Supplementary Information

Supplementary Information 1 - The archaeological sites

In addition to the genomic sequence from Yana RHS, ancient DNA sequences were obtained from multiple components of different ages at additional archaeological sites in Siberia and western Beringia. The age, archaeological assemblages and human remains from each are briefly summarized here. We begin with Yana RHS; the remainder follow in alphabetical order. To avoid confusion, the 'Young Yana' sample, which is from the Yana locality but dates to the last millennium, is listed and described separately.

1.1 Yana RHS – Vladimir Pitulko

1.1.1 Site description

Yana RHS (70°43'N, 135°25'E) is located in NE Siberia at more than 70° latitude. It is situated in the lower Yana River valley in the westernmost portion of the extensive coastal lowland that spans the area from the Yana River in the west to the Kolyma River in the east. In terms of paleogeography, this is the westernmost part of western Beringia. Although Yana RHS is denoted as a site, it is in fact a complex of seven geoarchaeological localities, including the Northern Point (NP) locality, where the sampled teeth were found (**Extended Data Fig. 1b**). These localities appear to represent separate but roughly contemporaneous archaeological occupations. All of them were discovered in a cultural layer within the second alluvial terrace (T2) above the Yana River, at an elevation of 16-18 m above the present-day average summer water level (Pitulko et al. 2004, Basilyan et al. 2011, Pitulko et al. 2014a).

The T2 terrace is Pleistocene in age, and the result of both alluvial and an aeolian deposition. Alluvial deposits of T2 accumulated from ~40 kya cal BP onwards, but ~13 kya cal BP the depositional process becomes aeolian. At the same time, alluvial deposits of the first alluvial terrace (T1) start accumulating. At the top of the T2 sequence younger erosional cuts filled with terminal Pleistocene and Holocene deposits were observed, but none of them reaches the cultural layer within T2.

The cultural layer occurs ~7.5 m above the average summer water level and is overlain by 8-11 m of frozen sediments (**Extended Data Fig. 1c, d**). The cultural layer is characterized by excellent preservation of organic material due to permafrost conditions. A polygonal network of syngenetic ice wedges >2 m in diameter results in polygons of the deposits containing the cultural layer of 5-7m width (**Extended Data Fig. 1c, d**). The sedimentary unit containing the cultural material is 12-15 cm thick and of bright reddish-brown colour. It includes

abundant remains of Pleistocene animals, ivory and bone artefacts, as well as lithic tools and debitage (Pitulko et al. 2014a).

The human teeth that are the subject of the genetic analysis reported here are from the NP locality at Yana RHS, a large (>2000 m^2) area systematically excavated, and which yielded more than 80,000 plotted objects. That includes three human teeth (two of which provided the genetic data discussed here), which were recovered from squares 2V26, X26 and H29 (all 1 *x* 1 m excavation units) in the western part of the excavation (**Extended Data Fig. 1e; Fig. S1.1**). The recovery of the teeth was the result of the methods required in excavating in permafrost conditions (Pitulko 2008, 2015): namely, excavation of the cultural layer, with low-pressure water screening (2 mm mesh) of the excavated sediments in order to recover the smallest fraction of finds (including lithic artefacts and bone pieces). The human teeth studied here were collected from the 2 mm mesh.

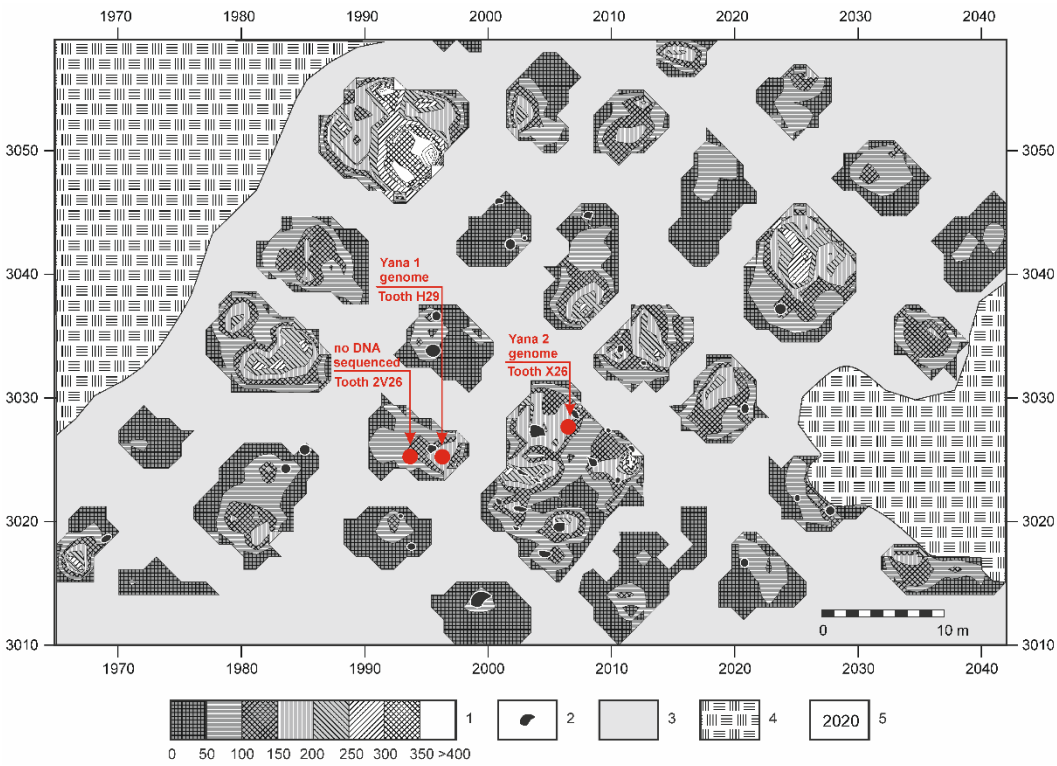


Figure S1.1 Yana RHS, Northern Point locality excavation map, with locations of human tooth finds indicated (tooth X26 and H29 produced genomes which are being used in this study while sample 2V26 failed to do so when analyzed in previous work by Lee et al. 2016). Densities of plotted finds per 1 sq. m. Legend: 1 – graded density of finds per 1 sq. m; 2 – hearth; 3 – polygonal ice-wedges, where finds are not present; 4 – unexcavated area; 5 – excavation grid numbers

The *in situ* cultural layer at the NP locality was securely sealed by the overlying, frozen deposits. Its age is constrained by 28 radiocarbon ages that support an age range for the layer of between ~28.9 and 26.9 ^{14}C kya BP (between ~33 and ~31 kya cal BP) (**Table S1.1**). The dates nearest to the studied human teeth are located in squares F31 and D27, and have produced ages of 28,060 ±180 radiocarbon years BP (Beta-271412, on bone with an embedded fragment of ivory artefact), and 27,850 ±150 radiocarbon years BP (Beta-230442, on plant remains). These two radiocarbon ages statistically overlap (by chi-square test), and average to 27,940 ± 115 radiocarbon years BP, which when calibrated (IntCal13) yields a 1 sigma age range of 31,450-31,774 cal BP, with a median probability of 31,630 cal BP.

Table S1.1. Radiocarbon ages from the Northern Point (NP) locality, Yana RHS

Radiocarbon age	Lab. No.	Material dated	Reference
1. Radiocarbon ages on sediments above the cultural layer			
10,590 ± 300	LE-7615	peat	Pitulko and Pavlova 2010
11,950 ± 70	Beta-223406	plant remains	Pitulko and Pavlova 2010
14,010 ± 80	Beta-243115	plant remains	Pitulko and Pavlova 2010
17,970 ± 100	Beta-243116	plant remains	Pitulko and Pavlova 2010
19,770 ± 100	Beta-243117	plant remains	Pitulko and Pavlova 2010
22,290 ± 150	Beta-204858	plant remains	Pitulko and Pavlova 2010
26,450 ± 160	Beta-191331	plant remains	Pitulko et al. 2007
2. Cultural layer radiocarbon ages			
26,680 ±160	Beta-191334	burnt bone fragment	Pitulko and Pavlova 2010
27,140 ± 180	Beta-191321	bone collagen from musk-ox metacarpal bone	Pitulko et al. 2007
27,200 ± 2400	LE-7668	charred material from the hearth	Pitulko and Pavlova 2010
27,250 ± 230	Beta-223413	charcoal	Pitulko and Pavlova 2010
27,440 ± 210	Beta-162233	collagen from woolly rhinoceros horn foreshaft*	Pitulko et al. 2004

27,510 ± 180	Beta-191332	plant remains	Pitulko and Pavlova 2010
27,620 ± 240	Beta-204863	bone collagen	Pitulko and Pavlova 2010
27,820 ± 190	Beta-191328	plant remains	Pitulko et al. 2007
27,850 ± 150	Beta-230442	plant remains	*This report*
27,890 ± 190	Beta-191335	plant remains	Pitulko and Pavlova 2010
27,900 ± 200	Beta-191333	charred material from the hearth	Pitulko and Pavlova 2010
28,000 ± 190	Beta-191329	plant remains	Pitulko et al. 2007
28,060 ± 180	Beta-271412	bone with embedded ivory fragment	*This report*
28,090 ± 200	Beta-191327	bone collagen from Pleistocene bison phalange I	Pitulko and Pavlova 2010
28,250 ± 170	Beta-173064	mammoth ivory artifact/foreshaft*	Pitulko and Pavlova 2010
28,500 ± 200	Beta-191326	Pleistocene bison hoof phalange	Pitulko and Pavlova 2010
28,570 ± 300	Beta-191322	Pleistocene hare humerus	Pitulko et al. 2007
29,130 ± 410	Beta-204864	soot	Pitulko and Pavlova 2010
3. Radiocarbon ages on sediments below the cultural layer			
29,610 ± 230	Beta-191330	plant remains	Pitulko et al. 2007
33,220 ± 520	Beta-204873	plant organic material	Pitulko and Pavlova 2010
34,820 ± 620	Beta-204875	plant organic material	Pitulko and Pavlova 2010

Note: All Beta- (Beta Analytic) dates are AMS ages; the LE date (Institute for the History of Material Culture, Russian Academy of Sciences, St. Petersburg) is a standard radiometric (conventional ^{14}C) age. The two dates marked with asterisks are on artefacts collected next to the bank exposure of the cultural layer. Other NP dates are from *in situ* position.

1.1.2 Archaeological material

As the artifact assemblage from Yana RHS has been described in detail elsewhere (e.g. Pitulko et al. 2004, Pitulko et al. 2014a, 2014b, Pitulko et al. 2015a, 2015b Pitulko et al. 2017), here we provide only a brief summary. The lithic industry (**Fig. S1.2)** is flake-based

and was crafted with a simple knapping technology. The assemblage is dominated by large scrapers and chopping / cutting tools, but there are also a number of specific micro-tools (many of which are backed pointed implements made of quartz crystal and regular local chert).

There were ~2,500 bone and ivory artefacts recovered during excavations between 2002 and 2016 (**Fig. S1.3)**. These include (i) hunting equipment like foreshafts and points, (ii) 'domestic' tools such as bone and ivory punches, bone needles, needle cases, and awls, (iii) personal adornments and decorations, and (iv) symbolic objects (Pitulko et al. 2004, 2012, 2014a, 2015). The ivory objects also include decorated ivory bowls (Pitulko et al. 2012). Pitulko & Pavlova (2015) have interpreted these artefacts as reflecting complex social behaviour and organization. Further, some exotic materials like amber were used to manufacture various pendants. The non-local origin of these raw materials suggests high mobility and/or extensive trade/exchange networks (Pitulko et al. 2012).

Personal ornaments are abundant and include the following categories: small beads or sewn-on adornments, teeth, soft stone, and ivory pendants, as well as diverse flat decorated ivory pieces commonly known as hair-bands or 'diadems' (Pitulko et al. 2012, 2014a, Pitulko & Pavlova 2015). With about 6,000 finished and more than 700 half-products, pre-forms, and incomplete items, beads are the most numerous category of personal ornaments (Pitulko et al. 2014a, 2014b).

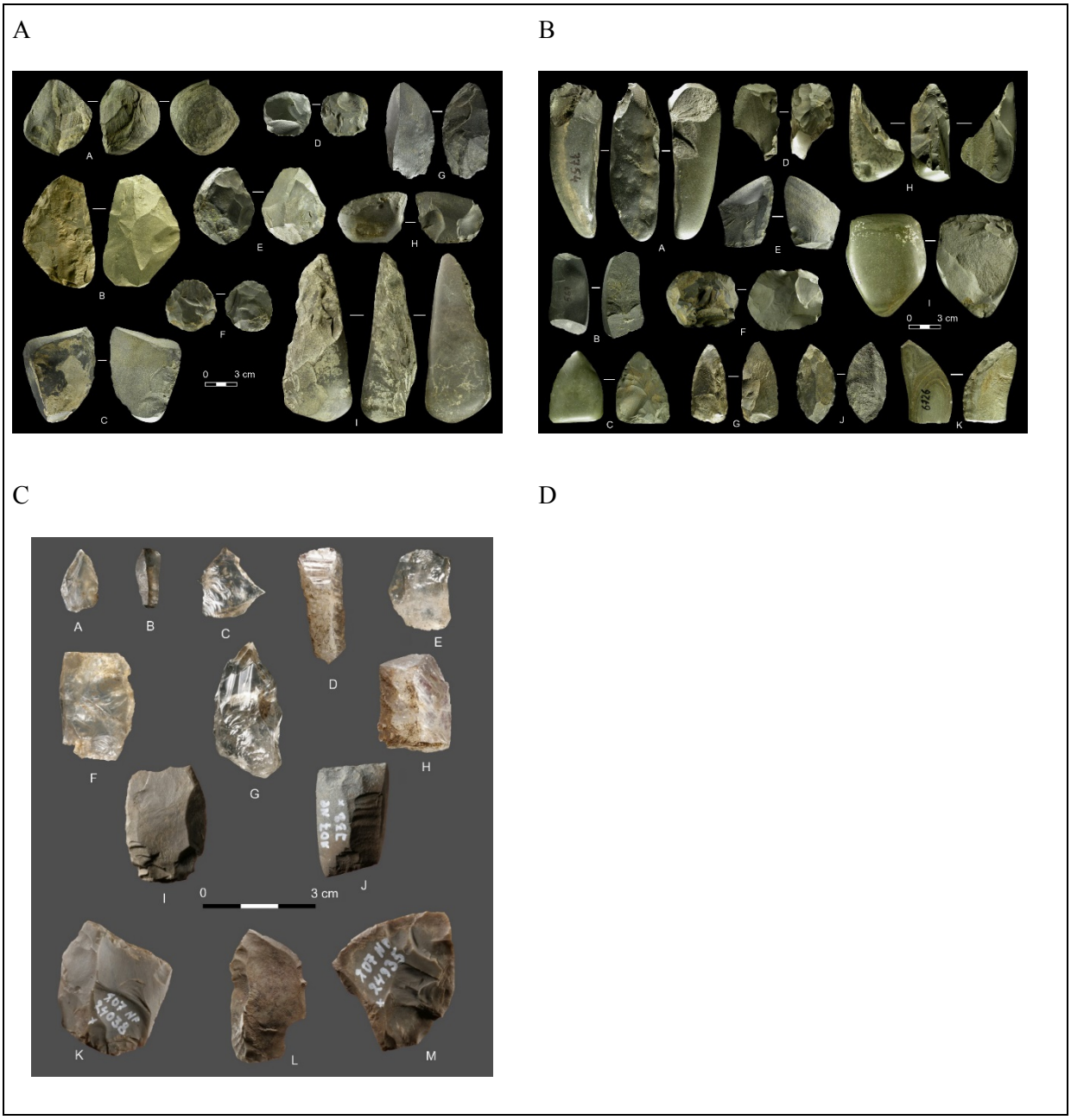


Figure S1.2. Yana site stone tools. **(A)** A, D - discoid cores; E, F - small rounded bifaces (exhausted discoid cores); C, G - naturally backed side-scrapers made on citrus slice flakes, B, H – combined side-scrapers on massive flakes, with partial ventral retouching; I - pick-like tool; (A) - (I) are made of greenish grey siliceous rock (argillite); **(B)** A - chisel-like tool; B, D - carinated end-scraper; C, G - convergent scraper; E - déjeté scraper; F - small rounded biface; H - massive tool with diagonal working edge; I - transverse scraper; J - limace; K - side scraper with a convex working edge; (A) - (K) are made of greenish grey siliceous rock (argillite); **(C)** A - pointed tool; B - backed blade; C - burin; D - chisel-like tool; E - small scraper; F - chisel-like tool; G - burin; H - chisel-like tool; I - small cutting tool; J - chisel-like tool; K - small cutting tool; L – pointed tool; M - small cutting tool; (A) - (H) are of rock crystal; (I), (J), and (L) - grey greenish siliceous rock (argillite); (K) and (M) are of good quality chert; **(D)** A-F, K - small narrow pointed tools made on backed 'bladelets' or flakes; G-I - 'triangles' shaped on flakes; H, P - small transverse scrapers; J - small déjeté-type scraper; L - small double déjeté-type scraper; M-O - backed 'bladelets'; (A) – (I), (L) – (P) are of greenish grey siliceous rock (argillite); (J), (K) are of good quality chert (Pitulko et al. 2012b).

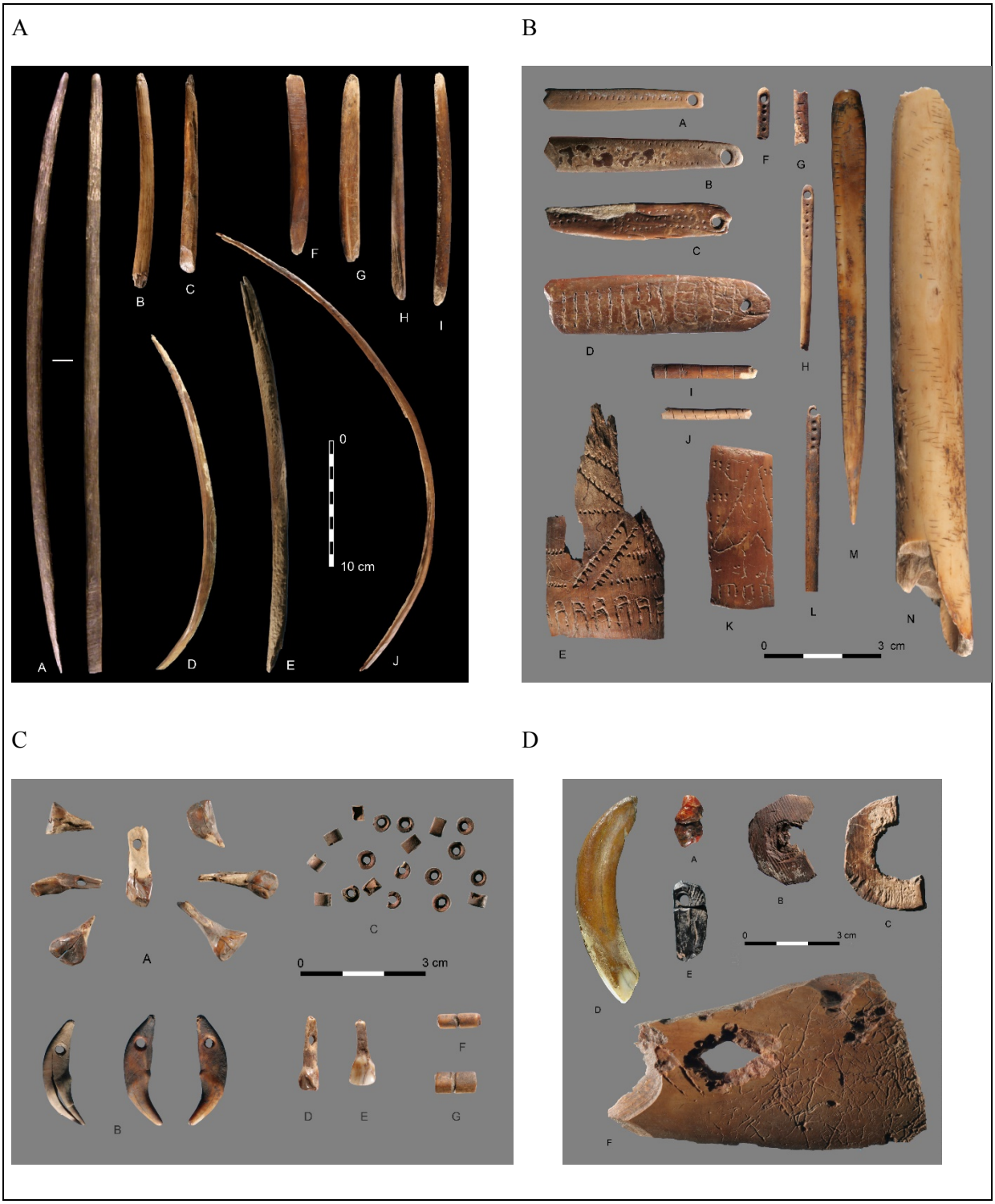


Figure S1.3. Yana site bone, horn and ivory materials. **(A)** A, E-G - beveled rods/foreshafts; D, J - ivory points (sagaie) with unilaterally flattened base part; B, C, H, I - fragments of points; (A) – wooly rhinoceros horn; (B) - (J) – mammoth ivory; **(B)** A to D; fragment of decorated object – E; needles and needle fragments with ownership marks – F, G, H, L; thin ivory shafts with circum incisions – I, J; ivory plaque with anthropomorphic image – K; an awl with regular incisions on the laterals – M; wolf (?) bone with spiral design – N; **(C)** Personal ornaments; A reindeer teeth with drilled holes; B drilled canines of small size carnivore (arctic fox?); C ivory beads; D drilled reindeer tooth; E reindeer tooth with a circumferential cut; F and G tubular beads with a circumferential cut decoration in the middle; **(D)** Amber pendant – A; ivory rings ('Kurtak-type' pendants) – B, C; a tooth pendant (horse incisor with a notch) – D; antraxolite 'mammoth head' pendant – E; 'Kurtak-type' pendant preform – F. (Pitulko & Nikolskiy 2012, Pitulko et al. 2012).

