# The population history of northeastern Siberia since the Pleistocene

Sikora, M.; Pitulko, V.V.; Sousa, V.C.; Allentoft, M.E.; Vinner, L.; Rasmussen, S.; ... ; Willerslev, E.

ARTICLE

https://doi.org/10.1038/s41586-019-1279-z

# The population history of northeastern Siberia since the Pleistocene

Martin Sikora[1,43]*, Vladimir V. Pitulko[2,43]*, Vitor C. Sousa[3,4,5,43], Morten E. Allentoft[1,43], Lasse Vinner[1], Simon Rasmussen[6,41], Ashot Margaryan[1], Peter de Barros Damgaard[1], Constanza de la Fuente[1,42], Gabriel Renaud[1], Melinda A. Yang[7], Qiaomei Fu[7], Isabelle Dupanloup[8], Konstantinos Giampoudakis[9], David Nogués-Bravo[9], Carsten Rahbek[9], Guus Kroonen[10,11], Michaël Peyrot[11], Hugh McColl[1], Sergey V. Vasilyev[12], Elizaveta Veselovskaya[12,13], Margarita Gerasimova[12], Elena Y. Pavlova[2,14], Vyacheslav G. Chasnyk[15], Pavel A. Nikolskiy[2,16], Andrei V. Gromov[17], Valeriy I. Khartanovich[17], Vyacheslav Moiseyev[17], Pavel S. Grebenyuk[18,19], Alexander Yu. Fedorchenko[20], Alexander I. Lebedintsev[18], Sergey B. Slobodin[18], Boris A. Malyarchuk[21], Rui Martiniano[22], Morten Meldgaard[1,23], Laura Arppe[24], Jukka U. Palo[25,26], Tarja Sundell[27,28], Kristiina Mannermaa[27], Mikko Putkonen[25], Verner Alexandersen[29], Charlotte Primeau[29], Nurbol Baimukhanov[30], Ripan S. Malhi[31,32], Karl-Göran Sjögren[33], Kristian Kristiansen[33], Anna Wessman[27,34], Antti Sajantila[25], Marta Mirazon Lahr[1,35], Richard Durbin[22,36], Rasmus Nielsen[1,37], David J. Meltzer[1,38], Laurent Excoffier[4,5]* & Eske Willerslev[1,36,39,40]*

**Northeastern Siberia has been inhabited by humans for more than 40,000 years but its deep population history remains poorly understood. Here we investigate the late Pleistocene population history of northeastern Siberia through analyses of 34 newly recovered ancient genomes that date to between 31,000 and 600 years ago. We document complex population dynamics during this period, including at least three major migration events: an initial peopling by a previously unknown Palaeolithic population of 'Ancient North Siberians' who are distantly related to early West Eurasian hunter-gatherers; the arrival of East Asian-related peoples, which gave rise to 'Ancient Palaeo-Siberians' who are closely related to contemporary communities from far-northeastern Siberia (such as the Koryaks), as well as Native Americans; and a Holocene migration of other East Asian-related peoples, who we name 'Neo-Siberians', and from whom many contemporary Siberians are descended. Each of these population expansions largely replaced the earlier inhabitants, and ultimately generated the mosaic genetic make-up of contemporary peoples who inhabit a vast area across northern Eurasia and the Americas.**

Northeastern Siberia (the modern Russian Far East) is one of the most remote and extreme of the environments that were colonized by humans in the Pleistocene epoch. Extending from the Taimyr Peninsula in the west to the Pacific Ocean in the east, and north from the border between China and Russia to the Arctic Ocean, the region is currently home to dozens of diverse ethnolinguistic groups. Recent genetic studies of the indigenous peoples of this land have revealed complex patterns of admixture, which are argued to have occurred largely within the past 10,000 years[1–3]. Humans have been in the region for far longer[4–6], but their origins and the demographic processes of this deeper population history are largely unknown. The earliest, most secure archaeological evidence for human occupation in this region comes from an artefact-rich, high-latitude (approximately 70° N) site on the Yana River (Siberia) named Yana RHS, which dates to 31,600 calibrated years before present (taken as AD 1950)[4] (Fig. 1). Yana RHS yielded a flake-based stone tool industry and sophisticated bone and ivory artefacts, which are reminiscent of technologies seen in the Upper Palaeolithic in Eurasia and southern Siberia[5,7] (Extended Data Fig. 1). By the time of the Last Glacial Maximum (LGM) at about 23–19 thousand years ago (ka)[8], the archaeological culture seen at Yana RHS had disappeared from northeastern Siberia. Instead, archaeological lithic assemblages at this time are dominated by a distinctive microblade technology, which

[1]Lundbeck Foundation GeoGenetics Centre, University of Copenhagen, Copenhagen, Denmark. [2]Palaeolithic Department, Institute for the History of Material Culture, Russian Academy of Science, St Petersburg, Russia. [3]Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal. [4]Institute of Ecology and Evolution, University of Bern, Bern, Switzerland. [5]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [6]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Copenhagen, Denmark. [7]Key Laboratory of Vertebrate Evolution and Human Origins, Center for Excellence in Life and Paleoenvironment, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, China. [8]Swiss Integrative Center for Human Health SA, Fribourg, Switzerland. [9]Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark. [10]Department of Nordic Studies and Linguistics, University of Copenhagen, Copenhagen, Denmark. [11]Leiden University Centre for Linguistics, Leiden University, Leiden, The Netherlands. [12]Institute of Ethnology and Anthropology, Russian Academy of Science, Moscow, Russia. [13]Russian State University for Humanities (RSUH), Moscow, Russia. [14]Polar Geography Department, Arctic & Antarctic Research Institute, St Petersburg, Russia. [15]St Petersburg Pediatric Medical University, St Petersburg, Russia. [16]Geological Institute, Russian Academy of Sciences, Moscow, Russia. [17]Peter the Great Museum of Anthropology and Ethnography, Russian Academy of Sciences, St Petersburg, Russia. [18]North-East Interdisciplinary Scientific Research Institute, Far East Branch, Russian Academy of Sciences, Magadan, Russia. [19]Northeast State University, Magadan, Russia. [20]Institute of Archaeology and Ethnography of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia. [21]Institute of Biological Problems of the North, Far East Branch, Russian Academy of Sciences, Magadan, Russia. [22]Department of Genetics, University of Cambridge, Cambridge, UK. [23]University of Greenland, Nuuk, Greenland. [24]Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland. [25]Department of Forensic Medicine, University of Helsinki, Helsinki, Finland. [26]Forensic Genetics Unit, National Institute for Health and Welfare, Helsinki, Finland. [27]Department of Cultures, Archaeology, University of Helsinki, Helsinki, Finland. [28]Institute of Biotechnology, University of Helsinki, Helsinki, Finland. [29]Laboratory of Biological Anthropology, Department of Forensic Medicine, University of Copenhagen, Copenhagen, Denmark. [30]Shejire DNA, Almaty, Kazakhstan. [31]Department of Anthropology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [32]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [33]Department of Historical Studies, University of Gothenburg, Gothenburg, Sweden. [34]Department of Archaeology, University of Turku, Turku, Finland. [35]Leverhulme Centre for Human Evolutionary Studies, Department of Archaeology, University of Cambridge, Cambridge, UK. [36]Wellcome Sanger Institute, Cambridge, UK. [37]Department of Integrative Biology, University of California, Berkeley, CA, USA. [38]Department of Anthropology, Southern Methodist University, Dallas, TX, USA. [39]GeoGenetics Groups, Department of Zoology, University of Cambridge, Cambridge, UK. [40]The Danish Institute for Advanced Study, The University of Southern Denmark, Odense, Denmark. [41]Present address: Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [42]Present address: Human Genetics Department, University of Chicago, Chicago, IL, USA. [43]These authors contributed equally: Martin Sikora, Vladimir V. Pitulko, Vitor C. Sousa, Morten E. Allentoft. *e-mail: martin.sikora@bio.ku.dk; pitulko.vladimir@gmail.com; laurent.excoffier@iee.unibe.ch; ewillerslev@bio.ku.dk

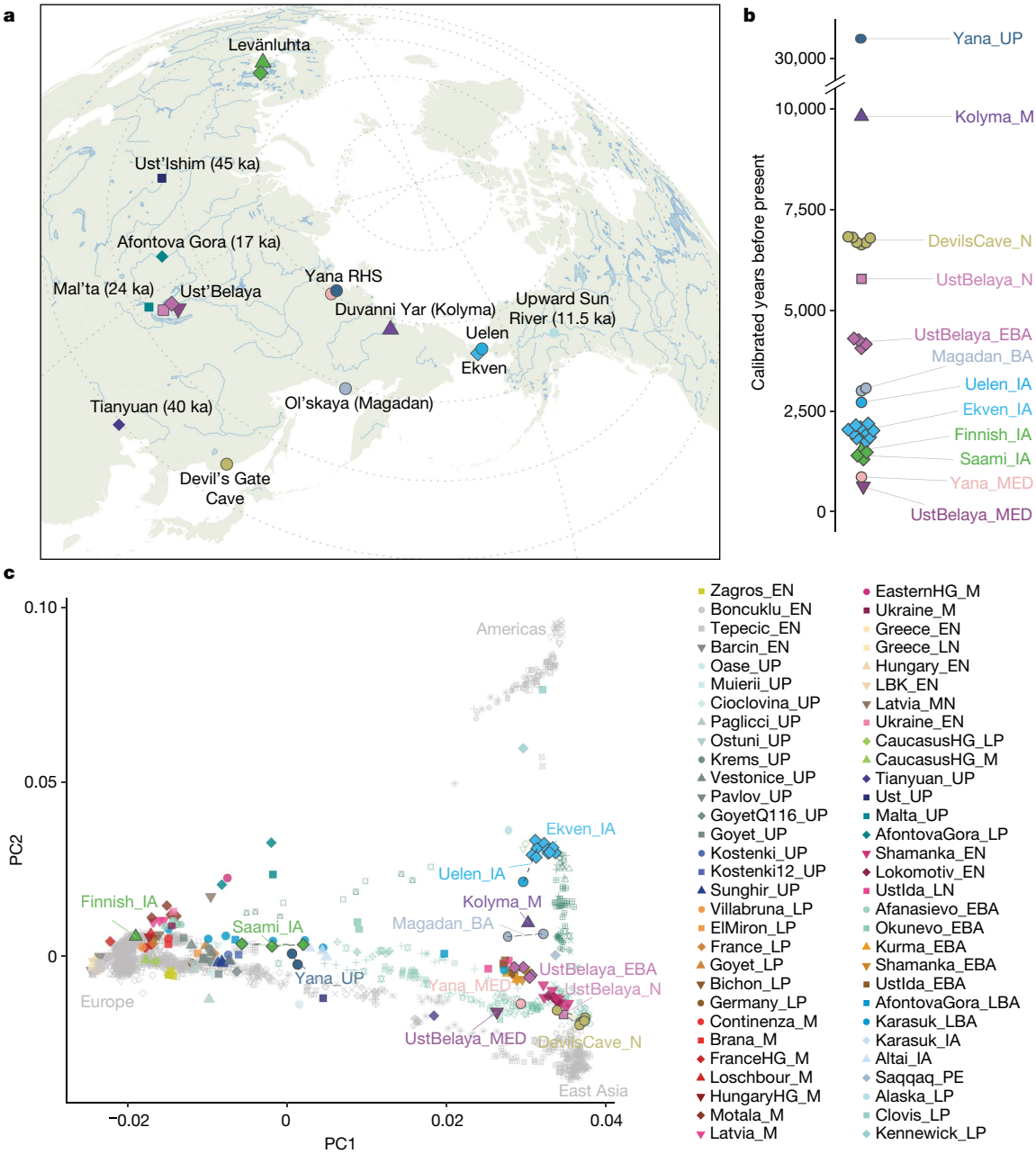



**Fig. 1 | Genetic structure of ancient northeast Siberians. a**, Sampling locations of newly reported individuals, as well as select previously published individuals with dates in parentheses. **b**, Sample dates. **c**, Principal component (PC) analysis of 257 ancient individuals projected onto a set of 1,541 modern Eurasian and American individuals. Abbreviations in group labels: UP, Upper Palaeolithic; LP, Late Palaeolithic; M, Mesolithic; EN, Early Neolithic; MN, Middle Neolithic; LN, Late Neolithic; N, Neolithic; EBA, Early Bronze Age; LBA, Late Bronze Age; IA, Iron Age; PE, Palaeo-Eskimo; MED, Medieval; HG, hunter-gatherer.

spread in a time-transgressive manner north and east out of the Amur region[9,10]. This tradition did not reach Chukotka or cross the Bering land bridge (Beringia) until the end of the Pleistocene—later than the earliest known sites in the Americas. Changes in material culture in northeastern Siberia continued into the late Holocene epoch, but it remains debated whether these successive cultural complexes represent in situ technological evolution or migrations of distinct groups of people. In the case of the latter, it is unclear how these groups may have been related to each other, to contemporary Siberians, or to Native Americans, whose ancestors may have emerged in this region (or, at the very least, traversed it en route to Beringia).

