The population history of northeastern Siberia since the Pleistocene

Martin Sikora[1,*], Vladimir V. Pitulko[2,*], Vitor C. Sousa[3,4,5,*], Morten E. Allentoft[1,*], Lasse Vinner[1], Simon Rasmussen[6,†], Ashot Margaryan[1], Peter de Barros Damgaard[1], Constanza de la Fuente Castro[1], 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[1,24], Laura Arppe[25], Jukka U. Palo[26,27], Tarja Sundell[28,29], Kristiina Mannermaa[28], Mikko Putkonen[26], Verner Alexandersen[30], Charlotte Primeau[30], Nurbol Baimukhanov[31], Ripan S. Malhi[32,33], Karl-Göran Sjögren[34], Kristian Kristiansen[34], Anna Wessman[28], Antti Sajantila[26], Marta Mirazon Lahr[1,35], Richard Durbin[22,23], Rasmus Nielsen[1,36], David J. Meltzer[1,37], Laurent Excoffier[4,5], Eske Willerslev[1,23,38,39]

1 - Lundbeck Foundation GeoGenetics Centre, University of Copenhagen, Øster Voldgade 5–7, 1350 Copenhagen, Denmark.
2 - Palaeolithic Department, Institute for the History of Material Culture RAS, 18 Dvortsovaya nab., 191186 St. Petersburg, Russia.
3 - Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal.
4 - Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland.
5 - Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland.
6 - Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kemitorvet, Building 208, 2800 Kongens Lyngby, Denmark.
7 - Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, CAS, Beijing 100044, China.
8 - Swiss Integrative Center for Human Health SA, 1700 Fribourg, Switzerland
9 - Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, Copenhagen, Denmark.
10 - Department of Nordic Studies and Linguistics, University of Copenhagen, Denmark.

11 - Leiden University Centre for Linguistics, Leiden University, The Netherlands
12 - Institute of Ethnology and Anthropology RAS (Russian Academy of Science) Lininsky pr. 32, 119991 Moscow, Russia
13 - Russian State University for Humanities (RSUH), Miusskaya pl. 6, 125993 Moscow, Russia
14 - Polar Geography Department, Arctic & Antarctic Research Institute, 38 Bering Str, 199397 St Petersburg, Russia.
15 - St Petersburg Pediatric Medical University, 2 Litovskaya Street, 194100 St. Petersburg, Russia.
16 - Geological Institute, Russian Academy of Sciences, Pyzhevsky per. 7, 119017 Moscow, Russia.
17 - Peter the Great Museum of Anthropology and Ethnography, Russian Academy of Sciences, 3 University emb., St. Petersburg, 199034, Russia
18 - North-East Interdisciplinary Scientific Research Institute, Far East Branch, Russian Academy of Sciences, 16 Portovaya street, Magadan, 685000, Russia.
19 - Northeast State University, 13 Portovaya street, Magadan, 685000, Russia.
20 - Institute of Archaeology and Ethnography of the Siberian Branch of the Russian Academy of Sciences, 17, Acad. Lavrentiev Avenue, Novosibirsk, 630090, Russia.
21 - Institute of Biological Problems of the North, Far East Branch, Russian Academy of Sciences, 18 Portovaya street, 685000 Magadan, Russia
22 - Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK.
23 - Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge CB10 1SA, UK.
24 - University of Greenland, Manuutoq 1, 3905 Nuuk, Greenland
25 - Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland.
26 - Department of Forensic Medicine, University of Helsinki, Helsinki, Finland.
27 - Forensic Genetics Unit, National Institute for Health and Welfare, Helsinki, Finland
28 - Department of Cultures, Archaeology, University of Helsinki, Helsinki, Finland.
29 - Institute of Biotechnology, University of Helsinki, Helsinki, Finland
30 - Laboratory of Biological Anthropology, Department of Forensic Medicine, University of Copenhagen, Frederik V's Vej 11, 2100 Copenhagen, Denmark
31 - Shejire DNA, Almaty, Kazakhstan
32 - Department of Anthropology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

33 - Carle R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
34 - Department of Historical Studies, University of Gothenburg, 40530 Göteborg, Sweden.
35 - Leverhulme Centre for Human Evolutionary Studies, Department of Archaeology, University of Cambridge, Cambridge CB2 1QH, UK.
36 - Department of Integrative Biology, University of California, Berkeley, California 94720, USA.
37 - Department of Anthropology, Southern Methodist University, Dallas, Texas 75275, USA.
38 - GeoGenetics Groups, Department of Zoology, University of Cambridge, Downing St., Cambridge, CB2 3EJ, UK.
39 - The Danish Institute for Advanced Study, The University of Southern Denmark, Odense, Denmark

† Present address: Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen, Denmark.

*These authors contributed equally to this work.

Northeastern Siberia has been inhabited by humans for more than 40,000 years, yet its deep population history remains poorly understood. Here, we investigate the region's late Pleistocene population history through analyses of 34 new ancient genomes from 31,000 to 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", distantly related to early West Eurasian hunter-gatherers; the arrival of East Asian peoples giving rise to Native Americans and "Ancient Paleosiberians", closely related to contemporary communities from far northeastern Siberia such as Koryaks; and a Holocene migration of East Asian peoples, named "Neosiberians", from which many contemporary Siberians descend. Each of these population expansions nearly replaced earlier inhabitants, ultimately generating the mosaic genetic make-up observed in contemporary peoples inhabiting a vast area across northern Eurasia and the Americas.

Northeastern Siberia (the modern Russian Far East) is one of the most remote and extreme environments colonised by humans in the Pleistocene. Extending from the Taimyr Peninsula in the west to the Pacific Ocean in the east, and north from the China/Russia border to the Arctic Ocean, the region is presently 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 last 10,000 years (kya)[1–3]. Yet humans have been in the region far longer[4–6], but their origins and the demographic processes of this deeper population history is largely unknown. The earliest, most secure archaeological evidence for human occupation comes from the artefact-rich, high-latitude (~70° N) Yana RHS site dated to 31,600 years cal BP (Figure 1)[4]. Yana RHS yielded a flake-based stone tool industry and sophisticated bone and ivory artefacts, reminiscent of technologies seen in the Eurasian Upper Palaeolithic (UP) and southern Siberia (Extended Data Fig. 1)[5,7]. By the time of the Last Glacial Maximum (LGM) ~23-19 kya[8], the Yana-related assemblage had disappeared. LGM and later artefact assemblages are dominated by a distinctive microblade stone tool technology, which spread in a time-transgressive manner north and east out of the Amur region[9,10], but did not reach Chukotka or cross the Bering Land Bridge (Beringia) until the end of the Pleistocene, and thus later than the earliest known sites in the Americas. Changes in material culture continued into the late Holocene, but it remains debated whether these successive cultural complexes represent *in situ* technological

