LETTER

doi:10.1038/nature12960

Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European

Iñigo Olalde[1]*, Morten E. Allentoft[2]*, Federico Sánchez-Quinto[1], Gabriel Santpere[1], Charleston W. K. Chiang[3], Michael DeGiorgio[4,5], Javier Prado-Martinez[1], Juan Antonio Rodríguez[1], Simon Rasmussen[6], Javier Quilez[1], Oscar Ramírez[1], Urko M. Marigorta[1], Marcos Fernández-Callejo[1], María Encina Prada[7], Julio Manuel Vidal Encinas[8], Rasmus Nielsen[9], Mihai G. Netea[10], John Novembre[11], Richard A. Sturm[12], Pardis Sabeti[13,14], Tomàs Marquès-Bonet[1,15], Arcadi Navarro[1,15,16,17], Eske Willerslev[2] & Carles Lalueza-Fox[1]

Ancient genomic sequences have started to reveal the origin and the demographic impact of farmers from the Neolithic period spreading into Europe[1–3]. The adoption of farming, stock breeding and sedentary societies during the Neolithic may have resulted in adaptive changes in genes associated with immunity and diet[4]. However, the limited data available from earlier hunter-gatherers preclude an understanding of the selective processes associated with this crucial transition to agriculture in recent human evolution. Here we sequence an approximately 7,000-year-old Mesolithic skeleton discovered at the La Braña-Arintero site in León, Spain, to retrieve a complete pre-agricultural European human genome. Analysis of this genome in the context of other ancient samples suggests the existence of a common ancient genomic signature across western and central Eurasia from the Upper Paleolithic to the Mesolithic. The La Braña individual carries ancestral alleles in several skin pigmentation genes, suggesting that the light skin of modern Europeans was not yet ubiquitous in Mesolithic times. Moreover, we provide evidence that a significant number of derived, putatively adaptive variants associated with pathogen resistance in modern Europeans were already present in this hunter-gatherer.

Next-generation sequencing (NGS) technologies are revolutionizing the field of ancient DNA (aDNA), and have enabled the sequencing of complete ancient genomes[5,6], such as that of Ötzi, a Neolithic human body found in the Alps[1]. However, very little is known of the genetic composition of earlier hunter-gatherer populations from the Mesolithic period (circa 10,000–5,000 years before present, BP; immediately preceding the Neolithic period).

The Iberian site called La Braña-Arintero was discovered in 2006 when two male skeletons (named La Braña 1 and 2) were found in a deep cave system, 1,500 m above sea level in the Cantabrian mountain range (León, Northwestern Spain) (Fig. 1a). The skeletons were dated to approximately 7,000 years BP (7,940–7,690 calibrated BP)[7]. Because of the cold environment and stable thermal conditions in the cave, the preservation of these specimens proved to be exceptional (Fig. 1b). We identified a tooth from La Braña 1 with high human DNA content (48.4%) and sequenced this specimen to a final effective genomic depth-of-coverage of 3.40× (Extended Data Fig. 1).

We used several tests to assess the authenticity of the genome sequence and to determine the amount of potential modern human contamination. First, we observed that sequence reads from both the mitochondrial

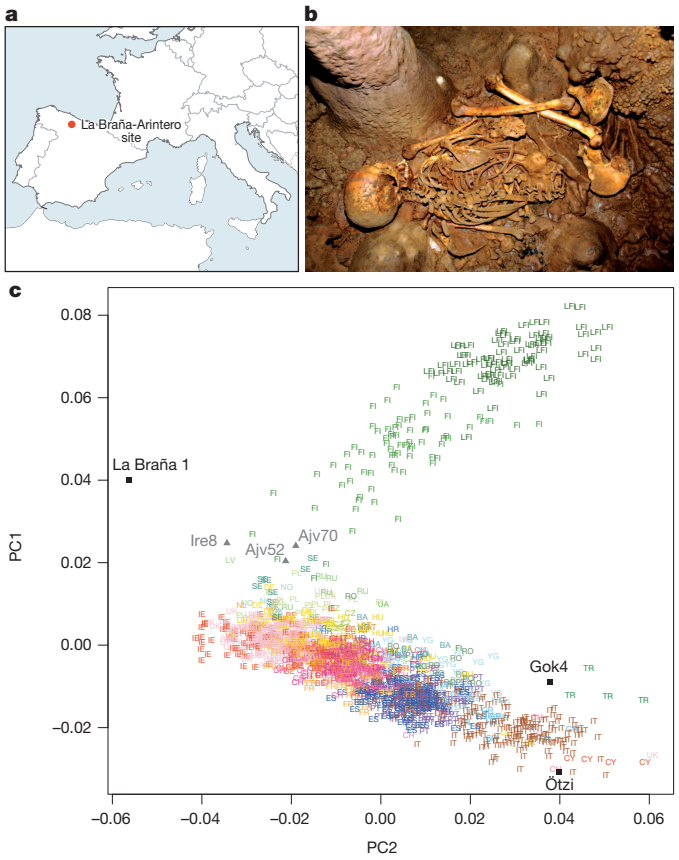


Figure 1 | Geographic location and genetic affinities of the La Braña 1 individual. a, Location of the La Braña-Arintero site (Spain). **b**, The La Braña 1 skeleton as discovered in 2006. **c**, PCA based on the average of the Procrustes transformations of individual PCAs with La Braña 1 and each of the five Neolithic samples[1,3]. The reference populations are the Finnish HapMap, FINHM and POPRES. Population labels with labelling of ref. 12 with the addition of FI (Finns) or LFI (late-settlement Finns). Ajv70, Ajv52, Ire8 and Gok4 are Scandinavian Neolithic hunter-gatherers and a farmer, respectively[3]. Ötzi is the Tyrolean Ice Man[1].

[1]Institut de Biologia Evolutiva, CSIC-UPF, Barcelona 08003, Spain. [2]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, DK-1350 Copenhagen K, Denmark. [3]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California 90095, USA. [4]Department of Integrative Biology, University of California, Berkeley, California 94720, USA. [5]Department of Biology, Pennsylvania State University, 502 Wartik Laboratory, University Park, Pennsylvania 16802, USA. [6]Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark. [7]I.E.S.O. 'Los Salados', Junta de Castilla y León, E-49600 Benavente, Spain. [8]Junta de Castilla y León, Servicio de Cultura de León, E-24071 León, Spain. [9]Center for Theoretical Evolutionary Genomics, University of California, Berkeley, California 94720, USA. [10]Department of Medicine and Nijmegen Institute for Infection, Inflammation and Immunity, Radboud University Nijmegen Medical Centre, 6500 Nijmegen, The Netherlands. [11]Department of Human Genetics, University of Chicago, Illinois 60637, USA. [12]Institute for Molecular Bioscience, Melanogenix Group, The University of Queensland, Brisbane, Queensland 4072, Australia. [13]Center for Systems Biology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [14]Broad Institute of the Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02142, USA. [15]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain. [16]Centre de Regulació Genòmica (CRG), Barcelona 08003, Catalonia, Spain. [17]National Institute for Bioinformatics (INB), Barcelona 08003, Catalonia, Spain.
*These authors contributed equally to this work.

DNA (mtDNA) and the nuclear DNA of La Braña 1 showed the typical ancient DNA misincorporation patterns that arise from degradation of DNA over time[8] (Extended Data Fig. 2a, b). Second, we showed that the observed number of human DNA fragments was negatively correlated with the fragment length ($R^2 > 0.92$), as expected for ancient degraded DNA, and that the estimated rate of DNA decay was low and in agreement with predicted values[9] (Extended Data Fig. 2c, d). We then estimated the contamination rate in the mtDNA genome, assembled to a high depth-of-coverage (91×), by checking for positions differing from the mtDNA genome (haplogroup U5b2c1) that was previously retrieved with a capture method[2]. We obtained an upper contamination limit of 1.69% (0.75–2.6%, 95% confidence interval, CI) (Supplementary Information). Finally, to generate a direct estimate of nuclear DNA contamination, we screened for heterozygous positions (among reads with >4× coverage) in known polymorphic sites (Single Nucleotide Polymorphism Database (dbSNP) build 137) at uniquely mapped sections on the X chromosome[6] (Supplementary Information). We found that the proportion of false heterozygous sites was 0.31%. Together these results suggest low levels of contamination in the La Braña 1 sequence data.

To investigate the relationship to extant European samples, we conducted a principal component analysis (PCA)[10] and found that the approximately 7,000-year-old Mesolithic sample was divergent from extant European populations (Extended Data Fig. 3a, b), but was placed in proximity to northern Europeans (for example, samples from Sweden and Finland)[11–14]. Additional PCAs and allele-sharing analyses with ancient Scandinavian specimens[3] supported the genetic similarity of the La Braña 1 genome to Neolithic hunter-gatherers (Ajv70, Ajv52, Ire8) relative to Neolithic farmers (Gok4, Ötzi) (Fig. 1c, Extended Data Figs 3c and 4). Thus, this Mesolithic individual from southwestern Europe represents a formerly widespread gene pool that seems to be partially preserved in some modern-day northern European populations, as suggested previously with limited genetic data[2,3]. We subsequently explored the La Braña affinities to an ancient Upper Palaeolithic genome from the Mal'ta site near Lake Baikal in Siberia[15]. Outgroup f_3 and D statistics[16,17], using different modern reference populations, support that Mal'ta is significantly closer to La Braña 1 than to Asians or modern Europeans (Extended Data Fig. 5 and Supplementary Information). These results suggest that despite the vast geographical distance and temporal span, La Braña 1 and Mal'ta share common genetic ancestry, indicating a genetic continuity in ancient western and central Eurasia. This observation matches findings of similar cultural artefacts across time and space in Upper Paleolithic western Eurasia and Siberia, particularly the presence of anthropomorphic 'Venus' figurines that have been recovered from several sites in Europe and Russia, including the Mal'ta site[15]. We also compared the genome-wide heterozygosity of the La Braña 1 genome to a data set of modern humans with similar coverage (3–4×). The overall genomic heterozygosity was 0.042%, lower than the values observed in present day Asians (0.046–0.047%), Europeans (0.051–0.054%) and Africans (0.066–0.069%) (Extended Data Fig. 6a). The effective population size, estimated from heterozygosity patterns, suggests a global reduction in population size of approximately 20% relative to extant Europeans (Supplementary Information). Moreover, no evidence of tracts of autozygosity suggestive of inbreeding was observed (Extended Data Fig. 6b).

To investigate systematically the timing of selection events in the recent history of modern Europeans, we compared the La Braña genome to modern populations at loci that have been categorized as of interest for their role in recent adaptive evolution. With respect to two recent well-studied adaptations to changes in diet, we found the ancient genome to carry the ancestral allele for lactose intolerance[4] and approximately five copies of the salivary amylase (*AMY1*) gene (Extended Data Fig. 7 and Supplementary Information), a copy number compatible with a low-starch diet[18]. These results suggest the La Braña hunter-gatherer was poor at digesting milk and starch, supporting the hypotheses that these abilities were selected for during the later transition to agriculture.

To expand the survey, we analysed a catalogue of candidate signals for recent positive selection based on whole-genome sequence variation from the 1000 Genomes Project[13], which included 35 candidate non-synonymous variants, ten of which were detected uniquely in the CEU (Utah residents with northern and western European ancestry) sample [19]. For each variant we assessed whether the Mesolithic genome carried the ancestral or derived (putatively adaptive) allele.

Of the ten variants, the Mesolithic genome carried the ancestral and non-selected allele as a homozygote in three regions: *C12orf29* (a gene with unknown function), *SLC45A2* (rs16891982) and *SLC24A5* (rs1426654) (Table 1). The latter two variants are the two strongest known loci affecting light skin pigmentation in Europeans[20–22] and their ancestral alleles and associated haplotypes are either absent or segregate at very low frequencies in extant Europeans (3% and 0% for *SLC45A2* and *SLC24A5*, respectively) (Fig. 2). We subsequently examined all genes known to be associated with pigmentation in Europeans[22], and found ancestral alleles in *MC1R*, *TYR* and *KITLG*, and derived alleles in *TYRP1*, *ASIP* and *IRF4* (Supplementary Information). Although the precise phenotypic effects cannot currently be ascertained in a European genetic background, results from functional experiments[20] indicate that the allelic combination in this Mesolithic individual is likely to have resulted in dark skin pigmentation and dark or brown hair. Further examination revealed that this individual carried the *HERC2* rs12913832*C single nucleotide polymorphism (SNP) and the associated homozygous haplotype spanning the *HERC2–OCA2* locus that is strongly associated

Table 1 | Mesolithic genome allelic state at 10 nonsynonymous variants recently selected in Europeans

Allelic state	Gene	Name	SNP	Amino-acid change	Function
La Braña 1 carries the derived allele	*PTX4*	Pentraxin 4	rs2745098	Arg281Lys	May be involved in innate immunity
	UHRF1BP1	UHRF1 binding protein 1	rs11755393	Gln454Arg	Risk locus for systemic lupus erythematosus
	GPATCH1	G patch domain containing 1	rs10421769	Leu520Ser	Receptor for OmpA expressed by *E. coli*
	WWOX	WW domain-containing oxidoreductase	rs12918952	Ala179Thr	Acts as a tumour suppressor and has a role in apoptosis
	CCDC14	Coiled-coil domain-containing protein 14	rs17310144	Thr365Pro	Unknown
La Braña 1 carries both the ancestral and the derived allele	*SETX*	Senataxin	rs1056899	Val2587Ile	Involved in spinocerebellar ataxia and amyotrophic lateral sclerosis
	TDRD12	Tudor domain containing 12	rs11881633	Glu413Lys	Unknown
La Braña 1 retains the ancestral allele	*C12orf29*	Chromosome 12 open reading frame 29	rs9262	Val238Leu	Unknown
	SLC45A2	Solute carrier family 45, member 2	rs16891982	Leu374Phe	Associated with skin pigmentation
	SLC24A5	Solute carrier family 24, member 5	rs1426654	Ala111Thr	Associated with skin pigmentation

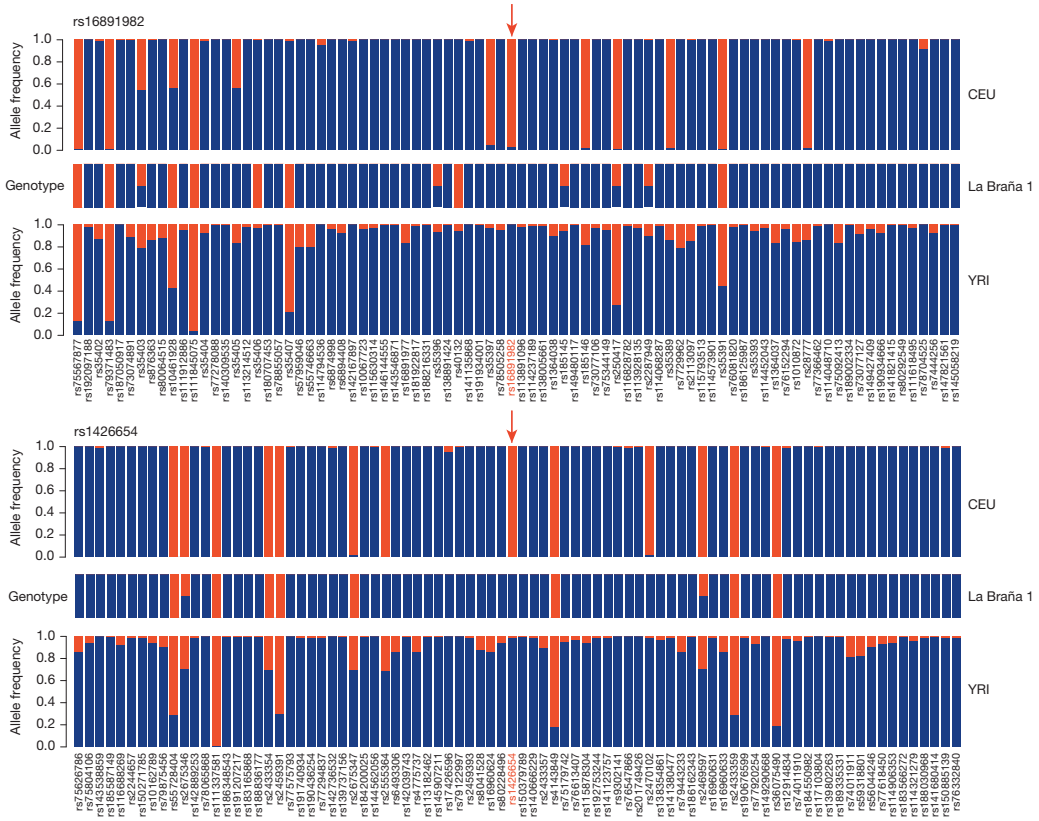


Figure 2 | Ancestral variants around the *SLC45A2* (rs16891982, above) and *SLC24A5* (rs1426654, below) pigmentation genes in the Mesolithic genome. The SNPs around the two diagnostic variants (red arrows) in these two genes were analysed. The resulting haplotype comprises neighbouring SNPs that are also absent in modern Europeans (CEU) (n=112) but present in Yorubans (YRI) (n=113). This pattern confirms that the La Braña 1 sample is older than the positive-selection event in these regions. Blue, ancestral; red, derived.

with blue eye colour[23]. Moreover, a prediction of eye colour based on genotypes at additional loci using HIrisPlex[24] produced a 0.823 maximal and 0.672 minimal probability for being non-brown-eyed (Supplementary Information). The genotypic combination leading to a predicted phenotype of dark skin and non-brown eyes is unique and no longer present in contemporary European populations. Our results indicate that the adaptive spread of light skin pigmentation alleles was not complete in some European populations by the Mesolithic, and that the spread of alleles associated with light/blue eye colour may have preceded changes in skin pigmentation.

For the remaining loci, La Braña 1 displayed the derived, putatively adaptive variants in five cases, including three genes, *PTX4, UHRF1BP1* and *GPATCH1 (*ref. 19), involved in the immune system (Table 1 and Extended Data Fig. 8). *GPATCH1* is associated with the risk of bacterial infection. We subsequently determined the allelic states in 63 SNPs from 40 immunity genes with previous evidence for positive selection and for carrying polymorphisms shown to influence susceptibility to infections in modern Europeans (Supplementary Information). La Braña 1 carries derived alleles in 24 genes (60%) that have a wide range of functions in the immune system: pattern recognition receptors, intracellular adaptor molecules, intracellular modulators, cytokines and cytokine receptors, chemokines and chemokine receptors and effector molecules. Interestingly, four out of six SNPs from the first category are intracellular receptors of viral nucleic acids (*TLR3*, *TLR8*, *IFIH1* (also known as *MDA5*) and *LGP2*)[25].

Finally, to explore the functional regulation of the genome, we also assessed the La Braña 1 genotype at all expression quantitative trait loci (eQTL) regions associated to positive selection in Europeans (Supplementary Information). The most interesting finding is arguably the predicted overexpression of eight immunity genes (36% of those with described eQTLs), including three Toll-like receptor genes (*TLR1*, *TLR2* and *TLR4*) involved in pathogen recognition[26].

These observations suggest that the Neolithic transition did not drive all cases of adaptive innovation on immunity genes found in modern Europeans. Several of the derived haplotypes seen at high frequency today in extant Europeans were already present during the Mesolithic, as neutral standing variation or due to selection predating the Neolithic. *De novo* mutations that increased in frequency rapidly in response to zoonotic infections during the transition to farming should be identified among those genes where La Braña 1 carries ancestral alleles.

To confirm whether the genomic traits seen at La Braña 1 can be generalized to other Mesolithic populations, analyses of additional ancient genomes from central and northern Europe will be needed. Nevertheless, this genome sequence provides the first insight as to how these hunter-gatherers are related to contemporary Europeans and other ancient peoples in both Europe and Asia, and shows how ancient DNA can shed light on the timing and nature of recent positive selection.

METHODS SUMMARY

DNA was extracted from the La Braña 1 tooth specimen with a previously published protocol[2]. Indexed libraries were built from the ancient extract and sequenced on the Illumina HiSeq platform. Reads generated were mapped with BWA[27] to the human reference genome (NCBI 37, hg19) after primer trimming. A metagenomic analysis and taxonomic identification was generated with the remaining reads using BLAST 2.2.27+ and MEGAN4 (ref. 28) (Extended Data Fig. 9). SNP calling was undertaken using a specific bioinformatic pipeline designed to account for ancient DNA errors. Specifically, the quality of misincorporations likely caused by ancient DNA damage was rescaled using the mapDamage2.0 software[29], and a set of variants with a minimum read depth of 4 was produced with GATK[30]. Analyses including PCA[10], Outgroup f_3[16] and D statistics[17] were performed to determine the population affinities of this Mesolithic individual (Supplementary Information).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 October; accepted 17 December 2013.

Published online 26 January 2014.

1. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature Commun.* **3,** 698 (2012).
2. Sánchez-Quinto, F. *et al.* Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Curr. Biol.* **22,** 1494–1499 (2012).
3. Skoglund, P. *et al.* Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336,** 466–469 (2012).
4. Laland, K. N., Odling-Smee, J. & Myles, S. How culture shaped the human genome: bringing genetics and the human sciences together. *Nature Rev. Genet.* **11,** 137–148 (2010).
5. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463,** 757–762 (2010).
6. Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334,** 94–98 (2011).
7. Vidal Encinas, J. M. & Prada Marcos, M. E. *Los hombres mesolíticos de La Braña-Arintero (Valdelugueros, León)* (León: Junta de Castilla y León, 2010).
8. Overballe-Petersen, S., Orlando, L. & Willerslev, E. Next-generation sequencing offers new insights into DNA degradation. *Trends Biotechnol.* **30,** 364–368 (2012).
9. Allentoft, M. E. *et al.* The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. B Biol. Sci.* **279,** 4824–4733 (2012).
10. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
11. Nelson, M. R. *et al.* The population reference sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am. J. Hum. Genet.* **83,** 347–358 (2008).
12. Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456,** 98–101 (2008).
13. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
14. Surakka, I. *et al.* Founder population-specific HapMap panel increases power in GWA studies through improved imputation accuracy and CNV tagging. *Genome Res.* **20,** 1344–1351 (2010).
15. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505,** 87–91 (2014).
16. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461,** 489–494 (2009).
17. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328,** 710–722 (2010).
18. Perry, G. H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nature Genet.* **39,** 1256–1260 (2007).
19. Grossman, S. R. *et al.* Identifying recent adaptations in large-scale genomic data. *Cell* **152,** 703–713 (2013).
20. Lamason, R. L. *et al.* SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310,** 1782–1786 (2005).
21. Norton, H. L. *et al.* Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* **24,** 710–722 (2007).
22. Sturm, R. A. & Duffy, D. L. Human pigmentation genes under environmental selection. *Genome Biol.* **13,** 248 (2012).
23. Sturm, R. A. *et al.* A single SNP in an evolutionary conserved region within intron 86 of the *HERC2* gene determines human blue-brown eye color. *Am. J. Hum. Genet.* **82,** 424–431 (2008).
24. Walsh, S. *et al.* The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet.* **7,** 98–115 (2013).
25. Aoshi, T., Koyama, S., Kobiyama, K., Akira, S. & Ishii, K. J. Innate and adaptive immune responses to viral infection and vaccination. *Curr. Opin. Virol.* **1,** 226–232 (2011).
26. Moresco, E. M. Y., LaVine, D. & Beutler, B. Toll-like receptors. *Curr. Biol.* **21,** R488–R493 (2011).
27. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
28. Huson, D. H., Mitra, S., Ruscheweyh, H.-J., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* **21,** 1552–1560 (2011).
29. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29,** 1682–1684 (2013).
30. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank L. A. Grau Lobo (Museo de León) for access to the La Braña specimen, M. Rasmussen and H. Schroeder for valid input into the experimental work, and M. Raghavan for early access to Mal'ta genome data. Sequencing was performed at the Danish National High-Throughput DNA-Sequencing Centre, University of Copenhagen. The POPRES data were obtained from dbGaP (accession number 2038). The authors are grateful for financial support from the Danish National Research Foundation, ERC Starting Grant (260372) to TM-B, and (310372) to M.G.N., FEDER and Spanish Government Grants BFU2012-38236, the Spanish Multiple Sclerosis Netowrk (REEM) of the Instituto de Salud Carlos III (RD12/0032/0011) to A.N., BFU2011-28549 to T.M.-B., BFU2012-34157 to C.L.-F., ERC (Marie Curie Actions 300554) to M.E.A., NIH NRSA postdoctoral fellowship (F32GM106656) to C.W.K.C., NIH (R01-HG007089) to J.N., NSF postdoctoral fellowship (DBI-1103639) to M.D., the Australian NHMRC to R.A.S. and a predoctoral fellowship from the Basque Government (DEUI) to I.O.

