Ancient mitochondrial lineages support the prehistoric maternal root of Basques in Northern Iberian Peninsula.

SUPPLEMENTARY INFORMATION

Leire Palencia-Madrid, Sergio Cardoso, Christine Keyser, Juan Carlos López-Quintana, Amagoia Guenaga-Lizasu, Marian M. de Pancorbo[1].

[1]Corresponding author e-mail:
marian.mdepancorbo@ehu.eus

Table of Contents

SUPPLEMENTARY INFORMATION

APPENDIX

1. Isotopic analysis of ancient samples from Santimamiñe

Dating ancient remains of prehistoric individuals by ^{14}C is essential to establish their antiquity and to contextualize their genetic data. In the same way, analysis of the stable isotopes of carbon $^{13}C/^{12}C$ ($\delta^{13}C$) and nitrogen $^{15}N/^{14}N$ ($\delta^{15}N$) in these samples contributes to the knowledge of their diet. Thus, the combined analysis of the stable isotopes with ^{14}C makes it possible to slightly infer the cultural period the individuals of Santimamiñe belong to.

As it is said in the paper, the analysis of the ^{14}C radioactive isotope dated the age of these remains in a wide period of time, ranging from 5,210 cal BCE to 390 calibrated years of Common Era (cal CE). Four of these individuals are comprised in a short time period, between 1,740 and 1,320 cal BCE.

The oldest sample found in the Cave of Santimamiñe (S.12N) was dated in the transition between the 5th and 6th millennia cal BCE, which is a critical time period in this area due to the controversy that exists around the arrival of the Neolithic period to the Franco-Cantabrian region (1, 2). The Neolithic culture appears to have been established around 5,500 – 5,200 BCE (years before Common Era) in some areas of the Iberian Peninsula (3), whereas in the Franco-Cantabrian region the first remains clearly associated to Neolithic culture appeared later on (4), showing that this region suffered a delay of several hundreds of years in comparison to surrounding regions (5, 6). The reason of this delay and the way in which Neolithic culture arrived at the Franco-Cantabrian region remain unclear. Thus, there could be some doubts about the cultural period the individual S.12N of the Cave of Santimamiñe belonged to. In this case, a more thorough analysis of the stable isotopes, combined with ^{14}C, makes it possible to infer the cultural period of the individual S.12N.

The study of stable isotopes of carbon $^{13}C/^{12}C$ ($\delta^{13}C$) and nitrogen $^{15}N/^{14}N$ ($\delta^{15}N$) provided information related to the alimentary habits of these individuals found in the cave, which varied over time (Fig. 2). Previous studies of human ancient remains from the Cantabrian coastline pointed to a high diet contrast between Mesolithic and Neolithic populations, as it was observed in other littoral regions of Europe (7-9). Through the analysis of the stable isotopes it has been known that food of marine origin had much more importance in the diet of littoral Mesolithic populations than in that of Neolithic populations, for whom it was almost irrelevant. The analysis and comparison of the stable isotope ratios between the remains from the Cave of Santimamiñe and those of other human remains from the Cantabrian region (Fig. 2b) pointed to a probable Mesolithic character of S.12N individual. Values higher than -18‰ indicate that this individual ingested more proteins from marine origin than the individuals analyzed from Chalcolithic and Bronze Age periods. Results of $\delta^{15}N$ fit with this diet, and values

slightly below 11‰ indicate that most of those proteins came preferentially from marine invertebrates instead of fishes. Ancient individuals from Santimamiñe were able to obtain marine foods and to incorporate them to their diet because this cave is very close to the estuary of the Oka river, which arose in the beginning of the Holocene (10), and not too distant from the coastline (Fig. 1). These results of isotopic analysis of S.12N are consistent with the stratigraphic level where this remain was found (11). In addition, archaeological studies show clear evidences of the systematic exploitation of the estuarial and coastal resources, which reached its highest level during the Mesolithic, when a big deposit of shells was formed in the Cave of Santimamiñe (12). Therefore, although the presence of Neolithic groups along Franco-Cantabrian region at the same time could not be discarded, the oldest individual from the Cave of Santimamiñe showed a Mesolithic diet, probably indicating the persistence of this subsistence way of life in the area at least until 5,210 – 4,950 cal BCE.

On the contrary, Neolithic individuals typically show lower values of $\delta^{13}C$, what could be due to the development of other subsistence strategies with a principal component of terrestrial origin and a very much reduced intake of seafood. In the Cave of Santimamiñe this lower values of $\delta^{13}C$ could be seen in samples dated on Chalcolithic and Bronze Age periods, when farming and ranching were better established livelihood strategies (Fig. 2a) (4, 13). In the case of the sample S.15I, dated in Late Roman period, the value of $\delta^{13}C$ is similar to that obtained on the Mesolithic individual. This $\delta^{13}C$ value could be due to the return to a marine food intake or to the consumption of some cereals that belong to C4 plants, such as millet (*Panicum miliaceum* and/or *Setaria italic*), which increases the value of $\delta^{13}C$ (14). This type of cereals could be consumed directly or through animals fed with them, and were cultivated in this region at least since Bronze Age (4).

2. Phylogenetic history

Analyzing mitochondrial lineages of the human remains of the Cave of Santimamiñe one by one shows that some of their haplogroups are common in other ancient populations of Europe. Lineages of H1, T2b, and U5b, observed in Santimamiñe between 2,200 and 1,610 cal BCE, are widely known along the European prehistory, with several human remains scattered through the continent (15-19). These subhaplogroups belong to those haplogroups that have the highest frequency in current European populations. Strikingly, with the exception of T2b, with its highest frequency peak located in the North of Italy (20), subhaplogroups H1 and U5b reach their highest frequencies in the North of the Iberian Peninsula, one of the regions where Paleolithic populations took refuge during the LGM (21-24).

On the contrary, two subhaplogroups found in Santimamiñe, U5a2a and U3a, are scarce in prehistory and virtually absent in current European populations. Their frequency is also very low or almost null in the Basque Country (23, 25) and, indeed, they were absent in the population from Busturialdea analyzed in this study. However, these samples are of great interest because they represent the most ancient remains of these haplogroups found to date in the western part of the European continent.

In the case of the subhaplogroup U5a2a, its first evidence is a Mesolithic individual of 6,743 ± 139 cal BCE found in Holhlenstein-Stadel (Germany) (23, 26). This individual is not too far in time from the Santimamiñe remain, belonging to the same subhaplogroup, dated in 5,210 – 4,950 cal BCE. These are the two first U5a2a remains found in Europe. Interestingly, they would be close to the moment of the origin of the subhaplogroup calculated by molecular clock, since Malyarchuk *et al*. (27) dated it between 5,700 and 9,300 years ago, while Behar *et al*. (28) set an older but wider date ratio of about 12,965.7 (± 5,917.8) years ago. Although the place where this lineage arose is not clear, Malyarchuk *et al.* (27) suggested East Europe as its place of origin, since its highest frequencies in current populations are found in East and Central Europe. However, despite very few ancient remains from southern and western Europe are available, this lineage appears to be more widely distributed over European Mesolithic populations than nowadays. Therefore, the place where subhaplogroup U5a2a originated cannot be clarified until more ancient remains of this lineage are analyzed. On the other hand, the presence of this subhaplogroup in Germany and in Franco-Cantabrian region in a relatively short span of time suggests a rapid spread and dispersion of this lineage, or it could also be due to an earliest date of origin in respect to that suggested by Malyarchuck *et al*. (27).

