatherers

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 [1]. 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 [1]. Interestingly, González-Fortes et al. [2] reported two new Mesolithic individuals from Iberia, named after the respective sites Chan (from Chan do Lindeiro) and Canes 1 (from Los Canes), noting that the Chan individual differed from the predominant WHG ancestry profile without exploring this observation further.

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 characterise 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, and (iv) Neolithic groups that followed different routes of expansion can be identified by the structure in their respective HG ancestry profiles.

We performed f3-outgroup statistics of the form f3(HG1, HG2, Mbuti) to measure the shared genetic drift among HG individuals in our dataset. In doing so, we obtained the same clustering pattern (Satsurblia, Mal'ta, Věstonice, El Mirón and Villabruna) concordant with Fu et al. [1] (Figure S1).

Most Iberian HG are grouped either with the previously defined El Mirón cluster or fall within the wider GoyetQ2 cluster, which formed the rationale for additional testing of the robustness of this cluster (e.g. our newly reported Moita do Sebastião [~8ky cal BP], Balma Guilanyà [~12ky cal BP], and Chan [~9ky cal BP], whereas La Braña 1 and Canes 1 cluster with Villabruna. Formal statistics to test the clustering of Iberian HG to the GoyetQ2 cluster are shown in the main manuscript. The positions of the Iberian HG on the PCA plot (Figure S2) corroborates this pattern, with the Iberian HG forming a cline between WHG (i.e. Villabruna) and the GoyetQ2 cluster.

S2 Modelling admixture proportions in Iberian HG and WHG with qpWave/qpAdm

As shown by analyses presented in the main manuscript El Mirón is best explained as being admixed between individuals related to the Villabruna and GoyetQ2 clusters (Figure 3), despite being chronologically older than the best representative of each cluster, respectively, at least given the currently available fossil genetic record. Here we explore the implications of our genetic findings in view of the archaeological record in southwestern Europe.

S2.1 Iberian HG population structure and its correlation with the Azilian techno-complex

We also repeated the same admixture models as presented in the main manuscript (Figure 3D) using El Mirón instead of Goyet Q-2 as a geographically closer proxy for the non-Villabruna ancestry and Loschbour as a source of Villabruna-like ancestry (Figure S3A). Most of these models are statistically well supported, which suggests that El Mirón can serve as a local proxy for Iberian HG and explains why Chan and Moita do Sebastião are genetically similar to El Mirón (Supplementary section S2.1, Figure S3A). This also implies that the admixture between
the GoyetQ2 and Villabruna clusters must have happened earlier, given the chronologically older date of El Mirón. We also observe a subtle El Mirón-related ancestry contribution outside the Iberian Peninsula in France (e.g. Rochedane 9.8 ± 8.3 %, and Iboussieres 25-1 20.7 ± 9.1 %) (Supplementary section S3.1, Figure S3A, Data S9), whereas Goyet Q-2-like ancestry is present in a higher number of individuals outside Iberia (Figure 3D). Inside Iberia, we notice an additional contribution of Villabruna-like ancestry visible in the 12,000-year-old Balma Guilanyà individual, which could reflect the transition to the Holocene and ‘Azilian’ techno-complexes in northeastern Iberia.

The so-called ‘Azilian’ is the dominant techno-complex in Vasco-Cantabria [4] and the Pre-Pyrenean region of the Iberian Peninsula [5,6], and in central and southern France [7,8] during the Pleistocene-Holocene transition around 14,000 cal BP. This geographic region formed a natural corridor, which connected Iberia and France and where the Azilian techno-complex was shared, but at the same time was absent in the rest of the Iberian Peninsula (Figure S3B). Traditionally, the transition to the Azilian has been viewed as a direct result of the climatic amelioration, which took place during the Bolling/Allerød warming phase (14,700 -12,700 BP) [9] when HG started to develop a broader subsistence spectrum, specializing in small prey hunting and combining terrestrial and aquatic resources; a trend, which then continued during the Holocene [10]. It is also in this period that the naturalistic art paintings, very common in the region, disappeared, and were replaced by geometric motifs mostly found in mobile art [11]. The material assemblage also presents changes with respect to the preceding Magdalenian culture. Classically, the Azilian techno-complex was defined by the presence of flat, asymmetrical harpoons, painted cobles, thumb-end scrapers, and curve-backed ‘Azilian’ points. This set has been re-defined, incorporating new elements and criteria such as a generalized simplification of the knapping method combined with a microlithization/miniaturlization effect, in particular of backed points and end scrapers [12–14]. However, in many of these Azilian sites the chrono-stratigraphic continuity between the Magdalenian and Azilian levels makes it difficult to establish a clear rupture between both phases (Figure S3B).