Author Contributions C.L.-F. and E.W. conceived and lead the project. M.E.P. and J.M.V.E. provided anthropological and archaeological information. O.R. and M.E.A. performed the ancient extractions and library construction, respectively. I.O., M.E.A., F.S.-Q., J.P.-M., S.R., O.R., M.F.-C. and T.M.-B. performed mapping, SNP calling, mtDNA assembly, contamination estimates and different genomic analyses on the ancient genome. I.O., F.S.-Q., G.S., C.W.K.C., M.D., J.A.R., J.Q., O.R., U.M.M. and A.N. performed functional, ancestry and population genetic analyses. R.N. and J.N. coordinated the ancestry analyses. M.G.N., R.A.S. and P.S. coordinated the immunological, pigmentation and selection analyses, respectively. I.O., M.E.A., T.M.-B., E.W. and C.L.-F. wrote the majority of the manuscript with critical input from R.N., M.G.N., J.N., R.A.S., P.S. and A.N.

Author Information Alignment data are available through the Sequence Read Archive (SRA) under accession numbers PRJNA230689 and SRP033596. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.W. (ewillerslev@snm.ku.dk) or C.L.-F. (carles.lalueza@upf.edu).

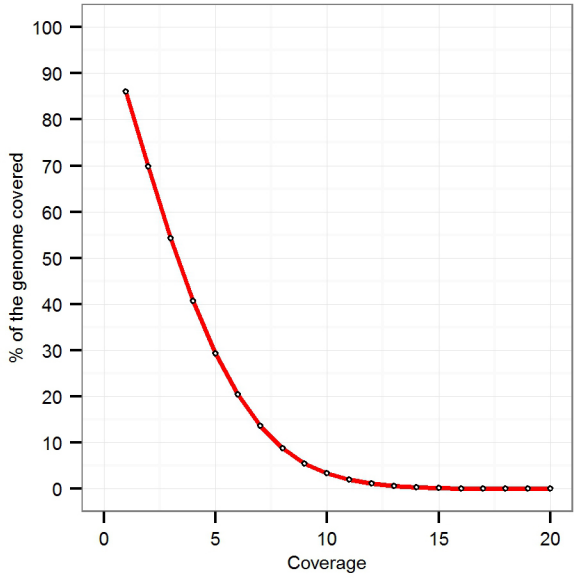
a

	La Braña 1
Raw sequences	2,158,245,715
After adapter removal	2,058,756,210
Mapped	1,045,510,409
Duplicates	827,618,790
Unique	217,891,619
MAPQ>25	138,514,623
Mean read length	66
Missmatch rate	0.0071
Indel rate	0.00017
Strand balance	0.50

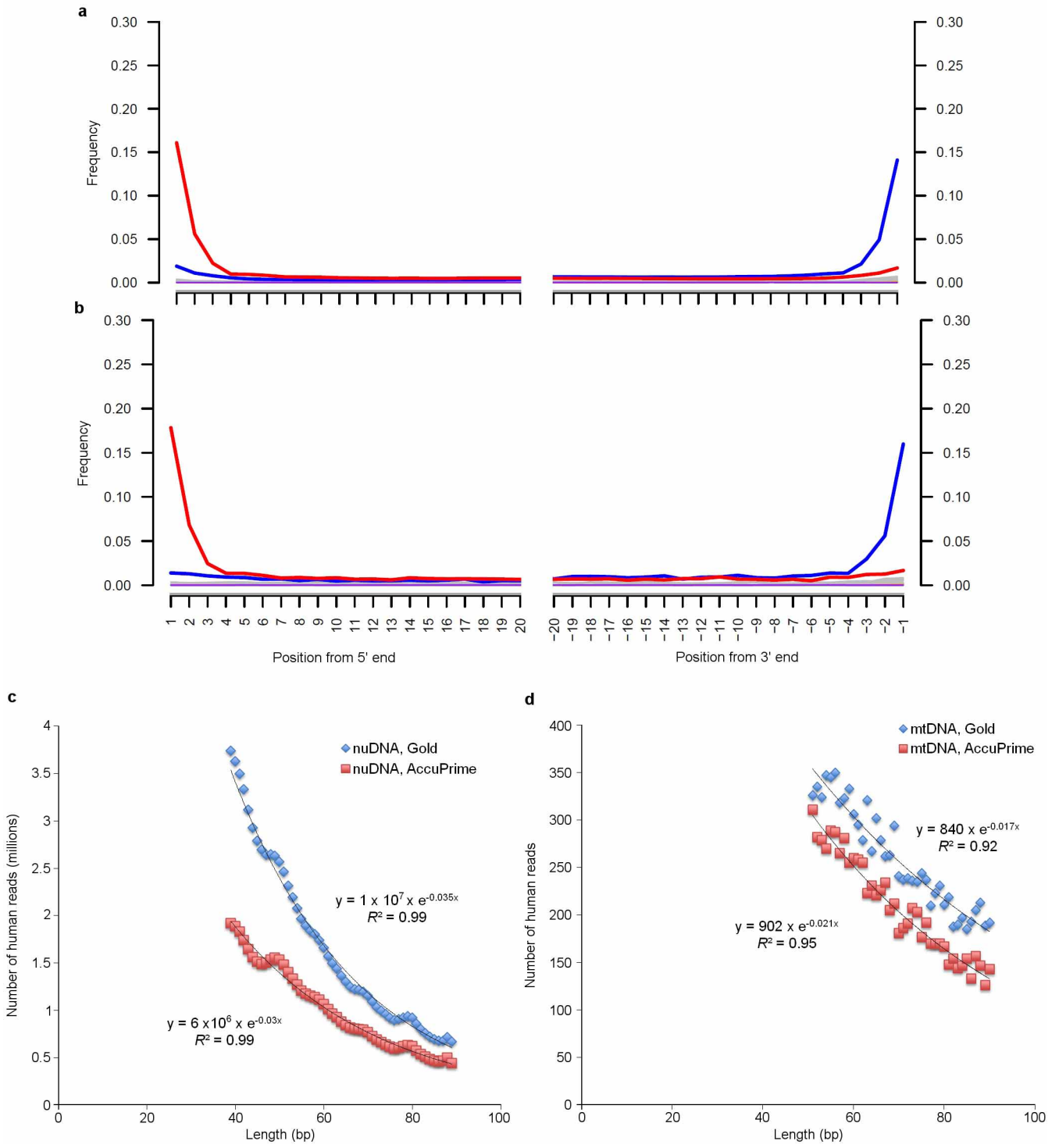
b

Chromosomes	% covered	Mean read depth	Mean read depth (only in covered positions)
1	88.26	3.68	4.17
2	86.57	3.37	3.89
3	86.09	3.24	3.77
4	82.99	2.91	3.5
5	85.66	3.22	3.75
6	85.68	3.23	3.77
7	86.33	3.38	3.92
8	86.52	3.34	3.86
9	87.92	3.61	4.1
10	88.5	3.66	4.14
11	88.45	3.68	4.16
12	87.18	3.45	3.95
13	83.37	2.98	3.58
14	87.13	3.47	3.99
15	89.34	3.76	4.21
16	91.6	4.28	4.67
17	92.03	4.45	4.84
18	86.09	3.28	3.8
19	94.41	5.09	5.39
20	91.98	4.22	4.59
21	85.76	3.53	4.11
22	94.05	4.98	5.29
X	69.36	1.57	2.26
Y	60.39	1.42	2.35
Total	86.17	3.4	3.94

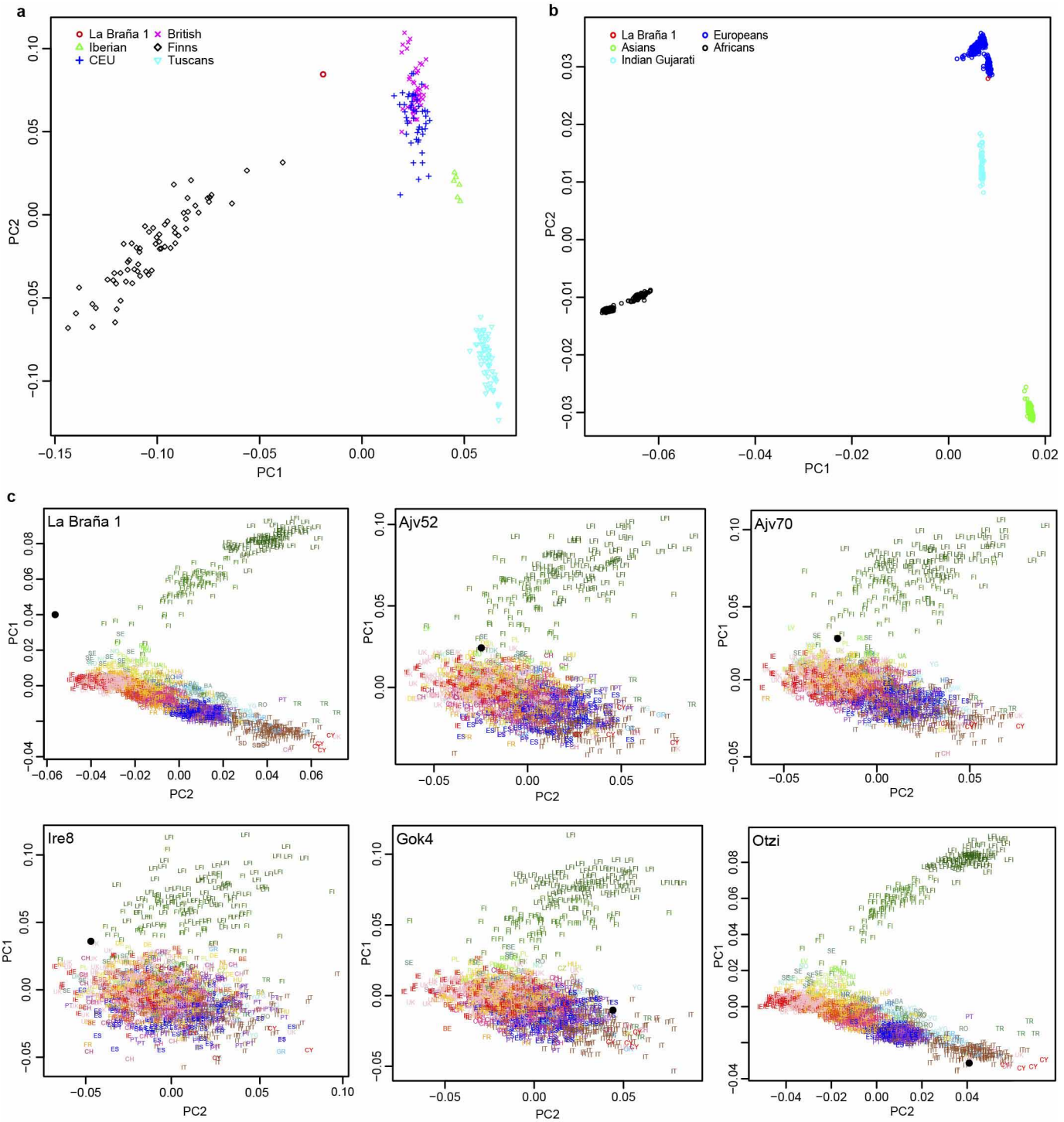
c



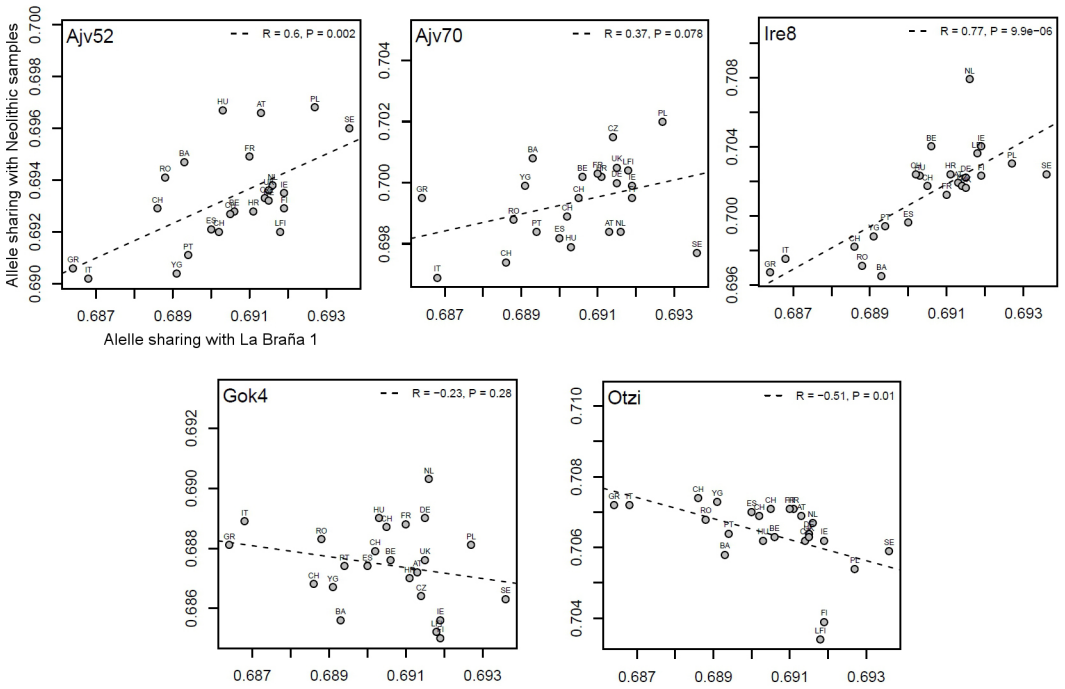
Extended Data Figure 1 | Alignment and coverage statistics of the La Braña 1 genome. **a**, Alignment summary of the La Braña 1 sequence data to hg19 assembly. **b**, Coverage statistics per chromosome. The percentage of the chromosome covered by at least one read is shown, as well as the mean read depth of all positions and positions covered by at least one read. **c**, Percentage of the genome covered at different minimum read depths.



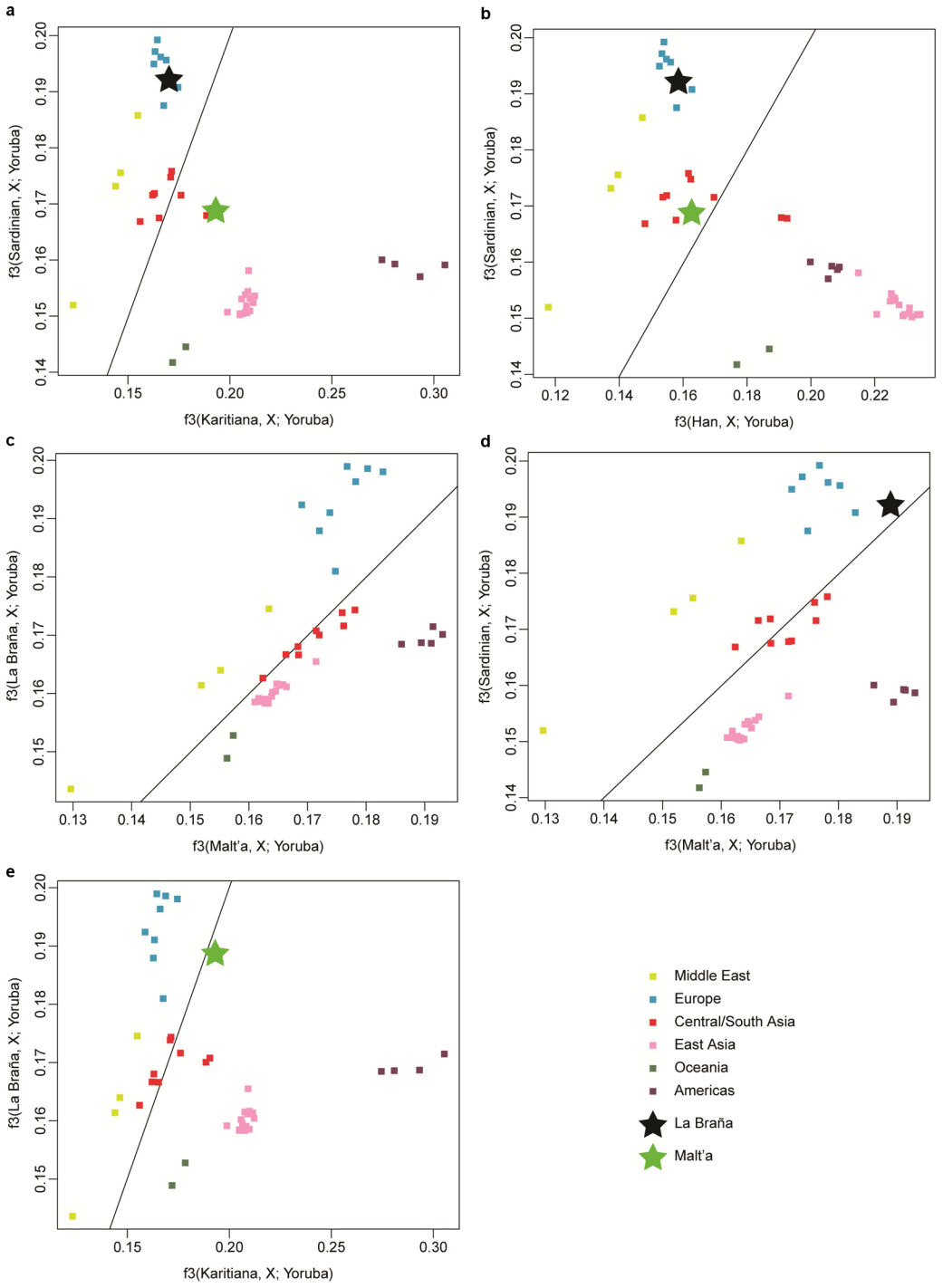
Extended Data Figure 2 | Damage pattern of La Braña 1 sequenced reads. **a**, **b**, Frequencies of C to T (red) and G to A (blue) misincorporations at the 5′ end (left) and 3′ end (right) are shown for the nuclear DNA (nuDNA) (**a**) and mtDNA (**b**). **c**, **d**, Fragment length distribution of reads mapping to the nuclear genome (**c**) and mtDNA genome (**d**). Coefficients of determination (R^2) for an exponential decline are provided for the four different data sets. The exponential coefficients for the four data sets correspond to the damage fraction (λ); e is the base of the natural logarithm.



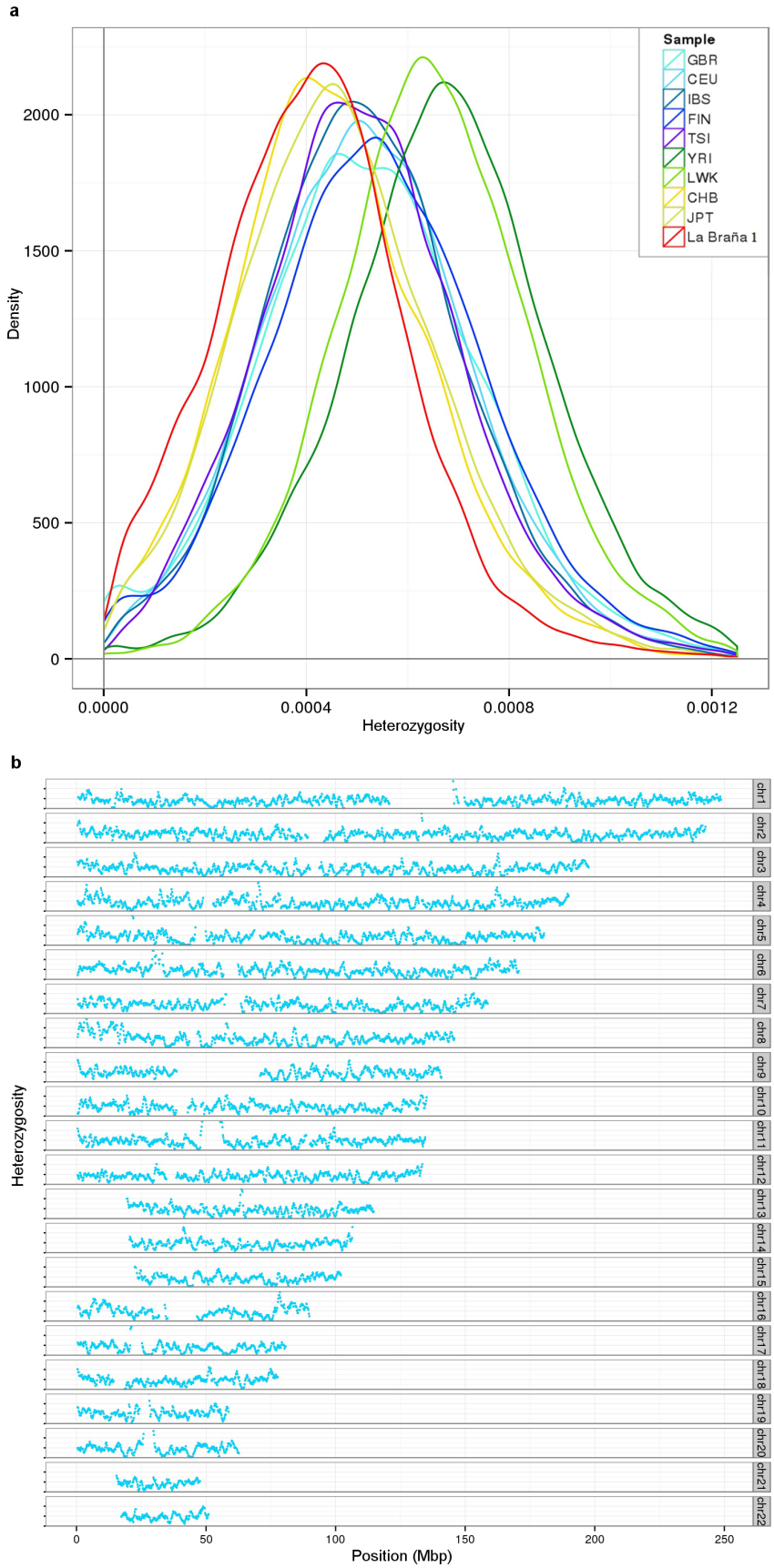
Extended Data Figure 3 | Genetic affinities of the La Braña 1 genome. **a**, PCA of the La Braña 1 SNP data and the 1000 Genomes Project European individuals. **b**, PCA of La Braña 1 versus world-wide data genotyped with the Illumina Omni 2.5M array. Continental terms make reference to each Omni population grouping as follows: Africans, Yoruba and Luyha; Asians, Chinese (Beijing, Denver, South, Dai), Japanese and Vietnamese; Europeans, Iberians, Tuscans, British, Finns and CEU; and Indian Gujarati from Texas. **c**, Each panel shows PC1 and PC2 based on the PCA of one of the ancient samples with the merged POPRES+FINHM sample, before Procrustes transformation. The ancient samples include the La Braña 1 sample and four Neolithic samples from refs 1 and 3.



