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

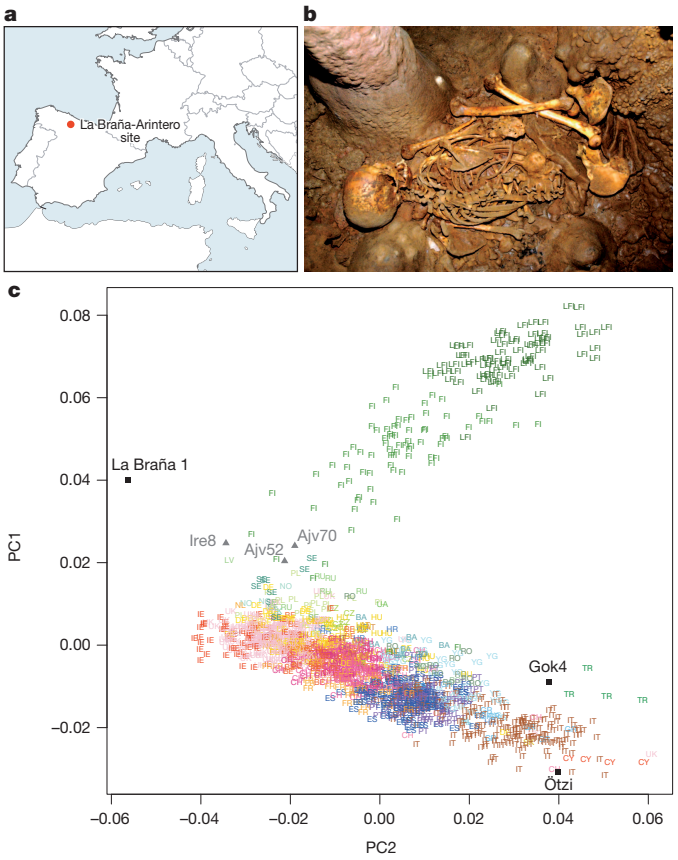


Figure 1 | Geographic location and genetic affinities of the La Braña 1 individual. **a**, Location of the La Braña-Arintero site (Spain). **b**, The La Braña 1 skeleton as discovered in 2006. **c**, PCA based on the average of the Procrustes transformations of individual PCAs with La Braña 1 and each 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

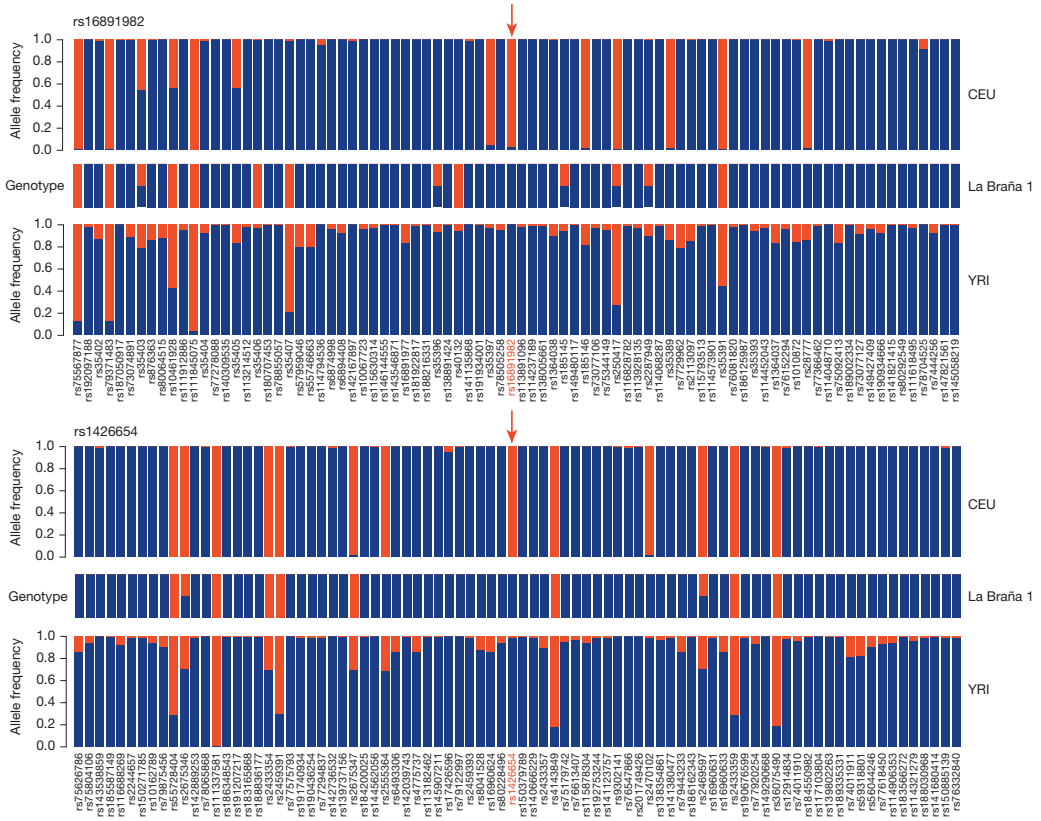


Figure 2 | Ancestral variants around the *SLC45A2* (rs16891982, above) and *SLC24A5* (rs1426654, below) pigmentation genes in the Mesolithic genome. The SNPs around the two diagnostic variants (red arrows) in these two genes were analysed. The resulting haplotype comprises neighbouring SNPs that are 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).