What makes Yana RHS distinctive are the excellent organic preservation conditions (due to permafrost) resulting in abundant faunal remains and a rich and highly diverse bone and ivory industry (Pitulko et al. 2004, 2012, 2014a, 2015). The faunal remains suggest that the occupants of Yana RHS hunted a number of species including mammoth (Nikolskiy & Pitulko 2014). Whereas, bison, horse, and reindeer are most abundant and represent the major food species (Pitulko et al. 2014a), mammoth was also exploited for its ivory, which was widely used as a raw material for tool production (including hunting weapons) and manufacturing of a variety of decorated objects. Small game species are not well represented suggesting that they were hunted rarely, for example there are few arctic fox remains, while Pleistocene hare was hunted more frequently. The majority of Pleistocene hare remains were found articulated, suggesting hare was hunted for fur rather than for its meat (Pitulko et al. 2014a).

1.1.3 Human remains – the Yana RHS teeth – Verner Alexandersen & Charlotte Primeau

The Yana RHS NP locality has produced three human teeth, one in each of the following squares: 2V26, X26, and H29. In this study only the teeth from H29 (Yana 1) and X26 (Yana 2) are used, the other tooth (from square 2V26) has been reported on by Lee et al. (2006). All three teeth can be considered as contemporaneous, because they were found in the same cultural layer. Thus noted, all were recovered during water-screening of the excavated sediments from adjacent polygons and among archaeological remains in the cultural layer, but are not part of burials. Prior to the destructive DNA analysis a morphologic description and measurement of the teeth was undertaken.

1.1.3.1 Deciduous second molar from the left side of the maxilla (H29)

The tooth crown (**Fig. S1.4**) is preserved, but the roots have been resorbed. The tooth had been shed in vivo. The occlusal surface is moderately worn with exposure of secondary dentin on the two lingual cusps. Thin edges of the root trunk are still preserved. Supragingival calculus is observed on the basal parts of the crown. All four cusps are well developed and the two mesial cusps are larger than the distal cusps. The oblique ridge connecting the mesiolingual cusp with the distobuccal cusp across the occlusal surface is still visible. Minor structures such as accessory tubercles on the marginal ridges or a metaconule

(intermediate tubercle on an upper molar between the hypocone and the metacone) on the oblique ridge cannot be observed. Dental wear may have removed such tubercles.

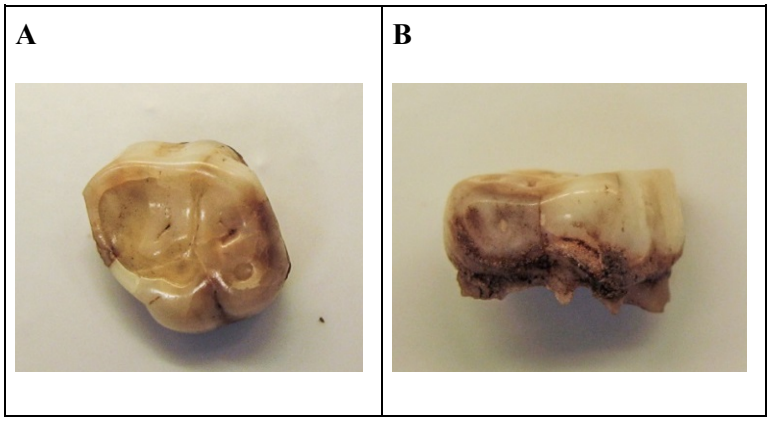


Figure S1.4. The deciduous second molar from the left side of the maxilla. (A) Shows the occlusal surface. Mesial side on the left, buccal side at the top, distal side at the right and lingual side at the bottom of the photo. (B) the same tooth seen at an angle from the distal side. Supragingival calculus is visible on the enamel surface near the base of the crown.

For comparison, these minor ridges and tubercles are present on an unworn European second deciduous molar (**Fig. S1.5.A**) while they are missing on an Inuit second deciduous molar (**Fig. S1.5.B**) that is worn just as much as the Yana tooth. The buccal surface of the crown on the Yana tooth is smooth with a larger mesial and a smaller distal bulging cusp. The lingual surface has a deep groove separating the two lingual cusps. On the lingual surface of the larger mesial cusp, a faint groove is seen near the corner to the mesial surface. This may be a faint evidence of a Carabelli's structure (a small or large additional structure at the mesiolingual angle of maxillary second deciduous molars or first permanent molars).

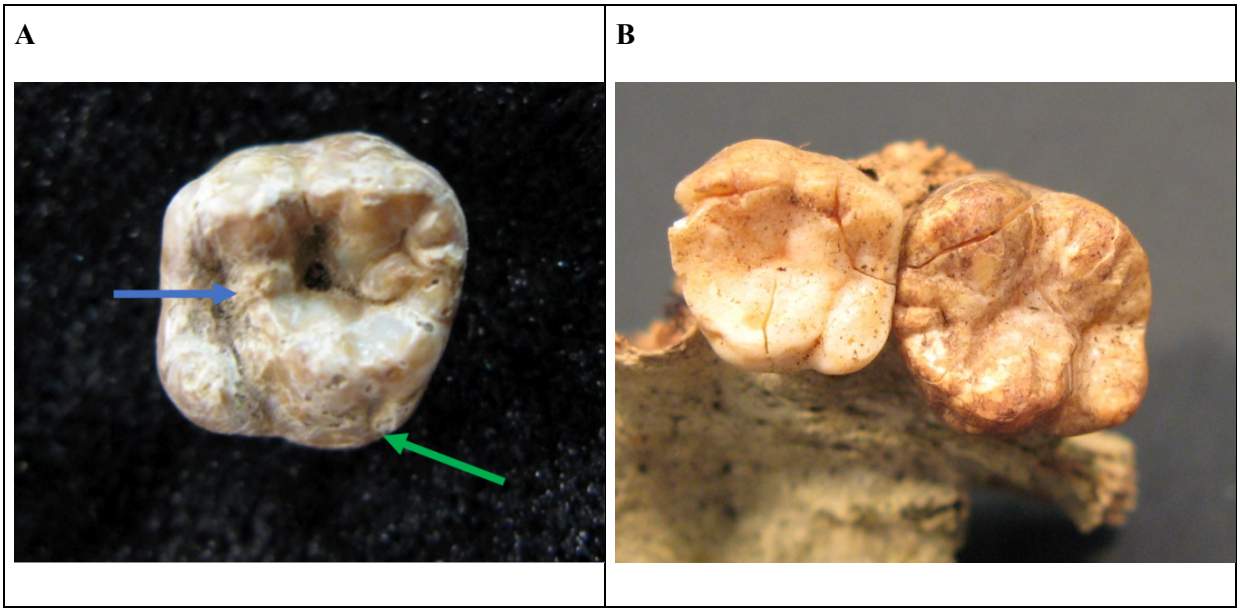


Figure S1.5 (A) Deciduous second molar from the right side of a maxilla, showing the occlusal surface. The buccal side is seen at the top of the image, distal side to the left, mesial side to the right and lingual side at the bottom of the image. Blue arrow points to the metaconule and green arrow to the Carabelli's structure. The material is curated at the Laboratory of Biological Anthropology, Copenhagen. The individual is from the Danish Viking Age, Slagelse, grave 63. (B) Deciduous second molar (on the left) and permanent first molar (on the right) of the upper right side of an Inuit child, showing the occlusal surfaces. The material is curated at the Laboratory of Biological Anthropology, Copenhagen. Individual is KAL-0038x02

On a European second deciduous molar from the right side of the jaw (**Figure S1.5.A**), from the Viking period, the Carabelli's structure is slightly more developed and visible at the base of the large mesiolingual cusp. The Inuit molar (**Figure S1.5.B**) again shows the faint groove associated with a protuberance that show the evidence of a Carabelli' structure. The cervical enamel line follows a wavy course around the base of the crown with no enamel extensions found on the small portion of the root trunk remaining on the tooth (**Figure S1.4.B**).

1.1.3.2 Comparative aspects of size of the Yana molar

The mesiodistal and buccolingual diameters of the Yana tooth were measured with a sliding caliper to the nearest 0.1mm. The mesiodistal diameter measured 9.15mm and the buccolingual diameter measured 10.0mm. Comparative diameters of selected individuals and groups are presented in **Table S1.2**. It is only teeth from the modern populations that has sex separate as this was known at the time of data collection. Sex of skeletal remains cannot be determined with any acceptable accuracy for younger children (Klales and Burns 2017).

Table S1.2 shows that crown diameters of second deciduous molars have variability across populations and time periods with no definitive pattern. The mesiodistal and buccolingual diameters of the Yana molar are within plus/minus one standard deviation from most comparable samples where the variance could be calculated. The crown size of the Yana second deciduous molar can therefore not be placed in a close relation to any of the populations from **Table S1.2**, either relative to time period or geographical region.

Table S1.2. Tooth size for deciduous second upper molars in selected individuals and populations

Period	Site / Population	N	Mesiodistal diameter ($\bar{x}$)	SD	N	Buccolingual diameter ($\bar{x}$)	SD	Author
Upper Paleolithic	Yana	1	9.15	-	1	10.0	-	*Present study*
Upper Paleolithic	European	5	9.60	0.78	10	10.5	0.51	Frayer, 1978
Upper Paleolithic	European	9	9.88	-	3	9.40	-	Brabant, 1969
Upper Paleolithic	Abri Pataud, France	1	9.20	-	1	9.40	-	Legoux, 1975
Upper Paleolithic	Abri Pataud France	1	9.40	-	1	11.2	-	Legoux, 1975
Upper Paleolithic	Pavlov, Czech Rep.	1	9.0	-	1	9.60	-	Klima et al. 1997
Upper Paleolithic	Pavlov Czech Rep.	1	8.0	-	1	9.80	-	Klima et al. 1997
Upper Paleolithic	Pavlov Czech Rep.	1	8.0	-	1	9.80	-	Klima et al. 1997
Upper Paleolithic	Pavlov Czech Rep.	1	8.0	-	1	10.0	-	Klima et al. 1997
Mesolithic	European	11	9.10	0.63	22	10.30	0.30	Frayer, 1978
Mesolithic[2]	Denmark	6	9.26	0.39	4	10.40	-	*unpublished*[1]
Pre-colonial	Greenland	37	9.27	0.57	38	9.83	0.44	*unpublished*[1]
Modern[3]	Greenland SW	30	8.89	0.31	29	9.86	0.33	*unpublished*[1]
Modern[3]	Greenland E	17	9.18	0.40	17	10.24	0.59	*unpublished*[1]
C. 3000 B.P	North America, Ohio Indian	35	9.20	0.68	34	10.30	0.48	Sciulli, 1990

Modern	Iceland	168 158	9.0 (boys) 8.97 (girls)	0.45 0.45	245 200	10.10 (boys) 9.88 (girls)	0.44 0.41	Axelsson & Kirveskari, 1984
Modern	Pima Indian	24 22	9.75 (boys) 9.53 (girls)	0.51 0.46	24 21	10.63 (boys) 10.43 (girls)	0.52 0.37	Alvrus, 2000
Modern[4]	American White	33 36	9.03 (boys) 8.97 (girls)	0.44 0.48	40 38	10.10 (boys) 9.73 (girls)	0.49 0.55	*unpublished*[1]

[1]Unpublished data measured by Verner Alexandersen
[2]Teeth from the Mesolithic period from Denmark are from Nederst (grav 9), Nivå Møllevej, Vedbæk, (Bøgebakken grav 19B) and Dobbeltgraven; Strøby Egede (grav E og B).
[3] Inuit teeth from Modern Greenlanders from the Southwest (S.W) and East (E.) coasts were measured from dental plaster casts (P.O. Pedersen's collection).
[4]Teeth from American White were measured from dental plaster casts of children from Madison, Wisconsin.

1.1.3.3 Morphological variability of the maxillary deciduous second molar.

Hanihara (1967), studied crown characteristics in the deciduous dentition and was able to distinguish between populations with a Caucasoid (White Americans) and an Asiatic (Eskimo, Pima Indians and Japanese) dental complex. Hanihara defined 8 variable traits for incisors, canines and molars. For the deciduous upper second molars Hanihara found marked differences in frequency and size of the Carabelli's structure, and the metaconule. Such regional differences in frequency of dental traits are well known from the permanent dentition and led to establishment of the Arizona State University Dental Anthropology System (ASUDAS) for recording the trait variability (Scott and Turner 1997).

Carabelli's structure. The Carabelli's structure is manifested as an extra mesiolingual protuberance, sometimes even a large cusp, on the deciduous second molar and the permanent first molar in the maxilla and more rarely on second and third permanent molars. When the structure is visible on the permanent first molar it will also be present on the second deciduous molar (Kieser 1984). In the ASUDAS system, 7 grades of development are distinguished (Scott and Turner 1997). European populations in general show a higher frequency of well-developed Carabelli's structures (Carabelli's tuberculum) compared to North-Asiatic populations (**Table S1.3**). The frequencies of individuals with the lower grades of the structure show great variability and the breaking point between absence and presence of a distinct Carabelli's structure in the ASUDAS varies between researchers. Grades 3-7 or 2-7 definitely indicate the presence of the structure.

Table S1.3. Frequencies (%) of presence of two dental traits in four ethnic groups

	Japanese	Pima Indian	Inuit	White American
Asiatic complex Metaconule	41.8	47.0	29.1	3.5
Caucasoid complex Carabelli's structure	11.9	0.0	0.0	35.1

(Adapted from Table 1 in Hanihara, 1967)

The Yana molar has a Carabelli's structure grade 1. This does not unambiguously point in the direction of the Asiatic pattern, because photos of deciduous second molars from European Upper Paleolithic also show that weak development, or even lack of the structure, occurred in those populations. This was the case for Sunghir, Child B (Turner 1986), Kostenki grav 15 (Haeussler 1995), Dolni Vestonice: individual 13, 14 and 15 as well as Pavlov individual 20, 28 (Hillson 2006). However, a large tuberculum is observed on Pavlov individual 36, which also has a large hypocone (Hillson 2006). Abri Pataud, individual no. 1 and 2 from Upper Paleolithic France, also has a faintly developed Carabelli's structure (Legoux 1975). Hence, the presence of a Carabelli's structure on the Yana molar does not point to either an Asiatic or European descent.

Accessory ridges on the cusps, tubercles on the marginal ridges and metaconule. The occlusal wear on the Yana molar has removed all signs of accessory ridges on the cusps and tubercles from the marginal ridges. Unworn permanent molars and premolars may have accessory ridges on the primary cusps and the marginal ridges Such ridges may be more common in Asiatic than in European populations. Ludwig (1957) showed this to be the case for mandibular second premolars. A special tubercle, the metaconule, can be found in the middle section of the oblique crista. It is more common in Asiatic than Caucasoid populations (**Table S1.3**).

The oblique ridge shows no remains of the metaconule on the Yana molar. It is surprisingly well developed and prominent on the Danish molar from the Viking period (**Figure S1.5.A**) although less than 5% out of 1200 modern Danish children showed a metaconule (Jørgensen 1956). On the two Eskimo Inuit teeth (**Fig. S1.5.B**) only the unworn permanent first molar shows remains of the metaconule as well as other accessory ridges. In North-Asiatic populations the metaconule can be very common. Kitagawa et al. (1995) even observed an incidence of 73.1% in their Jomon sample. Out of the eight dental traits in the deciduous dentition capable of discrimination between Caucasoid and North Asiatic populations, only

two are related to the upper second molar crown; the Carabelli's structure and the metaconule (**Table S1.3**). Due to the morphological variability existing in each human population it is impossible to refer the single Yana molar to any of the two dental complexes.

1.1.3.4 The maxillary deciduous canine (X26).

The crown is well preserved with the root missing (**Fig. S1.6.A**). The root is resorbed (**Fig. S1.6.B**) and the tooth was exfoliated and lost in vivo. The crown is worn to a considerable extent with exposure of secondary dentin on the horizontal occlusal surface. The tooth was difficult to identify with regard to position, but the open pulp chamber was oval with the longer axis directed mesiodistally. This indicates a maxillary canine. The position of being from the left or right side could not be determined. The size of the tooth crown could not be measured due to the severe wear of the tooth. The lingual tuberculum was smooth and moderately developed. Supragingival calculus is present on the facial surface.

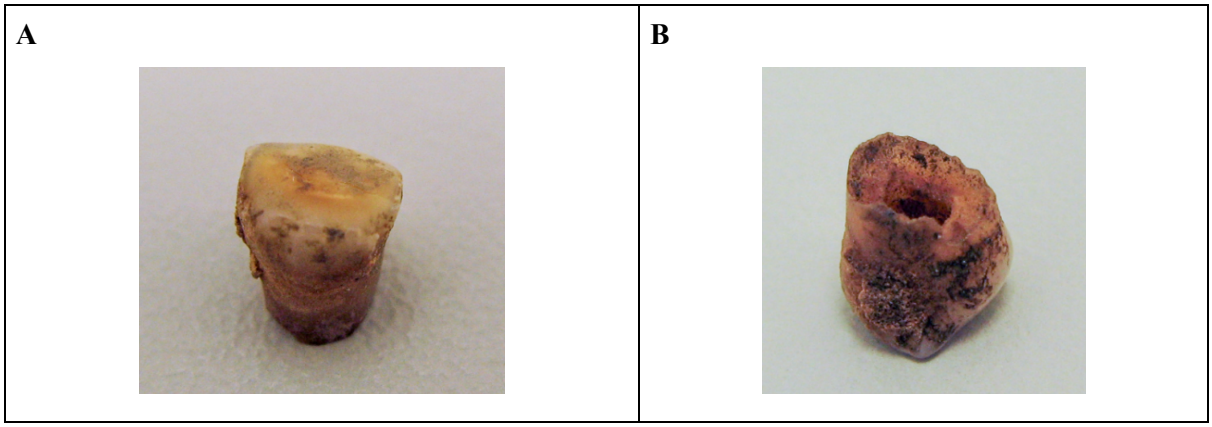


Figure S1.6. The deciduous canine, likely from the maxilla. (A) shows the canine at an angle from the facial side, with supragingival calculus visible on the enamel surface near the base of the crown. (B) shows the root of the canine.

