bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The maternal genetic make-up of the Iberian Peninsula between the Neolithic and the Early Bronze Age

Anna Szécsényi-Nagy[1*], Christina Roth[2], Guido Brandt[3], Cristina Rihuete-Herrada[4], Cristina Tejedor-Rodríguez[5], Petra Held[2], Íñigo García-Martínez-de-Lagrán[5], Héctor Arcusa Magallón[5], Stephanie Zesch[6], Corina Knipper[7], Eszter Bánffy[8], Susanne Friederich[9], Harald Meller[9], Primitiva Bueno Ramírez[10], Rosa Barroso Bermejo[10], Rodrigo de Balbín Behrmann[10], Ana M. Herrero-Corral[11], Raúl Flores Fernández[12], Carmen Alonso Fernández[13], Javier Jiménez Echevarria[13], Laura Rindlisbacher[14], Camila Oliart[4], María-Inés Fregeiro[4], Ignacio Soriano[4], Oriol Vicente[4], Rafael Micó[4], Vicente Lull[4], Jorge Soler Díaz[15], Juan Antonio López Padilla[15], Consuelo Roca de Togores Muñoz[15], Mauro S. Hernández Pérez[16], Francisco Javier Jover Maestre[16], Joaquín Lomba Maurandi[17], Azucena Avilés Fernández[17], Katina T. Lillios[18], Ana Maria Silva[19], Miguel Magalhães Ramalho[20], Luiz Miguel Oosterbeek[21], Claudia Cunha[22], Anna J. Waterman[23], Jordi Roig Buxó[24], Andrés Martínez[25], Juana Ponce Martínez[25], Mark Hunt Ortiz[26], Juan Carlos Mejías-García[26], Juan Carlos Pecero Espín[26], Rosario Cruz-Auñón Briones[26], Tiago Tomé[27], Eduardo Carmona Ballestero[28], João Luís Cardoso[29], Ana Cristina Araújo[31], Corina Liesau von Lettow-Vorbeck[32], Concepción Blasco Bosqued[32], Patricia Ríos Mendoza[31], Ana Pujante[32], José I. Royo-Guillén[33], Marco Aurelio Esquembre Beviá[34], Victor Manuel Dos Santos Goncalves[35], Rui Parreira[35], Elena Morán Hernández[35], Elena Méndez Izquierdo[36], Jorge Vega de Miguel[37], Roberto Menduiña García[37], Victoria Martínez Calvo[38], Oscar López Jiménez[38], Johannes Krause[3], Sandra L. Pichler[14], Rafael Garrido-Pena[32], Michael Kunst[39], Roberto Risch[4], Manuel A. Rojo-Guerra[40], Wolfgang Haak[3], Kurt W. Alt[41,42*]

[1]Laboratory of Archaeogenetics in the Institute of Archaeology, Research Centre for the Humanities, Hungarian Academy of Sciences, Budapest, Hungary

[2]Department of Anthropology, Johannes Gutenberg University, Mainz, Germany

[3]Max Planck Institute for the Science of Human History, Jena, Germany

[4]Departamento de Prehistoria, Universitat Autònoma de Barcelona, Barcelona, Spain

[5]Arcadia-General Foundation of Valladolid University, Valladolid, Spain

[6]German Mummy Project, Reiss-Engelhorn-Museen, Mannheim, Germany

[7]Curt-Engelhorn-Zentrum Archäometrie gGmbH, Mannheim, Germany

[8]Romano-Germanic Commission, German Archaeological Institute, Frankfurt am Main, Germany

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[9]State Office for Heritage Management and Archaeology Saxony-Anhalt and State Museum of Prehistory, Halle, Germany

[10]Departamento de Historia y Fil, Universidad de Alcalá de Henares, Alcalá de Henares, Spain

[11]Departamento de Prehistoria, Universidad Complutense de Madrid, Madrid, Spain

[12]Professional Archaeologist, Parla (Madrid), Spain

[13]Cronos SC Arqueología y Patrimonio, Burgos, Spain

[14]Integrative Prehistory and Archaeological Science, University of Basel, Basel, Switzerland

[15]Museo Arqeológico de Alicante-MARQ, Alicante, Spain

[16]Departamento de Prehistoria y Arqueología, Historia Antigua, Filología Griega y Filología Latina, Universidad de Alicante, Alicante, Spain

[17]Departamento de Prehistoria, Arqueología, Historia Antigua, Historia Medieval y Ciencias y Técnicas Historiográficas, Universidad de Murcia, Murcia, Spain

[18]Department of Anthropology, University of Iowa, Iowa City, Iowa, United States of America

[19]Research Center for Anthropology and Health, University of Coimbra, Coimbra, Portugal

[20]Laboratório Nacional de Energia e Geologia, I.P., Museu Geológico, Lisboa, Portugal

[21]Instituto Politécnico de Tomar, Escola Superior de Tecnologia de Tomar, Tomar, Portugal

[22]Bolsista PCI, Coordenação de Ciências Humanas, Museu Paraense Emílio Goeldi, Belém, Brazil

[23]Department of Natural and Applied Sciences, Mount Mercy University, Cedar Rapids, Iowa, United States of America

[24]Arrago Ltd., Barcelona, Spain

[25]Museo Arquelógico de Lorca, Lorca, Spain

[26]Departamento de Prehistoria y Arqueología, Universidad de Sevilla, Sevilla, Spain

[27]Universidade Federal do Pará, Campus Universitário do Guamá, Belém, Brazil

[28]Servicio Territorial de Cultura de Valladolid, Valladolid, Spain

[29]Departamento de Ciencias Sociais e de Gestao, Universidade Aberta, Lisboa, Portugal

[30]Research Centre in Biodiversity and Genetic Resources (CIBIO-InBIO) & Directorate-General for Cultural Heritage (DGPC), Lisboa, Portugal

[31]Departamento de Prehistoria y Arqueología, Universidad Autónoma de Madrid, Madrid, Spain

[32]Arqueología Estudios, Murcia, Spain

[33]Technical archaeologist, Gobierno de Aragón, Zaragoza, Spain

[34]Arpa Patrimonio Ltd., Alicante, Spain

[35]Centro de Arqueologia da Universidade de Lisboa (Uniarq), Lisboa, Portugal

[36]Grupo de Investigación HUM-949 TELLUS, Prehistoria y Arqueología en el Sur de Iberia, Universidad de Sevilla, Sevilla, Spain

[37]ARGEA Ltd., Madrid, Spain

[38]GIPSIA Ltd., Toledo, Spain

[39] German Archaeological Institute Madrid, Madrid, Spain

[40]Departamento de Prehistoria y Arqueología, Universidad de Valladolid, Valladolid, Spain

[41]Center for Natural and Cultural History of Man, Danube Private University, Krems, Austria

[42]Department of Biomedical Engineering and Department of Environmental Science, Basel University, Basel, Switzerland

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*Corresponding authors: szecsenyi-nagy.anna@btk.mta.hu, kurt.alt@unibas.ch

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Agriculture first reached the Iberian Peninsula around 5700 BCE. However, little is known about the genetic structure and changes of prehistoric populations in different geographic areas of Iberia. In our study, we focused on the maternal genetic makeup of the Neolithic (~ 5500-3000 BCE), Chalcolithic (~ 3000-2200 BCE) and Early Bronze Age (~ 2200-1500 BCE). We report ancient mitochondrial DNA results of 213 individuals (151 HVS-I sequences) from the northeast, middle Ebro Valley, central, southeast and southwest regions and thus on the largest archaeogenetic dataset from the Peninsula to date. Similar to other parts of Europe, we observe a discontinuity between hunter-gatherers and the first farmers of the Neolithic, however the genetic contribution of hunter-gatherers is generally higher and varies regionally, being most pronounced in the inland middle Ebro Valley and in southwest Iberia. During the subsequent periods, we detect regional continuity of Early Neolithic lineages across Iberia, parallel to an increase of hunter-gatherer genetic ancestry. In contrast to ancient DNA findings from Central Europe, we do not observe a major turnover in the mtDNA record of the Iberian Late Chalcolithic and Early Bronze Age, suggesting that the population history of the Iberian Peninsula is distinct in character.

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

The changeover from a hunter-gatherer lifestyle to a productive mode of subsistence first emerged around 10,000 BCE in the Near East[1–3]. This so-called 'Neolithic transition' brought about fundamental changes in economy, social structure, demography and human health, and laid the foundations for agrarian societies and thus for ancient civilizations. Over the course of the 7th and 6th millennia BCE, agriculture spread from the Balkans to Central Europe. Another route of dissemination ran along the Mediterranean coastlines of Greece, Italy and the south of France to the Iberian Peninsula and north to the Paris Basin and Central Europe[4]. However, the process of Neolithisation was non-linear, with the archaeological record documenting influences of local cultural traditions[5]. In Iberia, the Neolithic transition, which began around 5700 BCE, appears to have been complex, and Mesolithic and farming communities coexisted and interacted for as long as 2 millennia[6].

The artefactual remains, mainly ceramic, attest to the different origins and modes of Neolithisation on the Iberian Peninsula. On one hand stands a Mediterranean maritime colonization by Neolithic pioneers characterized by ceramic with clear parallels to the *Ligurian Impressa* collections of Italian origin[7–9]. Some Early Neolithic sites were also located in the hinterland, suggesting further routes of dissemination through the Pyrenees and/or along major rivers, such as the Ebro[10,11]. On the other hand, North African influences and contacts are tangible in the southern Iberian Neolithic[12,13]. All in all, Iberia appears a melting pot of influences and groups, combining Neolithic lifeways and with indigenous mechanisms of adaptation[14].

In the Early Neolithic, we can observe common features shared over large areas, but also some regionally restricted phenomena[15]. In the whole territory, for example, there existed sophisticated systems of agriculture and livestock handling, with adaptable crops[16,17] and seasonal strategies in flock management[18–20]. The groups of the *Franco-Iberian Cardial* and the *Epicardial* pottery styles appeared at this time and recent studies have revealed mutual diachronic influences in the material culture and economies of these cultures[15,21].

An increasing number of burials from that epoch have come to light in the last years[22–24]. The oldest are individual inhumations, sometimes grouped in cemeteries as in Los Cascajos[25]. It is also common to find human remains in caves, which were increasingly used for collective burials[26]. From the late 5th millennium BCE onward, burial monuments appeared, and megalithic tombs became widespread[27]. This phenomenon links Iberia with other parts of Europe, indicating long-distance networks of communication. Meanwhile, certain parts of northeast Iberia maintained individual inhumations in pits, mostly in small cemeteries[28]. Besides megalithic tombs, ditched enclosures

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

extending over more than 100 hectares started to dominate the landscapes of southern and central Iberia from 3300 to 3100 BCE, highlighting another widely spread European phenomenon[29,30]. During the Iberian Chalcolithic period (3000-2200 BCE), fortified settlements with stone walls and semi-circular bastions appeared in the western and southern parts of the Peninsula, while elsewhere, open settlements were still extant[31]. The diversity in settlement and burial types suggests the existence of social structures with different levels of complexity[32]. At the same time, as exchange networks, circulated precious goods such as ivory from Africa and even Asia to Iberia[33].

From ca. 2600 BCE onwards, the so-called 'Bell Beaker Phenomenon' became manifest with its characteristic pottery, copper weapons, gold ornaments, and other prestige goods, an archaeological reflection of important social and economic changes which spread across vast regions of western and central Europe[34]. Iberia's Bell Beaker assemblages are among the richest and most diverse in western Europe[35], both in terms of settlements and burials[36]. It has thus long been a focus of archaeological research, commencing from migrationist hypotheses and leading up to current social explanations, where the Bell Beaker phenomenon is perceived as a package of prestigious objects exchanged and consumed by elite groups and displayed on special occasions[37].