The subhaplogroup U3a found in the individual S.16G in Santimamiñe has been rarely seen in Europe. This subhaplogroup arose 14,818.8 (± 4,640.3) years ago (28), and it is thought that it was introduced in Europe by Neolithic migrations (29). The most ancient remains of subhaplogroup U3a brings us to an area currently encompassing Israel (4,585 – 4,245 cal BCE) (30), while in Europe, the oldest known remains of this subhaplogroup were found in Germany, in Salzmünde-Schiepzig site, with an age of 3,400 – 3,100/3,025 cal BCE (16). There were no more evidences in this continent until 200-270 CE, in Denmark (31). The finding of U3a in Santimamiñe, with an age of 1,450 – 1,320 cal BCE, provides the evidence of the arrival of this lineage to the Western Europe no later than the Bronze Age.

On the whole, the mitochondrial lineages of samples S.12N and S.16G are the most ancient evidence of the presence of subhaplogroups U5a2a and U3a in West Europe, respectively. Therefore, U5a2a was present in a Mesolithic population from Western Europe, around 7,000 years ago, at a little time after its appearance for the first time in Central Europe (26). However, nowadays U5a2a is only found at very low frequencies in central and northern Europe (32-34), and due to its absence in current Basque population it has been leaved out of the maternal lineages that hypothetically took part in the continent settlement which set off from the Franco-Cantabrian region after the LGM. On the other hand, we found that U3a, a subhaplogroup that probably arose in the Near East and was introduced into the European continent with Neolithic migrations (29), had reached the Western area of the European continent in the Bronze Age, about 3,300 years ago.

Thus, our results highlight the importance of analyzing ancient remains, especially those found at West and South Europe which still remain poorly known, to disclose lost ancient lineages that would go unnoticed through the analysis of exclusively current populations.

3. Current population of Busturialdea

The control region of mtDNA was analyzed in a sample of 158 autochthonous Basque individuals belonging to the population inhabiting the vicinity of the Cave of Santimamiñe, the region of Busturialdea. The global haplotype diversity of the autochthonous Basque individuals residing in Busturialdea was 0.9825 ± 0.0034. The most common control region haplotype in this population was 16129A, 16519C, 263G, 315.1C (6.96%), which was present in the four localities sampled. This haplotype has been found with a remarkable frequency among numerous populations of the North of the Iberian Peninsula, especially in the Basque area, or in populations of Basque ancestry (23, 25, 35, 36). Therefore, it could be considered a representative haplotype of this region. Actually, subhaplogroup H1j1, which only presents the polymorphism 16129A in the control region, was described by Behar et al. (25) as an autochthonous haplogroup of the Franco-Cantabrian region, based on its frequency in current populations from this region and its virtual absence in the rest of Europe (25, 37). In this population sample of Busturialdea eighteen individuals (11.39%) were classified into H1j1 subhaplogroup by the analysis of the coding region polymorphism T4733C.

The haplotype 16192T, 16270T, 16319A, 73G, 150T, 263G, 315.1C, 533G (5.06%), which belongs to subhaplogroup U5b1f1a, is also remarkable. This subhaplogroup represents 10% of the mitochondrial diversity of the current populations of the Franco-Cantabrian region, reaching a maximum frequency of ~24% in the populations next to the Pyrenees and decreasing to the West (37). Due to its frequency distribution pattern and its absence in populations out of the Franco-Cantabrian region, U5b1f1a was proposed as an autochthonous lineage of the Basque population, which could have originated during the Younger Dryas (12 800 – 11 500 years ago), with a splitting age of 11 985 years (37).

4. Median-joining network and time estimates

A median-joining network of haplogroup U5 was constructed (Fig. S3-a). Samples included were from both ancient Northern Iberian Peninsula and current population of Busturialdea. This phylogenetic branch was selected because it was the most frequent haplogroup among pre-Neolithic European populations (17, 38). Moreover, as it has a considerable frequency in current population of Busturialdea (12.66%), it is a suitable phylogenetic group for testing the continuity of the haplogroup U5 in this region.

Median-joining network distributed these samples following the expected phylogenetic arrangement, segregating lineages of U5a branch from those of U5b. In all the cases, current samples from Busturialdea arise from ancient samples or connecting median vectors. Two of the most ancient samples, ERR-1 (~12,000 years before present) and BRA-1 (~7,000 years before present), are important nodes from which the other samples spread. The remaining two Mesolithic samples (S.12N and AIZ-1) appeared in separated branches of the network, due to a more complex haplotypes, which seem to be extinct in current autochthonous population. Finally, sample S.7J from Bronze Age arise from Mesolithic BRA-1, in the same way that other current samples of haplogroup U5b do.

Time estimates obtained from this network, with a mutation rate of 1 substitution every 4.127 years, show that time expanded between ancient and current samples matches well with the expected age (Fig. S3). For instance, considering the sample ERR-1 as the ancestral node that gave rise to the remaining samples of U5 haplogroup in the network (Fig. S3-b), the age estimated is 17,540 ± 5,518.8 years, which is consistent with the radiocarbon date of that sample (12,310 years) (39). Time estimated for U5a branch (Fig. S3-c) of the network (15,132 ± 4,960 years) is also close to that calculated by Malyarchuk *et al*. (19.900 ± 6,000 years) for this haplogroup (27). On the other hand, age estimated for BRA-1 as ancestor (Fig. S3-d) of the samples that arose from its node in U5b branch of the network (9,433 ± 2,834.3 years), also matches with the radiocarbon date of this sample (~7,000 years before present)(40).

The short time gap presented between the real and estimated ages could be due to the small number of samples into the ancestral nodes, as well as, to an overestimation of the age caused by the mutation rate used. The result of this analysis agrees with the possibility of ancient samples of Northern Iberian Peninsula being the ancestors of current autochthonous population of Busturialdea, at least, for those lineages of haplogroup U5. Thus, the network and the ages estimated from it, reassert the hypothesis of genetic continuity of haplogroup U5 in Basque populations from pre-Neolithic times for some lineages.

5. Materials and methods

5.1 Ancient samples

5.1.1 Sample selection

Since the discovery of the Cave of Santimamiñe several human remains have been found at different stratigraphic units throughout numerous dig campaigns (1918-1926, 1960-1962) (41), (10). Last excavation work promoted a deeper study of them that led to an exhaustive anthropological study (42). Based on that study, seven human teeth, of apparently different individuals, were selected for genetic analysis (Dataset S1). Every selected tooth was attached to th jaw bone, and without visible damage signs, like crown abrasion, cracks and cavities.

Ancient human remains analyzed in this study were deposited in the collection of the Museum of Archaeology of Biscay. The genetic analysis of these ancient remains was granted by the institutions that are in charge of the collection, by the approval of the Cultural Heritage Service of the Council of the province of Biscay and the Director of Cultural Heritage of the Department of Culture of the Basque Government, as they were issued by the resolution of 16th June 2009 and 9th October 2009, respectively. After the analysis, all remaining material was returned to the Museum of Archaeology of Biscay where it is accessible by other researchers.