To investigate these questions, we used single-end shotgun sequencing to generate whole genomes of 34 ancient individuals with associated radiocarbon ages that range from 31,600 to 600 calibrated years before present (Fig. 1, Supplementary Information 1, 2, Supplementary Tables 1, 2). Our data include samples from ancient individuals that are key for understanding Siberian population history: two high-quality genomes sequenced from fragmented milk teeth (Supplementary Information 2) that were recovered from Yana RHS (Yana1 and Yana2, which produced genomes at 25× and 7× coverage, respectively), which are the oldest Pleistocene human remains found to date at such high latitude; a high-coverage genome (14× coverage) of an individual from the Duvanny Yar site at the Kolyma River (Kolyma1), dated to about 9.8 ka; 14 genomes from ancient individuals from sites in far eastern Chukotka (Ekven and Uelen) and the northern coast of the Sea of Okhotsk (Ol'skaya, near the city of Magadan), ranging in age between 3 and 2 ka; 6 individuals from the approximately 7,600-year-old site of Devil's Gate Cave in northern East Asia[11]; 6 genomes of individuals from southern Siberia (Ust'Belaya in the Lake Baikal region) dating to between 6.5 ka and 0.6 ka; an individual from the Yana River but from

a different locality than Yana RHS, here named 'Young Yana' (dated to 0.8 ka); as well as 4 individuals from the Levänluhta site in southwestern Finland, dated to approximately 1.5 ka. We analysed these data in the context of large panels of previously published ancient and present-day individuals (Supplementary Information 3, Supplementary Tables 3, 4).

## Upper Palaeolithic peoples at Yana RHS

The Yana RHS human remains represent the earliest direct evidence of human presence in northeastern Siberia, a population that we refer to as Ancient North Siberians (ANS). The two Yana RHS individuals were unrelated males who carry mitochondrial haplogroup U (which is predominant among ancient West Eurasian hunter-gatherers) and Y chromosome haplogroup P1, which is ancestral to haplogroups Q and R—which are widespread among present-day Native Americans and Eurasians, respectively[12,13] (Extended Data Fig. 2, Supplementary Information 4, 5, Supplementary Table 1). Genetic clustering using outgroup-$f_3$ statistics demonstrates broad genetic similarities with a wide range of present-day populations across Northern Eurasia and the Americas. This contrasts with other Upper Palaeolithic Eurasians, such as those from Sunghir[14] and Tianyuan[15]—who share similar amounts of genetic drift with present-day populations that are geographically more restricted to western Eurasia and East Asia, respectively (Extended Data Fig. 3). Despite their extreme northeastern Siberian geographical location, the Yana RHS individuals are genetically closer to West Eurasians, although symmetry tests using $f_4$ statistics reject tree-like clade relationships between them, early West Eurasians (Sunghir) and East Asians (Tianyuan) (Extended Data Fig. 3d, e, Extended Data Table 1, Supplementary Information 6). Using admixture graphs and outgroup-based estimation of mixture proportions, we find that ANS can be modelled as early West Eurasian with an approximately 22% contribution from early East Asians (Extended Data Fig. 3f, Supplementary Information 6). Demographic modelling of the high-coverage individual Yana1 using a site-frequency-spectrum-based framework indicates an early divergence of the ANS lineage at about 39 ka (95% confidence interval (CI) 32.2–45.8 ka), concomitant with substantial gene flow (approximately 29%; 95% CI 21.3–40.1%) from East Asians, an event that probably occurred very soon after the latter diverged from West Eurasians 43.1 ka (95% CI 33.4–48.6 ka) (Fig. 2a, Supplementary Information 7). Thus, the ANS population represents a distinct lineage with affinities to both early West Eurasians and early East Asians, albeit in a 2:1 ratio. These complex relationships among early Eurasian groups are also supported by the presence of East Asian ancestry and mitochondrial haplogroup M in Western Europe by 35 ka[16,17]. Finally, we estimate about 2% Neanderthal ancestry in the Yana RHS genomes, which is contained in longer genomic tracts than in present-day individuals, as also shown for other Upper Palaeolithic Eurasians[14,16] (Supplementary Information 6).

We next investigated how the Yana RHS individuals relate to an ancient Siberian population represented by the Mal'ta individual (dated to 24 ka) from the Lake Baikal region (previously termed 'Ancestral North Eurasians'), from which Native Americans derive around 40% of their ancestry[18]. We find that the Mal'ta individual shares more alleles with the Yana RHS individuals than with other West Eurasian hunter-gatherers (for example, $f_4$(Mbuti, Mal'ta; Sunghir, Yana) $> 0$; $Z = 3.99$) (Extended Data Table 1, Supplementary Information 6). The Mal'ta and Yana RHS individuals also exhibit a similar pattern of genetic affinities to both early West Eurasians and East Asians, consistent with previous studies[19,20]. In admixture graphs, the Mal'ta individual can be successfully fit as a descendant of the ANS lineage, with a minor contribution from an early Eurasian lineage that is ancestrally related to Late Palaeolithic hunter-gatherers from the Caucasus (Extended Data Fig. 3e, f). The Ancestral North Eurasian lineage of the Mal'ta individual can thus be considered a descendant of the ANS lineage, and our results therefore suggest that—by 31.6 ka—people related to the ANS were probably widespread across northeastern Eurasia.

The two Yana RHS individuals were contemporaneous, which provides an opportunity to investigate relatedness and levels of inbreeding at this remote Upper Palaeolithic settlement. We find that the two individuals were not closely related and did not exhibit signatures of recent inbreeding, with a moderately large estimate of recent effective-population size of up to 500 individuals (Extended Data Fig. 4, Supplementary Information 4, 5). Our results mirror those observed at Sunghir, an early (approximately 34,000-year-old) European Upper Palaeolithic site located about 4,500 km southwest of Yana RHS, which reinforces the view that wide-ranging mate exchange networks were present among Upper Palaeolithic foragers across the pre-LGM landscape[14].

## Ancient Palaeo-Siberians and Native Americans

Following the occupation at Yana RHS, there is an absence of archaeological sites in northeastern Siberia. The area is re-occupied in the later part of the LGM, about 20 ka, when sites preserving a very distinctive stone tool technology become widespread. This gap in the archaeological sequence is critical, as it was within the intervening period that the population ancestral to Native Americans emerged[18,21], although no genomes from individuals of this age have been recovered in northeastern Siberia to date. We find that the Kolyma1 individual (dated to 9.8 ka)—who represents a lineage that formed after about 30 ka, which we name 'Ancient Palaeo-Siberian'—documents the first major genetic shift that we observe in the region (Extended Data Fig. 5). Principal component analysis, outgroup-$f_3$ statistics and mitochondrial DNA and Y chromosome haplogroups (G1b and Q1a1b, respectively) demonstrate a close affinity between Ancient Palaeo-Siberians and present-day Koryaks, Itelmen and Chukchis, as well as with Native Americans (Extended Data Fig. 5, Supplementary Information 6). Admixture graph modelling shows that the Kolyma1 individual derives from a mixture of East Asian and ANS-related ancestry, similar to that found in Native Americans, although the East Asian contribution is greater in Kolyma1 than in Native Americans (75% versus 63%) (Extended Data Fig. 3f, Supplementary Information 6). For both Ancient Palaeo-Siberians and Native Americans, ANS-related ancestry is more closely related to Mal'ta than to the Yana individuals (Extended Data Fig. 3f), which rejects the hypothesis that the Yana lineage contributed directly to later Ancient Palaeo-Siberians or Native American groups.

We then estimated the demographic parameters of population history models that included the Ancient Palaeo-Siberian Kolyma1 and Native Americans, represented either by Ancient Beringians[21] (the Upward Sun River 1 (USR1) individual) or present-day Native Americans (Karitiana individuals; Supplementary Information 7). We find that their ancestors diverged at about 30 ka (95% CI 26.8–36.4 ka) from present-day East Asians (represented by Han individuals), consistent with previous results[21]; the subsequent divergence of Ancient Palaeo-Siberians from the population ancestral to Ancient Beringians or Native Americans occurred at about 24 ka (95% CI 20.9–27.9 ka) (Fig. 2, Supplementary Information 7). We infer that both Ancient Palaeo-Siberians and Ancient Beringians received ANS-related gene flow at a similar time (Kolyma1 at 20.2 ka (95% CI 15.5–23.7 ka) and USR1 at 19.7 ka (95% CI 13.3–23.5 ka)). This gene flow amounts to 16.6% (95% CI 7.5–22.2%) of ANS ancestry into the Kolyma1 individual and 18.3% (95% CI 9.8–20.3%) of ANS ancestry into USR1, which is comparable to the estimates we obtained using admixture graphs. An alternative model with a single admixture pulse in the ancestral population of the Kolyma1 and USR1 individuals showed a comparable likelihood (Supplementary Information 7), but differences in the estimated proportions of ANS-related ancestry between Kolyma1 and USR1 favour the two-independent-pulses model. The Kolyma1 individual thus represents the closest relative to the ancestral Native American population in northeastern Siberia that has been found to date.

Changes in climatic conditions are commonly put forward as principal drivers of Pleistocene population movement and regional abandonment in Siberia. We used palaeoclimatic modelling to infer geographical locations in Siberia that were suitable for human occupation between 48 and 12 ka to further investigate this hypothesis. When humans were present at Yana RHS, interstadial climatic conditions meant that a large stretch of the Arctic coast of northeastern Siberia was

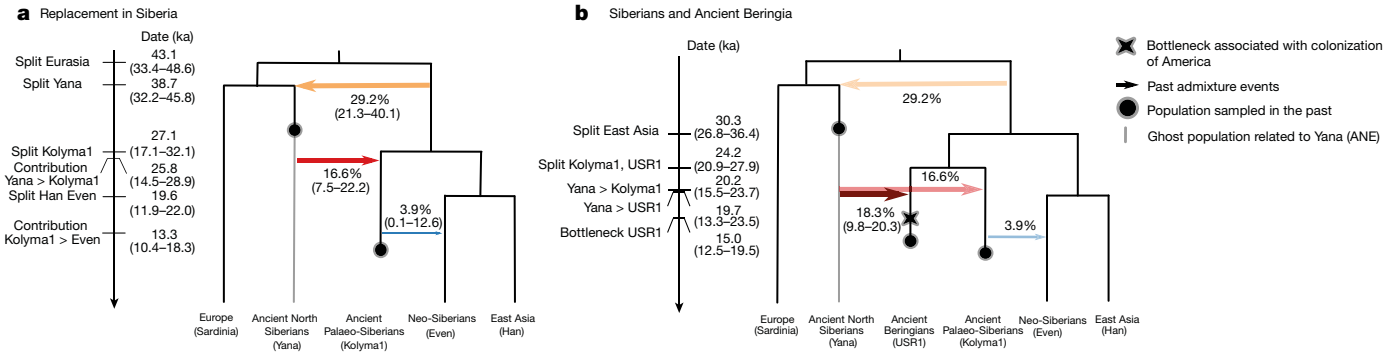



**Fig. 2 | Demographic modelling of Siberian and Native American populations. a**, **b**, Inferred parameters for models with ancient and modern Siberian populations (**a**) and Siberian and Ancient Beringian populations (**b**). Dates of events (ka) are indicated on the left, and admixture estimates are given as percentages along the arrows. Point parameter estimates are shown with associated 95% CI in parentheses. Neanderthal contribution was modelled as an unsampled ('ghost') Neanderthal population contributing 3% into the ancestors of all Eurasian populations, and an extra 0.5% into the Asian lineage. Neanderthal effective population size and split times were fixed according to recent estimates based on the genome-wide site frequency spectrum[34]. Shaded arrows for the model in **b** indicate admixture proportions that were fixed to values estimated under model **a**. ANE, Ancestral North Eurasians.

suitable for human occupation (Extended Data Fig. 6, Supplementary Information 8). Conditions in the region became harsher during the LGM, which is consistent with the absence of archaeological sites in the area at this time. However, our models suggest the existence of a refugium across southern Beringia during the LGM (for example, see Alaska in the '22 ka' panel of Extended Data Fig. 6a), consistent with previous reports[22]. Therefore, a possible scenario for gene flow during the formation of the early Native American and Ancient Palaeo-Siberian gene pools may have involved early ANS-related groups occupying southern Beringia during the LGM, and subsequently admixing with East Asian-related peoples who expanded northwards towards the end of the LGM. This scenario would also be consistent with a divergence of Ancient Beringians from ancestral Native Americans in eastern Beringia rather than in Siberia, which is supported by genetic data (scenario 2 in ref. [21]). Alternatively, the closer affinity of both Kolyma1 and Native Americans to Mal'ta, rather than the Yana RHS individuals, could suggest a more-southwesterly location (Lake Baikal region) for the admixture, with a northward expansion after the LGM. The alternative scenario is supported by archaeological evidence for a movement south during the LGM, although the genetic isolation that is observed between Asians and ancestral Native Americans after about 23 ka would have required the maintenance of a structured population during the LGM—this, in turn, implies that Ancient Palaeo-Siberians and ancestral Native Americans occupied different refugia. Regardless, our results support the broader implication that glacial and post-glacial climate change was a major driver of human population history across northern Eurasia.