evolution or distinct groups of people. In the case of the latter, it is unclear how the groups were related to each other, to contemporary Siberians, or to Native Americans, whose ancestors possibly 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 ranging from 31,600 to 600 years cal BP (Figure 1; Supplementary Information 1-2; Supplementary Data Table 1-2). Our data include samples from ancient individuals that are key for the understanding of Siberian population history: two high-quality genomes (25X and ~7X coverage) sequenced from fragmented milk teeth (Supplementary Information 2) recovered from the Yana RHS site (Yana1 and Yana2, respectively), which are the oldest, northernmost Pleistocene human remains found to date; a high coverage genome (14X coverage) of an individual from the Duvanny Yar site at the Kolyma River (Kolyma1), dated to ~9.8 kya; fourteen genomes from ancient individuals from sites in far eastern Chukotka (Ekven, Uelen) and the northern coast of the Sea of Okhotsk (Ol'skaya, Magadan), ranging from ~3 to 2 kya; six individuals from the ~7.6 kya site of Devil's Gate Cave in Primorskoye, northern East Asia[11]; seven genomes of individuals from northern and southern Siberia (six from Ust'Belaya, in the Lake Baikal region (6.5 kya – 0.6 kya), and the individual "Young Yana" (0.8 kya), a different locality than Yana RHS); as well as four ~1.5 kya individuals from the Levänluhta site in northwestern Eurasia (Finnish, Saami). We analysed these data in the context of large panels of previously published ancient and present-day individuals (Supplementary Information 3; Supplementary Data Table 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 we refer to as "Ancient North Siberians" (ANS). The two Yana RHS individuals were unrelated males, carrying mitochondrial haplogroup U, predominant among ancient West Eurasian hunter-gatherers, and Y chromosome haplogroup P1, ancestral to haplogroups Q and R, which are widespread among present-day Eurasians and Native Americans[12,13] (Extended Data Fig. 2; Supplementary Information 4, 5; Supplementary Data 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 UP Eurasians, such as those from Sunghir[14] and Tianyuan[15], who share overall similar amounts of genetic drift with present-day populations but are geographically more restricted to West Eurasia and East Asia, respectively (Extended Data Fig. 3). Symmetry tests using f_4 statistics reject tree-like clade relationships with both early West Eurasians (Sunghir) and East Asians (Tianyuan); however, Yana is genetically closer to West Eurasians, despite its geographic location in northeastern Siberia (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 Yana can be modelled as early West Eurasian with ~25% 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 (SFS)–based framework indicates an early divergence and mixture of the Yana lineage at ~39 kya (95% CI: 32.2-45.8), receiving ~29% (95% CI: 21.3-40.1) contribution from East Asians, which likely occurred very soon after their divergence from West Eurasians (95% CI: 33.4-48.6 kya) (Figure 2a; Supplementary Information 7). Thus, Yana represents a distinct lineage with affinities to both early West Eurasians and East Asians, documenting the complex population relationships among early Eurasian groups, also supported by the presence of East Asian ancestry and mitochondrial haplogroup M in Western Europe by 35 kya[16,17]. Finally, we estimate ~2% Neanderthal ancestry in Yana, which is contained in longer genomic tracts than in present-day individuals, comparable to other UP Eurasians (Supplementary Information 6)[14,16].

We next investigated how Yana relates to the ancient Siberian population represented by the 24 kya Mal'ta individual from the Lake Baikal region, previously termed "Ancestral North Eurasians" (ANE), from which Native Americans derive ~40% of their ancestry[18]. We find that Mal'ta shares more alleles with Yana than with other west Eurasian hunter-gatherers (e.g. f_4(Mbuti, Mal'ta; Sunghir, Yana) = 0.0019, Z = 3.99; Extended Data Table 1; Supplementary Information 6). Mal'ta and Yana 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, Mal'ta can be successfully fit as a descendant of the ANS lineage, with a minor contribution from an early Eurasian lineage ancestrally related to Caucasus hunter-gatherers (CHG) (Extended Data Fig. 3e-f). The ANE lineage of Mal'ta can thus be considered a descendant of the ANS lineage, and our results therefore suggest that by 31.6 kya ANS-related peoples were likely widespread across northern Eurasia.

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

Ancient Paleosiberians and Native Americans

Following the occupation at Yana RHS, there is an absence of archaeological sites in northeastern Siberia until the latter part of the LGM, when groups bearing a very distinctive stone tool technology appear (~20 kya). It was within that intervening period that the ancestral Native American population emerged[18,21], but to date no genomes from individuals of this age have been recovered in northeastern Siberia. We find that the 9.8 kya Kolyma1 individual, representing a lineage that formed after ~30 kya which we name "Ancient Paleosiberians" (AP), documents the first major genetic shift we observe in the region (Extended Data Fig. 5). Principal component analysis (PCA), outgroup f_3-statistics and mtDNA and Y chromosome haplogroups (G1b and Q1a1a, respectively) demonstrate a close affinity between AP 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 Kolyma1 derives from a mixture of East Asian and ANS-related ancestry similar to that found in Native Americans, although with a greater East Asian contribution in Kolyma1 (75% versus 63%) (Extended Data Fig. 3f; Supplementary Information 6). For both AP and Native Americans, the ANS-related ancestry is more closely related to Mal'ta than Yana (Extended Data Fig. 3f), therefore rejecting a direct contribution of Yana to later AP or Native American groups.