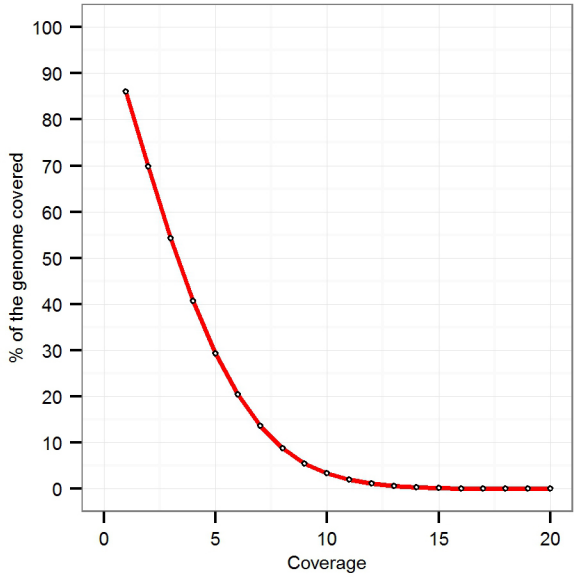
a

	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

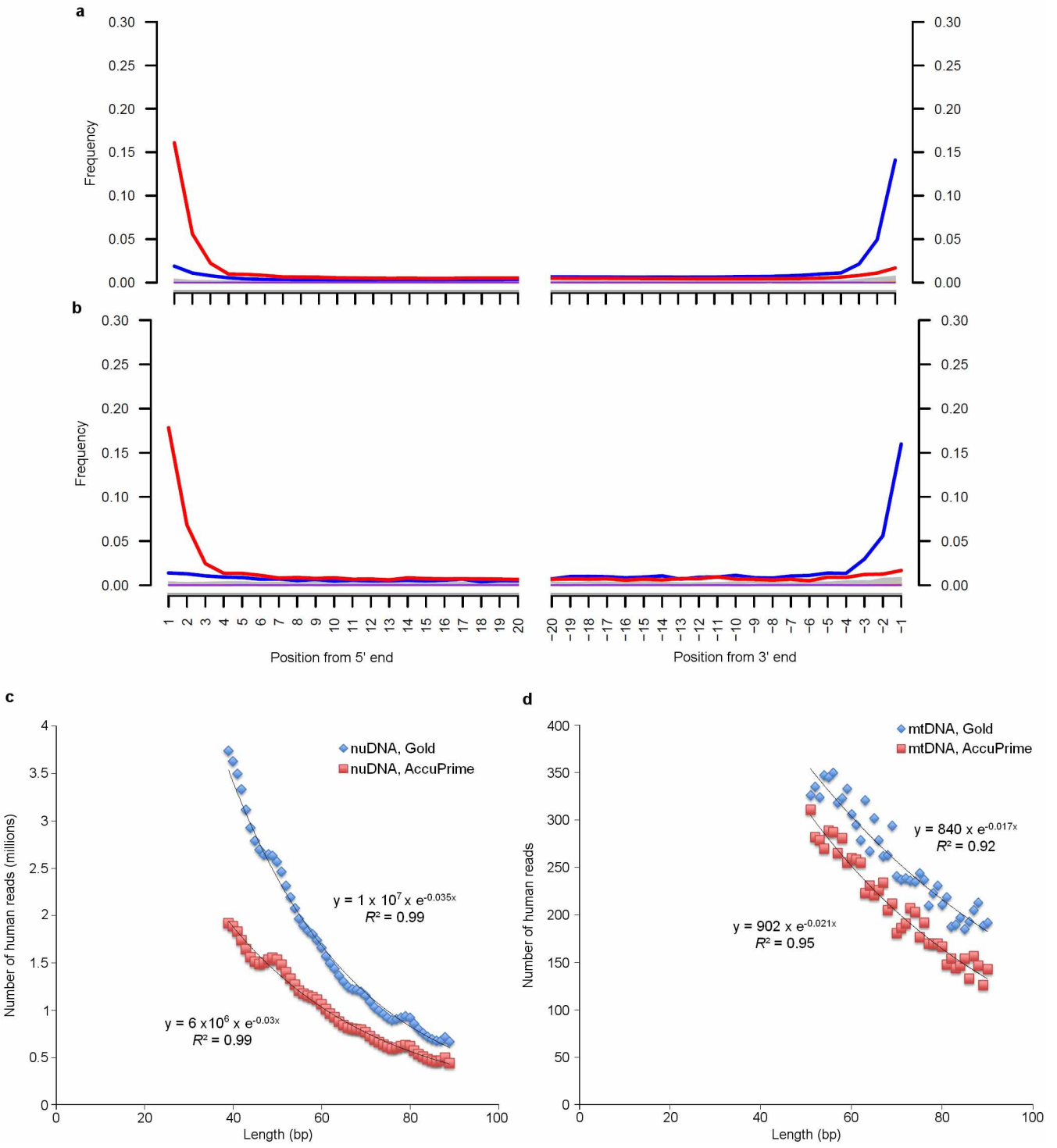
b

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

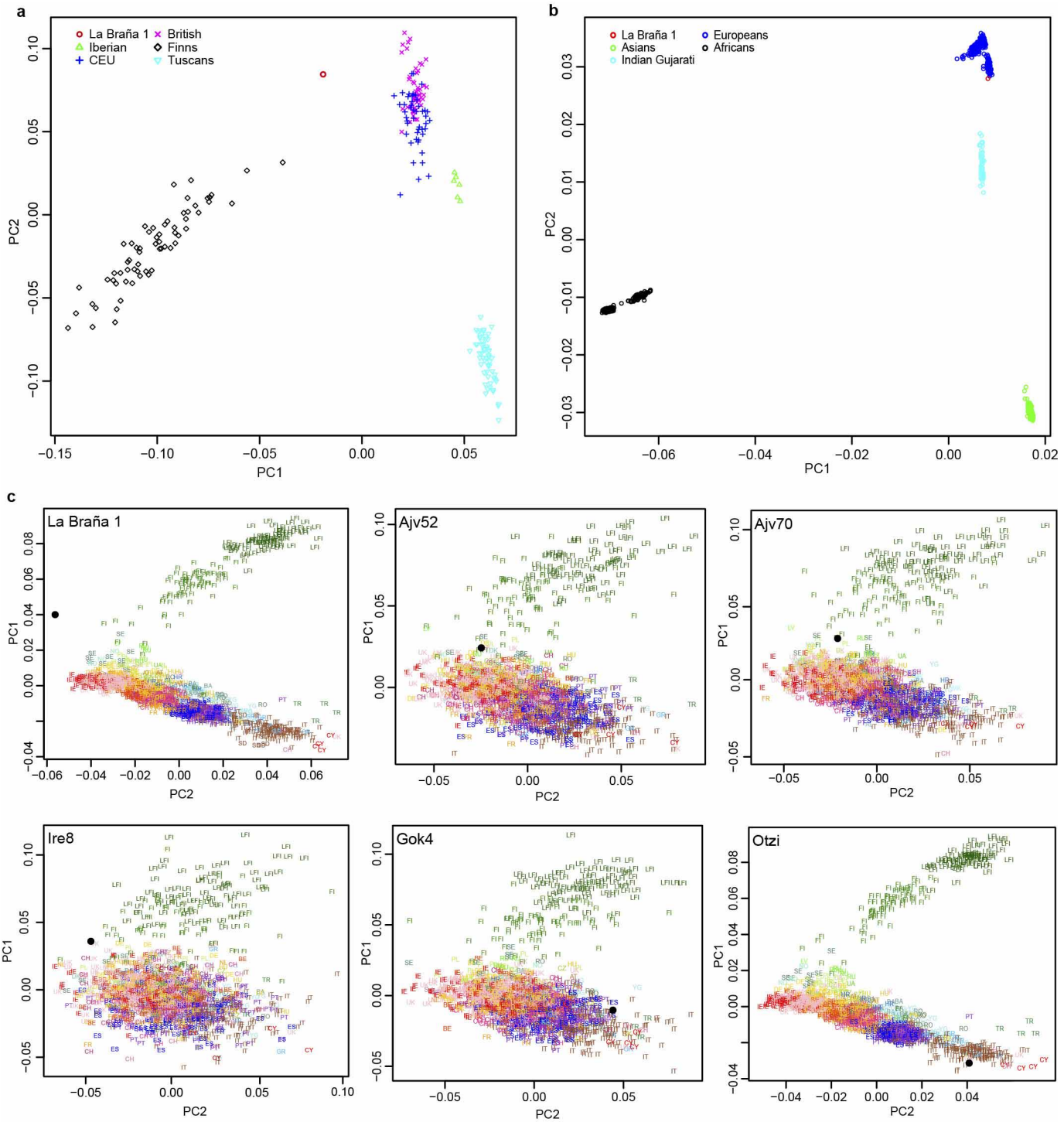
c



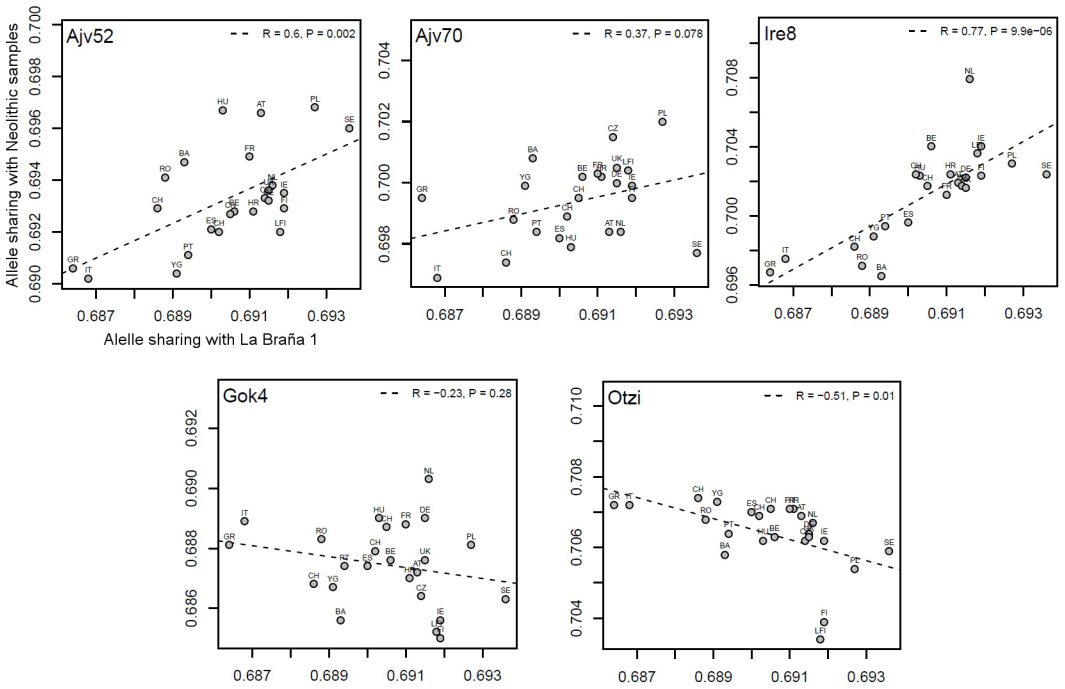
Extended Data Figure 1 | Alignment and coverage statistics of the La Braña 1 genome. **a**, Alignment summary of the La Braña 1 sequence data to hg19 assembly. **b**, Coverage statistics per chromosome. The percentage of the chromosome covered by at least one read is shown, as well as the mean read depth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