## Holocene transformations across Siberia and Beringia

Our genomic data provide further insights into the timing and the origins of peoples involved in more-recent gene flow during the Holocene, across what was by then the Bering Sea. The Saqqaq individual from Greenland (dated to 4 ka[23])—representing Palaeo-Eskimos—clusters with the Kolyma1 individual, but shows a greater affinity to East Asians than Kolyma1 (Fig. 1, Extended Data Table 1). By modelling the Saqqaq individual as a mixture of Ancient Palaeo-Siberians (represented by Kolyma1) and East Asians (represented by the Devil's Gate Cave individuals), we estimate that the Saqqaq individual contains around 20% East Asian ancestry (Extended Data Fig. 7a, b, Supplementary Information 6, Supplementary Table 5). Individuals from the Uelen and Ekven Neo-Eskimo sites (dated to 2.7–1.6 ka), located on the Siberian shore of the Bering Sea, cluster closely with contemporary Inuit individuals (Fig. 1, Extended Data Fig. 8a). We fit them as a mixture of 69% Ancient Palaeo-Siberian (Kolyma1) and 31% Native American ancestry (represented by the Anzick individual associated with Clovis assemblages), thereby documenting a 'reverse' gene flow across the Bering Sea (from northwestern North America to northeastern Siberia). This is consistent with linguistic evidence for a back-migration into Siberia of a population speaking a language from the Eskimo–Aleut family (Extended Data Table 1; Extended Data Fig. 7, Supplementary Information 6, 9, Supplementary Table 5). The source population of this gene flow post-dates the divergence of the USR1 individual from other Native Americans (at about 20.9 ka[21]), as the individuals at Ekven share more alleles with ancient Native Americans (individuals Anzick-1 and Kennewick Man) than with Ancient Beringians (USR1); this confirms previous results from present-day Inuits[24] (Extended Data Table 1). Using linkage-disequilibrium-based admixture dating[25] with the Saqqaq and Anzick-1 individuals as source populations, we find a significant signal of admixture linkage disequilibrium, with an estimated date of 100–200 generations before the age of the individuals (Supplementary Information 6). Although these estimates show considerable uncertainty (owing to the limited sample size and genomic coverage), they nevertheless indicate that gene flow from Native Americans into Siberia took place possibly as early as about 5 ka (about 100 generations before the earliest individual from Uelen and Ekven), well after the disappearance of Beringia. Finally, we investigated the genetic affinity between North American populations who speak Na-Dene languages (including Athabascans) and Siberian populations[26], which has previously been suggested to relate either to gene flow from a Palaeo-Eskimo source[27] or to an unknown source population that was more closely related to Koryaks[21]. We find that the Kolyma1 individual is a better proxy for this source population than the Saqqaq individual, using both admixture graph modelling (Supplementary Information 6) and chromosome-painting symmetry tests (Extended Data Fig. 5); this provides additional evidence against a genetic contribution from a migration of Palaeo-Eskimos (Saqqaq) to contemporary speakers of Na-Dene languages.

The Holocene archaeological record of northeast Siberia is marked by further changes in material culture. We used a temporal transect of ancient Siberians from about 6 ka to 500 years ago to investigate whether these cultural transitions were associated with genetic changes. We find that, in a principal component analysis of present-day non-African populations, most contemporary Siberian populations are arranged along two separate genetic clines. The majority of individuals (referred to as 'Neo-Siberians') lie on an east–west cline that is stretched out along principal component 1, between European individuals at one end and East Asian individuals at the other (Fig. 1). A secondary cline between East Asians and Native Americans along principal component 2 includes speakers of Palaeo-Siberian languages and Inuit populations (Extended Data Fig. 8c). Estimated mixture proportions show that Ancient Palaeo-Siberian ancestry (Kolyma1) was common in other Siberian regions until the early Bronze Age (Extended Data Fig. 7),

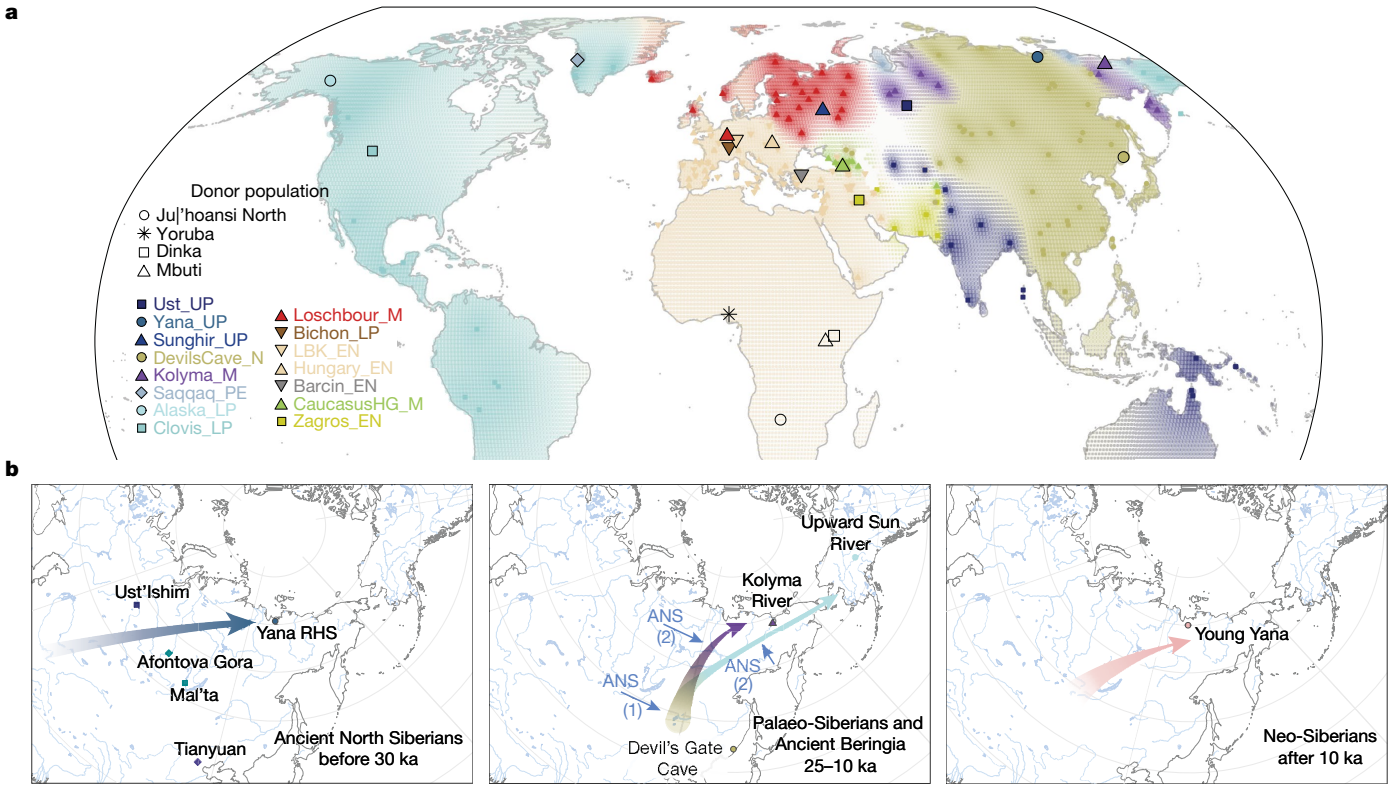



**Fig. 3 | Genetic legacy of ancient Eurasians. a**, Worldwide map of top haplotype donations inferred by chromopainter. Coloured symbols represent a modern recipient population; the colour and shape indicate the donor population that contributes the highest fraction of haplotypes to that recipient population. Geographical locations of donor populations used in this analysis (modern Africans and ancient Eurasians) are indicated by the corresponding larger symbols with a black outline added. Extended regions of shared top donors are visualized by spatial interpolation of the respective donor population colour. **b**, Major hypothesized migrations into northeast Siberia. Arrows indicate putative migrations that gave rise to ANS (left), Ancient Palaeo-Siberians (middle) and Neo-Siberians (right). Key sample locations for the respective time slice are indicated with symbols. Small blue arrows in the middle panel indicate possible ANS admixture scenarios: (1) admixture in southern Siberia and (2) admixture in Beringia.

but thereafter was largely restricted to the northeast; this is exemplified by an individual from Ol'skaya dated to 3 ka, who closely resembles present-day Koryaks and Itelmen. Using present-day Even individuals to represent Neo-Siberians in our demographic model, we find evidence for a divergence from East Asians at about 20 ka, with only low levels (about 4%) of Ancient Palaeo-Siberian gene flow at around 13 ka (Fig. 2, Supplementary Information 7). Thus, our data provide evidence for a second major population turnover in northeastern Siberia in which Neo-Siberians expanding northwards largely replaced Ancient Palaeo-Siberians; this pattern is also evident in chromosome-painting analyses of present-day populations (Fig. 3). A notable exception are the Ket (an isolated population that speaks a Yeniseian language), who have previously been described as rich in 'Ancestral North Eurasian' ancestry and as having genetic links to Palaeo-Eskimos[26]. The Ket fall on a secondary cline parallel to Neo-Siberians in the chromosome-painting analysis, and carry about 40% Ancient Palaeo-Siberian ancestry (Extended Data Figs. 8c, 7). Our findings are consistent with the proposed linguistic link between the Yeniseian-speaking Ket and Na-Dene-speaking Athabascan populations (Supplementary Information 9), likely through shared ancestry with an Ancient Palaeo-Siberian metapopulation that was more widespread across northern Eurasia before the expansion of Neo-Siberian peoples.

Our Holocene transect reveals additional complexity in recent times, with evidence for further episodes of gene flow and local population replacements. A notable example is found in the Lake Baikal region in southern Siberia; here, the genomes from Ust'Belaya and neighbouring Neolithic and Bronze Age sites show a succession of three distinct genetic ancestries over an approximately 6,000-year period. The earliest individuals show predominantly East Asian ancestry (represented by individuals from Devil's Gate Cave) (Fig. 1, Extended Data Figs. 7, 8, followed by a resurgence of Ancient Palaeo-Siberian ancestry (up to about 50% ancestry) in the early Bronze Age, as well as the influence of West Eurasian steppe ancestry (about 10% ancestry from individuals associated with the Afanasievo culture) (Extended Data Fig. 7, Supplementary Table 5). This is consistent with previous reports of gene flow from an unknown Ancestral North Eurasian-related source into Lake Baikal hunter-gatherers[28]. Our results suggest a southward expansion of Ancient Palaeo-Siberians as a possible source. This is consistent with the replacement of Y chromosome lineages observed at Lake Baikal, from predominantly haplogroup N in the Neolithic to haplogroup Q during the early Bronze Age[28]. Finally, the most recent individual from Ust'Belaya (about 600 years old) falls along the Neo-Siberian cline and is similar to the Young Yana individual from northeastern Siberia (about 760 years old), demonstrating the geographical extent of the Neo-Siberian demographic expansion in the recent past. We show that most populations on the Neo-Siberian cline can be modelled as predominantly East Asian, with varying proportions of West Eurasian steppe ancestry; the largest proportions of this latter ancestry are observed among Altaian populations, both of the present day and in more-recent Bronze Age and Iron Age archaeological contexts (Extended Data Fig. 7, Supplementary Table 5). Together, these findings demonstrate considerable population movement and admixture throughout southern and eastern Siberia during the Holocene, in which groups dispersed in multiple directions—but without clear evidence of the wholesale population replacement seen earlier in the Pleistocene.

Finally, we investigated the geographical extent of these processes of population flux across northern Eurasia. The notable spatial pattern of Ancestral Palaeo-Siberian and East Asian ancestry in present-day populations (Fig. 3) suggests that Ancient Palaeo-Siberian ancestry was once widespread, probably as far west as the Ural Mountains. At the western edge of northern Eurasia, genetic and strontium-isotope data from ancient individuals at the Levänluhta site (Supplementary Information 1) document the presence of Saami ancestry in south-western Finland in the late Holocene, at about 1.5 ka. This ancestry component is currently limited to the northern fringes of the region, which mirrors the pattern that is observed for Ancient Palaeo-Siberian ancestry in northeastern Siberia. However, although the ancient Saami individuals contain ancestry from an eastern source, we find that this is modelled better by East Asians than by Ancient Palaeo-Siberians, which suggests that the influence of Ancient Palaeo-Siberians probably did not extend across the Ural Mountains into western Eurasia (Extended Data Fig. 7, Supplementary Table 5). East–west gene flow continued to shape the gene pool of the Finnish population into the very recent past. We observe West Eurasian admixture in present-day Saami; by contrast, present-day Finns have greater Siberian ancestry than the ancient Levänluhta individual (Extended Data Table 1), who may represent the Scandinavian component in the dual-origin (Uralic and Scandinavian) gene pool of Finns today.

## Discussion

Our findings reveal that the population history of northeastern Siberia is far more complex than previously inferred from the contemporary genetic record. It involved, at a minimum, three major population expansions and subsequent large-scale replacements during the late Pleistocene and early Holocene, with smaller-scale population fluxes since then. These three major waves are also clearly documented in the archaeological record. The initial movement into the region represents a now-extinct ANS population diversifying at about 38 ka, soon after the basal split between West Eurasians and East Asians, which is represented by the archaeological culture found at Yana RHS[4,29]. This finding is consistent with other studies that have shown that this period was a time of rapid expansion of early modern humans across Eurasia[13]. The arrival of people who carried ancestry from East Asia and their admixture with descendants of the ANS lineage at about 20–18 ka led to the formation of the Ancient Palaeo-Siberian and Native American lineages. In the archaeological record, this is reflected by the spread of a microblade technology that accompanies the post-LGM contraction of the once-extensive mammoth steppe[10]. This group was, in turn, largely replaced by Neo-Siberians in the early to mid-Holocene. Our data suggest that the Neo-Siberians received ANS-related ancestry indirectly through admixture with Ancient Palaeo-Siberian groups at about 13 ka, and possibly later from Bronze Age groups from the central Asian steppe after around 5 ka. A signal of Australasian ancestry that has been observed at a very low frequency in some modern and ancient South American populations[30–32] is not evident in any of the ancient Siberian or Beringian samples sequenced here, or in previous studies[21].