Around 2200 BCE, the Chalcolithic settlement and funerary practices were suddenly discontinued, particularly in the western and southern part of Iberia, where most of the ditched and fortified settlements were abandoned and collective sites and megalithic tombs were replaced by individual burials[31]. El Argar groups began to emerge in southeast Iberia, with large and massively fortified urban centers like La Bastida (Murcia)[38], which managed to control a territory of over 35,000 km^2 during the following 650 years.

With regard to its population history, the Iberian Peninsula has been the focus of several recent archaeogenetic studies[39–46]**.** The mtDNA gene pools of pre-Neolithic hunter-gatherers along with early Neolithic farming populations were first compared in present-day Portugal[47]. More recent studies focusing not only on mtDNA but also on Y chromosome markers have supported the model of a pioneer colonization in the north-eastern coastal regions of the Iberian Peninsula at the onset of the Iberian Neolithic[39,48]. On the other hand, the northern part of Spain (Cantabrian fringe) attested to a rather complex Neolithic transition[42]. Recent mtDNA and genome-wide analyses have put an emphasis on the genetic affinity and shared Near Eastern ancestry between the early Iberian farmers and the contemporary Central European Linearbandkeramik (LBK) population[44,49,50]. The latest mtDNA and genomic studies have revealed increased subsequent admixture of hunter-gatherer elements during the local middle Neolithic (La Mina; Alto de Reinoso)[46,50] and the Chalcolithic (El Portalon; El Mirador) again reminiscent of processes observed in Central Europe[41,51].

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Despite numerous research projects being carried out over the past years, the Neolithic settlement history of Europe can still only be explained at a broad scale[50–55]. Regional transects though time detailing the developments and the course of the Neolithic in central Germany[56–58] and in the Carpathian Basin[59] have mostly been examined by mitochondrial DNA (mtDNA) control region data. Our project completes the latter series by focusing on the archaeological models and hypotheses that have been put forward for the Iberian Peninsula, and where diachronic (i.e. 'through time') sampling of ancient DNA allows the detection of demographic changes and discontinuities between 5700 and 1500 BCE. Essential research questions focus on three levels: i) the individual sites, ii) the Iberian Peninsula as a scene of Neolithic transition and iii) a comparison with contemporaneous ancient and modern-day Europeans. A key question of our study of the mtDNA diversity on the Iberian Peninsula through time was to examine to what extent regional and supra-regional cultural groups could be recognized as genetically identifiable entities, as shown in other areas of Europe[49,56]. A related question was whether the cultural breaks that can be seen, for instance, at the end of the Neolithic and Chalcolithic periods, were also accompanied by human population turnovers.

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

We processed ancient DNA samples of 318 human individuals from prehistoric Iberia and generated reproducible mitochondrial hypervariable region haplotypes (HVS-I, np 16020-16401) from 151 individuals following strict authentication criteria (see Methods). MtDNA haplogroup classification of further 62 samples was based on multiplex typing single nucleotide polymorphisms (SNPs) (Fig. 1, Supplementary Table S1-4). The DNA amplification and reproduction success of the HVS-I showed strong differences among samples from diverse regions of the Iberian Peninsula: the highest amplification rates were observed in northeast Spain (NEI), the middle Ebro Valley (MEV), and the geographically high-lying regions of central Spain (CI) (71-78%), while especially the southern part of the Iberian Peninsula (southeastern Iberian (SEI) group and southwest Iberia (SWI) had very low amplification success rates (20-43%).

We merged these data with previously published HVS-I mtDNA results from 182 prehistoric individuals (Supplementary Table S5), and separated the HVS-I dataset of 322, and mtDNA haplogroup information of 395 prehistoric Iberian individuals into geographically and chronologically defined groups (see Methods). The resulting haplogroup compositions of these groups are presented in Fig. 2.

Haplogroup based analyses

The Iberian pre-Neolithic mtDNA substrate is characterized by high frequencies of H (43.8%) and U5b (37.5%), whilst haplogroups N (12.5%) and U4 (6.3%) were also common. This composition is different from other continental hunter-gatherer groups due to the lack or low proportion of haplogroups U2 and U5a, and the high abundance of H[42,43,45,47].

The haplogroup composition of the Iberian Early Neolithic (EN) population shows similarities to the Early Neolithic data from Anatolia, the Carpathian Basin (6500-4900 BCE), and Central Europe (5500-4000 BCE, represented by the central German Mittelelbe-Saale region)[51,56,59]. Haplogroups K, J, T2, HV, V and X are observed in comparable frequencies in Iberian and Central European groups. However, the proportion of haplogroup H is higher in the Iberian Early Neolithic (40.4%) than in Central Europe (15%), while the frequency of N1a is very low (1.8% compared to 9.3% in Central Europe). Another difference with regards to Central Europe is the occurrence of the N* haplogroup in Neolithic Iberia, which is already present in pre-Neolithic times (Supplementary Tables S6).

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We used the haplogroup frequencies of the studied groups for principal component analyses (PCA, see Methods). When compared to published ancient DNA data from the Near East, the western Anatolian Neolithic population is most similar to Central European Neolithic populations and to the Early Neolithic of Northeast Iberia (Fig. 3, Supplementary Fig. S2-3), but it also shows some affinities with further Iberian farmers through common EN haplogroups (e.g. K, J, T2)[51]. The presence of haplogroup N* in the Pre-Pottery Neolithic in Syria and northeast Iberian Neolithic can also be interpreted in this light[39,60], but since N* is also observed in hunter-gatherers of western Iberia[47], the link is not clear.

The Neolithic farmers of the middle Ebro Valley (MEV) show more subsisted hunter-gatherer (HG) lineages (H, U5b, U in 76.1%) than the early farmers of northeast Iberia (Fig. 3, Supplementary Fig. S2-3). However, the haplogroup compositions of Iberian hunter-gatherers and Neolithic MEV farmers are significantly different in the Fisher's exact test (Supplementary Table S7), and also the test of population continuity does not support direct continuity between hunter-gatherers and farmers in this region (Supplementary Table S8). Following the MEV group, the southwest Iberian Neolithic dataset shows the second highest frequency of hunter-gatherer-type haplogroups, (62% H, U, U5a and U5b) compared to other Iberian regions, but the population continuity test nonetheless supports population change in this region and shows that the new haplogroups cannot be explained by drift. The number of Chalcolithic SWI samples is still too small to estimate the population dynamic of the latter period, when farmer-type haplogroups (K, J) become seemingly more common (with ~27.6%) (Supplementary Table S6).

As Middle Neolithic (~ 4500-3500 BCE) and Late Neolithic (~ 3500-3000 BCE) periods of most sites are not separable at the current state of archaeological research, we combine these data in our analyses (MLN). Fisher's exact test confirms the connection between the EN and MLN periods of Iberia in general, but the differences become significant (p= 0.0465) considering larger population size in the EN-MLN population continuity test. This signal indicates weak population genetic changes, likely explained by the assimilation of the local indigenous hunter-gatherer population into farming communities (Supplementary Table S7-8). During this interval haplogroups U4, U3, and I appeared in northeast Iberia. Curiously, haplogroup I appeared in Central Europe with the 3rd millennium BCE Corded Ware groups and was associated with ancestry derived from the eastern steppe[50,56]. It is possible that the I haplotypes in Spain are from a different source, since haplogroup I is also observed in the Early Neolithic of middle Ebro Valley[42] and in Late Neolithic NEI[61].

Haplogroup W was reported from a Late Neolithic dataset in northeast Spain[61], but has not been found anywhere else in the entire Iberian dataset of 395 prehistoric individuals (Supplementary Table

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S5). Besides W, the occurrence of T2 and U3 explains the Central European affinities of the NEI_MLN along the second component of PCA, but the third component and Ward clustering connect this group to the closer Late Neolithic south French region (Supplementary Fig. S2-3). The EN-MLN population continuity is supported in the northeast by the test of population continuity and also supported by the p=0.1383 value of the Fisher's exact test (Supplementary Table S7-8). Haplogroup U5a appears first in the 4000-3000 BCE period of southwest Iberia, and persists in the following central Iberian Chalcolithic. Our population continuity test supports continuity between the Neolithic (containing mainly Late Neolithic) and Chalcolithic (CHA) populations of central Iberia, and that connection is also supported by the non-significant (p=0.2264) differences in haplogroup compositions (Supplementary Table S7-8). Other typical Central European and Carpathian Basin Neolithic haplogroups such as H5, T1, and U8, which are observed in later Central European Neolithic periods, are missing from the entire prehistoric Iberian dataset. These differences account for the separation of most of the Iberian prehistoric groups from the Central European Neolithic and Early Bronze Age populations along the second component of the PCA, shown on Figure 3.

Some of the Iberian Neolithic mtDNA haplogroups (I, U2, W, N, N1a) are not observed in the successive Chalcolithic Iberian population (n=156), whereas others maintain a steady frequency (V, T2, X) throughout 3500 years. An interesting exception is haplogroup L1b in the Late Chalcolithic Central Iberia at the site Camino de las Yeseras, near Madrid. This group is most frequent in today's West-Central Africa[62], and hints at a connection to the North-West African coasts in prehistoric times.

Tests of population continuity and Fisher's exact test between the whole Iberian Early and Late Chalcolithic datasets result in non-significant p values (p=0.5444-0.5578 and 0.9779), which support population continuity between the two periods. Unfortunately, a clear chronological separation of the Chalcolithic dataset cannot be achieved for all samples and therefore we consider only a subset of the CHA dataset (n=71 for Early and n=47 for Late CHA) in this analysis (Supplementary Table S8). It is also not possible to resolve the Early and Late Chalcolithic transition at regional scale.

We do not observe connections between the Central European Late Neolithic (chronologically comparable to the Chalcolithic in Iberia) and the Iberian (entire and late) CHA groups at the haplogroup level, and thus cannot confirm a Late Chalcolithic expansion toward Central Europe as suggested by haplogroup H mitogenome data[63].

The Early Bronze Age (EBA) sample set in the Iberian prehistoric transect is still very small (n=37 haplogroups from all regions), therefore it has to be merged in PCA and clustering analyses with data from the earlier periods. At the mtDNA haplogroup level the EBA does not show new influences or

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

population changes at the onset of the Iberian Bronze Age according to results from Fisher's test and test of population continuity, when both the entire Peninsula and southeastern Iberia in particular were considered (Supplementary Table S7-8).

A more detailed analysis of H sub-haplogroups focuses on 17 SNPs in the coding region of the human mitogenome. The largest proportion of the H individuals belongs to the subhaplogroup H1 (65.1%), and the second largest group is subhaplogroup H3 (14%), while 18.6% of the H individuals cannot be assigned to any of the subgroups included in the H-PLEX assay (see Methods, Supplementary Table S3-4)[64]. H3 was detected in Chalcolithic individuals from central, southeast and southwest Iberia. H1 was observed in each period and region, but more frequently in the Chalcolithic and Early Bronze Age than in the Neolithic. The comparative ancient H data from the European prehistory is too sparse for in-depth statistical analyses. However it becomes apparent from our results that the H diversity in prehistoric Iberia is different from the H diversity of Central Europe[56,59,63,65], and more similar to the Neolithic populations in France[66,67]. Notably, common Central European subhaplogroups H5 and H7[63,65] have not yet been observed in Southwest Europe.

Haplotype and sequence based analyses

The mtDNA variation in prehistoric Iberia is further explored by HVS-I sequence and haplotype analyses of 322 individuals. The haplotype diversity is higher in the Iberian hunter-gatherer group (Hd=0.952) than in the hunter-gatherer populations of Central Europe (HG_Pleistocene: Hd=0.879, HG_Holocene: Hd=0.931). The diversity decreases in the Early Neolithic of Iberia (Hd=0.898) and increases in the following Middle-Late Neolithic (Hd=0.911) and Chalcolithic periods (Hd=0.944). The Early Bronze Age shows lower diversity (Hd=0.917), although this could be a consequence of the small number of investigated samples (n=16) (Supplementary Table S9).