We then estimated demographic parameters of population history models including Kolyma1, Ancient Beringians[21] (Upward Sun River 1 [USR1]), and present-day Native Americans (Karitiana). We find that the ancestors of all three diverged ~30 kya (95% CI: 26.8-36.4) from present-day East Asians (Han), in agreement with previous results[21], with a subsequent divergence of Kolyma1 from the Ancient Beringian / Native American ancestral population at ~24 kya (95% CI: 20.9-27.9) (Figure 2; Supplementary Information 7). Both Kolyma1 and

Native American ancestors received ANS-related gene flow at a similar time (Kolyma1 20.2 kya (95% CI: 15.5-23.7); USR1 19.7 kya (95% CI: 13.3-23.5)). This gene flow amounts to 16.6% (95% CI: 7.5%-22.2%) of ANS ancestry into Kolyma1, and 18.3% (95% CI: 9.8%-20.3%) into USR1, comparable to the estimates obtained using admixture graphs. An alternative model with a single admixture pulse in the ancestral population of Kolyma1 and USR1 showed a comparable likelihood (Supplementary Information 7), but differences in the estimated ANS-related ancestry proportions between Kolyma1 and USR1 favour the two-independent pulses model. Kolyma1 thus represents the closest relative to the ancestral Native American population in northeastern Siberia found to date.

Changes in climatic conditions are commonly put forward as a principal driver of Pleistocene population movement and regional abandonment in Siberia. We used paleoclimatic modelling to infer geographic locations suitable for human occupation from 48 kya to 12 kya to further investigate this hypothesis. When humans were present at Yana RHS, interstadial climatic conditions were suitable for human occupation across a large stretch of the Arctic coast of northeastern Siberia (Extended Data Fig. 8; Supplementary Information 8). Conditions in the region became harsher during the LGM, consistent with the absence of archaeological evidence of occupation of the area at the time. Interestingly, the models suggest the existence of a refugium across southern Beringia during the LGM (e.g. panel 22 kya, Alaska, Extended Data Fig. 8a), in line with previous reports[22]. A possible scenario for gene flow during the formation of the early Native Americans and AP gene pools might therefore have involved early ANS-related groups occupying that region during the LGM, and subsequently admixing with East Asian-related peoples arriving from the South 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[21]). Alternatively, the closer affinity of both Kolyma1 and Native Americans to Mal'ta rather than Yana could suggest a more southwestern location (Lake Baikal region) for the admixture, with a northward expansion following the LGM. While supported by archaeological evidence of a movement south during the LGM, the genetic isolation observed between Asians and ancestral Native Americans after ~23 kya would require the maintenance of a structured population during the LGM, implying distinct refugia for AP and Native American ancestors. 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 of and the origins of peoples involved in more recent gene flow across the (now) Bering Strait during the Holocene. The 4 kya Saqqaq individual from Greenland[23], representing Paleoeskimos, clusters with Kolyma1, but shows greater affinity to East Asians (Figure 1; Extended Data Table 1). Modelling Saqqaq as a mixture of AP (Kolyma1) and East Asians (Devil's Gate Cave), we estimate it harbours around 20% East Asian ancestry (Extended Data Fig. 7a-b; Supplementary Information 6; Supplementary Data Table 5). Individuals from the Uelen and Ekven Neo-Eskimo sites (2.7 – 1.6 kya), located on the Siberian shore of the Bering Sea, cluster closely with contemporary Inuit (Figure 1, Extended Data Fig. 6a). We fit them as a mixture of 69% AP (Kolyma1) and 31% Native American (Clovis) ancestry, thereby documenting a 'reverse' gene flow across the Bering Sea from northwestern North America to northeastern Siberia, in accordance with the linguistic evidence for a back-migration of Eskimo-Aleut (Extended Data Table 1; Extended Data Fig. 7, Supplementary Information 6, 9; Supplementary Data Table 5). The source population of this gene flow post-dates the divergence of USR1 from other Native Americans (~20.9 kya[21]), as the individuals at Ekven share more alleles with ancient Native Americans (Anzick-1, Kennewick) than with ancient Beringians (USR1), confirming previous results from present-day Inuit[24] (Extended Data Table 1). Using linkage-disequilibrium (LD) based admixture dating[25] with Saqqaq and Anzick-1 as source populations, we find significant admixture LD with an estimated date between 100 – 200 generations ago (Supplementary Information 6). While the estimates show considerable uncertainty due to the limited sample size and genomic coverage, they nevertheless indicate a time for gene flow from Native Americans into Siberia well after the disappearance of Beringia, but possibly as early as ~5 kya (~ 100 generations before the earliest individual from Uelen and Ekven). Finally, we investigated the genetic affinity between North American populations speaking Na-Dene languages (Athabascans) and Siberian populations[26], previously suggested to relate to either gene flow from a Paleoeskimo source[27] or an unknown source population more closely related to Koryaks[21]. We find that Kolyma1 is a better proxy for this source population than Saqqaq using both admixture graph modelling (Supplementary Information 6) and chromosome-painting symmetry tests (Extended Data Fig. 5), thereby providing additional evidence against a contribution to Na-Dene from a migration of Paleoeskimos as represented by Saqqaq.

The Holocene archaeological record of northeast Siberia is marked by further changes in material culture. We used a temporal transect of ancient Siberians from ~6 kya to 500 years ago to investigate whether these cultural transitions were associated with genetic changes. We find that in a PCA of present-day non-African populations, most contemporary Siberian populations are arranged along two separate genetic clines. The majority of individuals (referred to as "Neosiberians") lie on an East-West cline stretched out along PC1 between European individuals at one end, and East Asian individuals at the other (Figure 1). A secondary cline between East Asians and Native Americans along PC2 includes Paleosiberian speakers and Inuit populations (Extended Data Fig. 6c). Estimated mixture proportions show that AP ancestry (Kolyma1) was common in other Siberian regions until the early Bronze Age (Extended Data Fig. 7), but thereafter was largely restricted to the northeast, exemplified by a 3 kya individual from Ol'skaya (Magadan) who closely resembles present-day Koryaks and Itelmens. Using present-day Even individuals to represent "Neosiberians" in our demographic model, we find evidence for a recent divergence from East Asians ~13 kya, with only low levels (~6%) of AP gene flow at ~11 kya (Figure 2; Supplementary Information 7). Thus, our data provide evidence for a second major population turnover in northeastern Siberia, with Neosiberians arriving from the south largely replacing AP, a pattern also evident in chromosome painting analyses of present-day populations (Figure 3). A notable exceptions are the Ket, an isolated population that speaks a Yeniseian language, which has previously been described as rich in ANE-ancestry and with genetic links to Paleoeskimos[26]. The Ket fall on a secondary cline parallel to Neosiberians in the chromosome painting analysis and carry ~40% of AP ancestry (Extended Data Fig. 6c; Extended Data Fig. 7). Our findings are consistent with the proposed linguistic link between the Yeniseian speaking Ket and Na-Dene speaking Athabascan populations (Supplementary Information 9) through shared ancestry with an AP metapopulation that was more widespread across Northern Eurasia before the Neosiberian expansion.