We find that—despite the complex pattern of population admixture throughout the past 40,000 years—the first inhabitants of northeastern Siberia (represented by the Yana RHS individuals) were not the direct ancestors of either Native Americans or present-day Siberians, although traces of their genetic legacy can be observed in ancient and modern genomes across America and northern Eurasia. These earliest ancient Siberians (the ANS) are known from a handful of other ancient genomes (those of the Mal'ta and Afontova Gora individuals); they are the descendants of one of the early modern human populations that diversified as Eurasia was first settled by our species, and are thus highly distinct. The ANS were later partially assimilated with a group with East Asian affinity who formed the Ancient Palaeo-Siberians (represented by Kolyma1); this group also probably once had a wide geographical distribution across northern Eurasia. The genetic legacy of Ancient Palaeo-Siberians among present-day Siberians is more limited, being restricted to groups in northeastern Siberia. Importantly, this legacy is also evident in the Americas, which implies that the majority of Native American genetic ancestry is likely to have originated in northeastern Siberia rather than south-central Siberia, as has been inferred from modern mitochondrial and Y chromosome DNA[33]. The Neo-Siberians, who occupy much of the range that was previously inhabited by ANS-related and Ancient Palaeo-Siberian groups, represent a more recent arrival that originated further south. The replacement processes we have revealed for the northeastern portion of Siberia are mirrored in far-western Eurasia by the regional displacement and admixture of the Saami people during the late Holocene. This suggests that similar processes probably took place in many other parts of the northern hemisphere.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1279-z.

Received: 22 February 2018; Accepted: 7 May 2019;
Published online 5 June 2019.

1. Fedorova, S. A. et al. Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol. Biol.* **13**, 127 (2013).
2. Pugach, I. et al. The complex admixture history and recent southern origins of Siberian populations. *Mol. Biol. Evol.* **33**, 1777–1795 (2016).
3. Wong, E. H. M. et al. Reconstructing genetic history of Siberian and Northeastern European populations. *Genome Res.* **27**, 1–14 (2017).
4. Pitulko, V. V. et al. The Yana RHS site: humans in the Arctic before the last glacial maximum. *Science* **303**, 52–56 (2004).
5. Pitulko, V. V., Nikolskiy, P. A., Basilyan, A. & Pavlova, E. Y. in *Paleoamerican Odyssey* (eds Graf, K. E. et al.) Ch. 2, 13–44 (Texas A&M Univ. Press, 2014).
6. Pitulko, V. V. et al. Early human presence in the Arctic: evidence from 45,000-year-old mammoth remains. *Science* **351**, 260–263 (2016).
7. Pitulko, V., Pavlova, E. & Nikolskiy, P. Revising the archaeological record of the Upper Pleistocene Arctic Siberia: human dispersal and adaptations in MIS 3 and 2. *Quat. Sci. Rev.* **165**, 127–148 (2017).
8. Rasmussen, S. O. et al. A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: refining and extending the INTIMATE event stratigraphy. *Quat. Sci. Rev.* **106**, 14–28 (2014).
9. Derevi̇anko, A. P., Powers, W. R. & Shimkin, D. B. *The Paleolithic of Siberia: New Discoveries and Interpretations* (Institute of Archaeology and Ethnography, Siberian Division, Russian Academy of Sciences, 1998).
10. Pitulko, V. V. & Nikolskiy, P. A. The extinction of the woolly mammoth and the archaeological record in Northeastern Asia. *World Archaeol.* **44**, 21–42 (2012).
11. Siska, V. et al. Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Sci. Adv.* **3**, e1601877 (2017).
12. Dulik, M. C. et al. Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. *Proc. Natl Acad. Sci. USA* **109**, 8471–8476 (2012).
13. Poznik, G. D. et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat. Genet.* **48**, 593–599 (2016).
14. Sikora, M. et al. Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* **358**, 659–662 (2017).
15. Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
16. Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
17. Posth, C. et al. Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a Late Glacial population turnover in europe. *Curr. Biol.* **26**, 827–833 (2016).
18. Raghavan, M. et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
19. Lipson, M. & Reich, D. A working model of the deep relationships of diverse modern human genetic lineages outside of Africa. *Mol. Biol. Evol.* **34**, 889–902 (2017).
20. Yang, M. A. et al. 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. *Curr. Biol.* **27**, 3202–3208.e9 (2017).
21. Moreno-Mayar, J. V. et al. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* **553**, 203–207 (2018).
22. Hoffecker, J. F., Elias, S. A., O'Rourke, D. H., Scott, G. R. & Bigelow, N. H. Beringia and the global dispersal of modern humans. *Evol. Anthropol.* **25**, 64–78 (2016).
23. Rasmussen, M. et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
24. Reich, D. et al. Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
25. Loh, P.-R. et al. Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233–1254 (2013).
26. Flegontov, P. et al. Genomic study of the Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Sci. Rep.* **6**, 20768 (2016).

27. Flegontov, P. et al. Paleo-Eskimo genetic legacy across North America. Preprint at https://www.biorxiv.org/content/10.1101/203018v1 (2017).
28. Damgaard, P. de B. et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).
29. Pitulko, V. V., Pavlova, E. Y., Nikolskiy, P. A. & Ivanova, V. V. The oldest art of the Eurasian Arctic: personal ornaments and symbolic objects from Yana RHS, Arctic Siberia. *Antiquity* **86**, 642–659 (2012).
30. Skoglund, P. et al. Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015).
31. Raghavan, M. et al. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884 (2015).
32. Moreno-Mayar, J. V. et al. Early human dispersals within the Americas. *Science* **362**, eaav2621 (2018).
33. Dulik, M. C. et al. Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *Am. J. Hum. Genet.* **90**, 229–246 (2012).
34. Malaspinas, A.-S. et al. A genomic history of Aboriginal Australia. *Nature* **538**, 207–214 (2016).

**Acknowledgements** We thank F. Shidlovskiy (Ice Age Museum, Moscow) for providing access to the Kolyma1 sample. E.W., D.J.M. and M.S. thank St. John's College (Cambridge University) for providing a congenial environment for scientific discussions. This work was supported by the Lundbeck Foundation, the Danish National Research Foundation, KU2016, the Novo Nordisk Foundation and the Wellcome Trust (GeoGenetics). I.D. and V.C.S. were supported by Swiss NSF grants 310030B-166605 and 31003A-143393 to L.E., and V.C.S. was further supported by Portuguese FCT (UID/BIA/00329/2013). V.V.P., E.Y.P. and P.A.N. are supported by Russian Science Foundation project N 16-18-10265-RNF. P.A.N. is supported by the Federal research program no. 0135-2016-0024. V.V.P., E.Y.P. and V.G.C thank the Rock Foundation of New York, USA for long-term support of the Yana research. D.J.M. is supported by the Quest Archaeological Research Program. P.S.G., A.I.L. and B.A.M. are funded by RFBR (19-09-00144). A.Y.F. was supported by the IAET SB RAS project no. 0329-2019-0001. R.M. was supported by an EMBO Long-Term Fellowship (ALTF 133-2017) and R.D. by Wellcome grant WT207492. M. Peyrot is supported by an ERC starting grant ERC-2017-STG 758855. S.R. was supported by the Novo Nordisk Foundation (NNF14CC0001). Q.F. was supported by NSFC (91731303, 41672021, 41630102). V.I.K. and V.M. are funded by RFBR (18-09-00349), A.S. thanks Medicinska understödsföreningen Liv och Hälsa r.f. and the Finnish Society of Sciences and Letters for financial support. M.E.A. was funded by the Villum Foundation (Young Investigator, grant no. 10120). M.M.L. was supported by an ERC advanced grant (no. 295907).

**Reviewer information** *Nature* thanks Pontus Skoglund and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** E.W. initiated and led the study. V.V.P., S.V.V., E.V., M.G., E.Y.P., V.G.C., P.A.N., A.V.G., V.I.K., V.M., P.S.G., A.Y.F., A.I.L., S.B.S., B.A.M., M.M., L.A., J.U.P., T.S., K.M., M. Putkonen, N.B., K.-G.S., K.K., A.W., A.S. and E.W. excavated, curated, sampled and/or described analysed skeletons. M.E.A., L.V., A.M., P.d.B.D, C.d.l.F. and H.M. performed laboratory work. M.S., V.C.S., M.E.A., S.R., G.R., M.A.Y., Q.F., I.D., K.G., D.N.-B., G.K., M. Peyrot, R.M., V.A. and C.P. analysed or assisted in analysis of data. M.S., E.W., V.C.S., L.E., M.E.A., D.J.M. and V.V.P. interpreted results with considerable input from M.M.L. and R.S.M. E.W., L.E., R.N., R.D. and C.R. supervised analysis. M.S., E.W. and D.J.M. wrote the manuscript with considerable input from V.V.P., V.C.S., L.E. and M.M.L., and contributions from all other authors. All authors contributed to the final interpretation of data.

**Competing interests** The authors declare no competing interests.

**Additional information**
**Extended data** is available for this paper at https://doi.org/10.1038/s41586-019-1279-z.
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1279-z.
**Reprints and permissions information** is available at http://www.nature.com/reprints.
**Correspondence and requests for materials** should be addressed to M.S., V.V.P., L.E. or E.W.
**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

**Sample processing and DNA sequencing.** The ancient DNA work was conducted in dedicated ancient DNA clean-room facilities at the Centre for GeoGenetics (Natural History Museum, University of Copenhagen) according to strict ancient DNA standards. DNA was extracted from the samples following established protocols[35,36]. Sequencing libraries were built from the extracts and amplified as previously described[37,38], and sequenced on the Illumina platform. Raw reads were trimmed for Illumina adaptor sequences using AdapterRemoval[39], and mapped to the human reference genome build 37 using BWA[40] with seeding disabled[41]. Final analysis BAM files were obtained by discarding reads with mapping quality $\leq 30$, removing PCR duplicates with MarkDuplicates (http://picard.sourceforge.net) and local realignment using GATK[42] (Supplementary Information 2, 3).

**Authentication, mitochondrial DNA and chromosome Y analyses.** Authentication for ancient DNA was carried out by examining fragment-length distributions and nucleotide-substitution patterns that are characteristic for ancient DNA damage, using mapDamage[43]. Levels of contamination were estimated for all individuals on mitochondrial DNA sequences using schmutzi[44], as well as on chromosome X for male individuals using ANGSD[45]. Mitochondrial DNA sequences were reconstructed using endoCaller from schmutzi[44], and haplogroups assigned with HaploGrep[46]. Y chromosome haplogroups were assigned from reads overlapping single-nucleotide polymorphisms (SNPs) included in the Y-chromosome DNA haplogroup tree from the International Society of Genetic Genealogy (http://www.isogg.org, version 13.37), as previously described[14]. Phylogenetic analysis was carried out on haploid SNP calls from high-coverage individuals obtained with samtools and bcftools[47], using RAxML[48] with the ASC_GTRGAMMA model[13] (Supplementary Information 2, 4).

**Analysis panels.** Autosomal analyses were carried out on three analysis panels of ancient and modern individuals[3,14,16,18,20,21,23,28,30,31,35,49–69] and different sets of SNPs. Panel 1 ('HO 1240K') includes modern individuals from worldwide populations genotyped using the Affymetrix HumanOrigins array[50], merged with ancient individuals with data from shotgun sequencing or genomic capture (the 1240K panel[70]). Panel 2 ('SGDP/CGG 2240K') includes shotgun sequencing data for modern and ancient individuals, as well as selected ancient individuals with genomic capture, all genotyped at SNPs included in the 2240K capture panel[16,61]. Panel 3 ('CGG WGS') includes all genome-wide SNPs genotyped across high-coverage modern and ancient individuals with shotgun sequencing data. Genotyping was carried out separately for each diploid individual using samtools and bcftools[47], and filtered as previously described[14] (Supplementary Information 3). Pseudo-haploid genotypes for low-coverage ancient individuals were obtained by sampling a random high-quality read at each covered SNP position of the respective panels.

**Population structure and admixture modelling.** Population structure was investigated with principal component analysis using smartpca[71]. Principal components were inferred using modern as well as high-coverage ancient individuals, followed by projection of low-coverage individuals using lsqproject. Genetic affinities of ancient and modern individuals were investigated with the *f*-statistic framework[72], using outgroup-$f_3$ statistics for estimation of shared genetic drift[18] as well as $f_4$ statistics for allele-sharing analyses. Standard errors were estimated using a weighted block jackknife with five-megabase block size. Admixture graph modelling was carried out using qpGraph, and outgroup-based estimation of admixture components using qpAdm from the ADMIXTOOLS package[72] (Supplementary Information 6).

**Relatedness and identity-by-descent analyses.** Relatedness among the ancient individuals was quantified using the kinship coefficient estimator implemented in KING[73], obtained from a pairwise identity-by-state matrix inferred with realSFS, implemented in ANGSD[45] (Supplementary Information 5). Genomic segments that were homozygous-by-descent or identical-by-descent were inferred for all high-coverage individuals using IBDseq[74]. Distributions of number and total length of homozygous-by-descent segments for effective population sizes were obtained by simulating 100 haploid individuals from a simple two-population demography[14] using msprime[75].