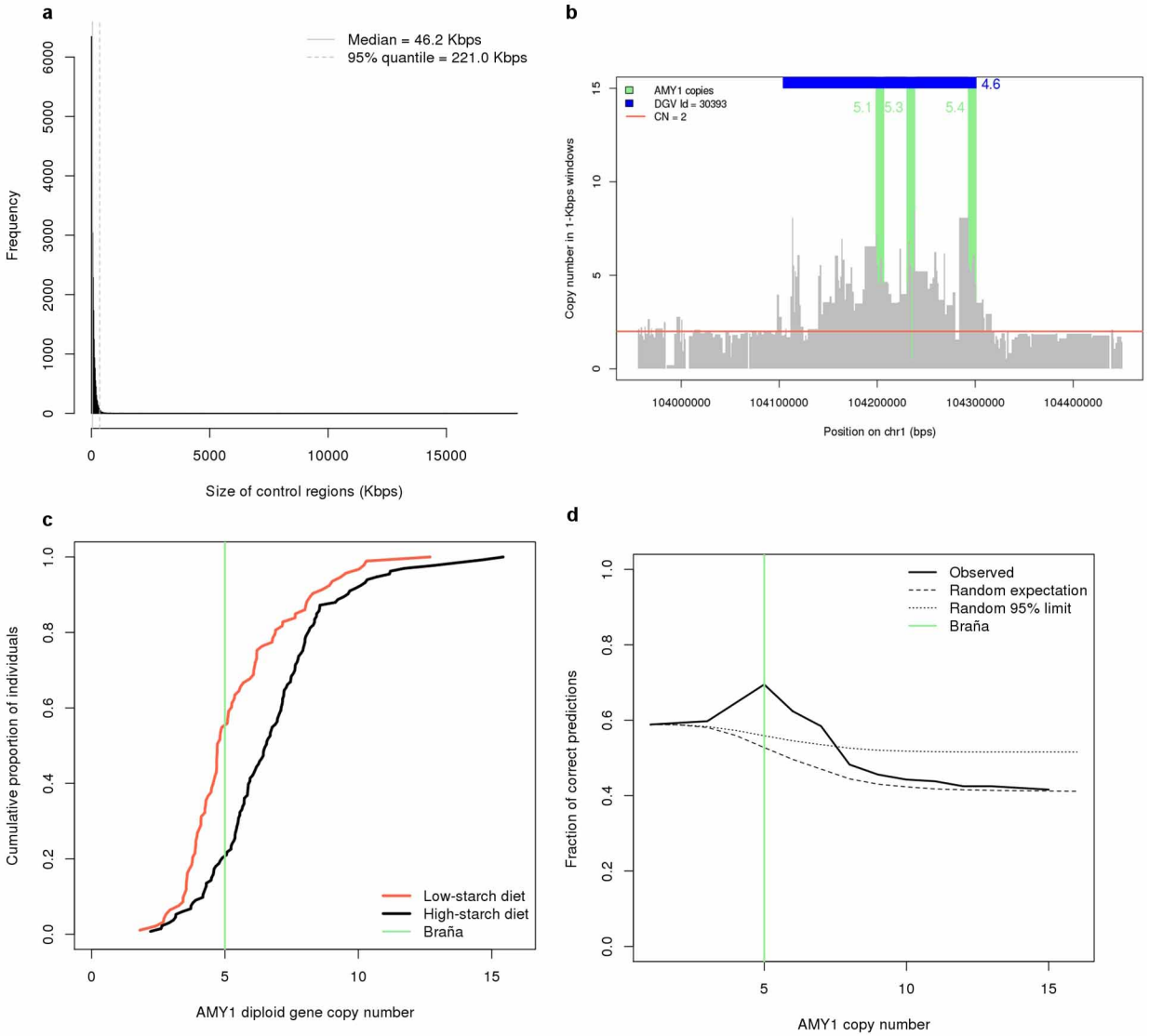
Extended Data Figure 4 | Allele-sharing analysis. Each panel shows the allele-sharing of a particular Neolithic sample from refs 1 and 3 with La Braña 1 sample. The sample IDs are presented in the upper left of each panel (Ajv52, Ajv70, Ire8, Gok4 and Ötzi). In the upper right of each panel, the Pearson's correlation coefficient is given with the associated *P* value.



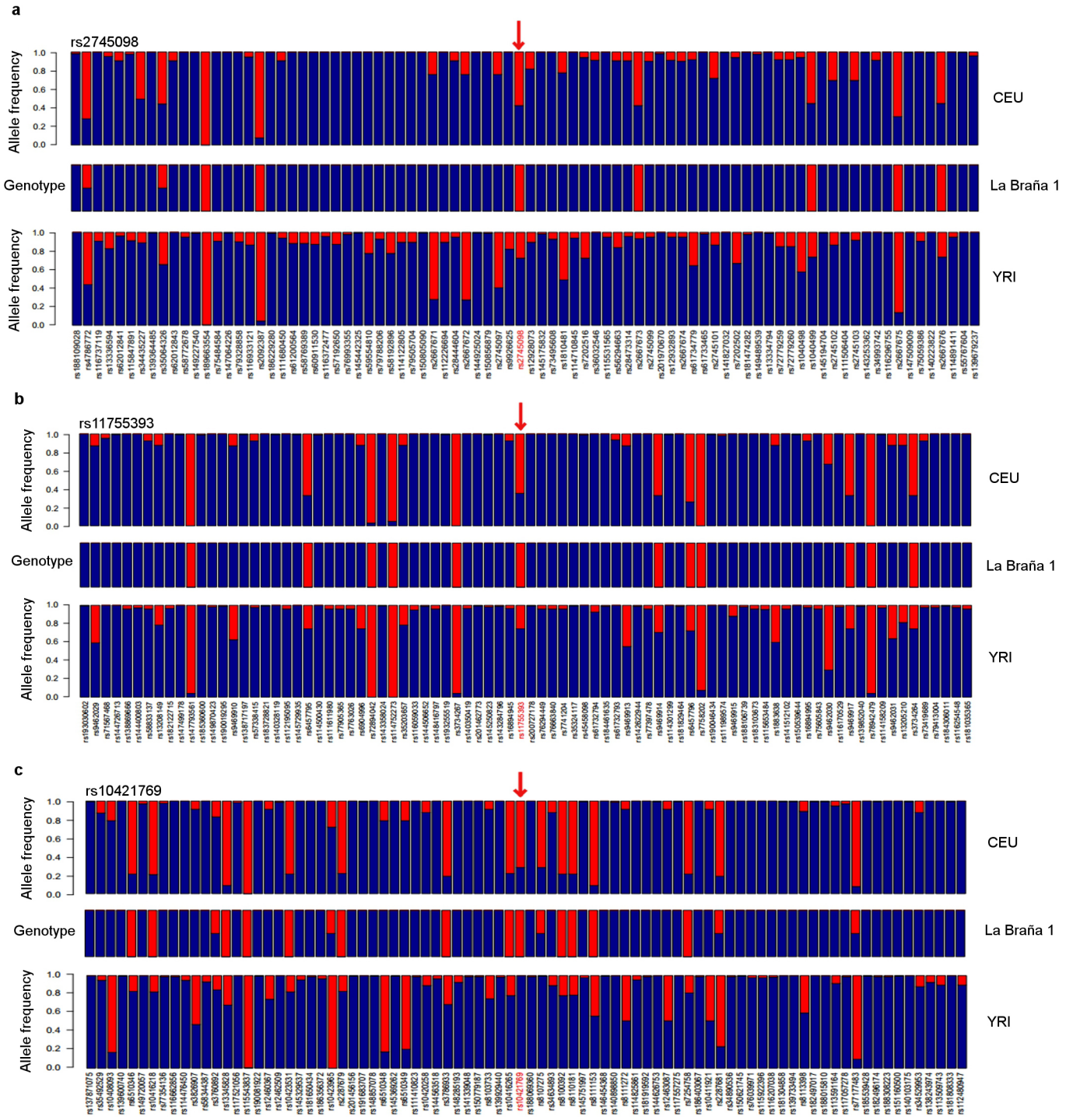
Extended Data Figure 5 | Pairwise outgroup f_3 statistics. **a**, Sardinian versus Karitiana. **b**, Sardinian versus Han. **c**, La Braña 1 versus Mal'ta. **d**, Sardinian versus Mal'ta. **e**, La Braña 1 versus Karitiana. The solid line represents $y = x$.



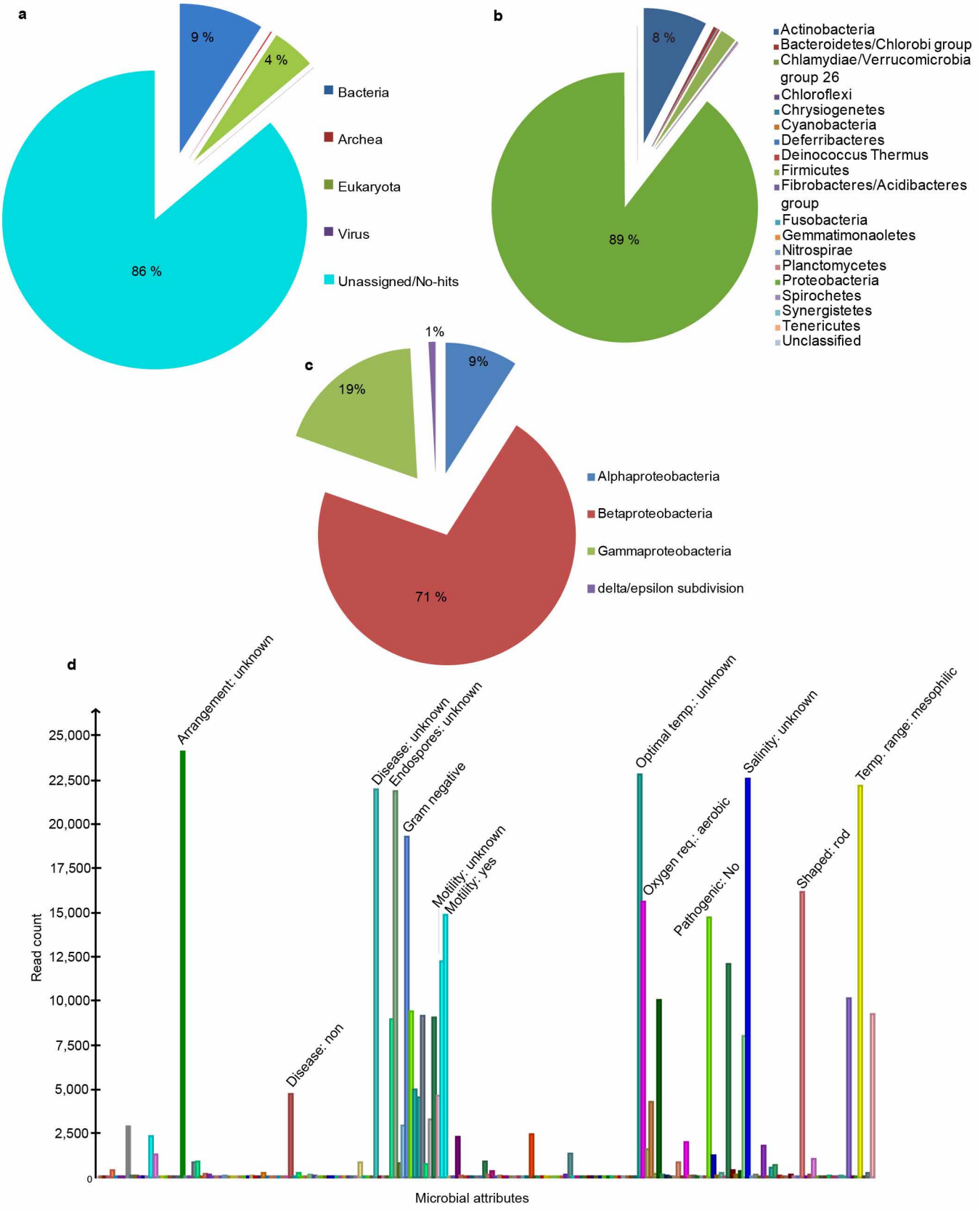
Extended Data Figure 6 | Analysis of heterozygosity. **a**, Heterozygosity distributions of La Braña 1 and modern individuals with similar coverage from the 1000 Genomes Project (using 1-Mb windows with 200 kb overlap). CEU, northern- and western-European ancestry. CHB, Han Chinese; FIN, Finns; GBR, Great Britain; IBS, Iberians; JPT, Japanese; LWK, Luhya; TSI, Tuscans; YRI, Yorubans. **b**, Heterozygosity values in 1-Mb windows (with 200 kb overlap) across each chromosome.



Extended Data Figure 7 | Amylase copy-number analysis. **a**, Size distribution of diploid control regions. **b**, *AMY1* gene copy number in La Braña 1. CN, copy number; DGV, Database of Genomic Variation. **c**, La Braña 1 *AMY1* gene copy number in the context of low- and high-starch diet populations. **d**, Classification of low- and high-starch diet individuals based on *AMY1* copy number. Using data from ref. 18, individuals were classified as in low-starch (less or equal than) or high-starch (higher than) categories and the fraction of correct predictions was calculated. In addition, we calculated the random expectation and 95% limit of low-starch-diet individuals classified correctly at each threshold value.



Extended Data Figure 8 | Neighbouring variants for three diagnostic SNPs related to immunity. **a**, rs2745098 (*PTX4* gene). **b**, rs11755393 (*UHRF1BP1* gene). **c**, rs10421769 (*GPATCH1* gene). For *PTX4*, *UHRF1BP1* and *GPATCH1*, La Braña 1 displays the derived allele and the European-specific haplotype, indicating that the positive-selection event was already present in the Mesolithic. Blue, ancestral; red, derived.



Extended Data Figure 9 | Metagenomic analysis of the non-human reads. **a**, Domain attribution of the reads that did not map to hg19. **b**, Proportion of different Bacteria groups. **c**, Proportion of different types of Proteobacteria. **d,** Microbial attributes of the microbes present in the La Braña 1 sample.

Supporting Online Material for

Derived Immune and Ancestral Pigmentation Alleles in a 7,000-Year-old Mesolithic European

Iñigo Olalde, Morten E. Allentoft, Federico Sánchez-Quinto, Gabriel Santpere, Charleston W. K. Chiang, Michael DeGiorgio, Javier Prado-Martinez, Juan Antonio Rodríguez, Simon Rasmussen, Javier Quilez, Oscar Ramírez, Urko M. Marigorta, Marcos Fernández, María Encina Prada, Julio Manuel Vidal Encinas, Rasmus Nielsen, Mihai G. Netea, John Novembre, Richard A. Sturm, Pardis Sabeti, Tomàs Marquès-Bonet, Arcadi Navarro, Eske Willerslev, Carles Lalueza-Fox

This file includes:

Contents

1-Archaeological context

1.1-The La Braña-Arintero site

The accidental finding of two skeletons at the site of La Braña-Arintero by spelunkers was published in a local newspaper in November 2006. As a testament to their great antiquity, the bones were found partly incorporated in stalagmites. Realizing the importance of the finding, the Servicio Territorial de Cultura of the autonomous government Junta de Castilla y León commissioned the excavation of the site to the archaeologist Julio Manuel Vidal Encinas.

La Braña-Arintero is a small karstic cave located in the municipal area of the village of Valdelugueros in the province of León (Spain). Access through the original cave opening is no longer practicable, but the present-day entrance to the cave is found at 1,489 m ASL in the Cantabrian Mountain range. It is close to a natural crossing (pass of Vegarada, at 1,567 m ASL) that connects the Cantabrian coast with the river Duero basin. The entrance consist of a narrow gallery extending downwards for about 20 meters until an almost vertical pit of four meters; after climbing down, there is another narrow gallery that ends in a ledge. The first skeleton (La Braña 1) was found here, arranged in a fetal position as a primary burial (Figure 1b). The second one (La Braña 2) was discovered a few meters away, at the bottom of a second 4-meters deep pit that had collapsed in the past. Traces of ochre and fragments of charred wood were found around both skeletons. In addition, La Braña 2 was found with 24 atrophic red deer canines. These are typical Mesolithic ornaments described from many other sites, especially from Central Europe[7]. The presence of these people in a mountainous environment could be associated with the progressive climatic 'improvement' after the Last Glacial Maximum and the specialised hunting of local chamois and mountain sheep.

Both individuals are relatively young males (30-35 years of age) of about 1.66-1.70 cm of height. Their close proximity in the cave and the AMS dates (see below) suggest they could be contemporaneous. The causes of death cannot be ascertained from the examination of the bones.

The area has a strong continental climate, with mild summers and very cold winters. The average min and max temperatures during winter are -4.3ºC and 5.5ºC respectively, and the average yearly temperature is 8.1ºC. It seems very likely that these

low temperatures, as well as the thermal stability favored by the deep location in the cave, have contributed to the observed remarkable DNA preservation. Because a previous genetic study established that La Braña 1 was be better preserved than La Braña 2[2] (the latter was likely submerged during floods of the cave), it was decided to focus exclusively on La Braña 1 for additional sampling and DNA sequencing.

1.2-Dating

Both skeletons were AMS radiocarbon-dated at Beta Analytic Radiocarbon Dating Laboratory (Miami, USA), yielding similar dates of 6,980±50 BP for La Braña 1 and 7,030±50 BP for La Braña 2 (both uncalibrated). Calibrated dates place the remains at 7,940-7,690 BP and 7,960-7,750 BP (2 sigma), respectively (using INTCAL 04).

2-Sampling and DNA extraction

Three samples (dental roots from the upper third molars and one femur cortical sample) of about 50 mg each were obtained at the Museo de León, and extracted in dedicated ancient DNA facilities at the Institute of Evolutionary Biology in Barcelona. We followed an extraction protocol previously described[2]. In brief, it is based on overnight incubation at 50 °C in 10 ml of lysis solution (0.5% SDS, 50 mM TRIS, 0.5M EDTA, 1 mg/mL of proteinase K in H_2O). The DNA is subsequently extracted in equal volumes of three steps of phenol–chloroform/isoamyl alcohol. The resulting aqueous phase is concentrated using Amicon filter columns, followed by purification with a silica extraction (Fermentas) and eluted in sterile water to a final volume of 50 µl.

3-Library preparation and sequencing

Two blunt-end DNA libraries were prepared using the NEBNext DNA Sample Prep Master Mix Set 2 (E6070) and Illumina specific adapters. The libraries were built according to a previously published protocol[31]. The initial nebulization step was skipped due to the fragmented nature of ancient DNA. End-repair was performed in 25 µl

reactions using 20 µl of DNA extract. This was incubated for 20 mins at 12°C and 15 mins at 37°C, and purified using PN buffer with Qiagen MinElute spin columns, and eluted in 15 µl. Following end-repair, Illumina-specific adapters (prepared as in[32]) were ligated to the end-repaired DNA in 25 µl reactions. The reaction was incubated for 15 mins at 20°C and purified with PB buffer on Qiagen MinElute columns, before eluted in 20 µl EB Buffer. The adapter fill-in reaction was performed in a final volume of 25 µl and incubated for 20 mins at 37°C followed by 20 mins at 80°C to inactivate the Bst enzyme. The entire DNA library (25 µl) was then amplified and indexed in a 50 µl PCR reaction.

To accommodate differences in efficiency and sequence bias introduced by polymerases in the amplification step[33], two libraries were built; one with AmpliTaq Gold DNA Polymerase, and one with AccuPrime Pfx polymerase (both Invitrogen). The libraries were amplified in a two-step manner, where 5 µl PCR product from the first amplification round was transferred into new 50 µl PCR reactions. To increase complexity, the second-round PCR's were set-up with 4 reactions in parallel. Following purification on Qiagen QiaQuick spin columns, these were eluted in 30 µl EB buffer, and pooled, yielding a total of 120 µl amplified and purified library, for each of the two polymerases.

The AccuPrime protocol was as follows:

10 µl AccuPrime reaction mix, 2 µl BSA (20mg/ml), 1 µl AccuPrime Taq, 3% DMSO, 2 µl index primer (10uM), 2µl inPE1.0 primer (10 µM), 25 µl library DNA, and H_2O to 50 µl. The thermocycling profile for the first-round-PCR was 4 mins at 95°C, followed by 12 cycles of 30s at 95°C, 30s at 60°C and 40s at 68°C, and a final elongation step at 7 mins for 68°C. For the second-round PCR, 6 cycles was used and the primers were replaced by IS5 and IS6[32].

The AmpliTaq Gold protocol was as follows:

5 µl Buffer Gold, 4 µl $MgCl_2$ (25mM), 2 µl BSA (20mg/ml), 1 µl AmpliTaq Gold , 3% DMSO, 0.5 µl dNTPs (25 mM), 2 µl index primer (10uM), 2µl inPE1.0 primer (10 µM), 25 µl library DNA, and H_2O to 50 µl. The thermocycling profile for the first-round-PCR was 4 mins at 95°C, followed by 12 cycles of 30s at 95°C, 30s at 60°C and 40s at 72°C, and a final elongation step at 7 mins for 72°C. For the second-round PCR, 11 cycles was used and the primers were replaced by IS5 and IS6[32].

The libraries were profiled on a Bioanalyzer (Agilent 2100 Bioanalyzer High Sensitivity DNA chip), pooled equimolarly, and sequenced (single read, 100 base pairs

(bp)) multiplexed in one lane to estimate the efficiency (ratio of human DNA versus environmental contaminant DNA) in each sample. The femur bone sample yielded an efficiency of 0.05%, the first tooth sample (an upper right third molar root) showed 1.7%, and the second (an upper left third molar root) a remarkably high efficiency of 48.4% (see details of mapping etc below). At present it is not possible to explain the discrepancies in the efficiency among the three samples from La Braña 1. They can be attributed to stochastic conservation processes within the same environment. It was decided to deep-sequence the high-efficiency sample. Nine lanes of data were then generated on an Illumina HiSeq 2000 platform, at the National High-throughput DNA Sequencing Centre in Copenhagen (http://seqcenter.ku.dk/).

4-Raw reads processing

We generated a total of 2,158,245,715 reads of 94 nucleotides. To process these raw data, we first trimmed stretches of N´s from the 3' and 5' ends of the reads. Stretches of consecutive bases with 0, 1, or 2 quality scores were identified and removed from both ends of the reads. Reads shorter than 30 nucleotides were discarded for further analyses. After these quality filters, we kept 2,154,010,213 reads with an average length of 93.4 bp.

Owing to *post-mortem* degradation, ancient DNA (aDNA) fragments are usually very short, resulting in sequencing of the adapter which has been ligated during library preparation. The presence of parts of the adapter in the sequencing reads can affect the correct mapping to the reference genome, and it can also bias the SNP calling. Therefore, we used AdapterRemoval[34] to remove adapter sequences from the reads, and we discarded sequences shorter than 30 bp after adapter trimming. We found 83% of the reads containing adapter sequence, finally keeping a total of 2,058,756,210 reads with an average length of 61 bp, which were aligned to the human reference genome (see below).

4.1-Mapping

We mapped all the sequences to the human reference genome (GRCh37/hg19) with the mitochondrion replaced by the revised Cambridge reference sequence

(rCRS)[35]. We used BWA[27] with the quality trimming parameter set to a Sanger quality score of 15. Furthermore, following recommendations in[36], the edit distance parameter was set to 0.02 and the seed region was disabled. With these criteria, a total of 1,045,510,409 reads (48.4%) could be aligned to the reference genome. The library constructed with AmpliTaq Gold yielded a higher ratio of mapped reads (55.7%) than the one constructed with AccuPrime (44.4%). Next, we used Picard tools (http://picard.sourceforge.net/) to remove PCR duplicates and performed indel realignment followed by base quality recalibration (using known polymorphic sites in dbSNP v137), both with GATK[30,37]. We found a duplicity of 79%, yielding a total of 217,891,619 unique reads.

Finally, we filtered out reads with mapping quality less than 25 and ambiguously mapped reads. This final set of 138,514,623 reads with an average length of 66 bp was used for the SNP calling (Extended Data Figure 1a). In order to calculate the average depth of coverage of this final set, we used the DepthOfCoverage tool implemented in GATK[30,37]. To avoid low complexity regions, we restricted the analysis to unique regions in hg19 assembly, downloaded from UCSC Genome Browser (http://genome.ucsc.edu). The proportion of the genome covered at different read depths can be seen in Extended Data Figure 1c. Around 86% of the genome was covered by at least one read. The average depth of coverage was 3.40x, and 3.94x when restricted to positions covered at least once. Extended Data Figure 1b shows the results for each of the chromosomes. In autosomes, the mean read depth in covered positions ranged from 3.50x on chromosome 4, to 5.39x on chromosome 19. Sex chromosomes had a mean read depth half of that of the autosomes, consistent with a haploid state (XY).