The haplotype sharing between the larger chronological groups reflects a low level of Iberian hunter-gatherer maternal lineages in the Middle and Late Neolithic periods (7.6%) (Table 1). The higher HG_I genetic contribution in the Early Neolithic (15.8%) is composed of southwest Iberian and MEV Early Neolithic H haplotypes. Note that the level of haplotype sharing between HG_I and EN might be overestimated, since this proportion of sharing includes 'H-rCRS' haplotypes that could not be categorized further in the original publications. During the MLN period several new hunter-gatherer-type U5 and H lineages appeared. Between the two successive Neolithic periods, the haplotype sharing amounts to 50.5%, and 39.0% of the MLN lineages are already present in the EN (detected in ancestral SHA). In the Chalcolithic period, we also observe ca. 39.8% ancestral EN contribution, and

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

only 15.3% of the lineages originate from later Neolithic periods. The continuity of lineages is also seen in the smaller EBA dataset, which shows close connections to MLN and CHA periods, and new lineages appear only in 18.8% (Table 1, Supplementary Table S10).

The amount of lineage sharing is generally high among the Iberian early farming groups (28-68.5%) (Supplementary Fig. S4). The diverse haplotypes of the Central European Early Neolithic are mostly shared with northeast Iberian groups, but the sharing is also high with prehistoric southeast Iberian groups. The lowest lineage sharing of the Central European Early Neolithic is observed in southwest Iberian and MEV groups, where the local autochthonous substrate could have made a larger genetic contribution compared to other study regions. Lineage sharing with Iberian hunter-gatherers is the highest in the Neolithic MEV (23.9%) and in the southeast Iberian (mostly Chalcolithic) group (25.7%). Concerning temporal succession in NEI and CI regions, the proportion of hunter-gatherer lineages increases (from 5.3 to 17.7 %) during the Neolithic in NEI, and also increases from 13.3 to 17.9 % during the Neolithic-Chalcolithic periods in CI (Supplementary Table S10).

Parallel to the increase of hunter-gatherer lineages, we observe permanence of Early Neolithic lineages in the NEI_MLN datasets (52.9% lineage sharing) and also in southeastern and southwestern territories of Iberia (54.3-62.9%). Contrarily, Early Neolithic NEI shares only a few lineages with the MEV and Central Iberian Neolithic (32.6-41.7%). Continuity between Neolithic and Chalcolithic of CI is reflected in the amount of shared lineages between the two periods (61.2%), but CI also shares many lineages with all other Iberian groups (Supplementary Table S10).

Genetic distances (F_{ST}) are generally low among the Iberian prehistoric groups, and none of them are significant (Supplementary Table S11). We observe the shortest distances between the Early Neolithic in northeast Iberia and Central Europe. By plotting the HVS-I based genetic distances after multidimensional scaling (MDS, Supplementary Fig. S5), we observe large-scale relationships. Here, the NEI populations are closest to the Central European EN, while hunter-gatherer-Neolithic/Chalcolithic and Central European Late Neolithic-Early Bronze Age groups cluster as well. Interestingly, MDS shows that some groups (e.g. Yamnaya, MEV_Neo) appear less differentiated from the remaining populations when compared to haplogroup-based PCA. The clustering of the groups on the MDS plot was therefore tested by the analysis of molecular variance. The best-supported groupings (i.e. with the highest variance among the clusters and lowest within the clusters) were calculated with different arrangements of the groups. The Iberian groups differentiated from an EN Central European- Carpathian Basin and a MLN-EBA Central European- LN French cluster, into a central Iberian (CI_Neo, CI_CHA), northeast Iberian (NEI_EN, NEI_MLN, NEI_CHA_EBA) and a geographically mixed Iberian clusters (SEI_Neo_CHA_EBA, SWI_Neo_CHA, MEV_NI_Neo).

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Groups within the latter clusters are also connected through higher haplotype sharing in pairs and elevated hunter-gatherer proportions (Supplementary Table S12).

We further compared HVS-I sequences of three larger chronological groups (EN, MLN and CHA) of the Iberian Peninsula with 133 modern populations. Genetic distances to modern-day populations are generally low, and restricted to certain region(s) of Europe (Supplementary Fig. S6, Supplementary Table S13). The Early Neolithic shows relatively high affinity to modern Lebanon, Georgia, Turkey and Palestine (F_{ST} =0.0034-0.0069), but even higher affinities to several European (Portuguese, French, Italian but also Polish, Irish) populations. The genetic distances to modern populations further increase with the MLN and CHA periods, generally similar to modern-day Europe. Recent genomic studies have highlighted the similarity of early farmers to modern South Europeans[68], especially to modern day Sardinians[50,51]. This picture was not reflected in our mtDNA dataset, where Sardinians ranked only the 57th closest to the Iberian EN group out of 133 modern populations, and even farther from the MLN group (Supplementary Table S13).

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

During the Last Glacial Maximum, the Iberian Peninsula, just as Southeast Europe and the Italian Peninsula, formed a classic Glacial Refuge Area for European populations, well documented by archaeological evidence[69–71]. Therefore, it is assumed that parts of today's European population are the descendants of the residents of these refugia. This theory is supported by coalescence dates (predating hypothesized expansion times) of some mtDNA lineages (H1, H3, V, U5b1), and their frequency peaks in Iberia[72–76]. Although, the Franco-Cantabrian glacial refuge theory has not yet been confirmed by ancient mtDNA or genomic analyses. In addition, the published hunter-gatherer mtDNA H data do not have the necessary resolution to differentiate subgroups of H[42,43,47]. Further Iberian Pleistocene and Holocene H samples need to be investigated at full mitogenomic resolution in order to define, which H lineages were indigenous, i.e. go back to the initial colonization of Europe and prevailed in glacial refugia, and which were brought to the Peninsula by later migrations.

With the end of the Last Ice Age and the beginning of the Holocene ~12,000 BCE, a common European Mesolithic population emerged, which also included a distinct signal of Near Eastern origin[77]. The indigenous Iberian late Upper Paleolithic and Mesolithic populations were represented by mitochondrial haplogroups H, U4, U5b and N* from Northern Spain and Portugal[42,43,45,47]. This currently described maternal genetic makeup of the Iberian pre-Neolithic population differs slightly from Central European hunter-gatherers, where U5b and U5a dominated, but also R, U2, U4, U8, U* were observed[54,58,59,78–80].

The Iberian Peninsula is characterized by diverse landscapes with distinct economic potentials. During the initial phase of Neolithisation, first farmers probably arrived in northeast Iberia primarily along the Mediterranean route and from there spread along the coastline and rivers (e.g., the Ebro Valley) into the hinterland[7–9,81]. Our genetic data support a substantial influx of Neolithic immigrants to northeastern Spain, where farmer lineages are most abundant and diverse, and on to other regions, like the SEI and MEV. In the Iberian Peninsula, the admixture of immigrants with local groups was detectable in the mtDNA haplogroup diversity from the very beginning. This is conspicuously different from the Carpathian Basin[59] and Central Europe[56], where the process of assimilation of hunter-gatherers was at times delayed for centuries[50]. We found the persistence of 'hunter-gatherer' mtDNA haplogroups in the Neolithic to be particularly strong in the middle Ebro Valley and in southwest Iberia. These genetic results match the archaeological data on the mode of Neolithisation described for these areas, so that the fully established Neolithic Iberian communities had distinct hunter-gatherer components[16,82,83]. The further the early farmers advanced into the northwestern and southwestern part of the Peninsula, the higher the proportion of indigenous HG lineages. Geography

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

appears to have been a decisive factor in the advance of Neolithic lifeways. Prehistoric central Iberia (southern and northern Meseta, Ambrona Valley), however, was never an isolated region surrounded by mountains, but rather an important hub for innovations and impulses coming into Iberia[84]. On a genetic level, Neolithic central Iberia showed corresponding mitogenomic connections to northeast and southeast Iberian regions, but significant differences were detected compared to the Neolithic MEV and SWI (Supplementary Table S7).

The Pyrenees are also an area of special interest with regard to the route of immigration of the first farmers. This study includes ten samples from the cave site of Els Trocs, dated to the Early and Middle Neolithic[20]. Due to the considerable temporal differences between the occupations, the respective sample sets can be considered as genetically independent. The ^{14}C data of the six individuals from the earliest phase cluster closely (Supplementary Information). Genome-wide analyses of five of these six individuals revealed the group to be early Neolithic immigrants[50,51], while the typical Iberian HG mtDNA haplogroups (H, U5b, U4 and N*) are missing. Therefore, we suspect the presence of a (still) isolated EN group at Els Trocs. The genetic profile and isolated geographic position in the Pyrenees suggest immigration of this community from north of the Pyrenees (or even from Central Europe via the Rhone Valley) rather than west up the Mediterranean. This alternative gateway to Spain has also been proposed based on archaeobotanical evidence[17,85]. The first and (to our knowledge) sole Iberian appearance of the mtDNA haplogroup N1a in an adult from Els Trocs, which matches an identical HVS-I N1a haplotype in Anatolia and Central Europe, might have arrived on the continental route[50,51,56,86]. These shared roots between Southwest and Central Europe were especially conspicuous in the Northeast Iberian EN group, as the agreement in the common mtDNA haplogroups such as V, J, K, T2 and X suggests. Furthermore, we observe genetic links between Middle Neolithic NE Iberia and SE France on the mtDNA haplogroup level (Supplementary Fig. S3), which fit the archaeological record of both areas showing mutual technological and typological influences[87–89]. However, the mtDNA data from the French site Treilles derived from a small and probably endogamous community, since only 13 haplotypes were present among the 29 typed individuals[67], so that the genetic link to the NEI groups is rather tentative given the small sample size.

The transition from the EN to the MLN is documented in recent ancient genomic studies, which described an increase of hunter-gatherer elements in the farming populations of central Iberia during the MLN period at the site of La Mina and the Chalcolithic sites of El Mirador and El Portalón[41,50,51]. The increase of HG maternal lineages is also observed in our Neolithic NEI data, where the F_{ST} from HG_I decreases and the amount of HG_I haplotypes increases from the early Neolithic to the Middle/Late Neolithic periods in general (5.3-17.7%). The same trend can be seen in our F_{ST} analyses

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where the central Iberian CHA (F_{ST} =0.0535) appear closer to the HG_I group than the CI_Neo group (F_{ST}=0.0685), and the number of shared lineages also increases between HG_I- CI_Neo and HG_I-CI_CHA sample sets (from 13.3 to 17.9%).

The Iberian Middle Neolithic genomic data are still too scarce to draw general conclusions about admixture with late hunter-gatherers[50], and therefore the connection between middle Neolithic and Chalcolithic groups and the dynamics of the hunter-gatherer resurgence must still be studied in greater detail. In this study, we could support the general continuity between MLN-CHA and the observed trend of increasing hunter-gatherer ancestry (Supplementary Table S7-8, S11).

According to the data currently available, the homogeneity of Chalcolithic ancient DNA results suggests that human mobility and genetic mixing had generally increased in Iberia by the Chalcolithic. Although we analyzed four geographically separated CHA groups in the northeast, central as well as in the southeast and southwest of the Iberian Peninsula, they did not exhibit any significant differences except for the comparison of the central and the southeastern group (Fisher's test results in Supplementary Table S7, Supplementary Fig. S1).