**Demographic modelling.** The parameters of alternative demographic scenarios were inferred on the basis of the joint site frequency spectrum, by approximating the likelihood of a given model with coalescent simulations using fastsimcoal2[76]. Demographic modelling was carried out on selected ancient individuals from the CGG WGS panel, merged with a set of genomes of present-day individuals from the Simon's Genome Diversity Project[68]. We discarded singleton SNPs for this analysis to minimize the influence of possible sequencing errors in the ancient individuals. Confidence intervals were obtained using a block-bootstrap approach, resampling blocks of one megabase. Parameters in coalescent time were scaled to time in years, assuming a mutation rate of $1.25 \times 10^{-8}$ per generation per site[77] and a generation time of 29 years[78] (Supplementary Information 7).

**Haplotype-sharing analyses.** Haplotype-based analyses of population structure were carried out using chromopainter[79] on all individuals with diploid genotypes in both the HO 1240K and WGS datasets. We used shapeit[80] to reconstruct phased haplotypes for each individual. Chromosome painting was then carried out as previously described[81]. We first estimated the parameters $N_e$ and $\theta$ on a subset of individuals (chosen from diverse modern and ancient groups) and chromosomes (2, 9, 16 and 22) using 10 iterations of the expectation–maximization algorithm, separately for each dataset. Chromosome painting for inferring global population structure related to the ancient individuals was then performed by painting all non-African modern individuals as recipients, using African as well as high-coverage ancient individuals as possible donors. Population structure was investigated by multidimensional scaling on the co-ancestry matrix obtained from chromopainter, both for length and number of shared chunks. For the analysis of Siberian ancestry in present-day Athabascan groups, a second analysis was carried out by painting all Native American groups, using modern Africans and ancient individuals from outside the Americas as potential donors. We quantified differential sharing of pairs of Native American populations A and B with a particular donor group using the symmetry statistic[30]

$$S(A,B) = \frac{\text{Chunk length recipient } A \;-\; \text{Chunk length recipient } B}{\text{Chunk length recipient } A \;+\; \text{Chunk length recipient } B}$$

Standard errors were estimated using a block jackknife, dropping each of the 22 chromosomes in turn.

**Palaeoclimatic modelling.** We used palaeoclimatic modelling to identify regions with the most suitable climatic conditions, in steps of 1,000 years from 48 to 12 ka. We collated a geo-referenced database of modern human fossil and archaeological dated remains, including 936 modern human occurrences across all time intervals. All palaeoclimatic data were gridded to a $1 \times 1$-degree resolution, and all occurrences within a grid cell were aggregated to a single occurrence. Palaeoclimatic conditions were simulated under the HadCM3 (Hadley Centre Coupled Model, version 3) atmospheric–ocean general circulation model, and we selected the three seasonal variables that maximized the climatic signal information: autumn total precipitation, summer average temperature and autumn average temperature. An ensemble of seven different algorithms was used to characterize the climatic niche of modern humans, using the package biomod2. We validated the accuracy of the climatic suitability predictions using cross-validation within each time period. To identify regions with the most suitable climatic conditions across all time periods (from 48 to 12 ka), we estimated the median suitability (and standard deviation) across time intervals for each grid cell (Supplementary Information 8).

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

## Data availability

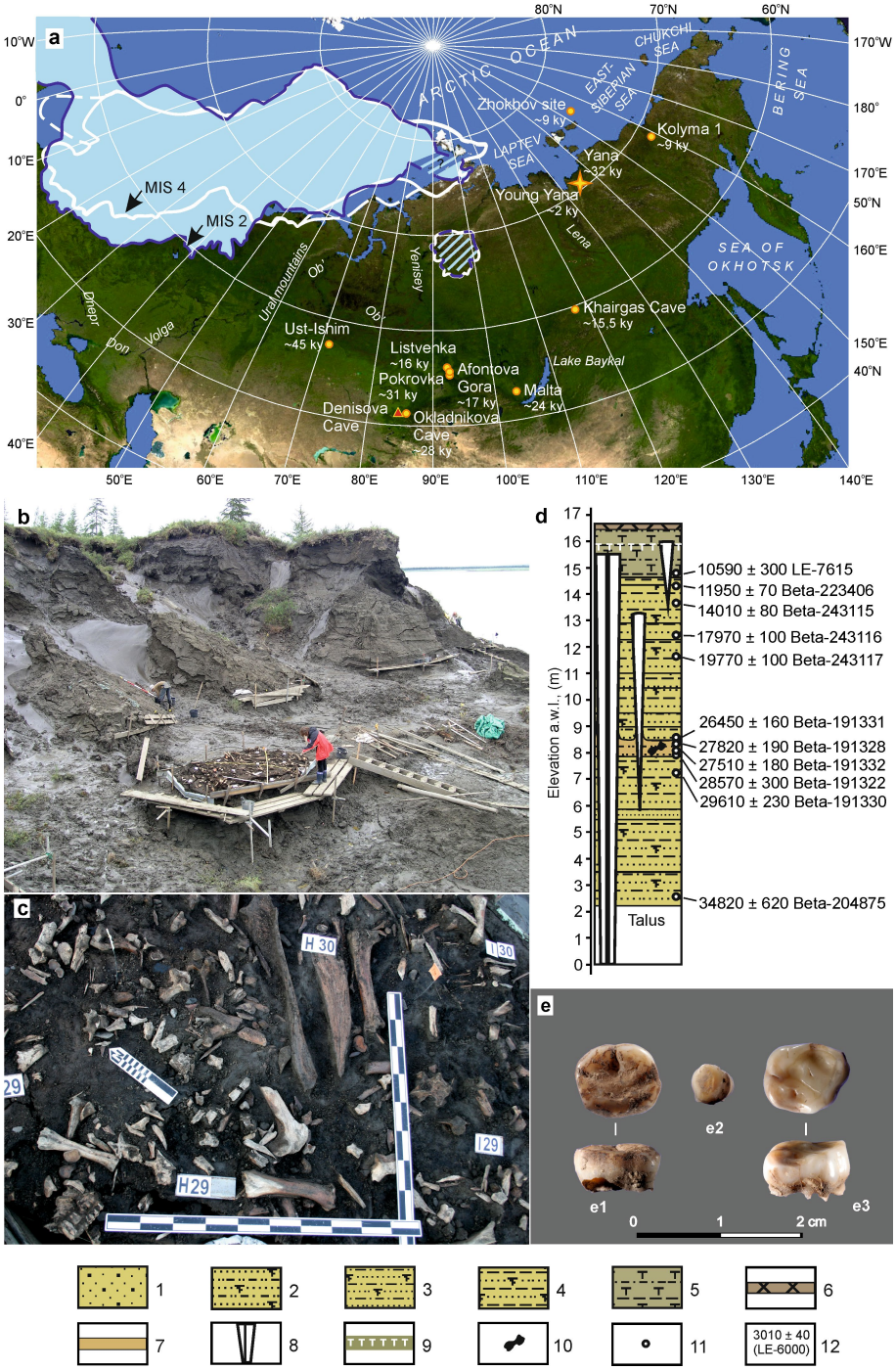
Sequence data have been deposited in the European Nucleotide Archive (ENA) under accessions PRJEB29700 and PRJEB26336.

## Code availability

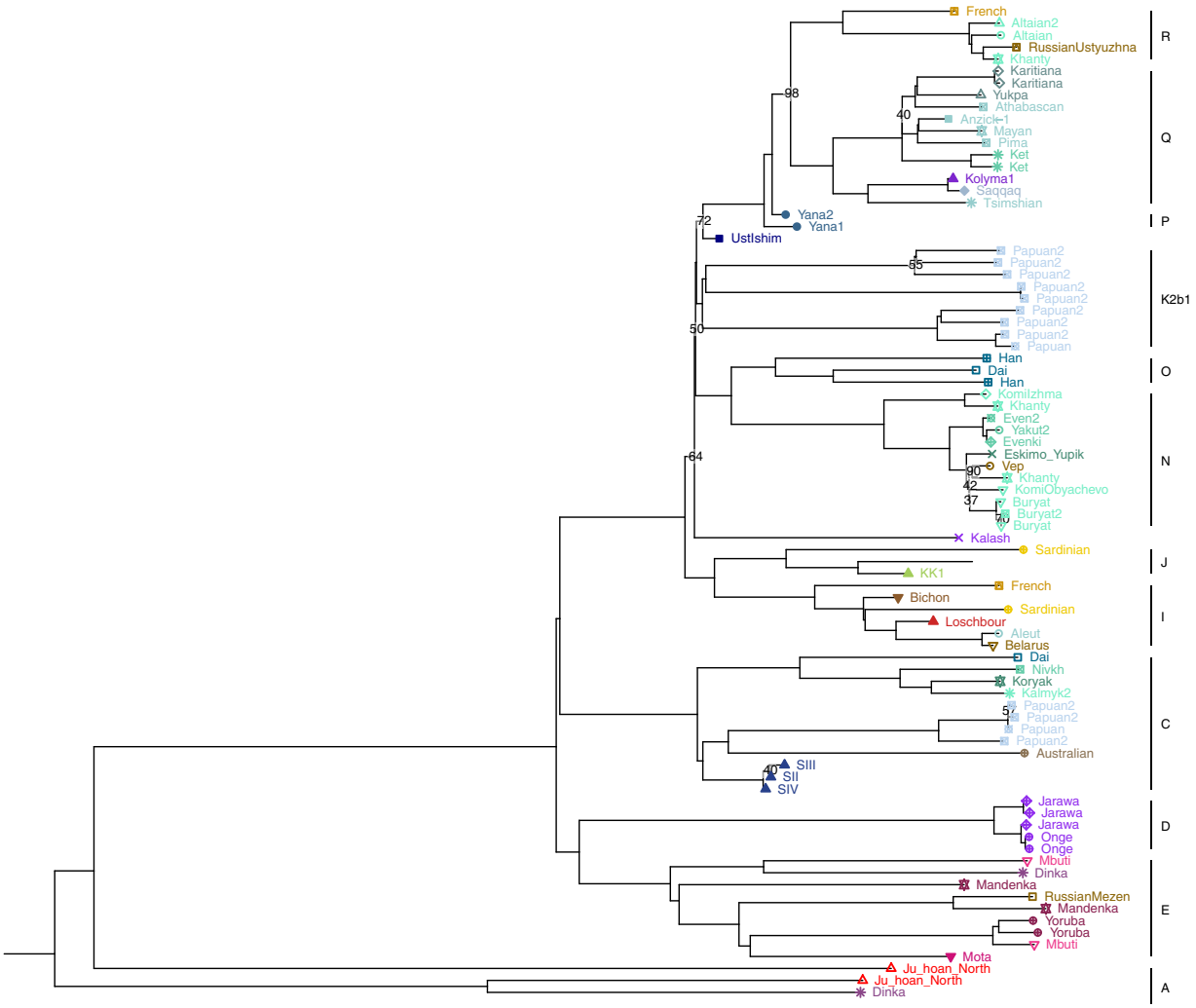
Source code with functions for calculating *f*-statistics is available as an R package at GitHub (https://github.com/martinsikora/admixr)


35. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
36. Damgaard, P. B. et al. Improving access to endogenous DNA in ancient bones and teeth. *Sci. Rep.* **5**, 11184 (2015).
37. Orlando, L. et al. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–78 (2013).
38. Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87–94 (2012).
39. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
40. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
41. Schubert, M. et al. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012).
42. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
43. 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).
44. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
45. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

