Published in final edited form as:
Nature. 2021 April 01; 592(7853): 253–257. doi:10.1038/s41586-021-03335-3.

Initial Upper Paleolithic humans in Europe had recent Neanderthal ancestry

Mateja Hajdinjak[1,2,*], **Fabrizio Mafessoni**[1], **Laurits Skov**[1], **Benjamin Vernot**[1], **Alexander Hübner**[1,3], **Qiaomei Fu**[4], **Elena Essel**[1], **Sarah Nagel**[1], **Birgit Nickel**[1], **Julia Richter**[1], **Oana Teodora Moldovan**[5,6], **Silviu Constantin**[7,8], **Elena Endarova**[9], **Nikolay Zahariev**[10], **Rosen Spasov**[10], **Frido Welker**[11,12], **Geoff M. Smith**[11], **Virginie Sinet-Mathiot**[11], **Lindsey Paskulin**[13], **Helen Fewlass**[11], **Sahra Talamo**[11,14], **Zeljko Rezek**[12,15], **Svoboda Sirakova**[16], **Nikolay Sirakov**[16], **Shannon P. McPherron**[11], **Tsenka Tsanova**[11], **Jean-Jacques Hublin**[12,17], **Benjamin M. Peter**[1], **Matthias Meyer**[1], **Pontus Skoglund**[2], **Janet Kelso**[1], **Svante Pääbo**[1,*]

[1] Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[2] Francis Crick Institute, London, United Kingdom

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

[4] Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, Center for Excellence in Life and Paleoenvironment, , Beijing, China

[5] Emil Racovita Institute of Speleology, Cluj Department, Cluj-Napoca, Romania

[6] Romanian Institute of Science and Technology, Cluj-Napoca, Romania

[7] Emil Racovita Institute of Speleology, Department of Geospeleology and Paleontology, Bucharest, Romania

[8] Centro Nacional de Investigatión sobre la Evolución Humana, CENIEH, Burgos, Spain

[9] National History Museum, Sofia, Bulgaria

[10] Archaeology Department, New Bulgarian University, Sofia, Bulgaria

[11] Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[*] Correspondence to: mateja_hajdinjak@eva.mpg.de, paabo@eva.mpg.de; **Correspondence and requests for materials** should be addressed to M.H. or S.P.

Author contributions M.H., E. Essel, S.N., B.N., J.R. and Q.F. performed ancient DNA lab work. O.T.M., S.C., E. Endarova, N.Z., R.S., F.W., G.M.S., V.S.-M., H.F., S.T., Z.R., S.S., N.S., S.P.M., T.T. and J.-J.H. provided and analysed archaeological material. M.H., F.M., L.S., B.V., A.H. and B.M.P. analysed DNA data. B.M.P., M.M., P.S., J.K. and S.P. supervised the study. M.H., J.K. and S.P. wrote the manuscript with input from all co-authors.

Competing interests The authors declare no competing interests.

Peer review information Nature thanks Carles Lalueza-Fox, Marie Soressi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Reviewer reports are available.

Europe PMC Funders Author Manuscripts

[12] Section for Evolutionary Genomics, GLOBE Institute, University of Copenhagen, Denmark

[13] Department of Archaeology, University of Aberdeen, Aberdeen, UK

[14] Department of Chemistry 'G. Ciamician', University of Bologna, Bologna, Italy

[15] University of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, PA, USA

[16] National Institute of Archaeology with Museum, Bulgarian Academy of Sciences, Sofia, Bulgaria

[17] Chaire de Paléoanthropologie, College de France, 75005 Paris, France

Europe PMC Funders Author Manuscripts

Summary

Modern humans appeared in Europe by at least 45,000 years ago[1–5], but the extent of their interactions with Neanderthals, who disappeared by about 40,000 years ago[6], and their relationship to the broader expansion of modern humans outside of Africa are poorly understood. We present genome-wide data from three individuals dated to between 45,930 and 42,580 years ago from Bacho Kiro Cave, Bulgaria[1,2]. They are the earliest Late Pleistocene modern humans that have been recovered in Europe thus far, and were found associated with an Initial Upper Palaeolithic artefact assemblage. Unlike two previously studied individuals of similar ages from Romania[7] and Siberia[8] that did not contribute detectably to later populations, these individuals are more closely related to present-day and ancient populations in East Asia and the Americas than to later west Eurasians. This indicates that they belonged to a modern human migration into Europe that was not previously known from the genetic record, and provides evidence that there was at least some continuity between the earliest modern humans in Europe and later people in Eurasia. Moreover, we find that all three individuals had Neanderthal ancestors a few generations back in their family history, confirming that the first European modern humans mixed with Neanderthals and suggesting that such mixing could have been common.

The transition between the Middle and Upper Palaeolithic periods in Europe, which started about 47,000 years before present (47 kyr **BP**)[1,2], overlapped with the spread of modern humans and the disappearance of Neanderthals, which occurred by about 40 kyr BP[6]. Analyses of the genomes of Neanderthals and modern humans have shown that gene flow occurred between the two hominin groups approximately 60–50 kyr BP[8–11], most likely in southwestern Asia. However, owing to the scarcity of modern human remains from Eurasia that are older than 40 kyr[1–5,12], genome-wide data are available for only three individuals of this age[7,8,13] (Fig. 1). Little is therefore known about the genetics of the earliest modern humans in Eurasia, the extent to which they interacted with archaic humans and their contribution to later populations. For example, while the roughly 42,000 to 37,000-year-old *Oase1* individual from Romania[7,14] and the roughly 45,000-year-old *Ust'Ishim* individual from Siberia[8] do not show specific genetic relationships to subsequent Eurasian populations, the approximately 40,000-year-old *Tianyuan* individual from China contributed to the genetic ancestry of ancient and present-day East Asians[13]. Another open question is the extent to which modern humans mixed with Neanderthals when they spread across Europe and Asia. Direct evidence of local interbreeding exists only for the *Oase1,* who had a recent Neanderthal ancestor[7] in his family history.

Here, we analyse genome-wide data from human specimens found in direct association with an Initial Upper Palaeolithic (IUP) assemblage of artefacts in Bacho Kiro Cave, Bulgaria[1] (Fig. 1), as well as from two more recent specimens from the same site (Supplementary Information 1). The IUP groups together assemblages that fall chronologically between the last Middle Palaeolithic assemblages and the first bladelet industries of the Upper Palaeolithic. The IUP spans a broad geographical area[15], from southwest Asia, central and eastern Europe to Mongolia[16] (Fig. 1, Supplementary Information 1). Although there are reasons to group these assemblages on the basis of their lithic technology, the IUP also shows great regional variability. Therefore, it is debated whether the IUP represents a dispersal of modern humans across middle-latitude Eurasia, the diffusion of certain technological ideas, instances of independent invention, or a combination of some or all of these[15]. The IUP is contemporaneous with late Neanderthal sites in central and western Europe[6] and precedes later Upper Palaeolithic (UP) techno-complexes in Europe, such as the Protoaurignacian and the Aurignacian, by several thousand years[5].

Five human specimens were recovered from Bacho Kiro Cave in recent excavations. They consist of a lower molar *(F6-620)* found in the upper part of the Layer J in the Main Sector, and four bone fragments *(AA7-738, BB7-240, CC7-2289* and *CC7-335)* from Layer I in Niche 1. They have been directly radiocarbon-dated to between 45,930 and 42,580 calibrated years before present (cal. BP)[1,2], and their mitochondrial genomes are of the modern human type, suggesting that they are the oldest Upper Palaeolithic modern humans that have been recovered in Europe[1]. One bone fragment was found in Layer B in the Main Sector *(F6-597)* and another one was among the finds from excavations in the 1970s, when it was retrieved in a position corresponding to the interface of Layers B and C *(BK1653).* The two latter bone fragments were directly dated to 36,320–35,600 cal. BP and 35,290–34,610 cal. BP[1,2], respectively. Although the lithic assemblages from the later layers are sparse, they are likely to be Aurignacian[1,2]. We also produced additional data from the *Oase1* mandible[7,14], which was found outside any archaeological context in Peştera cu Oase, Romania[14]. The mandible was directly dated to about 42–37 kyr BP[14], although this may be an underestimate as the dating was performed before recent technical improvements.

We extracted DNA from between 29.3 mg and 64.7 mg of tooth or bone powder from the specimens as described[1]. We also treated 15 mg of bone powder from *Oase1* with 0.5% hypochlorite solution to reduce bacterial and human contamination before DNA extraction[17]. Among DNA fragments sequenced from the DNA libraries constructed from the Bacho Kiro Cave and *Oase1* extracts, between 0.003% and 1.8% could be mapped to the human genome (Supplementary Information 2). Owing to the low fraction of hominin DNA, we used in-solution hybridization capture[18] to enrich the libraries for about 3.8 million single-nucleotide polymorphisms (SNPs) that are informative about modern human variation and archaic admixture[7,19] (excluding *F6-597*, which contained very little if any endogenous DNA; Supplementary Information 2).

For the six specimens, between 57,293 and 3,272,827 of the targeted SNPs were covered by at least one DNA fragment (Extended Data Table 1). Of these, between 11,655 and 2,290,237 SNPs were covered by at least one fragment showing C-to-T substitutions in the first three and/or the last three positions from the ends, suggesting the presence of

deaminated cytosine bases, which are typical of ancient DNA[20] (Extended Data Table 1, Extended Data Fig. 1). On the basis of the numbers of putatively deaminated fragments aligning to the X chromosome and the autosomes[21] (Supplementary Information 4), we conclude that specimens *F6-620, AA7-738, BB7-240* and *CC7-335* belonged to males, whereas *BK1653* and *CC7-2289* belonged to females, although the low amount of data makes this conclusion tentative for *CC7-2289* (Extended Data Fig. 2a).

Using an approach that makes use of DNA deamination patterns[22], we estimated that the overall nuclear DNA contamination was between 2.2% ± 0.5% *(F6-620)* and 42.4% ± 0.6% (CC7-2289). In the male specimens, we estimated contamination from polymorphisms on the X chromosome[23] to between 1.6% ± 0.1% and 3.4% ± 0.5% (Supplementary Information 2). Owing to the presence of present-day human contamination, we restricted all downstream analyses to putatively deaminated fragments for all specimens except *F6-620* (for which contamination was so low that we used all fragments). This left between 11,655 and 3,272,827 SNPs per specimen to be used for the subsequent analyses (Supplementary Information 2).

The molar *F6-620* and the bone fragment *AA7-738* have identical mitochondrial genome sequences[1] and both come from males. The pairwise mismatch rate between the two specimens at the SNPs[24] is 0.13, similar to the mismatch rate between libraries from the same specimen (Extended Data Fig. 2b). By contrast, this number is 0.23 (interquartile range: 0.22–0.25) for the other Bacho Kiro Cave specimens, similar to unrelated ancient individuals from other studies (Extended Data Fig. 2b). Thus, we conclude that specimens *F6-620* and *AA7-738* belonged to the same individual or to identical twins, which is much less likely.

We enriched the libraries from the male individuals using probes that targeted about 6.9 Mb of the Y chromosome[25] (Supplementary Information 4) and arrived at 15.2-fold coverage for *F6-620,* 2.5-fold for *BB7-240* and 1.5-fold for *CC7-335. F6-620* carries a basal lineage of the Y chromosome haplogroup F (F-M89), whereas *BB7-240* and *CC7-335* carry haplogroup C1 (C-F3393). Although haplogroup C is common among males from East Asia and Oceania, both haplogroups F and C1 are rare in present-day humans and are found only at low frequencies in mainland Southeast Asia and Japan[26,27].

We estimated the extent of genetic similarity among the Bacho Kiro Cave individuals and other early modern humans using outgroup *f*3-statistics[28]. The three roughly 45,000-year-old IUP individuals are more similar to one another than to any other ancient individual (Extended Data Fig. 3a). By contrast, *BK1653,* which is about 35,000 years old, is more similar to later Upper Palaeolithic individuals from Europe that are around 38,000 years old or younger[29,30] ($3.0 \leq |Z| \leq 17.4$; Extended Data Fig. 4, Supplementary Information 5) for example, to the roughly 35,000-year-old *GoyetQ116-1* individual from Belgium and members of the ‘Věstonice’ genetic cluster, which are associated with later Gravettian assemblages[29] (Extended Data Figs. 3a, 4b, c).