The detection of the 'African' Lb1 haplogroup at the Late Chalcolithic site Camino de las Yeseras (Madrid, central Iberia) is remarkable, given that ivory adornments of African origin have also been documented at this and other contemporaneous sites[90–92]. The observation of a western-central 'African' haplogroup alongside African artefacts and raw materials in Copper Age Iberia indicate long distance exchange that at least occasionally seem to have involved mobility of individuals and/or gene flow.

The Bell Beaker phenomenon was a decisive element in the Iberian Chalcolithic, lasting from the Late CHA to the EBA (2600 -1800 BCE)[35,36]. Our tests of population continuity and Fisher's tests between Early and Late CHA periods supported population continuity between the two phases. However, our data structure did not allow studying the Early to Late CHA population changes at a regional scale, so that further studies might reveal local variation with regard to individuals with Bell Beaker / non Bell Beaker cultural affiliations. Interestingly, we did not find novel signs of mtDNA connections between Chalcolithic Iberia and contemporaneous Central Europe. Links between the two regions had previously been suggested based in particular on Bell Beaker elements present across a wide geographic range[93–95], as well as by the maternal genetic analyses of the El Mirador site[40]. None of investigated Chalcolithic individuals show 'steppe ancestry', as seen in contemporaneous Central European Corded Ware and Bell Beaker groups, suggesting that eastern influxes did not reach the Iberian Peninsula until later periods[51]. An evaluation of this unexpected observation will require further in-depth palaeogenetic studies.

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Around 2200 BCE, the emergence of El Argar groups was evidently preceded by a break in Chalcolithic cultural traditions in southeast Iberia. Yet there are no apparent new influences or signals of substantial population change on the mtDNA haplogroup level at this time, so that the observed changes may either be due to an upheaval of existing social structures or an influx of groups that cannot be distinguished from the local population at the present level of genetic resolution, e.g., from southeastern Europe, as previously proposed for El Argar. Unraveling these apparently contradictory data will certainly require further in depth analyses both on the archaeological and the archaeogenetic level.

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conclusion

The present study, based on 213 new and 182 published mtDNA data of prehistoric Iberian individuals, reveals a markedly different mode of interaction between local hunter-gatherers and incoming early farmers during the Early and Middle Neolithic of the Iberian Peninsula, as compared to Central Europe. As a characteristic of Iberian population dynamics, the proportion of autochthonous hunter-gatherer haplogroups was already high among the early Neolithic groups and increased even further in relation to the distance to the Mediterranean coast. In contrast, the early farmers in Central Europe showed comparatively little admixture of contemporaneous hunter-gatherer groups. Already during the first centuries of Neolithic transition in Iberia, we observe a mix of female DNA linages of different origins. Earlier hunter-gatherer haplogroups were found together with a variety of new lineages, which ultimately derive from Near Eastern farming groups. On the other hand, some early Neolithic sites in northeast Iberia, especially the early group from the cave site of Els Trocs in the central Pyrenees, seem to exhibit affinities to Central European LBK communities. This variety of contemporaneous haplogroups suggests that Early Neolithic migrations were not conducted via maritime routes across the Mediterranean only, but could also have included movement along inland routes, e. g. across the Pyrenees from the north. The diversity of female linages in the Iberian communities continued even during the Chalcolithic, when populations became more homogenous, indicating higher mobility and admixture across different geographic regions. Even though the sample size available for Early Bronze Age populations is still limited, especially with regards to El Argar groups, we observe no substantial changes to the mitochondrial DNA pool until 1500 BCE. The expansion of groups from the eastern steppe[50,96], which profoundly impacted Late Neolithic and EBA groups of Central and North Europe, cannot (yet) be seen in the contemporaneous population substrate of the Iberian Peninsula at the present level of genetic resolution. This highlights the distinct character of the Neolithic transition both in the Iberian Peninsula and elsewhere and emphasizes the need for further in depth archaeogenetic studies for reconstructing the close reciprocal relationship of genetic and cultural processes on the population level.

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Materials

The studied sites are distributed across the Iberian Peninsula, with an emphasis on the archaeologically relevant regions on the Mediterranean coasts, in central and northern Spain and in southern Portugal (Fig. 1). We targeted representative sites from the Mesolithic, Neolithic, Chalcolithic and the Early Bronze Age to cover 4,000 years of prehistory on the Iberian Peninsula. Given that the southern European climate is usually unfavorable for DNA preservation, we implemented a flexible sampling scheme contingent on amplification successes, extending sampling from sites with good ancient DNA preservation and/or including additional sites to obtain sufficient data. Sample specific context information were supplied by our colleagues and project partners in Portugal and Spain or collected from previously published papers. All archaeological sites, relevant radiocarbon dates, and individuals incorporated in the present study are listed in Supplementary Table S1-2.

Altogether, 318 individuals from 57 archaeological sites in Spain, Portugal and Morocco were sampled and analyzed for this study. Whenever possible we preferred samples from recent excavations over those that had been held at museum collections for prolonged periods of time. Teeth and bone samples were taken under clean conditions in Iberian museums, at the Institute of Anthropology at Johannes Gutenberg University in Mainz, or directly on site in the case of La Mina, Arroyal, Els Trocs and in part at La Bastida. Mitochondrial profiles of 37 individuals from our project (Alto de Reinoso (n=27), La Mina (n=5) and Els Trocs (n=5) were previously published by our research team[46,50,51]. Here we also report additional samples from the sites of La Mina and Els Trocs.

The chronological classification of the burials was based on the archaeological data. In order to avoid possible problems due to terminological inconsistencies, we used temporal (from Mesolithic to Early Bronze Age; based on relative chronology and absolute dating) and geographic groupings to assess the population mitogenetic data from the Iberian Peninsula. We further distinguished groups based on contextual archaeological evidence, e. g. subsistence strategies, such as hunter-gatherers vs. early farmers. We targeted collecting two to three teeth from each skeleton and sampled bone only when teeth were not available.

Methods

Ancient DNA sample preparation

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Standard sample preparation protocols were used during the ancient DNA work in the Institute of Anthropology in Mainz[56]**.** Samples were first UV irradiated for 30 min each side and then, the surface was removed by shot-blasting with aluminium-oxide-abrasive. Samples were then ground to fine powder using a mixer mill (Retsch). The milling process was controlled by hydroxylapatite blank controls, which were treated as samples in the subsequent DNA extraction and amplification steps. Series of samples were tested from each archaeological site, and the amplification success defined any further ancient DNA work. In general, A and B samples from different teeth or bone fragments per individual were analyzed independently. In the sole case is the site of Cova del Cantal, A and B samples, i.e. repeat extractions, derived from the same bone fragment.

Ancient DNA extraction and amplification

Of the total 48 extraction batches, 25 were performed using the phenol-chloroform DNA extraction method[56] and 23 using the silica based DNA extraction protocol[63].

We combined different sets of primer pairs for the amplification of the HVS-I of the mitochondrial genome. Well-preserved samples were amplified in two segments (np 16046-16401), average preserved samples in four fragments (np 15997-16409) and samples with very fragmented DNA content were amplified with a combination of six overlapping primer pairs. All HVR I-II primers are listed in Table S11 in Alt et al.[46], PCR, amplicon purification and sequencing protocols were the same as reported in Brandt et al.[56]. 40% of the PCR products were cloned using pUC18 vectors and E.coli cells[56,86], and these clones were re-amplified and re-sequenced.

In order to get precise basal mitochondrial haplogroup determinations, we used the GenoCoRe22 multiplex PCR and ABI PRISM® SNaPshot® typing protocol[57], and typed 22 coding region positions of the mitochondrial genome. Samples assigned to the haplogroup H, were further sub-typed used the H-PLEX17 focusing on 17 sub-haplogroup H defining single nucleotide polymorphisms (SNPs)[64]. Note that recent mitochondrial phylogeny updates have complicated the interpretation of typing results: the H1 defining variant G3010A is now also reported for subhaplogroups H30b, H65a, H79a, H105a (Phylotree Build 17).

Samples belonging to haplogroup U were also tested for six coding region SNPs via multiplex amplification and single base extension protocols, and further categorized into subhaplogroups U2, U3, U4, U5, U5a, U5b, and U8 (Supplementary Table S4). Haplogroup T were also further typed in the coding region for information on T1 or T2 sub-haplogroups (C12633A-T1 and A11812G-T2) and sequenced separately[65]. Conditions of the U and T–Plexes are as follows: PCRs were set up in a

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

volume of 25 µI, containing 1 x PCR Gold buffer, 8 mM $MgCl_2$, 0.02 µg BSA, 500 µM dNTPs, 0.02 µM of each primer, 2 U AmpliTaq Gold® DNA polymerase, and 2- 3 µI DNA extract. Cycling conditions were 95°C for 10 minutes, and 35 cycles of 95°C for 30 seconds, 56°C for 30 seconds, and 65°C for 30 seconds, followed by a final adenylation step at 65°C for 10 minutes for the U-Plex and 32 cycles and an annealing temperature of 59°C for the H-PLEX17. Amplification success was checked via gel electrophoresis. Purification and SBE-reactions for the U-Plex was set up according to the GenoCoRe22 protocol.

Individuals with consistent HVS-I profiles from both samples were amplified over the whole HVS-I at least three times from at least two independent extracts, producing between 6-18 independent and overlapping amplicons (depending on the primer pairs used). PCR products that showed ambiguous nucleotide positions were cloned to monitor possible background contaminations and DNA damage. In addition, for those individuals of the sites Alto de Reinoso and La Mina that showed the same HVS-I haplotype, we also amplified and sequenced the HVS-II region via four overlapping primer pairs[46].

Authenticity and contamination

To monitor contamination, sample mix up and other potential errors we used the following authentication strategy: 1. All HVS-I haplotypes were replicated from at least two independent DNA extracts obtained from two skeletal elements from nearly all sampled individuals. The HVS-I fragment was amplified by numerous overlapping PCR amplicons. PCR products with ambiguous sequences were cloned, and 4-8 clones per PCR were sequenced. 2. Due to the applied primer-systems, amplification of the HVS-I enabled the identification of contiguous sequences with 6-10 SNP calls in the overlapping regions. 3. The final haplogroup call was based on several independent amplifications targeting different loci of the mitochondrial genome (HVS-I and coding region SNPs). In all reproducible and reported cases, the HVS-I sequences confirmed the haplogroup assignment by the GenoCoRe22 SNP typing (and vice versa), as did the results of the H-PLEX17, U-Plex and T-Plex assays. 4. All haplotypes could be placed in the mitochondrial phylogeny (www.phylotree.org). In all cases where ancient haplotypes differed from the current representative haplotypes on the mtDNA phylotree by private mutations, the latter were double-checked by several PCRs and by sequencing in combination with other SNPs. 5. Characteristic post-mortem DNA damage-derived substitutions (C>T, G>A) were observed, but these substitutions were neither reproducible nor consistent across sequence assemblies and could easily be distinguished from true SNPs. 6.

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Altogether 51 of the 2376 (2.15%) blank controls were contaminated during the course of the three-year investigation. No contamination could be detected in any of the multiplex PCRs. Of the 51 contaminated blanks, 0.96% (17 out of 1764) were PCR blanks, 6.4% (20 out of 311) hydroxyapatite milling blanks, and 4.7% (14 out of 301) extraction blanks. All contaminated blanks were sequenced and compared to haplotypes of samples analyzed along with these PCR extraction blanks. Fourteen individuals that were amplified from two extraction events had to be excluded from further analyses due to signs of potential contamination. 7. The HVS-I region was sequenced from all relevant co-workers (archaeologists, anthropologists and geneticists) (Supplementary Table S14) and compared with samples and blanks. None of the samples matched with the haplotype of the main processor of the samples (C.R.), although co-workers with common haplotypes (basal J, H-rCRS) cannot be ruled out as possible source of contamination. However, we find no signs of systematic contamination by these workers and consider it unlikely that samples that share these haplotypes were affected in various independent experiments whilst parallel samples were not.