Our Holocene transect reveals additional complexity in recent times, with evidence for further episodes of gene flow and local population replacements. A striking example is found in the Lake Baikal region in southern Siberia, where the genomes from Ust'Belaya and neighbouring Neolithic and Bronze Age sites show a succession of three distinct genetic ancestries over a ~6,000 year period. The earliest individuals show predominantly East Asian (Devil's Gate Cave) ancestry (Figure 1; Extended Data Fig. 6, 7), followed by a resurgence of AP ancestry (up to ~50% ancestry fraction) in the early Bronze

Age, as well as influence of West Eurasian Steppe ancestry (Afanasievo; ~10%) (Extended Data Fig. 7; Supplementary Data Table 5). This is consistent with previous reports of gene flow from an unknown ANE-related source into Lake Baikal hunter-gatherers[28]. Our results suggest a southward expansion of AP as a possible source, 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 ~600 year-old individual from Ust'Belaya falls along the Neosiberian cline, similar to the ~760 year-old 'Young Yana' individual from northeastern Siberia, demonstrating the geographic extent of the Neosiberian demographic expansion in the recent past. We show that most populations on the Neosiberian cline can be modelled as predominantly East Asian, with varying proportions of West Eurasian Steppe ancestry, the largest of which is observed among more recent ancient as well as present-day Altaian populations (Extended Data Fig. 7; Supplementary Data Table 5). Together, these findings demonstrate considerable population movement and admixture throughout southern and eastern Siberia during the Holocene, with groups dispersing in multiple directions, yet without clear evidence of the wholesale population replacement seen in earlier Pleistocene times.

Finally, we investigated the geographic extent of these processes of population flux across Northern Eurasia. The striking spatial pattern of Ancestral Paleosiberian and East Asian ancestry in present-day populations (Figure 3) suggests that AP ancestry was once widespread, likely as far west as the Urals. 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 Southern Finland in the late Holocene, ~1.5 kya. This ancestry component is currently limited to the northern fringes of the region, mirroring the pattern observed for AP ancestry in northeastern Siberia. However, while the ancient Saami individuals harbour ancestry from an eastern source, we find that this is better modelled by East Asians rather than AP, suggesting that AP influence likely did not extend across the Urals into Western Eurasia (Extended Data Fig. 7; Supplementary Data 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; in 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/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 migrations 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 ~38 kya, soon after the basal West Eurasian and East Asian split, represented by the archaeological culture found at Yana RHS[4,29]. This finding is consistent with other studies that have shown this was a time of rapid expansion of early modern humans across Eurasia[13]. The arrival of people carrying ancestry from East Asia, and their admixture with descendants of the ANS lineage ~20-18 kya, led to the rise of the AP and Native American lineages. In the archaeological record this is reflected by the spread of microblade technology that accompanies the post-LGM contraction of the once-extensive mammoth steppe[10]. This group was, in turn, largely replaced by Neosiberians in the early and mid-Holocene. Our data suggest that the Neosiberians received ANS-related ancestry indirectly through admixture with AP groups ~11 kya, and possibly later from Bronze Age groups from the central Asian steppe after ~5 kya. Intriguingly, a signal of Australasian ancestry that has been observed in very low frequency in some modern and ancient South Americans[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 last 40,000 years, the first inhabitants of northeastern Siberia, represented by Yana, 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 (ANS), who are known from a handful of other ancient genomes (Mal'ta and Afontova Gora), are the descendants of one of the early modern human populations that diversified as Eurasia was first settled by our species, and thus highly distinct. They were later partially assimilated by a group with East Asian affinity forming "Ancient Paleosiberians" (represented by Kolyma1), who likely also once had a wide geographic distribution across northern Eurasia. Its genetic legacy among present-day Siberians is more limited, restricted to groups in northeastern Siberia. Importantly, this legacy is also evident in the Americas, implying that the majority of Native American genetic ancestry is likely to have originated in northeastern Siberia, rather than south-central Siberia,

as inferred from modern mitochondrial and Y chromosome DNA[33]. The Neosiberians, occupying much of the range previously inhabited by ANS-related and AP 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, suggesting that similar processes likely took place in many other parts of the northern hemisphere.

References

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 Evolutionary Biology* **13**, 127 (2013).
2. Pugach, I. *et al.* The Complex Admixture History and Recent Southern Origins of Siberian Populations. *Mol Biol Evol* msw055 (2016). doi:10.1093/molbev/msw055
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. Human Habitation in Arctic Western Beringia Prior to the LGM. in *Paleoamerican Odyssey* (eds. Graf, K. E., Ketron, C. V. & Waters, M. R.) (Texas A&M University 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. *Quaternary Science Reviews* **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. *Quaternary Science Reviews* **106**, 14–28 (2014).
9. Derevianko, 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 Archaeology* **44**, 21–42 (2012).
11. Siska, V. *et al.* Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Science Advances* **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. *PNAS* **109**, 8471–8476 (2012).

13. Poznik, G. D. *et al.* Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nature Genetics* **48**, 593 (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. *PNAS* **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. *Current Biology* **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. *Current Biology* **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 (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. *Evolutionary Anthropology: Issues, News, and Reviews* **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. *Scientific Reports* **6**, 20768 (2016).
27. Flegontov, P. *et al.* Paleo-Eskimo genetic legacy across North America. *bioRxiv* 203018 (2017). doi:10.1101/203018
28. Damgaard, P. de B. *et al.* The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* eaar7711 (2018). doi:10.1126/science.aar7711

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* eaav2621 (2018). doi:10.1126/science.aav2621
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. *The American Journal of Human Genetics* **90**, 229–246 (2012).
34. Malaspinas, A.-S. *et al.* A genomic history of Aboriginal Australia. *Nature* **538**, 207–214 (2016).

Acknowledgements

We thank Fedor Shidlovskiy, the Ice Age Museum, Moscow, Russia, 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 most congenial environment for scientific discussions. This work was supported by The Lundbeck Foundation, The Danish National Research Foundation, and KU2016 (GeoGenetics). A. Y. F. was funded by the Russian Science Foundation (project No.14-50-00036). 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.P., E.Y.P., and P.A.N. are supported by Russian Science Foundation project N 16-18-10265-RNF. P.N. is supported by the Federal research program #0135-2016-0024. 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. Pe. is supported by an ERC starting grant ERC-2017-STG 758855. S.R. was supported by the Novo Nordisk Foundation (NNF14CC0001).