When comparing the Bacho Kiro Cave individuals to present-day populations[31], we found that the IUP individuals share more alleles (that is, more genetic variants) with present- day

populations from East Asia, Central Asia and the Americas than with populations from western Eurasia (Fig. 2a, Supplementary Information 5), whereas the later *BK1653* individual shares more alleles with present-day western Eurasians (Extended Data Figs. 3b, 4a).

We next investigated whether these observations could be due to the fact that present-day populations in western Eurasia derive part of their ancestry from 'Basal Eurasians'[32,33], an inferred population that diverged early from other non-Africans and may have 'diluted' allele sharing between western Eurasians and IUP individuals. To do this, we compared *Ust'Ishim*, *Oase1* and the IUP Bacho Kiro Cave individuals to western Eurasian individuals such as the approximately 38,000-year-old *Kostenki14* from Russia[29,30], which pre-dates the introduction of 'Basal Eurasian' ancestry to Europe around 8,000 cal. BP[32]. We found that *Ust'Ishim* and *Oase1* individuals showed no more affinity to western than to eastern Eurasians, suggesting that they did not contribute ancestry to later Eurasian populations, as previously shown[7,8] (Supplementary Information 5, Extended Data Fig. 5). By contrast, the IUP Bacho Kiro Cave individuals shared more alleles with the roughly 40,000-year-old *Tianyuan* individual[13] from China (Fig. 2b) and other ancient Siberians[34,35] and Native Americans[36–39] (Fig. 2c) than with *Kostenki14* ($3.6 \leq |Z| \leq 5.3$). Among other western Eurasian Upper Palaeolithic humans, the IUP Bacho Kiro Cave individuals shared more alleles with *Oase1* ($3.6 \leq |Z| \leq 4.3$) and the roughly 35,000-year-old *GoyetQ116-1* [29] individuals than with *Kostenki14* ($3.2 \leq |Z| \leq 4.3$; Fig. 2c, Supplementary Information 5). Notably, *GoyetQ116-1* has previously been shown to share more alleles with early East Asians than other individuals of similar age in Europe[13].

When we explored models of population history that are compatible with the observations above using admixture graphs[28], we found that the IUP Bacho Kiro Cave individuals were related to populations that contributed ancestry to *Tianyuan* individual in China as well as to a lesser extent to the *GoyetQ116-1* and *Ust'Ishim* individuals (all $|Z| < 3$; Fig. 2d, Supplementary Information 6). This resolves the previously unclear relationship between *GoyetQ116-1* and *Tianyuan* individuals[13] without the need for gene flow between these two geographically distant individuals. The models also suggest that the later *BK1653* individual belonged to a population that was related, but not identical, to that of *GoyetQ116-1* (Fig. 2d, Extended Data Fig. 4, Supplementary Information 6) and that the Vestonice cluster, whose members were found in association with Gravettian assemblages[29], derived part of their ancestry from such a population and the rest from populations related to the roughly 34,000-year-old *Sunghir* individuals[40] from Russia (Fig. 2d, Supplementary Information 6).

As the IUP Bacho Kiro Cave individuals lived at the same time as some of the last Neanderthals in Europe[6], we estimated the proportion of Neanderthal DNA in their genomes by taking advantage of two high-quality Neanderthal genomes[9,10,41]. We found that the IUP individuals *F6-620, BB7-240* and *CC7-335* carried 3.8% (95% confidence interval (CI): 3.3–4.4%), 3.0% (95% CI: 2.4–3.6%) and 3.4% (95% CI: 2.8–4.0%) Neanderthal DNA, respectively. This is more than the average of 1.9% (95% CI: 1.5–2.4%) found in other ancient or present-day humans, except for *Oase1*, who had a close Neanderthal relative (6.4% (95% CI: 5.7–7.1%); Extended Data Fig. 6, Supplementary Information 7). By contrast, the more recent *BK1653* individual carried only 1.9% (95% CI: 1.4–2.4%)

Neanderthal DNA, similar to other ancient and present-day humans[10,41] (Extended Data Fig. 6). As has been the case for all humans studied so far, the Neanderthal DNA in *BK1653* and the IUP Bacho Kiro Cave individuals was more similar to the *Vindija33.19*[10] and *Chagyrskaya8*[42] Neanderthals than to the *Altai* Neanderthal[9] ($2.8 \leq |Z| \leq 5.1$; Supplementary Information 7).

To study the spatial distribution of Neanderthal ancestry in the genomes of the Bacho Kiro Cave individuals, we used around 1.7 million SNPs at which Neanderthal[9] and/or Denisovan[43] genomes differ from African genomes[7] and an approach[44] that detects tracts of archaic DNA in ancient genomes. We found a total of 279.6 centiMorgans (cM) of Neanderthal DNA in *F6-620,* 251.6 cM in *CC7-335* and 220.9 cM in *BB7-240,* and these individuals carried seven, six and nine Neanderthal DNA segments longer than 5 cM, respectively (Fig. 3, Extended Data Fig. 7a, Supplementary Information 8). The longest introgressed Neanderthal segment in *F6-620* encompassed 54.3 cM, and the longest segments in *CC7-335* and *BB7-240* were 25.6 cM and 17.4 cM, respectively (Fig. 3, Extended Data Fig. 7a). By contrast, the total amount of Neanderthal DNA in the *BK1653* genome was 121.7 cM and the longest Neanderthal segment was 2.5 cM (Fig. 3, Extended Data Fig. 7a).

On the basis of the distribution of the long Neanderthal segments, we estimate that *F6-620* individual had a Neanderthal ancestor less than six generations back in his family tree (Extended Data Table 2, Supplementary Information 8). Both *CC7-335* and *BB7-240* individuals had Neanderthal ancestors about seven generations back in their families, with upper confidence intervals of ten and seventeen generations, respectively (Extended Data Table 2, Extended Data Fig. 7b, Supplementary Information 8). Thus, all IUP Bacho Kiro Cave individuals had recent Neanderthal ancestors in their immediate family histories.

To further explore the extent to which the Bacho Kiro Cave individuals contributed ancestry to later populations in Eurasia, we investigated whether the Neanderthal DNA segments in Bacho Kiro Cave genomes overlapped with Neanderthal segments in present-day populations more than expected by chance. We find more overlapping of segments between present-day East Asians and the IUP Bacho Kiro Cave individuals (lowest correlation coefficient of 0.09, 95% CI: 0.08–0.1) than with *BK1653* individual ($P = 0.02$, Wilcoxon test). By contrast, *BK1653* individual shows more overlapping of Neanderthal segments with present-day western Eurasians (a correlation coefficient of 0.11, 95% CI: 0.1–0.12) than do the IUP Bacho Kiro Cave individuals ($P < 1 \times 10^{-18}$, Wilcoxon test). This is compatible with the observation that the IUP Bacho Kiro Cave population contributed more ancestry to later populations in Asia and the Americas, whereas *BK1653* individual contributed more ancestry to populations in western Eurasia.

We next looked for overlap between parts of the human genome that carry little or no Neanderthal ancestry (Neanderthal ‘deserts’), which are thought to have been caused by purifying selection against Neanderthal DNA shortly after introgression[45,46]. We find almost no introgressed Neanderthal DNA in the previously described deserts in the IUP Bacho Kiro Cave and *Oase1* individuals (249 kb out of 898 Mb of introgressed sequence; $P = 0.0079$, permutation P value). When we restricted these comparisons to the more recent Neanderthal

contributions (that is, segments longer than 5 cM), we similarly found no overlap (0 Mb out of 415 Mb, $P = 0.15$, permutation P value), suggesting that the selection against Neanderthal DNA variants occurred within a few generations, although additional individuals with recent Neanderthal ancestry will be needed to fully resolve this question.

In conclusion, the Bacho Kiro Cave genomes show that several distinct modern human populations existed during the early Upper Palaeolithic in Eurasia. Some of these populations, represented by the *Oase1* and *Ust'Ishim* individuals, show no detectable affinities to later populations, whereas groups related to the IUP Bacho Kiro Cave individuals contributed to later populations with Asian ancestry as well as some western Eurasians such as *GoyetQ116-1* individual in Belgium. This is consistent with the fact that IUP archaeological assemblages are found from central and eastern Europe to present-day Mongolia[5,15,16] (Fig. 1), and a putative IUP dispersal that reached from eastern Europe to East Asia. Eventually populations related to the IUP Bacho Kiro Cave individuals disappeared in western Eurasia without leaving a detectable genetic contribution to later populations, as indicated by the fact that later individuals, including *BK1653* at Bacho Kiro Cave, were closer to present-day Europeans than to present-day Asians[29,30]. In Europe, the notion of successive population replacements is also consistent with the archaeological record, where the IUP is clearly intrusive against the Middle Palaeolithic background and where, apart from the common focus on blades, there are no clear technological connections between the IUP and the subsequent Aurignacian technologies. Finally, it is striking that all four of the European individuals who overlapped in time with late Neanderthals[7] and from whom genome-wide data have been retrieved had close Neanderthal relatives in their family histories (Fig. 3, Extended Data Figs. 7, 8). This suggests that mixing between Neanderthals and the first modern humans arriving into Europe was perhaps more common than is often assumed.

Online methods

Ethics declaration

All approvals for specimen handling have been obtained from the relevant institutions. For the *Oase1* specimen, the permission was granted to S.P. by the Emil Racovita Institute of Speleology, as the national authority in caves study. For the Bacho Kiro Cave specimens, the permission was granted by the Bulgarian Ministry of Culture and the National Museum of Natural History (Sofia, Bulgaria).

DNA extraction and library preparation

Data generation for the seven Bacho Kiro Cave specimens (specimen IDs: *F6-620, AA7-738, BB7-240, CC7-2289, CC7-335, F6-597* and *BK1653)* was based on DNA libraries prepared and described previously[1]. To obtain additional data from *Oase1* individual, we extracted DNA from 15 mg of bone powder from the specimen[7,14]. As it was previously found to be highly contaminated with microbial and present-day human DNA, the bone powder was treated with 0.5% hypochlorite solution before DNA extraction[17]. Four single-stranded DNA libraries were prepared from the resulting extract and two additional libraries were prepared, each using 5 μl of the two DNA extracts generated previously[7] as

described[48]. The pools of libraries were then sequenced directly on Illumina MiSeq and HiSeq 2500 platforms in a double index configuration (2 x 76 cycles)[49] and base calling was done using Bustard (Illumina).

DNA captures

We enriched the selected amplified libraries for about 3.7 million SNPs across the genome described in Supplementary Data 2 of ref. [19] (SNP Panel 1 or 390k array), and Supplementary Data 1, 2 and 3 of ref. [7] (SNP Panels 2, 3 and 4, or 840k, 1000k and Archaic admixture arrays, respectively). For the male individuals (Bacho Kiro Cave *F6-620, BB7-240* and CC7- *335),* an aliquot of each library was additionally enriched for about 6.9 Mb of the Y chromosome[25]. All of the enriched libraries were sequenced on the Illumina HiSeq 2500 platforms in a double index configuration (2 x 76 cycles)[49] and base calling was done using Bustard (Illumina).

Sequencing of capture products and data processing

For all sequencing runs we trimmed the adapters and merged overlapping forward and reverse reads into single sequences using leeHom[50] (version: https://bioinf.eva.mpg.de/leehom/). The Burrows-Wheeler Aligner[51] (BWA, version: 0.5.10-evan.9-1-g44db244; https://github.com/mpieva/network-aware-bwa) with the parameters adjusted for ancient DNA ("-n 0.01 –o 2 –l 16500")[43] was used to align the data from all sequencing runs to the human reference genome (GRCh37/1000 Genomes release; ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_seq uence/). Only reads that showed perfect matches to the expected index combinations were used for all downstream analyses. PCR duplicates were removed using bam-rmdup (version: 0.6.3; https://github.com/mpieva/biohazard-tools) and SAMtools (version: 1.3.1)[52] was used to filter for fragments that were at least 35 bp long and that had a mapping quality equal to or greater than 25. BAM files of the libraries enriched for the specific subset of the nuclear genome were further intersected with the BED files containing target SNP positions (390k, 840k, 1000k, Archaic admixture, a merged set of SNP Panels 1 and 2 or 1240k, and a merged set of SNP Panels 1, 2 and 3 or 2200k) and regions (Y chromosome) using BEDtools[53] (version: 2.24.0). In order to filter for endogenous ancient DNA or putatively deaminated fragments, we used elevated C-to-T substitutions relative to the reference genome at the first three and/or last three positions of the alignment ends[20]. We merged libraries originating from the same specimen using samtools merge[52] to produce the final datasets for downstream analyses (Extended Data Tab. 1, Supplementary Information 2).