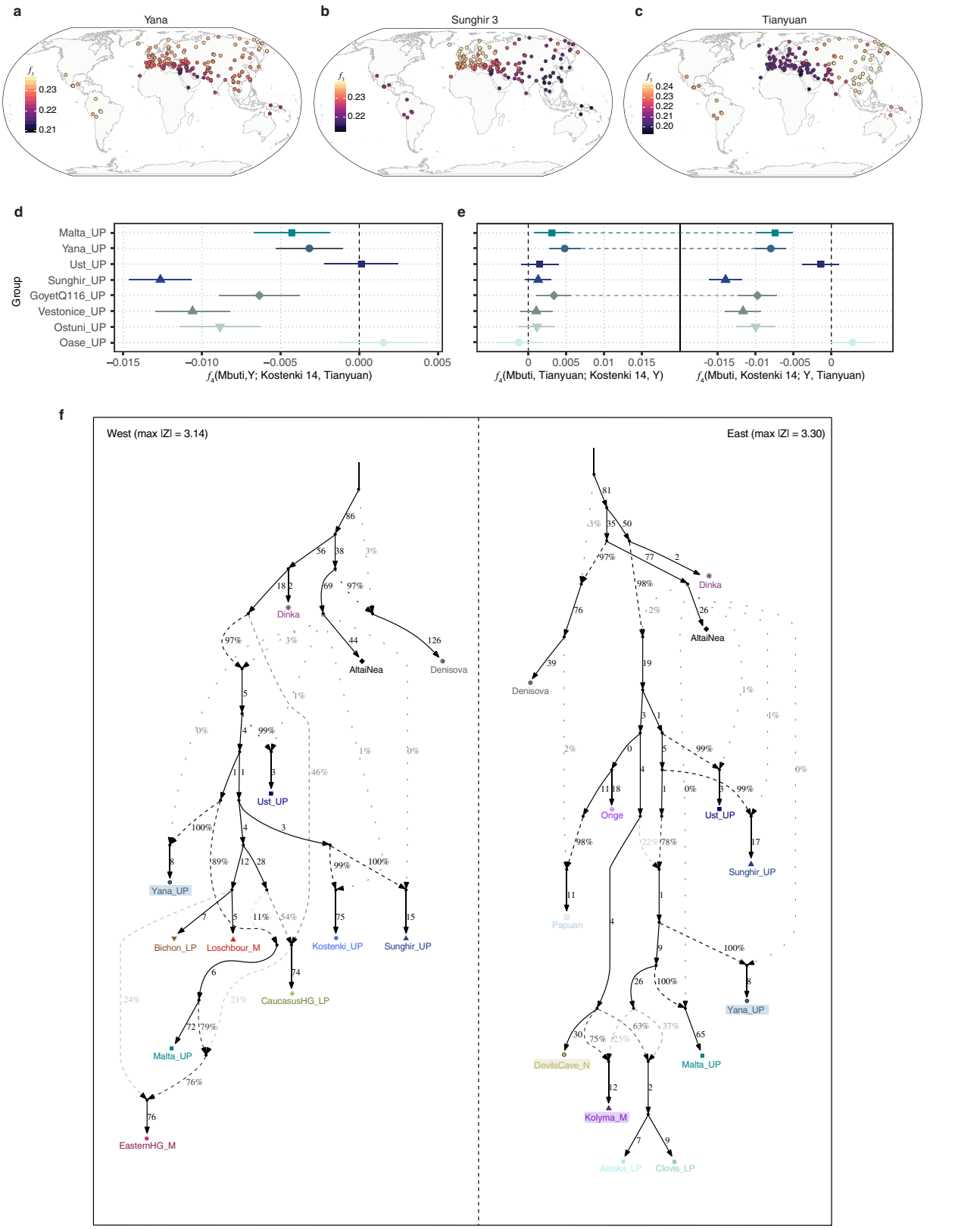
46. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
47. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
48. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
49. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
50. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
51. Rasmussen, M. et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014).
52. Raghavan, M. et al. The genetic prehistory of the New World Arctic. *Science* **345**, 1255832 (2014).
53. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
54. Fu, Q. et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
55. Seguin-Orlando, A. et al. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
56. Olalde, I. et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).
57. Gamba, C. et al. Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
58. Rasmussen, M. et al. The ancestry and affiliations of Kennewick Man. *Nature* **523**, 455–458 (2015).
59. Gallego Llorente, M. et al. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**, 820–822 (2015).
60. Ayub, Q. et al. The Kalash genetic isolate: ancient divergence, drift, and selection. *Am. J. Hum. Genet.* **96**, 775–783 (2015).
61. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
62. Jones, E. R. et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
63. Broushaki, F. et al. Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499–503 (2016).
64. Mondal, M. et al. Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nat. Genet.* **48**, 1066–1070 (2016).
65. Kılınç, G. M. et al. The demographic development of the first farmers in Anatolia. *Curr. Biol.* **26**, 2659–2666 (2016).
66. Jeong, C. et al. Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc. Natl Acad. Sci. USA* **113**, 7485–7490 (2016).
67. Hofmanová, Z. et al. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl Acad. Sci. USA* **113**, 6886–6891 (2016).
68. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
69. Jones, E. R. et al. The Neolithic transition in the Baltic was not driven by admixture with early European farmers. *Curr. Biol.* **27**, 576–582 (2017).
70. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
71. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
72. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
73. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
74. Browning, B. L. & Browning, S. R. Detecting identity by descent and estimating genotype error rates in sequence data. *Am. J. Hum. Genet.* **93**, 840–851 (2013).
75. Kelleher, J., Etheridge, A. M. & McVean, G. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput. Biol.* **12**, e1004842 (2016).
76. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust demographic inference from genomic and SNP data. *PLoS Genet.* **9**, e1003905 (2013).
77. Scally, A. The mutation rate in human evolution and demographic inference. *Curr. Opin. Genet. Dev.* **41**, 36–43 (2016).
78. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423 (2005).
79. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).
80. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
81. Hellenthal, G. et al. A genetic atlas of human admixture history. *Science* **343**, 747–751 (2014).
82. Akimova, E. et al. A new direct radiocarbon AMS date for an Upper Palaeolithic human bone from Siberia. *Archaeometry* **52**, 1122–1130 (2010).
83. Alekseev, V. P. in *The Palaeolithic in Siberia* (ed. Derev'anko, A. P.) 329–33 (Univ. of Illinois Press, 1998).
84. Chikisheva, T. et al. An Upper Paleolithic human mandible and a first cervical vertebra from Afontova Gora II. *Archaeol., Ethnol. Anthropol. Eurasia* **44**, 150–157 (2016).
85. Khaldeyeva, N. et al. An Upper Paleolithic mandible from Listvenka, Siberia: a revision. *Archaeol., Ethnol. Anthropol. Eurasia* **44**, 147–156 (2016).
86. Pitulko, V. V., Ivanova, V. V., Kasparov, A. K. & Pavlova, E. Y. Reconstructing prey selection, hunting strategy and seasonality of the early Holocene frozen site in the Siberian High Arctic: a case study on the Zhokhov site faunal remains, De Long Islands. *Environ. Archaeol.* **20**, 120–157 (2015).
87. Zubova, A. V. & Chikisheva, T. A. The morphology of human teeth from Afontova Gora II, Southern Siberia, and their status relative to the dentition of other Upper Paleolithic Northern Eurasians. *Archaeol., Ethnol. Anthropol. Eurasia* **43**, 135–143 (2015).
88. Zubova, A. V. Chikisheva, T. A. & Shunkov, M. V. The morphology of permanent molars from the Paleolithic layers of Denisova Cave. *Archaeol., Ethnol. Anthropol. Eurasia* **45**, 121–134 (2017).
89. Reich, D. et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
90. Hubberten, H. W. et al. The periglacial climate and environment in northern Eurasia during the Last Glaciation. *Quat. Sci. Rev.* **23**, 1333–1357 (2004).
91. Svendsen, J. I. Late Quaternary ice sheet history of northern Eurasia. *Quat. Sci. Rev.* **23**, 1229–1271 (2004).
92. Pitulko, V. V. et al. Landscape–climatic changes in Yana Palaeolithic site area on western part of Yana–Indighirka lowland during late Pleistocene–Holocene. *Bull. North-East Sci. Center Russ. Acad. Sci. Far East Branch* **1**, 16–29 (2013).



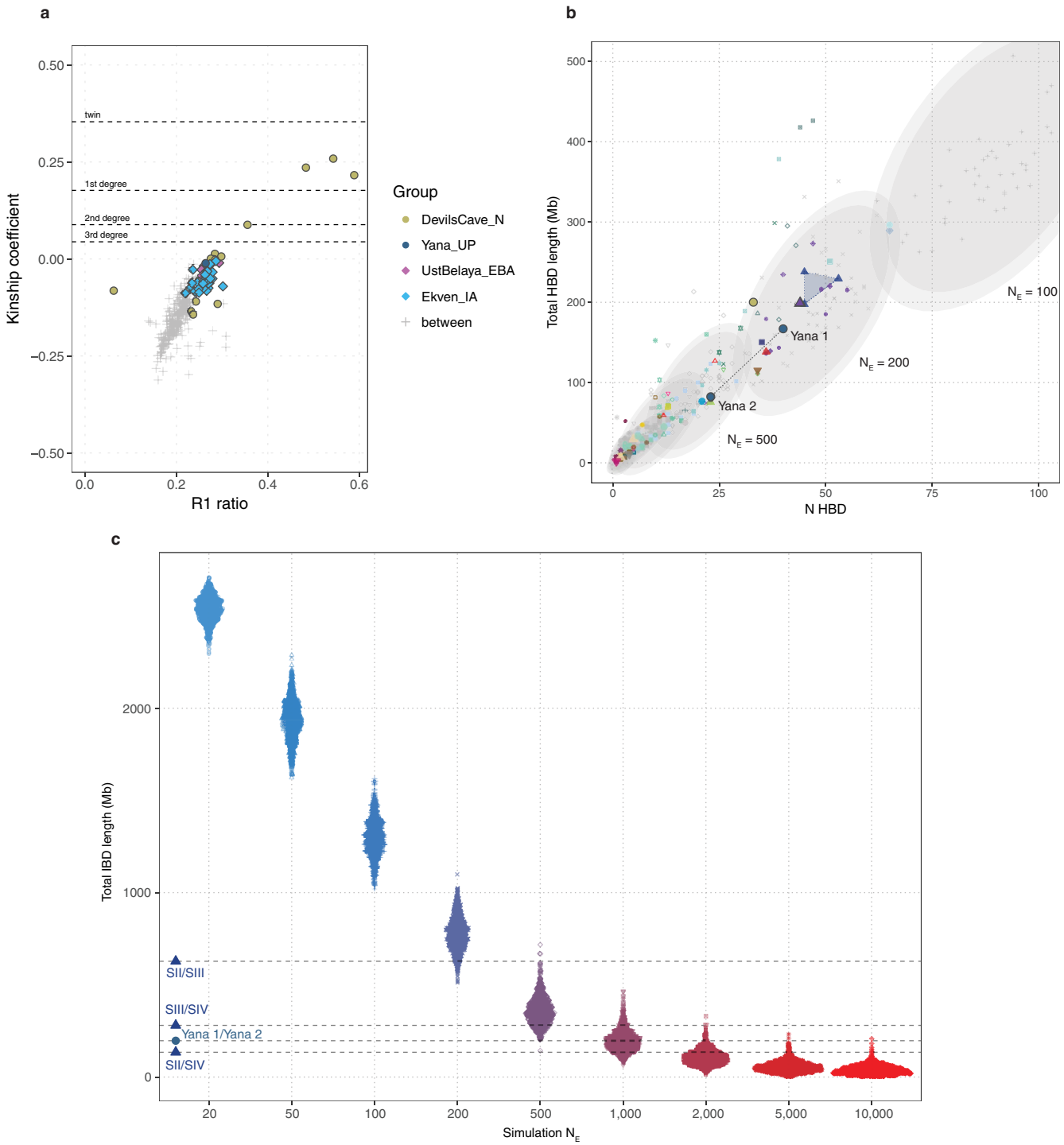

**Extended Data Fig. 1 | Geographical, chronological and archaeological context for the earliest human remains discovered in northern Siberia.** **a**, Map of known $^{14}$C-dated anatomically modern human fossils of late Pleistocene and early Holocene age (yellow dots) found in Siberia[54,82–87]. Yellow star, Yana RHS site; red triangle, Denisova Cave, which has yielded Neanderthal and Denisovan remains[88,89]; white line, reconstructed maximum ice sheet extent at about 60 ka; blue filling, ice sheet extent during the LGM, around 20 ka (refs [90,91]); potentially glaciated areas are cross-hatched. **b**, General view of the Northern Point excavation area at the Yana RHS site[4]. **c**, Cultural layer in H29 unit in which the human tooth was found. **d**, Cryolithological profile for Northern Point of Yana RHS[92]. **e**, Human tooth found during the excavations in unit 2V26 shown in occlusal and lateral view (e1); human tooth found in unit X26 shown in occlusal view (e2); and human tooth found in unit H29 shown in occlusal and lateral view (e3). The samples e2 (Yana2 genome) and e3 (Yana1 with high-coverage (25.6×) genome sequence) are used in this study. In the key for **d** (bottom): 1, sand with small pebbles; 2, sandy silt; 3, clayey-sand silt; 4, sandy-clayey silt; 5, interbedding of clayey silt bands and sandy-clayey silt with beds and lenses of peat; 6, soil–vegetation layer; 7, cultural layer; 8, polygonal ice wedges; 9, boundary of seasonally active layer; 10, location of bones of Pleistocene animals sampled for $^{14}$C dating; 11, location of $^{14}$C samples of plant remains; 12, radiocarbon date and laboratory code.



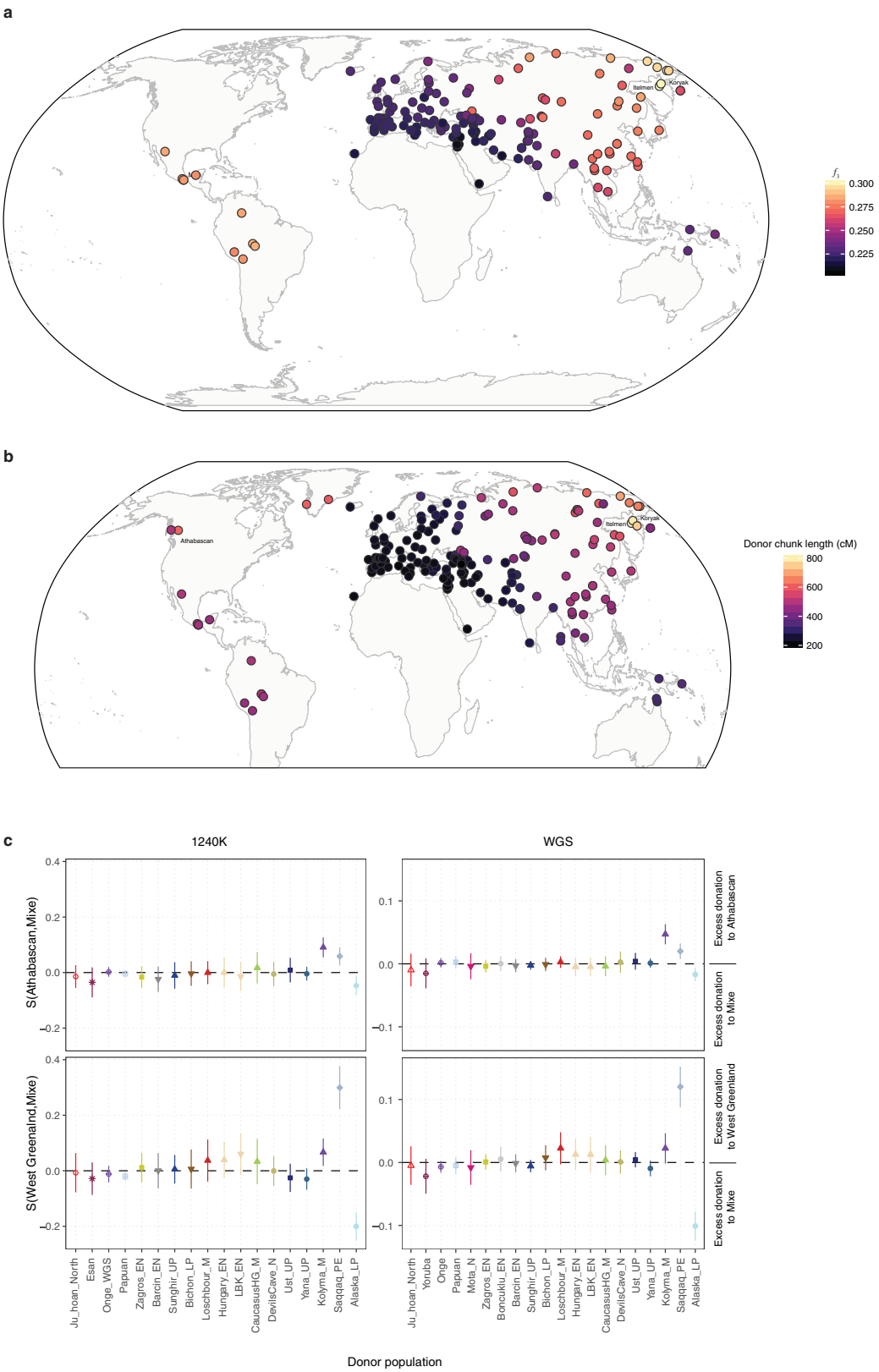

**Extended Data Fig. 2 | Y chromosome phylogeny.** Maximum likelihood tree of Y chromosome sequences for modern and ancient individuals, with major haplogroups highlighted. Numbers on internal nodes show bootstrap support values from 100 replicates for nodes with bootstrap values $< 100$.