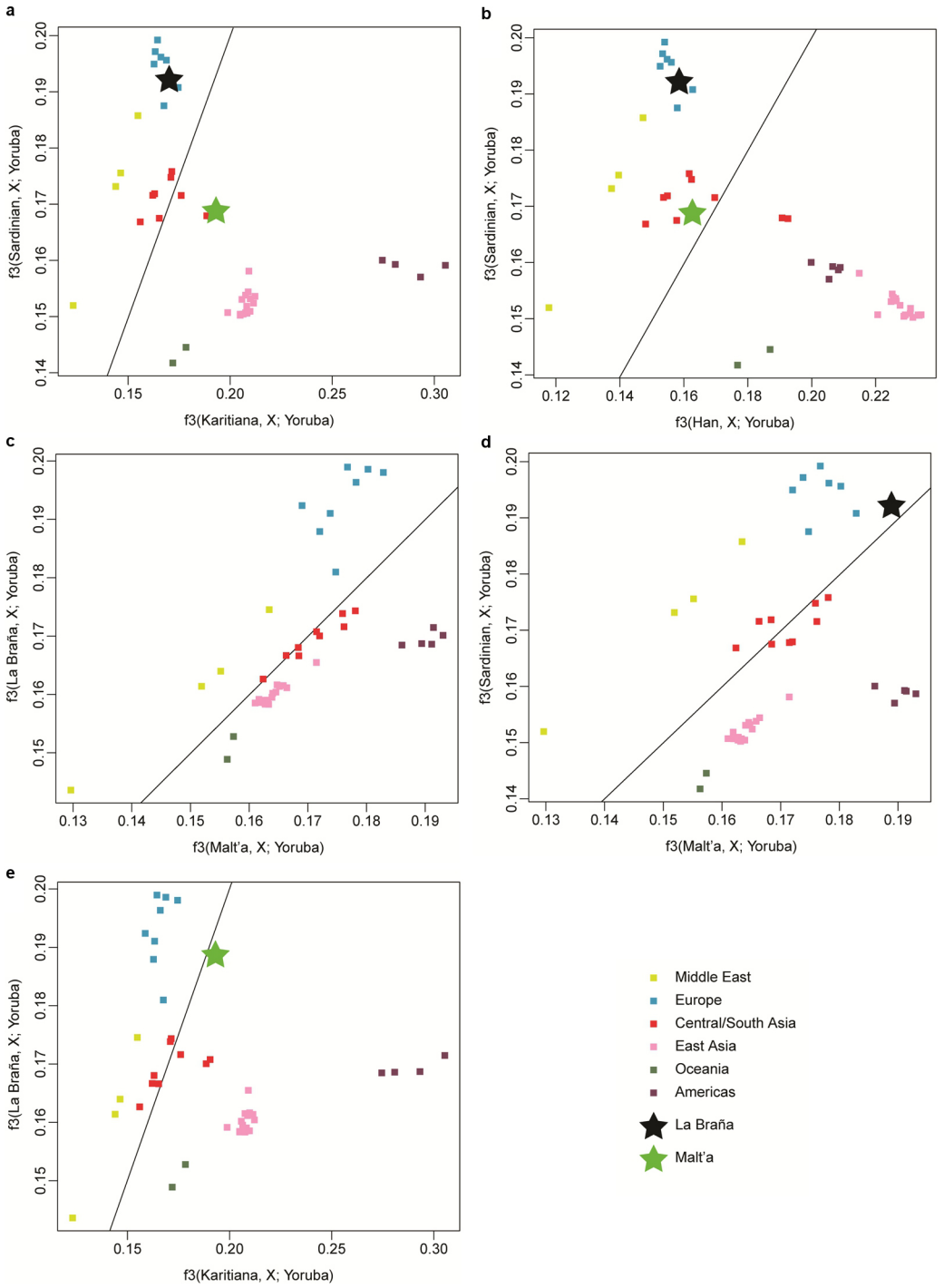
Extended Data Figure 2 | Damage pattern of La Braña 1 sequenced reads. **a**, **b**, Frequencies of C to T (red) and G to A (blue) misincorporations at the 5′ end (left) and 3′ end (right) are shown for the nuclear DNA (nuDNA) (**a**) and mtDNA (**b**). **c**, **d**, Fragment length distribution of reads 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.