Merging of the Bacho Kiro Cave and Oase1 data with other genomes

We performed random read sampling using bam-caller (https://github.com/bodkan/bam-caller, version: 0.1) by picking a base with a base quality of at least 30 at each position in the 1240k and 2200k SNP Panels that was covered by at least one fragment longer than 35 bp with a mapping quality equal to or higher than 25 ($L \geq 35$ bp, MQ ≥ 25, BQ ≥ 30). To mitigate the effect of deamination-derived substitutions on downstream analyses, we did not sample any Ts on the forward strands (in the orientation as sequenced) or any As on the reverse strands in the first five and/or last five positions from the alignment ends. Owing to the haploid nature of the Y chromosome, we called genotypes across the approximately 6.9

Europe PMC Funders Author Manuscripts

Mb of the Y chromosome for the enriched libraries of male individuals by calling a consensus allele at each position by majority call requiring a minimum coverage of 3 for specimens *F6-620* and *BB7-240* and of 2 for specimen *CC7-335* using using bam-caller (https://github.com/bodkan/bam-caller, version: 0.1) (Supplementary Information 2).

We merged the data from the newly sequenced specimens with datasets of previously published ancient and present-day humans, as well as archaic humans for three SNP panels (1240k, 2240k and Archaic admixture; Supplementary Information 3). Data from the 1240k panel include genotypes of 2,109 ancient and 2,974 present-day individuals compiled from published studies and available in the EIGENSTRAT format[28] at https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data (version 37.2, released 22 February 2019). Data from the 2240k panel include published genetic data of ancient modern humans obtained through hybridization captures[7,13,29] and a range of present-day[9,31] and ancient modern humans[8,30,32–37,54–59], as well as the archaics[9,10,42,43,60], for which whole-genome shotgun data of varying coverage are available (Supplementary Information 3). The Archaic admixture panel data include 21 ancient modern humans directly enriched for these sites[7,13,29], as well as the genotypes of present-day[9,31] and ancient modern humans[8,30,32–37,54–59], as well as the archaics[9,10,42,43,60], for which whole-genome shotgun data are available (Supplementary Information 3) and that were intersected with about 1.7 million SNPs of the Archaic admixture panel using BEDTools[53] (version: 2.24.0).

Population genetic analyses

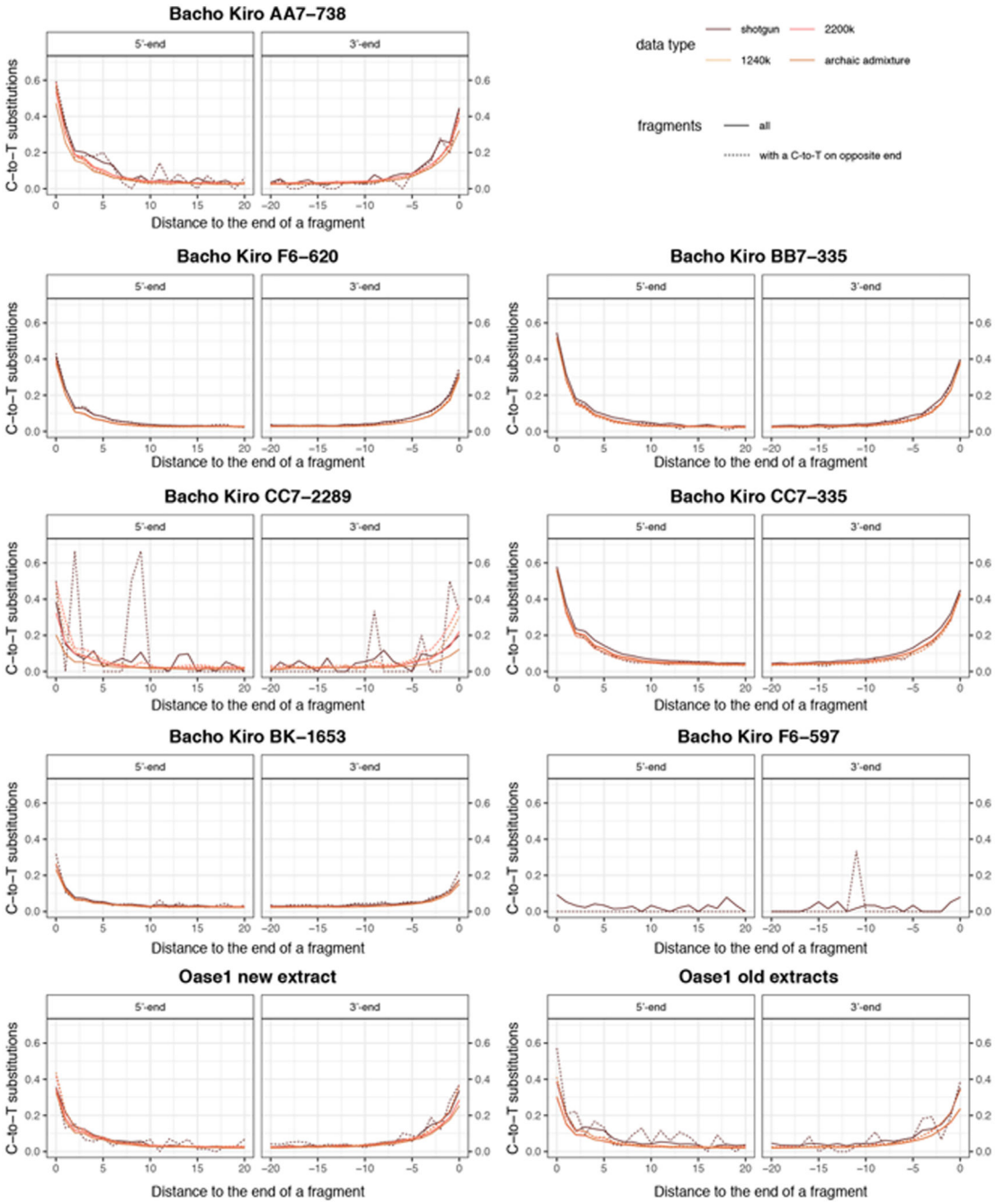
To determine the relationship of Bacho Kiro Cave individuals and *Oase1* to other modern and archaic humans we used a range of *f*-statistics from ADMIXTOOLS[28] (version: v5.1) and as implemented in the R package admixr[61] (version: 0.7.1; Supplementary Information 4). We used qpGraph program (Admixture Graph) from ADMIXTOOLS[28] (version: v5.1) to test models of the relationship among Initial Upper Palaeolithic Bacho Kiro Cave individuals, the roughly 34,000-year-old Bacho Kiro Cave individual *BK1653* and other ancient modern humans from Eurasia older than 30,000 years BP (Fig. 2d, Supplementary Information 6).

Neanderthal ancestry

We estimated the proportion of Neanderthal DNA in the genomes of present-day and ancient modern humans by computing a direct f_4 ratio[28] that takes advantage of the two high-quality Neanderthal genomes[9,10,42] (Extended Data Fig. 6, Supplementary Information 7). We used admixfrog[44] (version: 0.5.6, https://github.com/BenjaminPeter/admixfrog/) to detect archaic introgressed segments in the genomes of Bacho Kiro Cave individuals and *Oase1,* as well as in other ancient modern humans and 254 present-day non-Africans from the SGDP[31] as a direct comparison (Supplementary Information 8). We used these introgressed segments to estimate the number of generations since the most recent Neanderthal ancestor of IUP Bacho Kiro Cave individuals and *Oase1* (Supplementary Information 8), to investigate the overlap of Neanderthal segments in Bacho Kiro Cave individuals with those detected in present-day and ancient modern humans (Supplementary Information 9), and to investigate the overlap of Neanderthal segments in IUP Bacho Kiro

Cave individuals and *Oase1* with parts of the human genome devoid of Neanderthal ancestry[45,46] (Neanderthal deserts; Supplementary Information 10).

Extended Data

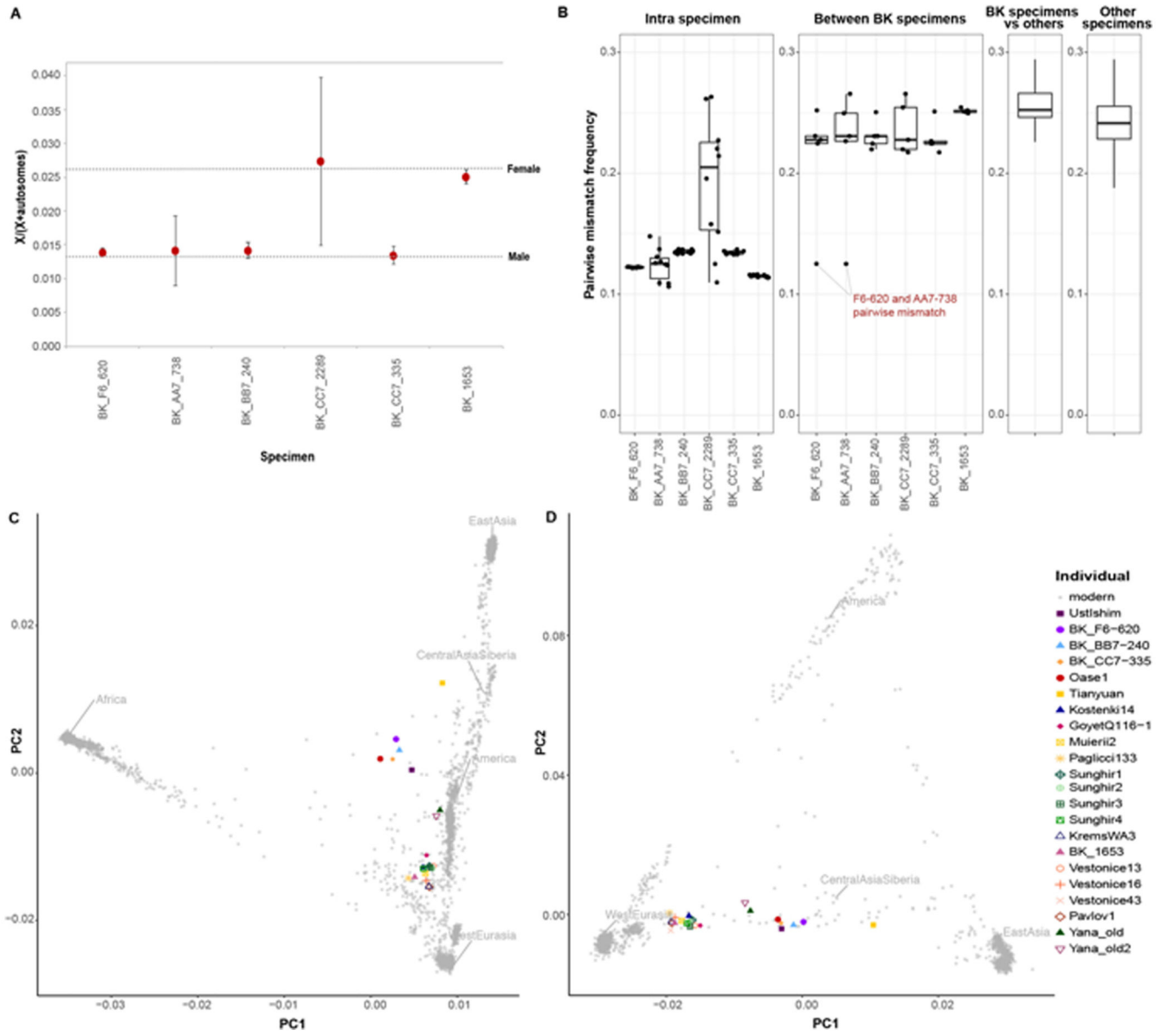


Extended Data Fig. 1. C-to-T substitution frequencies at the beginning and end of nuclear alignments for the merged libraries of Bacho Kiro Cave specimens and *Oase1*.

Only fragments of at least 35 bp that mapped to the human reference genome with a mapping quality of at least 25 (MQ > 25) were used for this analysis. Solid lines depict all

Europe PMC Funders Author Manuscripts

fragments and dashed lines the fragments that have a C-to-T substitution at the opposing end (conditional substitutions).