1.1.3.5 Did the teeth belong to one or two children?

The missing roots and the appearance of what remains of them as thin irregular edges is evidence of normal tooth shedding of the Yana molar and canine at the time of eruption of the succeeding permanent second premolar and permanent canine. Both teeth will normally be shed at an age around 10-12 years Nyström et al. (2001). In any population there is a large variation in time and order in the eruption of teeth **Table S1.4,** and a deciduous second molar and canine will usually be shed months apart in the same child (Nyström et al. 2001).

The molar was shed due to the eruption of the permanent second premolar. It only showed

moderate attrition. As more attrition would be expected at the time of shedding, this could mean the child had an early eruption of the permanent teeth. The canine was shed due to the eruption of the permanent canine. However, this canine showed extensive attrition, and should be exfoliated about the same time as the molar if it belonged to the same child. It is not unlikely that the anterior teeth including the canine becomes more affected by attrition than the second molars in children from the Upper Paleolithic period, but we suggest it is more likely that the teeth belonged to two individuals.

Table S1.4. Dental eruption for a modern Finnish sample

Boys				**Girls**			
Tooth	5%	Median	95%	Tooth	5%	Median	95%
Upper perm. canine	9.5	11.41	13.5	Upper perm. canine	9.0	10.77	13.5
Upper second premolar	9.9	12.11	14.7	Upper second premolar	9.5	11.37	14.0

(Calculations were based on a combination of exact emergence ages and ages based on clinical examinations at half-year or 1-year intervals. Adapted from Nyström et al. 2001).

Deciduous teeth from the same child can be found at rare occasions. Hillson (2006), examined several shed deciduous teeth in an Upper Paleolithic site in the Czech Republic and found a first and second deciduous molar supposedly belonging to the same child. However, considering that the archaeological site of Yana-RHS was large with many artifacts and animal bones (Pitulko et al. 2004), it seems likely that several families and several children stayed at the same time on the site. This is one more reason to suggest that the Yana teeth belonged to two children.

1.1.3.6 Conclusions regarding the dental material from Yana RHS

The second deciduous molar from the left side of the maxilla cannot be unambiguously assigned to a European or Asiatic population from its morphological appearance. The deciduous canine, probably from the maxilla, is likewise too worn to contribute to any conclusions. Both teeth would have been shed at an age corresponding to about 10-12 years. It is concluded that the two teeth most likely derived from two children due to the difference in dental attrition, but it cannot be excluded that they may from a single child.

References

Alvrus, A. Sex dimorphism in the deciduous dentition of modern Pima. *Dental Anthropology* **14**, 9-14 (2000).

Axelsson, G. & Kirveskari P. Crown Size of Deciduous Teeth in Icelanders. *Acta Odontologica Scandinavica* **42**, 339-343 (1984).

Basilyan A.E., Anisimov M.A., Nikolskiy P.A., & Pitulko V.V. Wooly mammoth mass accumulation next to the Paleolithic Yana RHS site, Arctic Siberia: its geology, age, and relation to past human activity. *Journal of Archaeological Science* 38, 2461-2474 (2011).

Brabant, H. Observations sur L'evolution de la Denture Temporaire Humaine en Europe Occidentale. *Bulletin du Groupement International pour la Recherche scientifique en Stomatologie* **8**, 235-302 (1965).

Floss, H. *Rohmaterialversorgung im Paläolithikum des Mittelrheingebietes.* RGZM Monographie. Verlag des RGZM, Mainz (1994).

Frayer, D.W. *Evolution of the Dentition in Upper Paleolithic and Mesolithic Europa.* Publications in Anthropology, 10 (Lawrence: University of Kansas 1978).

Gakkel, Y.Y. (Ed.). *Novosibirskie ostrova* [New Siberian islands]. Gidrometeoizdat, Leningrad (1967).

Gureev, A. Fauna of the USSR, mammals, Lagomorpha. Nauk, Moscow, USSR 3: 1–276 (1964).

Haeussler, A.M. *Upper Paleolithic teeth from the Kostenki sites on the Don River, Russia.* In: (J. Moggi-Cecchi ed.) Aspects of dental biology: paleontology, anthropology and evolution (International Institute for the Study of Man 1995).

Hanihara, K. Racial Characteristics in the Dentition. Journal of Dental Research **46**, Supplement to No. **59**, 23-26 (1967).

Hillson, S. *Dental Morphology, Proportions, and Attrition.* In: (Trinkaus E. and J. Svoboda eds.) Early Modern Human Evolution in Central Europe. Human Evolution Series. (Oxford University Press 2006).

Jørgensen, K.D. The Deciduous Dentition. A descriptive and comparative anatomical study. *Acta Odontologica Scandinavica* **44**, Suppl. 20 (1956).

Kieser, J.A. An analysis of the Carabelli trait in the mixed deciduous and permanent human dentition. *Archives of Oral Biology* **29**, 403-406 (1984).

Kitagawa, Y., Manabe, Y., Oyamada, J. and Rokutanda, A. Dental Morphology of the Prehistoric Jomon People of Japan: Comparison of Nonmetric Characters. *American Journal of physical Anthropology* **97**, 101-113 (1995).

Klales, A.B., and Burns, T.L. Adapting and Applying the Phenice (1969) Adult Morphological Sex Estimation Technique to Subadults. *Journal of Forensic Sciences* **62**(3), 747-752 (2017).

Klíma, B., Vlcek, E., Adovasio, J.M., Damblon, F., Hyland, D.C., Jarosova, L., Musil, R., van der Plicht, J., Soffer, O., Svoboda, J., Skrdla, P., Trinkaus, R., Vandiver, P. and Verpoorte A. *Pavlov I – Northwest. The Upper Paleolithic Burial and its Settlement Context* (Brno: Academy of Sciences of the Czech Republic, Institute of Archaeology 1997).

Lee E.J., Merriwether D.A., Kasparov A.K., Khartanovich V.I., Nikolskiy P.A., Shidlovskiy F.K., Gromov A.V., Chikisheva T.A., Chasnyk V.G., Timoshin V.B., Pavlova E.Y. & Pitulko V.V. A genetic perspective of prehistoric hunter-gatherers in the Siberian Arctic: ancient DNA analysis of human remains from 8,000 years ago. *Journal of Archaeological Science: Reports.* DOI:10.1016/j.jasrep.2016.06.001 (2016).

Legoux, P. *Présentation des Dents des Pestes Humains de L'Abri Pataud*. In: (Movius, H.L.J ed.) Excavation of the Abri Pataud, Les Eyzies (Dordogne). (American School of Prehistoric Research. Bulletin No. 30. 1975).

Ludwig, F.J. The mandibular second premolars: Morphologic variation and inheritance. *Journal of Dental Research* **36,** 263-73 (1957).

Medvedev, G.I. Art from central Siberian Palaeolithic sites. In A.P. Derevianko (Ed.) *The Paleolithic of Siberia: new discoveries and interpretations*: 132–37. Institute of Archaeology& Ethnography SB RAS, Novosibirsk; University of Illinois Press, Urbana (IL) (1998).

Nikolskiy, P.A. and Pitulko, V.V. Evidence from the Yana Palaeolithic site, Arctic Siberia, yields clues to the riddle of mammoth hunting. *Journal of Archaeological Science* 40, 4189-4197 (2013).

Nyström, M.E., Kleemola-Kujala, M., Evälahti, L., Peck, L. and Kataja, M. Emergence of permanent teeth and dental age in a series of Finns. *Acta Odontologica Scandinavica* 59, 49-56 (2001).

Pitulko V.V. Chapter 16. Digging through permafrost in Siberia. In Carver M., Gaydarska B., Monton-Subias S. (Eds.). Field Archaeology from Around the World. Ideas and Approaches. Springer International Publishing Switzerland. P. 111-113. (2015).

Pitulko V.V. Principal excavation techniques under permafrost conditions (based on Zhokhov and Yana sites, Northern Yakutia). Archaeology, Ethnology & Anthropology of Eurasia 34, 26–33 (2008).

Pitulko, V., Nikolskiy, P., Basilyan, A. & Pavlova, E. Human habitation in the Arctic Western Beringia prior the LGM. In Graf, K.E., Ketron, C.V. and Waters, M.R. (eds), *Paleoamerican Odyssey*, CSFA, Dept. of Anthropology, Texas A & M University, Texas A&M University Press, College Station, pp. 13–44 (2014a).

Pitulko V.V., Pavlova E.Y. & Ivanova V.V. Upper Paleolithic art of the Arctic Siberia: personal adornments from excavations of the Yana site. *Ural Historical Journal* 2, 6-18 (2014b).

Pitulko, V.V. & Nikolskiy, P.A. Extinction of wooly mammoth in Northeastern Asia and the archaeological record. *World Archaeology* 44, 21-42 (2012).

Pitulko, V.V., Nikolsky, P.A., Girya E. Yu., Basilyan, A.E., Tumskoy, V.E., Koulakov, S.A., Astakhov, S.N., Pavlova, E.Yu. and Anisimov, M.A. The Yana RHS Site: Humans in the Arctic Before the Last Glacial Maximum. *Science* **303**, 52-56 (2004).

Pitulko, V.V., Pavlova, E.Y., Nikolskiy, P.A. & Ivanova, V.V. The Oldest Art of Eurasian Arctic. *Antiquity* 86: 642-659 (2012).

PitulkoV.V. & Pavlova E.Y. Art of the Yana site: Mammoth Ivory Diadems and Bracelets (preliminary analysis of the collection). In N.V. Fedorova (Ed). Archaeology of the Arctic. Vol. 2. Ekaterinburg: Delovaya Pressa. P.140-161 (2015a).

PitulkoV.V., Pavlova E.Y., & Nikolskiy, P.A. Mammoth Ivory Technologies in the Upper Palaeolithic Arctic Siberia: a Case Study based on the materials from Yana RHS site. *World Archaeology* 47, 333-389 (2015b).

Pitulko, V.V., Pavlova E.Y., & Nikolskiy, P.A 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 (2018)

Sciulli, P.W. Deciduous Dentition of a Late Archaic Population of Ohio. *Human Biology* **62**, (2), 221-245 (1990).

Scott, G.R. and Turner II, C.G. *The anthropology of modern human teeth*. Dental morphology and its variation in recent human populations (Cambridge University Press 1997).

Sinitsyn, A.A. in *New aspects of the Central and Eastern European Upper Palaeolithic - methods, chronology, technology and subsistence* (eds Neugebauer-Maresch, C. & Owen, L.R.) 27-48 (Österreichische Akademie der Wissenschaften, 2010).

Soffer, O. *The Upper Paleolithic of the central Russian plain.* New York: Academic Press (1985).

Svoboda, J.A. Raw material sources in Early Upper Paleolithic Moravia. The concept of lithic exploitation areas. *Anthropologie XXI/2*, 147-158 (1983).

Svoboda, J.A., Ložek, V. & Vlèek, E. *Hunters between East and West. The Paleolithic of Moravia* (Plenum Press, New York, 1996).

Turner II, C.G. The First Americans: The Dental Evidence. *National Geographic Research* **2**(1), 37-46 (1986).

White, R. Intégrer la complexité sociale et opérationnalle: la construction matériale d'identité sociale à Sungir. H. Camps-Fabrer (Ed.). *Préhistoire d'os*. L'Universite de Provence, Provence, pp. 319-333 (1997).

White, R. Ivory personal ornaments of Aurignacian age: technological, social, and symbolic perspective. In J. Hahn, M. Menu, Y. Taborin, P. Walter, F. Widemann (eds).*Le Travail et l'usage de l'ivoire au Paleolithique Supereiur*. Centro Universitario Europeo per I Bene Culturali. Actes de la Table Ronde, Ravello, pp. 29-62 (1995).

1.2. Devil's Gate Cave

1.2.1 Site description

Devil's Gate Cave (Chertovy Vorota) is located on a steep limestone cliff some 35 m above Krivaya River – a tributary of the Rudnaya River – in Primorye Province in the Russian Far East (44°29'N, 135°30'E) (Kuzmin et al. 2012). The site is ~12 km southwest of the city of Dal'negorsk, and ~30 km to the present day coastline of the Sea of Japan.

The main chamber of Devil's Gate cave is ~45 m deep and 5-10 m wide. Excavations in the cave were carried out in the summer of 1973, extended over 170 m^2, and revealed a cultural layer 20-40 cm thick belonging to the Primorye – Rudnaya Neolithic archaeological culture (Andreyeva, 1974; Kuzmin et al. 2002; Kuzmin et al. 2012; Zhuchikhovskaya, 2006). In the central part of the site, carbonized wood remains were found from a ~45 m^2 rectangular pit dwelling that had burned (Kuzmin et al. 2012: Figure 4).

The majority of the cultural material from the site was found within the structure, and included hundreds of stone, bone and antler tools (including projectile points, knives, scrapers, harpoon heads, and daggers), pottery vessels, bone and marine shell ornaments, as well as other organic material such as wooden artefacts and textiles. The textile material is of particular significance, representing as it does some of the earliest textiles dated in the world, ranging in age from 7900 to 8200 radiocarbon years BP (Kuzmin et al. 2012: Table 1). The textiles included carbonized remains of ropes and strings, nets and woven textiles, made from sedge grass (Kuzmin et al. 2012).

Recovered animal bones were a variety of prey species, including brown and black bears, wild boar, red deer, and badger, multiple species of birds (grouse, ptarmigan, dove and duck), along with freshwater and anadromous fish (salmon) (Andreeva 1991; Kuzmin et al. 2012). Stable isotope analyses of human bone indicated ~25% of the diet was based on marine foods (Kuzmin et al. 2002). The assemblage does not include any domestic plants or animals, which suggests that the inhabitants of the site belonged to a hunter-gatherer-fisher population who also manufactured pottery.

1.2.2 Human remains

On the floor of the dwelling, disarticulated human remains were found. In total, bones from 7 individuals could be identified (Siska et al 2017):

- NEO236 (Skull Б [B], Devils Gate 2) – adult female (misidentified originally as a male), 20-25 years old. Remains date to 6793 ± 43 radiocarbon years BP (UBA-33764).
- NEO237 (Skull Д [D], Devils Gate 3) – adult female, 50-60 years old. Remains date to 6687 ± 46 radiocarbon years BP (UBA-33765).
- NEO238 (Skull A, Devils Gate 4) – juvenile female, 6-7 years. Remains date to 6678 ± 48 radiocarbon years BP (UBA-33766).
- NEO240 (Skull E, Devils Gate 1) – adult female, 40-50 years old. Remains date to 6627 ± 45 radiocarbon years BP (UBA-33768).
- NEO235 (Skull Г [G]) – adult female (misidentified originally as a male), ~50 years old. Remains date to 6826 ± 47 radiocarbon years BP (UBA-33763).
- NEO239 (Skull Ж) (Devils Gate 5) – complete skull of a young person, 18-20 years old (believed to be a male based on morphology but genetically determined to be female)
- Skull B [V]– sub-adult, 12-13 years old

Anthropological research was conducted by T. S. Balueva (1978). In her view, the morphological combination peculiar to the modern Tunguso-Manchurian population of Priamurye and Primorye was already created in the Neolithic era.

The description of the male skull of 18-20 years which is the best preserved: Giperbrakhikraniya (l – 172 mm; b – 158mm). Height of the skull – 144 mm. High and wide facial skeleton. Big orbits. Massive lower jaw. Ortognatic profile.

Fragments of other skulls also confirm a Mongoloid complex. Probable accessory to the Baikal group of North mongoloid anthropological type.

References

Andreeva, Z. V., V.A. Tatarnikov, Peshera “Chertovy Vorota v Primor’e” (The “Devil’s Gate” Cave in Russian Far East). Arheol. Otkrytiya 1973 Goda. Mosc. Nauka, 180–181 (1974).

Andreeva, Z.V. (ed.) 1991. *Neolit Dalnego Vostoka: drevnee poselenie v peshchere Chertovy Vorota* [The Neolithic of the Far East: ancient settlement in Chertovy Vorota Cave].Moscow: Nauka (in Russian).

Baluyeva T. S. Kraniologichesky material of a neolytic layer of a cave “Devil’s Gate” (Primorye). *Questions of anthropology*, issue 58. 1978. page 184-187.

Kuzmin, Y V., Charles T. Keally, A.J. Timothy Jull, George S. Burr, 2012. The earliest surviving textiles in East Asia from Chertovy Vorota Cave, Primorye Province, Russian Far East. *Antiquity* 86, Issue 332, pp. 325-337.

Zhushchikhovskaya, I. S. Archaeology of the Russian Far East: Essays in Stone Age Prehistory, S. M. Nelson, A. P. Derevianko, Y. V. Kuzmin, R. L. Bland, Eds. (Archaeopress, 2006), pp. 101–122.

1.3 Ekven & Uelen sites

The Ekven and Uelen sites are both located within 25 km of one another on the Bering Sea coast, in the region of Cape Dezhnev (East Cape). Both sites have yielded a large number of burials of the Neoeskimo tradition of Siberia and northern North America. That tradition is divided into several subgroups, starting with the Old Bering Sea (OBS) group which is later replaced by the Punuk and Birnirk cultures (for a discussion of the cultural chronology, see Fitzhugh 2009). Most of the remains in these two cemeteries date to the OBS. We briefly describe each site in turn.

1.3.1 Ekven site description

The Ekven site, one of the largest sites in Chukotka, was found in 1961, and is located on two hills (Western and Eastern) in close proximity to one another on the shore of the Bering Sea, 15 km south of Cape Dezhnev (East Cape) (66° 01' 17.5" N, 170° 06' 25" W) (Bronshtein and Dneprovsky 2009). The site was occupied principally during the first millennium, and includes cultural remains associated with the Old Bering Sea (OBS), Birnirk and early Punuk cultures (Bronshtein and Dneprovsky 2009; Fitzhugh 2009).

The Ekven site includes an extensive cemetery, investigated by Arutiunov and Sergeev from 1961-1970 (Arutiunov and Sergeev 2006a [original, 1975]), and which led to the recovery of 210 burials, including a considerable number of male-female pair and group burials. A second round of investigations in the 1980s-1990s by Bronshtein and Dneprovsky recovered an additional 120 burials, and located a nearby settlement (~1 km distant) comprised of ~20-30 houses constructed of whale bones, driftwood and turf for insulation (Bronshtein and Dneprovsky 2009; Csonka 2009).

The Ekven villagers exploited ringed and bearded seals, walruses, whales (gray and bowhead), foxes, polar bears, birds, fish, and mollusks (Csonka 2009). Stable isotope analyses from 11 individuals at Ekven were reported by Kuzmin (2010). The average values were -11.8 ‰ for $\delta^{13}C$ and 19.5 ‰ for $\delta^{15}N$. For the samples in the present project the mean values were -11.4 ± 0.3 ‰ for $\delta^{13}C$ and 21.5 ± 0.3 ‰ for $\delta^{15}N$ (below). The slight difference from the previously reported results may be due to different individuals having been analyzed. These values are quite extreme and suggest that most of the protein in the diet came from marine sources on a high trophic level. This is consistent with archaeological evidence suggesting a specialization on hunting marine mammals such as walrus and whales (Bronshtein and Dneprovsky 2009). Likewise, many of the Ekven burials were accompanied

by rich grave goods, which included sea mammal hunting weaponry such as bows, atlatl weights, and toggling harpoons (many hundreds of the latter were recovered). There were in addition a range of other artifacts, including walrus tusk vessels, snow goggles, adze and ulu handles (Arutiunov and Sergeev 2006a).