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.P., 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.C., 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.D., D.N.-B., G.K., M.Pe., R.M., V.A., 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.M. E.W., L.E., R.N., R.D., 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 final interpretation of data.

Author information

Reprints and permissions information is available at www.nature.com/reprints.

The authors declare no competing financial interests.

Correspondence and requests for materials should be addressed to E. W. (ewillerslev@snm.ku.dk), L. E. (laurent.excoffier@iee.unibe.ch) or M.S. (martin.sikora@snm.ku.dk)

Figure legends

Figure 1. Genetic structure of ancient northeast Siberians. **a,** Sampling locations of newly reported and selected previously published individuals (italics). **b,** Sample ages. **c,** PCA 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; EBA – Early Bronze Age; LBA - Late Bronze Age; IA – Iron Age; PE – Paleoeskimo; MED - Medieval

Figure 2. Demographic modelling of Siberian and Native American populations. Inferred parameters for models with: **a,** Ancient and modern Siberian populations, **b,** Siberian and Ancient Beringian. Point parameter estimates are shown in bold and 95% confidence intervals within square brackets. Times of events in kya indicated in the left, and admixture estimates in percentage in the arrows. 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 size and split times were fixed according to recent estimates based on genome-wide SFS[34]. Shaded arrows for the "Siberia and Ancient Beringia" model (**b**) indicate admixture proportions that were fixed to values estimated under model (**a**).

Figure 3. Genetic legacy of ancient Eurasians. **a,** World-wide map of top haplotype donations inferred by chromopainter. Coloured symbols represent a modern recipient population, with the colour and shape indicating the donor population contributing the highest fraction of haplotypes to that recipient population. Geographic locations of donor populations used in this analysis (modern Africans and ancient Eurasians) are indicated by the corresponding larger symbols with black outline added. Extended regions of shared top donors are visualized by spatial interpolation of the respective donor population color. **b,** Major hypothesized migrations into northeast Siberia. Arrows indicate putative migrations giving rise to Ancient North Siberians (left), Ancient Paleosiberians (middle) and Neosiberians (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 (2) admixture in Beringia.

547

Methods

Sample processing and DNA sequencing

The ancient DNA (aDNA) work was conducted in dedicated aDNA clean-room facilities at Centre for GeoGenetics, Natural History Museum, University of Copenhagen according to strict aDNA 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 ≤ 30, removing PCR duplicates with MarkDuplicates (http://picard.sourceforge.net) and local realignment using GATK[42] (Supplementary Information 2 and 3).

Authentication, mitochondrial DNA and chromosome Y analyses

Authentication for ancient DNA was carried out by examining fragment length distributions and nucleotide substitution patterns 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 SNPs included in the Y-DNA haplogroup tree from the International Society of Genetic Genealogy (ISOGG; 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/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[3,23,49,50,18,51–58,30,59,60,35,31,61–63,16,64–69,14,20,21,28] and different sets of SNPs. Panel 1 ("HO 1240K") includes modern individuals from world-wide 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 separately for each diploid individual using samtools/bcftools[47], and filtered as previously described[14] (Supplementary Information section 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 PCA 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 5 megabase (Mb) 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 (IBS) matrix inferred with realSFS implemented in angsd[45] (Supplementary Information 5).
Genomic segments homozygous-by-descent (HBD) and identical-by-descent (IBD) were inferred for all high-coverage individuals using IBDseq[74]. Distributions of number and total length of HBD 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 based on the joint site frequency spectrum (SFS), 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 1 Mb. Parameters in coalescent time were scaled to time in years assuming a mutation rate of 1.25 x 10^{-8} / generation / 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 θ on a subset of individuals (chosen from diverse modern and ancient groups) and chromosomes (2, 9, 16, 22) using 10 iterations of the Expectation-Maximization (E-M) 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 (MDS) on the co-ancestry matrix obtained from chromopainter, both for length and number of shared chunks. For the analysis of the 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]

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

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

Paleoclimate modelling

We used paleoclimatic modelling to identify regions with the most suitable climatic conditions, in steps of 1,000 years from 48 to 12kya. We collated a geo-referenced database of modern human fossil and archaeological dated remains, including 936 modern human

occurrences across all time intervals. All paleoclimatic data were gridded to a 1x1 degree resolution, and all occurrences within a grid cell were aggregated to a single occurrence. Paleoclimatic conditions were simulated under the HadCM3 (Hadley Centre Coupled Model, version 3) Atmospheric– Ocean General Circulation Model (AOGCM), 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 characterise 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 12ka, we estimated the median suitability, and standard deviation, across time intervals for each grid cell (Supplementary Information 8).

Methods references

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 (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**, (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).

694 49. Meyer, M. *et al.* A High-Coverage Genome Sequence from an Archaic Denisovan
695 Individual. *Science* **338**, 222–226 (2012).
696 50. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for
697 present-day Europeans. *Nature* **513**, 409–413 (2014).
698 51. Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial
699 site in western Montana. *Nature* **506**, 225–229 (2014).
700 52. Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345**,
701 1255832 (2014).
702 53. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai
703 Mountains. *Nature* **505**, 43–49 (2014).
704 54. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western
705 Siberia. *Nature* **514**, 445–449 (2014).
706 55. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200
707 years. *Science* **346**, 1113–1118 (2014).
708 56. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-
709 old Mesolithic European. *Nature* **507**, 225–228 (2014).
710 57. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European
711 prehistory. *Nat Commun* **5**, 5257 (2014).
712 58. Rasmussen, M. *et al.* The ancestry and affiliations of Kennewick Man. *Nature* **523**,
713 455–458 (2015).
714 59. Llorente, M. G. *et al.* Ancient Ethiopian genome reveals extensive Eurasian
715 admixture throughout the African continent. *Science* **350**, 820–822 (2015).
716 60. Ayub, Q. *et al.* The Kalash Genetic Isolate: Ancient Divergence, Drift, and Selection.
717 *The American Journal of Human Genetics* **96**, 775–783 (2015).
718 61. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal
719 ancestor. *Nature* **524**, 216–219 (2015).
720 62. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians.
721 *Nat Commun* **6**, 8912 (2015).
722 63. Broushaki, F. *et al.* Early Neolithic genomes from the eastern Fertile Crescent.
723 *Science* **353**, 499–503 (2016).
724 64. Mondal, M. *et al.* Genomic analysis of Andamanese provides insights into ancient
725 human migration into Asia and adaptation. *Nat Genet* **48**, 1066–1070 (2016).
726 65. Kılınç, G. M. *et al.* The Demographic Development of the First Farmers in Anatolia.
727 *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. *PNAS* **113**, 7485–7490 (2016).
67. Hofmanová, Z. *et al.* Early farmers from across Europe directly descended from Neolithic Aegeans. *PNAS* **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. *Current Biology* **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. *Current Opinion in Genetics & Development* **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 Meth* **10**, 5–6 (2013).
81. Hellenthal, G. *et al.* A Genetic Atlas of Human Admixture History. *Science* **343**, 747–751 (2014).