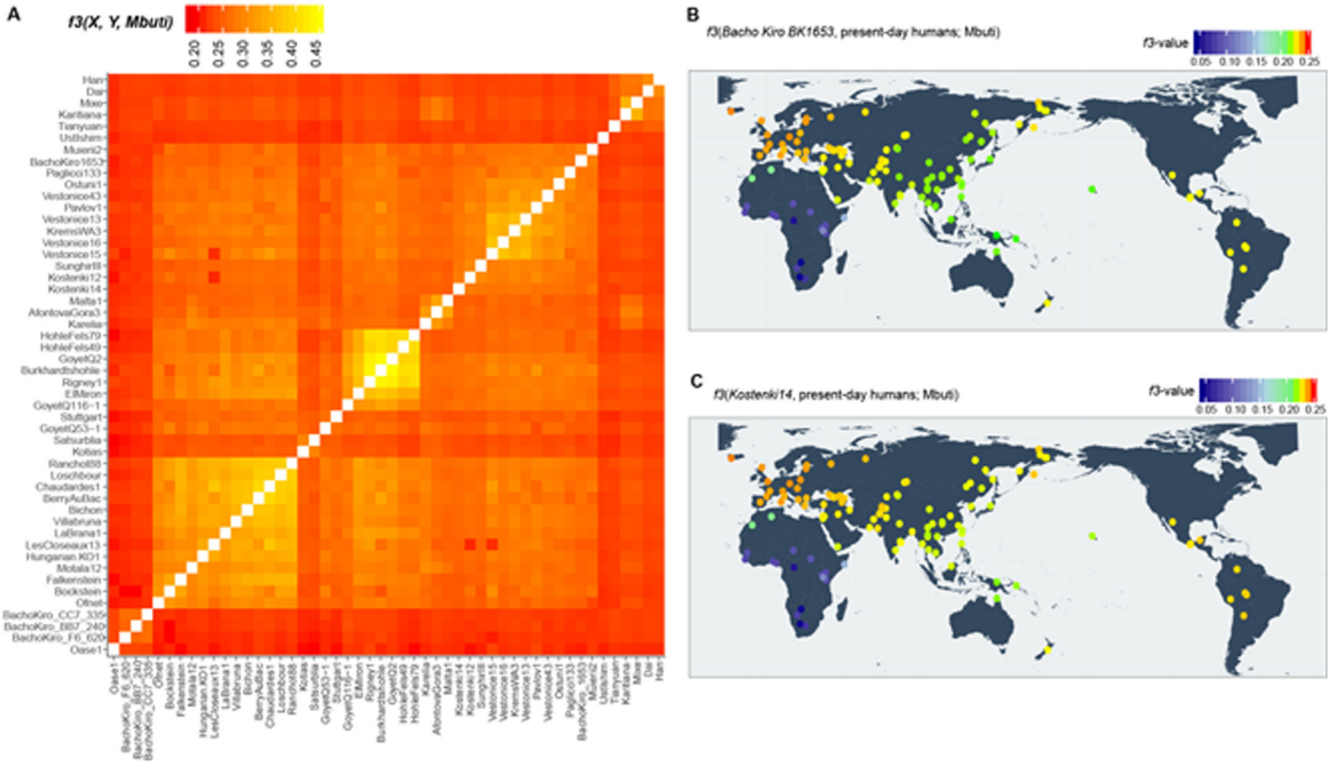


Extended Data Fig. 2. Sex determination, pairwise mismatch rate between specimens and principal component analyses (PCA).

a, Sex determination for Bacho Kiro Cave specimens. Only fragments that showed C-to-T substitutions in the first three and/or last three positions and overlapping 2200k Panel SNPs were used for this analysis (for the number of deaminated fragments per specimen, see Supplementary Table 2.8). The expected ratios of X to (X + autosomal) fragments for a female and a male individual are depicted as dashed lines, and circles correspond to the calculated values for each of the Bacho Kiro Cave specimens. Whiskers correspond to 95% binomial confidence intervals. **b**, Pairwise mismatch rate between different libraries from the

Europe PMC Funders Author Manuscripts

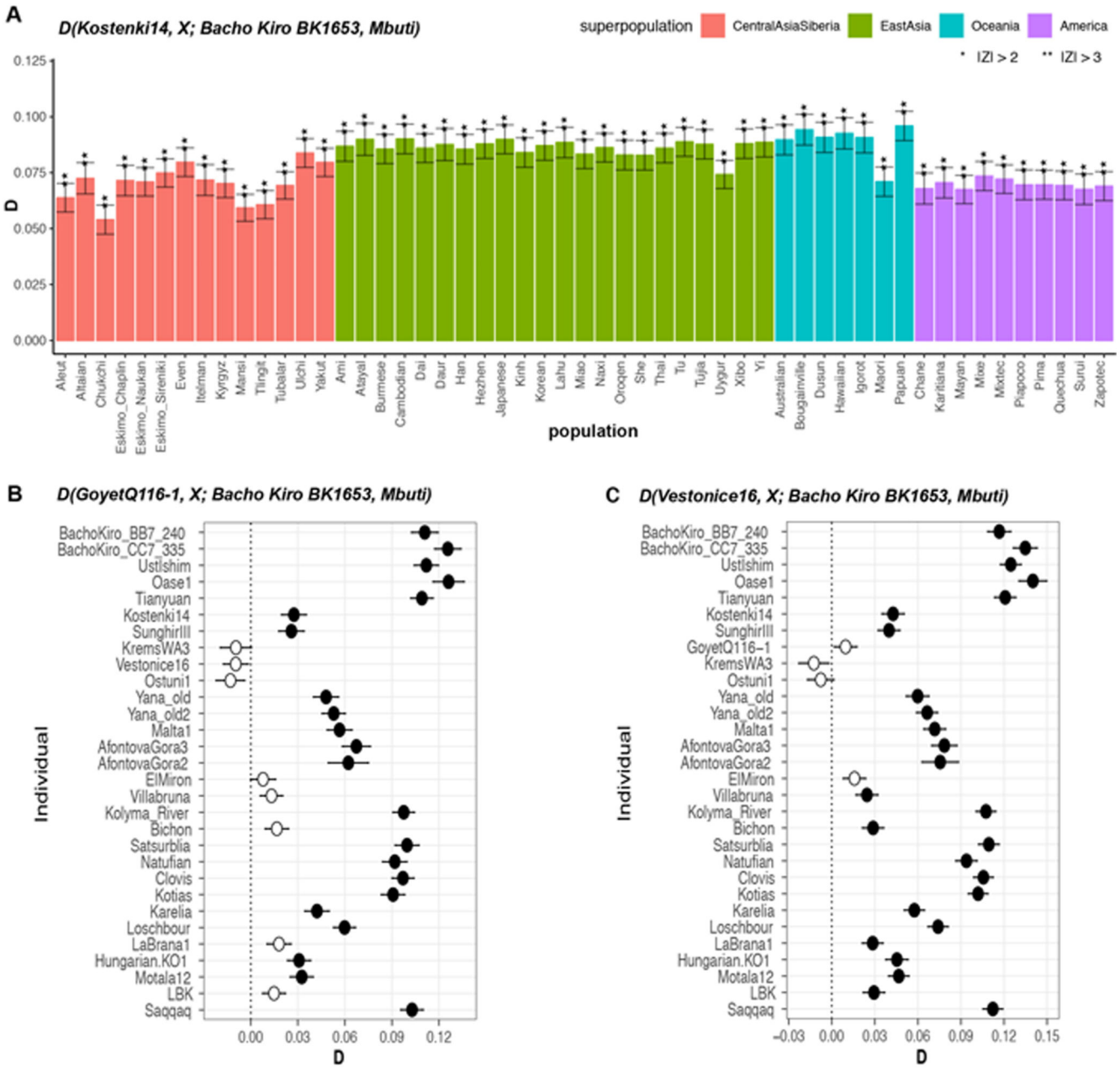
same specimen (intra-specimen), between different Bacho Kiro Cave specimens (inter-specimen) and between other ancient modern humans older than 30,000 cal. BP. The boxplots were drawn using the summary statistics geom_stat from the R-package ggplot; lower and upper hinges, first and third quartiles; whiskers, maximum value of 1.5x the interquartile range; centre line, median. SNPs across all autosomes of the 2200k Panel were used for the calculations (number of snps (nsnps) = 2,056,573) **c**, A PCA of 2,970 present-day humans genotyped on 597,573 SNPs with 22 ancient individuals older than 30,000 cal. BP projected onto the plane. **d**, A PCA of 1,444 present-day Eurasian and Native American individuals genotyped on 597,573 SNPs with 22 ancient individuals older than 30,000 cal. **BP**projected onto the plane. **c**, **d**, Grey dots denote present-day human genomes.



Extended Data Fig. 3. Heatmaps of outgroupf3-statistics corresponding to the amount of shared genetic drift between individuals and/or populations.

a, Genetic clustering of ancient individuals, including IUP Bacho Kiro Cave individuals, *BK1653* and *Oase1* based on the amount of shared genetic drift and calculated as *f3*(ancient$_1$, ancient$_2$; Mbuti). Lighter colours in this panel indicate higher *f3* values and correspond to higher shared genetic drift (nsnps = 2,056,573). **b**, **c**, Shared genetic drift between approximately 35,000-year-old *BK1653* (**b**; nsnps = 825,379) or approximately 38,000-year-old *Kostenki14* individuals[29,30] (**c**; nsnps = 1,676,430) and present-day human populations from the SGDP[31] calculated as f3(Bacho Kiro *BK1653/Kostenki14,* present-day humans; Mbuti). Three Mbuti individuals from the same panel[31] were used as an outgroup. Higher *f3* values[47] are indicated with warmer colours and correspond to higher shared genetic drift. Plotted f values were calculated using ADMIXTOOLS[28] as implemented in

admixr[61]. Coordinates for present-day humans were previously published[31]. The heatmap scale is consistent with those in Fig. 2a, Supplementary Figs. 5.1, 5.2.



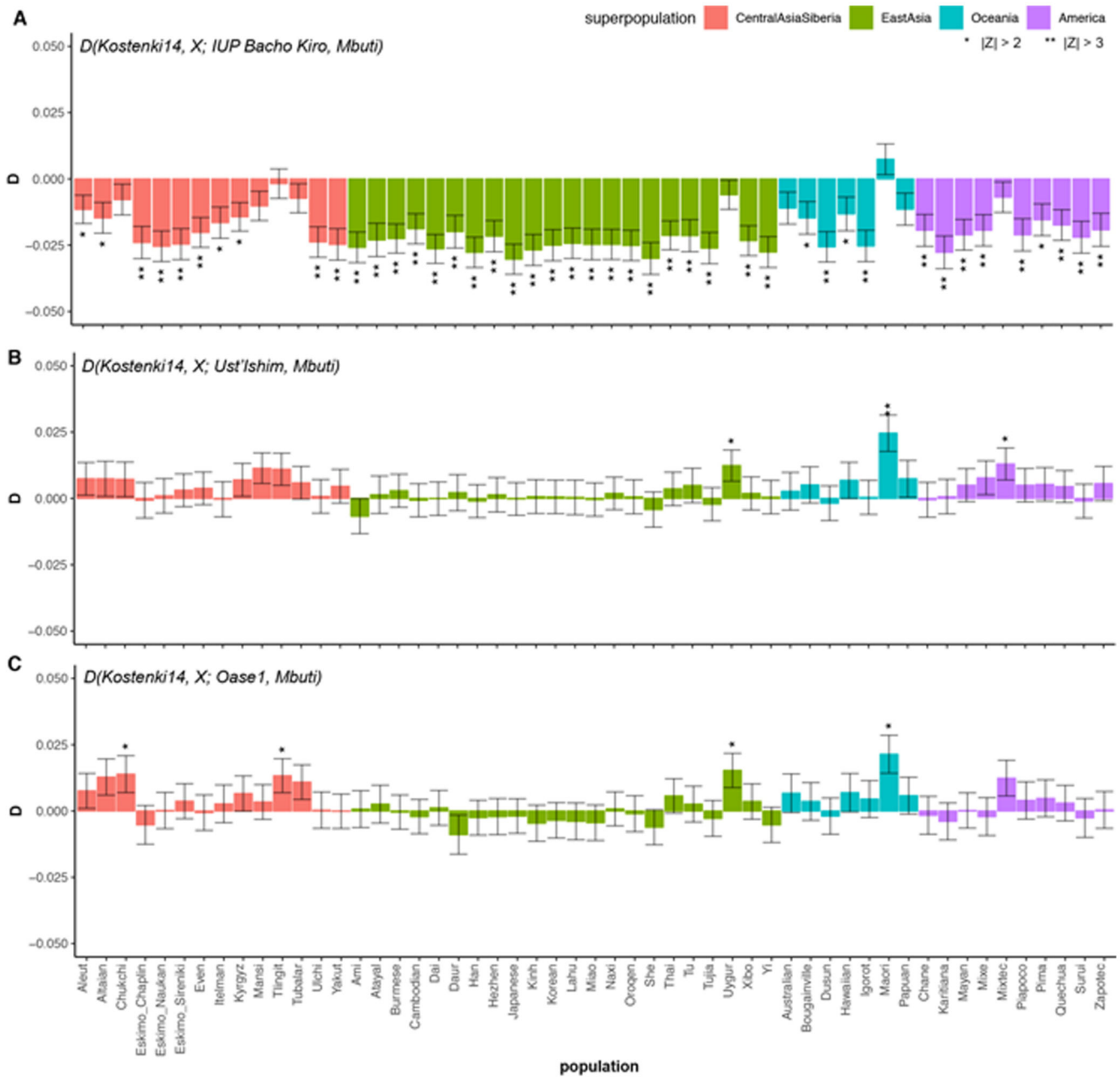
Extended Data Fig. 4. Population affinities of the approximately 35,000-year-old *BK1653* individual.