The dominance of marine hunting leads to problems in the calibration of radiocarbon ages, as the marine reservoir effect in the area seems to be high and locally variable. Dumond & Griffin (2002) and Kuzmin et al. (2002) report reservoir ages ranging ca. 370-750 years in the eastern Bering Sea and Primorye Province of the Russian Far East; local values for the Chukotka coast are not known, however. It is therefore difficult to make a meaningful formal calibration, but in general we would suggest that a rough estimate would be that the radiocarbon ages for the sampled Ekven individuals, which range from ~1730-2200 radiocarbon years BP, probably date from sometime in the early centuries of the first millennium.

1.3.2 Ekven human remains

Of the 330 human remains recovered from the Ekven site, 189 were described in more or less detail by Arutiunov and Sergeev (2006a). From the remains they recovered, 13 samples for genomic analyses were obtained, all from teeth, and representing 11 individuals. Brief comments on each burial and associated grave goods follow (after Arutiunov and Sergeev 2006).

Burial 3 (Western Hill). Skeleton of an adult male (Debets 2006), with the head to the south in a supine position, and placed directly above two other burials (Burials 4 and 5). The lower legs of this individual were covered by a whale scapula, and found with him were multiple artifacts made of bone and tusk, toggling and non-toggling harpoons, and slate adzes and a few other stone tools (Arutiunov and Sergeev 2006:11-12). Burial 3 dates to 1846 ± 28 radiocarbon years BP (UBA-34050).

Burial 5 (Western Hill). An adult female (Debets 2006) from a double burial (Burials 5 and 6) which, as noted was found stratigraphically below Burials 4 and 3 (in that order). Burials 5 and 6 were oriented in opposite directions, but because of their close proximity elements of the two skeletons were mixed, and not in anatomical position. A variety of artifacts were found with the remains, included a variety of stone, tusk and ivory tools, including carved items (Arutiunov and Sergeev 2006a:12). Burial 5 dates to 2186 ± 35 radiocarbon years BP (UBA-34051).

Burial 7 (Western Hill). The skeleton of a male, its bones in anatomical order, with traces of wood above the burial and under the skeleton. Traces of ochre were found to the right of the skull, with a bone arrowhead in the same place, and various other artifacts including a bird bone, several other stone and slate artifacts, a walrus-tusk bow-drill and mattock, and a figurine in the shape of a seal head (Arutiunov and Sergeev 2006a:14) Burial 7 dates to 1733 ± 34 radiocarbon years BP (UBA-34054).

Burial 9 (Western Hill). Skeleton of an adult female (Debets 2006), found in a lower level of structure with another individual (Burial 8). Burial 9 was laid out in a supine position, the body covered with whale ribs, and accompanied by multiple artifacts of bone, ivory and stone artifacts, principally hunting implements such as harpoon heads and spears (Arutiunov and Sergeev 2006a:14). Burial 9 yielded two dates: 1900 ± 31 radiocarbon years BP (UBA-34060) and 2007 ± 35 radiocarbon years BP (UBA-34057). The two dates are statistically distinct and cannot be averaged. One possible explanation for the discrepant ages is that Burial 9 was in the same structure as Burial 8, although deeper (Arutiunov and Sergeev 2006a:14). Perhaps the more recent age came from the Burial 8 individual, which was stratigraphically higher and presumably more recent.

Burial 10-11 (Western Hill). Double burial with two adult males (Debets 2006), lying in opposite directions in supine position. The burial pit was lined with vertically standing whale bones, and wood remains were found below and above the skeletons. Both individuals were accompanied with rich sets of bone, slate and stone artefacts. (Arutiunov and Sergeev 2006a:15-16). Burial 10 yielded two dates: 2091 ± 47 radiocarbon years BP (UBA-34056) and 2101 ± 49 radiocarbon years BP (UBA-34052). Burial 11 dates to 2070 ± 44 radiocarbon years BP (UBA-34053). All three radiocarbon ages statistically overlap (by chi-square test), and average to 2090 ± 27 radiocarbon years BP.

Burial 12 (Eastern Hill). Skeleton of an adult male (Debets 2006), lying in supine position on a log platform in a grave line with stone slabs, and a accompanied by a large number of bone, slate, stone and wooden objects. There was also a fragmented child's skull found stratigraphically higher in the burial (Arutiunov and Sergeev 2006a:16-17). Burial 12 dates to 1822 ± 35 radiocarbon years BP (UBA-34061).

Burial 13 (Eastern Hill). A double burial, the upper skeleton lying on its back and in anatomical order (save for a missing skull), and the lower, smaller skeleton, a juvenile female (Debets 2006) below but in poorer condition. A number of bone, tusk, slate and stone objects

were found, along with stray human skeletal elements. (Arutiunov and Sergeev 2006a:17-18). Burial 13 dates to 2014 ± 35 radiocarbon years BP (UBA-34049).

Burial 18 (Western Hill). An adult female (Debets 2006) found in the center of a rock slab-lined triple burial, and though it was found in anatomical order it was missing a few skeletal elements and in general was in poor condition. The skeleton was found with multiple harpoon heads and arrowpoints, along with a variety of other objects (Arutiunov and Sergeev 2006a:18-20). Burial 18 dates to 1952 ± 29 radiocarbon years BP (UBA-34055).

Burial 28 (Western Hill). Skeleton of an adult male (Debets 2006). The skeleton was supine, in anatomical order. Grave goods were relatively few, and more diverse, including a carved ulu handle, an arrowhead a toggling harpoon head, a small stone axe, an antler burin handle, and a flake scraper (Arutiunov and Sergeev 2006a:22). Burial 28 dates to 2145 ± 36 radiocarbon years BP (UBA-34058).

Burial 32 (Western Hill). A burial of an adult female (Debets 2006), with a small inventory of grave goods, including a few flakes and a bow drill (Arutiunov and Sergeev 2006a:23). Burial 32 dates to 2035 ± 31 radiocarbon years BP (UBA-34059).

1.3.3 Uelen site description

The Uelen burial ground was discovered in the mid-1950s, and is located on the coast of the Bering Sea on Cape Dezhnev (East Cape) (66°09’ 30” N, 169°48’W), only 170 m from the modern settlement of Uelen (Arutiunov & Sergeev 2006b [original 1969]; Fitzhugh et al. 2009). The Uelen site is ~25 km north of Ekven, and just a few hundred meters from the Old Uelen village site, discovered a decade earlier; the village was only partially excavated, while the burial ground was completely excavated from 1957-1960 (Arutiunov & Sergeev 2006b:37).

The cemetery covered an area of ~50 *x* 20 m and contained more than 78 burials, with rich accompaniments of grave goods, principally from the Old Bering Sea cultural tradition. The position of the Uelen in the broader context of the archaeology and anthropology of the Bering Sea region is provided in Arutiunov & Sergeev (2006b), and especially the ‘Prologue 2006’ of that volume, written by Arutiunov, which updates the findings of the original 1969 monograph.

1.3.4 Uelen human remains

Teeth from two burials from Uelen were sampled, NEO233 and NEO234, corresponding to burials 12/58 and 13/58 in Arutiunov & Sergeev (2006b:47).

Burial 12/58 was a double burial with a male (B) and a female (A), of which the female was sampled. The grave was surrounded by a wall made of stones and whale bones. Both skeletons were in anatomical order, lying supine, one on top of the other. Only the mandible of the lower skeleton (B) was found. The upper skeleton (A) seemed to be female. There were traces of a wooden floor between the two skeletons. The grave contained a rich inventory of bone harpoons, needles and other objects, and also a pottery vessel (Arutiunov & Sergeev 2006b:47, Figure 6). Burial 12 dates to 1837 ± 35 radiocarbon years BP (UBA-34047).

Burial 13/58 contained the skeleton of a young woman. The skeleton was on its side, the bones somewhat displaced and fragmented. The grave contained two zoomorphic bone carvings, a bone scraper, and an awl near the right thigh and pelvis. A fragmented child's skull and ribs were found near the skull. There was a mattock and traces of pottery near the right leg (Arutiunov & Sergeev 2006b:47, Figure 7). Burial 13 dates to 2714 ± 44 radiocarbon years BP (UBA-34048).

1.3.5 General notes on the Ekven and Uelen human remains

Anthropological materials from Ekven and Uelen provide almost the only data available on the physical anthropology of the population of the OBS culture. Based on this material, Levin and Debets have made the conclusion about rather extreme antiquity of that combination of signs which has received the name “Eskimo” or “Arctic” type (Levin, Sergeyev, 1964, Debets 2006). The face is characterized by considerable width in comparison with a brain box, long, narrow and high. Height considerably exceeds width. The face of Eskimos is high and wide, though concedes in the size to Yakuts and Buryats. Also the massive lower jaw is very big.

Odontological materials from the Ekven and, in a smaller measure Uelen, burial grounds show strong likeness with modern Eskimos (especially from the territory of Alaska). At similarity both series among themselves the Uelen group nevertheless approaches the Siberian continental Mongoloids slightly more. In both burial grounds archaic odontological traits are found (Zubov, in Arutiunov and Sergeev 2006b).

On several skeletons (>20%) from Ekven, osteoarthritis and spondylitis deformations were observed (Lebedinskaya 2006). Difficult living conditions, in particular marine hunting in which all members of the collective including teenagers took part, could possibly cause overload in a not quite fully developed organism and lead to degenerate changes in joints. However, Lebedinskaya suggests another possibility, Kashin-Beck disease (KBD), which is endemic to present-day populations in the area. Its cause is not yet clear. Suggested causes of KBD include mycotoxins present in grain, trace mineral deficiency in nutrition, and high levels of fulvic acid in drinking water, but recently genetic causes have also been proposed (Shi et al 2015).

References

Arutiunov, S.A. and D.A. Sergeev (2006a) *Problems of ethnic history in the Bering Sea: the Ekven cemetery*. Translated and edited by Richard Bland from *Problemy etnicheskoi istorii Beringomor'ya (Ekvenskii mogil'nik)*, originally published 1975. Shared Beringian Heritage Program, National Park Service.

Arutiunov, S.A. and D.A. Sergeev (2006b) *Ancient cultures of the Asiatic Eskimos: the Uelen cemetery*. Translated and edited by Richard Bland from *Drevnie kul'tury aziatskikh eskimosov (Uzlenskii mogilnyk)*, originally published 1969. Shared Beringian Heritage Program, National Park Service.

Bronshtein, M.M. and K.A. Dneprovsky (2009) Archaeology at Ekven, Chukotka. In *Gifts from the Ancestors. Ancient Ivories of the Bering Strait*, W. Fitzhugh, J. Hollowell and A. L. Crowell, eds., pp. 94-95. New Haven: Yale University Press.

Csonka, Y. (2009) Life at Ekven: a paleoethnography. In *Gifts from the Ancestors. Ancient Ivories of the Bering Strait*, W. Fitzhugh, J. Hollowell and A. L. Crowell, eds.., pp. 96-109. New Haven: Yale University Press.

Csonka, Y. (2014) *The Ekven Settlement. Eskimo beginnings on the Asian shore of Bering Strait*. BAR International Series 2624.

Debets, G.F. (2006) Paleoanthropological materials from the Old Bering Sea cemeteries of Uelen and Ekven. In *Problems of ethnic history in the Bering Sea: the Ekven cemetery*, pp. 209-252. Translated and edited by Richard Bland from *Problemy etnicheskoi istorii*

Beringomor'ya (Ekvenskii mogil'nik), originally published 1999. Shared Beringian Heritage Program, National Park Service.

Dumond, D.E. & D. Griffin (2002) Measurements of the Marine Reservoir Effect on Radiocarbon Ages in the Eastern Bering Sea. *Arctic* 55:77–86.

Fitzhugh, W. (2009) Notes on art styles, cultures, and chronology. In *Gifts from the Ancestors. Ancient Ivories of the Bering Strait*, W. Fitzhugh, J. Hollowell and A. L. Crowell, eds.., pp. 88-93. New Haven: Yale University Press.

Fitzhugh, W., A. Crowell and J. Hollowell (2009) Art styles, culture and dating. In *Gifts from the Ancestors. Ancient Ivories of the Bering Strait*, W. Fitzhugh, J. Hollowell and A. L. Crowell, eds., pp. 31-41. New Haven: Yale University Press.

Kuzmin, Y.V. (2010) Holocene Radiocarbon-Dated Sites in Northeastern Siberia: Issues of Temporal Frequency, Reservoir Age, and Human-Nature Interaction. *Arctic Anthropology* 47/2, 104-115.

Lebedinskaya, G.V. (2006) Pathological Changes in the Skeletal Remains at Ekven Cemetery. In *Ancient cultures of the Asiatic Eskimos: the Uelen cemetery*, by S.A. Arutiunov and D.A. Sergeev, pp. 213-220. Translated and edited by Richard Bland from *Drevnie kul'tury aziatskikh eskimosov (Uzlenskii mogilnyk)*, originally published 1969. Shared Beringian Heritage Program, National Park Service.

Levin, M.G. and D.A. Sergeev (1964) Drevnie mogil'niki Chukotki i nekotorye aspekty eskimosskoi problemy [Early cemeteries of Chukotka and some aspects of the Eskimo problem]. Doklady na VII Mezhdunarodnom kongresse antropologicheskikh i etnograficheskikh nauk. Moscow.

Shi X., F. Zhang, A. Lv, Y. Wen & X. Guo (2015) COL9A1 Gene Polymorphism Is Associated with Kashin-Beck Disease in a Northwest Chinese Han Population. PLoS ONE10(3): e0120365.

Zubov, A.A. (2006) Odontological analysis of the cranial series from the Ekven and Uelen Cemeteries. . In *Ancient cultures of the Asiatic Eskimos: the Uelen cemetery*, by S.A. Arutiunov and D.A. Sergeev, pp. 199-212. Translated and edited by Richard Bland from *Drevnie kul'tury aziatskikh eskimosov (Uzlenskii mogilnyk)*, originally published 1969. Shared Beringian Heritage Program, National Park Service.

1.4 Kolyma1

1.4.1 Site description

The Kolyma1 DNA sample (accession number DY-1sk, Institute for the History of Material Culture, St. Petersburg) was obtained from a partial cranium. The specimen was recovered from an erosional slope along the right bank of the lower Kolyma river, in arctic northeastern Siberia (68° 37' 48.1" N, 159° 08' 20.9" E), some 35-45 km downstream from the mouth of the Omolon River (Kaplina et al. 1978; Sher et al. 1979) (**Extended Data Figure 1.3.a, b)**. The locality where it was found is the well-known and extensively studied Duvannyy Yar section, a 12 km stretch of the river which has yielded rich faunal remains (Sher 1974; Yashina et al., 2012; Zimov et al. 2012; Lee et al. 2015), provides a thick exposure of the Yedoma Suite ice complex (Giterman et al. 1982), and serves as the stratotype for what Hopkins (1982) designated as the Duvannyy Yar Interval (Hopkins 1982; Sher et al. 1987).

Broadly speaking, the Duvannyy Yar stratigraphic sequence consists of four major units visible above the water line of the river (Giterman et al. 1982; Sher et al. 1987). The lowest, Unit I, is represented by lacustrine silts filling in ice-wedge pseudomorphs, indicative of deposition under a climatic regime cold enough for permafrost. The overlying deposits of Unit II (from the 'Boutellier Interval of Hopkins [1982]) are comprised of woody, peaty sediments likely representing thaw-lake debris (Giterman et al. 1982). Both Units I and II predate the LGM, and were deposited under interstadial conditions (Hopkins 1982). Unit III, the most massive portion of the stratigraphic sequence, marks the periglacial Duvannyy Yar Interval, when conditions in Beringia were extremely dry and cold. The Unit III deposits are principally ice-rich finely-bedded organic silt and loess, interbedded at the base with alluvial sand and fine gravel, the stratum marked throughout by ice wedges (Giterman et al. 1982; Hopkins 1982; Vasil'chuk et al. 2001; Zanina et al. 2011). The uppermost stratum in the section, Unit IV, was deposited during warmer post-LGM conditions, and is comprised of silts deposited under lacustrine and boggy conditions, and which in places fill erosional depressions that resulted from the partial thawing of ice-wedges and permafrost deposits (Giterman et al. 1982).

The Kolyma 1 human cranial fragment was collected from the slope of an exposure of Holocene Unit IV. Two radiocarbon ages are available on the specimen: 8,765 ± 35 (UCIAMS-14768) and 8,779 ± 41 (AAR-20354) radiocarbon years BP. These dates overlap statistically (as determined by chi-square test), and can be averaged, resulting in a mean age

estimate of 8,770 ± 27 radiocarbon years BP. The mean age, when calibrated (IntCal13), yields a 2 sigma age range of 9,668-9,904 cal BP, with a median probability of 9,769 cal BP.

Kolyma1 was an isolated specimen, and has no associated archaeological context.

1.4.2 Human remains

The Kolyma1 sample is a braincase with lacking the left temporal bone, preserving only part of the right temporal portion, and lacking the facial portion of the skull (**Fig. S1.7 c,d**). Although its partial preservation limits the number of measurements that can be obtained from the specimen, a sufficient portion remains to determine its length (195 mm) and width (134), along with a number of other measurements.

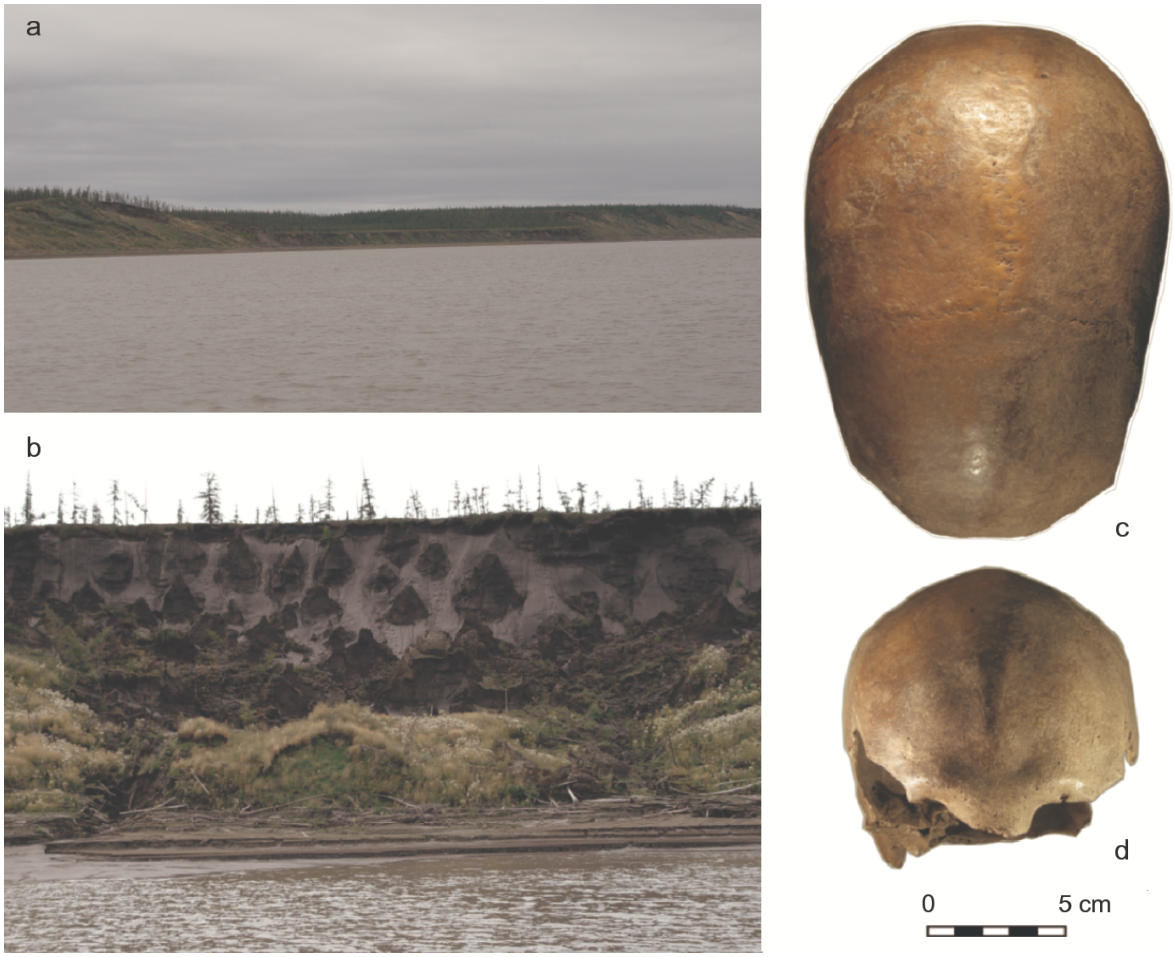


Figure S1.7. Duvanny Yar frozen bluff in low reaches of the Kolyma River, northeastern Siberia. **a**, general view from the West; **b**, central part of the exposure; **c** and **d**, human skull from Duvanny Yar frozen bluff, surface find (**c**, plan view; **d**, frontal view).