4.2-SNP calling

To ensure that DNA damage would not contribute to downstream analyses, we used the rescale parameter implemented in the mapDamage2.0 software[29]. Taking into account initial qualities, positions in the reads and damage patterns, the software re-scales the quality of misincorporations likely caused by aDNA damage. Using the GATK Unified Genotyper, we produced a preliminary set of variants (1,063,187 SNPs) with a minimum read depth of 4. Unified Genotyper was used with default parameters, except for the *stand_call_conf* parameter that was set to 20 in order to allow more

borderline calls. This preliminary set of SNPs was refined using the variant quality score recalibration (VQSR) procedure, also implemented in GATK. This procedure uses a set of training SNPs (typically Hapmap v3.3 and the Omni chip array from the 1,000 Genomes Project (1000G)) to build a Gaussian mixture model of high quality variants, based on different annotations (mapping quality, total depth, strand balance, etc). From the preliminary set, using VQSR, we kept a subset of SNPs such that 99% of the positions in the training set were retained. Table S1 shows the frequencies of all possible nucleotide changes found in the resulting set of variants, comprising 990,167 SNPs. Of these variants, 98.5% overlapped with dbSNP v137, resulting in 14,703 novel SNPs (Table S2). Given that La Braña 1 is a male individual, we filtered out all the heterozygous positions found in the X and Y chromosomes. Heterozygous calls in sexual chromosomes are likely a consequence of sequencing errors, structural variation between the reference and targeted genome, modern human contamination (discussed below) and mapping errors, specially common in the Y-chromosome[5]. This final set of SNPs with a minimum read depth of 4x was used for analyses in sections 5.3, 5.4, 7.1 and 8. Similarly, using the Unified Genotyper, we produced a set of homozygous-reference positions with a minimum read depth of 4.

5-Authenticity

5.1-Patterns of post-mortem damage

In order to determine whether the genome of the Mesolithic individual showed the typical aDNA *post-mortem* degradation pattern, we used the mapDamage2.0 software[29]. This new version of the software extends the original features of mapDamage by incorporating a Bayesian statistical framework of DNA damage that, given some key features of the DNA molecules, calculates the probability that a C to T or a G to A mutation might be true damage misincorporations. We ran the mapDamage2.0 software for both nuclear (Extended Data Figure 2a) and mitochondrial reads (Extended Data Figure 2b), and observed clear ancient DNA patterns, typified by an increase of C to T and G to A changes towards the ends 5' and 3' ends of the reads, respectively, as well as an observed excess of purine bases in the reference genome

before the reads start and after the reads end (data not shown)[38]. The frequency of C to T changes at the first position of mitochondrial reads was 18% and 16% for nuclear data, which is congruent with values from other ancient samples of similar age[39].

5.2-Rate of DNA decay

It is well known from ancient DNA studies that there is a negative exponential correlation between the number of DNA molecules present in a sample and their length[40,41]. This pattern is a reflection of *post-mortem* DNA fragmentation (facilitated mainly by depurination), occurring at random places across the genome, leaving few long molecules and many short ones. If the DNA in the La Braña 1 sample is ancient, it should conform to this relationship, and we therefore examined the lengths of all the reads that mapped to the human reference genome. Following Allentoft et al.[9], we only examined the declining part of the distribution knowing that the decreasing number of shortest fragments is a result of these being lost in the DNA extraction.

In the analysis, we distinguished between mitochondrial DNA (mtDNA) and nuclear DNA (nuDNA), and further, between the two polymerases used in the library build (Taq Gold and AccuPrime). In all four cases, the plots displayed an exponential correlation (R^2 = 0.92-0.99) (Extended Data Figure 2 c, d), providing support for the authenticity of La Braña 1 DNA.

As demonstrated in Deagle et al.[42], the decay constant (λ) in this relationship represents the damage fraction (i.e. the fraction of broken bonds in the DNA backbone). Solving the equations for the four functions, we find λ-values of 3% and 3.5% for nuDNA, and of 1.7% and 2.1% for mtDNA (Extended Data Figure 2 c, d). This difference corresponds with previous findings that mtDNA breaks at a slower rate than nuDNA[9], possibly due to its circular structure and the double membrane of the mitochondria. Whether a random effect or not, it is clear that the two polymerases generated slightly different fragment length distributions (Extended Data Figure 2 c, d) and hence slightly different λ-values. In the further calculations we therefore used the average values of 0.0325 and 0.019 for nuDNA and mtDNA respectively.

Based on ancient mtDNA in fossil megafauna bones, it was recently demonstrated that long-term *post-mortem* DNA fragmentation can be described as a rate process, and that the damage fraction (λ, per site) can be converted to a decay rate (k,

per site per year), when the age of the sample is known[9]. Given that La Braña 1 was dated to 6980 BP (uncalibrated), we calculated the corresponding DNA decay rates (Table S3).

We see that the molecular half-life ($T_{1/2}$ = ln 2/k) in the La Braña 1 tooth is 1,507 and 2,476 years for 100 bp fragments of nuDNA and mtDNA respectively (Table S3), implying that after this time, half the molecules of this length will be gone due to one or more strand breaks. Further, it can be deduced that with this rate of DNA decay, it will take c. 12,000 years to reach an average mtDNA length of 30 bp (which is arguably the minimum mappable length) and c. 357,000 years for the mtDNA to be completely disintegrated (Table S3). Although the nuDNA degrades faster, these estimates provide optimism of much older sequenceable DNA being retrieved from sites with preservation conditions that are comparable to La Braña-Arintero cave

5.2.1-Predictions

A general model has been proposed for the temperature-dependency of the rate of mtDNA decay in bone[9]:

$$\ln k = 41.2 - 15267.6 * 1/T,$$

where T is the absolute temperature.

Given that the average ambient temperature for La Braña-Arintero cave was measured to 8.1°C, the predicted mtDNA decay rate (k) is 2.8E-6 per site per year. We note that this is not far off from the 2.1E-6 per site per year rate that we estimated for La Braña 1 (Table S3). This shows that it is no longer impossible to produce qualified predictions of the rate of DNA degradation, and hence the theoretical DNA survival limits in fossils. This may prove extremely beneficial to ancient DNA research and related disciplines.

Although in the same order of magnitude, the DNA decay rate at La Braña-Arintero site is slightly faster than the predicted value. This is expected, as the model in[9] was based on quantitative PCR data, and could only account for the long-term and pure chemical DNA degradation. In contrast, decay rates calculated from fragment length distributions (i.e. La Braña 1 data) will also incorporate the signal from the initial *post-mortem*, DNA degradation phase, facilitated by enzymatic activity. Since this initial

decay is likely much faster, the overall rate averaged over time will be faster than predicted from the model.

Clearly, with more data available the current DNA decay model should be updated accordingly and verified across a range of temperatures, preservation environments, and substrates.

5.3-mtDNA contamination estimates

The number of raw reads that aligned to the rCRS was 399,350. After applying the same processing steps described above for the autosomal portion of the genome, we kept 19,265 unique mtDNA reads with an average length of 78 bp, representing a depth of coverage of 91x for the mitochocondrial genome. The fact that the mtDNA reads are longer than their autosomal counterparts, confirms the observation in[9] that mitochondrial DNA degrades at a slower rate.

The La Braña 1 mtDNA haplogroup was previously classified as U5b2c1[2]. Thus, we called an mtDNA consensus sequence and checked whether it could be assigned to this haplogroup again, using HaploGrep[43] and Phylotree version 15 (http://www.phylotree.org;[44]). The consensus sequence displayed all the mutations expected for haplogroup U5b2c1.

Next, to assess the presence of putative modern contamination in the sequenced reads, we calculated the proportion of sequences not matching the La Braña 1 consensus base in every position defining the haplogroup U5b2c1 (Table S4). For the 17 positions analyzed, 1.69% (0.75% - 2.67%, 95% C.I.) of the reads did not match La Braña 1 consensus haplogroup. The percentage of contamination provided here is an upper limit, as some of the discordant reads could be a consequence of heteroplasmic sites, nuclear-mitochondrial sequences (numts), sequencing errors, or DNA damage. In fact, 82% of these discordant reads display either a C to T change, or a G to A change, both characteristic misincorporations of ancient DNA.

5.4-Nuclear DNA contamination estimates

To obtain an estimate of contamination based on nuclear DNA, we used our knowledge that La Braña 1 was a male; thus, no heterozygous positions are expected in

the sex chromosomes. However, this assessment is not straightforward due to regions with high similarity between X and Y chromosomes and due to presence of many repetitive regions, especially in the Y chromosome. We first defined uniquely mapped regions in both chromosomes by removing repeats in the GRCh37/hg19 assembly masked with RepeatMasker and Tandem Repeats Finder (downloaded from the UCSC Genome Browser, http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/). Additionally, we masked k-mers in these two chromosomes that also mapped elsewhere in the genome. Both sex chromosomes were partitioned into 36-bps k-mers (with adjacent k-mers overlapping 5 bps) and these k-mers were aligned to hg19 using mrsFast[45]. The coordinates of k-mers on chromosomes X and Y mapping more than once along the genome, were also removed (Table S5).

Next, we counted the number of SNPs present in those unique regions of the X and Y chromosomes, and displaying a minimum coverage of 4. We also counted the number of homozygous-reference positions in the same regions that met the following criteria:

- Minimum depth of 4
- Overlap with known polymorphic sites (dbSNP v137)

A contamination estimate was obtained by dividing the number of heterozygous positions with the total number of positions. While the small number of positions available in the Y chromosome did not allow us to estimate this figure, the estimate for the X chromosome (a better estimate since it can monitor contamination from both females and males) proved to be very low: 0.31% (Table S6).

6-Diversity analysis

To explore the genome wide distribution of genetic variability in this Mesolithic individual, we looked at the distribution of heterozygosity across the genome in 1 Mb overlapping sliding windows (200 kb overlap). For this analysis, the rescale step of putative misincorporations was not applied, as it would also mask some real C-T and G-A heterozygotes, thus leading to the underestimation of La Braña 1 real heterozigosity. Instead, we removed all the C-T and G-A heterozygous positions with only one T or A allele, respectively, as a portion of them could be a consequence of *post-mortem* damage. For each window, the number of heterozygous positions in regions with

coverage between 4x and 10x was computed, and divided by the number of all callable positions with the same criteria (heterozygous + homozygous) (Extended Data Figure 6b). No evidence of tracts of autozygosity was observed, indicating that inbreeding was not common in these hunter-gatherer populations as opposed to for example the ancient Saqqaq people of Greenland[5,46].

To put the global heterozygosity value from La Braña 1 in context to modern genomes, we performed the same analysis on nine modern individuals with low coverage genomes from the 1000G Project[13] from nine different populations: five Europeans (CEU, GBR, FIN, IBS, and TSI), two Africans (LWK and YRI) and two Asians (CHB and JPT). To minimize differences in the power to call heterozygous positions, we selected individuals with similar coverage values to that of the La Braña 1 genome (~ 3x) (Table S7). As it was done for the La Braña 1 individual, C-T and G-A heterozygous positions with only one T or A allele, respectively, were removed. As expected from previous genetic studies, African samples showed the highest heterozygosity, followed by Europeans and Asians (Extended Data Figure 6a). The overall genomic heterozygosity for La Braña 1 was 0.042%, lower than the values observed in present day Asians (0.046-0.047%), Europeans (0.051-0.054%), and Africans (0.066-0.069%). The effective population size, estimated from heterozygosity patterns, suggests a global reduction in population size of ~20% relative to extant Europeans. (Table S8).

7-Ancestry

7.1-La Braña 1 in the context of modern populations

To investigate the genetic similarities of La Braña 1 with modern European populations we performed several principal component analyses (PCA).

First we compared the genotype calls of the La Braña 1 SNP data (minimum read depth of 4x) against data from European individuals of the 1000G Project[13]. From the 236 European individuals sequenced exclusively with Illumina technology, only SNPs with a MAF >0.5% were considered. Next, sites were pruned using PLINK software [47], with the LD-based SNP pruning option with default parameters (pairwise

genotypic r2 > 0.5 within sliding windows of 50 Kb and overlap of 5 SNPs per sliding window) to retain only informative positions and avoid the possible linkage biases. This produced a final set of 1,282,286 SNPs. Furthermore, we used 10 outlier interactions in EIGENSOFT with an outlier sigma threshold of 6.0 within the first six eigenvectors to identify possible outliers. Five individuals (HG00119, HG00124, HG00186, HG00276 and HG00323) were removed after the first iteration on the first, third, fourth and fifth eigenvectors. Finally the PCA was plotted using Rplot (Extended Data Figure 3a). La Braña 1 individual clusters near, but not exactly within, Northern European populations such as CEU, Finns and Great Britons. Thus, we can rule out that the genetic variants of this ancient individual can be drawn from the distribution of extant Southern European populations such as Iberians or Tuscans.

In addition we compared La Braña 1 with Omni chip data[48] to further ascertain the context of the sample with world-wide populations. In general, the filtering process was identical to the one for the previous PCA (regarding MAF, LD and outlier iterations), with the sole difference that since this is chip data, no individuals had to be filtered due to its sequencing technology. After the filtering process, 1,250 individuals were used, producing a final set of 505,410 SNPs (Extended Data Figure 3b).

7.2-La Braña 1 in the context of other ancient samples

We aimed to compare the genetic ancestry of the La Braña 1 individual with other ancient individuals and a merged dataset of modern individuals, using PCA analysis. This merged dataset of modern individuals consisted of POPRES Europeans (dbGAP accession number 2039[12]) and Finnish HapMap (FINHM) individuals (ftp.fimm.fi/pub/FIN_HAPMAP3/[14]). The POPRES dataset comprises 1,387 unrelated European individuals that was previously curated[12]. The FINHM dataset comprises 162 Finnish individuals, of which 82 were from the Late-Settlement Finnish founder population. Seven FINHM individuals were subsequently identified as population outliers in PCA and thus removed from further analysis. In total, the merged reference dataset of POPRES and FINHM consists of 1,542 individuals with genotypes at 331,168 SNPs genome-wide. We downloaded previously published sequences of three Neolithic hunter-gatherers and two Neolithic farmers[1,3], one of which was Ötzi the Iceman. Due to the very low coverage of some of the ancient genomes (coverage ~0.5%

to 3% of the nuclear genome) there was little overlap across datasets for a joint PCA, so we combined the data using Procrustes transformation, as previously described[3,49].

First, we merged the combined POPRES+FINHM dataset with each of the six ancient genomes separately. Due to the low coverage of the ancient genomes (specially Ire8, Ajv52, Ajv70 and Gok4), in order to account for genotype uncertainties we set the ancient individual as well as each of the reference individuals to be homozygous for a randomly selected allele at each SNP in the genome. We also removed A/T and C/G SNPs, and filtered SNPs in which a T or an A was found in the ancient sample and a C or a G, respectively, were found in the POPRES+FINHM dataset. Finally, we filtered one SNP from each pair of SNPs with $r^2 > 0.2$ in windows of size 100 SNPs and steps of 5 SNPs using PLINK[47]. In total, 160,044, 1,832, 4,094, and 6,659 SNPs were available for analysis for individuals La Braña 1, Ire8, Ajv52, and Ajv70 (hunter-gatherers), respectively. A total of 4,326 and 99,499 SNPs were available for analysis for individuals Gok4 and Ötzi (Neolithic farmers). We performed PCA on each of the POPRES+FINHM+ancient genome dataset using EIGENSTRAT version 4.2[50] (Extended Data Figure 3c), and then transformed PC1 and PC2 from each analysis to match the PC1 and PC2 configuration obtained from the POPRES+FINHM+La Braña 1 dataset. Assuming that the ancient individuals do not heavily influence the loadings of the reference individuals, we then take the average of PC1 and PC2 obtained in each separate analysis to represent the PC values for those individuals, but retained the transformed PC values for the ancient individuals for combined visualization (Figure 1c).

7.3-Alelle sharing analysis

To further investigate the genetic similarities between the La Braña 1 individual and each of the Neolithic individuals from Skoglund et al.[3] and Keller et al.[1] with the extant European populations, we conducted an allele-sharing analysis. For each one of the ancient genomes, we find the set of overlapping SNPs with the POPRES+FINHM dataset described above. We restricted the analysis to the 24 populations with sample sizes greater than eight. Then, for each allele observed in the ancient sample, we calculated the frequency of that allele in 8 randomly sampled chromosomes (without replacement) in each extant European population, and replicated this process 100 times. We then computed the mean frequency across all overlapping SNPs and all replicates,

stratified by population. The allele-sharing analysis shows that for the La Braña 1 individual as well as the Neolithic hunter-gatherers (Ajv52, Ajv70, Ire8) there is a general trend of increased sharing as one moves northward across European populations. The degree of sharing is positively correlated between the La Braña 1 individual and the Neolithic hunter-gatherers (Extended Data Figure 4). On the other hand, higher degree of sharing with the La Braña 1 individual is correlated with lower degree of sharing with the Neolithic farmer samples, Gok4 and Ötzi (Extended Data Figure 4), corroborating our observations in PCA.

7.4-Y-chromosome analysis

The mean coverage obtained for the Y chromosome (1.4x) prevented us from recovering phylogenetically relevant SNPs at high coverage. However, using unfiltered data, we were able to narrow down the paternal lineage affiliation of La Braña 1 individual (Table S9). The presence of the derived allele in many different mutations defining haplogroups A1, A1b, BT, CT and CF suggests La Braña 1 sample belongs to either haplogroup C or F. When mutations defining those haplogroups were checked, only ancestral alleles were found in the haplogroup F-defining mutations, whereas seven C-defining mutations (M130, M216, P255, P260, V183, V199 and V232) showed only derived alleles. Thus, La Braña 1 most likely belonged to haplogroup C. The actual distribution of haplogroup C is thought to be a consequence of a single out of Africa migration through Southern Asia, followed by a northward migration that eventually reached Siberia and the Americas[51]. The fact that we found ancestral alleles in mutations defining C1, C2, C3 and C4 (Table S9), together with their actual phylogeographic distribution restricted to Asia, Oceania and the Americas suggests that our individual does not belong to any of these branches. Rather, a new branch within haplogroup C (C6, originally named C7) has recently been identified in several men from Southern Europe, suggesting this could be an ancient European clade[52]. Importantly, mutation V20 showed one read with the derived allele (A), which points to C6 as the most probable sub-clade for La Braña 1 sample. It could also be possible that this G to A mutation is a result of DNA damage. Other less likely haplogroup affiliations are C* and C5 (no read covered SNP M356), both found mainly in present-day India.

Besides the V20 mutation, four other positions could have potentially been assigned wrongly due to the presence of DNA damage. However, their allele state is phylogenetically coherent with the rest of the SNPs studied. The precise affiliation of La Braña 1 in the Y-chromosome phylogeny could be better determined in the future with more data and increased genomic coverage.

7.5-Outgroup f_3 and D statistic analyses

We wanted to examine the genetic relationship between the La Braña 1 sample and another ancient genome from the Mal'ta site near Lake Baikal[15]. For all the analyses in this section involving La Braña 1 and Mal'ta genomes, we sampled reads at genomic positions rather than performing genotype calling. That is, for each of the ancient samples, we counted the number k_i of reference reads and the number n_i-k_i of non-reference reads covering site i that the sample displayed at the site. We then assigned a single allele to that sample at site i by uniformly choosing at random the reference allele with probability k_i/n_i and the non-reference allele with probability 1-k_i/n_i. To avoid errors due to *post-mortem* deamination, we considered the site as missing data in the ancient sample (La Braña 1 or Mal'ta) if we observed a C→T or G→A transition relative to the reference allele.

We next merged the two ancient genomes with the modern HGDP SNP dataset of unrelated individuals. Only biallelic sites were retained, and sites for which the two observed alleles were a C and T or were G and A were assigned as missing data in the ancient genomes. To match the sampling of a single allele (read) at a site for the ancient sample, we sampled a single allele for each of the modern individuals uniformly at random.

The first statistic we calculated was an outgroup f_3 statistic[16] of the form f_3 (population 1, population 2; outgroup). Here, the outgroup is the target population of the traditional f_3 statistic. We use the Yoruban population from the HGDP dataset as the outgroup and choose populations 1 and 2 to be ancient genomes and non-African modern populations. Because the Yoruban individual is genetically an outgroup to all the individuals that we will choose as populations 1 and 2, then these outgroup f_3 statistics are proportional to the amount of genetic drift shared between populations 1 and 2 since their divergence with the outgroup. The higher the value of the outgroup f_3

statistic, the more genetic drift they have shared since their divergence with the Yoruban population.

Extended Data Figure 5 displays pairwise outroup f3 statistics. In general, individuals from Oceania, East Asia, and Americas exhibit a genetic affinity to the Mal'ta sample, whereas individuals from the Middle East and Europe show a genetic affinity to the La Braña 1 sample. Further, though the comparison between Sardinians and the Mal'ta samples suggests that the Mal'ta individual is closer to the La Braña 1 sample than it is to any of the other modern European populations. We now examine this affinity of the Mal'ta and La Braña 1 individuals further.

We next computed *D* statistics[17] of the form *D* (Outgroup, H3; H2, H1). Significantly negative values of *D* indicate that H3 is closer to H2 than it is to H1 and significantly positive values of *D* indicate that H3 is closer to H1 than it is to H2. The test is significant if the *D* statistic is associated with a *Z* score with absolute value greater than 3. We wished to test whether the Mal'ta individual is closer to the La Braña 1 individual than it is to modern Europeans. To test this, we chose the outgroup as the modern Yoruban population, H3 as the Mal'ta individual, H2 as the La Braña 1 individual, and H1 as a modern European population from the HGDP.

Table **S10** presents result for this analysis, indicating that the Mal'ta individual is closer to the La Braña 1 individual than it is to any European population, with the exception of Orcadian and Russian populations. We also calculated *D* statistics using all the East Asian population from the HGDP, and found that the Mal'ta individual is significantly closer to the La Braña 1 individual than it is to the East Asians (Table S11).

One worry with these *D* statistic results is that there could be an affinity of the Mal'ta and La Braña 1 individuals due to some shared error structure in the ancient DNA sequencing. To address this, we reran these two *D* statistics analyses using a previously-published ancient genome of an Aboriginal Australian individual for H3 in place of the Mal'ta individual. We performed read sampling and deamination mitigation akin to the previous ancient DNA analyses described above. Table S12 shows comparisons with Europeans as H1, indicating that the Aborigine individual is not significantly closer to La Braña 1 than it is to any of the European populations. In contrast, Table S13 shows that the Aborigine individual is significantly closer to all sampled East Asian population than it is to the La Braña 1 individual. These results indicate that the affinity of the Mal'ta individual and the La Braña 1 individual is not

due to some ancient DNA attraction, and it is instead due to a shared genetic ancestry of the Mal'ta individual and the La Braña 1 individual.

8-Functional analysis

Ensembl's Variant Effect Predictor v.2.5 (Ensembl 67 annotation) was used to annotate the functional effects of the La Braña 1 individual's genetic profile. We restricted our analysis to variant changes within the longest transcripts of CCDS-verified genes (Consensus Coding Sequence Project of EBI, NCBI, WTSI, and UCSC - April 30th 2013 release) and to variants that had a minimum coverage of 4 reads. The main functional characterization of both dbSNP v137 shared and private (not shared with dbSNP v137) changes are summarized in Table S14**.** Furthermore, Polyphen-2 software[53] (standalone HumDiv model) was used in order to predict the effect of nonsynonymous amino acid changes. Nonsense and missense positions were analyzed in detail to investigate changes that could have a functional consequence in the La Braña 1 individual's phenotype and fitness. We looked either for mutations occurring within genes associated with known Human Mendelian Diseases[54] or for sites where La Braña 1 was homozygous for the alternative allele (Table S15-18). We identified a stop codon placed within the *POMT1* gene, and this genotype can be associated with the Walker-Warburg syndrome[55], a rare form of autosomal recessive congenital muscular dystrophy. The condition causes symptoms such as muscle weakness, hydrocephaly and eye abnormalities. However, it is clear from the skeletal examination that La Braña 1 was not suffering any obvious pathological condition, besides a possible ancient traumatism affecting the face and the L5 vertebrae.

For the nonsynonymous changes, we followed a series of steps to ensure that we only retained the "most deleterious" changes, starting with those that, according to Polyphen-2, had a >0.95 probability of being damaging and showing a false positive rate of less than 0.05. We filtered the sites using chemical classification following the Grantham Scores (GS) [56], and confined our analysis to sites that had a radical Grantham Score (>= 150). Finally, we retained only positions that had a GERP score >4, correlated to the evolutionary constrains on that site[57], and therefore changes in these positions are expected to be deleterious (Table S17 and Table S18).