5.1.2 Working conditions

Every step involving ancient samples was carried out in laboratories exclusively dedicated to ancient DNA analysis. These rooms have positive pressure airflow, never have housed modern DNA samples and every working materials and surfaces are repeatedly cleaned with bleach (50%), milliQ autoclaved H_2O and ethanol (80%), and subjected to UV radiation for their sterilization. Access to these laboratories is limited and the use of specific clothes, as well as lab coat, face mask, bouffant cap, and double pair of gloves, which are frequently changed, is mandatory.

5.1.3 Cleaning and powdering

Teeth were removed from jaw bones with a Dremel tool. Then, they were cleaned to eliminate any traces of modern contamination from their surface by scratching the surface of each tooth with a scalpel, washing with bleach (10%) and rinsing with milliQ autoclaved H_2O. Each face of the tooth surface was irradiated with UV during 30 minutes into a CL-1000 UltraViolet CrossLinker (UVP, Upland, CA) at 450 μJ/cm^2, and finally each tooth was ground to a fine powder with a 6750 Freezer/Mill® (SPEX SamplePrep, Metuchen, NJ).

5.1.4 DNA extraction and quantification

Along each DNA extraction process we only worked with two different tooth samples and an extraction blank to avoid and detect possible contamination during the process. In addition, direct manipulation of the samples and reagents was always carried out into a laminar flow cabinet.

For DNA extraction 0.5 g of tooth powder per sample were used, and modified protocol from that described by (43), to adapt it to our laboratory, was followed. Prior to DNA extraction, we incubated the samples with EDTA 0.5M pH8 overnight on a rotary mixer to ease the demineralization. Digestion of tooth powder was performed by incubating the samples on the rotary mixer overnight at 55°C with a lysis buffer composed of EDTA 0.5M pH8, sodium lauroyl sarcosinate 1% and proteinase K (1mg/ml). The DNA of the supernatant was isolated using silica adsorption columns of Generon Hi-Flow® DNA Purification Spin Columns (Generon Ltd., Berkshire, UK) following the instructions provided by the manufacturer and using buffers of Qiagen (Germantown, MD). Elution step was carried out four times to maximize recovery of DNA, which was finally purified and concentrated with Amicon YM30 columns (MilliPore, Billerica, MA). Finally, DNA samples were quantified by qPCR with Quantifiler® Human DNA Quantification kit (Applied Biosystems, Foster City, CA) into an ABI Prism 7000 Sequence Detection System (Applied Biosystems, Foster City, CA). Additionally, this process allowed us to determine the presence of inhibition that could interfere in subsequent analyses.

5.1.5 Analysis of mtDNA control region

Hypervariable segments HVS-I and HVS-II of the mtDNA control region were PCR amplified by using five overlapping fragments of ~175 base pairs (bp) each one: two fragments for HVS-I and three fragments for HVS-II (Table S1). Bases 16159 to 16406 for HVS-I and 12 to 408 for HVS-II were edited.

PCR was carried out into a BioRad C-1000 thermal cycler (BioRad, Hercules, CA), with an initial denaturalization step of 3 minutes at 95°C, followed by 37 cycles of 45 seconds at 95°C, 60 seconds at 55°C and 60 seconds at 72°C, and a final elongation of 10 minutes at 72°C. The amplified products were cleaned with a mix of exonuclease I and shrimp alkaline phosphatase enzymes (TaKaRa Bio Inc., Shiga, Japan), at 37°C during 30 minutes, followed by 15 minutes at 80°C in the same thermal cycler. Then, PCR products were sequenced using BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) into an ABI3130 Genetic Analyzer (Applied Biosystems, Foster City, CA).

Sequences were edited with software ChromasPro v1.5 (Technelysium Pty Ltd, AU), and aligned and compared to the revised Cambridge Reference Sequence (44), rCRS (RefSeq NC_012920.1), with Clustal X v2 (45). mtDNA haplogroup assignment was performed according to PhyloTree Build 17 (http://www.phylotree.org) (46).

5.1.6 Analysis of mtDNA coding region

Haplogroup diagnostic polymorphisms located in the mtDNA coding region were analyzed to confirm, or even determine more specifically, haplogroup classification based on control region sequences. The analysis of coding region polymorphisms was performed by direct sequencing, as previously described for control region fragments, by using the primers showed in table S1.

5.1.7 Analysis of isotopes

A small amount of powder, between 30 and 100 mg, of each tooth was used to analyze its isotopic composition, both carbon (^{14}C and $^{13}C/^{12}C$) and nitrogen ($^{15}N/^{14}N$), with the aim of dating the antiquity of these human remains and obtaining additional information about the diet of the individuals. These analyses were performed by Beta Analytic (Miami, FL). Radiocarbon dates provided in this study are 2σ calibrated (95% probability).

5.2 Authentication of ancient DNA

5.2.1 Cloning

To ensure the reliability of the DNA results obtained from ancient samples, and following the recommendations made by Cooper and Poinar (47), 32% of the ancient DNA amplification products were cloned. Samples S.7J and S.16G were selected and their five control region PCR

products previously sequenced were cloned using the pGEM®-T Easy Vector System II (Promega, Madison, WI). Cloned products were purified with PureYield™ Plasmid Miniprep System (Promega, Madison, WI). These products were firstly digested with restriction enzyme *Pvu*II (New England Biolabs, Ipswich, MA) and then sequenced with M13 primers, in forward and reverse direction, using BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) and an ABI3130 Genetic Analyzer (Applied Biosystems, Foster City, CA).

5.2.2 Independent analysis by an external laboratory

Following authentication criteria set by Cooper and Poinar (47) two samples (S.11N and S.12N) were analyzed in parallel by the Laboratoire d'Anthropologie Moléculaire of the Institute of Legal Medicine of the Strasbourg University (France). The process of DNA extraction and analysis used is as the described in Lacan *et al*. (48). Additionally, in this laboratory sample S.11N was analyzed by MALDI-TOF mass spectrometry, using iPlex Gold technology (Sequenom, San Diego, CA), to study subhaplogroups defining polymorphisms of the coding region (49).

5.3 Samples of current autochthonous population

Samples from extant population were recruited after participants had provided their written informed consent, following procedures in accordance with the ethical standards of the Declaration of Helsinki. This study counts with the favorable ethical report from the Faculty of Pharmacy of the University of the Basque Country, signed at 26th September 2008.

Saliva samples of 158 autochthonous Basques of the current population inhabiting the surrounding area of the Cave of Santimamiñe were collected with mouthwash. Selection criteria were mainly the Basque origin of the surnames of the voluntary donors, and the geographical origins of their parents and grandparents within the area, verifying no direct kinship among them. The samples were collected in four villages close to the cave and historically related to it, to be named, Ereño, Gautegiz-Arteaga, Kortezubi and Nabarniz, which are located in the region of Biscay known as Busturialdea (Fig. 1).