Population Genetic Analyses

Reference populations and clustering of the Iberian data

We split pre-Neolithic samples of Western and Central Europe into three spatio-temporal groups: a Pleistocene group (HG_PLEI, before 10,000 BCE)[78,97], a Holocene group (HG_HOL, after 10,000 BCE) from Central and North Europe[54,58,59,78–80], and into an Iberian hunter-gatherer group (HG_I)[42,43,45,47]. As comparative ancient populations we used Neolithic (5800-4900 BCE) datasets from the Carpathian Basin (CB_EN)[59,80], Early (GER_EN, 5500-4000 BCE), Middle (GER_MN, 4000-2650 BCE) and Late Neolithic (GER_LN, 2800-2050 BCE) series from Central Germany[54,56,57,86,98,99] and from Central European Early Bronze Age period (CEU_EBA, 2200-1550 BCE)[56,96]. From France, we included two Neolithic groups from the 4th and 3rd millennium BCE (FRA_GUR and TRE)[66,67]. From the Eastern European steppe, we included an mtDNA dataset associated with the Yamnaya culture (YAM, 3000-2500 BCE)[50,96,100].

From the Iberian Peninsula we used all published mtDNA data[39–42,44,46–48,50,51,61], and combined them with our results. The dataset from Cova de Montanissel was only used in haplogroup based tests, due to insufficiently reproduced HVS-I results[101]. We divided the ancient Iberian dataset into the following geographical groups: southwest Iberia (SWI), southeast Iberia (SEI), central Iberia (CI), northeast Iberia (NEI), and north Iberia with middle Ebro Valley (MEV). For sufficiently large sample numbers, we further sub-divided these into the following chronological groups: pre-Neolithic

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

hunter-gatherers (abbreviated as HG_I from the Upper Paleolithic and Mesolithic), Early Neolithic (EN, ~ 5700-4500 BCE), Middle and Late Neolithic (MLN, ~ 4500-3000 BCE), Early Chalcolithic (~ 3000-2600 BCE), Late Chalcolithic (~ 2600-2200 BCE), and Early Bronze Age (EBA, ~ 2200-1500 BCE). Where the chronology could not be defined precisely, we classified the samples into Neolithic (Neo) and Chalcolithic (CHA) periods. Note the differences in chronological terminology between Central Europe and the Iberian prehistory, mostly due to the absence of a distinct 'Chalcolithic' period in the archaeological terminology used in Central Europe (Fig. 1).

MtDNA haplogroup frequency based tests

We used principal component analysis (PCA) to visualize the relationships between relative haplogroup compositions of the Iberian and other prehistoric datasets. PCA has the benefit of reducing the complexity of the entire dataset while maintaining an accurate representation of its variability. PCA was performed with 20 ancient populations comparing haplogroup frequencies of 25 mtDNA haplogroups (see Supplementary Table S6), and calculated in R version 3.1.3[102], using a customized *prcomp* function.

Hierarchical clustering was used to construct clusters of the predefined prehistoric datasets, based on similarities (distances) between their haplogroup compositions. This analysis helps interpreting the PCA plots in form of dendrograms. Clustering of haplogroup frequencies was performed using Ward type algorithm and Euclidean distance measurement method. The result was visualized as a dendogram with the *pvclust* package in R.2.13.1[103]. Cluster significance was evaluated by 10,000 bootstrap replicates. Significance of each cluster was given as an AU (Approximately Unbiased) p-value, in percentage.

Fisher's exact test is a nonparametric test for statistical significance[104]. The null hypothesis for each Fisher's exact test was that the Iberian groups belonged to the same population. The resulting p-value thus described the probability of obtaining the observed haplogroup compositions if both groups were part of the same metapopulation. Fisher's exact test was performed with absolute frequencies of the haplogroups observed in the Iberian dataset. We used all haplogroup results per group in this analyses and tested a series of pairwise comparisons of Iberian chronological and geographic groups, using *fisher.test* function in R v.3.0.3[102].

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The test of population continuity explores whether genetic differences between two populations from two or more consecutive time periods can be adequately explained by genetic drift alone or whether a discontinuity between the two populations has to be assumed[56]. The tests were performed for pairs of chronological groups in specific regions (central Iberia, middle Ebro Valley, northeast Iberia, and southwest Iberia). It was also performed for the entire Iberian time transect, using 'metapopulations' from consecutive pairs of chronological periods (HG_I, EN, MLN, CHA (early and late), EBA). Effective population sizes Ne=1000, 10000 were tested, calculating with differences between terminal dates of the chronological periods. Further parameters of the tests are presented in Supplementary Table S8.

MtDNA HVS-I sequence based tests

We also characterized the Iberian populations with standard diversity indices that were calculated in DnaSP[105]**.** The HVS-I sequence range between np 16048-16410 was used for the calculations (Supplementary Table S9).

Shared haplotype analysis[106] examines the occurrence of identical lineages in different populations. For the ancestral shared haplotype analysis (ancestral SHA), we took the relative chronology of the studied groups into account and tracked first appearance of any given lineage and their transmission (or presence/absence) in subsequent, i.e. later populations. SHA and ancestral SHA were performed as described in[59]. Levelplot of the percentage of shared lineages was visualized in R version 3.1.3, using level plot function. First, we considered the same groups that were counted in the F_{ST} analysis. In the next step, we assembled all Iberian regions, and we differentiated chronological groups of the Iberian prehistory (Supplementary Table S10). These groups were put into chronological order and ancestral haplotypes were determined. A major drawback of this process is the limited resolution of H-rCRS haplotypes. Even though we could differentiate H1 and H3 subhaplogroups in our dataset, the published comparative hunter-gatherer H-rCRS types have not been typed at this resolution[42,43,47]. Therefore, it is likely that the contribution of hunter-gatherer lineages to later time periods is overestimated. On the other hand, available haplogroup definitions of H, V, HV and U haplotypes were used to separate these HVS-I sequences.

F_{ST} values ('fixation indexes') measure the amount of genetic structuring/differentiation in populations. With F_{ST} values we aimed to test whether pairs of populations were genetically distinct, and panmixia could be ruled out as basal assumption. F_{ST} values between pairs of prehistoric populations were calculated using reproduced HVS-I sequences (np 16056-16390) in Arlequin v. 3.5,

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

applying a Tamura and Nei substitution model with a gamma parameter of 0.117[106]. F_{ST} p values were calculated based on 10000 permutations, and post hoc adjusted to correct for multiple comparison by the Benjamini and Hochberg method, using the function *p. adjust* in R. 3.0.3[107].

Multidimensional scaling (MDS) was used to translate the distances/dissimilarities between multidimensional objects, represented by F_{ST} values, into spatial (Euclidean) distances and to visualize them in two-dimensional space. MDS was performed in R v. 3.1.3 using *metaMDS* function in vegan library[108]. Distance matrix was calculated based on Slatkin F_{ST} values, which F_{ST} values were calculated using HVS-I sequences (np 16056-16390) in Arlequin v. 3.5, applying a Tamura and Nei substitution model with a gamma parameter of 0.117[106]. We included 18 ancient groups in this analysis (Supplementary Table S11), but excluded the Holocene group, which had the greatest genetic distances from the other Central European and Iberian populations, and thus compressed the other parts of the MDS plot.

Analysis of molecular variance (AMOVA), based on HVS-I sequences (np 16056-16390), was performed between a subset of the Iberian and the Central European prehistoric groups used in the MDS (HG groups and YAM were left out). The "among clusters" and "within clusters" variance, F_{CT}, F_{SC} and their p values were computed in Arlequin v. 3.5[106]. Geographic and chronological groups were arranged into different models, consisted of 2-5 clusters that were assumed as plausible from the MDS plot, and AMOVA was conducted for each arrangement (Supplementary Table S12).

We calculated F_{ST} values between modern and three prehistoric Iberian populations and mapped the genetic distances by interpolating between the geographic coordinates of modern populations. The comparative modern mtDNA dataset with detailed information was described in Csősz et al.[109]. We calculated genetic distances (F_{ST} values) randomly choosing a maximum of 140 sequences per population (n = 17494 sequences altogether), in order to balance the differences in sample sizes, and calculated F_{ST} values between prehistoric Iberian (EN, MLN, CHA) and 133 present-day populations. The analysis was performed in Arlequin v. 3.5, using a uniform sequence length (np 16068–16365), and a Tamura & Nei substitution model with a gamma value of 0.177 (Supplementary Table S13)[106]. F_{ST} values were combined with longitudes and latitudes according to context information from the literature. The F_{ST} values and coordinates were interpolated with the Kriging method implemented in Arcmap ArcGIS version 10.3.

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Lemmen, C., Gronenborn, D. & Wirtz, K. W. A simulation of the Neolithic transition in Western Eurasia. *J. Archaeol. Sci.* **38,** 3459–3470 (2011).

2. Bellwood, P. *First farmers: the origins of agricultural societies*. (Blackwell, 2005).

3. Özdogan, M. The beginning of the Neolithic economies in southeastern Europe: an Anatolian perspective. *J. Eur. Archaeol.* **5,** 1–33 (1997).

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

5. Bocquet-Appel, J.-P., Naji, S., Linden, M. Vander & Kozlowski, J. K. Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of 14C dates. *J. Archaeol. Sci.* **36,** 807–820 (2009).

6. Jiménez Guijarro, J. *Cazadores y campesinos: la neolitización del interior de la Península Ibérica.* (Real Academia de la Historia., 2010).

7. Bernabeu-Aubán, J., Molina-Balaguer, L., Esquembre-Bebiá, M. A., Ortega, J. R. & Boronat-Soler, J. de D. in *De Méditerranée et d'ailleurs... Mélanges offers à Jean Guilaine* 83–95 (Archives d'Écologie Préhistorique, 2009).

8. García Borja, P., Aura Tortosa, J. E., Bernabeu Aubán, J. & Jordá Pardo, J. F. Nuevas perspectivas sobre la neolitización en la Cueva de Nerja (Málaga-España): La cerámica de la sala del vestíbulo. *Zephyrus* **66,** 109–132 (2010).

9. Guilaine, J., Manen, C. & Vigne, J.-D. *Pont de Roque-Haute. Noveaux regards sur la néolithisation de la France méditerranéenne*. (Archives d´Ecologie Préhistorique, 2007).

10. Fernández-Eraso, J. El Neolítico inicial en el País Vasco meridional. Datos recientes. *Kobie. Ser. Anejos* **6,** 181–190 (2004).

11. Utrilla-Miranda, P. in *El Paisaje en el Neolítico mediterráneo* 179–208 (Saguntum Extra-5, 2002).

12. Manen, C., Marchand, G. & Carvallo, A. F. Le Néolithique ancien de la péninsule Ibérique : vers une nouvelle évaluation du mirage africain? in *Congrés du centenaire: Un siècle de construction du discours scientifique en Préhistoire* **3,** 133–151 (Société préhistorique française, 2007).

13. Morales, J. *et al.* The origins of agriculture in North-West Africa: macro-botanical remains from Epipalaeolithic and Early Neolithic levels of Ifri Oudadane (Morocco). *J. Archaeol. Sci.* **40,** 2659–2669

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(2013).

14. García-Martínez de Lagrán, Í. Recent Data and Approaches on the Neolithization of the Iberian Peninsula. *Eur. J. Archaeol.* **18,** 429–453 (2015).

15. Rojo-Guerra, M. A., Garrido, R. & García Martínez de Lagrán, Í. *El Neolítico en la Península Ibérica y su contexto europeo*. (Ed. Cátedra, 2012).