a, In contrast to the IUP Bacho Kiro Cave individuals, *BK1653* individual is significantly closer to the approximately 38,000-year-old *Kostenki14* individual[29,30] than to present-day non-African populations from Central Asia and Siberia, East Asia, South Asia, Oceania or the Americas, as calculated by D(*Kostenki14*, present-day humans; BachoKiro *BK1653,* Mbuti). D values for each comparison, plotted as barplots, were calculated using

Europe PMC Funders Author Manuscripts

ADMIXTOOLS[28] as implemented in admixr[61]. Present-day human genomes from the SGDP[31] were used in these statistics, and three Mbuti individuals from the same panel were used as an outgroup. **|Z| ≥ 3, *|Z| ≥ 2. **b**, **c**, *BK1653* shares significantly more alleles with the approximately 35,000-year-old *GoyetQ116-1* [29] (**b**) and the approximately 31,000-year-old *Vestonice16* [29] individuals (**c**) than with most other ancient modern humans. *D* values calculated as in **a.** Filled circles correspond to |Z| ≥ 3, and open circles indicate a |Z|<3 (not significant). Error bars in all panels show s.e. calculated using a weighted block jackknife[28] across all autosomes on the 2200k Panel (nsnps (Bacho Kiro *BK1653)* = 825,379) and a block size of 5 Mb.

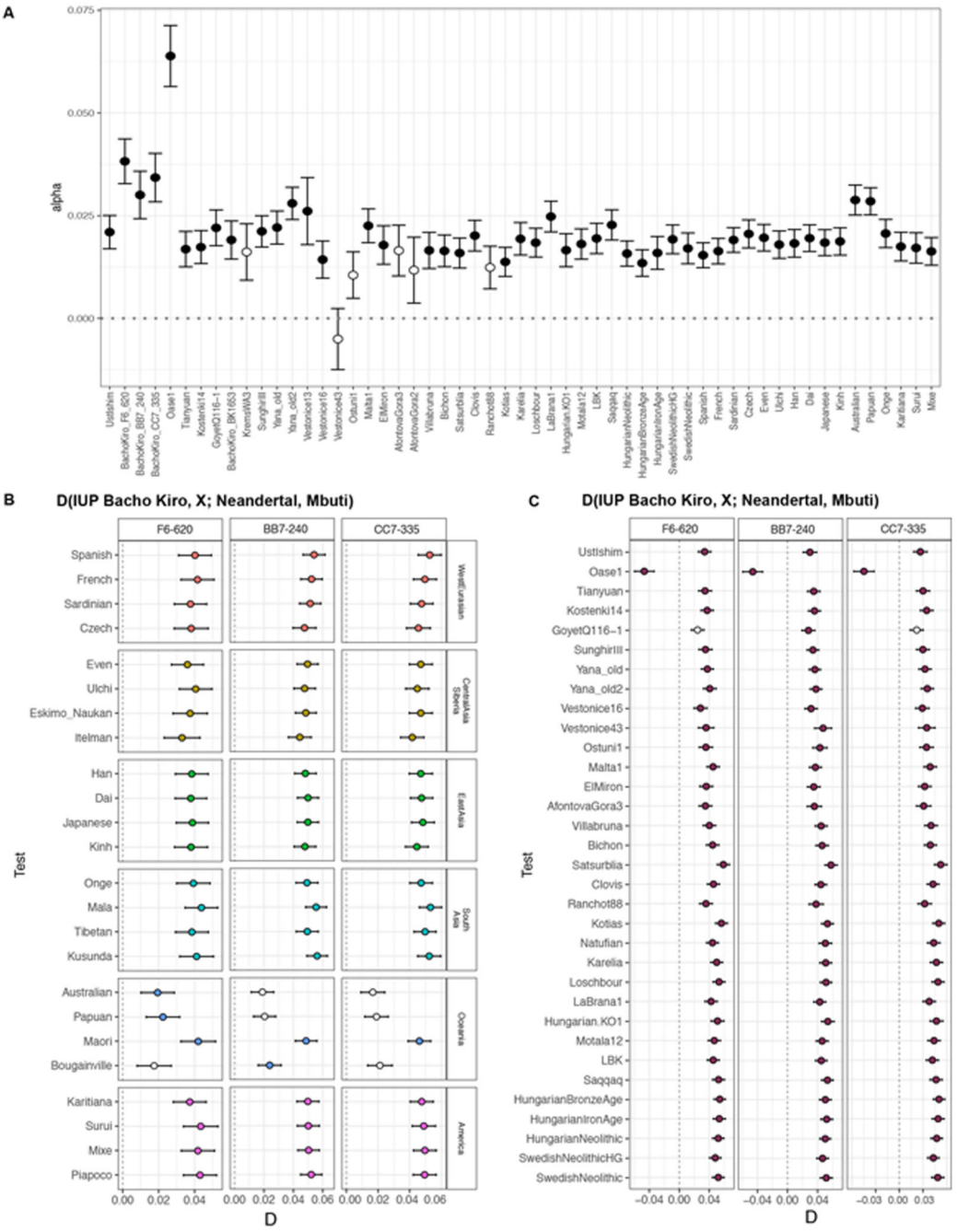
Europe PMC Funders Author Manuscripts



Extended Data Fig. 5. *D(Kostenki14,* X; IUP Bacho Kiro Cave individuals/*lUst'Ishiml Oase1,* Mbuti) where X is a present-day non-African population from Central Asia and Siberia, East Asia, South Asia, Oceania or the Americas.

D values for each comparison, plotted as barplots, were calculated using ADMIXTOOLS[28] as implemented in admixr[61]. Present-day human genomes from the SGDP[31] were used in these statistics, and three Mbuti individuals from the same panel were used as an outgroup. **|*Z*| > 3, *|*Z*| > 2. Error bars denote s.e. calculated using a weighted block jackknife[28] across all autosomes on the 2200k Panel and a block size of 5 Mb. **a**, A pool of three IUP Bacho Kiro Cave individuals (nsnps = 1,813,821). **b**, *Ust'Ishim* (nsnps = 1,951,462). **c**, *Oase1* (nsnps = 402,526).

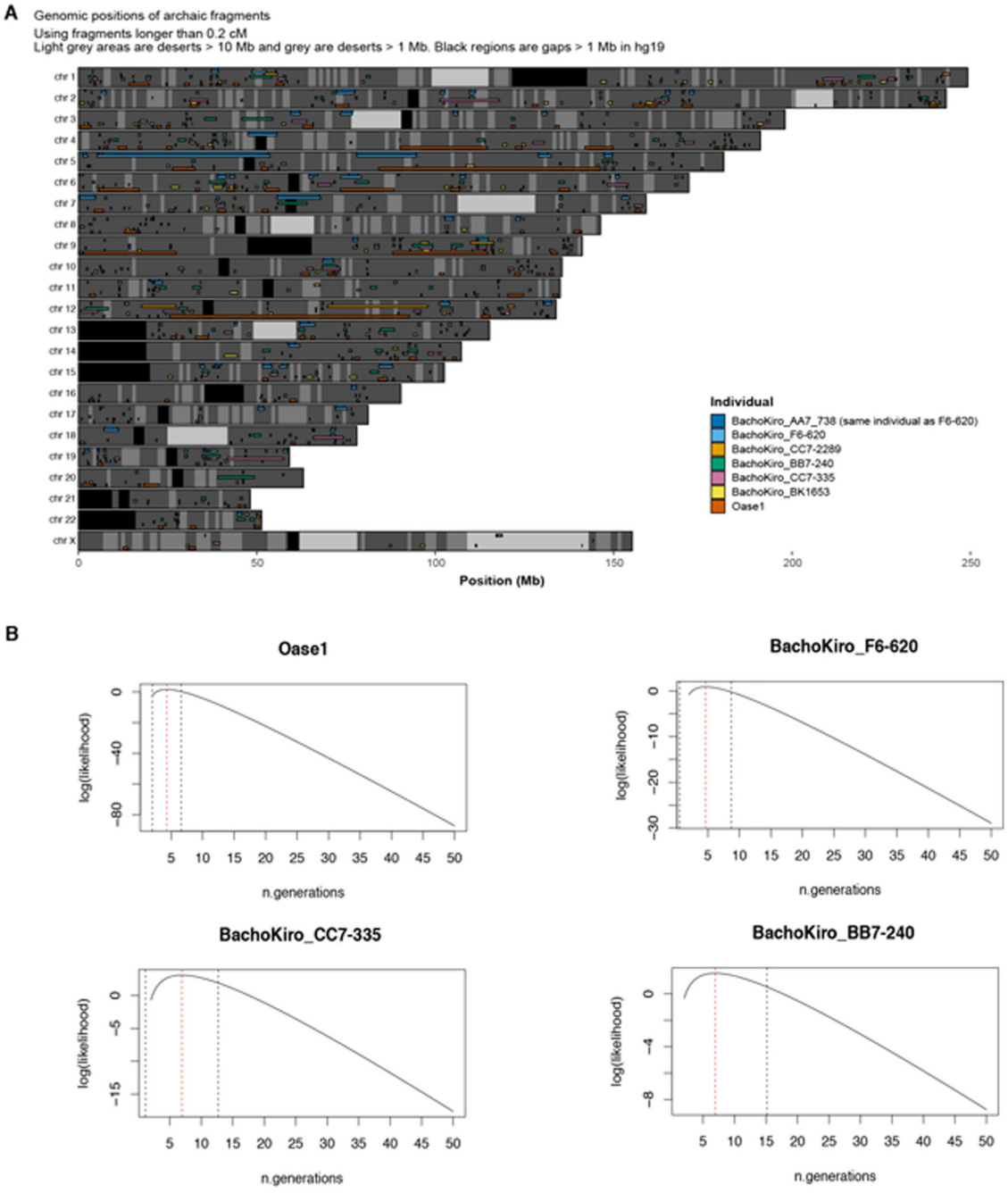
Europe PMC Funders Author Manuscripts



Extended Data Fig. 6. Neanderthal ancestry in IUP Bacho Kiro Cave individuals.

a, The proportion of Neanderthal ancestry in Bacho Kiro Cave individuals and other ancient and present-day modern humans calculated with a direct *f4* ratio that takes advantage of the two high-coverage Neanderthal genomes[9,10,41]. *f*4 ratio (alpha) values calculated using ADMIXTOOLS[28] as implemented in admixr[61]. **b, c**, Neanderthals[9,10,42] share significantly more derived alleles with IUP Bacho Kiro Cave individuals than with most present-day[31] (**b**) or ancient modern humans (**c**). *D* values calculated using ADMIXTOOLS[28] as implemented in admixr[61]. Filled circles correspond to $|Z| > 3$; open circles indicate $|Z|<3$ (not significant).

Error bars in all panels show s.e. calculated using a weighted block jackknife across all autosomes on the 2200k Panel (nsnps = 2,056,573) and a block size of 5 Mb .