DNA from saliva samples was extracted following a standard phenol-chloroform protocol (50) and quantified with the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA) in a DTX880 Multimode Detector (Beckman Coulter, Fullerton, CA). The entire control region of mtDNA was amplified and sequenced as described by Cardoso *et al*. (51). Coding region polymorphism determinant of H1j1 haplogroup was analyzed in the same way, using primers shown in table S1. These sequences were edited, aligned and compared to rCRS using SeqScape v2.5 (Applied Biosystems, Foster City, CA). The mtDNA haplogroup assignment was performed according to the mtDNA tree Build 17 of the Phylotree (http://www.phylotree.org) (46). All the control region sequences of the population of Busturialdea are available online at

GenBank under accession numbers KR697593 - KR697750. Likewise, sequences have been deposited into EMPOP (http://empop.online) (52) under accession number EMP00668.

5.4 Statistical and phylogenetic analyses

Multivariate analysis of ancient European populations through Principal Component Analysis (PCA) was performed with PAST v3.01 (53). Population statistics of current population of Busturialdea were calculated with Arlequin v3.5 (54). Software SplitsTree v4 (55) was used to draw the neighbour-joining tree which represented the relation between the haplotypes of modern and ancient samples. Median-joining network was performed with software Network v5 (http://www.fluxus-engineering.com) (56), and used to estimate the time elapsed between ancient and modern lineages. Mutation rate used for this analysis was one mutation every 4,127 years. To calculate this value we took into account the length of the sequences compared in the network (771 bp) and the mutation rate of $3.14x10^{-7}$ mut/site/years. This mutation rate, calculated for segments HVS-I and HVS-II, and calibrated with ancient DNA samples, was recently published by Rieux et al. (57) and is very close to that previously calculated by Sigurðadóttir et al. (58) and Endicott and Ho (59) for same segments and control region respectively. For both, neighbor-joining tree and median-joining network, the sequence of the mitochondrial Most Recent Common Ancestor (mtMRCA) was used as outgroup. This sequence is supposed to be on the root of the human mitochondrial phylogenetic tree and matches with Reconstructed Sapiens Reference Sequence (RSRS) (28).

References

1. González Morales MR, Straus LG, Diez Castillo A, Ruiz Cobo J. La costa y el interior en la Época Postglacial: las Transiciones Epipaleolítico-Mesolítico-Neolítico en la Región Vasco-Cantábrica. Munibe Antropologia - Arkeologia. 2004(56):61-78.
2. Alday Ruiz A. El final del Mesolítico y los inicios del Neolítico en la Península Ibérica: cronología y fases. Munibe Antropologia - Arkeologia. 2009(60):157-73.
3. Zilhao J. Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. Proceedings of the National Academy of Sciences of the United States of America. 2001;98(24):14180-5.
4. Zapata Peña L. Agricultura prehistórica en el País Vasco litoral. Munibe Antropologia - Arkeologia. 2005(57):553-61.
5. Arias Cabal P, Altuna Etxabe J, Armendáriz Gutiérrez A, González Urquijo JE, Ibáñez Estévez JJ, Ontañón Peredo R, et al. La transición al Neolítico en la región cantábrica: estado de la cuestión. 3º Congresso de Arqueología Peninsular : UTAD, Vila Real, Portugal, setembro de 1999. 32000. p. 115-34.
6. Fernández Eraso J, Blas Cortina MÁd, Delibes de Castro G, García Gazólaz J, Iriarte Chiapusso MJ, Rodanés Vicente JM. Mesolítico y Neolítico: pasado, presente y futuro de las investigaciones. Medio siglo de arqueología en el Cantábrico Oriental y su Entorno: actas del Congreso Internacional: Vitoria-Gasteiz : Instituto Alavés de Arqueología = Arkeologarako Arabar Institua, D.L. 2009; 2009. p. 73-114.

7. Richards MP, Price TD, Koch E. The mesolithic and neolithic subsistence in Denmark : new stable isotope data. Current anthropology. 2003;44(2):288-94.
8. Richards MP, Schulting RJ, Hedges RE. Archaeology: sharp shift in diet at onset of Neolithic. Nature. 2003;425(6956):366.
9. Arias Cabal P. Determinaciones de isótopos estables en restos humanos de la región Cantábrica: aportación al estudio de la dieta de las poblaciones del Mesolítico y el Neolítico. Munibe Antropologia - Arkeologia. 2005(57):359-74.
10. López Quintana JC. La ocupación humana de Santimamiñe (Kortezubi): paisaje, recursos y estrategias de explotación del medio desde el Magdaleniense inferior al Calcolítico-Edad del Bronce. La cueva de Santimamiñe: revisión y actualización (2004-2006): [Bilbao] : Diputación Foral de Bizkaia, 2011; 2011. p. 421-46.
11. Barandiarán JM. Recapitulación y apéndice. In: Barandiarán JM, editor. Obras Completas. Tomo IX1976. p. 421 - 75.
12. Gutiérrez Zugasti FI. Los moluscos alimenticios de la cueva de Santimamiñe (Kortezubi, Bizkaia): campañas de excavación 2004-2006. La cueva de Santimamiñe: revisión y actualización (2004-2006): [Bilbao] : Diputación Foral de Bizkaia, 2011; 2011. p. 247-66.
13. Llanos Ortiz de Landaluze A, Urteaga Artigas MM. Notas sobre el pastoreo durante la prehistoria en el País Vasco Peninsular. Estudios de Arqueología Alavesa. 2002(19):82-95.
14. Tafuri MA, Craig OE, Canci A. Stable isotope evidence for the consumption of millet and other plants in Bronze Age Italy. Am J Phys Anthropol. 2009;139(2):146-53.
15. Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Jane Adler C, et al. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nat Commun. 2013;4:1764.
16. Brandt G, Haak W, Adler CJ, Roth C, Szecsenyi-Nagy A, Karimnia S, et al. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. Science (New York, NY). 2013;342(6155):257-61.
17. Lacan M, Keyser C, Crubezy E, Ludes B. Ancestry of modern Europeans: contributions of ancient DNA. Cellular and molecular life sciences : CMLS. 2013;70(14):2473-87.
18. Gomez-Sanchez D, Olalde I, Pierini F, Matas-Lalueza L, Gigli E, Lari M, et al. Mitochondrial DNA from El Mirador cave (Atapuerca, Spain) reveals the heterogeneity of Chalcolithic populations. PLoS One. 2014;9(8):e105105.
19. Gunther T, Valdiosera C, Malmstrom H, Urena I, Rodriguez-Varela R, Sverrisdottir OO, et al. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(38):11917-22.
20. Pala M, Olivieri A, Achilli A, Accetturo M, Metspalu E, Reidla M, et al. Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. Am J Hum Genet. 2012;90(5):915-24.
21. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, 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. 2004;75(5):910-8.
22. Alvarez-Iglesias V, Mosquera-Miguel A, Cerezo M, Quintans B, Zarrabeitia MT, Cusco I, et al. New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. PLoS One. 2009;4(4):e5112.
23. Cardoso S, Alfonso-Sanchez MA, Valverde L, Odriozola A, Perez-Miranda AM, Pena JA, et al. The maternal legacy of Basques in northern navarre: New insights into the mitochondrial DNA diversity of the Franco-Cantabrian area. Am J Phys Anthropol. 2011;145(3):480-8.
24. Garcia O, Fregel R, Larruga JM, Alvarez V, Yurrebaso I, Cabrera VM, et al. Using mitochondrial DNA to test the hypothesis of a European post-glacial human recolonization from the Franco-Cantabrian refuge. Heredity. 2011;106(1):37-45.