16. Rojo-Guerra, M. A., Kunst, M., Garrido Pena, R., García-Martínez de Lagrán, I. & Morán Dauchez, G. *Paisajes de la memoria: asentamientos del Neolítico antiguo en el Valle de Ambrona (Soria, España)*. (Servicio de Publicaciones de la Universidad de Valladolid, 2008).

17. Zapata, L., Peña-Chocarro, L., Pérez-Jordà, G. & Stika, H.-P. Early neolithic agriculture in the Iberian Peninsula. *J. World Prehistory* **18,** 283–325 (2004).

18. Fernández Eraso, J. La actividad pecuaria en la Rioja Alavesa durante la Prehistoria Reciente. *Cuad. Arqueol. la Univ. Navarra* **18,** 159–171 (2010).

19. Oms-Arias, F. X. Caracterització tècnica, tipològica i cronològica de les ceràmiques del Neolític antic de la Cova Colomera (Prepirineu de Lleida). *Arch. Prehisotira Levantina* **XXVII,** 51–80 (2008).

20. Rojo-Guerra, M. Á. *et al.* 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). *Boletín del Semin. Arte y Arqueol.* **LXXIX,** 9–55 (2013).

21. Bernabeu-Aubán, J. & Martí-Oliver, B. in *La transition Néolitique en Méditerranée* (eds. Manen, C., Perrin, T. & Guilaine, J.) 419–438 (Archives d'Écologie Préhistorique, 2014).

22. Rojo-Guerra, M. *et al.* in *Del neolític a l'edat del bronze en el mediterrani occidental: Estudis en homenatge a Bernat Martí Oliver* 181–210 (Valencia (Provincia). Diputación Provincial, 2016).

23. Bernabeu-Auban, J. in *Restos de vida, restos de muerte: la muerte en la Prehistoria* (eds. Soler Mayor, B. & Pérez Fernández, A.) 45–54 (Museu de Prehistoria, 2010).

24. Garrido-Pena, R. & Rojo-Guerra, M. in *Funerary Practices in the Iberian Peninsula from the Mesolithic to the Chalcolithic. BAR International Series 2417* (eds. Gibaja, J. F., Carvalho, A. F. & Chambon, P.) 21–28 (2012).

25. García-Gazolaz, J. & Sesma Sesma, J. in *La Tierra te sea leve, Arqueología de la muerte en Navarra. Museo de Navarra, Pamplona, exposición, 27 de noviembre de 2007-30 de abril de 2008.* 49–58 (Institución Príncipe de Vian, 2007).

26. Garrido-Pena, R., Rojo Guerra, M., Tejedor-Rodriguez, C. & García-Martínez de Lagrán, I. in *El neolítico*

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

en la Península Ibérica y su contexto europeo (eds. Rojo Guerra, M. Á., Garrido Pena, R. & García-Martínez de Lagrán, I.) 143–174 (Cátedra, 2012).

27. Rojo-Guerra, M. A. in *Iberia: protohistory of the Far West of Europe: from Neolithic to roman conquest* 43–69 (Universidad de Burgos, Servicio de Publicaciones e Imagen Constitucional, 2014).

28. Martin, A. in *Sépultures et sociétés. Du Néolithique à l'Histoire* (ed. Guilaine, J.) 45–67 (2009).

29. Márquez Romero, J. E. & Jiménez Jáimez, V. *Recintos de fosos: Geneaología y significado deuna tradición en la Prehistoria del suroeste de la Península Ibérica (IV-III milenios AC)*. (Universidad de Malaga. Servicio de Publicaciones, 2010).

30. Valera, V. in *2200 BC – A climatic breakdown as a cause for the collapse of the old world? Tagungen des Landesmuseums für Vorgeschichte Halle 12* (eds. Meller, H., Arz, H. W., Jung, R. & Risch, R.) 409–427 (Landesamt für Denkmalpflege und Archäologie Sachsen-Anhalt, Landesmuseum für Vorgeschichte, 2015).

31. Lull, V., Micó, R., Rihuete Herrada, C. & Risch, R. in *2200 BC – A climatic breakdown as a cause for the collapse of the old world? Tagungen des Landesmuseums für Vorgeschichte Halle 12* (eds. Meller, H., Arz, H. W., Jung, R. & Risch, R.) 365–407 (Landesamt für Denkmalpflege und Archäologie Sachsen-Anhalt, Landesmuseum für Vorgeschichte, 2015).

32. Garcia Sanjuán, L. & Murillo-Barroso, M. Social complexity in copper age southern Iberia (ca. 3200-2200 cal b.c.): Reviewing the 'state' hypothesis at Valencina de la Concepción (Seville, Spain). *Prehistory Iberia. Debating Early Soc. Strat. State.* 119–140 (2013). doi:10.4324/9780203072035

33. Schuhmacher, T. X. Elfenbeinobjekte des Chalkolithikums und der Frühen Bronzezeit auf der Iberischen Halbinsel. Studien zu Herkunft, Austausch, Verarbeitung und sozialer Bedeutung von Elfenbein. *Iberia Archaeologica* **16,** 610 (2012).

34. Besse, M. & Strahm, C. The components of the Bell Beaker complex. in *Bell Beakers today: pottery, people, culture, symbols in prehistoric Europe. Proceedings of the International colloquium Riva del Garda (Trento, Italy) 11-16 May 1998 / Vol. 1.* (ed. Nicolis, F.) (2001).

35. Rojo Guerra, M., Garrido-Pena, R. & García-Martínez de Lagrán, I. *El Campaniforme en la Península Ibérica y su contexto europeo = Bell beakers in the Iberian Peninsula and their European context*. (Universidad de Valladolid, Secretariado de Publicaciones e Intercambio Editorial, 2005).

36. Garrido-Pena, R. in *Iberia. Protohistory of the far west of Europe: from Neolithic to Roman conquest* (ed. Almagro, M.) 113–124 (Universidad de Burgos. Fundación Atapuerca, 2014).

37. Garrido-Pena, R. in *Social Inequality in Iberian Late Prehistory. BAR International Series 1525.* (eds.

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Díaz-del-Río, P. & García Sanjuán, L.) 81–96 (2006).

38. Lull, V., Micó, R., Rihuete, C. & Risch, R. The La Bastida fortification: new light and new questions on Early Bronze Age societies in the western Mediterranean. *Antiquity* **88,** 395–410 (2014).

39. Gamba, C. *et al.* Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. *Mol. Ecol.* **21,** 45–56 (2012).

40. Gómez-Sánchez, D. *et al.* Mitochondrial DNA from El Mirador cave (Atapuerca, Spain) reveals the heterogeneity of Chalcolithic populations. *PLoS One* **9,** e105105 (2014).

41. Günther, T. *et al.* Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc. Natl. Acad. Sci.* **112,** 11917–11922 (2015).

42. Hervella, M. *et al.* Ancient DNA from hunter-gatherer and farmer groups from Northern Spain supports a random dispersion model for the Neolithic expansion into Europe. *PLoS One* **7,** e34417 (2012).

43. de-la-Rua, C., Izagirre, N., Alonso, S. & Hervella, M. Ancient DNA in the Cantabrian fringe populations: A mtDNA study from Prehistory to Late Antiquity. *Quat. Int.* **364,** 306–311 (2015).

44. Olalde, I. *et al.* A common genetic origin for early farmers from mediterranean cardial and central european LBK cultures. *Mol. Biol. Evol.* **32,** 3132–3142 (2015).

45. Sánchez-Quinto, F. *et al.* Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Curr. Biol.* **22,** 1494–9 (2012).

46. Alt, K. W. *et al.* A Community in Life and Death: The Late Neolithic Megalithic Tomb at Alto de Reinoso (Burgos, Spain). *PLoS One* **11,** e0146176 (2016).

47. Chandler, H., Sykes, B. & Zilhão, J. in *Actas dell III Congreso del Neolítico en la Península Ibérica, Santander, Monografías del Instituto internacional de Investigaciones Prehistóricas de Cantabria 1* (eds. Arias, P., Ontañón, R. & García-Moncó, C.) **1,** 781–786 (2005).

48. Lacan, M. *et al.* Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proc Natl Acad Sci U S A.* **108,** 18255–9 (2011).

49. Brandt, G., Szécsényi-Nagy, A., Roth, C., Alt, K. W. & Haak, W. Human paleogenetics of Europe – The known knowns and the known unknowns. *J. Hum. Evol.* **79,** 73–92 (2014).

50. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522,** 207–211 (2015).

51. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528,** 499–503

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(2015).

52. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507,** 225–8 (2014).

53. Skoglund, P. *et al.* Genomic diversity and admixture differs for Stone-age Scandinavian foragers and farmers. *Science* **344,** 747–750 (2014).

54. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,** 409–413 (2014).

55. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536,** 419–424 (2016).

56. Brandt, G. *et al.* Ancient DNA reveals key stages in the formation of Central European mitochondrial genetic diversity. *Science* **342,** 257–261 (2013).

57. Haak, W. *et al.* Ancient DNA from European early Neolithic farmers reveals their Near Eastern affinities. *PLoS Biol.* **8,** e1000536 (2010).

58. Bramanti, B. *et al.* Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326,** 137–40 (2009).

59. Szécsényi-Nagy, A. *et al.* Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. *Proc. R. Soc. B* **282,** 20150339 (2015).

60. Fernández, E. *et al.* Ancient DNA Analysis of 8000 B.C. Near Eastern Farmers Supports an Early Neolithic Pioneer Maritime Colonization of Mainland Europe through Cyprus and the Aegean Islands. *PLoS Genet.* **10,** e1004401 (2014).

61. Sampietro, M. L. *et al.* Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc. R. Soc. B* **274,** 2161–7 (2007).

62. Rosa, A. & Brehm, A. African human mtDNA phylogeography at-a-glance. *J. Anthropol. Sci.* **89,** 25–58 (2011).

63. Brotherton, P. *et al.* Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat. Commun.* **4,** 1764 (2013).

64. Martínez-Cruz, B. *et al.* Evidence of pre-Roman tribal genetic structure in Basques from uniparentally inherited markers. *Mol. Biol. Evol.* **29,** 2211–22 (2012).

65. Keerl, V. A River Runs Through It- Ancient DNA Data on the Neolithic Populations of the Great Hungarian Plain. Doctoral thesis. (Johannes Gutenberg University Mainz, 2015). at

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

<http://ubm.opus.hbz-nrw.de/volltexte/2015/4115/>

66. Rivollat, M. *et al.* When the Waves of European Neolithization Met: First Paleogenetic Evidence from Early Farmers in the Southern Paris Basin. *PLoS One* **10,** e0125521 (2015).

67. Lacan, M. *et al.* Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc Natl Acad Sci U S A.* **108,** 9788–91 (2011).

68. Skoglund, P. *et al.* Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336,** 466–469 (2012).

69. Dolukhanov, P. *The early Slavs*. (Longman, 1996).

70. Housley, R. A., Gamble, C. S., Street, M. & Pettitt, P. Radiocarbon evidence for the Lateglacial human recolonisation of northern Europe. *Proc. Prehist. Soc.* **63,** 25–54 (1997).

71. Gamble, C., Davies, W., Pettitt, P., Hazelwood, L. & Richards, M. The Archaeological and Genetic Foundations of the European Population during the Late Glacial: Implications for. *Cambridge Archaeol. J.* **15,** 193 (2005).

72. Torroni, a *et al.* A signal, from human mtDNA, of postglacial recolonization in Europe. *Am. J. Hum. Genet.* **69,** 844–52 (2001).

73. Achilli, A. *et al.* The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am. J. Hum. Genet.* **75,** 910–8 (2004).