The mastoid process is rather robust (2 points according to Martin scale). Supraorbital ridge and muscular relief are moderately developed. While precise sex determination can hardly be done because of lack of diagnostic areas of the skull, it appears more likely that it belongs to female. The sutures on the interior surface of the braincase are completely fused, while the exterior traces of closed sutures are partly detectible only near lambda and bregma landmarks. According to pattern of suture closure, the individual is estimated to have been an adult, possibly as much as 50-55 years at time of death. Although the Kolyma1 skull is incomplete, the bone preservation itself was sufficient for successful DNA extraction.

References

Giterman, R.E., A.V. Sher, and J.V. Matthews (1982) Comparison of the development of tundra-steppe environments in west and east Beringia: Pollen and macrofossil evidence from key sections. In *Paleoecology of Beringia*, D.M. Hopkins, J.V. Matthews, C.E. Schweger and S.B. Young, editors, pp. 43-73. Academic Press: San Diego.

Hokins, D.M. (1982) Aspects of the paleogeography of Beringia during the Late Pleistocene. In *Paleoecology of Beringia*, D.M. Hopkins, J.V. Matthews, C.E. Schweger and S.B. Young, editors, pp. 3-28. Academic Press: San Diego.

Kaplina, T.N., Giterman, R.E., Lakhtina, O.V., Abrashov, B.A., Kiselyov, S.V., Sher, A.V., (1978) Duvannyi Yar - a key section of Upper Pleistocene deposits of the Kolyma Lowland. Bulletin of the Commission on Quaternary Period Research, USSR Academy of Sciences, 48, 49-65 (In Russian).

Lee, E.J., D.A. Merriwether, A.K. Kasparov, P.A. Nikolskiy, M.V. Sotnikova, E.Y. Pavlova, and V.V. Pitulko (2015) Ancient DNA Analysis of the Oldest Canid Species from the Siberian Arctic and Genetic Contribution to the Domestic Dog. *PLoS ONE* 10(5), e0125759. DOI:10.1371/journal.pone.0125759

Sher A.V. (1974) Pleistocene mammals and stratigraphy of the Far Northeast USSR and North America. *International Geology Review* 16 (7-10), 1974, 1-284.

Sher, A.V., T.N. Kaplina, R.E. Giterman, A.V. Lozhkin, A.A. Arkhangelov, S.V. Kiselyov, Y.V. Kouznetsov, E.I. Virina, V.S. Zazhigin (1979) Late Cenozoic of the Kolyma Lowland. XIV Pacific Science Congress, Tour Guide XI, Khabarovsk, August 1979. Moscow, Academy of Sciences of the USSR: XIV Pacific Science Congress, 1-116.

Sher, A.V., Kaplina, T.N., Ovander, M.G. (1987) Unified Regional Stratigraphic Chart for the Quaternary deposits in the Yana-Kolyma Lowland and its mountainous surroundings. Explanatory Note. In Shilo N.A. (Ed.) Resheniya Mezhvedomstvennogo stratigraficheskogo soveshchaniya po chetvertichnoy sisteme Vostoka SSSR (Magadan, 1982). Ob"yasnitel'nye zapiski k regional'nym stratigraficheskim skhemam [Resolution of Interagency Stratigraphic Conference on the Quaternary System of the Eastern regions of the USSR (Magadan, 1982). Explanatory notes for regional stratigraphic charts]. Magadan, SVKNII DVO AN SSSR, 29-69 (In Russian).

Vasilchuk, Y.K., A.C. Vasil'chuk, D. Rank, W. Kutschera, J.G. Kim (2001) Radiocarbon dating of δ18O-δD plots in late Pleistocene ice-wedges of the DuvannyYar (lower Kolyma River, northern Yakutia). *Radiocarbon* 43:541-553.

Yashina, S., S. Gubin, S. Maksimovich, A. Yashina, E. Gakhova, D. Gilichinsky (2012) Regeneration of whole fertile plants from 30,000-y-old fruit tissue buried in Siberian permafrost. *Proceedings of the National Academy of Sciences* 109, 4008-4013.

Zanina, O.G., S.V. Gubin, S.A. Kuzmina, S.V. Maximovich, D.A. Lopatina (2011) Late-Pleistocene (MIS 3–2) palaeoenvironments as recorded by sediments, palaeosols, and ground-squirrel nests at Duvanny Yar, Kolyma lowland, northeast Siberia. *Quaternary Science Reviews* 30:2107–2123.

Zimov S.A., N.S. Zimov, A.N. Tikhonov and F.S. Chapin (III) (2012) Mammoth steppe: a high-productivity phenomenon. *Quaternary Science Reviews* 57:26-45.

1.5 Magadan / Olskaya site

Pavel Grebenyuk, Alexander Fedorchenko, Alexander Lebedintsev, Boris Malyarchuk

1.5.1 Site description

The Magadan specimen is from the Olskaya site (59° 28′ 57″ N, 150° 57′ 29″ E) is one of the reference sites of the Tokarev archaeological culture in the northern Priokhotye. It is coast of the Sea of Okhotsk ~10 km south of Magadan on Olsky (Vostochny) Cape, perched on a 3 m high shore terrace. The site was discovered in 1982, and was excavated during the mid-1980s and again in 1999. The overall settlement is estimated to have been ~9000 m^2.

A thorough inspection of the shore terrace allowed researchers to identify 26 circular dwelling depressions 5 to 9 m in diameter and 40–70 cm deep. Of these, one dwelling was excavated reported as having been circular-shaped, 0.6 m deep and 6 m in diameter. Its cultural layer was marked by a stratum of brown loam ~80 cm thick, which yielded stone and bone artifacts, and fragments of animal bones and mollusk shells. A rectangular hearth 100 *x* 120 cm in size was recovered; it was lined with vertically set stones, and filled with yellowish compacted ash (14-17 cm thick) interleaved with ~2cm thick layers of carbonaceous sediment (Lebedintsev 1990, 1999).

Through long-term studies of the Tokarev settlements on Olsky Cape, nine radiocarbon dates were obtained ranging from 2480 ± 20 to 1800 ± 60 radiocarbon years BP. Three age determinations were obtained by virtue of conventional method on charcoal (2380 ± 30 and 1830 ± 50 radiocarbon years BP) and on a processed wood fragment. Five AMS-based determinations were obtained on charcoal: 2250 ± 60, 2160 ± 60, 2110 ± 60, 1950 ± 60 and 1800 ± 60 radiocarbon years BP. One AMS date was obtained on carbonaceous organic material preserved on the outer surface of a ceramic vessel from Olskaya site: 2480 ± 20 radiocarbon years BP. Maritime sites in northern Priokhotye yield radiocarbon dates on organic material of ceramics older than those on bone fragments and charcoal. According to earlier study results, carbon from the Tokarev ceramics is partly of marine origin. As a result, the observed difference in radiocarbon determinations may be directly due to the reservoir effect (Takase et al. 2012). In terms of the calendar chronology, the time-span of the settlement on Olsky Cape ranges within 2443 – 1629 years BP (485 BC – 322 AD).

The economic life of ancient inhabitants of the Olskaya site was based on an active marine resource procurement including sea hunting and mollusk shells gathering, also hunting of reindeer and birds. The stone artifact inventory is comprised of spear heads and javelins (leaf-

shaped, truncated leaf-shaped and stem-shaped), arrow points (triangular and subtriangular, leaf-shaped, truncated leaf-shaped and leaf-stemmed), knives (leaf-shaped, truncated leaf-shaped, rounded, stemmed, polished), also oval, subtriangular or tanged scrapers, adzes and subrectangular and almond-shaped abrasion-processed axes, borers, burins, retouched flakes, hammer stones, abrasive tools, decoration items and personal ornaments.

The bone industry is represented by points and heads of spears and arrows, leisters and harpoons including toggling type, adzes, hoes, shovels, fishhooks, combs, awls, needles, borers and needle cases. The handles of burins or engraving tools are of a peculiar interest as they have slots for iron inserts, one of which is ornamented with scratched parallel inclined lines. The bone tools from Olskaya site reveal their close similarity with those reported from ancient sites located on Zavyalova and Spafaryeva islands (Lebedintsev 1999). The Olskaya ceramic industry is represented by fragments of a vessel having a false-textile imprint, its rim ornamented with circular pressed impressions, slant comb stamps over a sharp edge and inclined comb stamps below it.

1.5.2 Human remains

In 1983, the upper part of a cranium was found in the cultural layer, and in 1999 a left femur and a rib. The cranium was given a preliminary study by Yu. K. Chistov and V.I. Khartanovich. They determined that it belonged to a female of about 30–35 years age. The rib was assigned to an adult man, who had a wide flat chest. The cranium was transferred to the Peter the Great Museum of Anthropology and Ethnography in Saint-Petersburg in 1990 for its anthropologic examination, and was unavailable for genomic analysis for this project.

The Magadan1 DNA sample (M9984) was obtained from a part of femur. The left femur was found in 1999 at 70 cm below the modern surface. It was examined by G.S. Beloborodov and E.E. Shubert. After taking detailed measurements and making a careful examination of its morphology, they determined the specimen came from a 35-40 year old female with well-developed muscles who stood about 151-154 cm in height (Lebedintsev 2001). The specimen dates to 2920 ± 49 radiocarbon years BP (UBA-34732). At 1 sigma the calibrated age has multiple intercepts of 2995-3084 (60.8%), 3088-3145 (36.5%), and 3150-3156 (2.7%), with a median probability of 3066 cal BP. Given the heavy marine diet, however, were the local reservoir effect to be determined the calibrated age would likely be younger, perhaps by a

few centuries. The Magadan2 DNA sample (M0831) was obtained from a part of radius bone. The radius bone of 22,6 cm in length was found in the cultural layer in the shell midden in 2000. The bone preservation was sufficient for successful DNA extraction.

The anatomy, morphology and size of bones from Olskaya are the oldest ones found to date in the northern Priokhotye. Human bones found in the cultural layer of the Olskaya site can testify to the absence of cremation practice in the Tokarev archeological culture unlike the traditions of ancient Koryaks for whom it was typical. One scenario is that the Tokarev people disposed of the bodies of their dead within their middens, perhaps in a practice similar to that reported from the Neolithic of Kamchatka and the Okhotian Culture in Sakhalin and Hokkaido.

References

Lebedintsev, A. I. (1990) Early Maritime Cultures of North-Western Priokhot'e [In Russian].

Lebedintsev, A. I. (1999) in History, archeology and ethnography of the North-East of Russia, A. I. Lebedintsev, Ed. [In Russian] (NEISRI FEB RAS Press), pp. 77–96.

Lebedintsev, A. I., Beloborodov, G. S., Schubert, E. E. (2001) in Dikov Readings, A. I. Lebedintsev, Ed. [In Russian] (NEISRI FEB RAS Press), pp. 99–105.

Reimer, P. J. et al. (2013) IntCal13 and marine13 radiocarbon age calibration curves 0–50,000 years cal bp. Radiocarbon 55, 1869–1887.

Takase, K., Lebedintsev, A. I, Ptashinsky, A. V. (2012) in VII Dikov Readings, A. I. Lebedintsev, Ed. [In Russian] (NEISRI FEB RAS Press), pp. 139–143.

1.6 Ust-Belaya

1.6.1 Site description

Ust-Belsky II (or Ust-Belaya II, or Shumilikha) burial ground was discovered at Shumilikha locality on the left bank of the river Angara in the entry of the river Belaya. It was studied during 1972-1973, under the guidance of G. N. Mikhnyuk and I.L. Lezhnenko, and after O.I. Goryunova (see ”The Bronze Age of the Baikal region”, 1981. In Russian). The works were emergency and rescue, since the terrain did not exceed 1.5 m above the modern water level and was constantly flooded. 46 burials were discovered. It is possible that this was already a peripheral section of a large burial ground, the length of which was about 1 km along the coastline. More than 10 burials were found washed away on the foreshore. This site is a very interesting phenomenon. In most of the burials deceased were found in sited position, with legs bent at the knees, hands on the chest and head dropped on the chest or knees, in shaft-like cuts. Only six burials with deceased in sited position in the Cisbaikalia were published before this discovery (Okladnikov, 1974, 1975). The main set of burial goods of male burials of the II Ust-Belsky burial ground consists of bone harpoons, arrowheads and spearheads made of bone and stone, jade axes and adzes, spoons made of horn, grindstones, and also remarkable examples of fine art: zoomorphic and anthropomorphic images. Female burials contained bone needles, awls, and scrapers. Both male and female burials were accompanied by jewellery made of maral teeth, canines of wild boar, and rings of white jade. The bronze socketed axe was found among the burial goods of deceased found in sited position for the first time in Cisbaikalia. Its morphological features are similar to that of the finds from the Seymin-Turbin sites, dated back to the 9th-8th centuries BC. Besides burials with deceased in sited position, different burials were found: with deceased laying on the back, elongated, on the back with partial cremation, on the back with a full or partial backfilling with ochre, on the back wrapped with a birch bark, secondary burials with dismemberment and partial cremation. The authors of the study of Ust-Belsky II burial ground (Goryunova, Smotrova, 1981, p. 28) believe that in the burial goods of the burials with deceased in sited position genetic links with the Glazkovo complexes of the Cisbaikalia Eneolithic-Early Bronze Age are traced. The authors explain the appearance of plots of fine arts and socketed axe of the Karasuk type, widespread in Western Siberia, by the outside influence, rather than by the arrival of a new population. An analysis of scarce bronze items, including socketed axe, showed their local production on the basis of native copper in copper sandstones of the Irkutsk amphitheatre (Sergeeva, 1981. P.29).

M.M. Gerasimova carried out the study of the human remains (Gerasimova, 1981. P.32-37). The morphology of the 13 men's, 8 women's, and 5 child's skulls from the burials with deceased in sited position of Ust-Belsky (Shumilikha) burial ground was analysed. This group of skulls exhibited the following distinctive features: significantly flattered face (no less than that of the Naukan Eskimos, the Nivkhs, Ulchi, Tunginsky Buryats, Yakuts), face height is less than that of many Mongoloid groups, with the exception of Orochi, Selkup and Mansi. Since the initial date of the burial ground was determined as Glazkovo, the search for the place of studied group was carried out among the Serov, Kitoy and Glazkovo skull series and series from individual burial grounds - Verkholensky (Lena River), Ulyarba II (lake Baikal), Fofanovo (Selenga River). The comparison based on skull morphology showed a very distant similarity of the series from Shumilikha with all these series, especially with a series from the Ulyarba II burial ground, the closest one in chronology and culture. Since the Shumilikha series was characterized by increased variability of the metric traits and the presence of morphologically opposite variants, it was concluded that, in addition to the West Siberian "impulse", infiltration of small human groups from various regions of Transbaikalia and Mongolia continued, which led to a mosaic combination of the anthropological features in the humans of the Cisbaikalia in the Bronze Age.

1.6.2 Human remains

Genomic data of a total of six individuals from the site were analyzed in this study, directly dated to three different time periods.

Neolithic

NEO229, 5779 ± 60 radiocarbon years BP (UBA-33759).

Bronze Age

NEO298, 4060 ± 38 radiocarbon years BP (UBA-33769)

NEO231, 4160 ± 37 radiocarbon years BP (UBA-33761).

NEO232, 4259 ± 38 radiocarbon years BP (UBA-33762).

NEO230, 4298 ± 39 radiocarbon years BP (UBA-33760).

Medieval

NEO299, 632 ± 29 radiocarbon years BP (UBA-33770).

1.7 Young Yana

Vladimir Pitulko

1.7.1 Site description

The Young Yana sample, a human fibula that dates to the late Holocene, was discovered in 2015 at the Yana Mass Accumulation of Mammoth (YMAM) locality (Pitulko et al. 2014). YMAM, one of the several localities that comprise the Yana RHS Paleolithic site, was discovered in 2008. Unusually low levels of the Yana River allowed miners in search of ivory using high pressure water jets to bore out a tunnel (**Fig. S1.8 a,b**). The tunnel extended several tens of meters into the hillside, where it broke into the mammoth bone accumulation (Basilyan 2011). Hydraulic mining of the deposit continued until 2016, when the mammoth bone-bearing unit was finally exhausted. Over that eight-year period, YMAM yielded many hundreds of disarticulated mammoth bones, along with a small collection of artifacts (stone tools, and lithic and ivory flakes).

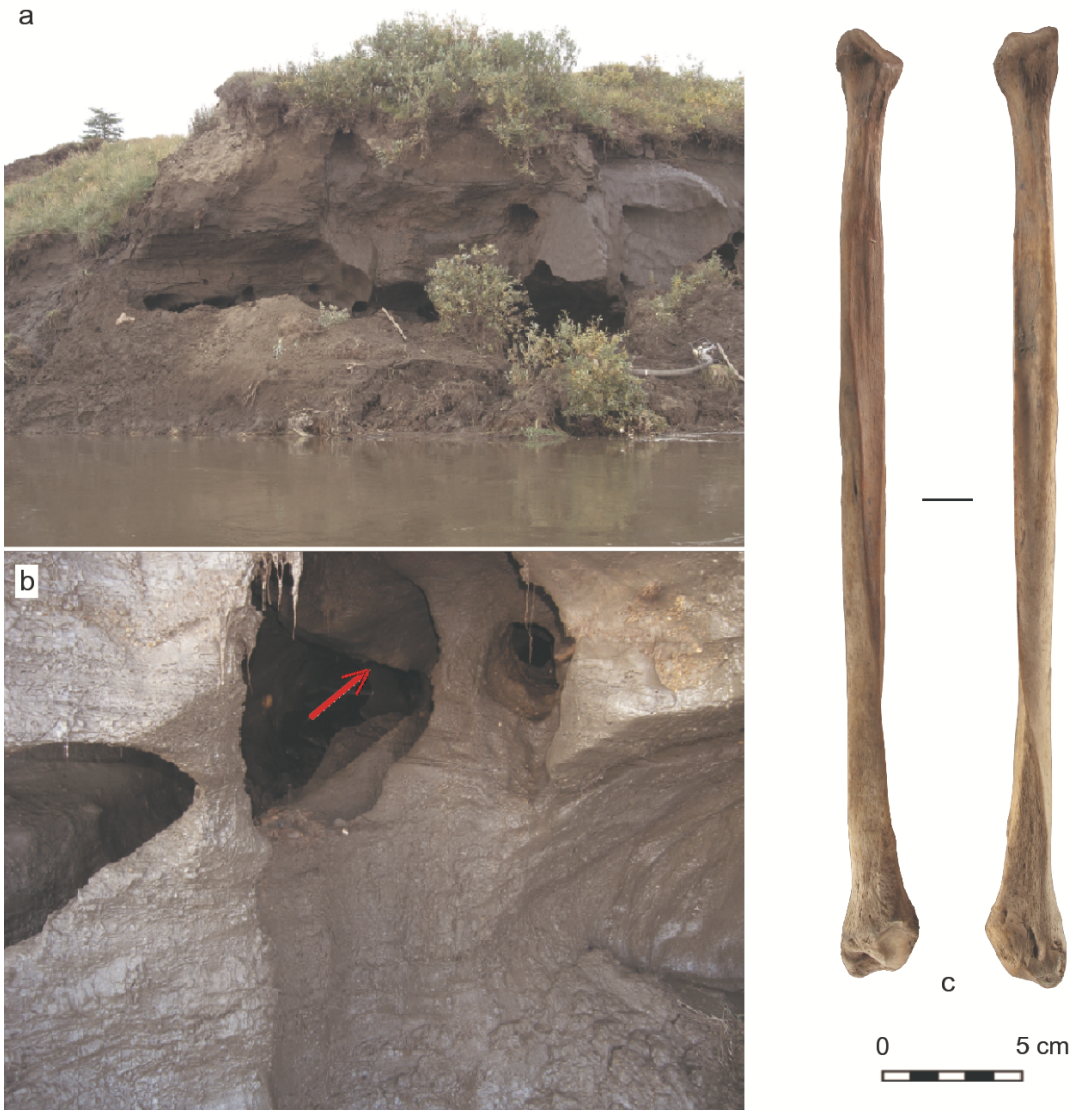


Figure S1.8. Yana River low reaches (northern Siberia). Yana RHS site, Yana A locality where fossil human bone was collected, sampled for genetics research and then directly dated and sequenced (for this study, it is termed 'Young Yana'). **a**, general view of the eroded river bank next to the Yana mass accumulation of mammoth (Basilyan et al. 2011); **b**, holes washed in the frozen deposits by local residents mining for the mammoth ivory, in the roof of one of such holes the human bone was found; **c**, human fibula (young Yana sample in this study).