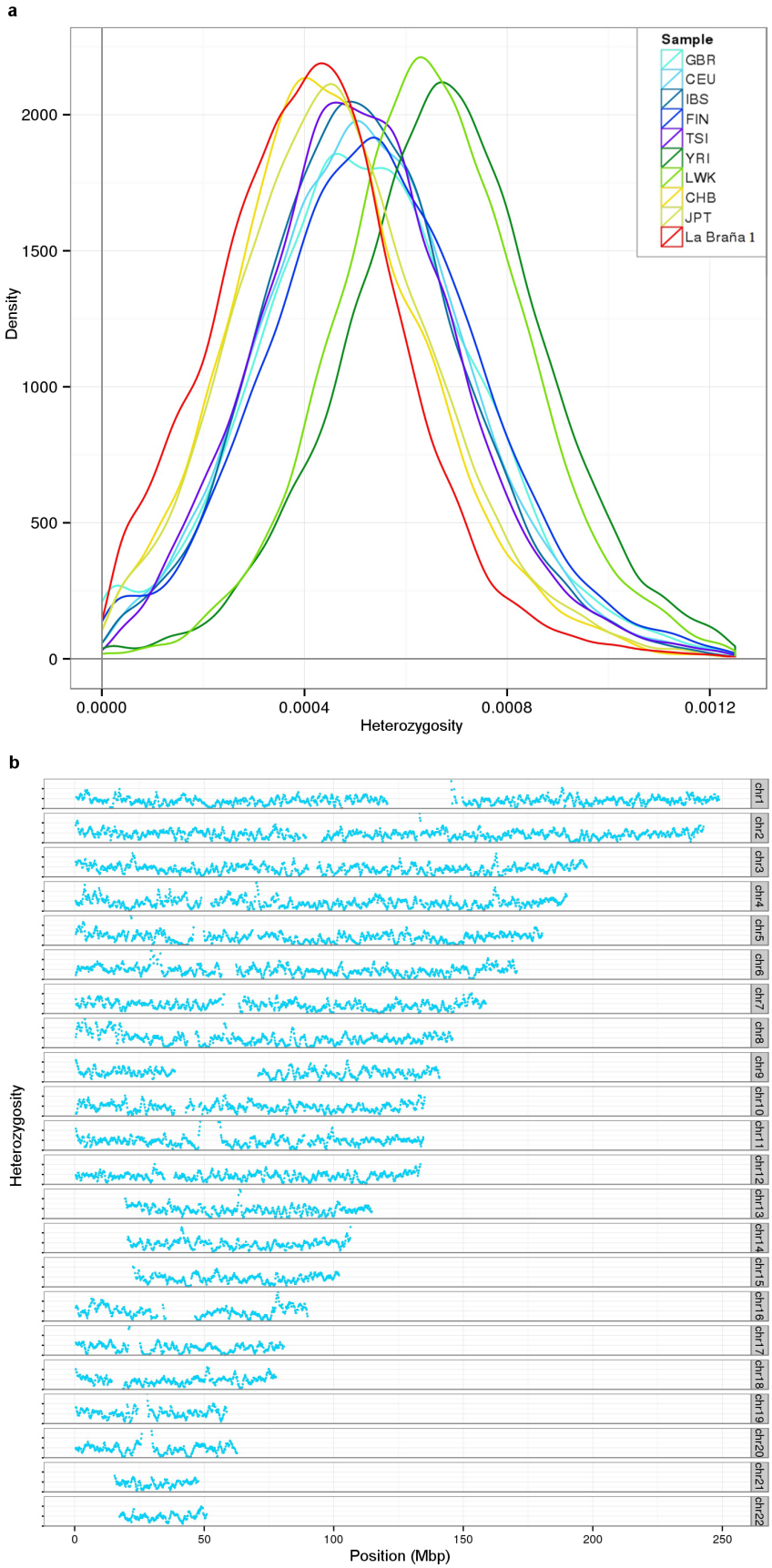
Extended Data Figure 3 | Genetic affinities of the La Braña 1 genome. **a**, PCA of the La Braña 1 SNP data and the 1000 Genomes Project European individuals. **b**, PCA of La Braña 1 versus world-wide data genotyped with the Illumina Omni 2.5M array. Continental terms make reference to each Omni population 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.