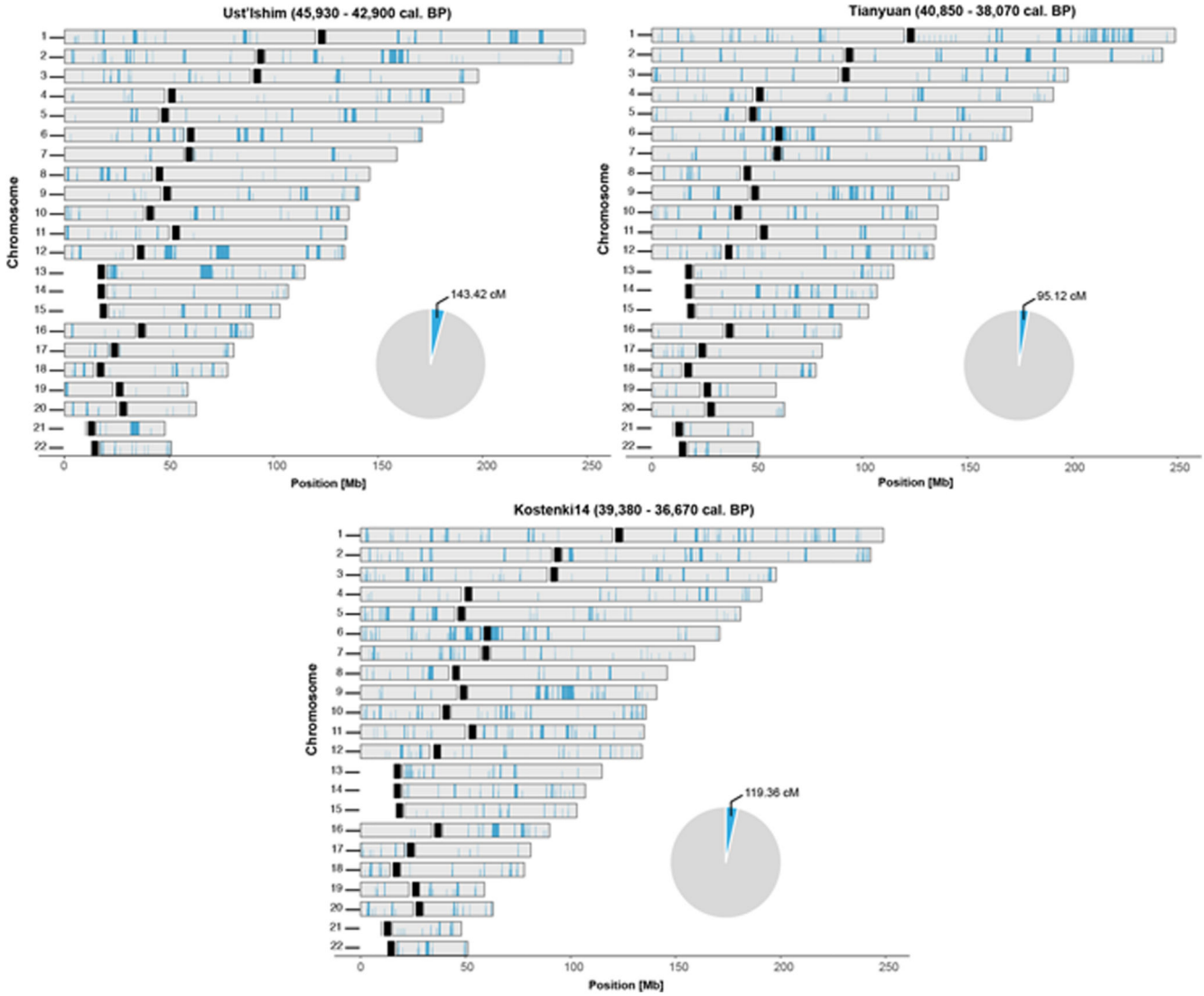


Extended Data Fig. 7. Segments of Neanderthal ancestry and estimates of the number of generations since the most recent Neanderthal ancestor.

a, A combined plot of the inferred Neanderthal segments in the genomes of IUP Bacho Kiro Cave individuals, *BK1653* and *Oase1,* including chromosome X, using a hidden Markov model approach (admixfrog)[44]. **b**, Maximum likelihood estimates (dashed red lines) of the

Europe PMC Funders Author Manuscripts

number of generations since a recent additional Neanderthal introgression into IUP Bacho Kiro Cave individuals and *Oase1.* Dashed black lines show 95% confidence intervals.



Extended Data Fig. 8. Spatial distribution of Neanderthal DNA in the *Ust'Ishim, Tianyuan* and *Kostenki14* genomes.

Segments corresponding to Neanderthal ancestry inferred using a hidden Markov model approach (admixfrog)[44] longer than 0.2 cM are indicated in blue. Centromeres are indicated in black. Pie charts indicate the total amount of Neanderthal DNA identified in each genome.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Extended Data Table 1

Amount of data generated for Bacho Kiro Cave and *Oase1* libraries for each SNP panel

Table shows number of fragments after merging all of the sequencing libraries together for each specimen. Fragments longer than 35 bp with a mapping quality of at least 25 that overlapped different target sites were used for reporting the number of SNPs on target.

SNP panel	Specimen	Number of sequenced fragments	Number of fragments ≥35bp	Number of mapped fragments ≥35bp, MQ≥25	Number of mapped fragments on target ≥35bp, MQ≥25	Number of unique fragments on target ≥35dp, MQ≥25	Number of SNPs on target	% of SNPs on target	Number of deaminated fragments on target ≥35dp, MQ≥25	Number of SNPs on target
390k	F6-620	42,944,497	31,607,676	10,664,089	5,417,563	3,610,135	372,571	94.61	879,932	294,287
	AA7-738	43,905,093	22,343,746	579,771	68,137	29,434	26,299	6.68	9,975	9,629
	BB7-240	48,679,210	32,531,469	5,428,952	1,666,056	893,179	311,587	79.13	307,661	191,684
	CC7-2289	16,999,931	12,126,017	51,649	19,810	9,097	8,659	2.2	1,828	1,816
	CC7-335	46,334,184	33,761,817	6,255,866	1,557,930	707,423	290,832	73.85	263,187	174,782
	BK1653	68,575,246	42,484,607	13,844,793	4,326,253	2,510,542	393,788	93.41	381,268	215,802
	Oasel	59,102,089	38,286,426	1,846,880	1,320,207	223,401	155,508	39.49	48,556	44,591
840k	F6-620	68,861,832	52,019,777	19,021,767	13,169,868	6,350,393	757,051	89.84	1,560,036	558,143
	AA7-738	62,671,011	46,507,342	790,242	478,993	64,038	58,037	6.89	21,929	21,376
	BB7-240	72,044,536	49,699,585	9,815,710	6,412,040	1,650,434	610,752	72.48	571,213	362,810
	CC7-2289	58,615,440	42,560,335	150,916	71,215	18,721	18,151	2.15	3,716	3,731
	CC7-335	68,794,275	51,231,565	10,390,207	6,599,859	1,290,666	561,737	66.66	479,284	326,926
	BK1653	66,744,975	49,155,219	14,281,786	9,230,491	3,840,020	694,592	82.43	594,397	356,855
	Oasel	56,745,317	37,030,590	3,031,789	1,920,548	368,252	270,601	32.11	80,709	75,569
1000k	F6-620	69,608,818	53,289,756	19,230,002	13,384,508	5,313,451	808,722	81.05	1,255,469	537,525
	AA7-738	61,551,239	42,095,902	677,982	398,801	41,360	39,660	3.97	13,567	13,677
	BB7-240	75,390,565	51,844,259	8,990,852	5,821,596	1,209,288	565,852	56.71	402,866	294,403
	CC7-2289	54,944,426	40,633,321	167,548	64,293	9,745	10,003	1	1,874	1,965
	CC7-335	72,464,739	54,221,955	10,907,573	6,935,567	1,062,963	536,521	53.77	383,734	286,514
	BK1653	72,387,295	49,998,398	14,529,900	9,637,535	2,911,357	726,130	72.77	434,351	305,619
	Oasel	59,348,519	38,586,583	2,347,235	1,434,618	272,943	221,020	22.15	56,420	55,309
Archaic admixture	F6-620	153,782,894	113,385,420	31,863,327	20,457,629	8,262,261	1,405,078	80.32	2,004,756	927,570
	AA7-738	74,474,294	55,404,191	741,606	316,366	72,962	67,495	3.86	22,510	24,423
	BB7-240	159,324,081	104,745,219	11,377,514	6,359,192	1,852,195	947,950	54.19	632,972	500,142
	CC7-2289	77,644,299	56,438,550	351,596	78,994	28,065	19,703	1.13	4,048	3,959
	CC7-335	156,715,471	112,448,992	14,570,648	8,134,133	1,650,897	907,904	51.9	606,177	488,615
	BK1653	159,168,237	111,693,225	22,866,297	13,162,057	4,591,355	1,257,365	71.87	702,281	533,274
	Oasel	52,319,743	33,767,952	1,719,613	890,910	306,637	278,234	15.9	64,447	71,112
“2200k”	F6-620	181,415,147	136,917,209	48,915,858	35,088,657	15,022,650	1,867,749	87.09	3,639,282	1,362,667

SNP panel	Specimen	Number of sequenced fragments	Number of fragments ≥35bp	Number of mapped fragments ≥35bp, MQ≥25	Number of mapped fragments on target ≥35bp, MQ≥25	Number of unique fragments on target ≥35dp, MQ≥25	Number of SNPs on target	% of SNPs on target	Number of deaminated fragments on target ≥35dp, MQ≥25	Number of SNPs on target
	AA7-738	168,127,343	110,946,990	2,047,995	1,328,039	131,193	123,265	5.75	44,375	44,610
	BB7-240	196,114,311	134,075,313	24,235,514	16,377,366	3,662,866	1,451,175	67.67	1,252,736	840,421
	CC7-2289	130,559,797	95,319,673	370,113	169,597	37,238	37,590	1.75	7,372	7,696
	CC7-335	187,593,198	139,215,337	27,553,646	18,181,311	2,967,786	1,353,413	63.11	1,093,477	776,283
	BK1653	207,707,516	141,638,224	42,656,479	29,863,646	9,098,112	1,728,159	80.59	1,387,333	874,287
	Oasel	175,195,925	113,903,599	7,225,904	4,722,435	850,586	646,646	30.15	183,143	177,336

Extended Data Table 2

Estimates of the number of generations before the most recent Neanderthal introgression in Bacho Kiro Cave individuals and *Oase1* (tracts >5 cM) obtained by calculating the complementary cumulative distribution (CCD) of the lengths of Neanderthal tracts

Estimates for tracts <5 cM and therefore representative of older introgression events are also shown.

Individual	tracts > 5cM n.generations (+/- 95% CI)	tracts < 5cM n.generations (+/- 95% CI)
BachoKiro BB7-240	11.9 (6.9-17.0)	89.9 (87.9-92.0)
BachoKiro CC7-2289	7.7(3.4-11.9)	60.3 (50.7- 70.0)
BachoKiro CC7-335	7.4 (5.0 -9.8)	99.3 (95.7- 102.6)
BachoKiro F6-620	3.9 (1.9-6.0)	87.7 (85.8- 89.7)
Oasel	3.9 (3.4-4.6)	89.0(86.6-91.2)

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank A. Weihmann and B. Schellbach for their help with DNA sequencing; R. Barr, P. Korlević and S. Tüpke for help with graphics; D. Reich and M. Slatkin for discussions and input; the Tourist Association STD “Bacho Kiro” in Dryanovo; the History Museum in Dryanovo; the Regional History Museum in Gabrovo; the National Museum of Natural History (NMNH) in Sofia; and N. Spassov. M.H. is supported by a Marie Sklodowska Curie Individual Fellowship (no. 844014). Q.F. was supported by the Strategic Priority Research Program (B) (XDB26000000) of CAS, NSFC (41925009, 41630102, 41672021). O.T.M. and S.C. were supported by a grant from the Ministry of Research and Innovation, CNCS - UEFISCDI, project number PN-III-P4-ID-PCCF-2016-0016, within PNCDI III and the EEA Grants 2014-2021, under Project contract no. 3/2019. F.W. has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 948365). P.S. was supported by the Vallee Foundation, the European Research Council (grant no. 852558), the Wellcome Trust (217223/Z/19/Z) and Francis Crick Institute core funding (FC001595) from Cancer Research UK, the UK Medical Research Council and the Wellcome Trust. This study was funded by the Max Planck Society and the European Research Council (grant agreement no. 694707 to S.P.).

Europe PMC Funders Author Manuscripts

Data availability

The aligned sequences of Bacho Kiro Cave individuals and *Oase1* have been deposited in the European Nucleotide Archive under the accession number PRJEB39134. Comparative data of present-day human genomes from the SGDP that were used in this study are available at https://www.simonsfoundation.org/simons-genome-diversity-project/. Comparative data used in this study, which include genotypes of 2,109 ancient and 2,974 present-day individuals compiled from published studies, are available in the EIGENSTRAT file format at https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data (version 37.2, released 22 February 2019). To determine the Y chromosome haplogroups of male individuals in this study, we used the Y-haplogroup tree from the International Society of Genetic Genealogy (ISOGG, available at http://www.isogg.org, version: 13.38).