The mammoth remains and artifacts were deposited in a now-deeply buried paleochannel. Based on extensive radiocarbon dating, the YMAM deposits are contemporaneous with the cultural layer seen at the other Yana RHS localities, and together these mark multiple occupations of the area between ~29,000-27,000 radiocarbon years BP, with perhaps sporadic occupations over the millennium that followed (Pitulko et al. 2014: Table 2.2).

Subsequent to the deposition of the mammoth remains within the paleochannel (designated Member 3 of Terrace II at the site [Basilyan et al. 2011; Pitulko et al. 2014]), it was buried by alluvium and later sub-aerial aeolian deposits. Ultimately, the paleochannel was overlain by 8-11 m of permafrost deposits (Basilyan et al. 2011). The overlying deposits are marked by ice-wedges, and experienced multiple cut-and-fill episodes, which developed as terrace surface drainages evolved, shifted, and were pirated over time.

These cut-and-fill episodes began ~15,000 radiocarbon years ago, and continued into the late Pleistocene and Holocene. The channels that formed were typically filled by silt and peat-like deposits, and contain archaeological material of different ages. Over the course of the continued hydraulic mining at the locality, along with erosion by the Yana River, the bank of the YMAM locality was stepped back ~40m. In the process, more artifacts and bones appeared, and because of the complex history of deposition and re-deposition, their ages vary. The youngest of the Upper Paleolithic archaeological material known from this locality – mammoth ivory artifacts – date to roughly 18,000 radiocarbon years ago (Pitulko et al. 2014: Table 2.2). More recent material was recovered as well, including a large canid mandible dated to 884 ± 25 radiocarbon years BP (AAR-21020), and the human fibula designated here as 'Young Yana.'

1.7.2 Human remains

A group of local residents provided the Young Yana sample in 2015, which they reportedly collected from the ceiling of an artificial channel that was washed into the frozen deposits of the river bank (**Fig. S1.8 c,d**). That channel intersected the bottom of an older natural erosion channel, which was filled with recent secondary deposits that included twigs, grass, and pieces of driftwood, together with silt and allochthonous peat.

The Young Yana sample was embedded in these deposits. The specimen is a left fibula, with a maximum length of 331mm, a central diameter that ranges from a minimum of 11 mm to a maximum of 14.7 mm, and central circumference of 41 mm. It appears to have been derived from adult, perhaps 20-35 years of age, sex unknown (if the individual was male, he would have been ~160cm in height; if female ~157cm in height [Trotter and Gleser 1958]). The Young Yana specimen is directly dated to 862 ± 26 radiocarbon years BP (OxA-32884)

References

Basilyan, A.E., M.A. Anisimov, P.A. Nikolskiy, and V.V. Pitulko (2011) Woolly mammoth mass accumulation next to the Palaeolithic Yana RHS site, Arctic Siberia: Its geology, age, and relation to past human activity. *Journal of Archaeological Science* 38:2461–74.

Pitulko V., P. Nikolskiy, A. Basilyan, E. Pavlova. Human habitation in the Arctic Western Beringia prior the LGM. In K.E. Graf, C.V. Ketron, M.R. Waters (eds). *Paleoamerican Odyssey*. CSFA, Dept. of Anthropology, Texas A&M University, 2014, pp. 13–44.

Trotter M., Gleser G.C (1958) A re-evaluation of estimation of stature based on measurements of stature taken during life and of long bones after death. *American Journal of Physical Anthropology* 16:79-123.

1.8 Levanlühta

1.8.1. Archaeological context of Levanlühta

Anna Wessmann, Kristiina Mannermaa, Peter de Barros Damgaard

The Levänluhta wetland site located in southwestern Finland is one of the largest Scandinavian Iron Age burial sites with uncremated remains. The excavated findings at Levänluhta consist of unburnt commingled human remains from 98 individuals buried along with artefacts and animal bones. At the time of burials (400-800 CE) this spring site was a small lake, and it has been archaeologically investigated periodically since the 1800s (Wessman 2009). These excavations have yielded an impressive range of finds, including precious copper alloy brooches, arm rings and other dress implements that suggest that most of the buried individuals were women (Wessman 2009). Osteological studies have confirmed that it was a cemetery for women and children (Niskanen 2006). Past interpretations regarding the Levänluhta site have ranged from it being a place for human sacrifice (Hackman 1913) to a mass grave for slaves or those who had perished from famine, plague or war (Meinander 1946; 1950; Leppäaho 1949, Seger 1982, Niskanen 2006), but the nature of these burials remain clouded in speculations.

We generated genomic sequences (0.1 – 2.2 X genomic coverage from four human skeletons excavated from Levänluhta and investigated their genetic ancestry in relation to past and present genetic diversity in Fennoscandia and Eurasia. We radiocarbon-dated the four skeletons, confirming their age to c. 450 CE (Extended Data Table 1). Lastly, we conducted strontium ratio analyses to contextualize the genetic findings with possible mobility patterns of the individuals.

References

Leppäaho, Jorma. ”Kalevala vertailevan muinaistieteen valaisemana”. In: F. A. Heporauta, and M. Haavio, eds. Kalevala kansallinen aarre. Helsinki, Finland: WSOY (1994): 49–81.

Meinander, Carl-Fredrik. ”Förutsättningar för den förhistoriska bebyggelsen i södra Österbotten”. Nordenskiöld-samfundets tidskrift (1946): 70–101.

Meinander, Carl-Fredrik. ”Etelä-Pohjanmaan Esihistoria”. Etelä-Pohjanmaan Historia I-II, Helsinki (1950).

Niskanen, Markku. “Stature of the Merovingian-period inhabitants from Levänluhta, Finland”. Fennoscandia Archaeologica 23 (2006): 24–36.
Seger, Tapani. “The plague of Justinian and other scourges”. Fornvännen77 (1982):184–198.
Wessman, Anna. “Levänluhta – a place of punishment, sacrifice or just a common cemetery?” Fennoscandia Archaeologica 26 (2009): 47–71.

1.8.2 Strontium isotope analysis of the Levänluhta individuals

Laura Arppe

Material and Methods

To explore the local $^{87}Sr/^{86}Sr$ baseline values, we measured the $^{87}Sr/^{86}Sr$ ratio of water samples and skeletal parts of modern rodents captured near the site. The water sample taken from the nearby Kyrönjoki river is likely to represent a wider regional average of baseline $^{87}Sr/^{86}Sr$ values, whereas the samples (n=2) from the Levänluhta springs reflect more localized values describing the immediate surroundings of the burial site. While there is a possibility that the rodents’ diets may have been affected by Sr contamination from fertilizers, several studies use modern small mammals (e.g. Hoppe et al., 1999; Beard and Johnson, 2000; Price et al. 2011).

Samples for $^{87}Sr/^{86}Sr$ analysis were taken from tooth enamel of four Levänluhta individuals. Prior to dissolution, the samples of human dental enamel were pretreated with 5% CH3COOH. In addition to acetic acid leaching, the samples of modern rodents were additionally treated with a sodium hypochlorite solution to eliminate organic residues. The water samples were filtered upon collection and concentrated HNO3 added on site. Before analysis, they were evaporated to dryness and taken up in HNO3 for Sr separation. Sr was chromatographically extracted and purified using micro-columns filled with 200 µL of Sr-specific resin (TrisKem Sr Resin 50-100 µm). $^{87}Sr/^{86}Sr$ ratios were measured at the Geological Survey of Finland Espoo, on a VG SECTOR 54 thermal ionization mass spectrometer in dynamic mode. The international SRM978 standard gave a value of 0.710247 during the course of measurements. Long-term repeat measurements of the standard indicate a precision of 0.002% (2SE). Process blanks were below 1 ng.

Results

The local baseline values vary between 0.7238 and 0.7384. The Levänluhta springs show $^{87}Sr/^{86}Sr$ levels of 0.7238-0.7239. These values, which can be expected to approximate the isotopic composition of diagenetic Sr at the site of burial, are somewhat lower than the Sr isotope ratio of river Kyrönjoki (0.7288) and the rodent specimens (0.7279-0.7384). The human dental enamels gave $^{87}Sr/^{86}Sr$ values ranging from 0.7107 to 0.7305 (Table S1.5).

Table S1.5. 87Sr/86Sr values of the Levänluhta individuals, local modern rodents and environmental waters.

Sample ID	**Sample type**	**Element**	**87Sr/86Sr**	**± 2SE**
JK1968/DA234	human dental enamel	Left lower M1	0.720686	0.000016
JK2065/DA236	human dental enamel	Right lower M3	0.710782	0.000014
JK1963/DA238	human dental enamel	Right lower M3	0.730489	0.000015
JK2067/DA237	human dental enamel	Left lower M3	0.717454	0.000014
Baseline samples:				
19-m-Sr	rodent (*Clethrionomys*)	teeth and jaw	0.732432	0.000019
20-m-Sr	rodent (*Clethrionomys*)	teeth and jaw	0.727940	0.000016
21-m-Sr	rodent (*Clethrionomys*)	teeth and jaw	0.730113	0.000016
22-m-Sr	rodent (*Clethrionomys*)	teeth and jaw	0.729584	0.000015
23-p-Sr	rodent (*Sorex*)	teeth and jaw	0.731776	0.000016
24-p-Sr	rodent (*Sorex*)	teeth and jaw	0.738395	0.000016
LL-spring1	Levänluhta spring 1 water		0.723785	0.000017
LL-spring2	Levänluhta spring 2 water		0.723865	0.000017
LL-Kyrönjoki	river Kyrönjoki water		0.728770	0.000016

Discussion

Reliability of archeological Sr-data

Enamel $^{87}Sr/^{86}Sr$ values have been shown to be highly resistant to alteration (Chiaradia et al. 2003; Budd et al. 2000; Hoppe et al. 2003; Sponheimer and Lee-Thorp, 2006; Bocherens et al., 1994; Kohn et al., 1999), and weak acid leaching is effective in removing possible traces of diagenetic Sr from enamel (Budd et al., 2000; Hoppe et al., 2003; Sponheimer and Lee-Thorp, 2006). Ancient human populations tend to show surprising homogenous Sr-isotopic signatures (1σ stdev 0.00006-0.00080; compiled by Bentley, 2006), and diagenesis tends to further reduce the standard deviation of the sample. In the case of Levänluhta, most samples show $^{87}Sr/^{86}Sr$ values distinct from the burial environment, and a large standard deviation

(0.0082, 1σ), suggesting that they have retained their individual original values. It can thus be concluded that the analysed $^{87}Sr/^{86}Sr$ values are representative of the in vivo signatures.

Locals or non-locals?

One of the analysed individuals, JK1963/DA238 displayed a $^{87}Sr/^{86}Sr$ value (0.73049) consistent with the local baseline as determined by the $^{87}Sr/^{86}Sr$ values of the water samples and rodent specimens. Thus this individual likely spent their adolescence in the region. For the three remaining individuals with $^{87}Sr/^{86}Sr$ values between 0.71078 and 0.72069, outside the range of the local baseline, three possible scenarios emerge.

In *Scenario 1,* they spent their childhood (JK1968/DA234) or adolescence (JK2065/DA236 and JK2067/DA237) in areas where lower bedrock (and bioavailable) $^{87}Sr/^{86}Sr$ values are expected, such as the Baltic countries, northern Poland and Germany, southern Sweden, and Denmark, all underlain predominantly by Phanerozoic sedimentary rocks. For example, rivers draining the southeasternmost part of Sweden show Sr isotope compositions ranging from 0.711 to 0.720 (Löfvendahl et al 1990), and baseline $^{87}Sr/^{86}Sr$ values from 0.707 to 0.712 have been reported for Denmark (Frei and Frei, 2011; Frei and Price, 2012), from 0.709 to 0.720 for Northern Estonia and Saaremaa island (Oras et al. 2016; Price et al. 2016), and from 0.707 to 0.714 for Poland (Voerkelius et al. 2010; Buko et al., 2013; Gregoricka et al 2014).

In *Scenario 2*, they were derived from Finnish terrains further afield. Relatively high $^{87}Sr/^{86}Sr$ ratios common for the Precambrian rocks making up the bulk of the Fennoscandian Shield are expected throughout Finland, and indeed have been observed in Finnish river waters (0.72226-0.74515; Löfvendahl et al. 1991) and the most prevalent rock types (Kaislaniemi, 2011). However, according to a modeling study by Kaislaniemi (2011) larger areas of bedrock $^{87}Sr/^{86}Sr$ values from 0.715 to 0.721 can be expected a few tens of kilometers to the southeast from Levänluhta. Values below 0.715 occur in small patches some tens of kilometers from the site, while more extensive areas are located 200-250 km to the east. In the absence of bioavailable baseline data confirming these hypothesized occurrences of lower $^{87}Sr/^{86}Sr$ values, it is difficult to estimate the validity of this scenario.

In *Scenario 3*, these individuals incorporated a varying degree of marine, or – in the case of the Baltic Sea – brackish water dietary resources. The coast of the Bothnian Sea (=the northern part of the Baltic Sea, between Sweden and Finland) lay only 25-30 km to the northwest, and accessible to the Iron Age people of the Levänluhta region via the Kyrönjoki river. The Sr-isotope composition of the present-day Bothnian Sea is ca. 0.7094-0-7096 (Löfvendahl et al. 1990; Andersson et al., 1992; Widerlund and Andersson, 2006), and

possibly somewhat lower during the Iron Age based on the trend of Late Holocene freshening of the Baltic Sea (Widerlund and Andersson, 2011). A steady incorporation of Bothnian Sea resources would result in lower consumer $^{87}Sr/^{86}Sr$ ratios compared to locals relying on terrestrial dietary items only.

When combined with the stable isotope ($\delta^{13}C$, $\delta^{15}N$) information in Supplementary Section 3, suggesting a clear influence of "brackish" dietary items for the same three individuals with lower, "non-local" $^{87}Sr/^{86}Sr$ values we find that Scenario 2 might very well be plausible in their case. This holds true especially for individuals JK1968/DA234 and JK2067/DA237, with $^{87}Sr/^{86}Sr$ values relatively closer to the local baseline suggesting mixed use of local terrestrial and Baltic dietary resources. For individual JA2065/DA236, the low $^{87}Sr/^{86}Sr$ value (0.71078) would imply an exceptionally heavy reliance on Baltic Sea resources. The $\delta^{13}C$ and $\delta^{15}N$ values of the individual are near comparable (especially considering within-Baltic latitudinal gradients in $\delta^{13}C$; Torniainen et al. 2017) to the $\delta^{13}C$ and $\delta^{15}N$ values of a Middle Neolithic population on the Baltic island of Gotland (Eriksson, 2004) interpreted to have subsisted primarily on seals.

In summary, the $^{87}Sr/^{86}Sr$ data suggest, that individual JK1963/DA238 was probably a local relying on terrestrial foodstuffs, while individuals JK1968/DA234 and JK2067/DA237 might well have been locals incorporating a component of Baltic sea resources in their diet. Individual JA2065/DA236 could have been of foreign origin, or a local using almost exclusively Baltic Sea dietary items.

References

Andersson, P.S., Wasserburg, G.J., Ingri, J., 1992. The sources and transport of Sr and Nd isotopes in the Baltic Sea. Earth and Planetary Science Letters 113, 459–472.

Beard, B. L., Johnson C. M., 2000. Strontium isotope composition of skeletal material can determine the birth place and geographic mobility of humans and animals. Journal of Forensic Sciences 45, 1049–1061.

Bentley, R.A. 2006. Strontium isotopes from the Earth to the archaeological skeleton: a review. Journal of Archaeological Method and Theory 13, 135-187.

Bocherens, H., Brinkman,D.B., Dauphin,Y., Mariotti, A., 1994.Microstructural and geochemical investigations on Late Cretaceous archaeosaur teeth from Alberta, Canada. Canadian Journal of Earth Science 31, 783–792.

Budd P., Montgomery J., Barreiro B., Thomas R.G., 2000. Differential diagenesis of strontium in archaeological human dental tissues. Applied Geochemistry 15, 687-694.

Buko, A., Kara, M., Price, T.D., Duczko, W., Frei, K.M., Sobkowiak-Tabaka, I. 2013) A unique medieval cemetery from the 10th/11th century with chamber-like graves from Bodzia

(central Poland): Preliminary result of the multidisciplinary research. Archäologisches Korrespondenzblatt 43, 423–441.

Chiaradia, M., Gallay, A., Todt, W., 2003. Differential lead and strontium contamination styles of prehistoric human teeth at a Swiss necropolis (Sion, Valais). Applied Geochemistry 18, 353–370.

Frei, K.M. and Frei, R., 2011. The geographic distribution of strontium isotopes in Danish surface waters – a base for provenance studies in archaeology, hydrology and agriculture. Applied geochemistry 26, 326–340.

Frei, K.M. and Price, T.D., 2012. Strontium isotopes and human mobility in prehistoric Denmark. Anthropological and archaeological sciences 4, 103–114

Gregoricka, L. A., Betsinger, T. K., Scott, A. B. & Polcyn, M., 2014. Apotropaic practices and the undead: a biogeochemical assessment of deviant burials in post-medieval Poland. PLoS ONE 9, 1–24.

Hoppe, K., Koch, P.L., Carlson, R.W., Webb, S.D., 1999. Tracking mammoths and mastodons: reconstruction of migratory behavior using strontium isotope ratios. Geology 27, 439-442.

Hoppe K.A., Koch, P.L., Furutani, T.T, 2003. Assessing the preservation of biogenic strontium in fossil bones and tooth enamel. International Journal of Osteoarchaeology 13, 20-28.

Kaislaniemi, L., 2011. Estimating the distribution of strontium isotope ratios (87Sr/86Sr) in the Precambrian of Finland. Bulletin of the Geological Society of Finland 83, 95-113.

Kohn, M.J., Schoeninger, M.J, Barker, W.W., 1999. Altered states: Effects of diagenesis on fossil tooth chemistry. Geochimica et Cosmochimica Acta 63, 2737–2747.