25. Behar DM, Harmant C, Manry J, van Oven M, Haak W, Martinez-Cruz B, et al. The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times. Am J Hum Genet. 2012;90(3):486-93.
26. Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, et al. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science (New York, NY). 2009;326(5949):137-40.
27. Malyarchuk B, Derenko M, Grzybowski T, Perkova M, Rogalla U, Vanecek T, et al. The peopling of Europe from the mitochondrial haplogroup U5 perspective. PLoS One. 2010;5(4):e10285.
28. Behar DM, van Oven M, Rosset S, Metspalu M, Loogvali EL, Silva NM, et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet. 2012;90(4):675-84.
29. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, et al. Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet. 2000;67(5):1251-76.
30. Salamon M, Tzur S, Arensburg B, Zias J, Nagar Y, Weiner S, et al. Ancient mtdna sequences and radiocarbon dating of human bones from the chalcolithic caves of wadi el-makkukh. Mediterranean Archaeology and Archaeometry. 2010;10(2):1-14.
31. Melchior L, Gilbert MT, Kivisild T, Lynnerup N, Dissing J. Rare mtDNA haplogroups and genetic differences in rich and poor Danish Iron-Age villages. Am J Phys Anthropol. 2008;135(2):206-15.
32. Hedman M, Brandstätter A, Pimenoff V, Sistonen P, Palo JU, Parson W, et al. Finnish mitochondrial DNA HVS-I and HVS-II population data. Forensic Sci Int. 2007;172(2-3):171-8.
33. Tillmar AO, Coble MD, Wallerström T, Holmlund G. Homogeneity in mitochondrial DNA control region sequences in Swedish subpopulations. Int J Legal Med. 2010;124(2):91-8.
34. Lembring M, van Oven M, Montelius M, Allen M. Mitochondrial DNA analysis of Swedish population samples. Int J Legal Med. 2013;127(6):1097-9.
35. Alfonso-Sanchez MA, Cardoso S, Martinez-Bouzas C, Pena JA, Herrera RJ, Castro A, et al. Mitochondrial DNA haplogroup diversity in Basques: a reassessment based on HVI and HVII polymorphisms. American journal of human biology : the official journal of the Human Biology Council. 2008;20(2):154-64.
36. Cardoso S, Villanueva-Millán MJ, Valverde L, Odriozola A, Aznar JM, Piñeiro-Hermida S, et al. Mitochondrial DNA control region variation in an autochthonous Basque population sample from the Basque Country. Forensic Sci Int Genet. 2012;6(4):e106-8.
37. Cardoso S, Valverde L, Alfonso-Sanchez MA, Palencia-Madrid L, Elcoroaristizabal X, Algorta J, et al. The expanded mtDNA phylogeny of the Franco-Cantabrian region upholds the pre-neolithic genetic substrate of Basques. PLoS One. 2013;8(7):e67835.
38. Brandt G, Szecsenyi-Nagy A, Roth C, Alt KW, Haak W. Human paleogenetics of Europe--the known knowns and the known unknowns. J Hum Evol. 2015;79:73-92.
39. Hervella M, Izagirre N, Alonso S, Fregel R, Alonso A, Cabrera VM, 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. 2012;7(4):e34417.
40. Sanchez-Quinto F, Schroeder H, Ramirez O, Avila-Arcos MC, Pybus M, Olalde I, et al. Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. Current biology : CB. 2012;22(16):1494-9.
41. Aranzadi Td, Barandiarán JMd, Eguren Ed, Vizcaya (España). Diputación Foral. Junta de Cultura. Exploraciones de la Caverna de Santimamiñe : (Basondo, Cortézubi)2a Memoria, Los niveles con cerámica y el conchero. Bilbao: Imprenta de la Exma. Diputación de Vizcaya; 1931. 114 p., XLI p. de lám. p.

42. Herrasti Erlogorri L, Etxeberria Gabilondo F. Estudio de los restos humanos de la cueva de Santimamiñe (Kortezubi, Bizkaia): campañas de 2004 a 2006. La cueva de Santimamiñe:

revisión y actualización (2004-2006): [Bilbao] : Diputación Foral de Bizkaia, 2011; 2011. p. 375-82.
43. Marshall PL, Stoljarova M, Schmedes SE, King JL, Budowle B. A high volume extraction and purification method for recovering DNA from human bone. Forensic Sci Int Genet. 2014;12:155-60.
44. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature genetics. 1999;23(2):147.
45. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics (Oxford, England). 2007;23(21):2947-8.
46. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Human mutation. 2009;30(2):E386-94.
47. Cooper A, Poinar HN. Ancient DNA: do it right or not at all. Science (New York, NY). 2000;289(5482):1139.
48. Lacan M, Keyser C, Ricaut FX, Brucato N, Tarrus J, Bosch A, et al. Ancient DNA suggests the leading role played by men in the Neolithic dissemination. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(45):18255-9.
49. Mendisco F, Keyser C, Hollard C, Seldes V, Nielsen AE, Crubezy E, et al. Application of the iPLEX Gold SNP genotyping method for the analysis of Amerindian ancient DNA samples: benefits for ancient population studies. Electrophoresis. 2011;32(3-4):386-93.
50. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual: Cold Spring Harbor Laboratory; 1989.
51. Cardoso S, Zarrabeitia MT, Valverde L, Odriozola A, Alfonso-Sanchez MA, de Pancorbo MM. Variability of the entire mitochondrial DNA control region in a human isolate from the Pas Valley (northern Spain). Journal of forensic sciences. 2010;55(5):1196-201.
52. Parson W, Dür A. EMPOP—a forensic mtDNA database. Forensic Science International: Genetics. 2007;1(2):88-92.
53. Hammer Ø, Harper D, Ryan P. PAST-Palaeontological statistics. www uv es/~ pardomv/pe/2001_1/past/pastprog/past pdf, acessado em. 2001;25(07):2009.
54. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10(3):564-7.
55. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution. 2006;23(2):254-67.
56. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16(1):37-48.
57. Rieux A, Eriksson A, Li M, Sobkowiak B, Weinert LA, Warmuth V, et al. Improved calibration of the human mitochondrial clock using ancient genomes. Mol Biol Evol. 2014;31(10):2780-92.
58. Sigurğardóttir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P. The mutation rate in the human mtDNA control region. Am J Hum Genet. 2000;66(5):1599-609.
59. Endicott P, Ho SY. A Bayesian evaluation of human mitochondrial substitution rates. Am J Hum Genet. 2008;82(4):895-902.
60. Cardoso S, Alfonso-Sanchez MA, Valverde L, Sanchez D, Zarrabeitia MT, Odriozola A, et al. Genetic uniqueness of the Waorani tribe from the Ecuadorian Amazon. Heredity. 2012;108(6):609-15.
61. Grignani P, Peloso G, Achilli A, Turchi C, Tagliabracci A, Alu M, et al. Subtyping mtDNA haplogroup H by SNaPshot minisequencing and its application in forensic individual identification. Int J Legal Med. 2006;120(3):151-6.
62. Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, et al. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet. 2004;75(6):966-78.