One nonsynonymous change was identified in the *PHKB* gene, known to be associated to a Human Mendelian disease. Other mutations at *PHKB* cause a form of glycogen storage disease type IX (GSD IX) called GSD IXb. This form of disorder affects the liver and the muscles, causing muscle weakness. Again, we do not have any external evidence giving us reason to suspect that La Braña 1 individual was suffering from any such serious condition, and it seems very unlikely considering that the individual survived until an estimated age of 30-35 years.

9-Amylase copy number analysis

Copy number variation in the human gene encoding for salivary amylase (*AMY1*), an enzyme involved in starch degradation, has been reported as a remarkable example of recent positive selection associated to diet changes[18]. Using an *AMY1*-specific real-time quantitative PCR (qPCR) assay, the authors found that individuals from populations with high-starch diets have, on average, more *AMY1* copies than those with traditionally low-starch diets. Because La Braña 1 was a mesolithic hunter-gatherer, we expected the copy number at this locus to be consistent with a low-starch diet. To assess the raw copy number of the amylase, we employed a previously described method[58]. In short, Illumina reads without adapter sequences were trimmed to exclude the lower-quality ends of the reads, and then split into 36-bps reads. These reads were mapped against the human genome reference assembly using mrFast[58]; as reference assembly we used hg19 masked for repeats detected with RepeatMasker (www.repeatmasker.org) and Tandem Repeat Finder[59]. Next, we calculated the mean read depth per base pair in 1 kb non-overlapping windows of non-repetitive sequence. Finally, read depth values in the 1 kb windows were converted into GC-corrected estimates of absolute copy number using a set of diploid control regions.

In order to obtain a set of control regions, we excluded from hg19 all chromosomes other than autosomes and all known structural variants from several sources: (i) segmental duplications from the UCSC Genome Browser (http://genome.ucsc.edu/), (ii) variants from the Database of Genomic Variants (http://projects.tcag.ca/variation/), (iii) calls from the 1000G Project[13] present in the Database of genomic structural variation (dbVar) (http://www.ncbi.nlm.nih.gov/dbvar/).

This resulted in 22,864 control regions with a total size of 1,658,646,294 bps (52.9% of hg19) and the majority being smaller than ~200 kb (Extended Data Figure 7a).

We estimated copy number greater than two on the region in hg19 comprising the three copies of *AMY1*, a region in which structural variation has been consistently reported[60–64]. Using the coordinates for the structural variant from one of these studies [18] we estimated a copy number of ~5 copies in our ancient genome. Moreover, the predicted copy number was consistent across each of the three *AMY1* gene copies (Extended Data Figure 7b).

We then compared the *AMY1* number of copies with the data from[18] (Extended Data Figure 7c). Interestingly, 54% of the low-starch diet individuals in the original study had an *AMY1* copy number equal or lower than our sample (i.e. 5 copies) whereas 78.9% of the individuals with a high-starch diet had >5 copies. There is considerable overlap between the *AMY1* copy number distributions between low- and high-starch diet populations (Extended Data Figure 7c). However, using the data from[18] we showed that a threshold of five copies of the *AMY1* gene for classifying individuals as either in the low- or high-starch group maximizes the fraction of correct predictions (g = 69.5%), and that this value is significantly higher than the expected by chance (Extended Data Figure 7d). Altogether, these results suggest that the amylase copy number in the individual sequenced is likely to correspond to a low starch diet environment, as expected from a Mesolithic, pre-agriculture sample.

10- Lactose intolerance

The SNP rs4988235, known to influence the lactase LCT gene[65], showed the variant (G) associated with lactose intolerance (8x). This is expected, as lactose tolerance is believed to have arisen with the introduction of farming at the posterior Neolithisation process. Hence, the Mesolithic hunter-gatherer population, to which the La Braña 1 individuals belonged, was probably lactose intolerant.

One of the 8 reads showed an A at this position, but being next to another G to A change, it is very likely a result of *post-mortem* damage in this particular read.

11- Dating recent adaptations

This first genome of a Mesolithic individual should allow us to date, for the first time, candidate adaptative events described in Europeans that might have been associated to the transition to the agriculture. A novel scan for signatures of recent adaptations in humans, using a composite for multiple signals test in 1000G data, identified 35 high-scoring recent candidate nonsynonymous variants, 10 of which were detected in CEU individuals[19]. We determined the state of these 10 nonsynonymous SNPs in the Mesolithic individual (Table S19).

When working with ancient genomes at low coverage, one must be aware of the problem introduced by nucleotide misincorporations, which are mainly caused by C to T and G to A changes. However, the probability of identical misincorporations accumulating by chance in different reads at the same genomic position is very low, and hence we focused on alleles observed in multiple reads[66].

We observed three SNPs diagnostic of possible adaptative events with only one read at La Braña 1: rs17310144, rs10421769 and rs12918952 (Table S19). The SNP rs17310144 (in the immunity gene *GPATCH1*) involves a T to G change that cannot be interpreted as *post-mortem* damage. However, the latter two can be problematic as they involve potential *post-mortem* damage changes (one T allele in rs10421769 and one A allele in rs12918952).

Focusing on the SNPs supported by more than one read, four variants (rs11755393, rs2745098, rs11881633 and rs1056899) presented clearly the selected allele in Europeans. Two of them, rs2745098 and rs11755393, have been associated with a function in the immune system. On the other hand, rs9262, rs1426654 and rs16891982 showed only the non-selected allele in Europeans. Interestingly, rs1426654 and rs16891982 are located in two genes related to pigmentation, SLC24A5 and SLC45A2 respectively[20,67]. For both genes, La Braña 1 carries alleles absent or extremely rare in modern Europeans.

11.1- Haplotype analysis

In order to provide more power to the previous observations, we wanted to study if current European or African haplotypes around these variants could be observed in

the ancient sample. For each of the 10 candidate nonsynonymous SNPs; we looked for variants in strong linkage disequilibrium (LD) (r^2> 0.8) in CEU and YRI from the 1000G. Next, we determined the state of these variants in the Mesolithic genome. Since we have no phased data, we focused only in homozygous SNPs, where La Braña 1 haplotype can be unambiguously identified.

For SNPs rs16891982, rs1426654 and rs9262 (in *SLC45A2, SLC24A5 and C12orf29* respectively) for which La Braña 1 carries the ancestral allele, only two SNPs were found in LD with rs1426654 in YRI, and only one with rs9262 in CEU and YRI (Table S20). The reason is that most variants surrounding them are completely or almost completely fixed in CEU, whereas in YRI LD (r^2) did not surpass the 0.8 thresholds. We analyzed a window of around 40 variants at each side of the selected SNPs and analyzed their allele frequencies in extant CEU and YRI and compared them with La Braña 1 genotypes (Figure 2). For rs16891982 and rs1426654 most neighboring SNPs that are segregating in YRI appear fixed in CEU, while La Braña 1 carries genotypes not compatible with the present European's at many sites. On the contrary, combinations of alleles for these SNPs found in La Braña 1 can be present in the extant YRI population, which suggests that those sites with ancestral alleles are not back mutations, but that the Mesolithic sample contains partial ancestral haplotypes. These observations are not clear for rs9262, since La Braña 1 carries and ancestral haplotype that can be found frequently among present CEU individuals.

In contrast, for the SNPs carrying the derived allele in La Braña 1: rs11755393 (*UHRF1BP1*) and rs2745098 (*PTX4*), we were able to retrieve many SNPs in strong LD independently in CEU and YRI. We observed that, for these SNPs, haplotypes found in La Braña 1 are currently common in extant CEU population and less frequent in YRI (Table S20). A similar snapshot of the allele frequencies of neighboring SNPs in CEU and YRI (Extended Data Figure 8a, b) visually demonstrates that most of the alleles found in La Braña 1 correspond to the major allele found in CEU and to the minor allele found in YRI. Genotypes found in La Braña 1 are more compatible with a CEU rather than a YRI combination of alleles.

Similarly, the derived allele for heterozygous SNPs in La Braña 1: rs11881633 (*TDRD12*) and rs1056899 (*SETX*), is combined in haplotypes (from SNPs in LD>0.8) that are much more common in present CEU than among YRI individuals (Table S20). The ancestral allele, also present in La Braña 1 at these two heterozygous SNPs, is forming haplotypes that are completely absent from both CEU and YRI populations.

These might have been transitional haplotypes that are now vanished from extant populations.

Finally, La Braña 1 carries also the derived allele for SNPs rs17310144 (*CCDC14*), rs12918952 (*WWOX*) and rs10421769 (*GPATCH1*) (Extended Data Figure 8c). However, in these three cases haplotypes displayed by La Braña 1 are absent, or almost absent, from present Europeans (Table S20). This may indicate that despite La Braña 1 presented the derived alleles, they were not conforming the haplotypes that later increased in frequency in Europeans. Alternatively, since these SNPs are represented by only one read in La Braña 1, we cannot rule out the presence of the ancestral allele and hence a heterozygous state. Moreover, for SNPs in WWOX and GPATCH1 genes, the presence of the derived allele could be a consequence of *post-mortem* DNA damage. If these 3 SNPs carried the ancestral allele, WWOX and GPATCH1 SNPs would show a haplotype at frequencies of 0.38 and 0.25 in CEU, respectively, and 0.08 and 0 in YRI. The ancestral state of a putative heterozygous rs17310144 in CCDC14 would imply a La Braña 1 haplotype present at frequencies of 0.4 and 0.87 in CEU and YRI, respectively. Nevertheless, allele frequencies of SNPs surrounding these two SNPs show a pattern more similar to CEU than to YRI.

12-Pigmentation

In order to further investigate the pigmentation phenotype in this Mesolithic sample, we determined the ancestral/derived state in a set of additional pigmentation genes described to have a role in extant Europeans' phenotype[22,68].

The coverage in La Braña 1 is stated for each SNP, ranging from 1x to 12x (Table S21). The low coverage in some of these SNPs can underestimate the heterozygosity and also can be subjected to potential confounding effects from *post-mortem* DNA damage.

La Braña 1 shows ancestral alleles in several of them, including *TYR*, *ASIP* and *MC1R*. Interestingly, HERC2/OCA2 SNPs related to eye color on chromosome 15 (rs4778241, rs1129038, rs12593929, rs12913832, rs7183877, rs3935591, rs7170852, rs2238289, rs3940272, rs8028689, rs2240203, rs11631797, rs916977)[69] consistently displayed a haplotype that is highly associated with blue eye color (Table S21). Furthermore, using the alleles present in La Braña 1 for the six most informative SNPs

for eye color prediction[70] and for other SNPs associated to hair color, the HIrisPlex model[24] was applied (Table S22). We obtained a probability of 0.780 and 0.202 for having black and brown hair respectively. Regarding the eye color prediction, no data could be recovered for SNP rs12896399. Thus, accounting for the three possible genotypes in this SNP, we obtained a probability range of 0.706-0.458, 0.214-0.117 and 0.328-0.177 for blue, intermediate and brown color respectively. If it can be confirmed that La Braña 1 had blue eyes with the genotyping of the SNP rs12896399, it would be so far the oldest known individual carrying this trait.

The combination of seven SNPs has also been shown to be useful in the prediction of "not-dark" and "not white" skin color in a global human sample[71]. However, some of these SNPs are not variable between Asians and Europeans and thus, its usefulness for skin color prediction in an ancient, European-specific genetic background is debatable. Moreover, the model, developed for forensics purposes, is based on the number of these SNPs present in a particular individual. La Braña 1 has the non-dark allele in rs6119471 (two copies), rs12913832 (two copies) and rs12203592 (one copy).

Regarding the final phenotypic effect of each of these pigmentation genes, a semi-quantitative assessment of different functional studies classifies both *SLC45A2* and *SLC24A5* with the strongest effects in Europeans; *MC1R*, *KITLG*, *IRF4* with medium effects and *ASIP*, *OCA2*, *TYR* and *TYRP1* with weak effect[22]. All these genes show alleles that are present in high frequencies in Europeans but absent of rare in Africans; numerous association and functional studies have shown they play a role in light skin color in Europeans influencing the melanin synthesis, either as a membrane transporters (*SLC45A2*, *SLC24A5*, *OCA2*), growth factor (*KITLG*), melanogenic enzymes (*TYR*, *TYRP1*), G-protein coupled receptor (*MC1R*) or *MC1R* antagonist protein (*ASIP*).

The only derived alleles found in La Braña 1 are present in *IRF4* and two of the weak-effect genes, *TYRP1* and *ASIP*. Interestingly, one of the SNPs (rs1408799) is basically associated to blue eye color in Europeans[22]. This is again in agreement with the prediction of dark skin pigmentation and non-brown eyes in this particular individual. The existence of many association studies on pigmentation and the understanding of the biochemical pathways involved in the melanosomes indicate that any new discovery on gene pigmentation will represent a very minor proportion of the observed phenotypic variation.

As a conclusion, it is very difficult to ascertain the precise skin pigmentation phenotype of La Braña 1 individual from the comparison with modern human populations because the two critical pigmentation SNPs (at *SLC45A2* and *SLC24A5* genes*)* are essentially fixed in extant Europeans. Although some neighbouring populations such as Middle Easterns or North Africans can carry these ancestral alleles, their general genetic background is still quite different from that of modern Europeans and especially from Northern Europeans, to which La Braña 1 displays clear affinities. It is likely that La Braña 1 represents a unique phenotype that is no longer present in modern Europeans.

13-Immunity

Pathogens are expected to have shaped the human genome as a result of selective pressures acting upon genes involved in host resistance and immunity response. The intimate contact with domesticated animals during the Neolithic is believed to have increased considerable the zoonotic infections (that is, the transmission of specific infectious diseases from an animal to humans). Indeed, it is believed that a large number of the common modern infections encountered in humans are zoonotic post-Neolithic events[72]. Several genome-wide scans for positive selection studies have detected different genes that could be involved in the Neolithic remodeling of the human immune system[73]. Due to the inherent difficulties of detecting recent positive selective events, some of these genetic targets have not been replicated in all studies; however, they represent instances of significantly large allele frequency differences between Europeans and the rest of the world populations.

We have determined the ancestral/derived allele at the La Braña 1 Mesolithic genome in several genes that fulfill the following criteria: they show evidence of positive selection in European populations and they harbour polymorphisms shown to influence susceptibility to infections in European populations[74–79] (Table S23). In total, we ascertained 63 SNPs from 40 genes. For several of these genes it has been argued that selective events driven by migration patterns and local infections have represented an important factor shaping the modern-day representation in European populations[80]. Nevertheless, the precise role of each gene in a general picture in relation to the possible zoonosis at the Neolithic is still not well understood.

The assessment of these polymorphisms shows that a significant number of these genes present derived variant. The presence of the inactive *CASP12* variant in the La Braña 1 individual support the earlier hypothesis of a pre-Neolithic origin for the spread and fixation of this mutation[81,82]. These genes encode proteins that are pattern-recognition receptors (*TLR1*, *TLR3*, *TLR8*, *CD14*, *LGP2* and *IFIH1)*, intracellular adaptor molecules (*LRRK2*), cytokines or cytokine modulators (*SOCS2*, *TOLLIP*, *IFNG*, *CASP12*, *IL-29*, *TGFB2*), chemokine and chemokine receptors (*CCL18*) and effector molecules (*NOS2A*), all with central roles in the host defense against major groups of viral or bacterial pathogens. Indeed, the fundamental role played by these genes in innate immune host defense and protection against major infectious diseases (e.g. sepsis), may explain the advent of protective genotypes in European populations independent and anterior to the spread of zoonotic infections due to domesticated animals. In line with this, the La Braña 1 individual carries the heterozygous haplotype at *Mal/TIRAP* gene that has been described as protective to a broad range of infections such as pneumococcal infection, bacteremia, malaria and tuberculosis[83]. The pre-Neolithic presence of these major infections (including malaria, known to have been present in the Southern part of the continent) in Europe is likely to have driven this pattern.

In addition to the group of genes that show the presence of the derived protective allele in the La Braña 1 individual, other genes show however the presence of ancient variants. Important examples are *SH2B3*, that has been reported a strong recent selective event[84]. The derived allele that has been demonstrated to induce a stronger antibacterial immune response is absent in the La Braña 1 individual. The ancient allele is also present for the *TLR6* gene, in which the derived variant has been suggested to protect against colonization and infection with lung pathogens[85,86]. The ancient haplotype is present for the well-known Asp299Gly and Thr399Ile *TLR4* polymorphisms, although the impact of these gene variants on susceptibility to infections in more controversial[87]. CCR5delta32 polymorphism has been suggested to be selected in Europeans by either plague or smallpox[88]. Only one read of the gene was available in the La Braña 1 individual, nevertheless showing the absence of the deletion.

In conclusion, the genotype in several important immune genes in the La Braña 1 individual shows that protective derived alleles present in extant Europeans have occurred before the Neolithic period for several important genes that have broad immune effects for bacterial and viral infections. Nevertheless, protective variants in

several other immune genes are not present in this Mesolithic individual, arguing for a mixed picture of pre- and post-Neolithic events shaping the current immune system of Europeans.

14- e-QTL analysis

We determined the La Braña 1 individual ancestral/derived allele state in a variety of regulatory SNPs (eQTL) candidates for adaptive evolution. There is increasing evidence for the importance of eQTLs in the detection of positive selection[89]. We ascertained 14 regions positively selected in Europeans that have been shown to contain eQTL SNPs in the same population, as identified by Grossman et al.[19]. On average, each region spans 41.5 kb (range: 0 – 207.9 kb) and harbours 10 eQTLs (range: 1 – 45). To increase our power to look upon genes whose expression is controlled by the ascertained eQTL, we used a recent meta-analysis of eQTL effects detected upon ~8,000 European individuals[90]. In total, we found 32 genes whose expression levels are controlled by the candidate eQTLs detected by Grossman et al.[19]. For every gene and candidate adaptive eQTL, we determined the ancestral/derived allelic state of the Mesolithic individual, as well as its predicted effect (higher/lower) on gene expression (Table S24). Interestingly, La Braña 1 possessed the positive selected haplotype in Europeans that increases the expression of *SPINT2*, whose product has been associated to several Mendelian diarrhea associated to sodium excretion.

Furthermore, we looked for eQTL SNPs that have been shown to control the expression of the 40 genes studied in section 13 and determined their allelic state in the La Braña 1 individual (Table S25).

15-Metagenonomic analysis

15.1- General taxonomic units discovery and analysis

We retrieved all reads that did not mapped to the human reference genome (around 1,000 million reads). We then subsampled one million reads and used BLAST 2.2.27+[91] with *megablast* settings against the non-redundant nucleotide database (nr).

The blast output was loaded on MEGAN4[28] to perform the taxonomy assignments and analyses. Complementarily, we produced contigs from unmapped reads using Velvet[92]. Contigs were then assigned to taxons via BLAST 2.2.27+ and MEGAN4.

First, we assessed whether we were infra-representing taxonomy diversity by downsampling the whole dataset of unmapped reads to one million reads. We performed rarefaction curves by counting the number of detected leaves by cumulatively increasing the number of reads in bins of 100.000. We observed that, when around 80% of our subset was analyzed, a plateau was reached with poor increase of new leaves.

Around 86% of all reads did not match any subject in nr database (Extended Data Figure 9a). Most reads appeared unassigned or with no blast matches, some of them probably representing repetitive or low-complexity sequences. Of the 14% of aligned reads, 65% corresponded to bacterial organisms, 34% to *Eukaryota* and less than 1% to *Archea*. Almost no viral DNA was detected. The Bacteria group was mostly enriched of *Proteobacteria* (around 89%) and some presence of *Acinobacteria* (around 8%) (Extended Data Figure 9b). Similar represented organisms were found when *blasting* contigs instead of raw reads.

Looking into the group of *Proteobacteria* (Extended Data Figure 9c), around 71% corresponded to *Betaproteobacteria*. This class comprises several aerobic groups of bacteria, many of them inhabiting wastewater or soil. The most represented group in *Betaproteobacteria* is *Comamonadaceae*, with many reads mapping on *Acidovorax*, a known plant pathogen. There is also 19% of *Gammaproteobacteria,* mostly *Pseudomonadales, Xanthomonadales (including Xanthomonas* and *Stenotrophomonas,* which are very ubiquitous organisms) and also different *Enterobacteriales*. *Actinobacteria*, the second best represented group of *Bacteria*, includes around 600 reads divided among different *Mycobacterium* species.

Regarding the microbial attributes of the microbes present in the La Braña 1 sample, most of them are gram negative, rod-shaped, mesophiles and aerobic organisms with unknown involvement in disease (Extended Data Figure 9d).

15.2- Viral metagenomics

To specifically assess the presence of integrated or episomal viral DNA, we mapped with *bwa-0.7.4*[27] *mem* all reads non-mapping to the reference human genome

against two different sequence databases. The first one included all known genomes of DNA viruses and the other one included all known RNA-virus genomes. Then, reads that aligned to any of the two databases were mapped again this time against individualized viral genomes using *bwa-0.7.4*[27] *aln*. Finally, the number of uniquely mapping reads was counted.

We only found viruses (Table S26) either used as positive control in DNA sequencing (PhiX) or bacterial phages infecting organisms already detected in the previous analyses (i.e *Pseudomonas phage vB_PaeS_PMG1* or *Burkholderia phage BcepMu*).

References

31. Orlando, L. *et al.* Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499,** 74–8 (2013).

32. Meyer, M. & Kircher, M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harb. Protoc.* pdb.prot5448 (2010). doi:10.1101/pdb.prot5448

33. Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52,** 87–94 (2012).

34. Lindgreen, S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes* **5,** 337 (2012).

35. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23,** 147 (1999).

36. Schubert, M. *et al.* Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13,** 178 (2012).

37. DePristo, M. a *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–8 (2011).

38. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 14616–21 (2007).

39. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7,** e34131 (2012).

40. Brotherton, P. *et al.* Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res.* **35,** 5717–28 (2007).

41. Schwarz, C. *et al.* New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Res.* **37,** 3215–29 (2009).

42. Deagle, B. E., Eveson, J. P. & Jarman, S. N. Quantification of damage in DNA recovered from highly degraded samples--a case study on DNA in faeces. *Front. Zool.* **3,** 11 (2006).

43. Kloss-Brandstätter, A. *et al.* HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32,** 25–32 (2011).

44. Van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30,** E386–94 (2009).

45. Hach, F. *et al.* mrsFAST: a cache-oblivious algorithm for short-read mapping. *Nat. Methods* **7,** 576–7 (2010).

46. Pemberton, T. J. *et al.* Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* **91,** 275–92 (2012).

47. Purcell, S. *et al.* PLINK : A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).

48. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–73 (2010).

49. Wang, C. *et al.* Comparing spatial maps of human population-genetic variation using Procrustes analysis. *Stat. Appl. Genet. Mol. Biol.* **9,** Article 13 (2010).

50. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–9 (2006).

51. Zhong, H. *et al.* Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. *J Hum Genet* **55,** 428–435 (2010).

52. Scozzari, R. *et al.* Molecular dissection of the basal clades in the human Y chromosome phylogenetic tree. *PLoS One* **7,** e49170 (2012).

53. Adzhubei, I. a *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–9 (2010).

54. Blekhman, R. *et al.* Natural selection on genes that underlie human disease susceptibility. *Curr. Biol.* **18,** 883–9 (2008).

55. Reeuwijk, J. Van *et al.* The Expanding Phenotype of POMT1 Mutations : From Walker-Warburg Syndrome to Congenital Muscular Dystrophy , Microcephaly , and Mental Retardation. **27,** 453–459 (2006).

56. Grantham, R. Amino acid difference formula to help explain protein evolution. *Science* **185,** 862–864 (1974).

57. Davydov, E. V *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* **6,** e1001025 (2010).

58. Alkan, C. *et al.* Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat. Genet.* **41,** 1061–7 (2009).

59. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27,** 573–80 (1999).

60. Matsuzaki, H., Wang, P.-H., Hu, J., Rava, R. & Fu, G. K. High resolution discovery and confirmation of copy number variants in 90 Yoruba Nigerians. *Genome Biol.* **10,** R125 (2009).

61. Perry, G. H. *et al.* The Fine-Scale and Complex Architecture of Human Copy-Number Variation. 685–695 (2008). doi:10.1016/j.ajhg.2007.12.010.

62. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444,** 444–54 (2006).

63. Shaikh, T. H. *et al.* High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res.* **19,** 1682–90 (2009).

64. Wong, K. K. *et al.* ARTICLE A Comprehensive Analysis of Common Copy-Number Variations in the Human Genome. **80,** 91–104 (2007).