Code availability

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

Data availability

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

Extended data legends

Extended Data Figure 1. Geographical, chronological and archaeological context for the earliest human remains discovered in Northern Siberia. **a,** map of known 14C dated anatomically modern human fossils of late Pleistocene and early Holocene age (yellow dots) found in Siberia (Akimova et al. 2010; Alexeev 1998; Chikisheva et al. 2016; Fu et al. 2014; Khaldeyeva et al. 2016; Pitulko et al. 2015; Zubova and Chikisheva 2015) and Yana RHS finds (yellow star), Denisova Cave that yielded Neanderthal/Denisovan remains, red triangle (Chikisheva, Shunkov 2017; Reich et al. 2010) and the reconstructed maximum ice sheet extent at about 60,000 years ago (white line) and during the Last Glacial Maximum (LGM) around 20,000 years ago (ice-blue filling) (Hubberten et a. 2004; Svendsen et al. 2004); potentially glaciated areas are cross-hatched; **b,** general view of the Northern Point excavation area at the Yana site (Pitulko et al. 2004); **c,** cultural layer in H29 unit where the human tooth was found; **d**, cryolithological profile for Northern Point of Yana RSH (Pitulko et al. 2013); **e,** human teeth found during the excavations in unit 2V26, occlusal and lateral view (e1), unit X26 (e2), occlusal view, and H29 (e3), occlusal and lateral view, samples e2 (Yana 2 genome) and e3 (Yana 1 with high coverage (25.6X) genome sequence) are being used in this study. Legend for (c): 1 – sand with small pebbles; 2 – sandy silt; 3 – claey-sand silt; 4 – sandy-clayey silt; 5 – interbedding of clayey silt bands and sandy-clayey silt with beds and lenses of peat; 6 – soil-vegetable layer; 7 – culture layer; 8 – polygonal ice wedges; 9 – boundary of seasonal active layer; 10 – location of bones of Pleistocene animals sampled for 14C dating; 11 – location of 14C samples of plant remains; 12 – radiocarbon date and lab code.

Extended Data Figure 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 Figure 3. Genetic affinities of Yana. a-c Geographic heat maps depicting outgroup-f_3 statistic for **a,** Yana1, **b,** Tianyuan and **c** Sunghir3 with 167 world-wide populations**. d,** f_4-statistics contrasting allele sharing of Yana and other selected UP groups with early West Eurasians (Kostenki) or East Asians (Tianyuan). **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 standard errors 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 coloured background. Early Upper Palaeolithic individuals were modelled allowing for a possible additional Neanderthal contribution to account for higher level of Neanderthal ancestry (dotted lines).

Extended Data Figure 4. Relatedness and identity-by-descent (IBD). a, Kinship coefficients 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 IBD lengths for simulations of varying effective population sizes. Observed values for pairs from Sunghir and Yana are indicated by dashed lines.

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

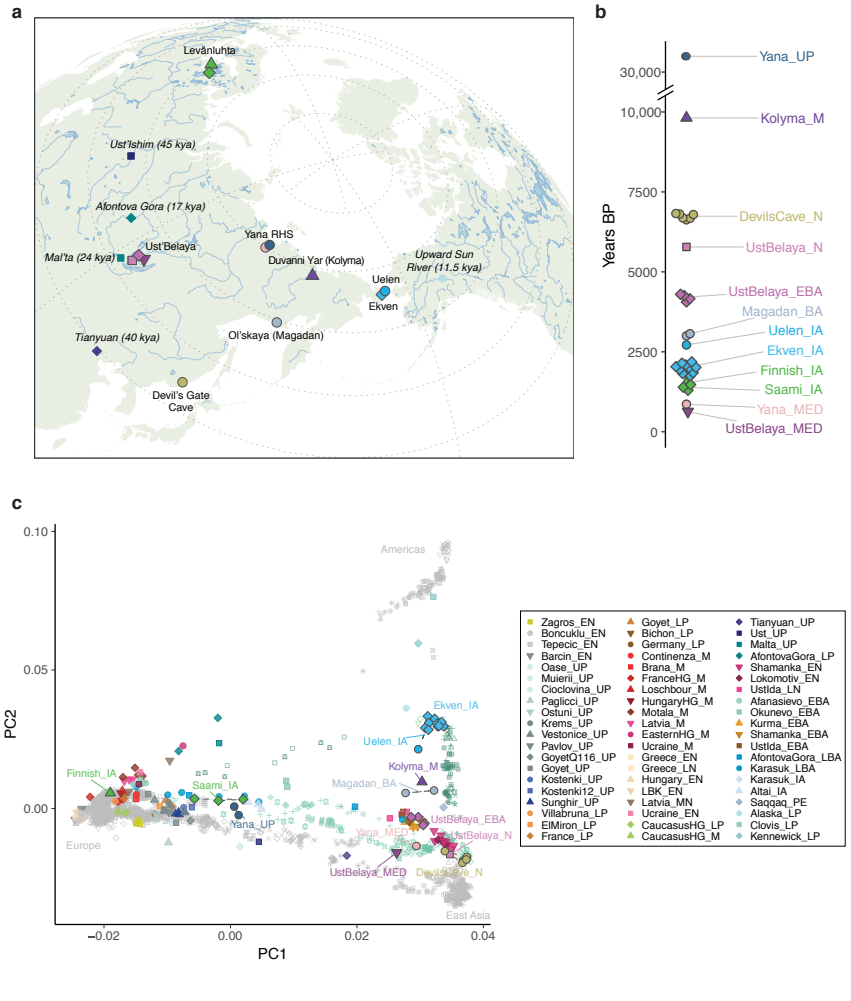
Extended Data Figure 6. Genetic diversity in Northern Eurasia related to ancient genomes. a, PCA of 93 ancient individuals projected onto a set of 587 modern Asian and American individuals. **b, c** MDS plots of 715 individuals from 91 modern populations, obtained from the chromosome painting co-ancestry matrix using modern Africans and high

coverage ancient individuals as donors, based on **(b)** total length of chunks, or **(c)** total number of chunks.