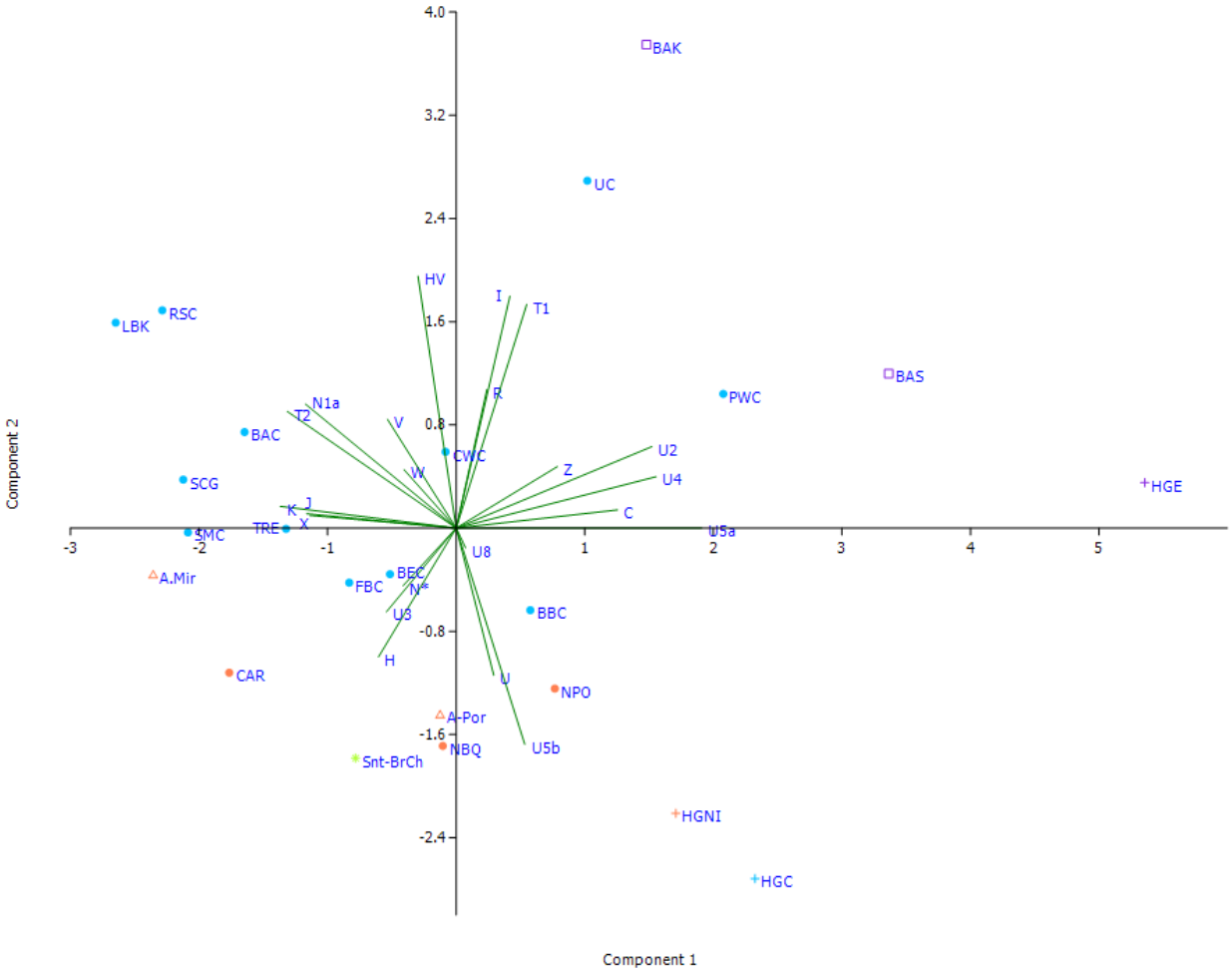

Fig. S1. Biplot representation of the PCA of European prehistoric populations based on their haplogroup frequencies. Each population is represented by different symbols and colours. The symbol refers to the cultural period being cross for hunter-gatherers and pre-Neolithic groups; circle for Neolithic populations; triangle for Chalcolithic period and square for Bronze Age. The colour refers to the geographical origin, with purple for Eastern European and Asian populations, blue for central Europe, turquoise for Southern France and orange for populations of Iberian Peninsula. The group of samples from Santimamiñe of Chalcolithic and Bronze Age is represented by a green asterisk. Data is shown in dataset S3.

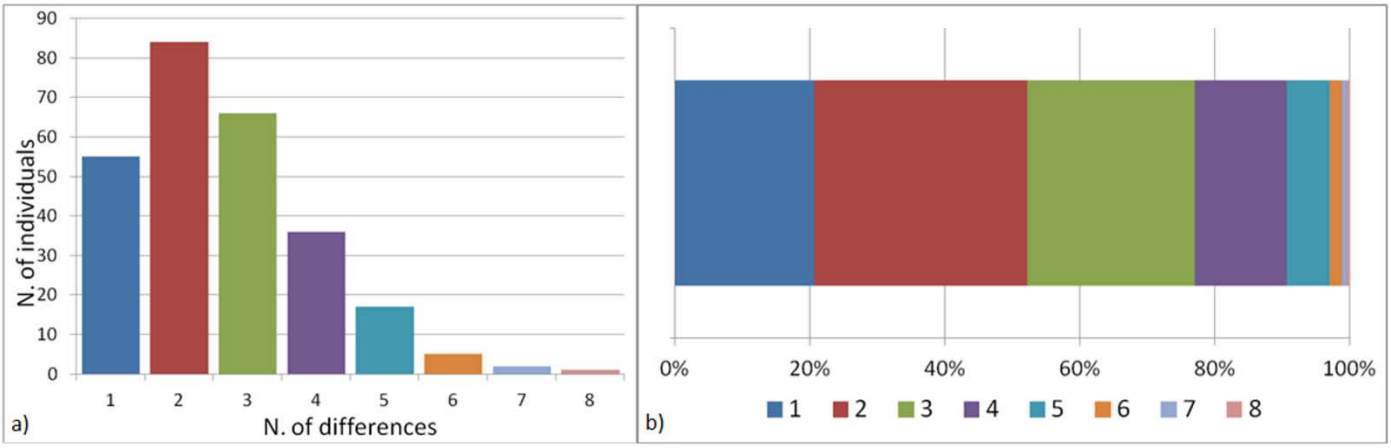

Fig. S2. Number of differences found between ancient samples of the Cave of Santimamiñe and the individuals of the population sample of Busturialdea that belong to the same mtDNA phylogenetic branches. a) Number of individuals against total number of differences. b) Frequency of the number of differences that present individuals of Busturialdea that are phylogeneticaly close to the remains of the Cave of Santimamiñe.