65. Bersaglieri, T. *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74,** 1111–20 (2004).

66. Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, a & Pääbo, S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **29,** 4793–9 (2001).

67. Soejima, M. & Koda, Y. Population differences of two coding SNPs in pigmentation-related genes SLC24A5 and SLC45A2. *Int. J. Legal Med.* **121,** 36–9 (2007).

68. Beleza, S. *et al.* Genetic architecture of skin and eye color in an African-European admixed population. *PLoS Genet.* **9,** e1003372 (2013).

69. Eiberg, H. *et al.* Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum. Genet.* **123,** 177–87 (2008).

70. Liu, F. *et al.* Eye color and the prediction of complex phenotypes from genotypes. *Curr. Biol.* **19,** R192–3 (2009).

71. Spichenok, O. *et al.* Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic Sci. Int. Genet.* **5,** 472–8 (2011).

72. Wolfe, N. D., Dunavan, C. P. & Diamond, J. Origins of major human infectious diseases. *Nature* **447,** 279–83 (2007).

73. Barreiro, L. B. & Quintana-Murci, L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat. Rev. Genet.* **11,** 17–30 (2010).

74. Casals, F. *et al.* Genetic adaptation of the antibacterial human innate immunity network. *BMC Evol. Biol.* **11,** 202 (2011).

75. Manry, J. *et al.* Evolutionary genetics evidence of an essential, nonredundant role of the IFN-γ pathway in protective immunity. *Hum. Mutat.* **32,** 633–42 (2011).

76. Barreiro, L. B. *et al.* Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genet.* **5,** e1000562 (2009).

77. Vasseur, E. *et al.* The selective footprints of viral pressures at the human RIG-I-like receptor family. *Hum. Mol. Genet.* **20,** 4462–74 (2011).

78. Manry, J. *et al.* Evolutionary genetic dissection of human interferons. *J. Exp. Med.* **208,** 2747–59 (2011).

79. Ferrer-Admetlla, A. *et al.* Balancing Selection Is the Main Force Shaping the Evolution of Innate Immunity Genes. *J. Immunol.* **181 ,** 1315–1322 (2008).

80. Netea, M. G., Wijmenga, C. & O'Neill, L. a J. Genetic variation in Toll-like receptors and disease susceptibility. *Nat. Immunol.* **13,** 535–42 (2012).

81. Xue, Y. *et al.* Spread of an inactive form of caspase-12 in humans is due to recent positive selection. *Am. J. Hum. Genet.* **78,** 659–70 (2006).

82. Hervella, M. *et al.* The loss of functional caspase-12 in Europe is a pre-neolithic event. *PLoS One* **7,** e37022 (2012).

83. Khor, C. C. *et al.* A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. *Nat. Genet.* **39,** 523–8 (2007).

84. Zhernakova, A. *et al.* Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. *Am. J. Hum. Genet.* **86,** 970–7 (2010).

85. Misch, E. A., Verbon, A., Prins, J. M., Skerrett, S. J. & Hawn, T. R. A TLR6 polymorphism is associated with increased risk of Legionnaires/' disease. *Genes Immun* (2013). at <http://dx.doi.org/10.1038/gene.2013.34>

86. Winters, A. H. *et al.* Single Nucleotide Polymorphism in Toll-like Receptor 6 Is Associated With a Decreased Risk for Ureaplasma Respiratory Tract Colonization and Bronchopulmonary Dysplasia in Preterm Infants. *Pediatr. Infect. Dis. J.* **32,** (2013).

87. Ferwerda, B. *et al.* Functional consequences of toll-like receptor 4 polymorphisms. *Mol. Med.* **14,** 346–52 (2007).

88. Galvani, A. P. & Slatkin, M. Evaluating plague and smallpox as historical selective pressures for the CCR5-Delta 32 HIV-resistance allele. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 15276–9 (2003).

89. Fraser, H. B. Gene expression drives local adaptation in humans. *Genome Res.* **23,** 1089–96 (2013).

90. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45,** 1238–43 (2013).

91. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).

92. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18,** 821–9 (2008).

93. Kong, A. *et al.* Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488,** 471–5 (2012).

Tables

Table S1. Frequency of all possible type of nucleotide changes in the final SNP set. Ref: reference allele. Alt: alternative allele.

Ref-Alt	Number of SNPs	Frequency (%)
A-T	26,514	2.7
A-C	47,318	4.8
A-G	203,684	20.6
T-A	26,474	2.7
T-C	203,257	20.5
T-G	46,903	4.7
C-A	38,291	3.9
C-T	121,086	12.2
C-G	58,358	5.9
G-T	38,641	3.9
G-A	121,111	12.2
G-C	58,527	5.9

Table S2. Filtered set of SNPs with a minimum read depth of 4x. Hom: homozygous positions. Het: heterozygous positions.

Chr.	SNPs	Hom.	Het.	% Het.	Present in dbSNP	Novel	% Novel
1	80,884	47,556	33,328	41.2	79,641	1,243	1.5
2	79,168	45,582	33,586	42.4	77,980	1,188	1.5
3	62,400	36,871	25,529	40.9	61,478	922	1.5
4	61,569	37,956	23,613	38.4	60,775	794	1.3
5	55,279	32,667	22,612	40.9	54,524	755	1.4
6	59,789	32,030	27,759	46.4	58,927	862	1.4
7	53,171	30,842	22,329	42.0	52,256	915	1.7
8	51,561	29,062	22,499	43.6	50,867	694	1.3
9	42,227	23,746	18,481	43.8	41,650	577	1.4
10	53,745	32,141	21,604	40.2	52,960	785	1.5
11	52,319	31,115	21,204	40.5	51,566	753	1.4
12	45,842	26,748	19,094	41.7	45,061	781	1.7
13	33,174	20,847	12,327	37.2	32,779	395	1.2
14	31,397	18,176	13,221	42.1	30,926	471	1.5
15	31,973	19,446	12,527	39.2	31,513	460	1.4
16	38,447	21,892	16,555	43.1	37,855	592	1.5
17	33,999	20,170	13,829	40.7	33,393	606	1.8
18	27,563	16,359	11,204	40.6	27,217	346	1.3
19	28,908	14,579	14,329	49.6	28,401	507	1.8
20	26,457	15,061	11,396	43.1	26,057	400	1.5
21	15,543	9,155	6,388	41.1	15,298	245	1.6
22	17,731	9,592	8,139	45.9	17,465	266	1.5
X	6,721	6,721	0	0	6,628	93	1.4
Y	274	274	0	0	221	53	19.3
Mt	26	26	0	0	26	0	0
Total	990,167	578,614	411,553	41.6	975,464	14,703	1.5

Table S3. DNA decay rates per site (k), the corresponding DNA half-life ($T_{1/2}$) for a 100 bp fragment, and the estimated time until the average fragment length in the sample is 30 bp and 1 bp respectively. Predicted values are based on equation in[9].

	k	**$T_{1/2}$, 100 bp**	**Yrs, av. length = 30 bp**	**Yrs, av. length = 1 bp**
La Braña 1, nuDNA	4.60E-06	1,507	7,246	217,391
La Braña 1, mtDNA	2.80E-06	2,476	11,905	357,143
Predicted for mtDNA at 8.1°C	2.10E-06	3,466	15,873	476,190

Table S4. Mitochondrial DNA contamination estimates by assessing the nucleotide substitutions present at the U5b2c1 haplogroup diagnostic positions.

Position	rCRS	La Braña 1	Coverage	Haplogroup	La Braña 1 / Total	Contamination %
11467	A	G	135	U	132/135	2.22
12308	A	G	125	U	123/125	1.60
12372	G	A	109	U	106/109	2.75
3197	T	C	129	U5	129/129	0
9477	G	A	47	U5	47/47	0
13617	T	C	108	U5	107/108	0.93
16192	C	T	131	U5	130/131	0.76
16270	C	T	119	U5	118/119	0.84
150	C	T	117	U5b	116/117	0.85
7768	A	G	118	U5b	115/118	2.54
14182	T	C	66	U5b	62/66	6.06
1721	C	T	137	U5b2	137/137	0
13637	A	G	112	U5b2	110/112	1.79
723	A	G	118	U5b2c	116/118	1.69
13017	A	G	146	U5b2c	137/146	6.16
6920	C	A	115	U5b2c1	115/115	0
13434	A	G	115	U5b2c1	114/115	0.87
				Total	1,914/1,947	1.69

Table S5. Chromosomes X and Y statistics after masking over-represented k-mers. The hg19 assembly was downloaded from the UCSC Genome Browser (http://genome.ucsc.edu/) with hard-masking of repeats reported by RepeatMasker and Tandem Repeat finder.

	Chromosome X	**Chromosome Y**
Size (Bps)	155,270,560	59,373,566
Ns in hg19 (including gaps)	95,898,879	49,783,032
Ns in hg19 (after kmer masking)	102,920,656	56,843,824
Bps masked (% chromosome)	7,021,777 (4.5)	7,060,792 (11.9)

Table S6. Contamination estimates for La Braña 1 (Heterozygous positions/Total) from the analysis of heterozygous positions in uniquely mapped regions on chromosome X. Minimum read depth of 4x. Ref: reference. Alt: Alternative.

	Homozygous Ref (dbSNP positions)	Homozygous Alt	Heterozygous	Total	Heterozygous/Total
chrX	61,906	2,583	206	64,695	0.0031

Table S7. Coverage values for the 1000G Project samples used for heterozygosity analyses. MQ: mapping quality.

Population	Sample	Number of reads (MQ > 25)	Mean read depth
YRI	NA18934	110,171,416	3.85
LWK	NA19456	78,386,431	2.85
CEU	NA12272	101,639,094	3.53
TSI	NA20792	114,718,783	3.10
IBS	HG02224	84,595,293	3.03
GBR	HG00114	95,651,536	3.44
FIN	HG00308	89,765,021	3.30
CHB	NA18618	85,740,358	2.93
JPT	NA18969	87,097,439	2.80

Table S8. Effective population sizes (Ne) estimated from heterozygosity values. μ = Mutation rate per nucleotide per generation = 1.20E-08 [93].

Sample	Θ (Heterozygosity)	Ne = Θ/(4*μ)
La Braña 1	0.000419	8,728
IBS	0.000520	10,839
TSI	0.000514	10,705
CEU	0.000525	10,927
GBR	0.000530	11,041
FIN	0.000545	11,349
YRI	0.000687	14,308
LWK	0.000658	13,718
CHB	0.000457	9,514
JPT	0.000469	9,762

Table S9. Phylogenetically relevant Y-chromosome SNPs studied in the La Braña 1 sample.

SNP name	SNP ID	Position in hg19	Haplogroup [a]	Ancestral	Derived	La Braña 1	Coverage
V241	rs184836128	7658712	A1	C	T	T [b]	1x
V174	rs2563344	4957608	A1	A	G	G	2x
V238	rs186764807	7651556	A1	G	T	T	2x
V250	rs191193873	17645335	A1	A	G	G	1x
L985		7374927	A1	A	C	C	1x
L989		23551003	A1	T	A	A	2x
P108		15426248	A1b	C	T	T	2x
L413	rs192939307	6932831	BT	G	A	A	3x
L418	rs78351457	10008803	BT	C	G	G	7x
M42	rs2032630	21866840	BT	A	T	T	1x
M299	rs13447347	22748506	BT	T	G	G	2x
M94	rs2032647	21938158	BT	C	A	A	4x
M168	rs2032595	14813991	CT	C	T	T	2x
M294	rs9341317	22744945	CT	C	T	T [b]	1x
P143	rs4141886	14197867	CF	G	A	A	4x
M89	rs2032652	21917313	F	C	T	C	2x
P134	rs9786877	7395806	F	C	G	C	1x
P135	rs9786502	21618856	F	C	T	C	1x
P142	rs4988808	7218079	F	G	A	G	1x
P145	rs17842387	8424089	F	G	A	G	2x
P148	rs16980396	19349615	F	C	T	C	2x
P187	rs17174528	9108252	F	G	T	G	1x
M130	rs35284970	2734854	C	C	T	T	2x
M216	rs2032666	15437564	C	C	T	T [b]	1x
P184		7218128	C	T	C	no data	
P255		8685038	C	G	A	A	3x
P260		17286006	C	A	C	C	2x
Page85	rs35906235	14924643	C	G	T	no data	
V77		17947542	C	C	T	no data	
V183		14263271	C	G	A	A	2x
V199		2772928	C	C	A	A	1x
V232		7629098	C	T	C	C	4x
M8	rs3899	7291534	C1	G	T	G	3x
M105	rs2032612	21866491	C1	C	T	no data	

P122		14850732	C1	C	A	C	8x
M38		21742158	C2	T	G	T	1x
M208	rs2032659	15576203	C2a	C	T	C	2x
M217	rs2032668	15437333	C3	A	C	A	1x
P44		14495251	C3	G	A	G	1x
Z1453		19020366	C3	A	G	A	1x
M347		2877479	C4	A	G	A [b]	1x
M210	rs2032660	15575780	C4a	A	T	A	1x
M356		2888203	C5	C	G	no data	
P92		14850853	C5a	C	T	no data	
V20	rs182352067	6845955	C6	G	A	A [b]	1x
V86		6909957	C6	G	A	G	1x
V182		14249991	C6	C	T	C	1x
V184		14263293	C6	C	T	C	2x
V219		6664018	C6	T	C	no data	
V222		7589937	C6a	G	C	G	4x

[a] According to: International Society of Genetic Genealogy (2013). Y-DNA Haplogroup Tree 2013, Version: 8.85, Date: 9 December 2013, http://www.isogg.org/tree/, Date of access: 09 December 2013.
[b] Possible *post-mortem* DNA damage

Table S10. *D* statistics based on Europeans from the HGDP SNP dataset and the La Braña 1 and Mal'ta ancient genomes. For all comparisons, Yoruba was the outgroup population, H3 was the Mal'ta individual, H2 was the La Braña 1 individual, and H1 was a European individual from the HGDP dataset. Statistically significant *D* statistics (indicated in bold) are those that have the absolute value of their *Z* score greater than 3.

Outgroup	**H3**	**H2**	**H1**	**D statistic**	**Z score**
Yoruba	Mal'ta	La Braña 1	Orcadian	-0.0197	-2.630
Yoruba	**Mal'ta**	**La Braña** 1	Adygei	-0.0317	**-4.264**
Yoruba	Mal'ta	La Braña 1	Russian	-0.0129	-1.763
Yoruba	**Mal'ta**	**La Braña** 1	Basque	-0.0277	**-3.726**
Yoruba	**Mal'ta**	**La Braña** 1	French	-0.0238	**-3.248**
Yoruba	**Mal'ta**	**La Braña** 1	Italian	-0.0347	**-4.609**
Yoruba	**Mal'ta**	**La Braña** 1	Sardinian	-0.0460	**-6.222**
Yoruba	**Mal'ta**	**La Braña** 1	Tuscan	-0.0378	**-4.866**

Table S11. D statistics based on East Asians from the HGDP SNP dataset and the La Braña 1 and Mal'ta ancient genomes. For all comparisons, Yoruba was the outgroup population, H3 was the Mal'ta individual, H2 was the La Braña 1 individual, and H1 was an East Asian individual from the HGDP dataset. Statistically significant *D* statistics (indicated in bold) are those that have the absolute value of their *Z* score greater than 3.

Outgroup	**H3**	**H2**	**H1**	***D* statistic**	**Z score**
Yoruba	**Mal'ta**	**La Braña** 1	Han	-0.0585	**-7.295**
Yoruba	**Mal'ta**	**La Braña** 1	Han (North)	-0.0600	**-7.421**
Yoruba	**Mal'ta**	**La Braña** 1	Dai	-0.0569	**-7.037**
Yoruba	**Mal'ta**	**La Braña** 1	Daur	-0.0535	**-6.460**
Yoruba	**Mal'ta**	**La Braña** 1	Hezhen	-0.0529	**-6.507**
Yoruba	**Mal'ta**	**La Braña** 1	Lahu	-0.0585	**-7.008**
Yoruba	**Mal'ta**	**La Braña** 1	Miao	-0.0589	**-7.219**
Yoruba	**Mal'ta**	**La Braña** 1	Oroqen	-0.0535	**-6.728**
Yoruba	**Mal'ta**	**La Braña** 1	She	-0.0620	**-7.467**
Yoruba	**Mal'ta**	**La Braña** 1	Tujia	-0.0573	**-6.907**
Yoruba	**Mal'ta**	**La Braña** 1	Tu	-0.0555	**-6.808**
Yoruba	**Mal'ta**	**La Braña** 1	Xibo	-0.0513	**-6.420**
Yoruba	**Mal'ta**	**La Braña** 1	Yi	-0.0577	**-7.020**
Yoruba	**Mal'ta**	**La Braña** 1	Mongola	-0.0494	**-6.104**
Yoruba	**Mal'ta**	**La Braña** 1	Naxi	-0.0549	**-6.767**
Yoruba	**Mal'ta**	**La Braña** 1	Uygur	-0.0377	**-4.879**
Yoruba	**Mal'ta**	**La Braña** 1	Cambodian	-0.0592	**-7.434**
Yoruba	**Mal'ta**	**La Braña** 1	Japanese	-0.0576	**-7.348**
Yoruba	**Mal'ta**	**La Braña** 1	Yakut	-0.0390	**-4.986**

Table S12. *D* statistics based on Europeans from the HGDP SNP dataset and the La Braña 1 and Aboriginal Australian ancient genomes. For all comparisons, Yoruba was the outgroup population, H3 was the Aborigine individual, H2 was the La Braña 1 individual, and H1 was a European individual from the HGDP dataset. Statistically significant *D* statistics are those that have the absolute value of their *Z* score greater than 3.

Outgroup	**H3**	**H2**	**H1**	***D* statistic**	***Z* score**
Yoruba	Aborigine	La Braña 1	Orcadian	-0.0102	-1.350
Yoruba	Aborigine	La Braña 1	Adygei	-0.0109	-1.433
Yoruba	Aborigine	La Braña 1	Russian	-0.0012	-0.156
Yoruba	Aborigine	La Braña 1	Basque	-0.0138	-1.790
Yoruba	Aborigine	La Braña 1	French	-0.0130	-1.711
Yoruba	Aborigine	La Braña 1	Italian	-0.0148	-1.892
Yoruba	Aborigine	La Braña 1	Sardinian	-0.0183	-2.369
Yoruba	Aborigine	La Braña 1	Tuscan	-0.0191	-2.452

Table S13. *D* statistics based on East Asians from the HGDP SNP dataset and the La Braña 1 and Aboriginal Australian ancient genomes. For all comparisons, Yoruba was the outgroup population, H3 was the Aborigine individual, H2 was the La Braña 1 individual, and H1 was an East Asian individual from the HGDP dataset. Statistically significant *D* statistics (indicated in bold) are those that have the absolute value of their *Z* score greater than 3.

Outgroup	**H3**	**H2**	**H1**	***D* statistic**	***Z* score**
Yoruba	**Aborigine**	La Braña 1	**Han**	0.0564	**6.710**
Yoruba	**Aborigine**	La Braña 1	**Han (North)**	0.0536	**6.313**
Yoruba	**Aborigine**	La Braña 1	**Dai**	0.0588	**6.985**
Yoruba	**Aborigine**	La Braña 1	**Daur**	0.0539	**6.412**
Yoruba	**Aborigine**	La Braña 1	**Hezhen**	0.0546	**6.360**
Yoruba	**Aborigine**	La Braña 1	**Lahu**	0.0571	**6.560**
Yoruba	**Aborigine**	La Braña 1	**Miao**	0.0559	**6.636**
Yoruba	**Aborigine**	La Braña 1	**Oroqen**	0.0570	**6.587**
Yoruba	**Aborigine**	La Braña 1	**She**	0.0577	**6.679**
Yoruba	**Aborigine**	La Braña 1	**Tujia**	0.0573	**6.584**
Yoruba	**Aborigine**	La Braña 1	**Tu**	0.0490	**5.795**
Yoruba	**Aborigine**	La Braña 1	**Xibo**	0.0538	**6.426**
Yoruba	**Aborigine**	La Braña 1	**Yi**	0.0595	**6.964**
Yoruba	**Aborigine**	La Braña 1	**Mongola**	0.0467	**5.530**
Yoruba	**Aborigine**	La Braña 1	**Naxi**	0.0541	**6.271**
Yoruba	**Aborigine**	La Braña 1	**Uygur**	0.0259	**3.280**
Yoruba	**Aborigine**	La Braña 1	**Cambodian**	0.0553	**6.615**
Yoruba	**Aborigine**	La Braña 1	**Japanese**	0.0546	**6.498**
Yoruba	**Aborigine**	La Braña 1	**Yakut**	0.0481	**5.944**

Table S14. Functional analysis results for the La Braña 1 genome. This analysis was restricted to sites with a coverage minimum of four reads per position. The meaning of each functional category can be found at: http://www.ensembl.org/

All synonymous-coding sites	5,680
All missense SNPs	4,544
All canceledstart SNPs	8
All readthrough SNPs	4
All nonsense SNPs	16
NOVEL missense SNPs	98
NOVEL canceled start SNPs	1
NOVEL readthrough SNPs	0
NOVEL nonsense SNPs	1
SNPs in Splice Acceptor	8
SNPs in Splice Donor	9
SNPs in 5'UTR	2,460
SNPs in 3'UTR	7,288
SNPs in UTR-Splicesites	20
SNPs in Introns	343,378
Polyphen-2 Predictions	4,332
Polyphen-2 Probably damaging	443
Polyphen-2 Possibly damaging	360
Polyphen-2 Benign	3,529
SNPs within CpG Islands	8,068

Table S15. Stop codons within genes associated with Mendelian diseases in the La Braña 1 individual. Hom: homozygous, Het: heterozygous.

Gene	Chr.	Position	Nucleotide change	Protein position	Aminoacid change	Exon	Private	Hom/Het	Description	Mendelian Disease related to gene	OMIM ID
POMT1	chr9	134385435	C/T	251	Q/*	8/20	No	Het	Protein-O-mannosyltransferase 1	Walker-warburg syndrome, Muscular dystrophy, limb-girdle, type 2k	607423

Table S16. Stop codons occurring in genes not associated to Mendelian diseases but present as homozygous for the alternative allele in La Braña 1 individual.

Gene	Chr.	Position	Nucleotide change	Protein position	Aminoacid change	Exon	Private	Description
MOB3C	chr1	47080679	G/A	24	R/*	1/4	No	MOB kinase activator 3C
ZNF117	chr7	64438667	G/A	428	R/*	4/4	No	Zinc finger protein 117
OR5AR1	chr11	56431216	C/T	19	Q/*	1/1	No	Olfactory receptor, family 5, subfamily AR, member 1
CASP12	chr11	104763117	G/A	125	R/*	3/8	No	Caspase 12
USP29	chr19	57642782	C/A	913	Y/*	4/4	No	Ubiquitin specific peptidase 29

Table S17. Nonsynonymous changes within genes associated to Mendelian diseases in La Braña 1 individual. Hom: homozygous, Het: heterozygous.

Gene	Chr.	Position	Nucleotide change	Protein position	Aminoacid change	Exon	Grantham Score	GERP score	Private	Hom/Het	Description	Mendelian Disease related to gene	OMIM ID
PHKB	chr16	47727305	C/T	928	R/C	28/31	180	5,81	Yes	Het	Phosphorylase kinase, beta	Phosphorylase kinase deficiency of liver and muscle, autosomal recessive	172490

Table S18. Nonsynonymous changes occurring in genes not associated to Mendelian diseases but present as homozygous for the alternative allele in La Braña 1 individual. Hom: homozygous, Het: heterozygous.