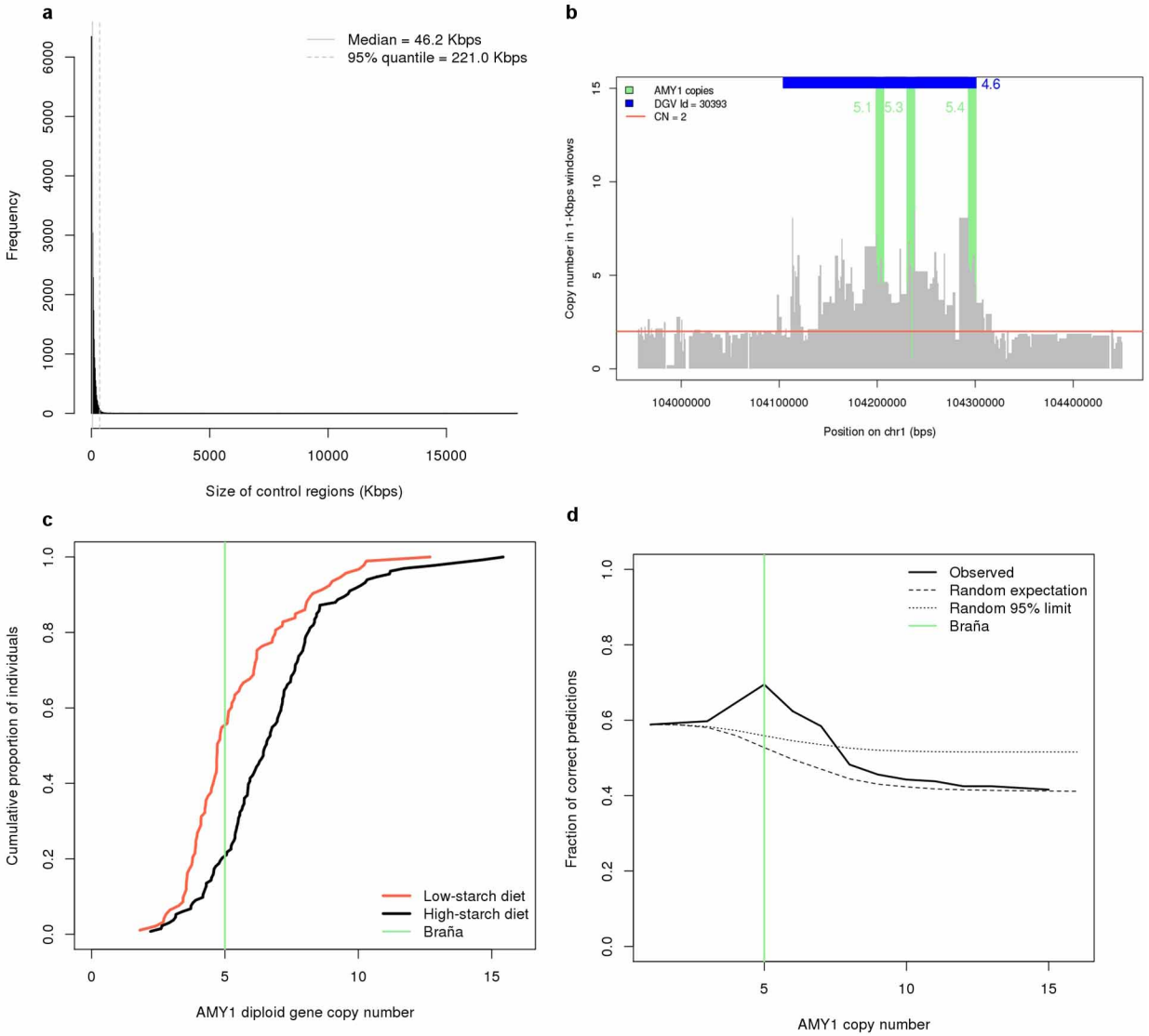
Extended Data Figure 4 | Allele-sharing analysis. Each panel shows the allele-sharing of a particular Neolithic sample from refs 1 and 3 with La Braña 1 sample. The sample IDs are presented in the upper left of each panel (Ajv52, Ajv70, Ire8, Gok4 and Ötzi). In the upper right of each panel, the Pearson's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