Löfvendahl, R., Åberg, G., Hamilton, P.J., 1990. Strontium in rivers of the Baltic Basin. Aquatic Sciences 52, 315-329.

Oras, E., Lang, V., Rannamäe, E., Varul, L., KOnsa, M., Limbo-Simovart, J., Vedru, G., Laneman, M., Malve, M., Price, T.D., 2016. Tracing prehistoric migration: isotope analysis of Bronze and Pre-Roman Iron Age coastal burials in Estonia. Estonian Journal of Archaeology 20, 3–32.

Price, D.T., Frei, K.M., Dobat, A.S., Lynnerup, N., Bennike, P., 2011. Who was in Harold Bluetooth's army? Strontium isotope investigation of the cemetary at the Viking Age fortress at Trelleborg, Denmark. Antiquity 85, 476-489.

Price, D.T., Peets, J., Allmäe, R., Maldre, L., Oras, E., 2016. Isotopic provenancing of the Salme ship burials in Pre-Viking Age Estonia. Antiquity 90, 1022-1037.

Sponheimer, M., and Lee-Thorp, J.A., 2006. Enamel diagenesis at South Africal Australopith sites: implications for paleoecological reconstruction with trace elements. Geochimica et Cosmochimica Acta 70, 1644-1654.

Torniainen, J., Lensu, A., Vuorinen, P.J., Sonninen, E., Keinänen, M., Jones, R.I., Patterson, W.P., Kiljunen M., 2017. Oxygen and carbon isoscapes for the Baltic Sea: testing their applicabilty in fish migration studies. Ecology and Evolution 7, 2255-2267.

Voerkelius S., Lorenz G. D., Rummel, S., Quétel, C., Heiss G, Baxter, M., Brach-Papa, C., Deters-Itzelsberger, P., Hoelzl, S., Hoogewerff, J., Ponzevera, E., Van Bocxstaele, M., Ueckermann, H., 2010. Strontium isotope signatures of natural mineral waters, the reference

to a simple geological map and its potential for authentication of food. Food chemistry 118, 933-940.

Widerlund, A., Andersson, P.S., 2006. Strontium isotopic composition of modern and Holocene mollusc shells as a palaeosalinity indicator for the Baltic Sea. Chemical Geology 232, 54-66.

Widerlund, A., Andersson, P.S., 2011. Late Holocene freshening of the Baltic Sea derived from high-resolution strontium isotope analyses of mollusk shells. Geology 39, 187-190.

1.8.3 Radiocarbon datings

Karl-Göran Sjögren

Four human tooth samples from Levänluhta were dated at the Chrono Centre, Queens University, Belfast. Collagen extraction and other laboratory methods used at the Chrono Centre are described in detail in Hoper et al (2015).

Details of the datings are given in Extended Table 1. 14C values were calibrated to 2 sigma intervals at the Belfast laboratory using the Calib software, rev 7.0.0, and the intcal13 calibration curve. δ13C and δ15N were measured on all samples, as well as C/N ratio. C/N for all samples was between the accepted standard for good collagen quality, i.e. between 2.9 and 3.6.

The calibrated values in Extended Table 1 do not take account of possible reservoir effects. However, three of the samples (JK1968, JK2065 and JK 2067) have elevated δ13C values indicating consumption of marine proteins. This suggests that a marine reservoir effect (MRE) likely affects these dates. The size of the MRE in the Baltic is highly variable both in time and space. For the coastal region in the vicinity of Levänluhta, modern MRE in the order of 100-200 years have been estimated (Lougheed et al 2013). It is therefore reasonable to assume that the calibrated ages of these are somewhat too old and should be reduced by a fraction of the maximum MRE. Due to lack of isotopic baselines for the time and region, this fraction is difficult to estimate, and we therefore do not attempt a detailed calibration taking MRE into account.

Another situation applies to the sample JK1963. Here, the δ13C value indicates a terrestrial diet, so no MRE should apply. However, a rather high δ15N value suggest a possible intake of protein from freshwater fish. This could imply a possible freshwater reservoir effect.

Again, as this problem has not been studied in the area and no local isotopic baselines are available, it is not possible to model this at present.

References

Hoper, S T, McDonald, J, Reimer, P J, Reimer, R, Svyatko, S, Thompson, M, 2015. The Queen's University, Belfast: Laboratory protocols used for AMS radiocarbon dating at the 14CHRONO Centre. Report number 5/2015. English Heritage.

Lougheed, B. C., H. L. Filipsson, I. Snowball, 2013. Large spatial variations in coastal 14C reservoir age – a case study from the Baltic Sea. Climate of the Past 2013/9, 1015–1028.

Supplementary Information 2 – Laboratory procedures and sequencing

aDNA extraction

aDNA work was conducted in dedicated clean-room facilities at Centre for GeoGenetics, Natural History Museum, University of Copenhagen following strict aDNA standards[1,2]. Human teeth, tooth enamel, femur (and petrous bone) samples were used to extract DNA from ancient individuals (Supplementary Data Table 1).

The whole tooth crowns were used to extract DNA from Yana2 and Yana_young specimens. For the Yana1 sample the ancient DNA extraction was conducted from tooth enamel. In case of other teeth samples the endogenous DNA rich outer cementum layer of the tooth roots were used[3]. To increase the yield of endogenous DNA we targeted the otic capsule when extracting DNA from the petrous bones[4].

Drilled bone material was briefly digested (predigestion step) by incubating in digestion solution (4.65 mL 0.5M EDTA, 50 mL recombinant Proteinase K, 50 mL 100x TE and 250 mL 10% N-Laurylsarcosyl) for 45 min at 40º C[3]. After this step the samples were centrifuged, the supernatant was removed and an identical digestion buffer was applied for a full overnight digestion at 40º C. The DNA isolation for most of the samples (except M9984/Kolyma 8467a) was conducted from 2 mL digested solution using a silica-powder-based extraction method. The silica suspension was prepared by mixing 6g of SiO_2 with 50 mL H_2O. After 1 h of sedimentation, 48 mL supernatant was transferred to a new 50 mL tube followed by another 5 h sedimentation. Then the top 43 mL was removed and the silica was re-suspended and activated with 60 mL 37% HCL. To each of 2mL digested sample, 20 mL of the binding buffer (19.54 mL QIAGEN buffer PB, 360 mL 5M sodium acetate, 100 mL 5M sodium chloride) and 100 mL silica suspension was added and adjusted to pH 4-5 with 37% HCl[5]. After an hour of incubation at room temperature the supernatant was removed after a brief centrifugation step at 2000 g for 2 min and the pelleted silica was re-suspended in 1 mL binding buffer and washed twice with 80% ice-cold ethanol. The DNA was eluted from silica particles in 60 mL QIAGEN EB buffer. Extraction blanks were included with each round of extractions.

For the M9984/Kolyma 8467a sample the liquid fraction was concentrated using a 30 kDA centrifugal unit (Millipore) to 200-250 µL. A special buffer development by Dabney et al (2013) was used for the purification step following the publication guidelines for its preparation. A total of 13 volumes of the buffer was added to each sample and spins were

done at 8,000 rpm using a silica column (MinElute PCR purification Kit, Qiagen). The washing step was done using the Qiagen Washing buffer and spin down at 13,000 rpm for one minute. In the elution step, the columns were incubated twice with 55 µL of Qiagen EB at 37º for 10 minutes. Extraction blank were included with each round of extraction.

NGS library preparation and sequencing of ancient samples

20 ml DNA extracts were build into blunt-end libraries using Illumina-specific adapters and NEBNext DNA Sample Pre Master Mix Set 2 (E6070) kit according to manufacturer's instructions with some modification mentioned below: Since ancient DNA molecules are naturally fragmented, the nebulization step was skipped. The end-repair step was conducted in 25 mL reactions using 20 mL of DNA. This was incubated for 20 min at 12º C and 15 min at 37º C, and purified using PB buffer with QIAGEN MinElute spin columns, and eluted in 17 ml. Next, Illumina-specific adapters according to Meyer and Kircher 2010[6] were ligated to the end-repaired DNA fragments in 25 mL reactions. The reaction was incubated for 15 min at 20º C and purified with PB buffer on QIAGEN MinElute columns, before eluted in 20 mL EB Buffer. The adaptor fill-in reaction was carried out in a final volume of 30 mL and incubated for 20 min at 65º C followed by 20 min at 80º C. To assess the amount of DNA libraries in each sample and therefore the optimal number of PCR cycle for library amplification, qPCR was performed using SYBR green MIX (Roche) according to manufacturer's instructions and the same forward and reverse primers used for the following index PCR step. The DNA library (12 ml) was then amplified and indexed in a 50 mL PCR reaction, mixing with 25 mL 2X Kapa U+, 1 mL of each primer (10 mM, inPE forward primer + indexed reverse primer) and 11 mL H2O. Thermocycling conditions were 45 s at 98º C, followed by number of cycles (based on qPCR values) of 15 s at 98º C, 30 s at 65º C and 30 s at 72º C, and a final 1 min elongation step at 72º C. The amplified library was purified with PB buffer on QIAGEN MinElute columns, before being eluted in 50 mL EB. Negative library controls based on EB as well as libraries constructed on the negative extractions controls were included.

Sequencing of aDNA

The quantification of purified DNA libraries was conducted using Agilent Bioanalyzer 2100. The library pools were sequenced (80 bp in most cases, single read) on Illumina HiSeq 2500 system at the Danish National High-throughput DNA Sequencing Centre as well as at the Wellcome Trust Sanger Center Illumina X10 machine (Supplementary Data Table S1).

Basecalling and sequence sorting by sample-specific indexes were performed by the Sequencing Centre using CASAVA v.1.8.2.

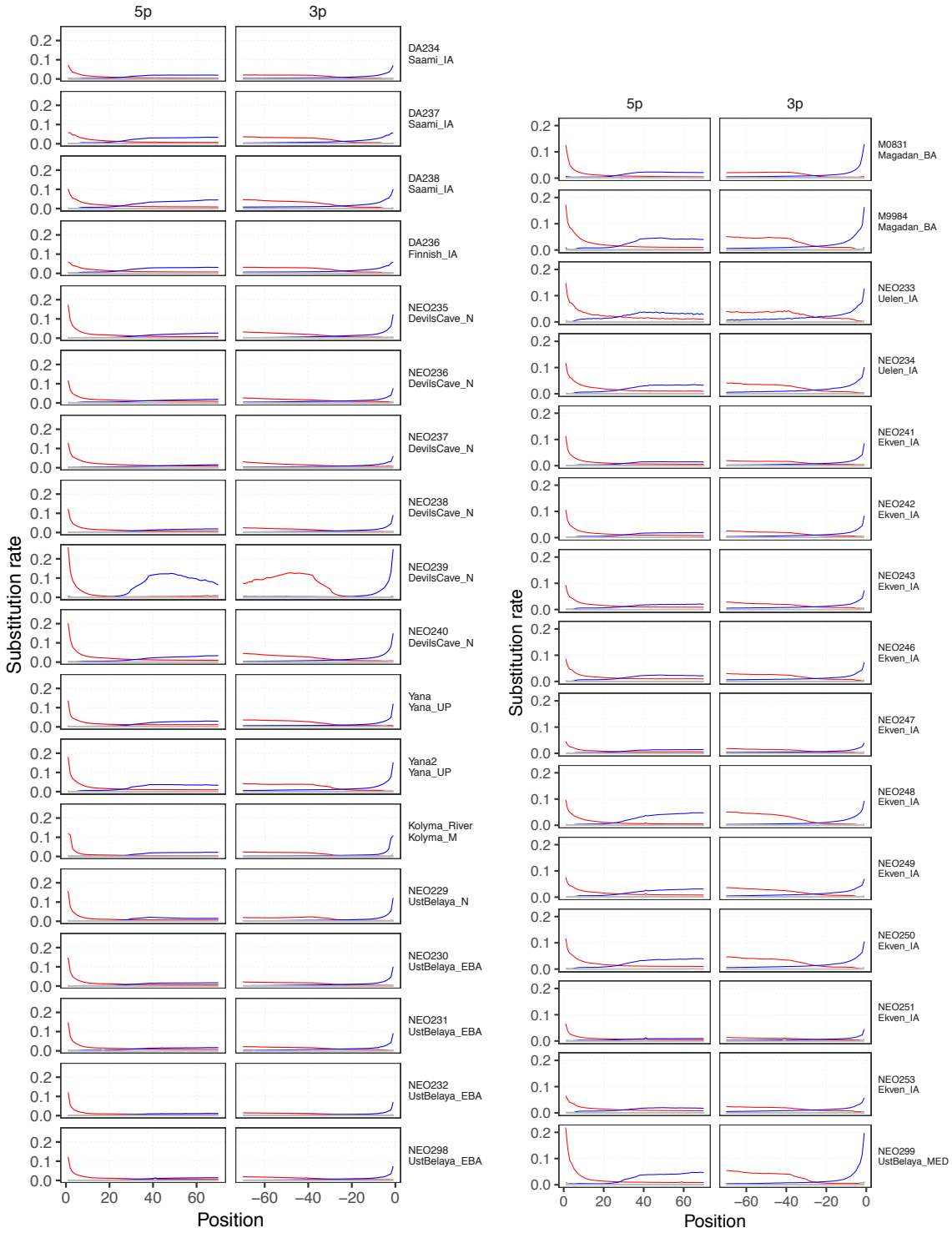


Fig. S2.1. Nucleotide misincorporation patterns in newly reported ancient individuals

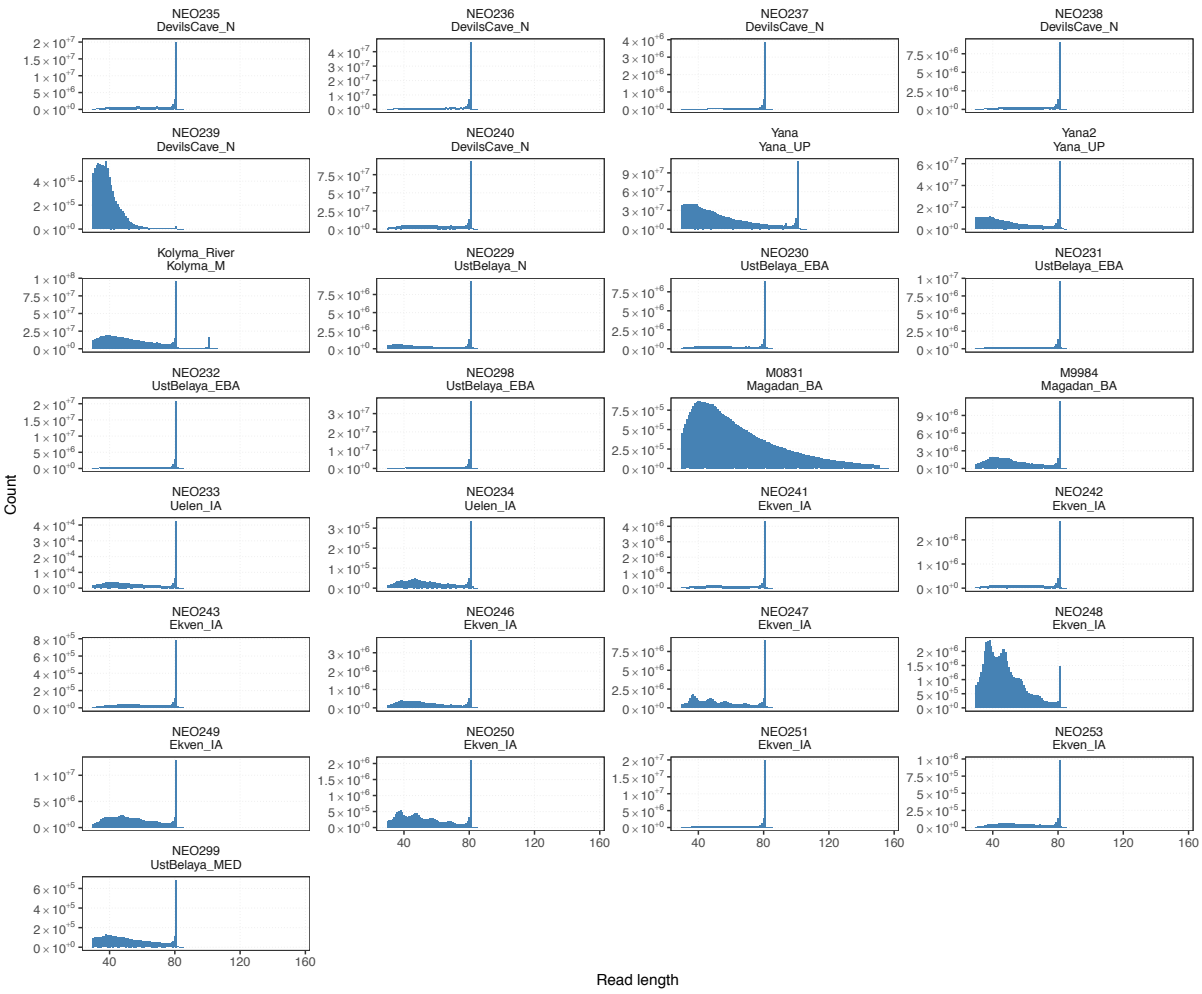


Fig. S2.2. Read length distributions for samples sequenced on Illumina HiSeq 2500

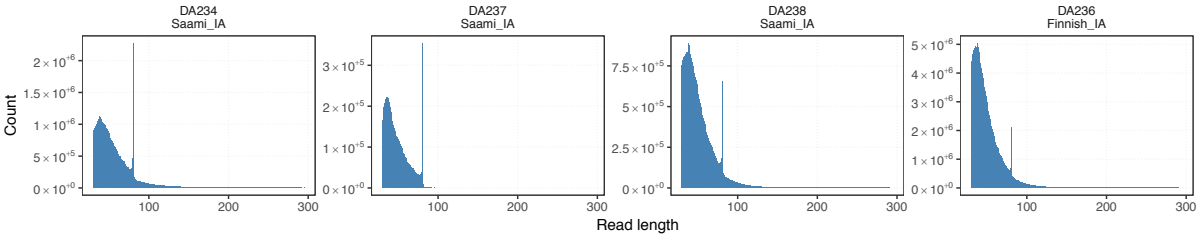


Fig. S2.3. Read length distributions for samples sequenced on both Illumina HiSeq 2500 and X10

References

1. Willerslev, E. & Cooper, A. Review Paper. Ancient DNA. *Proceedings of the Royal Society B: Biological Sciences* **272,** 3–16 (2005).
2. Gilbert, M. T. P., Bandelt, H.-J., Hofreiter, M. & Barnes, I. Assessing ancient DNA studies. *Trends Ecol. Evol. (Amst.)* **20,** 541–544 (2005).
3. Damgaard, P. B. *et al.* Improving access to endogenous DNA in ancient bones and teeth. *Sci Rep* **5,** 11184 (2015).
4. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat Comms* **5,** 5257 (2014).
5. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522,** 167–172 (2015).
6. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* (2010).

Supplementary Information 3 – Datasets

3.1 Overview of dataset processing

Analysis panels for population genetics analyses were constructed by merging the newly sequenced individuals with reference datasets of previously published ancient and modern individuals (Supplementary Data Tables 2, 3). For all datasets, genotypes for ancient individuals were obtained by randomly sampling an allele passing filters (mapping quality ≥ 30 and base quality ≥ 30) and matching one of the two alleles at the respective reference panel SNP positions. Ancient individuals with coverage ≥ 4X were additionally genotyped using the same samtools / bcftools[1] pipeline as modern WGS individuals, as previously described[2]. For WGS datasets, analyses were restricted to regions within the 1000 Genomes Phase 3 strict accessible genome mask (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/accessible_genome_mask s/20141020.strict_mask.whole_genome.bed), as well as outside repeat regions (UCSC genome browser simpleRepeat table).