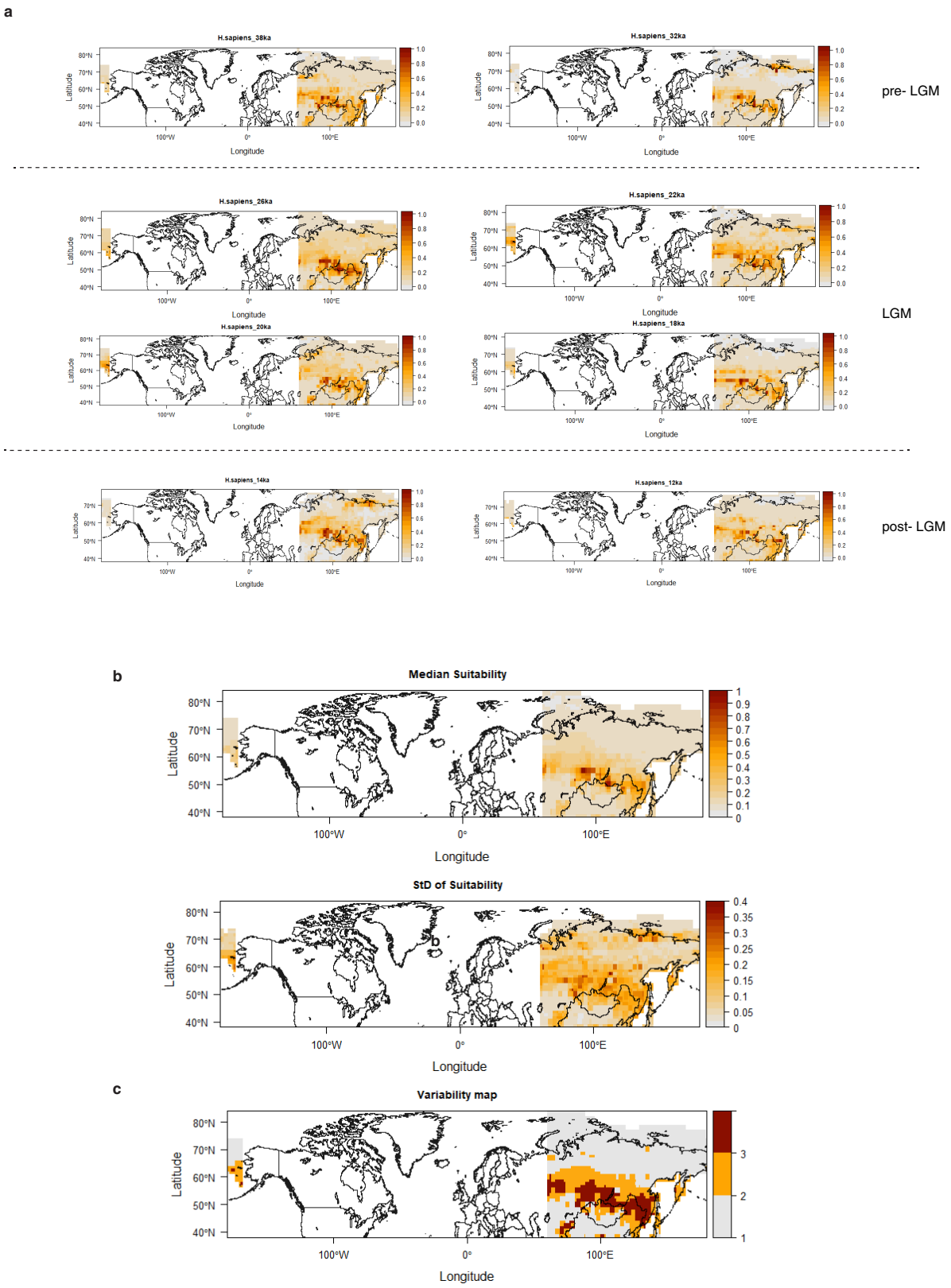

**Extended Data Fig. 3 | Genetic affinities of Yana RHS individuals.** **a**–**c**, Geographical heat maps depicting the outgroup-$f_3$ statistics for Yana1 (**a**), Tianyuan (**b**) and Sunghir3 (**c**) individuals with 167 worldwide populations. **d**, $f_4$-statistics contrasting allele sharing of Yana RHS individuals and other selected Upper Palaeolithic groups with early West Eurasians (represented by the Kostenki 14 individual) or East Asians (represented by the Tianyuan individual). **e**, $f_4$-statistics for highlighting groups with affinities to both early West Eurasians and East Asians (joined with dashed lines). Error bars indicate ± 3 s.e., obtained using a block jackknife (Methods). **f**, Admixture graph models of ancient and modern populations for western Eurasia (left) and East Asia and the Americas (right). Newly reported individuals are highlighted with a coloured background. Early Upper Palaeolithic individuals were modelled allowing for a possible additional Neanderthal contribution to account for higher levels of Neanderthal ancestry (dotted lines).



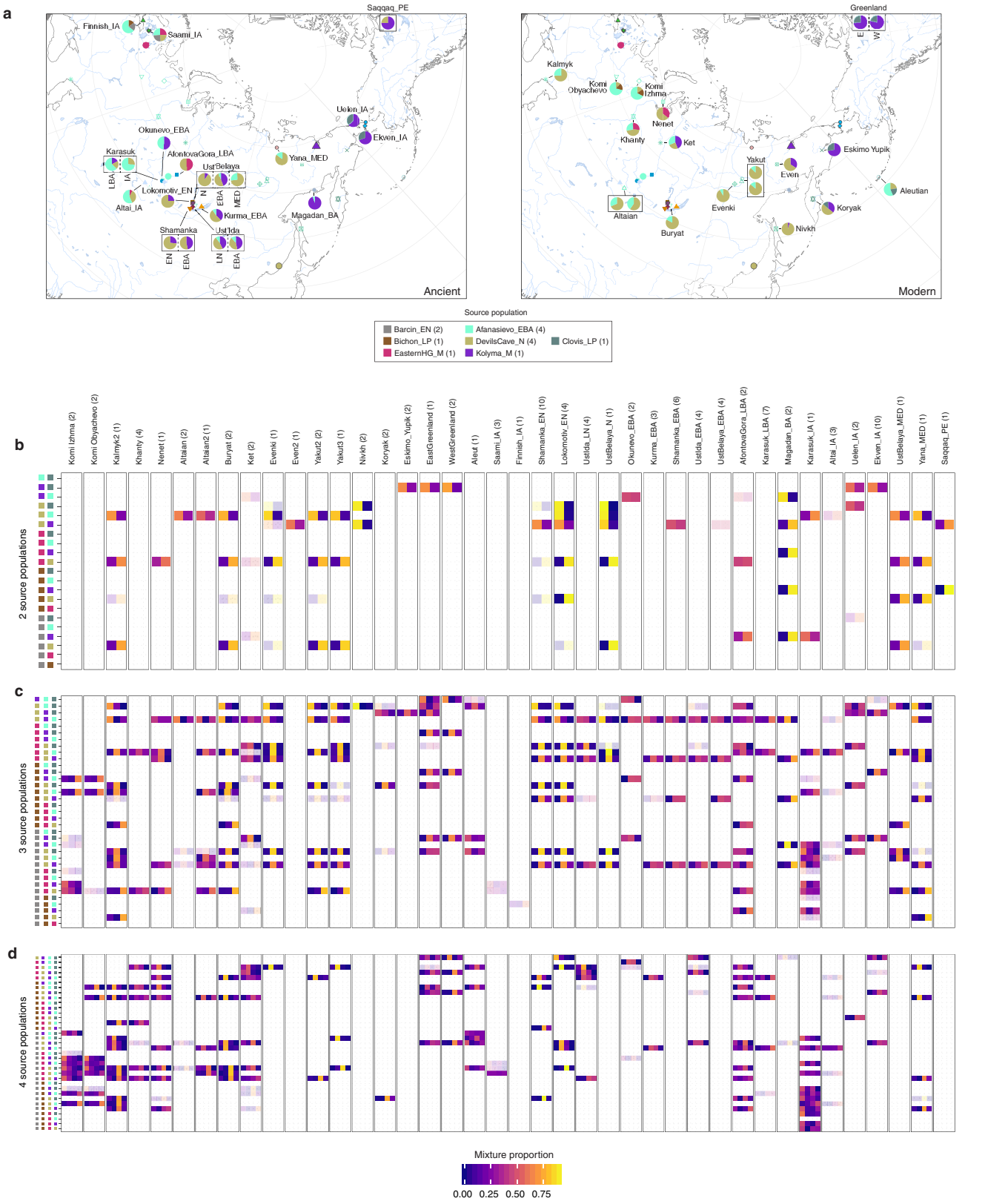

**Extended Data Fig. 4 | Relatedness and identity-by-descent. a**, Kinship coefficient and R1 ratio (number of double-heterozygous (Aa/Aa) sites divided by the total number of discordant genotypes) for newly reported ancient groups with multiple individuals per site. **b**, Number and length of homozygosity-by-descent (HBD) segments in ancient and modern individuals. Grey ellipses indicate 95% confidence region obtained from simulations of 100 haploid genomes of indicated effective population size. **c**, Distribution of total identity-by-descent (IBD) lengths for simulations of varying effective population sizes. Observed values for pairs from the Sunghir and Yana individuals are indicated by dashed lines.