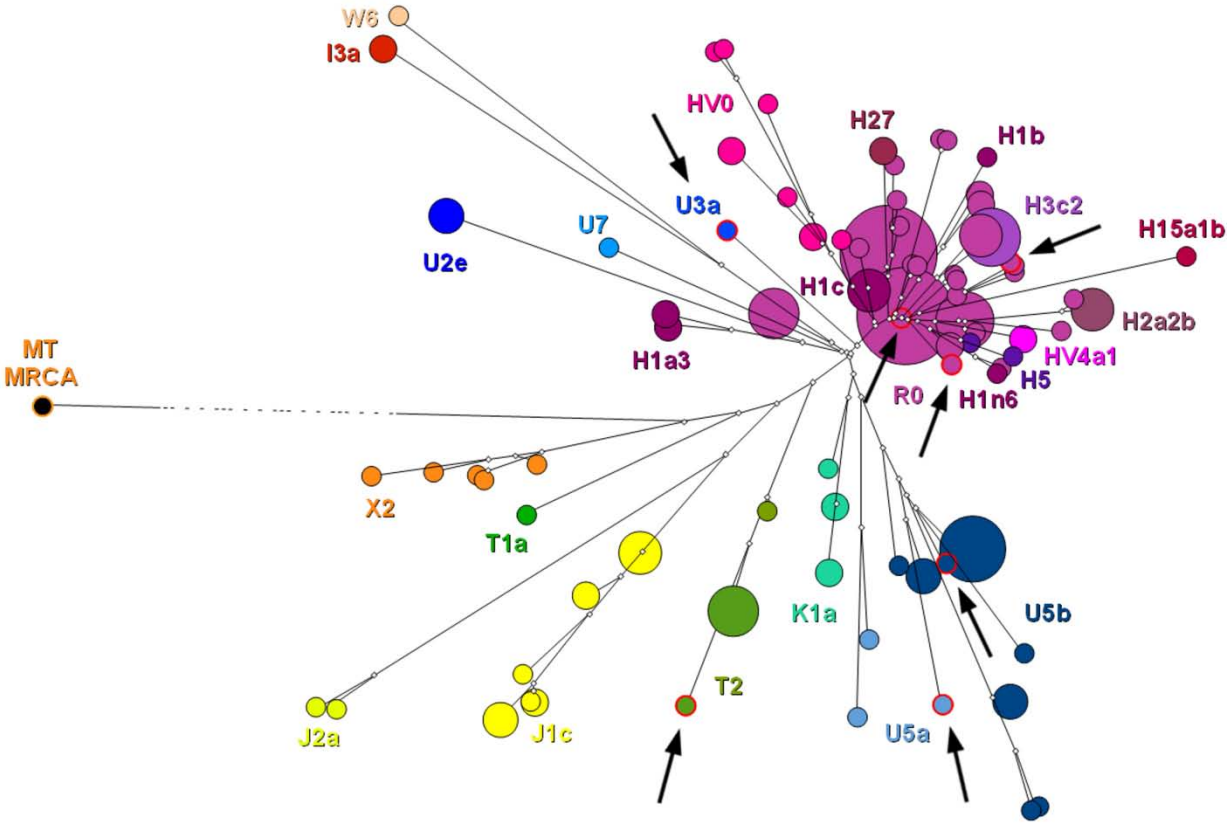

Fig. S3. Neighbour-joining tree that represents the phylogenetic relationship between ancient samples of Santimamiñe (nodes with red edge and marked with arrows) and the extant population of Busturialdea (nodes with black edge). The mitochondrial control region haplotype of the most recent common matrilineal ancestor (mtMRCA) as can be inferred from Phylotree (46) was used as outgroup. The size of each node is proportional to the number of samples that share their haplotype.

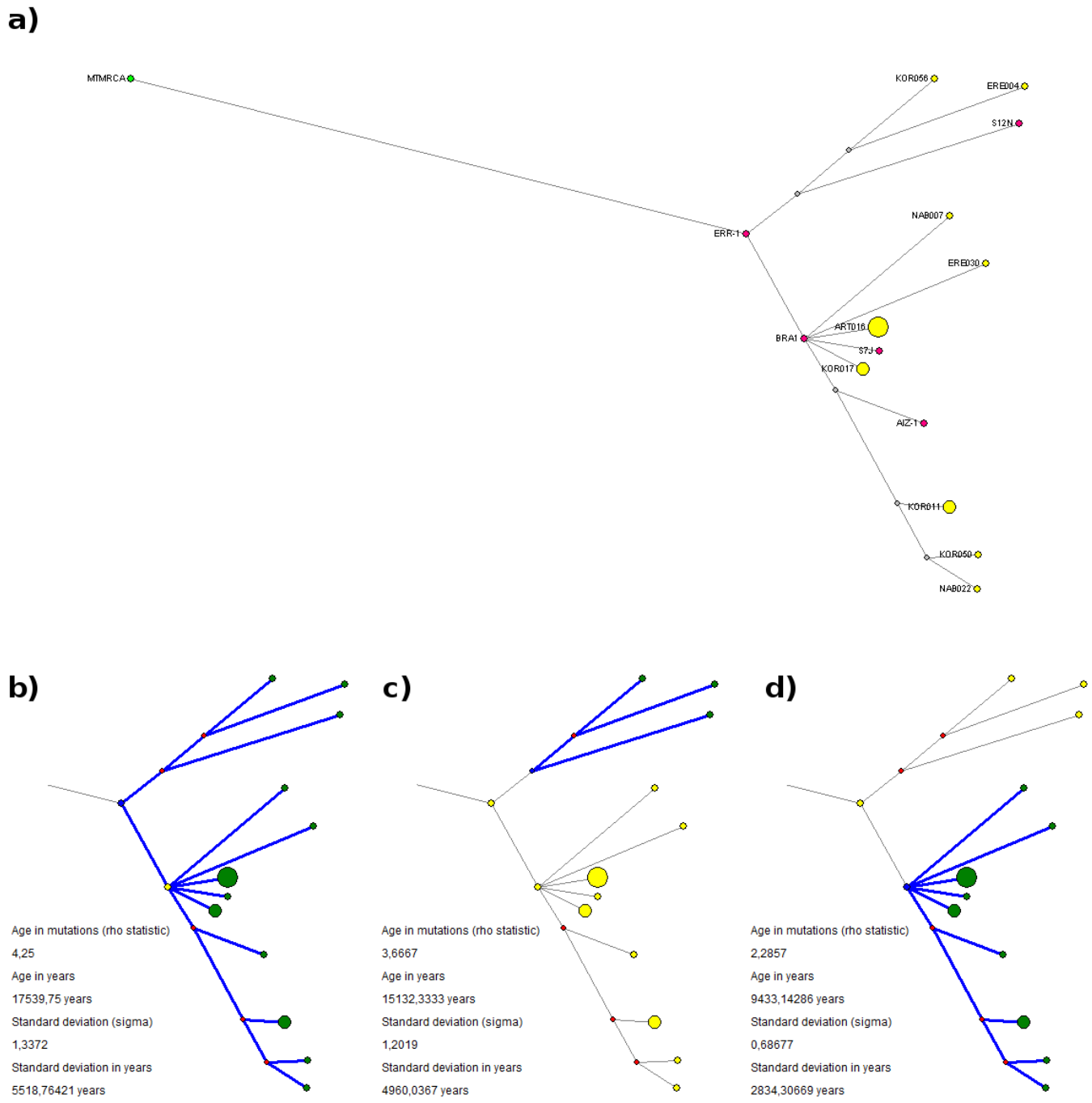

Fig. S4. Median-joining network and time estimates obtained from the analysis of haplogroup U samples. a) Median-joining network draw with samples of haplogroups U from ancient individuals of the northern Iberian Peninsula and current individuals from Busturialdea. Ancient samples are represented by pink-fuchsia nodes and modern samples from Busturialdea by yellow nodes. The outgroup node, in light green, is the mitochondrial control region haplotype of the most recent common matrilineal ancestor (mtMRCA) as can be inferred from Phylotree. The size of each node is proportional to the number of samples that share their haplotype. b) Age estimated for U branch, taking ERR-1 as ancestral node and whole remaining samples of haplogroup U as descendant nodes. c) Age estimated for U5a

branch, taking the convergent vector as ancestral node and the remaining U5a samples of Santimamiñe and Busturialdea as descendant nodes. d) Age estimated for U5b branch, taking BRA-1 as ancestral node and the remaining samples of subhaplogroup U5b as descendants.