74. Pereira, L. *et al.* High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res.* **15,** 19–24 (2005).

75. Soares, P. *et al.* The archaeogenetics of Europe. *Curr. Biol.* **20,** R174–183 (2010).

76. Achilli, A. *et al.* Saami and Berbers--an unexpected mitochondrial DNA link. *Am. J. Hum. Genet.* **76,** 883–6 (2005).

77. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534,** 200–205 (2016).

78. Posth, C. *et al.* Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Curr. Biol.* **26,** 827–33 (2016).

79. Bollongino, R. *et al.* 2000 years of parallel societies in Stone Age Central Europe. *Science* **342,** 479–81 (2013).

80. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5,** 5257 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

81. Martins, H. *et al.* Radiocarbon dating the beginning of the Neolithcis in Iberia: new results, new problems. *J. Mediterr. Archaeol.* **28,** 105–131 (2015).

82. Alday Ruiz, A. The Neolithic in the Iberian Peninsula: An explanation from the perspective of the participation of Mesolithic communities. *Zephyrus* **69,** 75–94 (2012).

83. Carvalho, A. F. in *On Pre- and Earlier History of Iberia and Central Europe* (eds. Armbruester, T. & Hegewisch, M.) 45–61 (Studien zur Archäologie Europas Band 11, 2010).

84. Bueno, P., Barroso, R. & Balbín, R. in *The Megalithic Architectures of Europe* (eds. Laporte, L. & Scarre, C.) 157–166 (Oxbow Books, 2015).

85. Stika, H. P. Early Neolithic agriculture in Ambrona, Provincia Soria, central Spain. *Veg. Hist. Archaeobot.* **14,** 189–197 (2005).

86. Haak, W. *et al.* Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* **310,** 1016–1018 (2005).

87. Gassin, B., Astruc, L., Léa, V., Philibert, S. & Gibaja Bao, J. F. Burins du Chasséen Méridional. *ArchéoLogiques* 319–341 (2006).

88. Gibaja-Bao, J. F. & Clop García, X. in *El Neolítico en la Península Ibérica y su contexto europeo* (eds. Rojo, M. A., Garrido, R. & García, I.) 333–370 (2012).

89. Léa, V. Raw, pre-heated or ready to use: discovering specialist supply systems for flint industries in mid-Neolithic (Chassey culture) communities in southern France. *Antiquity* **79,** 51–65 (2005).

90. Liesau, C., Banerjee, A. & Schwarz, J. O. in *Yacimientos calcolíticos con campaniforme de la región de Madrid: nuevos estudios. Madrid.* (eds. Blasco, C., Liesau, C. & y Ríos, P.) 381–386 (Universidad Autónoma de Madrid, 2011).

91. Liesau von Lettow-Vorbeck, C. in *Analysis of the economic foundations of the social supremacy of the Beaker groups. Proceedings of the XVII UISPP World Congress (1-7 September, Burgos, Spain). Volume 6. Session B36.* (eds. Guerra, E. & Liesau, C.) 69–93 (Archaeopress Archaeology, 2016).

92. Schuhmacher, T. X., Cardoso, J. L. & Banerjee, A. Sourcing African ivory in Chalcolithic Portugal. *Antiquity* **83,** 983–997 (2009).

93. Delibes-De-Castro, G. Reinterpretación del ajuar campaniforme de Villar del Campo: Nuevos elementos de juicio para la valoración de la incidencia centro europea en el mundo de Ciempozuelos. *Celtiberia* **XXVIII,** 267–286 (1978).

94. Garrido-Pena, R. *El Campaniformeen la Meseta Central de la PenínsulaIbérica (c. 2500-2000 A.C.). BAR*

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

International Series **892,** (2000).

95. Fitzpatrick, A. P. in *Celtic from the West 2: Rethinking the Bronze Age and the Arrival of Indo-European in Atlantic Europe,* (eds. Koch, J. T. & Cunliffe, B.) 41–70 (Oxbow, 2013).

96. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522,** 167–172 (2015).

97. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23,** 553–559 (2013).

98. Lee, E. J. *et al.* Emerging genetic patterns of the European Neolithic: perspectives from a late Neolithic Bell Beaker burial site in Germany. *Am. J. Phys. Anthropol.* **148,** 571–579 (2012).

99. Haak, W. *et al.* Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. *Proc Natl Acad Sci U S A.* **105,** 18226–31 (2008).

100. Wilde, S. *et al.* Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc Natl Acad Sci U S A.* **111,** 4832–4837 (2014).

101. Simon, M. *et al.* The Presence of Nuclear Families in Prehistoric Collective Burials Revisited: The Bronze Age Burial of Montanissell Cave (Spain) in the Light of aDNA. *Am. J. Phys. Anthropol.* **146,** 406–413 (2011).

102. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.r–project.org/ (2012).

103. Suzuki, R. & Shimodaira, H. pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. R package version 1.3-2. http://CRAN.R–project.org/package=pvclust (2014).

104. Fisher, R. A. On the interpretation of χ2 from contingency tables, and the calculation of P. *J. R. Stat. Soc.* **85,** 87–94 (1922).

105. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25,** 1451–2 (2009).

106. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10,** 564–567 (2010).

107. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57,** 289–300 (1995).

108. Oksanen, J. *et al.* vegan: Community Ecology Package. R package version 2.2-1. http://CRAN.R–project.org/package=vegan (2015).

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

109. Csősz, A. *et al.* Maternal Genetic Ancestry and Legacy of 10th Century AD Hungarians. *Sci. Rep.* **6,** 33446 (2016).

Financial statement

This study was funded by the German Research Foundation (Grant no. Al 287/14-1).

Author contribution

K.W.A., R.R. and W.H. conceived the study.

C.R. performed the ancient DNA work. A.S-N. analyzed the data.

C. T-R., Í. G-M-d-L., H. A. M., P. B. R., R. B. B., R. d. B. B., A. M. H-C., R. F. F., C. A. F., J. J. E., L. R., C. O., M. I. F., I. S., O. V., R. M., V. L., J. S. D., J. A. L. P., C. R-d-T. M., M. S. H. P., F. J. J. M., J. L. M., A. A. F., K. T. L., A. M. S., M. M. R., L. M. O., V. M. D. S. G., C. C., A. W., J. R. B., A. M., J. P. M., M. H. O., J. C. M-G., J. C. P. E., R. C-A. B., T. T., E. C. B., J. L. C., A. C. A., C. L. v. L-V., C. B. B., P. R. M., A. P., J. I. R-G., M. A. E. B., R.P., E. M. H., E. M. I., J. V. d. M., R. M. G., V. M. C., O. L. J., and M.K. provided samples, anthropological and archaeological information.

S. Z., C. K. and P. H. performed further bioarchaeological analyses.

G. B., E. B., S. F., H. M., J. K., C. R. H. gave critical input to the version to be submitted.

A.S-N., W. H., R.R., R.G-P., M.A.R-G., A. M. H-C., S. L. P. and K.W.A wrote the paper. All authors have read and commented the manuscript.

Competing interest

The authors declare no competing interest.

Data and materials availability

Consensus HVS-I DNA sequences were uploaded to NCBI GenBank under the accession numbers KY446554 - KY446703.

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures

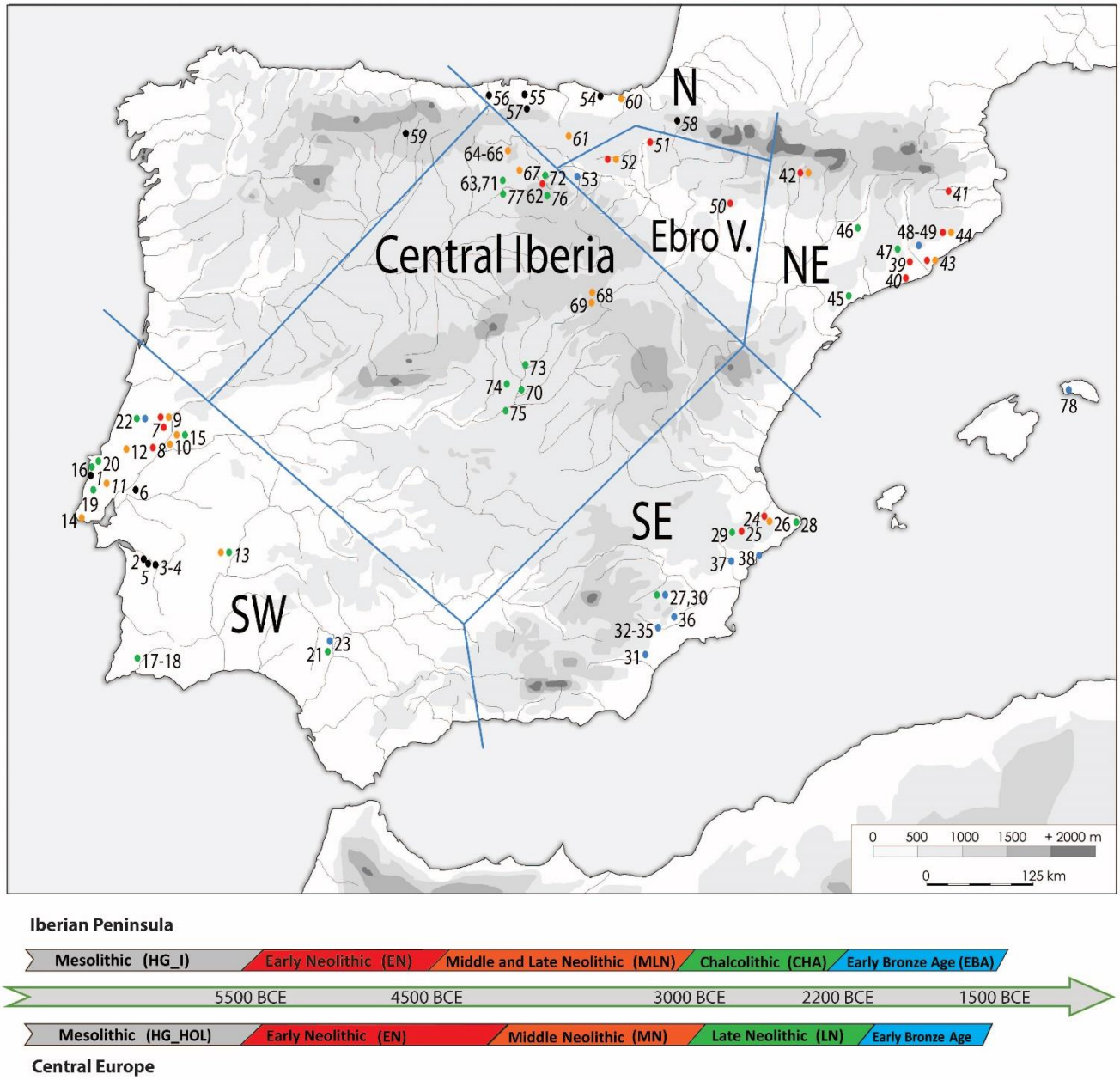

Figure 1. Map of the studied sites, including the published reference data and timing of archaeological periods on the Iberian Peninsula and in Central Europe.