The genomic data available for Iberian HG up to this time supports a gradual replacement of the genomic structure present in northern/northeastern regions of Iberia during the Magdalenian period (represented by the El Mirón individual). Indeed, the Iberian HG individuals with more Villabruna-like ancestry are from the Vasco-Cantabrian and Pre-Pyrenean regions (Iberia) and the increasing shift to Villabruna-like ancestry could be associated geographically and chronologically with the dispersion of new HG technologies along the Cantabrian corridor during Azilian and Mesolithic times, exemplified by individuals such as Azilian-associated Balma Guilanyà. Villabruna-like ancestry becomes even stronger during the Mesolithic in Cantabria (La Braña 1 and Canes 1) suggesting additional HG flux into north/northeastern Iberia, which again must have had a higher impact in this region. Mesolithic HG outside this region retain more ancestry shared with the GoyetQ2 cluster and have a higher, significant amount of El Mirón-like-ancestry (Chan from Galicia or Moita do Sebastião in Portugal) when modelled with El Mirón as a local source (Figure S3A and S3D).

S2.2 No African-Iberian connections during Mesolithic times

Another question that is heavily debated among archaeologists and that can be addressed based on our results is the potential African-Iberian connection during Mesolithic times, or more precisely during the period of the Geometric Mesolithic. The main characteristic technological change during the Late Mesolithic is the shift to the Blade and Trapeze industry, in Iberia commonly grouped as Mesolithic Geometric (Trapezes and Triangles phases). Although the distribution of this phenomenon is spread widely along eastern, western and North Africa, its origin is still debated [15]. 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 [16]. Other authors posit Crimea as the place of origin for the Blade and Trapeze industries [17]. Inside
Iberia this cultural phenomenon is distributed along both the Mediterranean [18] and Atlantic [19] coastlines. This contrasts with the situation in northern Iberia where Mesolithic Geometric flint tools are found at significantly lower numbers [20] with the exception of the Ebro Valley, a natural corridor between the Mediterranean region and Cantabria [21]. Interestingly, this is the geographic region and the time period where and when the Villabruna/Loschbour-related ancestry increases (e.g. shown in individuals from La Braña 1 and Canes 1), in contrast with the individual Moita do Sebastião from Portugal where the Geometric Mesolithic was predominant [22] and which shows a similar genetic ancestry to El Mirón.

Importantly, our genomic results from Moita do Sebastião, attributed to the Geometric Mesolithic, show a strong affinity with El Mirón but lack any genetic evidence in support of the hypothesis of an African origin for the Mesolithic Geometric (Figure S3C). Here, the f4-statistics of the form f4(Moita do Sebastião, WHG; Taforalt, Mbuti) in Figure S3C show no extra genetic affinity of Moita do Sebastião to Pleistocene Iberomaurusian HG from Taforalt (North Africa) [23]. Taforalt individuals are considered 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 [24]. The fact that the genomic structure at the Atlantic coast has not significantly changed since the Magdalenian currently supports the idea of cultural diffusion or analogous technological changes rather than a direct gene-flow. However, more data, especially from the Mediterranean Mesolithic Geometric, is needed to address this question in more detail.

S3. Phenotypic traits

We extracted a list of 36 SNPs of functional importance or related to known phenotypic traits [23] and estimated the probability of carrying the ancestral or derived alleles based on the number of reads after using a quality filter q30. We checked different SNPs positions on the gene OCA2 related to light eye colour. We obtain the ancestral allele (rs12913832, 3 reads) in the CHA001 and ELT002 individuals, and heterozygous values for ELT006, suggesting dark colour eyes for all of them. The SNP coverage for the remaining individuals was not sufficient. Another allele from the same gene (rs1800404) related to eye pigmentation could support the 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 colour 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 [25], 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).

S4. Kinship analysis.

We used the published tool ‘READ’ (Relationship Estimation from Ancient DNA) [26] to estimate the degree of genetic kinship relatedness among individuals (STAR Methods). We split our HG and Neolithic individuals 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.

S5. Y-chromosomal and mitochondrial 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 of the samples ranges from 88.86-99.99% (Data S4). Iberian HG individuals from Balma Guilanyà and Moita do Sebastião belong to the haplogroup U, together with the two Middle Neolithic (MN) individuals CHA004 and ELT006 (Data S6). Individual BAL003 could be assigned to U2’3’4’7’8’9’, also found in the Paglicci 108 (~27,000 yrs cal BP, Italy), Rigney 1 (~15,500 yrs cal BP, France) [1], and Grotta d’Oriente C HG (~14,000 yrs cal BP, Italy) [27]. Individual BAL0051 belongs to U5b2a, also found in Neolithic Scotland [28]. Moita do Sebastião (CMS001) carries haplogroup U5b1, which was also reported from Middle Neolithic, Bell Beaker and Middle Bronze Age individuals from Portugal and Spain [28,29], and in the Ranchot 88 HG (~10,000 yrs cal BP, France).