Gene	Chr.	Position	Nucleotide change	Protein position	Aminoacid change	Exon	Grantham Score	GERP score	Private	Description
C1orf177	chr1	55273580	G/T	126	G/C	4/10	159	4,45	No	Chromosome 1 open reading frame 177
NBEAL2	chr3	47045846	C/T	2054	S/F	37/54	155	4,98	No	Neurobeachin-like 2
ADD1	chr4	2906707	G/T	460	G/W	10/15	184	5,55	No	Adducin 1 (alpha)
PRSS48	chr4	152201053	G/A	53	C/Y	2/5	194	5,2	No	Protease, serine, 48
WBSCR27	chr7	73249299	G/C	171	S/W	6/6	177	5,09	No	Williams Beuren syndrome chromosome region 27
CCDC129	chr7	31683410	G/A	809	C/Y	11/14	194	4,07	No	Coiled-coil domain containing 129
NRAP	chr10	115410234	T/C	249	Y/C	8/42	194	6,17	No	Nebulin-related anchoring protein
OR2D3	chr11	6942726	G/C	165	W/S	1/1	177	5,17	No	Olfactory receptor, family 2, subfamily D, member 3
OR56B1	chr11	5758062	T/C	106	C/R	1/1	180	4,78	No	Olfactory receptor, family 56, subfamily B, member 1
GEMIN4	chr17	648186	G/A	1033	R/C	2/2	180	5,71	No	Gem (nuclear organelle) associated protein 4
KRT37	chr17	39580739	C/A	13	G/C	1/7	159	4	No	Keratin 37

Table S19. La Braña 1 alleles for 10 SNPs with signatures of recent positive selection present in modern Europeans[19]. The ancestral state together with frequencies in Europeans for the derived allele are shown, as obtained from the 1000G Browser based on Ensembl v69.

SNP	Gene	Reference (hg19)	Ancestral	Derived	Frequency of the derived allele in Europeans (%)	La Braña 1	Coverage
rs16891982	*SLC45A2*	C	C	G	97	CC	6x
rs1426654	*SLC24A5*	A	G	A	100	GG	3x
rs9262	*C12orf29*	G	G	C	59	GG	5x
rs11755393	*UHRF1BP1*	A	G	A	65	AA	13x
rs2745098	*PTX4*	C	T	C	60	CC	4x
rs11881633	*TDRD12*	A	G	A	45	GA	10x,4x
rs1056899	*SETX*	T	C	T	73	CT	4x,2x
rs12918952	*WWOX*	G	G	A	60	A[a]	1x
rs10421769	*GPATCH1*	T	C	T	65	T[a]	1x
rs17310144	*CCDC14*	T	T	G	56	G	1x

[a] Possible *post-mortem* DNA damage.

Table S20. Probability (P) of finding La Braña 1 haplotypes in extant CEU and YRI populations. SNP are considered to be in LD with the selected SNP if r2 ≥ 0.8. When the SNP is heterozygous in La Braña1, probabilities are shown for haplotypes carrying both alleles ordered: Ancestral/Derived.

SNP	Gene	Allele state in La Braña 1	# SNP in LD in CEU	P (La Braña 1 haplotype in CEU)	# SNP in LD in YRI	P (La Braña 1 haplotype in YRI)
rs16891982	*SLC45A2*	*Ancestral*	0	-	0	-
rs1426654	*SLC24A5*	*Ancestral*	0	-	2	0.98
rs9262	*C12orf29*	*Ancestral*	1	0.41	1	0.85
rs11755393	*UHRF1BP1*	*Derived*	38	0.64	42	0.22
rs2745098	*PTX4*	*Derived*	6	0.55	3	0.27
rs11881633	*TDRD12*	*Ancestral/Derived*	20	0/0.5	9	0/0.23
rs1056899	*SETX*	*Ancestral/Derived*	2	0/0.74	2	0/0.20
rs12918952	*WWOX*	*Derived*	1	0.03	0	-
rs10421769	*GPATCH1*	*Derived*	6	0	0	-
rs17310144	*CCDC14*	*Derived*	5	0	4	0

Table S21. Additional SNPs related to pigmentation recovered in the La Braña 1 genome. The ancestral state of each SNP together with allele frequencies are shown, as obtained from the 1000G Browser based on Ensembl v69.

						Frequency (%) of the allele present in La Braña 1		
SNP	**Gene**	**Amino acid change**	**Ancestral**	**Derived**	**La Braña 1**	**Africa**	**Asia**	**Europe**
rs1800414	*OCA2*	His615Arg	T	C	T(3x)	100	41	100
rs1800407	*OCA2*	Arg419Gln	C	T	C(12x)	100	100	92
rs1545397	*OCA2*	-	A	T	A(1x)	97	8	92
rs4424881	*APBA2*		T	C	TC(1x,4x) [b]	10(C)	45(C)	87(C)
rs6119471	*ASIP*	-	G	C	C(3x)	36	100	100
rs2378249	*ASIP/PIGU*	-	A	G	A(6x)	86	85	88
rs4911442	*ASIP*	-	A	G	A(1x)[b]	100	100	91
rs1042602	*TYR*	Ser192Tyr	C	A	C(6x)	99	100	63
rs1126809	*TYR*	Arg402Gln	G	A	G(6x)	99	100	75
rs1393350	*TYR*	-	G	A	G(6x)	99	100	76
rs10831496	*GRM5*	-	G	A	G(2x)	80	76	31
rs683	*TYRP1*	-	A	C	A(1x)	13	1	63
rs2733831	*TYRP1*	-	A	G	G(3x)	6	1	60
rs1408799	*TYRP1*	-	T	C	C(1x)	22	1	66
rs1407995	*DCT*	-	T	C	C(1x)	82	24	82
rs12203592	*IRF4*	-	C	T	CT(1x,2x)	1(T)	0(T)	11(T)
rs12821256	*KITLG*	-	T	C	T(8x)	100	100	86
rs642742	*KITLG*	-	T	C	No data			
rs28777	*SLC45A2*	-	C	A	C(8x)	80	88	2
rs35395	*SLC45A2*	-	T	C	T(1x)[b]	76	88	2

rs2402130	*SLC24A4*	-	G	A	GA(2x,4x)	41(A)	93(A)	77(A)
rs12896399	*SLC24A4*	-	G	T	No data			
rs2470102	*SLC24A5*	-	G	A	G(8x)	92	75	1
rs4959270	*EXOC2*	-	C	A	C(7x)	81	66	53
rs1805009	*MC1R*	Arg294His	G	C	G(5x)	100	100	99
rs2228479	*MC1R*	Val92Met	G	A	G(11x)	100	76	93
rs885479	*MC1R*	Arg163Gln	G	A	G(6x)	99	35	92
N29insA	*MC1R*	-	C	A	C(10x)	-		
rs11547464	*MC1R*	Arg142His	G	A	G(7x)	100	100	99
rs885479	*MC1R*	Arg163Gln	G	A	G(6x)	99	35	92
rs1805008	*MC1R*	Arg160Trp	C	T	C(6x)	99	100	93
rs1805005	*MC1R*	Val60Leu	G	T	G(12x)	99	100	89
rs1805006	*MC1R*	Asp84Glu	C	A	C(11x)	100	100	99
rs1805007	*MC1R*	Arg151Cys	C	T	C(8x)	100	100	92
Y152OCH	*MC1R*	-	C	A	C(7x)	-		
rs1110400	*MC1R*	Ile155Thr	T	C	T(8x)	100	100	100
rs12913832	*OCA2-HERC2*	-	A	G	G(3x)	4	0	71
rs3935591	*OCA2-HERC2*	-	T	C	C(9x)	21	27	82
rs2240203	*OCA2-HERC2*	-	C	T	T(3x)	36	61	93
rs4778241	*OCA2-HERC2*	-	A	C	C(7x)	40	23	78
rs1129038	*OCA2-HERC2*	-	C	T	T(5x)	4	0	71
rs12593929	*OCA2-HERC2*	-	G	A	A(7x)	37	58	92
rs7183877	*OCA2-HERC2*	-	C	A	C(1x)	92	70	90
rs7170852	*OCA2-HERC2*	-	T	A	A(9x)	20	27	81
rs2238289	*OCA2-HERC2*	-	A	G	A(3x)	32	29	84
rs3940272	*OCA2-HERC2*	-	T	G	G(1x)	21	27	81
rs8028689	*OCA2-HERC2*	-	C	T	T(3x)	59	61	93
rs11631797	*OCA2-HERC2*	-	A	G	G(4x)	18	26	80

rs916977	*OCA2-HERC2*	-	T	C	C(9x)	10	26	81

[b] Possible *post-mortem* DNA damage.

Table S22. Number of minor alleles present in the La Braña 1 genome for the 24 SNPs in the HIrisPlex model[24], together with predicted probabilities[a].

SNP	Minor allele	Number of minor alleles
N29insA	A	0
rs11547464	A	0
rs885479	T	0
rs1805008	T	0
rs1805005	T	0
rs1805006	A	0
rs1805007	T	0
rs1805009	C	0
Y152OCH	A	0
rs2228479	A	0
rs1110400	C	0
rs28777	C	2
rs16891982	C	2
rs12821256	G	0
rs4959270	A	0
rs12203592	T	1
rs1042602	T	0
rs1800407	A	0
rs2402130	G	1
rs12913832	T	0
rs2378249	C	0
rs12896399	G	No data
rs1393350	T	0
rs683	G	0

[a] Black hair: 0.78, brown hair: 0.20, blond hair: 0.02, blue eyes: 0.706-0.458, intermediate eye color: 0.214-0.117, brown eyes: 0.328-0.177.

Table S23. Additional SNPs related to immunity recovered in the La Braña 1 genome.

Function	SNP	Gene	Amino acid change	Ancestral	Derived	La Braña 1	Frequency (%) in Europe[a]	Frequency (%) in Africa[a]
Pattern recognition receptors	rs4833095	*TLR1*	Asn248Ser	C	T	T(3x)	C=24, T=76	C=89, T=11
	rs5743618	*TLR1*	Ile602Ser	A	C	No data	A=25, C=75	A=95, C=5
	rs121917864	*TLR2*	Arg677Trp	C	T	C(2x)	-	-
	rs5743708	*TLR2*	Arg753Gln	G	A	G(7x), A(1x)[b]	G=97, A=3	G=100, A=0
	rs3775291	*TLR3*	Leu412Phe	C	T	T(3x)	C=68, T=32	C=96, T=4
	rs5743315	*TLR3*	-	C	A	C(3x)	C=99, A=1	C=86, A=14
	rs4986790	*TLR4*	Asp299Gly	A	G	A(1x)[b]	A=94, G=6	A=93, G=7
	rs4986791	*TLR4*	Thr399Ile	C	T	C(3x)	C=94, T=6	C=100, T=0
	rs5743810	*TLR6*	Pro249Ser	G	A	G(2x)	G=57, A=43	G=98, A=2
	rs179008	*TLR7*	Gln11Leu	A	T	A(2x)	A=74, T=26	A=87, T=13
	rs3764880	*TLR8*	Met1Val	G	A	A(2x)	G=27, A=73	G=31, A=69
	rs5743836	*TLR9*	-	G	A	G(2x), A(3x)	G=13, A=87	G=40, A=60
	rs11096957	*TLR10*	Asn241His	T	G	T(3x)	T=63, G=37	T=45, G=55
	rs2066842	*NOD2*	Pro268Ser	C	T	C(8x)	C=76, T=24	C=96, T=4
	rs9302752	*NOD2*	-	C	T	C(5x)	C=75, T=25	C=65, T=35
	rs3135499	*NOD2*	-	C	A	C(8x)	C=44, A=56	C=59, A=41
	rs8057341	*NOD2*	-	G	A	G(6x)	G=71, A=29	G=81, A=19
	rs2569190	*CD14*	-	G	A	A(6x)	G=53, A=47	G=65, A=35
	rs61752945	*RIG-1*	Arg546Gln	C	T	C(2x)	C=99, T=1	C=98, T=2

	rs17217280	*RIG-1*	Asp580Glu	A	T	A(3x), T(2x)	A=86, T=14	A=99, T=1
	rs2074158	*LGP2*	Gln425Arg	T	C	C(9x)	T=82, C=18	T=22, C=78
	rs10930046	*IFIH1/MDA5*	Arg460His	C	T	T(4x)	C=1, T=99	C=48, T=52
	rs4804803	*DC-SIGN*	-	G	A	No data	G=22, A=78	G=41, A=59
Intracellular adaptor molecules	rs42490	*RIPK2*	-	G	A	G(4x)	G=39, A=61	G=29, A=71
	rs1873613	*LRRK2*	-	T	C	C(5x)	T=29, C=71	T=52, C=48
	rs4251545	*IRAK4*	Ala428Thr	G	A	G(6x), A(4x)	G=92, A=8	G=67, A=33
	rs8177374	*Mal/TIRAP*	Ser180Leu	C	T	C(3x), T(6x)	C=83, T=17	C=98, T=2
	rs3802814	*Mal/TIRAP*	-	G	A	G(4x), A(6x)	G=84, A=16	G=98, A=2
	rs3802813	*Mal/TIRAP*	Ser55Asn	G	A	G(10x), A(5x)	G=96, A=4	G=94, A=7
	rs3184504	*SH2B3*	Arg262Trp	C	T	C(7x)	C=53, T=47	C=97, T=3
Intracellular modulators	rs10745657	*SOCS2*	-	A	G	G(7x)	A=48, G=52	A=87, G=13
	rs7137054	*SOCS2*	-	C	G	C(5x)	C=84, G=16	C=58, G=42
	rs3750920	*TOLLIP*	-	C	T	T(8x)	C=54, T=46	C=72, T=28
	rs5743899	*TOLLIP*	-	C	T	T(6x)	C=21, T=79	C=42, T=58
	rs497116	*CASP12*	Arg125Stop	G	A	A(4x)	G=0, A=100	G=16, A=84
Cytokine and cytokine receptors	rs1800629	*TNF*	-	G	A	G(5x)	G=86, A=14	G=90, A=10
	rs361525	*TNF*	-	G	A	G(7x)	G=93, A=7	G=97, A=3
	rs10982385	*TNFSF15*	-	T	G	T(7x)	T=41, G=59	T=78, G=22
	rs4574921	*TNFSF15*	-	T	C	T(1x)[b], C(3x)	T=74, C=26	T=93, C=7
	rs10114470	*TNFSF15*	-	C	T	C(1x), T(4x)	C=68, T=32	C=89, T=11
	rs6478108	*TNFSF15*	-	T	C	T(3x), C(1x)	T=68, C=32	T=87, C=13
	rs3753344	*TNFRSF18*	-	G	A	G(9x)	G=95, A=5	G=80, A=20

	rs2043055	*IL-18BP*	-	A	G	A(2x)	A=63, G=37	A=50, G=50
	rs2039381	*IFNE*	Gln71Stop	G	A	G(4x)	G=99.74, A=0.26	G=92, A=8
	rs30461	*IL-29*	Asn188Asp	G	A	A(2x)	G=11, A=89	G=68, A=32
	rs2069727	*IFNG*	-	T	C	C(6x)	T=54, C=46	T=84, C=16
	rs2430561	*IFNG*	-	T	A	A(3x)	T=57, A=43	T=82, A=19
	rs11754268	*IFNGR1*	-	C	T	C(3x), T(6x)	C=77, T=23	C=93,T=7
	rs1800872	*IL-10*	-	G	T	G(1x), T(2x)	G=77, T=23	G=58, T=42
	rs900	*TGFB2*	-	A	T	A(3x)	A=73, T=27	A=65, T=35
	rs1891467	*TGFB2*	-	G	A	A(5x)	G=23, A=77	G=56, A=44
Chemokine and chemokine receptors	rs2015086	*CCL18*	-	G	A	A(1x)[b]	G=15, A=85	G=45, A=55
	rs2015070	*CCL18*	-	C	T	C(6x)	C=90, T=10	C=98, T=2
	rs14304	*CCL18*	-	C	T	C(1x), T(4x)	C=68, T=32	C=98, T=2
	rs4073	*IL-8*	-	A	T	A(1x), T(2x)	A=41, T=59	A=84, T=16
	rs2228428	*CCR4*	-	C	T	C(13x)	C=70, T=30	C=95, T=5
	rs333 (CCR5-delta32)	*CCR5*	-	-	-	1 read without the deletion	-	-
	rs1024611	*CCL2/MCP-1*	-	A	G	A(2x)	A=67, G=33	A=76, G=24
Effector molecules	rs8078340	*NOS2A*	-	G	A	G(6x)	G=87, A=13	G=78, A=22
	rs2779249	*NOS2A*	-	C	A	C(7x)	C=68, A=32	C=59, A=41
	rs2274894	*NOS2A*	-	G	T	G(3x), T(3x)	G=60, T=40	G=87, T=13
	rs7215373	*NOS2A*	-	C	T	T(3x)	C=46, T=54	C=83, T=17
	rs2333227	*MPO*	-	C	T	C(3x)	C=78, T=22	C=66, T=34

[a] From the 1000G Browser based on Ensembl v69.
[b] Possible *post-mortem* DNA damage.

Table S24. Expression evidence in the La Braña 1 individual for eQTLs from selected regions in Europeans.

Reg	Chr	Gene	eQTL SNP	Hg19	Exp. All[a]	Z-score[a]	FDR[a]	Anc	Der	Der[b]	Anc_Br[c]	Der_Br[c]
1	2	*MRPL53*	rs3771744	74578536	A	-23.13	0	A	G	↑	-	5X
		MTHFD2	rs3771744	74578536	A	-3.81	0.05	A	G	↑	-	5X
		LBX2	rs3771744	74578536	A	2.99	0.45	A	G	↓	-	5X
	La Braña 1 shows higher expression of *MRPL53* and *MTHFD2*, and lower expression of *LBX2*											
2	2	*MCM6*	rs7570283	136447707	C	15.44	0	T	C	↑	2X	1X
			rs584226	136641382	C	18.09	0	T	C	↑	1X[d]	-
			rs16855656	136642433	G	18.08	0	A	G	↑	-	-
			rs309126	136643900	C	18.25	0	T	C	↑	3X	2X
			rs3099429	136655582	T	13.49	0	A	T	↑	-	8X
		DARS	rs584226	136641382	C	-6.23	0	T	C	↓	1X[d]	-
			rs16855656	136642433	G	-6.23	0	A	G	↓	-	-
			rs309126	136643900	C	-5.49	0	T	C	↓	3X	2X
			rs3099429	136655582	A	27.11	0	A	T	↓	-	8X
		CXCR4	rs3099429	136655582	A	-4.79	0	A	T	↑	-	8X
	Higher expression of *CXCR4*, and likely heterozygous at several eQTL for *MCM6* and *DARS*											
3	Region around chr2:~178.2 Mb failed to replicate in Westra et al.											
4	4	*TMEM156*	rs2711968	39023992	C	11.6	0	G	C	↑	2X	-
			rs12642359	39027082	T	-8.8	0	T	C	↑	-	2X
			rs2254075	39030183	A	17.46	0	G	A	↑	4X	-
			rs2254212	39031842	C	17.44	0	T	C	↑	2X	-
			rs2711976	39034786	G	17.39	0	C	G	↑	1X	-
			rs2566094	39036374	A	17.24	0	G	A	↑	4X	-

rs2566097	39037336	C	16.99	0	T	C	↑	2X	-
rs2711981	39039258	C	17.08	0	T	C	↑	10X	-
rs2711982	39043997	T	12.41	0	C	T	↑	-	-
rs2711983	39044920	G	12.77	0	A	G	↑	3X	-
rs6813885	39045907	T	12.77	0	T	C	↓	-	4X
rs1430362	39048464	G	12.71	0	C	G	↑	1X	-
rs2566114	39052210	A	12.73	0	G	A	↑	3X	-
rs2566116	39052838	T	12.73	0	T	C	↓	-	2X
rs2566117	39053022	G	12.73	0	G	A	↓	-	2X
rs2566119	39054643	G	12.72	0	G	T	↓	-	2X
rs2711956	39055854	T	12.72	0	C	T	↑	4X	-
rs2711955	39055891	C	12.69	0	G	C	↑	3X	-
rs2711954	39056268	G	12.69	0	G	A	↓	-	2X
rs2711953	39056721	G	12.7	0	G	C	↓	2X	-
rs2711952	39056818	A	12.69	0	G	A	↑	9X	-
rs1367298	39057666	C	12.7	0	T	C	↑	3X	-
rs1367299	39057791	G	12.7	0	A	G	↑	3X	-
rs2711950	39058337	A	12.7	0	A	G	↓	-	3X
rs2566122	39058365	T	12.7	0	G	T	↑	2X	-
rs2566124	39060991	T	12.7	0	C	T	↑	2X	-
rs2711946	39061079	G	12.7	0	C	G	↑	3X	-
rs2566126	39061820	T	12.69	0	C	T	↑	-	-
rs2711943	39062436	T	12.69	0	C	T	↑	2X	-
rs1367301	39062805	A	12.69	0	G	A	↑	1X	-
rs2566128	39063377	C	12.69	0	T	C	↑	2X	-
rs2711941	39064162	A	12.69	0	C	A	↑	1X	-
rs2711940	39065091	G	12.69	0	A	G	↑	-	-
rs1036043	39066865	T	12.69	0	T	G	↓	-	1X

	rs891434	39070630	T	12.65	0	T	C	↓	-	6X
	rs1006403	39073908	A	12.67	0	G	A	↑	8X	-
	rs2566138	39075945	C	12.67	0	T	C	↑	1X[d]	-
	rs2712012	39077136	G	12.66	0	A	G	↑	2X	-
	rs2252519	39077904	G	12.66	0	A	G	↑	1X[d]	-
	rs2566142	39081681	G	11.87	0	A	G	↑	4X	-
	rs1036044	39084992	A	11.48	0	C	A	↑	1X	-
	rs2566146	39089450	A	11.58	0	A	G	↓	1X[d]	2X
	rs2711997	39102010	C	11.65	0	C	T	↓	-	7X
WDR19	rs2566097	39037336	C	-5.06	0	T	C	↓	2X	-
	rs2711981	39039258	C	-5.2	0	T	C	↓	10X	-
	rs2711982	39043997	T	-7.37	0	C	T	↓	-	-
	rs2711983	39044920	G	-8.74	0	A	G	↓	3X	-
	rs6813885	39045907	T	-8.74	0	T	C	↑	-	4X
	rs1430362	39048464	G	-8.68	0	G	C	↑	-	1X
	rs2566114	39052210	A	-8.67	0	G	A	↓	3X	-
	rs2566116	39052838	T	-8.67	0	T	C	↑	-	2X
	rs2566117	39053022	G	8.66	0	G	A	↓	-	2X
	rs2566119	39054643	G	-8.67	0	G	T	↑	-	2X
	rs2711956	39055854	T	-8.66	0	C	T	↓	4X	-
	rs2711955	39055891	C	-8.63	0	G	C	↓	3X	-
	rs2711954	39056268	G	-8.63	0	G	A	↑	-	2X
	rs2711953	39056721	G	-8.63	0	G	C	↑	2X	-
	rs2711952	39056818	A	-8.63	0	G	A	↓	9X	-
	rs1367298	39057666	C	-8.64	0	T	C	↓	3X	-
	rs1367299	39057791	G	-8.64	0	A	G	↓	3X	-

	rs2711950	39058337	A	-8.64	0	A	G	↑	-	3X
	rs2566122	39058365	T	-8.64	0	G	T	↓	2X	-
	rs2566124	39060991	T	-8.63	0	C	T	↓	2X	-
	rs2711946	39061079	G	-8.63	0	C	G	↓	3X	-
	rs2566126	39061820	T	-8.63	0	C	T	↓	-	-
	rs2711943	39062436	T	-8.62	0	C	T	↓	2X	-
	rs1367301	39062805	A	-8.63	0	G	A	↓	1X	-
	rs2566128	39063377	C	-8.63	0	T	C	↓	2X	-
	rs2711941	39064162	A	-8.64	0	C	A	↓	1X	-
	rs2711940	39065091	G	-8.65	0	A	G	↓	-	-
	rs1036043	39066865	T	-8.66	0	T	G	↑	-	1X
	rs891434	39070630	T	-8.65	0	T	C	↑	-	6X
	rs1006403	39073908	A	-8.67	0	G	A	↓	8X	-
	rs2566138	39075945	C	-8.67	0	T	C	↓	1X[d]	-
	rs2712012	39077136	G	-8.67	0	A	G	↓	2X	-
	rs2252519	39077904	G	-8.67	0	A	G	↓	1X[d]	-
	rs2566142	39081681	G	-9.06	0	A	G	↓	4X	-
	rs1036044	39084992	A	-9.3	0	C	A		1X	-
	rs2566146	39089450	A	-9.37	0	A	G	↑	1X[d]	2X
	rs2711997	39102010	C	-9.41	0	C	T	↑	-	7X
KLHL5	rs2254075	39030183	A	-5.7	0	G	A	↓	4X	-
	rs2254212	39031842	C	-5.72	0	T	C	↓	2X	-
	rs2711976	39034786	G	-5.72	0	C	G	↓	1X	-
	rs2566094	39036374	A	-5.73	0	G	A	↓	4X	-
	rs2566097	39037336	C	-5.67	0	T	C	↓	2X	-
	rs2711981	39039258	C	-5.69	0	T	C	↓	10X	-