Geographic regions, also differentiated in the mtDNA analyses, are indicated as: N: north, NE: northeast, SE: southeast, SW: southwest, Ebro V: middle Ebro Valley. Numbers on the map are colored according to the chronological periods, represented in the lower part of the figure. For the Central European chronology we used records from the most important comparative region of Central German Mittelelbe-Saale[56]*. Italicized numbers on the map show published archaeogenetic data (Supplementary Table S5). Site codes: 1. Toledo, 2. Arapouco, 3. Cabeço das Amoreiras, 4. Cabeço de Pez, 5. Poças de São Bento, 6. Moita do Sebastião, 7. Gruta do Caldeirão, 8. Galeria da*

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cisterna (Almonda cave), 9. Gruta de Nossa Senhora das Lapas, 10. Gruta do Cadaval, 11. Algar do Bom Santo, 12. Gruta das Alcobertas, 13. Perdigões, 14. Gruta do Poço Velho, 15. Gruta dos Ossos, 16. Tholos de Pai Mogo I, 17. Hipogeu de Monte Canelas I, 18. Hipogeu de Monte Canelas III, 19. Bolores, 20. Gruta de Malgasta, 21. Señorío de Guzman, 22. Gruta do Carvalhal de Turquel, 23. Cobre las Cruces, 24. Cova de l`Or, 25. Cova de la Sarsa, 26. Cova d'en Pardo, 27. Molinos del Papel, 28. Cova del Barranc del Migdia, 29. Cova del Cantal, 30. Camino de Molino, 31. Fuente Alamo, 32. Lorca-Los Tintes, 33. Lorca-Madre Mercedarias, 34. Lorca-Castillo de Lorca, 35. Rincón de Moncada, 36. La Bastida, 37. Tabayá, 38. Illeta dels Banyets, 39. Cova Bonica, 40. Can Sadurní, 41. Cova d'Avellaner, 42. Els Trocs, 43. Sant Pau de Camp, 44. Camí de Can Grau, 45. Barranc d'en Rifà, 46. Balma de Sargantana, 47. Cova de la Ventosa, 48. Miguel Vives, 49. Can Gambús, 50. Chaves, 51. Paternanbidea, 52. Los Cascajos, 53. Valdescusa, 54. Erralla, 55. La Chora, 56. La Pasiega, 57. El Miron, 58. Aizpea, 59. La Brana, 60. Marizulo, 61. Fuente Hoz, 62. Alto de Rodilla, 63. Fuente Celada, 64. Fuente Pecina 1, 65. Fuente Pecina 2, 66. Fuente Pecina 4, 67. Alto de Reinoso, 68. La Mina, 69. La Tarayuela, 70. El Juncal, 71. Arroyal I, 72. El Hundido, 73. Camino de las Yeseras, 74. Humanejos, 75. Valle de las Higueras, 76. El Portalon, 77. El Mirador, 78. Es Forat de ses Aritges.

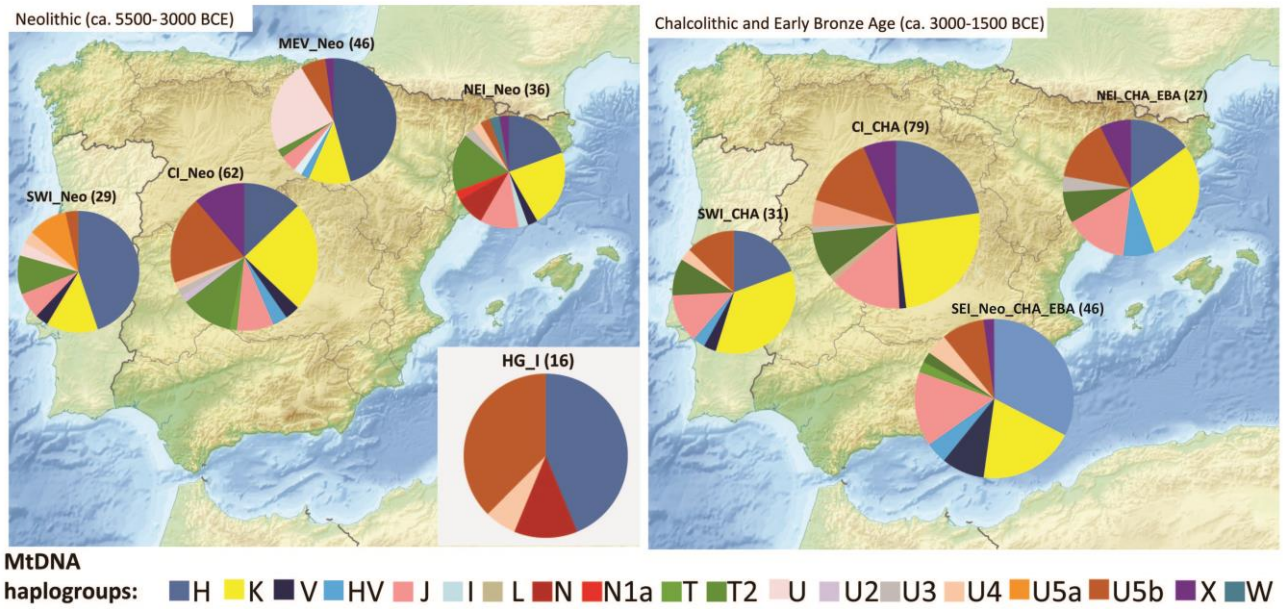

***Figure 2. MtDNA haplogroup composition of prehistoric groups**. Abbreviations: Iberian hunter-gatherers (HG_I), northeast Iberian Neolithic (NEI_Neo), northeast Iberian Chalcolithic and Early Bronze Age (NEI_CHA_EBA), Neolithic in the middle Ebro Valley and north Iberia (MEV_Neo), central Iberian Neolithic (CI_Neo), central Iberian Chalcolithic (CI_CHA), southwest Iberian Neolithic (SWI_Neo), southwest Iberian Chalcolithic (SWI_CHA), southeast Iberian Neolithic, Chalcolithic and Bronze Age (SEI_Neo_CHA_EBA). Relative haplogroup frequencies are presented*

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in Supplementary Table S6. The background map of Iberian Peninsula was downloaded from Wikipedia (https://en.wikipedia.org/wiki/File:Relief_Map_of_Spain.png#file) and modified in Adobe Illustrator CS6 software.

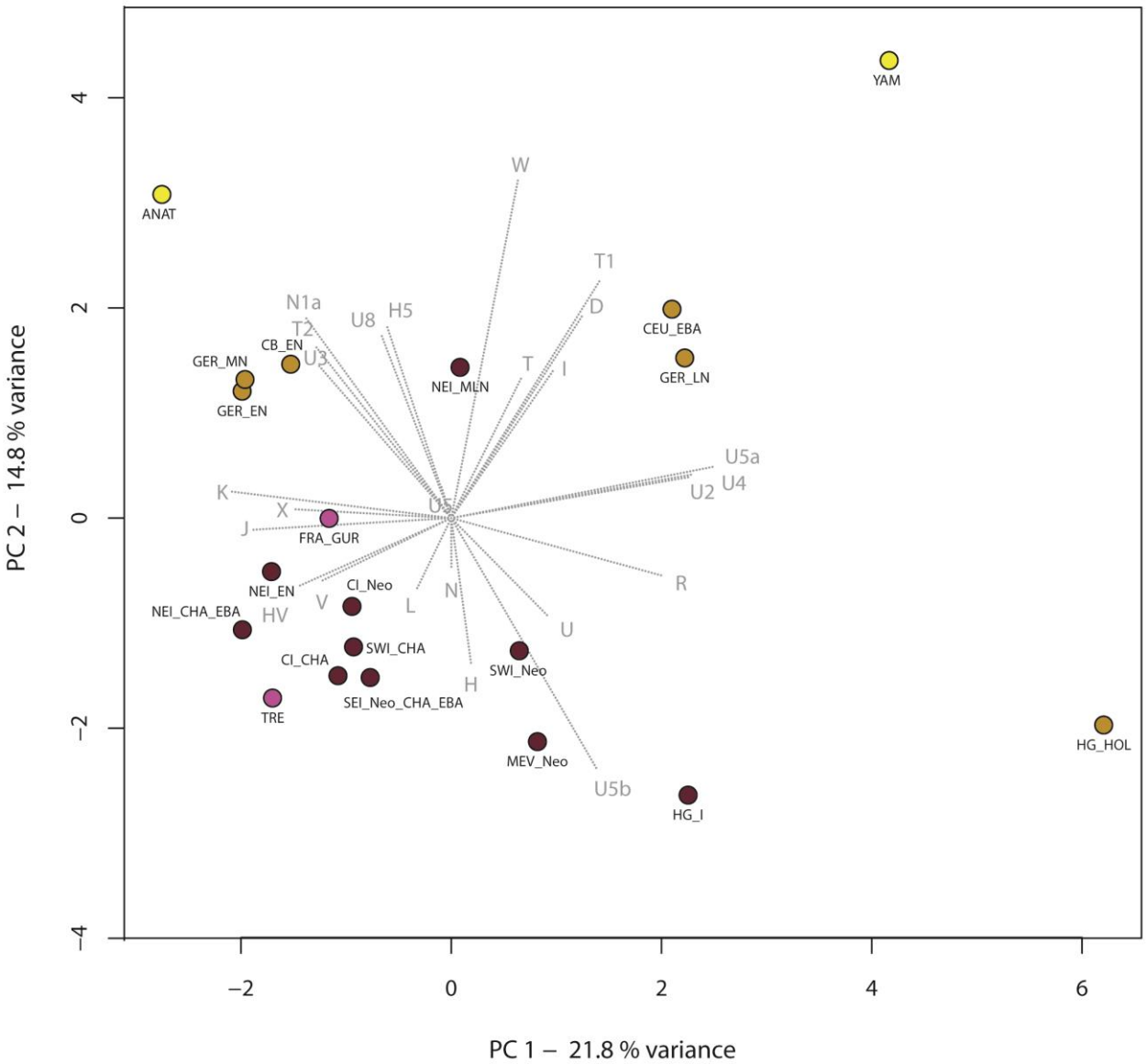

Figure 3. Principal component analysis based on haplogroup frequencies of 1046 individuals from 20 prehistoric groups. *The first two components display 36.6% of the total variance. Groups are colored according to their geographical positions: brown: Iberia, purple: France, ochre: Central and East-Central Europe, yellow: Eastern Europe and Near East. For abbreviations of Iberian groups, see legend of* ***Figure 2****. Further abbreviations: Central and North European hunter-gatherers from the Holocene (HG_HOL), Neolithic Anatolia (ANAT) Yamnaya (YAM), Early Neolithic Carpathian Basin (CB_EN), Early Neolithic Germany (GER_EN), Middle Neolithic Germany (GER_MN), Late Neolithic Germany (GER_LN), Early Bronze Age Central Europe*

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(CEU_EBA), Neolithic Gurgy site in France (FRA_GUR), Neolithic Treilles culture in France (TRE). For further information see Supplementary Table S6.

bioRxiv preprint doi: https://doi.org/10.1101/106963; this version posted February 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tables

A: SHA			**Detected in the following populations:**				
		n sample	**HG_I**	**EN**	**MLN**	**CHA**	**EBA**
mtDNA lineages:	**HG_I**	15	100	15.79	7.62	5.09	25.00
	EN	57	13.33	100	50.48	55.08	43.75
	MLN	105	60.00	57.89	100	67.80	68.75
	CHA	118	33.33	42.11	60.95	100	81.25
	EBA	16	13.33	12.28	20.95	34.75	100
B: ancestral SHA			**Detected in the following populations:**				
		n sample	**HG_I**	**EN**	**MLN**	**CHA**	**EBA**
mtDNA lineages:	**HG_I**	15	100				
	EN	57	21.05	78.95			
	MLN	105	11.43	39.05	49.52		
	CHA	118	16.95	39.83	15.25	27.97	
	EBA	16	25.00	31.25	18.75	6.25	18.75

Table 1. Results of shared haplotype analysis (SHA): percentage of shared HVS-I haplotypes among the Iberian chronological groups (A), and ancestral haplotype analysis with the studied Iberian groups (B). *Abbreviations: HG_I- hunter-gatherers, EN-Early Neolithic, MLN-Middle and Late Neolithic, CHA-Chalcolithic, EBA-Early Bronze Age, all in Iberia. For further details, see Supplementary Table 10 and Methods - mtDNA HVS-I sequence based tests.*