Table S1. List of primers used to analyze ancient samples of the Cave of Santimamiñe.

Primer name	Sequence	Primer length (bp)	Amplified product	Target region or polymorphisms	Reference
L16159	TACTTGACCACCTGTAGTAC	20	191 bp	HVS-I	Present study
H16311	TACGGTAAATGGCTTTATGT	20			
L16287	CACCCACTAGGATACCAACAAAC	23	164 bp	HVS-I	Present study
H16406	CGGGATATTGATTTCACGGAGG	22			
L12	ACATCACGATGGATCACAGGTC	22	177 bp	HVS-II	Present study
H142	GTGCGATAAATAATAGGATGAGGC	24			
L135	CTATGTCGCAGTATCTGTCTT	21	192 bp	HVS-II	Present study
H285	GGGGTTTGGTGGAAATTTTTTG	22			
L281	AGCCACTTTCCACACAGACATC	22	172 bp	HVS-II	Present study
H408.2	ACTGTTAAAAGTGCATACCGCC	22			
L467	ACATTATTTTCCCCTCCCACT	21	169 bp	A533G (U5b1f1)	Present study
H589	TTTCAGTGTATTGCTTTGAGGAGGT	25			
L1692	CTAAACCTAGCCCCAAACCC	20	121 bp	C1721T (U5b2)	Present study
H1769	CCTTGCGGTACTATATCTATTGC	23			
L5379	AATCTACTCCACCTCAATCACA	22	124 bp	C5437T (U5b1d)	Present study
H5462	GTAGGAGTAGCGTGGTAAG	19			
L14724	AATGATATGAAAAACCATCGTTG	23	153 bp	A14793G (U5a)	(60)
H14834	GAGTGAGCCGAAGTTTCATCA	21			
Fw456	CACCAGCCTAACCAGATTTCA	21	125 bp	G456A (H5'36)	(61)
Rv456	CGGGGGTTGTATTGATGAGA	20			
Fw3936	TAGCAGAGACCAACCGAACC	20	158 bp	C3915T (H6a), C3936T (H12), C3992T (H4)	(61)
Rv3936	GAAGATTGTAGTGGTGAGGGTGT	23			
Fw4745	CCGGACAATGAACCATAACC	20	118 bp	A4745G (H13a1), C4769T (H2a), T4793C (H7)	(61)
Rv4745	TGGGTAACCTCTGGGACTCA	20			
Fw6253	TTTCCCCGCATAAACAACAT	20	113 bp	T6253C (H15)	(61)
Rv6253	GGAGGGTAGACTGTTCAACCTG	22			
Fw8448	ATGGCCCACCATAATTACCC	20	170 bp	A8448G (H11)	(61)
Rv8448	GCAATGAATGAAGCGAACAG	20			
Fw11377	CTTCCCCTACTCATCGCACT	20	193 bp	C11377T (H1bh)	(61)
Rv11377	CGACATGGGCTTTAGGGAGT	20			
Fw13020-13101	AAACGCTAATCCAAGCCTCA	20	182 bp	T13020C (H9), A13101C (H8)	(61)
Rv13020-13101	GGTGGAAGCGGATGAGTAAG	20			
Fw14470	CCCCATGCCTCAGGATAC	18	154 bp	T14470A (H10)	(61)
Rv14470	TGATTGTTAGCGGTGTGGTC	20			
Fw2642	CGAGGGTTCAGCTGTCTCTTAC	22	411 bp	G2706A (H), G3010A (H1)	(23)
Rv3052	TTAATCGTTGAACAAACGAACC	22			
Fw6702	CCATTTGGATACATAGGTATGGT	23	377 bp	T7028C (H), A6776G (H3)	(23)
Rv7078	CCTCCTATGATGGCAAATACAG	22			
L4663	AGTATTTCCTCACGCAAGCAAC	22	479 bp	T4733C (H1j)	(60) (62)
H5099	GGAATGCGGTAGTAGTTAGG	20			

Table S2. Haplogroup distribution of population sample from Busturialdea (Biscay).

Haplogroup	Arteaga	Ereño	Kortezubi	Nabarniz	Total: Busturialdea	Frequency (%)
H1	8	1	15	2	26	16,456
H2	0	4	1	0	5	3,165
H3	2	1	2	1	6	3,797
H5	1	0	0	1	2	1,266
H15	0	0	0	1	1	0,633
H27	1	0	1	0	2	1,266
HV0	4	1	1	3	9	5,696
HV4	1	0	0	1	2	1,266
I3a	0	0	2	0	2	1,266
J1c	5	2	6	2	15	9,494
J2a	1	0	1	0	2	1,266
K1a	2	1	0	2	5	3,165
R0	14	10	14	6	44	27,848
T1a	1	0	0	0	1	0,633
T2	1	3	1	1	6	3,797
U2e1	0	2	1	0	3	1,899
U5a1	0	1	1	0	2	1,266
U5b	3	4	8	3	18	11,392
U7	0	0	0	1	1	0,633
W6	0	0	1	0	1	0,633
X2	3	1	1	0	5	3,165
Total	47	31	56	24	158	100

Table S3. Summary statistics for 158 control region (16024-576) sequences from the region of Busturialdea (Biscay). Insertions at 16193, 309 and 573 were ignored for all calculations.

Population statistics	Busturialdea (n=158)	Kortezubi (n=56)	Ereño (n=31)	Gautegiz-Arteaga (n=47)	Nabarniz (n=24)
Random match probability (sum of squares)	0.0237	0.0395	0.0527	0.0448	0.0473
Power of discrimination	0.9763	0.9605	0.9473	0.9552	0.9527
Haplotypes	83 (27 shared)	39 (10 shared)	23 (7 shared)	32 (10 shared)	22 (1 shared)
Mean pairwise diferences	8.762154 ± 4.064500	9.598701 ± 4.467086	8.310484 ± 3.952981	8.166512 ± 3.856499	8.735178 ± 4.184190
Genetic diversity (H)	0.9825 ± 0.0034	0.9779 ± 0.0104	0.9778 ± 0.0131	0.9759 ± 0.0117	0.9960 ± 0.0142
Nucleotide diversity	0.007775 ± 0.003992	0.008525 ± 0.004401	0.007400 ± 0.003915	0.007259 ± 0.003805	0.007758 ± 0.004144