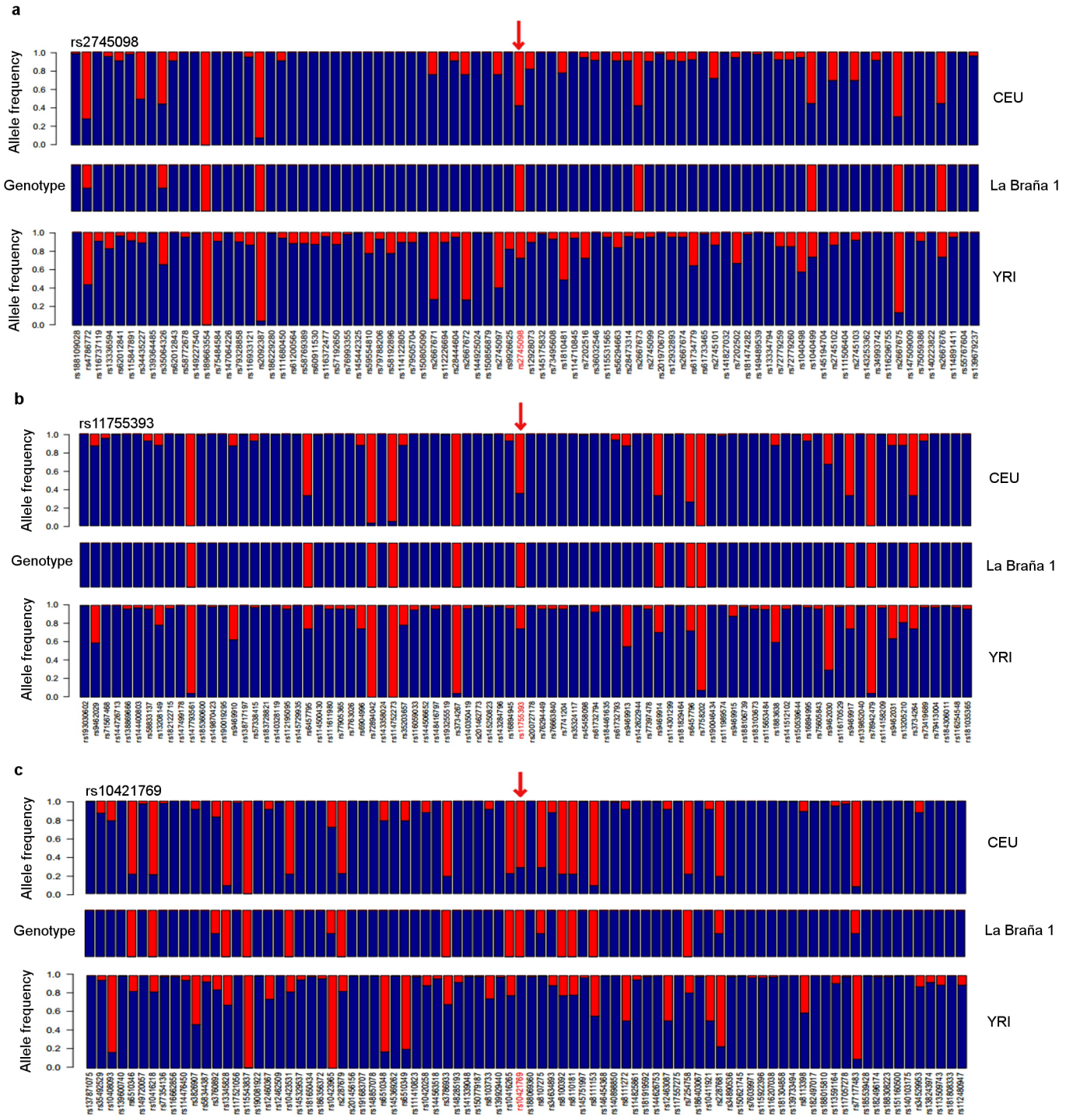
Extended Data Figure 5 | Pairwise outgroup f_3 statistics. **a**, Sardinian versus Karitiana. **b**, Sardinian versus Han. **c**, La Braña 1 versus Mal'ta. **d**, Sardinian versus Mal'ta. **e**, La Braña 1 versus Karitiana. The solid line represents $y = x$.