References

1. Hublin J-J, et al. Initial Upper Palaeolithic Homo sapiens from Bacho Kiro Cave, Bulgaria. Nature. 2020; 581:299–302. DOI: 10.1038/s41586-020-2259-z [PubMed: 32433609]
2. Fewlass H, et al. A 14 C chronology for the Middle to Upper Palaeolithic transition at Bacho Kiro Cave, Bulgaria. Nat Ecol Evol. 2020; 4:794–801. DOI: 10.1038/s41559-020-1136-3 [PubMed: 32393865]
3. Higham T, et al. The earliest evidence for anatomically modern humans in northwestern Europe. Nature. 2011; 479:521–524. DOI: 10.1038/nature10484 [PubMed: 22048314]
4. Benazzi S, et al. Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. Nature. 2011; 479:525–528. DOI: 10.1038/nature10617 [PubMed: 22048311]
5. Hublin J-J. The modern human colonization of western Eurasia: when and where? Quat Sci Rev. 2015; 118:194–210. DOI: 10.1016/j.quascirev.2014.08.011
6. Higham T, et al. The timing and spatiotemporal patterning of Neanderthal disappearance. Nature. 2014; 512:306–309. DOI: 10.1038/nature13621 [PubMed: 25143113]
7. Fu Q, et al. An early modern human from Romania with a recent Neanderthal ancestor. Nature. 2015; 524:216–219. DOI: 10.1038/nature14558 [PubMed: 26098372]
8. Fu Q, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature. 2014; 514:445–449. DOI: 10.1038/nature13810 [PubMed: 25341783]
9. Prufer K, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014; 505:43–49. DOI: 10.1038/nature12886 [PubMed: 24352235]
10. Prufer K, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science. 2017; 358:655–658. DOI: 10.1126/science.aao1887 [PubMed: 28982794]
11. Green RE, et al. A draft sequence of the Neandertal genome. Science. 2010; 328:710–722. DOI: 10.1126/science.1188021 [PubMed: 20448178]
12. Benazzi S, et al. The makers of the Protoaurignacian and implications for Neandertal extinction. Science. 2015; 348:793–796. DOI: 10.1126/science.aaa2773 [PubMed: 25908660]
13. Yang MA, et al. 40,000-Year-Old Individual from Asia Provides Insight into Early Population Structure in Eurasia. Curr Biol. 2017; 27:3202–3208. DOI: 10.1016/j.cub.2017.09.030 [PubMed: 29033327]
14. Trinkaus E, et al. An early modern human from the Pe§tera cu Oase, Romania. Proc Natl Acad Sci USA. 2003; 100:11231–11236. DOI: 10.1073/pnas.2035108100 [PubMed: 14504393]
15. Kuhn SL, Zwyns N. Rethinking the initial Upper Paleolithic. Quat Int. 2014; 347:2938.doi: 10.1016/j.quaint.2014.05.040
16. Zwyns N, et al. The northern Route for Human dispersal in central and northeast Asia: new evidence from the site of Tolbor-16, Mongolia. Sci Rep. 2019; 9doi: 10.1038/s41598-019-47972-1

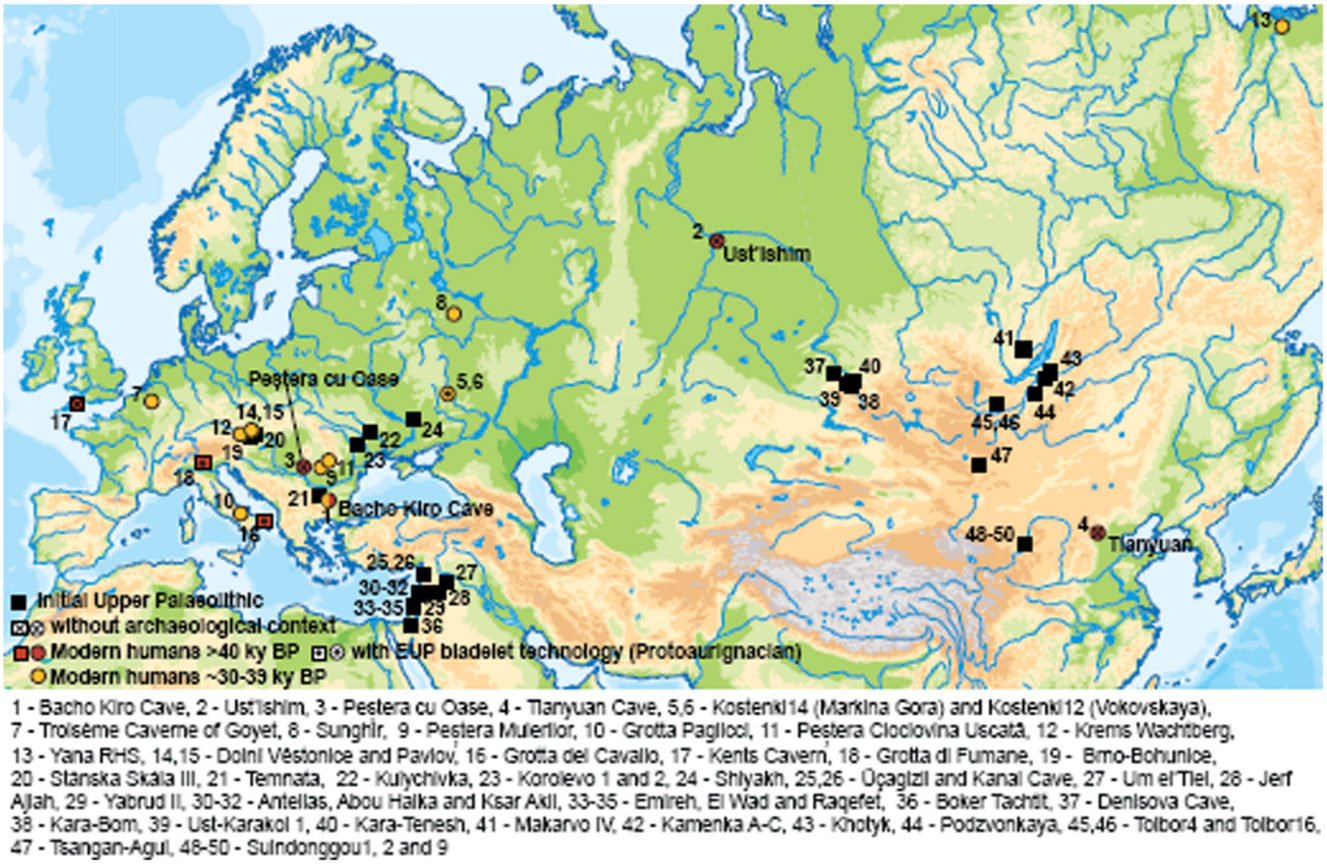
17. Korlevic P, et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. Biotechniques. 2015; 59:87–93. DOI: 10.2144/000114320 [PubMed: 26260087]
18. Fu Q, et al. DNA analysis of an early modern human from Tianyuan Cave, China. Proc Natl Acad Sci USA. 2013; 110:2223–2227. DOI: 10.1073/pnas.1221359110 [PubMed: 23341637]
19. Haak W, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015; 522:207–211. DOI: 10.1038/nature14317 [PubMed: 25731166]
20. Briggs AW, et al. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci USA. 2007; 104:14616–14621. DOI: 10.1073/pnas.0704665104 [PubMed: 17715061]
21. Meyer M, et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. Nature. 2016; 531:504–507. DOI: 10.1038/nature17405 [PubMed: 26976447]
22. Peyregne S, Peter BM. AuthentiCT: a model of ancient DNA damage to estimate the proportion of present-day DNA contamination. Genome Biol. 2020; 21:246.doi: 10.1186/s13059-020-02123-y [PubMed: 32933569]
23. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. BMC bioinformatics. 2014; 15:356.doi: 10.1186/s12859-014-0356-4 [PubMed: 25420514]
24. Mittnik A, et al. Kinship-based social inequality in Bronze Age Europe. Science. 2019; 366:731–734. DOI: 10.1126/science.aax6219 [PubMed: 31601705]
25. Petr M, et al. The evolutionary history of Neanderthal and Denisovan Y chromosomes. Science. 2020; 369:1653–1656. DOI: 10.1126/science.abb6460 [PubMed: 32973032]
26. Poznik GD, et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nat Genet. 2016; 48:593–599. DOI: 10.1038/ng.3559 [PubMed: 27111036]
27. Kutanan W, et al. Contrasting paternal and maternal genetic histories of Thai and Lao populations. Mol Biol Evol. 2019; 36:1490–1506. DOI: 10.1093/molbev/msz083 [PubMed: 30980085]
28. Patterson N, et al. Ancient admixture in human history. Genetics. 2012; 192:1065–1093. DOI: 10.1534/genetics.112.145037 [PubMed: 22960212]
29. Fu Q, et al. The genetic history of Ice Age Europe. Nature. 2016; 534:200–205. DOI: 10.1038/nature17993 [PubMed: 27135931]
30. Seguin-Orlando A, et al. Genomic structure in Europeans dating back at least 36,200 years. Science. 2014; 346:1113–1118. DOI: 10.1126/science.aaa0114 [PubMed: 25378462]
31. Mallick S, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 2016; 538:201–206. DOI: 10.1038/nature18964 [PubMed: 27654912]
32. Lazaridis I, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 2014; 513:409–413. DOI: 10.1038/nature13673 [PubMed: 25230663]
33. Lazaridis I, et al. Genomic insights into the origin of farming in the ancient Near East. Nature. 2016; 536:419–424. [PubMed: 27459054]
34. Sikora M, et al. The population history of northeastern Siberia since the Pleistocene. Nature. 2019; 570:182–188. DOI: 10.1038/s41586-019-1279-z [PubMed: 31168093]
35. Rasmussen M, et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature. 2010; 463:757–762. DOI: 10.1038/nature08835 [PubMed: 20148029]
36. Moreno-Mayar JV, et al. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. Nature. 2018; 553:203–207. DOI: 10.1038/nature25173 [PubMed: 29323294]
37. Moreno-Mayar JV, et al. Early human dispersals within the Americas. Science. 2018; 362doi: 10.1126/science.aav2621
38. Rasmussen M, et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature. 2014; 506:225–229. DOI: 10.1038/nature13025 [PubMed: 24522598]
39. Rasmussen M, et al. The ancestry and affiliations of Kennewick Man. Nature. 2015; 523:455–458. DOI: 10.1038/nature14625 [PubMed: 26087396]
40. Sikora M, et al. Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. Science. 2017; 358:659–662. DOI: 10.1126/science.aao1807 [PubMed: 28982795]
41. Petr M, Paabo S, Kelso J, Vernot B. Limits of long-term selection against Neandertal introgression. Proc Natl Acad Sci USA. 2019; 116:1639–1644. DOI: 10.1073/pnas.1814338116 [PubMed: 30647110]

42. Mafessoni F, et al. A high-coverage Neandertal genome from Chagyrskaya Cave. Proc Natl Acad Sci USA. 2020; 117:15132–15136. DOI: 10.1073/pnas.2004944117 [PubMed: 32546518]
43. Meyer M, et al. A high-coverage genome sequence from an archaic Denisovan individual. Science. 2012; 338:222–226. DOI: 10.1126/science.1224344 [PubMed: 22936568]
44. Peter BM. 100,000 years of gene flow between Neandertals and Denisovans in the Altai mountains. bioRxiv. 2020
45. Vernot B, et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science. 2016; 352:235–239. DOI: 10.1126/science.aad9416 [PubMed: 26989198]
46. Sankararaman S, Mallick S, Patterson N, Reich D. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. Curr Biol. 2016; 26:1241–1247. DOI: 10.1016/j.cub.2016.03.037 [PubMed: 27032491]
47. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2013. URL http://www.R-project.org/
48. Gansauge M-T, Aximu-Petri A, Nagel S, Meyer M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. Nat Protoc. 2020; 15:2279–2300. DOI: 10.1038/s41596-020-0338-0 [PubMed: 32612278]
49. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 2012; 40:e3.doi: 10.1093/nar/gkr771 [PubMed: 22021376]
50. Renaud G, Stenzel U, Kelso J. leeHom: adaptor trimming and merging for Illumina sequencing reads. Nucleic Acids Res. 2014; 42:e141.doi: 10.1093/nar/gku699 [PubMed: 25100869]
51. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26:589–595. DOI: 10.1093/bioinformatics/btp698 [PubMed: 20080505]
52. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. DOI: 10.1093/bioinformatics/btp352 [PubMed: 19505943]
53. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
54. Raghavan M, et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature. 2014; 505:87–91. DOI: 10.1038/nature12736 [PubMed: 24256729]
55. Olalde I, et al. Derived immune and ancestral pigmentation alleles in a 7,000-year- old Mesolithic European. Nature. 2014; 507:225–228. DOI: 10.1038/nature12960 [PubMed: 24463515]
56. Gallego Llorente M, et al. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. Science. 2015; 350:820–822. DOI: 10.1126/science.aad2879 [PubMed: 26449472]
57. Keller A, et al. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nat Commun. 2012; 3doi: 10.1038/ncomms1701
58. Gamba C, et al. Genome flux and stasis in a five millennium transect of European prehistory. Nat Commun. 2014; 5doi: 10.1038/ncomms6257
59. Jones ER, et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat Commun. 2015; 6doi: 10.1038/ncomms9912
60. Hajdinjak M, et al. Reconstructing the genetic history of late Neanderthals. Nature. 2018; 555:652–656. DOI: 10.1038/nature26151 [PubMed: 29562232]
61. Petr M, Vernot B, Kelso J. admixr—R package for reproducible analyses using ADMIXTOOLS. Bioinformatics. 2019; 35:3194–3195. [PubMed: 30668635]