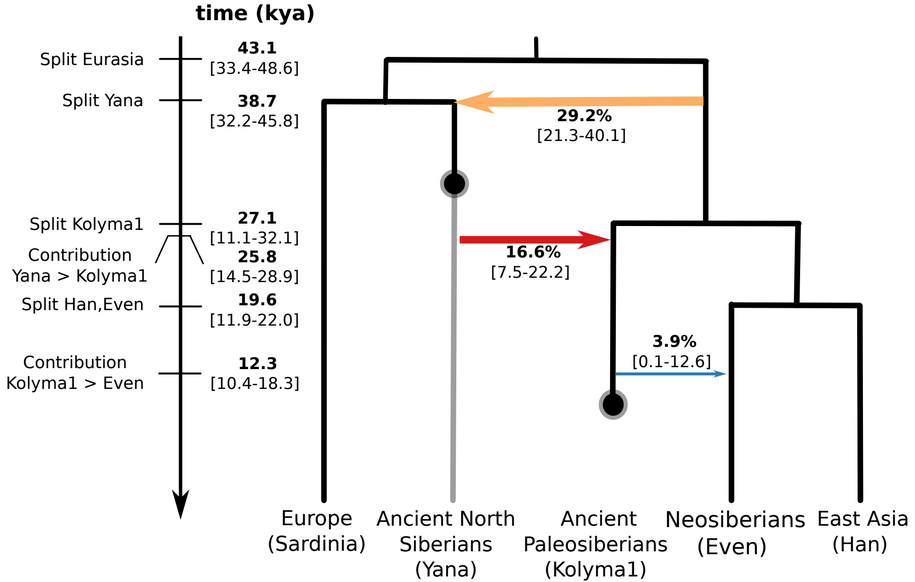
Extended Data Figure 7. Admixture modelling using qpAdm. a, Maps showing locations and ancestry proportions of ancient (left) and modern (right) groups. **b-d,** Ancestry proportions and fit for all possible 2-way (b), 3-way (c) and 4-way (d) reference population combinations. Transparent shading indicates model fit, with lighter transparency indicating models accepted with $0.05 > p \geq 0.01$ in qpAdm. Number of individuals for source and target populations are given in brackets.

Extended Data Figure 8. Paleo-climatic niche modelling. Maps showing climatically suitable regions for human occupation across temporal and spatial dimensions. Projections are bounded between 60 E to 180 E and from 38 N to 80N. Colour-key represents suitability values, with darker (lighter) colours corresponding to higher (lower) suitability values. **a,** Examples of climatic suitability for human occupation for different time slices. **b,** Median and standard deviation of climatic suitability across 23 climatic periods of millennial or bi-millennial time resolution. **c,** Regions highly climatically suitable for humans (red), low (grey), and regions with both periods of high and low suitability (orange)

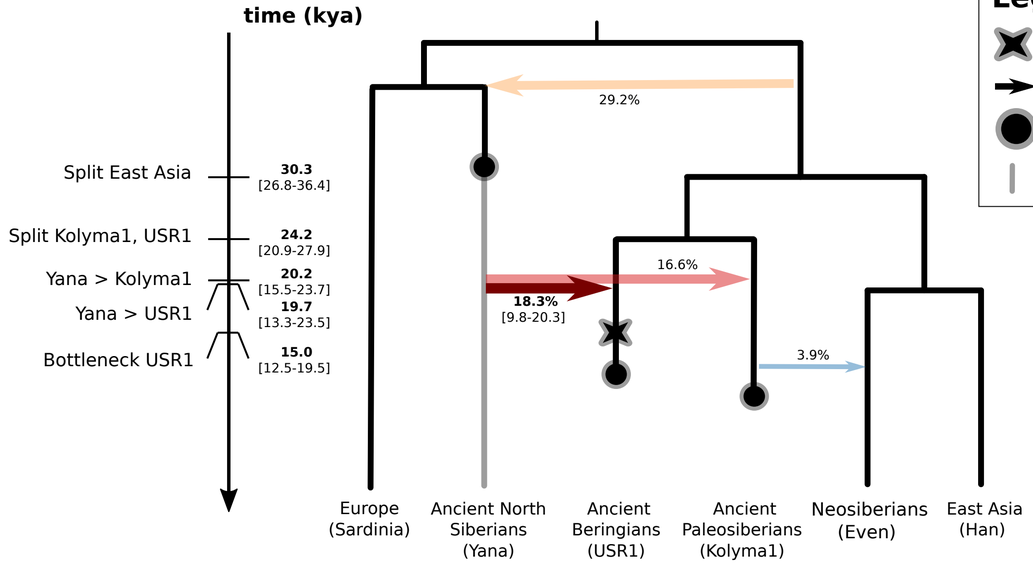
Extended Data Table 1. Key f-statistics. Z-scores were obtained using a block jackknife.