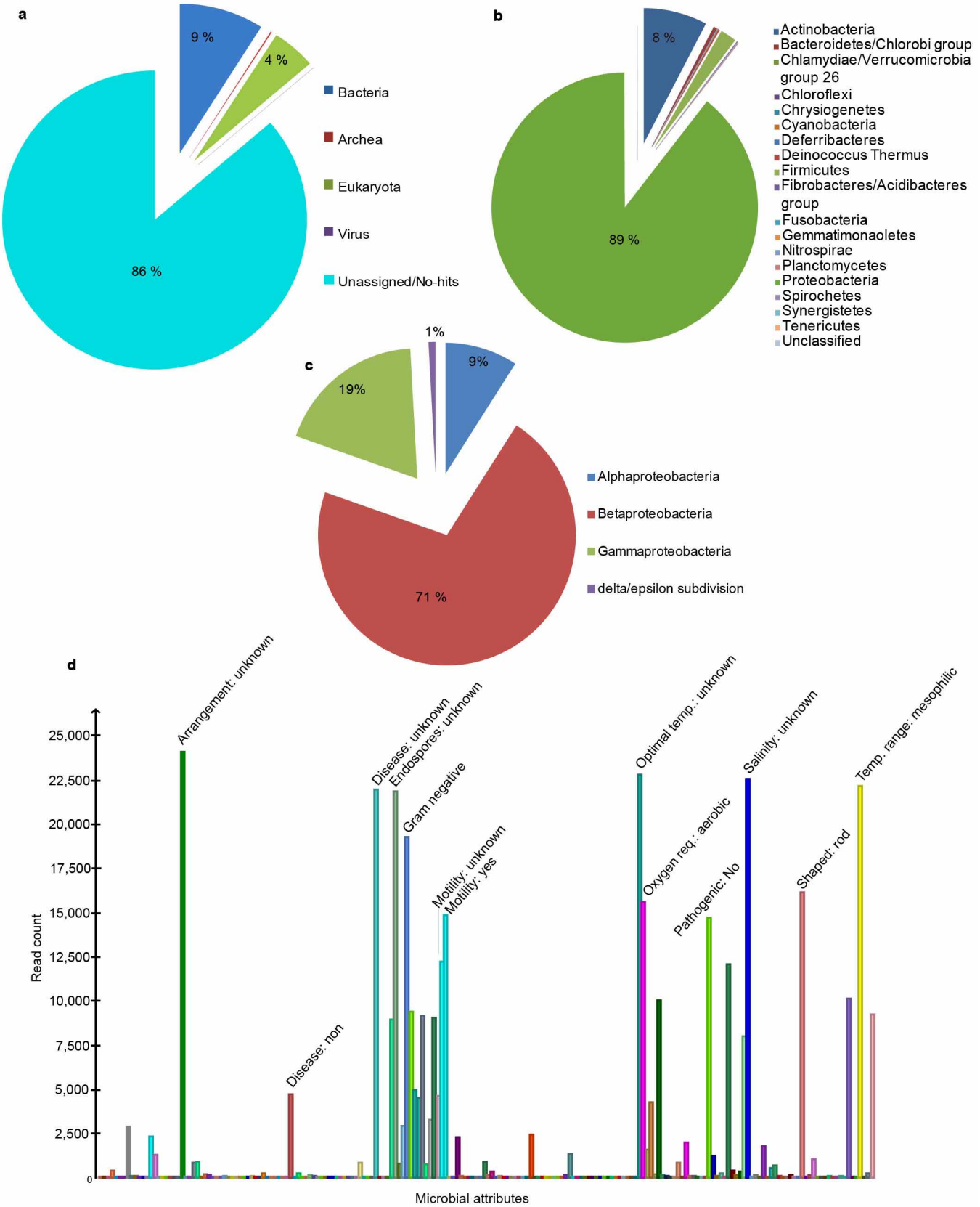
Extended Data Figure 6 | Analysis of heterozygosity. **a**, Heterozygosity distributions of La Braña 1 and modern individuals with similar coverage from the 1000 Genomes Project (using 1-Mb windows with 200 kb overlap). CEU, northern- and western-European ancestry. CHB, Han Chinese; FIN, Finns; GBR, 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.