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Europe PMC Funders Author Manuscripts



1 - Bacho Kiro Cave, 2 - Ust'Ishim, 3 - Peştera cu Oase, 4 - Tianyuan Cave, 5,6 - Kostenki14 (Markina Gora) and Kostenki12 (Vokovskaya), 7 - Troisème Caverne of Goyet, 8 - Sunghir, 9 - Peştera Muierilor, 10 - Grotta Paglicci, 11 - Peştera Cioclovina Uscată, 12 - Krems Wachtberg, 13 - Yana RHS, 14,15 - Dolní Věstonice and Pavlov, 16 - Grotta del Cavallo, 17 - Kents Cavern, 18 - Grotta di Fumane, 19 - Brno-Bohunice, 20 - Stanska Skála III, 21 - Temnata, 22 - Kulychivka, 23 - Korolevo 1 and 2, 24 - Shlyakh, 25,26 - Üçağızlı and Kanal Cave, 27 - Um el Tlel, 28 - Jerf Ajjah, 29 - Yabrud II, 30-32 - Antellas, Abou Halka and Ksar Akil, 33-35 - Emireh, El Wad and Raqefet, 36 - Boker Tachtit, 37 - Denisova Cave, 38 - Kara-Bom, 39 - Ust-Karakol 1, 40 - Kara-Tenesh, 41 - Makarvo IV, 42 - Kamenka A-C, 43 - Khotyk, 44 - Podzvonkaya, 45,46 - Tolbor4 and Tolbor16, 47 - Tsangan-Agui, 48-50 - Sulndonggou1, 2 and 9

Figure 1. Archaeological sites that have yielded genetic data and/or IUP assemblages.

Sites with modern human genome-wide data older than 40 kyr BP (red circles) or older than 30 kyr BP (yellow circles), sites in Europe with modern human remains older than 40 kyr BP (red squares) and sites with IUP assemblages (black squares).

Europe PMC Funders Author Manuscripts

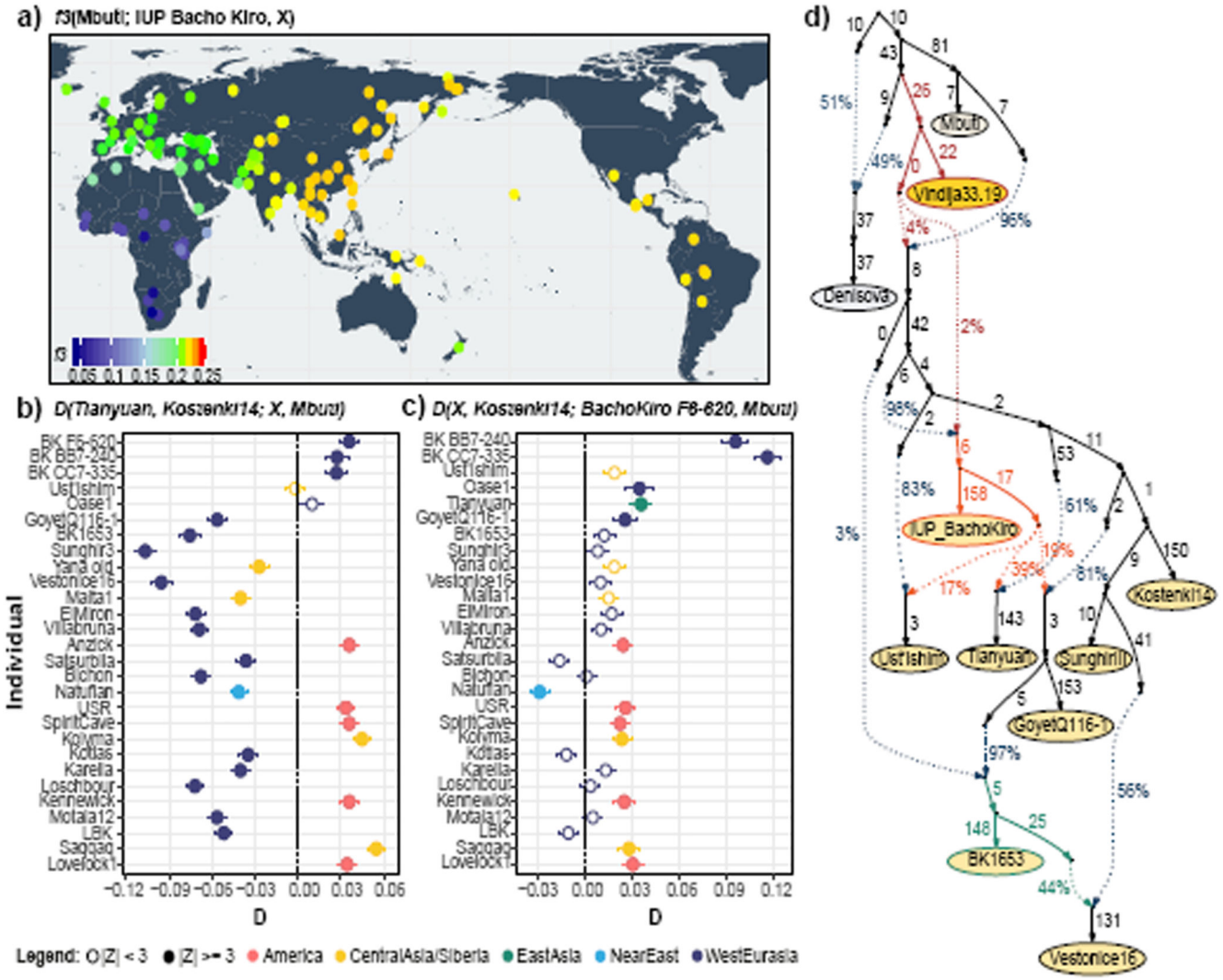


Figure 2. Population affinities of the IUP Bacho Kiro Cave individuals.

a, Allele sharing (*f3*) between the IUP Bacho Kiro Cave individuals and present-day populations (X) from the Simons Genome Diversity Project (SGDP)[31] after their separation from an outgroup (Mbuti). Warmer colours on the map[47] correspond to higher *f3* values (higher shared genetic drift). **b**, IUP Bacho Kiro Cave individuals share significantly more alleles (proportions of allele sharing or *D* values plotted on x-axis) with the roughly 40,000-year-old *Tianyuan* individual[13] than with the approximately 38,000-year-old *Kostenki14* individual[29,30]. Calculated as *D(Tianyuan, Kostenki14*; X, Mbuti). **c)** *F6-620* shares significantly more alleles with *Oase1* [7], *GoyetQ116-1* [29] and ancient Siberians and Native American individuals than with *Kostenki14*. Calculated as D(X, *Kostenki14*; *F6-620*, Mbuti). **b**, **c**, Filled circles indicate a significant value ($|Z| \geq 3$); open circles, $|Z| < 3$. Whiskers correspond to 1 s.e. calculated across all autosomes (1,813,821 SNPs) using a weighted block jackknife[28] and a block size of 5 Mb. BK, Bacho Kiro. **d**, Admixture graph relating Bacho Kiro Cave individuals and ancient humans older than 30 kyr **BP**. This model uses 281,732 overlapping SNPs in all individuals and fits the data with a single outlier ($Z =$

3.22). Numbers on solid branches correspond to the estimated drift in *f2* units of squared frequency difference; labels on dotted edges give admixture proportions.

Europe PMC Funders Author Manuscripts

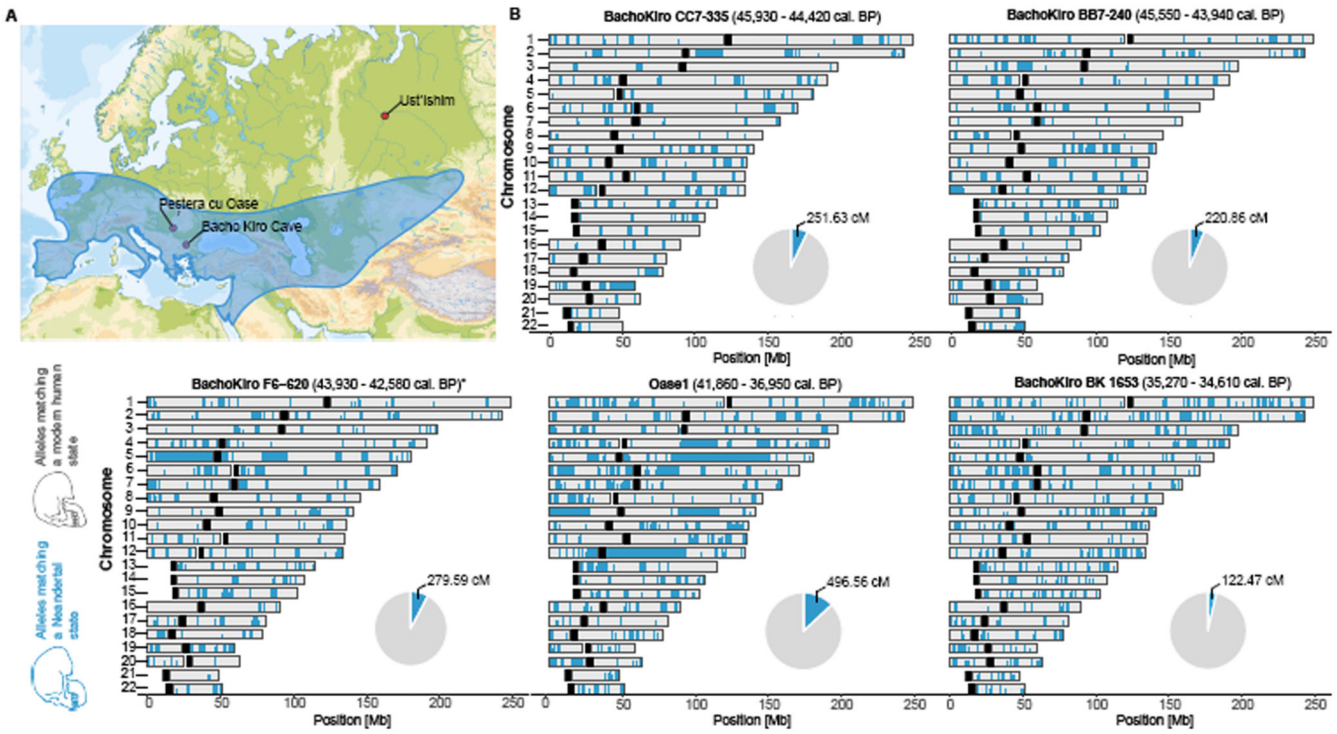


Figure 3. Geographical distribution of Neanderthal archaeological sites and genome–wide distribution of Neanderthal alleles in the genomes of ancient modern humans.

a, Neanderthal geographical range (blue) and the locations of Pe§tera cu Oase, Bacho Kiro Cave and where the femur of Ust'Ishim individual was found. **b**, Distribution of Neanderthal DNA in ancient modern human genomes. Neanderthal DNA segments longer than 0.2 cM are indicated in blue. Pie charts indicate the total proportion of Neanderthal DNA identified in each genome. Centromeres are shown in black.

Europe PMC Funders Author Manuscripts