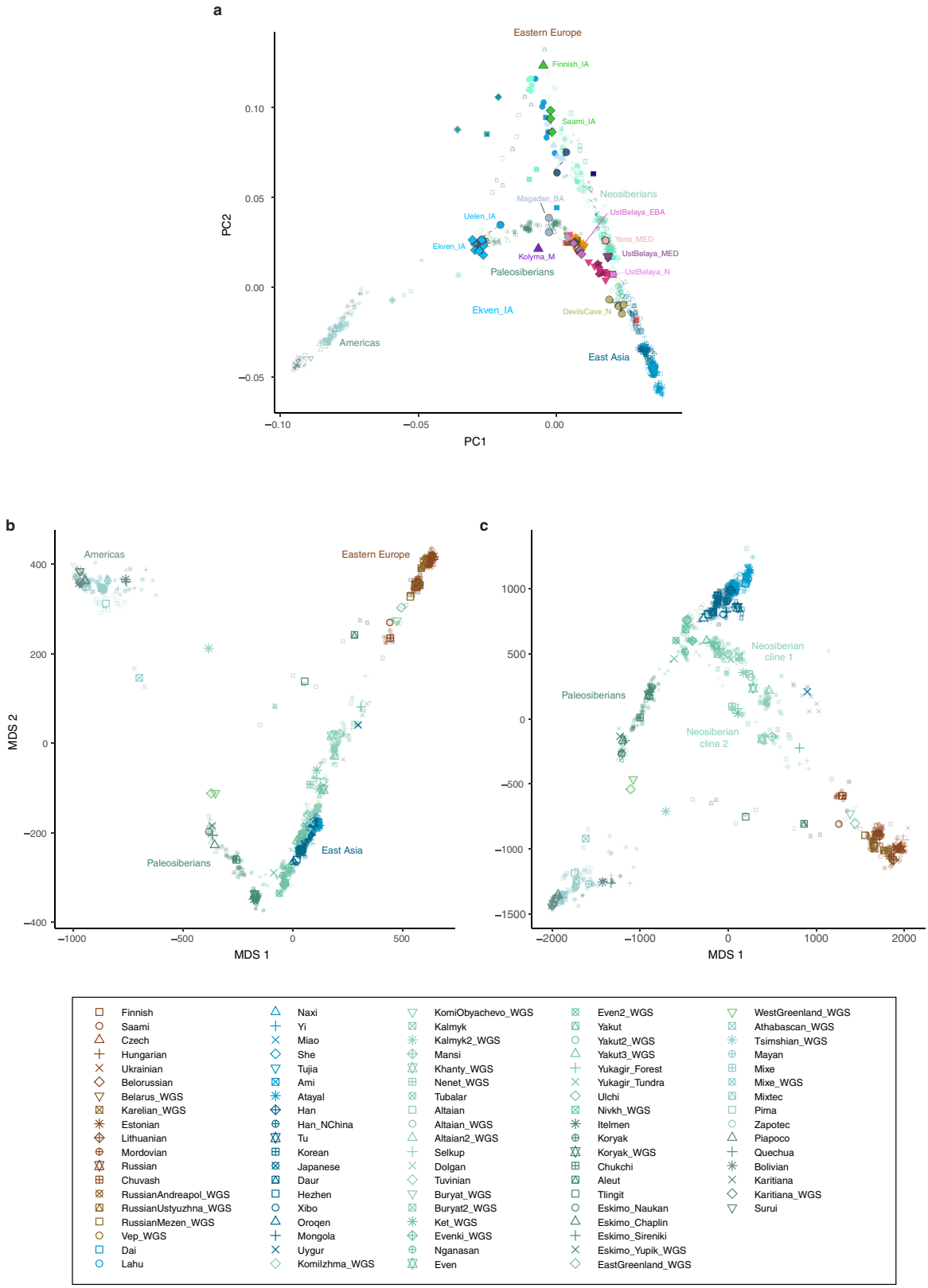

**Extended Data Fig. 5 | Genetic affinities of the Kolyma1 individual.** **a**, **b**, Geographical heat maps depicting genetic affinities of the Kolyma1 individual using outgroup-$f_3$ statistics with 167 modern populations (**a**) and total length of haplotype chunks donated to 206 modern populations in chromosome painting (**b**). **c**, Chromosome-painting symmetry statistic contrasting the total length of haplotypes donated from ancient and modern non-American donor groups to pairs of Native American populations, for two different datasets (1240K and WGS; Supplementary Information 3). The top panels show a greater excess in donations to Athabascans from Kolyma1. The bottom panels show the same statistic for West Greenland Inuit, a population with known affinity to Palaeo-Eskimos, which is reflected in the excess donations observed from the Saqqaq individual. Error bars indicate ± 3 s.e., obtained using a block jackknife.




**Extended Data Fig. 6 | Palaeoclimatic niche modelling.** Maps showing regions that are climatically suitable for human occupation, across temporal and spatial dimensions. Projections are bounded from 60° E to 180° E and from 38° N to 80° N. The colour key represents suitability values; darker colours correspond to higher—and lighter colours to lower—suitability values. **a**, Examples of climatic suitability for human occupation for different time slices. **b**, Median and s.d. of climatic suitability across 23 climatic periods of millennial or bimillennial time resolution. **c**, Regions that are highly climatically suitable for humans are shown in red; regions of low suitability are shown in grey; and regions with periods of both high and low suitability are shown in orange.




**Extended Data Fig. 7 | Admixture modelling using qpAdm. a**, Maps showing locations and ancestry proportions of ancient (left) and modern (right) groups. **b**–**d**, Ancestry proportions and fit for all possible two-way (**b**), three-way (**c**) and four-way (**d**) reference population combinations. Transparent shading indicates model fit, with lighter transparency indicating models accepted with $0.05 > P \geq 0.01$ in qpAdm. Numbers of individuals for source and target populations are given in parentheses.




**Extended Data Fig. 8 | Genetic diversity in northern Eurasia related to ancient genomes. a**, Principal component analysis of 93 ancient individuals projected onto a set of 587 modern Eurasian and Native American individuals. **b**, **c**, Multidimensional scaling (MDS) plots of 715 individuals from 91 modern populations, obtained from the chromosome-painting co-ancestry matrix using modern Africans and high-coverage ancient individuals as donors, based on total length of chunks (**b**) or total number of chunks (**c**).

**Extended Data Table 1 | Key *f*-statistics**

| P1 | P2 | P3 | P4 | $f_4$ | Z | Analysis panel | Note |
|---|---|---|---|---|---|---|---|
| Mbuti | Yana_UP | Kostenki_UP | Tianyuan_UP | -0.0030 | -4.63 | 2240k | Yana shares more ancestry with EWE than with EEA |
| | | Sunghir_UP | | -0.0032 | -5.93 | | |
| | | Vestonice_UP | | -0.0033 | -5.43 | | |
| | | ElMiron_LP | | -0.0033 | -5.16 | | |
| | | Malta_UP | | -0.0046 | -7.82 | | |
| Mbuti | Tianyuan_UP | Kostenki_UP | Yana_UP | 0.0041 | 6.88 | 2240k | Yana has increased EEA ancestry compared to EWE |
| | | Sunghir_UP | | 0.0032 | 6.66 | | |
| | | Vestonice_UP | | 0.0037 | 6.25 | | |
| | | ElMiron_LP | | 0.0029 | 4.90 | | |
| Mbuti | Malta_UP | Kostenki_UP | Yana_UP | 0.0023 | 3.69 | wgs no transitions | Mal'ta shares more ancestry with Yana than with EWE |
| | | Sunghir_UP | | 0.0019 | 3.99 | | |
| | | Vestonice_UP | | 0.0011 | 1.85 | | |
| | | ElMiron_LP | | 0.0017 | 2.66 | | |
| Mbuti | Malta_UP | DevilsCave_N | Kolyma_M | 0.0019 | 9.68 | wgs no transitions | Kolyma1 has increased ANS ancestry compared to East Asians |
| | | Onge | | 0.0020 | 9.71 | | |
| | | Han | | 0.0018 | 9.26 | | |
| Mbuti | Kolyma_M | Yakut | Koryak | 0.0017 | 10.50 | wgs no transitions | Kolyma1 shares more ancestry with Paleosiberians than with Neosiberians |
| | | Evenki | | 0.0009 | 5.40 | | |
| | | Buryat | | 0.0018 | 11.36 | | |
| Mbuti | DevilsCave_N | Kolyma_M | Saqqaq_PE | 0.0008 | 5.17 | wgs no transitions | Increased East Asian affinity of Saqqaq compared to Kolyma1 |
| | Han | | | 0.0006 | 4.13 | | |
| Mbuti | Clovis_LP | Saqqaq_PE | Ekven_IA | 0.0012 | 6.77 | wgs no transitions | Native American gene flow into Neoeskimos |
| | | | Eskimo_Yupik | 0.0011 | 6.11 | | |
| | Kennewick_LP | | Ekven_IA | 0.0015 | 8.80 | | |
| | | | Eskimo_Yupik | 0.0010 | 5.56 | | |
| Mbuti | Ekven_IA | Alaska_LP | Clovis_LP | 0.0012 | 7.74 | wgs no transitions | Source of Native American gene flow in Neoeskimos post-dates divergence of USR1 |
| | | | Kennewick_LP | 0.0010 | 6.41 | | |
| Mbuti | Kolyma_M | Mixe | Athabascan | 0.0009 | 6.49 | wgs no transitions | Kolyma1 is a better proxy for Siberian ancestry in Athabascans |
| | Saqqaq_PE | | | 0.0005 | 3.65 | | |
| | Athabascan | Kolyma_M | Saqqaq_PE | -0.0002 | -1.30 | | |
| Mbuti | Malta_UP | UstBelaya_N | UstBelaya_EBA | 0.0013 | 6.58 | wgs no transitions | Increased ANS ancestry in Bronze Age Lake Baikal |
| | | Shamanka_EN | Shamanka_EBA | 0.0006 | 5.56 | | |
| Mbuti | Magadan_BA | Kolyma_M | Koryak | 0.0015 | 7.00 | wgs no transitions | Magadan Bronze Age shares more ancestry with Koryak than with other Siberians |
| | | Saqqaq_PE | | 0.0011 | 5.67 | | |
| | | Ekven_IA | | 0.0019 | 12.70 | | |
| | | Eskimo_Yupik | | 0.0013 | 7.79 | | |
| Mbuti | LBK_EN | Saami_IA | Saami | 0.0013 | 3.56 | 2240k | West Eurasian gene flow in modern Saami after the Iron Age |
| | RussianUstyuzhna | | Saami | 0.0010 | 3.31 | | |
| | Finnish | | Saami | 0.0007 | 2.49 | | |
| Mbuti | Yakut3 | Finnish_IA | Finnish | 0.0014 | 3.44 | 2240k | Increased Siberian ancestry in modern Finns than in the Iron Age Finn |
| | UstBelaya_N | | Finnish | 0.0013 | 3.06 | | |
| | Saami_IA | | Finnish | 0.0000 | -0.03 | | |

*Z*-scores were obtained using a block jackknife.

nature research

Corresponding author(s): Eske Willerslev
Laurent Excoffier
Martin Sikora

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |
| ☐ | ☒ | Clearly defined error bars<br>*State explicitly what error bars represent (e.g. SD, SE, CI)* |

*Our web collection on statistics for biologists may be useful.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | No software was used |
| Data analysis | AdapterRemoval v2<br>bwa 0.7.10-r789<br>picard-tools 1.127<br>AdmixTools 4.1<br>Eigensoft 6.1.3<br>samtools 1.7<br>bcftools 1.7<br>plink v1.90b3w<br>GATK 3.7<br>mapDamage 2.0.8<br>schmutzi 1.5.1<br>HaploGrep 2<br>RAxML 8.2.11<br>BEDTools 2.33.0<br>snakemake 4.3.0<br>angsd 0.917<br>Arlequin 3.5.2.2 |

fastsimcoal2 2.6.0.3
ibdseq r1206
shapeit v2.r790
Chromopainter v2
admixr 0.1.0
OxCal v4.3
Calib 7.0.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence data were deposited in the European Nucleotide Archive (ENA) under accessions PRJEB29700 and PRJEB26336.

# Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | The number of genomes included in this study was determined by the availability of ancient remains with sufficient preservation for genome sequencing. The sample size is sufficient for the study, as even a single genome provides information from thousands of independent loci. |
| Data exclusions | Related individuals were removed for all population genetic analyses, according to pre-established relatedness cutoffs (1st and 2nd degree relatives) |
| Replication | Replication was not attempted |
| Randomization | No randomization was carried out. Individuals were assigned to groups based on genetic similarity as described |
| Blinding | Blinding was not relevant to this study, Individuals were assigned to groups based on genetic similarity as described |

# Reporting for specific materials, systems and methods

### Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Unique biological materials |
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☐ | ☒ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |

### Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Palaeontology

| | |
|---|---|
| Specimen provenance | Samples from Yana, Kolyma River |

Sample provenance: Institute for the History of Material Culture, Russian Academy of Sciences, 18 Dvortsovaya nab., St. Petersburg 191186, Russia.

Issuing authority: Institute for the History of Material Culture, Russian Academy of Sciences, 18 Dvortsovaya nab., St. Petersburg 191186, Russia.

Permit by: Director V.A. Lapshin

Date: 24.11.2017.

Based on: AGREEMENT FOR JOINT SCIENTIFIC RESEARCH between Institute for the History of Material Culture, Russian Academy of Science (St. Petersburg, Russia) and Centre for Geogenetics of Natural History Museum University of Copenhagen (Copenhagen, Denmark).

Samples from Uelen, Ekven, Ust'Belaya and Devil's Gate cave

Sample provenance: Centre of Physical Anthropology, Institute of Ethnology and Anthropology Russian Academy of Science, Vavilova street, 37a, Moscow, 119991, Russia.

Issuing authority: Centre of Physical Anthropology, Institute of Ethnology and Anthropology Russian Academy of Science, Vavilova street, 37a, Moscow, 119991, Russia.

Permit by: Director of Centre S.V. Vasilyev

Date: 24.11.2016 r.

Based on: AGREEMENT FOR JOINT SCIENTIFIC RESEARCH between Centre of Physical Anthropology, Institute of Ethnology and Anthropology Russian Academy of Science (Moscow, Russia) and Centre for Geogenetics of Natural History Museum University of Copenhagen (Copenhagen, Denmark).

Samples from Magadan

Sample provenance: North-East Interdisciplinary Scientific Research Institute, Far East Branch, Russian Academy of Sciences, 16 Portovaya street, Magadan, 685000, Russia.

Issuing authority: North-East Interdisciplinary Scientific Research Institute, Far East Branch, Russian Academy of Sciences, 16 Portovaya street, Magadan, 685000, Russia.

Permit by: Director V.V. Akinin

Date: 24.11.2016 r.

Based on: AGREEMENT FOR JOINT SCIENTIFIC RESEARCH between North-East interdisciplinary research institute (Magadan, Russia) and Centre for Geogenetics of Natural History Museum University of Copenhagen (Copenhagen, Denmark).

Specimen deposition

All specimens are in deposition at the respective institutions providing access for this study

Dating methods

New datings for this project were made in at the Chrono laboratory, Queens University, Belfast, UK; Aarhus AMS Centre, Aarhus, Denmark; KCCAMS Facility at UC Irvine, Irvine, USA; and the Oxford Radiocarbon Accelerator Unit, Oxford, UK. using their standard protocols. Lab numbers and raw and calibrated dates are available in Supplementary Data Table SD1. Calibrations were made in Oxcal 4.3 using the Intcal13 calibration curve.

☒ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.