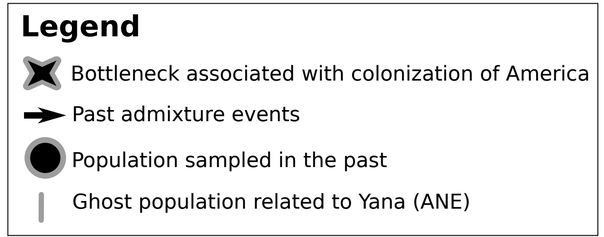

a
Levänluhta
Ust'Ishim (45 kya)
Alontova Gora (17 kya)
Mal'ta (24 kya)
Ust'Belaya
Yana RHS
Duvanni Yar (Kolyma)
Upward Sun River (11.5 kya)
Uelen
Ekven
Ol'skaya (Magadan)
Tianyuan (40 kya)
Devil's Gate Cave
b
Years BP
30,000
10,000
7500
5000
2500
0
Yana_UP
Kolyma_M
DevilsCave_N
UstBelaya_N
UstBelaya_EBA
Magadan_BA
Uelen_IA
Ekven_IA
Finnish_IA
Saami_IA
Yana_MED
UstBelaya_MED
c
PC2
PC1
0.10
0.05
0.00
−0.02
0.00
0.02
0.04
Americas
Europe
East Asia
Ekven_IA
Uelen_IA
Kolyma_M
Magadan_BA
Finnish_IA
Saami_IA
Yana_UP
Yana_MED
UstBelaya_EBA
UstBelaya_N
UstBelaya_MED
DevilsCave_N
Zagros_EN
Boncuklu_EN
Tepecik_EN
Barcin_EN
Oase_UP
Muierii_UP
Cioclovina_UP
Paglicci_UP
Ostuni_UP
Krems_UP
Vestonice_UP
Pavlov_UP
GoyetQ116_UP
Goyet_UP
Kostenki_UP
Kostenki12_UP
Sunghir_UP
Villabruna_LP
ElMiron_LP
France_LP
Goyet_LP
Bichon_LP
Germany_LP
Continenza_M
Brana_M
FranceHG_M
Loschbour_M
HungaryHG_M
Motala_M
Latvia_M
EasternHG_M
Ucraine_M
Greece_EN
Greece_LN
Hungary_EN
LBK_EN
Latvia_MN
Ucraine_EN
CaucasusHG_LP
CaucasusHG_M
Tianyuan_UP
Ust_UP
Malta_UP
AfontovaGora_LP
Shamanka_EN
Lokomotiv_EN
Ustlda_LN
Afanasievo_EBA
Okunevo_EBA
Kurma_EBA
Shamanka_EBA
Ustlda_EBA
AfontovaGora_LBA
Karasuk_LBA
Karasuk_IA
Altai_IA
Saqqaq_PE
Alaska_LP
Clovis_LP
Kennewick_LP

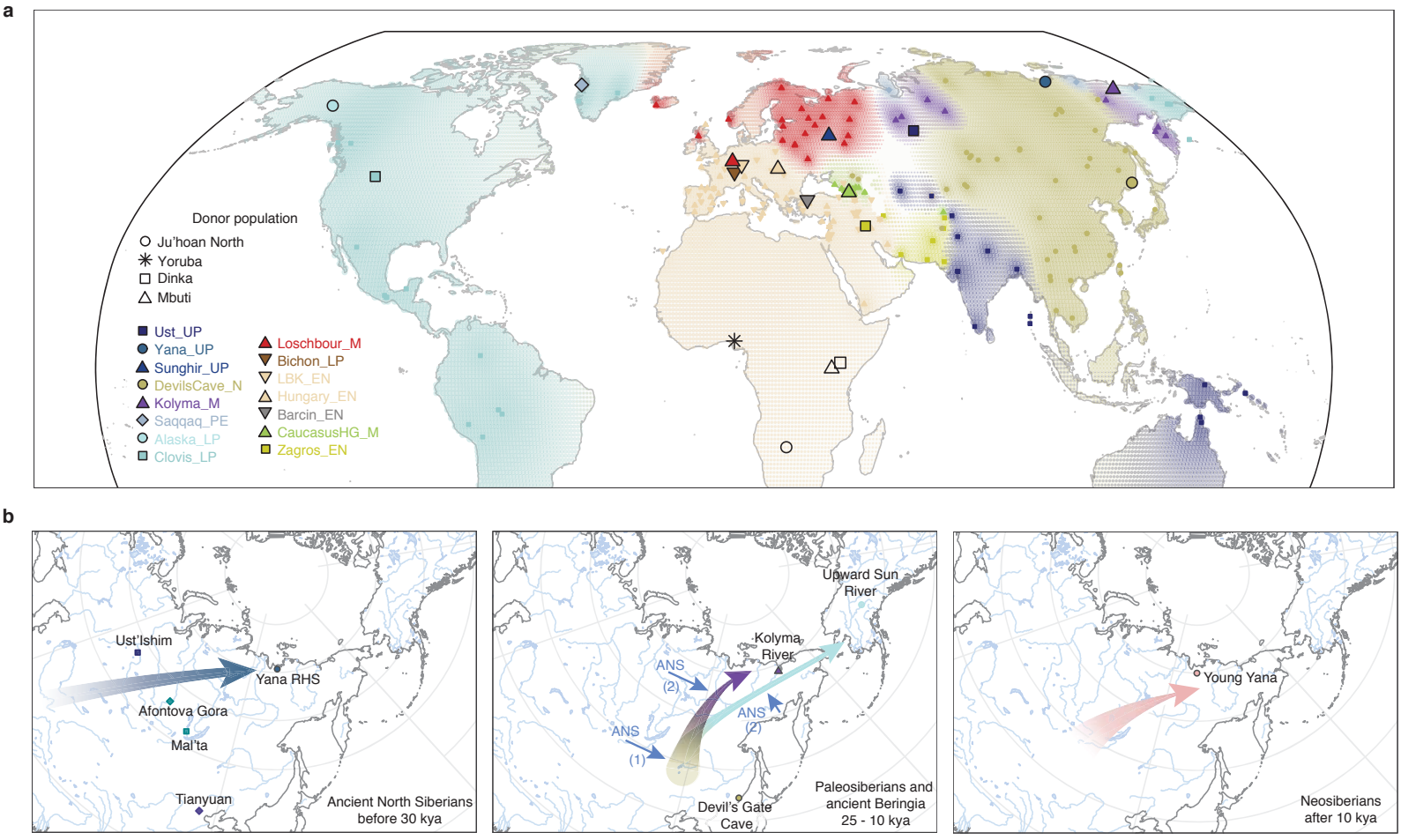
a) Replacement in Siberia



b) Siberians and Ancient Beringia






a
Donor population
Ju'hoan North
Yoruba
Dinka
Mbuti
Ust_UP
Yana_UP
Sunghir_UP
DevilsCave_N
Kolyma_M
Saqqaq_PE
Alaska_LP
Clovis_LP
Loschbour_M
Bichon_LP
LBK_EN
Hungary_EN
Barcin_EN
CaucasusHG_M
Zagros_EN
b
Ust'Ishim
Yana RHS
Afontova Gora
Mal'ta
Tianyuan
Ancient North Siberians
before 30 kya
Upward Sun
River
Kolyma
River
ANS
(2)
ANS
(2)
ANS
(1)
Devil's Gate
Cave
Paleosiberians and
ancient Beringia
25 - 10 kya
Young Yana
Neosiberians
after 10 kya