	La Braña 1 shows lower expression of *TMEM156* and higher expression of *WDR19* and *KLHL5*											
5	5	*AMACR*	rs250414	33990623	C	4.69	0	C	T	↓	2X	-
			rs7721230	33992529	A	5.2	0	A	G	↓	5X	-
			rs10472909	33994116	A	4.25	0.01	T	A	-	-	-
	La Braña 1 shows eQTLs for higher expression of *AMACR*											
6	5	*NNT*	rs3776423	43487874	G	-11.68	0	T	G	-	-	-
		PAIP1	rs3776423	43487874	G	-2.96	0.47	T	G	-	-	-
	Lack of reads for rs3776423											
7	6	*NCUBE1*	rs6454741	89873182	C	2.98	0.46	T	C	↑	4X	-
		PNRC1	rs1757815	89879296	A	6.17	0	T	A	↑	-	2X
			rs404943	89891582	C	3.38	0.19	T	C	↑	1X[d]	-
		PM20D2	rs1757815	89879296	A	3.54	0.13	T	A	↑	-	2X
			rs9362624	89894119	T	-3.02	0.43	G	T	↓	6X	-
			rs7751440	89897536	T	-3.31	0.23	T	C	↑	-	3X
			rs9362625	89898097	G	-3.05	0.4	T	G	↓	5X	-
			rs452667	89903320	T	-3.16	0.33	C	T	↓	7X	-
	La Braña 1 harbors alleles for higher expression of *RM20D2* and lower expression of *NCUBE1*											
8	Region around chr7:~150.7 Mb failed to replicate in Westra et al.											
9	9	*C9orf156*	rs2417575	100628642	G	-14.73	0	A	G	↓	3X	-
			rs2417576	100628707	C	-14.77	0	C	T	↑	-	2X
			rs7357707	100630949	T	30.96	0	G	T	-	-	-
			rs1443436	100631298	T	-14.77	0	A	T	↓	6X	-
			rs7033315	100636240	G	30.96	0	C	G	↑	8X	-
			rs10119795	100637339	T	31	0	C	T	↑	6X	-
			rs10984103	100639275	A	-14.76	0	A	C	↑	-	12X
			rs2120264	100645728	A	-14.76	0	A	G	↑	-	2X
			rs7853349	100650259	T	32.81	0	A	T	↑	9X	-

			rs12238579	100652058	C	32.81	0	G	C	↑	3X	1X
			rs6586	100666931	T	32.78	0	C	T	↑	-	-
			rs1561958	100669799	T	32.64	0	G	T	↑	2X	-
			rs2282192	100672338	T	24.8	0	C	T	↑	3X	-
			rs2120263	100675944	T	24.72	0	C	T	↑	2X	-
		HEMGN	rs7357707	100630949	T	-4.29	0.01	G	T	-	-	-
			rs7033315	100636240	G	-4.29	0.01	C	G	↓	8X	1X
			rs10119795	100637339	T	-4.28	0.01	C	T	↓	6X	-
			rs7853349	100650259	T	-4.86	0	A	T	↓	9X	-
			rs12238579	100652058	C	-4.86	0	G	C	↓	3X	1X
			rs6586	100666931	T	-4.89	0	C	T	-	-	-
			rs1561958	100669799	T	-5.03	0	G	T	↓	2X	-
			rs2282192	100672338	T	-6.18	0	C	T	↓	3X	-
			rs2120263	100675944	T	-6.21	0	C	T	↓	2X	-
		CORO2A	rs10984103	100639275	A	4.7	0	A	C	↓	-	12X
			rs2120264	100645728	A	4.67	0	A	G	↓	-	2X
La Braña 1 harbors alleles for higher expression of *HEMGN* and lower expression of *CORO2A*												
10	11	*TSPAN32*	rs756920	2310870	G	11.76	0	G	C	↓	-	9X
			rs7941928	2315404	A	22.65	0	G	A	↑	1X	5X
			rs2244891	2319355	C	18.02	0	C	T	↓	-	9X
			rs2074021	2323528	T	16.64	0	T	C	↓	-	14X
			rs10831733	2324440	T	18.07	0	C	T	↑	6X	6X
			rs937614	2326920	G	15.41	0	C	G	↑	2X	4X
			rs800140	2333675	A	19.24	0	A	G	↓	-	9X
			rs2651803	2349462	G	18.82	0	G	A	↓	-	4X
			rs2651802	2350322	A	18.82	0	G	A	↑	9X	-
		ASCL2	rs756920	2310870	G	-10.5	0	G	C	↑	-	9X
			rs7941928	2315404	A	10.14	0	G	A	↑	1X	5X

			rs2244891	2319355	C	4.75	0	C	T	↓	-	9X
			rs2074021	2323528	T	-10.25	0	T	C	↑	-	14X
			rs10831733	2324440	T	10.92	0	C	T	↑	6X	6X
			rs937614	2326920	G	9.18	0	C	G	↑	2X	4X
			rs800140	2333675	A	6.34	0	A	G	↓	-	9X
			rs2651803	2349462	G	5.93	0	G	A	↓	-	4X
			rs2651802	2350322	A	5.91	0	G	A	↑	9X	-
		CD81	rs7941928	2315404	A	4.69	0	G	A	↑	1X	5X
			rs2244891	2319355	C	8.99	0	C	T	↓	-	9X
			rs10831733	2324440	T	4.56	0	C	T	↑	6X	6X
			rs937614	2326920	G	4.92	0	C	G	↑	2X	4X
			rs800140	2333675	A	9.2	0	A	G	↓	-	9X
			rs2651803	2349462	G	8.89	0	G	A	↓	-	4X
			rs2651802	2350322	A	8.91	0	G	A	↑	9X	-
	La Braña 1 likely heterozygous at several eQTLs for *TSPAN32*, *ASCL2* and *CD81*											
11	17	*CARKL*	rs161370	3530253	G	-13.4	0	T	G	↓	-	3X
			rs222790	3531490	A	-13.5	0	G	A	↓	-	9X
		CTNS	rs161370	3530253	G	-5.22	0	T	G	↓	-	3X
			rs222790	3531490	A	-5.18	0	G	A	↓	-	9X
		TAX1BP3	rs161370	3530253	G	2.95	0.48	T	G	↑	-	3X
			rs222790	3531490	A	2.98	0.45	G	A	↑	-	9X
	Lower expression of *CARKL* and *CTNS* and higher expression of *TAX1BP3*											
12	Region around chr19: ~20.9 Mb failed to replicate in Westra											
13	19	*SPINT2*	rs7253245	38753786	T	-9.07	0	T	G	↑	-	8X
			rs3786870	38755864	T	-9.17	0	C	T	↓	7X	-

		RASGRP4	rs7253245	38753786	T	-4.94	0	T	G	↑	-	8X
			rs3786870	38755864	T	-5.04	0	C	T	↓	7X	-
		C19orf33	rs7253245	38753786	T	4.9	0	T	G	↓	-	8X
			rs3786870	38755864	T	4.81	0	C	T	↑	7X	-
		PPP1R14A	rs7253245	38753786	T	4.34	0.01	T	G	↓	-	8X
			rs3786870	38755864	T	4.39	0.01	C	T	↑	7X	-
	Lower expression of *C19orf33* and *PPP1R14A*, and higher expression of *SPINT2* and *RASGRP4*											
14	20	*EDEM2*	rs6088713	33693857	T	5.88	0	C	T	-	-	-
			rs6088721	33706011	A	6.05	0	A	C	↓	6X	-
			rs2273805	33706575	T	6.12	0	T	A	↓	8X	-
			rs2065108	33706822	C	6.12	0	C	T	↓	3X	-
			rs6088724	33712915	C	6.12	0	C	T	↓	7X	-
			rs6088727	33713639	G	6.1	0	G	A	↓	4X	-
			rs6087677	33714805	T	6.11	0	T	C	↓	2X	-
			rs1535466	33718706	G	6.46	0	A	G	↑		2X
			rs3746427	33730464	A	5.75	0	G	A	↑	5X	1X[d]
			rs1415771	33734493	A	-6.35	0	G	A	↓	-	9X
			rs945959	33734905	C	-6.36	0	C	G	-	-	-
			rs1124511	33736697	C	-6.39	0	C	A	↑	2X	-
			rs6088747	33754604	G	5.82	0	T	G	↑	5X	-
		EIF6	rs6088713	33693857	T	-3.61	0.1	C	T	-	-	-
			rs6088721	33706011	A	10.38	0	A	C	↓	6X	-
			rs2273805	33706575	T	10.4	0	T	A	↓	8X	-
			rs2065108	33706822	C	10.4	0	C	T	↓	3X	-
			rs6088724	33712915	C	10.4	0	C	T	↓	7X	-
			rs6088727	33713639	G	10.41	0	G	A	↓	4X	-
			rs6087677	33714805	T	10.42	0	T	C	↓	2X	-

	rs1535466	33718706	G	-3.44	0.17	A	G	↓	-	2X
	rs3746427	33730464	A	-9.15	0	G	A	↓	5X	1X[d]
	rs1415771	33734493	A	-4.33	0.01	G	A	↓	-	9X
	rs945959	33734905	C	-4.44	0	C	G	-	-	-
	rs1124511	33736697	C	-4.46	0	C	A	↑	2X	-
	rs6088747	33754604	G	-8.92	0	T	G	↓	5X	-
ACSS2	rs6088713	33693857	T	-5.34	0	C	T	-	-	-
	rs6088721	33706011	A	-7.87	0	A	C	↑	6X	-
	rs2273805	33706575	T	-7.87	0	T	A	↑	8X	-
	rs2065108	33706822	C	-7.87	0	C	T	↑	3X	-
	rs6088724	33712915	C	-7.85	0	C	T	↑	7X	-
	rs6088727	33713639	G	-7.85	0	G	A	↑	4X	-
	rs6087677	33714805	T	-7.82	0	T	C	↑	2X	-
	rs1535466	33718706	G	-4.08	0.02	A	G	↓	2X	-
	rs3746427	33730464	A	9.56	0	G	A	↑	5X	1X[d]
	rs1415771	33734493	A	-6.52	0	G	A	↓	-	9X
	rs945959	33734905	C	-6.58	0	C	G	-	-	-
	rs1124511	33736697	C	-5.75	0	C	A	↑	2X	-
	rs6088747	33754604	G	9.91	0	T	G	↑	5X	-
TRPC4AP	rs6088721	33706011	A	-7.33	0	A	C	↑	6X	-
	rs2273805	33706575	T	-7.44	0	T	A	↑	8X	-
	rs2065108	33706822	C	-7.44	0	C	T	↑	3X	-
	rs6088724	33712915	C	-7.42	0	C	T	↑	7X	-
	rs6088727	33713639	G	-7.4	0	G	A	↑	4X	-
	rs6087677	33714805	T	-7.42	0	T	C	↑	2X	-
	rs1535466	33718706	G	-3.44	0.22	A	G	↓	2X	-
	rs1415771	33734493	A	9.45	0	G	A	↑	-	9X

rs945959	33734905	C	9.42	0	C	G	-	-	-
rs1124511	33736697	C	9.47	0	C	A	↓	2X	-

Lower expression of *ACSS2*, higher expression of *TRPC4AP*. Mixed evidence for *EDEM2* and *EIF6*

[a] Exp.all indicates the minor allele tested in Westra et al.[90] study, along with its Z-score and False Discovery Rate.

[b] Direction of expression effect for derived allele in Westra et al.[90] study.

[c] Number of reads found covering the ancestral and derived allele in the La Braña 1 individual.

[d] Possible post-mortem DNA damage.

Table S25. Expression evidence in the La Braña 1 individual for eQTLs of 40 genes that influence susceptibility to infections.

Function	Gene	Chr	eQTL SNP	Hg19	Exp. All[a]	Z-score[a]	FDR[a]	Anc	Der	Der[b]	Anc_Br[c]	Der_Br[c]
Pattern recognition receptors	*TLR1*	4	rs2101521	38811551	A	-11.61	0	A	G	↑	-	4X
			rs6815814	38816338	C	-11.04	0	C	A	↑	-	5X
			rs17616434	38812876	C	-11.03	0	C	T	↑	-	5X
			rs4543123	38792524	G	-10.8	0	G	A	↑	-	4X
			rs10004195	38784724	A	-10.76	0	A	T	↑	-	1X
		La Braña 1 shows high expression of *TLR1*										
	TLR2	4	rs1816702	154609523	T	-5.34	0	T	C	↑	-	6X
			rs4696483	154619255	T	-5.21	0	T	C	↑	-	1X
			rs17030307	154588602	G	-3.88	0.04	A	G	↓	10X	-
			rs2289318	154633734	C	-3.83	0.05	G	C	↓	3X	3X
			rs7656411	154627655	G	3.73	0.07	G	T	↓	-	1X
		La Braña 1 is homozygous for higher expression at three top eQTLs for *TLR2*										
	TLR3	4	rs4452470	186948168	G	-4.65	0	A	G	↓	-	3X
			rs11938703	186944884	T	4.03	0.02	C	T	↑	4X	2X
			rs12646803	186964309	A	-3.98	0.03	-	-	-	-	-
			rs4862625	186946612	T	3.93	0.03	A	T	↑	3X	-
			rs2139315	186954900	T	-3.82	0.05	A	T	↓	5X	-
		La Braña 1 shows mixed evidence for eQTLs of *TLR3* expression										
	TLR4	9	rs2770150	120463139	G	-34.4	0	A	G	↓	6X	-
			rs11789302	120459526	G	-34.39	0	T	G	↓	4X	-
			rs2737191	120462715	G	-34.38	0	A	G	↓	2X	-
			rs10491851	120444525	A	-34.22	0	C	A	↓	4X	-
			rs13296664	120446696	T	-34.18	0	C	T	↓	1X	-

	La Braña 1 shows high expression of *TLR4*										
TLR6	4	rs2174284	38579158	T	-7.61	0	-	-	-	-	-
		rs10776482	38494604	G	-7.56	0	G	A	↑	-	2X
		rs10776483	38494859	G	-7.56	0	A	G	↓	2X	-
		rs11096956	38495999	A	-7.55	0	A	C	↑	1X	-
		rs17582830	38587246	G	-7.48	0	A	G	↓	6X	-
	La Braña 1 shows mixed evidence in *TLR6*-associated eQTLs										
TLR7	-[d]										
TLR8	-[d]										
TLR9	-[d]										
TLR10	-[d]										
CARD15	16	rs9302752	50719103	T	62.66	0	C	T	↑	5X	-
		rs2287195	50714979	T	62.64	0	C	T	↑	8X	-
		rs1420685	50719674	T	62.61	0	T	C	↓	-	4X
		rs9933594	50724087	G	62.47	0	C	G	↑	3X	-
		rs7202124	50714029	G	62.23	0	G	A	↓	-	2X
	La Braña 1 shows eQTL genotypes for lower expression of *NOD2/CARD15*										
CD14	5	rs778587	140006430	T	20.4	0	C	T	↑	8X	-
		rs7711117	139988753	T	-20.02	0	C	T	↓	6X	1X[e]
		rs12517200	140019921	G	19.68	0	A	G	↑	6X	-
		rs2563298	140011315	A	19.64	0	C	A	↑	4X	-
		rs3822356	140022436	G	-19.63	0	A	G	↓	7X	-
	La Braña 1 harbours a composite of expression increasing / decreasing alleles										
RIG-1 / LGP2	-[d]										
IFIH1/ MDA5	-[d]										
DC-SIGN	-[d]										

Intracellular adaptor molecules	*RIPK2*	8	rs13274635	90758337	G	-15.52	0	G	A	↑	3X	-
			rs428484	90776033	C	-15.4	0	C	T	↑	3X	-
			rs39766	90811369	C	-15.39	0	T	C	↓	-	2X
			rs411310	90808193	A	-15.38	0	A	T	-	-	-
			rs42343	90808212	A	-15.38	0	C	A	-	-	-
		La Braña 1 shows low expression of *RIPK2*										
	LRRK2	12	rs10748014	40615069	T	-8.03	0	C	T	↓	6X	3X
			rs10878224	40614601	T	-8.02	0	C	T	↓	2X	2X
			rs12827541	40609979	T	-7.98	0	C	T	↓	1X	1X[e]
			rs10878222	40610718	A	-7.98	0	G	A	-	-	-
			rs10878220	40608880	T	-7.95	0	C	T	↓	3X	3X
		La Braña 1 heterozygous at LRRK2 eQTL *SNPs*										
	IRAK4	12	rs11182244	44138539	C	-7.56	0	T	C	↓	-	1X
			rs4251535	44179210	C	-7.4	0	G	C	↓	3X	2X
			rs4251444	44158665	G	-7.39	0	C	G	↓	1X	2X
			rs4251446	44158875	A	-7.38	0	G	A	↓	2X	-
			rs4251425	44153909	T	-7.38	0	C	T	↓	4X	9X
		La Braña 1 is heterozygous at many eQTLs that control *IRAK4* expression										
	TIRAP	11	rs588361	126084834	G	6.01	0	G	A	↓	-	2X
			rs580145	126079939	C	6.01	0	C	T	↓	-	1X[e]
			rs2282580	126081403	C	-5.88	0	T	C	↓	-	10X
			rs3851111	126083218	T	-5.88	0	G	T	↓	-	2X
			rs1893352	126160687	G	-5.61	0	G	A	↑	3X	3X
		La Braña 1 shows eQTL alleles that lower *TIRAP* expression										
	SH2B3	12	rs2239195	111881309	T	-12.1	0	G	T	↓	5X	-
			rs848132	111989979	A	-11.71	0	C	A	↓	3X	-

			rs1029388	111926901	C	-11.67	0	C	T	↑	-	1X[e]
			rs2339816	111912188	C	-11.63	0	T	C	↓	2X	-
			rs607316	111969448	T	-11.6	0	T	C	↑	-	1X
		La Braña 1 shows high expression of *SH2B3*										
Intracellular modulators	*SOCS2*	-[d]										
	TOLLIP	12	rs3793966	1302948	T	-9.02	0	C	T	↑	-	8X
			rs3829223	1300406	T	-8.99	0	C	T	↑	-	9X
			rs4963031	1264823	T	-7.04	0	T	C	-	-	-
			rs12285466	1293201	T	-6.53	0	C	T	↑	8X	-
			rs3793964	1301982	T	6.25	0	C	T	↓	6X	-
		La Braña 1 shows tendency towards higher expression of *TOLLIP*										
	CASP12	-[d]										
Cytokine and cytokine receptors	*TNF*	6	rs2516479	31528326	G	16.94	0	G	C	↓	1X	-
			rs2239704	31540141	A	16.92	0	C	A	↑	5X	2X
			rs928815	31531215	T	16.9	0	G	T	↑	-	1X
			rs2857602	31533378	G	16.9	0	A	G	↑	2X	5X
			rs2844484	31536224	A	16.83	0	G	A	↑	2X	-
		La Braña 1 shows a mixture of increasing/decreasing alleles at eQTLs for *TNF* expression										
	TNFSF15	-[d]										
	TNFRSF18	1	rs7515488	1163804	T	-15.05	0	C	T	↓	6X	-
			rs11721	1152631	A	-14.7	0	A	C	↑	-	12X
			rs3813199	1158277	A	-13.29	0	G	A	↓	5X	1X[e]
			rs3766186	1162435	A	-13.29	0	C	A	↓	7X	-
			rs6603783	1181751	C	-12.59	0	C	T	↑	-	8X
		La Braña 1 harbours eQTL alleles that increase *TNFRSF18* expression										
	IL18BP	-[d]										

	IFNE	-[d]										
	IL29	19	rs570880	39774530	T	-7.72	0	A	T	↓	-	-
			rs1375910	39850982	A	-5.48	0	G	A	↓	5X	2X
			rs16973374	39869540	G	-5.48	0	G	C	↑	-	10X
			rs36265	39862881	T	-4.5	0	T	C	↑	-	8X
			rs1368442	39835169	G	-3.9	0.04	G	T	↑	-	4X
		La Braña 1 shows high expression of *IL29*										
	IFNG	12	rs12818107	68483259	T	4.06	0.02	C	T	↑	-	-
			rs11610401	68487317	A	4.04	0.02	T	A	↑	3X	-
			rs11610754	68486587	G	4.04	0.02	C	G	↑	4X	-
			rs12825700	68492980	A	4.04	0.02	G	A	↑	5X	-
			rs1558744	68504592	A	4.04	0.02	A	G	↓	-	2X
		La Braña 1 shows low expression of *IFNG*										
	IFNGR-CD119	-[d]										
	IL10	-[d]										
	TGFB2	-[d]										
Chemokine and chemokine receptors	*CCL18*	-[d]										
	IL8	12	rs3184504	111884608	T	-5.32	0.01	C	T	↓	7X	-
			rs653178	112007756	C	-5.26	0.02	T	C	↓	4X	-
			rs10774625	112072424	A	-5.17	0.03	G	A	↑	-	8X
			rs11065987	111910219	G	-4.97	0.06	G	A	↑	-	-
		La Braña 1 shows high expression of *IL8*										
	CCR4	3	rs7613051	33065339	A	6.02	0	A	G	↓	3X	5X
			rs13080709	33053282	T	5.97	0	C	T	↑	2X	4X
			rs7629880	33058811	G	5.97	0	A	G	↑	2X	-
			rs4328758	33057576	A	5.36	0	G	A	↑	1X	4X

			rs4350874	33057595	C	5.3	0	T	C	↑	3X	2X
		La Braña 1 is heterozygous at many eQTL influencing *CCR4* expression										
	CCR5	-[d]										
	CCL2	17	rs2041190	32634541	G	-3.09	0.37	A	G	↓	2X	4X
		La Braña 1 is heterozygous at eQTL of *CCL2*										
Effector molecules	*NOS2A*	-[d]										
	MPO	17	rs12451466	56378067	A	12.9	0	G	A	↑	3X	-
			rs16942910	56363513	T	12.86	0	T	C	↓	-	3X
			rs2877875	56375765	T	12.86	0	T	C	↓	-	2X
			rs11650865	56364741	T	12.86	0	T	G	↓	-	2X
			rs11867412	56382512	T	12.86	0	C	T	↑	7X	-
		La Braña 1 shows low expression of *MPO*										

[a] Exp.all indicates the minor allele tested in Westra et al.[90] study, along with its Z-score and False Discovery Rate.

[b] Direction of expression effect for derived allele in Westra et al.[90] study.

[c] Number of reads found covering the ancestral and derived allele in the La Braña 1 individual.

[d] No eQTLs associated to the expression of this gene in Westra et al.[90] dataset.

[e] Possible post-mortem DNA damage.

Table S26. Virus representation in the La Braña 1 sample.

Virus ID	Virus length	Number of uniquely mapping reads	Virus full name
NC_001422	5386	2099925	Enterobacteria phage phiX174
NC_007821	6068	36012	Enterobacteria phage WA13
NC_007856	5486	11267	Enterobacteria phage ID18
NC_001330	6087	5710	Enterobacteria phage alpha3
NC_007817	5486	4954	Enterobacteria phage ID2 Moscow/ID/2001
NC_012868	6094	3332	Enterobacteria phage St-1
NC_001416	48502	2982	Enterobacteria phage lambda
NC_016765	54024	1823	Pseudomonas phage vB_PaeS_PMG1
NC_005887	42415	1455	Burkholderia phage BcepC6B
NC_019723	47090	1868	Enterobacteria phage HK630
NC_001420	5577	1340	Enterobacteria phage G4
NC_019711	47288	1545	Enterobacteria phage HK629
NC_005882	36748	707	Burkholderia phage BcepMu
NC_019715	39578	464	Enterobacterial phage mEp234
NC_003315	31508	456	Haemophilus phage HP2
NC_001697	32355	456	Haemophilus phage HP1
NC_002167	39732	433	Enterobacteria phage HK97
NC_011357	62147	355	Stx2-converting phage 1717