Early Neolithic individuals from Cueva de Chaves present non-U haplogroups. Individual CHA001 carries haplogroup HV0+195, but was previously reported as K based on PCR-based results [30]. This haplogroup was previously found in MN Ireland [31], as well as MN Scotland and Bell Beaker individuals from England [28]. Individual CHA002 was assigned to K1a2a, which is common in Early Neolithic Iberia, e.g. Cova Bonica [32], Cova de els Trocs [33] and Cueva del Toro [24], but also in Chalcolithic and Bell Beaker individuals from Iberia and Italy [25,28]. CHA003 was assigned to K1a3a, so far reported from Neolithic Anatolia [25], Neolithic Scotland, and Chalcolithic Scotland, and Bell Beaker individuals from Sicily [28]. Middle Neolithic individual CHA004 carried haplogroup U4a2f, found in HG from the Iron Gates, Romania, and Lithuania [27,34]. The Middle Neolithic ELT002 individual carries haplogroup J1c1b, present in the Koros Early Neolithic [35], Neolithic from Scotland [28], Iberian Late Neolithic-Chalcolithic [29], and Bronze Age from Italy and Germany [36]. ELT006 was assigned to haplogroup U3a1, which has been reported from Middle Neolithic France and Germany [27,28], and Chalcolithic Iberia [25]. Early Neolithic Fuente Celada carries haplogroup X2b+226, found in MN Hungary [37] and Middle Bronze Age Iberia [29]. X2b was also found in Iberian Late Neolithic [38], Chalcolithic [35] and Bell-Beaker individuals [28], Neolithic England [28] and Greece [39], and the Early and Late Neolithic from Morocco [24].

Y chromosome haplogroups were called from the list of Y-SNPs including in the 1240K capture assay using the script yhaplo [40]. BAL0051 could be assigned to haplogroup I1, while BAL003 carries the C1a1a haplogroup. To the limits of our typing resolution, Early and Middle Neolithic individuals CHA001, CHA003, ELT002 and ELT006 share haplogroup I2a1b, which was also reported for the Loschbour [41] and Motala HG [25], and other Late Neolithic and Chalcolithic individuals from Iberia [28,29], as well as Neolithic Scotland, France, England [28], and Lithuania [34]. Both C1 and I1/ I2 are considered typical male HG lineages prior to the arrival of farming in Europe. Interestingly, CHA002 was assigned to haplogroup R1b-M343, which together with an Early Neolithic 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 [25,27,28,33,36]. 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 Early Neolithic males from Neolithic Anatolia [25], Starčevo, LBK Hungary [35], Impressa from Croatia and Serbia Neolithic [27] and Czech Neolithic [28], but also in Middle Neolithic Croatia [27] and Chalcolithic Iberia [28].
Figure S1. Heat plot showing the genetic distances between Eurasian HG. Genetic distances were calculated using f3-outgroup statistics of the form $f_3(X;Y, Mbuti)$, 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. [1].
Figure S2. PCA analysis calculated with 777 present day West Eurasians [3] with option shrinkmode:YES on which HG individuals were projected.
Figure S3. A) Modelling European HG as admixture of Villabruna-like and El Mirón-like ancestry using Loschbour and El Mirón as proximal sources, respectively. B) Map showing the distribution of the main sites related to the Azilian techno-complex, adapted from Strauss [4]. Note that Iberian HG restricted to the Cantabrian and the Pre-Pyrenean regions carry more Loschbour-like ancestry (e.g. Balma Guilanyà, La Braña 1, and Canes 1) than Iberian HG from outside this area (Chan and Moita do Sebastião individuals). The increase of Villabruna ancestry is first attested with the Balma Guilanyà individual, recovered from an Azilian archaeological level. C) f4-statistics showing no affinity between Geometric Mesolithic Moita do Sebastião from Portugal and Iberomaurusian HG from Taforalt, Morocco, North Africa. D) Correlation between Villabruna-like ancestry (dark blue; Figure 3D) and Loschbour-like ancestry (light blue; Figure S3A) and time. 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. 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.
Supplementary References


http://doi.org/10.1073.pnas.1717762115