3.2 Analysis panels

3.2.1 Panel 1 "1240K"

This panel includes SNPs targeted in two panels of in-solution capture used in previous ancient DNA studies[40,4540,453,4], for a total of 1,153,042 autosomal markers. This dataset includes the following individuals:

- the entire set of 170 ancient individuals included for analysis in this study (Supplementary Data Table 2, "1240K")
- SNP array genotype data of 2,287 modern individuals (the "fully public" samples from[5,6]) genotyped at a subset of 593,609 SNPs of the 1240K panel
- Genotypes of 104 modern WGS samples (Supplementary Data Table 3) at the full 1240K SNP set

3.2.2 Panel 2 "2240K"

This panel includes an extended panel of SNPs targeted in the 1240K panel as well as the 1000K panel ("panel 3" from[4]) for a total of 2,043,687 autosomal markers. This dataset includes the following individuals:

- A subset of 154 ancient individuals with 2240K capture or WGS data (Supplementary Data Table 2, "2240K")
- Genotypes of 104 modern WGS samples (Supplementary Data Table 3)
- Genotypes of 300 modern individuals from the SGDP[7]

3.2.3 Panel 3 "WGS"

This panel includes a total of 12,047,779 SNPs called after filters, by genotyping the following individuals as described above

- 25 ancient WGS individuals with coverage ≥ 4X (Supplementary Table 1, "diploid")
- 104 modern WGS individuals (Supplementary Data Table 3)

Genotypes for all 139 ancient individuals with WGS data (Supplementary Data Table 2, "WGS") were then obtained by randomly sampling an allele at the set of SNPs called from the higher coverage samples.

References

1. 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).
2. Sikora, M. et al. Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. Science 358, 659–662 (2017).
3. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522, 207–211 (2015).
4. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. Nature 524, 216–219 (2015).
5. Lazaridis, I. et al. Ancient human genomes suggest three an
cestral populations for present-day Europeans. Nature 513, 409–413 (2014).
6. Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. Nature 536, 419–424 (2016).
7. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538, 201–206 (2016).

Supplementary Information 4 – Y chromosome analyses

4.1 Sex determination

Sex determination for all individuals was carried out by analysing the fraction of high quality reads (MAPQ ≥ 30) mapping to the Y chromosome[1]. All study individuals were unambiguously assigned to either male or female (Table S3.1)

Table S4.1. Sex determination

Sample ID	Group	Number of reads X + Y	Y	RY	SE	Assignment
Yana1	Yana_UP	40,050,627	3,361,636	0.084	0.0000	XY
Yana2		10,264,112	856,272	0.083	0.0001	XY
Kolyma1	Kolyma_M	21,641,387	1,836,508	0.085	0.0001	XY
NEO235	DevilsCave_N	2,968,995	12,124	0.004	0.0000	XX
NEO236		5,618,882	19,344	0.003	0.0000	XX
NEO237		387,340	1,814	0.005	0.0001	XX
NEO238		1,183,859	5,119	0.004	0.0001	XX
NEO239		209,309	17,442	0.083	0.0006	XY
NEO240		15,582,686	70,065	0.005	0.0000	XX
NEO229	UstBelaya_N	1,449,874	6,907	0.005	0.0001	XX
NEO230	UstBelaya_EBA	1,196,884	5,314	0.004	0.0001	XX
NEO231		1,136,943	4,907	0.004	0.0001	XX
NEO232		1,366,528	121,985	0.089	0.0002	XY
NEO298		4,080,005	17,295	0.004	0.0000	XX
M9984	Magadan_BA	3,543,939	19,057	0.005	0.0000	XX
NEO234	Uelen_IA	87,410	438	0.005	0.0002	XX
NEO233		8,007	53	0.007	0.0009	XX
NEO241	Ekven_IA	339,123	31,022	0.092	0.0005	XY
NEO242		235,613	21,490	0.091	0.0006	XY
NEO243		67,595	6,137	0.091	0.0011	XY
NEO246		450,646	40,426	0.090	0.0004	XY
NEO247		1,279,655	112,710	0.088	0.0003	XY
NEO248		1,448,057	122,985	0.085	0.0002	XY
NEO249		2,414,994	212,053	0.088	0.0002	XY
NEO250		424,782	36,692	0.086	0.0004	XY
NEO251		2,300,997	9,147	0.004	0.0000	XX
NEO253		97,588	8,694	0.089	0.0009	XY
DA234	Saami_IA	2,076,545	1,437	0.001	0.0000	XX
DA237		300,983	214	0.001	0.0000	XX
DA238		1,358,499	693	0.001	0.0000	XX
DA236	Finnish_IA	7,154,940	3,336	0.001	0.0000	XX
NEO299	UstBelaya_MED	238,311	1,269	0.005	0.0001	XX
Yana_young	Yana_MED	2,629,655	234,046	0.089	0.0002	XY

4.2 Y chromosome haplogroup assignment

To determine Y chromosome haplogroups, we extracted genotypes for male individuals at SNPs included in the Y-DNA haplogroup tree from the International Society of Genetic Genealogy (ISOGG, http://www.isogg.org, version 13.38). We genotyped each individual using samtools/bcftools with the haploid option, and recorded associated variant quality (QUAL tag) and read depth (DP, DP4) at each SNP call. In Table S4.2 we report for each individual the most derived haplogroup with multiple derived alleles observed, which also shows consistent derived alleles for all parent haplogroups along the tree.

Table 4.2. Y chromosome haplogroups

Group	Sample ID	Haplogroup		Observed mutations			
		Major clade	ISOGG	Derived	Total	Fraction	IDs
Yana_UP	Yana1	P-M45	P1	12	21	0.57	L138;P237/PF5873;L721/PF6020;P243/PF5874;F1857/P337/Page83/PF5901;L779/PF5907/YSC0000251;P228/PF5927;P239/PF5930;P281/PF5941;M45/PF5962;L768/PF5976/YSC0000274;L471/PF5989
	Yana2	P-M45	P1	13	21	0.62	L138;P237/PF5873;L721/PF6020;P243/PF5874;F1857/P337/Page83/PF5901;L779/PF5907/YSC0000251;P228/PF5927;P239/PF5930;P281/PF5941;M45/PF5962;P284;L768/PF5976/YSC0000274;L471/PF5989
Kolyma_M	Kolyma1	Q-NWT01	Q1a1a	2	68	0.03	Y693;Y703
DevilsCave_N	NEO239	C-L1373	C2b	3	4	0.75	F6273/FGC16321/Y4554/Z16745;FGC16425/Z18153/Y6704;F12453/FGC16446/Y6708/Z16758
UstBelaya_EBA	NEO232	Q-L53	Q1b1	15	15	1.00	CTS1330/M1070;M1073;L213;M1108/CTS4105;CTS5104/M1116;CTS5647/M1120;CTS6106/M1124;CTS6357/M1127;L475;L55;CTS10434/M1151;M1161/Y768;M1162;CTS11920/M1176;M1177/Y761
Ekven_IA	NEO241	Q-NWT01	Q1a1	7	9	0.78	F1298;Y619;Z19172;F5013/SK1918;Y699;Y643;F5425/Z19212
	NEO242	Q-L53	Q1b1	3	3	1.00	CTS2068/M1089;M1108/CTS4105;Y763
	NEO243	Q-M242	Q1	1	1	1.00	F2656/M1140
	NEO246	Q-L53	Q1b1	5	6	0.83	CTS1330/M1070;CTS2122/M1090/Z799;CTS6857/M1129;CTS10434/M1151;L53/S326
	NEO247	Q-M3	Q1b1a1a2	3	3	1.00	FGC8456/Y4303;FGC8466/Y4325;FGC8472/Y6163
	NEO248	Q-B143	Q1a1b	19	19	1.00	YP1470/Z36007;YP1471/Z36008;FGC25513.2;Z36011;Z36012;Z36013;Z36014;YP1473/Z36015;Z36018;Z36019;YP1478/Z36020;YP1481/Z36021;Z36022;Z36025;Z36026;YP1500/Z36028;YP1632/Z36030;Z36031;YP1502/Z36036
	NEO249	C-L1373	C2b1a1b2	2	2	1.00	Z32841;Z32842
	NEO250	Q-B143	Q1a1b	7	7	1.00	B143/YP1469;Z36009;FGC25513.2;YP1473/Z36015;Z36026;Z36032;Z36035
	NEO253	C-L1373	C2b	2	2	1.00	F3805;FGC16511/Z18156/Y4648
Yana_MED	Yana_young	N-M178	N1a1a1a1a4a1	7	7	1.00	M1993;M1987;M2077;M1991;M2103;M2108;M2122

<u>Yana</u>

The two UP individuals at Yana were assigned to haplogroup P1 (P-M45), which is ancestral to haplogroups Q and R. Both individuals show the derived state at ~60% of P1-definining SNPs in the ISOGG database, with one additional mutation in Yana 2 not found in Yana 1 (P284). Neither of them shows any derived allele at either Q- or R-defining mutations, and can thus be considered basal P1* (P1xQ,R). The 24,000 year old Mal'ta individual was found to be basal R*, which is a subclade of P1[2].

The medieval individual from the other Yana site was assigned to a subclade of haplogroup N1a1a (N-M178), a common haplogroup among present-day Siberian populations[3].

Kolyma

Kolyma1 was assigned to haplogroup Q1a1a, a subclade of haplogroup Q-M242 whose descendants are widespread among present-day Siberians and Native Americans. This haplogroup is closely related to the one observed in the 4,000-year-old Saqqaq Paleoeskimo[4] individual (Extended Data Fig. 2), and has been observed among some present-day populations of the North American Arctic[5].

Devil's Gate Cave

The male individual from Devil's Gate cave was assigned to haplogroup C2b based on three observed derived mutations for the haplogroup.

Ust'Belaya

The male individual from early Bronze Age Ust'Belaya was assigned to haplogroup Q1b1 (Q-L53).

Ekven

The male individuals recovered from the Neo-eskimo cemetary at Ekven were found to exhibit predominantly haplogroup Q, with the exception of two individuals assigned to C2b. A number of different subclades of Q-M242 were found, including at least one individual with a subclade of Q-M3 (NEO247), one of the founding haplogroups of Native Americans observed in ancient American individuals[6,7].

References

1. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40,** 4477–4482 (2013).
2. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505,** 87–91 (2014).
3. Derenko, M. *et al.* Y-chromosome haplogroup N dispersals from south Siberia to Europe. *J. Hum. Genet.* **52,** 763–770 (2007).

4. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463,** 757–762 (2010).
5. 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.* **109,** 8471–8476 (2012).
6. Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506,** 225–229 (2014).
7. Rasmussen, M. *et al.* The ancestry and affiliations of Kennewick Man. *Nature* **523,** 455–458 (2015).

Supplementary Information 5 – Relatedness

SI 5.1 IBS analysis

We used the estimator introduced by[1], implemented in the KING package (equation (9) in[1]) to infer kinship coefficients between the study individuals. The coefficients were calculated from a matrix of IBS sharing between two individuals, obtained using a 2D-SFS inferred by realSFS from the ANGSD package[2], on polymorphic sites from the 1000 Genomes project[3]. We additionally used a related statistic called R1, the ratio of double heterozygous (Aa/Aa) sites divided by the total number of discordant genotypes. Using the notation of[1]:

$$R1 = \frac{N_{Aa,Aa}}{N_{AA,Aa} + N_{AA,aa} + N_{Aa,AA} + N_{Aa,aa} + N_{aa,AA} + N_{aa,Aa}}$$

Results for this analysis are shown in Figure ED4 and Table S6.1. We found two first-degree relative pairs as well as a second-degree relationship among four individuals from Devil's Gate Cave. All related individuals were females, and mitochondrial DNA haplogroups were consistent with the inferred relationships (Table S6.1). We therefore excluded the individuals with lower coverage for each first-degree pair (NEO235, NEO237) for population-level analyses throughout the study.

Table S5.1 Relative pairs at Devil's Gate Cave

Sample ID						Coverage		MT Haplogroup	
Individual 1	Individual 2	Group	Kinship	ratioR1	Degree	Individual 1	Individual 2	Individual 1	Individual 2
NEO235	NEO240	DevilsCave_N	0.26	0.54	1	1.263	6.563	D4m	D4m
NEO236	NEO237		0.22	0.59	1	2.456	0.173	D4	D4
NEO237	NEO238		0.24	0.48	1	0.173	0.515	D4	D4
NEO236	NEO238		0.09	0.36	2	2.456	0.515	D4	D4

References

1. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinforma. Oxf. Engl.* **26,** 2867–2873 (2010).
2. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15,** 356 (2014).
3. The 1000 Genomes Project Consortium,. A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

Supplementary Information 6 – Admixture modelling

Martin Sikora, Melinda Yang, Qiaomei Fu

6.1 Overview of analyses

In this section we provided more detailed descriptions of the admixture modelling carried out on the ancient and modern groups. For these analyses, we used methods of the *f*-statistic framework[1] implemented in the ADMIXTOOLS package. We quantified genetic affinities among individuals and groups using f_3- and f_4-statistics, as well as the more model-based approaches of admixture graph fitting (*qpGraph*) and estimation of mixture proportions (*qpWave/qpAdm*)[2]. The bulk of these analyses are based on the "WGS" dataset (panel 3) using only transversion polymorphisms, to minimize the effects of shared ancient DNA damage. Results based on other panels (e.g. those including individuals with SNP capture data only) are noted where appropriate.

6.2 Ancient North Siberians (Yana)

Affinities with early Eurasians

We quantified the genetic affinities of the two Yana individuals with other ancient individuals using outgroup-f_3 statistics. Both individuals share higher affinity with each other than with other individuals (Table S6.1; Extended Data Fig. 3d; Fig. S6.1), and we confirmed with D-statistics D(Mbuti,X;Yana1,Yana2) that they form a clade to the exclusion of all other individuals (all |Z| < 3).

Table S6.1 Top outgroup-f_3 statistics for Yana individuals (1240K panel)

f_3(Mbuti;Yana1,X)				f_3(Mbuti;Yana2,X)			
X	f_3	SE	N SNPs	X	f_3	SE	N SNPs
Yana2	0.312	0.005	381,772	Yana	0.312	0.005	381,772
MA1	0.267	0.004	403,588	MA1	0.268	0.005	258,695
GoyetQ116-1	0.265	0.005	383,180	GoyetQ116-1	0.264	0.004	248,673
SI	0.262	0.004	412,118	Vestonice43	0.263	0.005	38,705
Loschbour	0.262	0.004	574,857	Karelia	0.262	0.004	314,434
SII	0.262	0.004	563,647	AfontovaGora3	0.261	0.005	74,503
SIII	0.262	0.004	545,288	Vestonice16	0.261	0.004	235,703
Karelia	0.261	0.004	492,138	Loschbour	0.261	0.004	363,849
Motala2	0.261	0.005	82,936	Ostuni1	0.261	0.005	96,842
SIV	0.260	0.004	561,979	Bichon	0.261	0.004	237,441
Bichon	0.259	0.004	368,563	SI	0.260	0.004	263,481
Vestonice16	0.258	0.005	362,289	SII	0.258	0.004	358,179
Vestonice43	0.255	0.005	58,632	Motala2	0.257	0.005	53,154
Ostuni1	0.254	0.005	145,632	SIV	0.257	0.004	357,104
AfontovaGora3	0.253	0.005	112,425	SIII	0.256	0.004	345,301

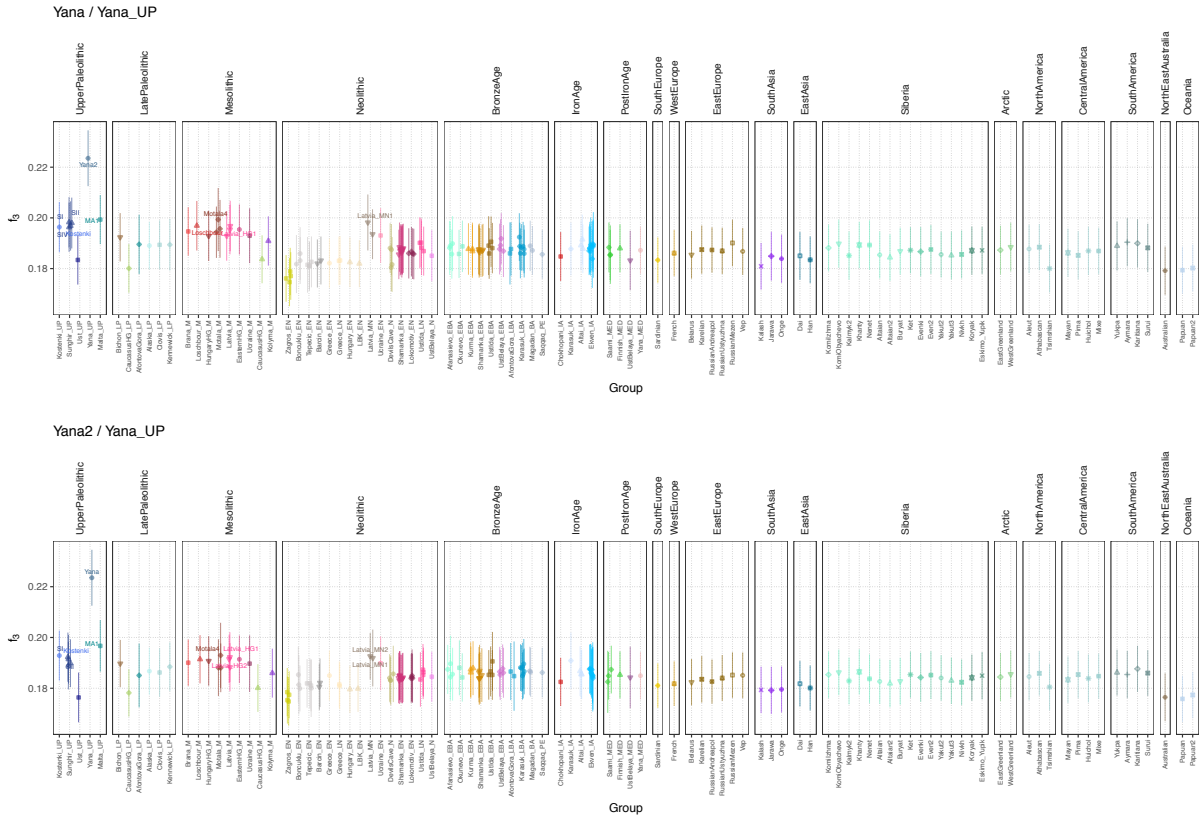


Figure S6.1 Outgroup-f_3 statistics for Yana individuals (WGS panel, no transitions).

While genetic affinities for Yana were highest for early West Eurasians (EWE), we found that they do not form a clade with EWE to the exclusion of early East Asians (EEA) (Extended Data Fig. 7e-f). This was also reflected in the outgroup-f_3 statistics, as both individuals shared most genetic drift with Mal'ta and GoyetQ116-1, two individuals that have also previously been shown to harbor both EWE and EEA affinities[3,4]. We further investigated the population structure among Yana and other early Eurasians using admixture graphs. We used the model proposed by Lipson et al.[4] as a base for this analysis, modified to include Upper Paleolithic hunter-gatherers from Sunghir[5] as representatives of EWE as well as Devil's Gate Cave individuals as EEA. This provides a good fit to the data with $|Z|<3$ for all predicted *f*-statistics (Figure S6.1). Yana was succesfully added to this base model as a mixture of EWE and EEA, with an estimated 27% of its ancestry derived from EEA ($\max(|Z|) = 3$; Figure S6.2). This estimate was consistent with mixture proportions obtained using *qpAdm* and a set of 6 outgroups (Mbuti, Yoruba, Mota, Onge, Papuan, Ust'Ishim), where Yana was estimated as a mixture of 69 ± 5% EWE (Kostenki) and 31 ± 5% EEA (Devil's Gate Cave) ($p = 0.44$).

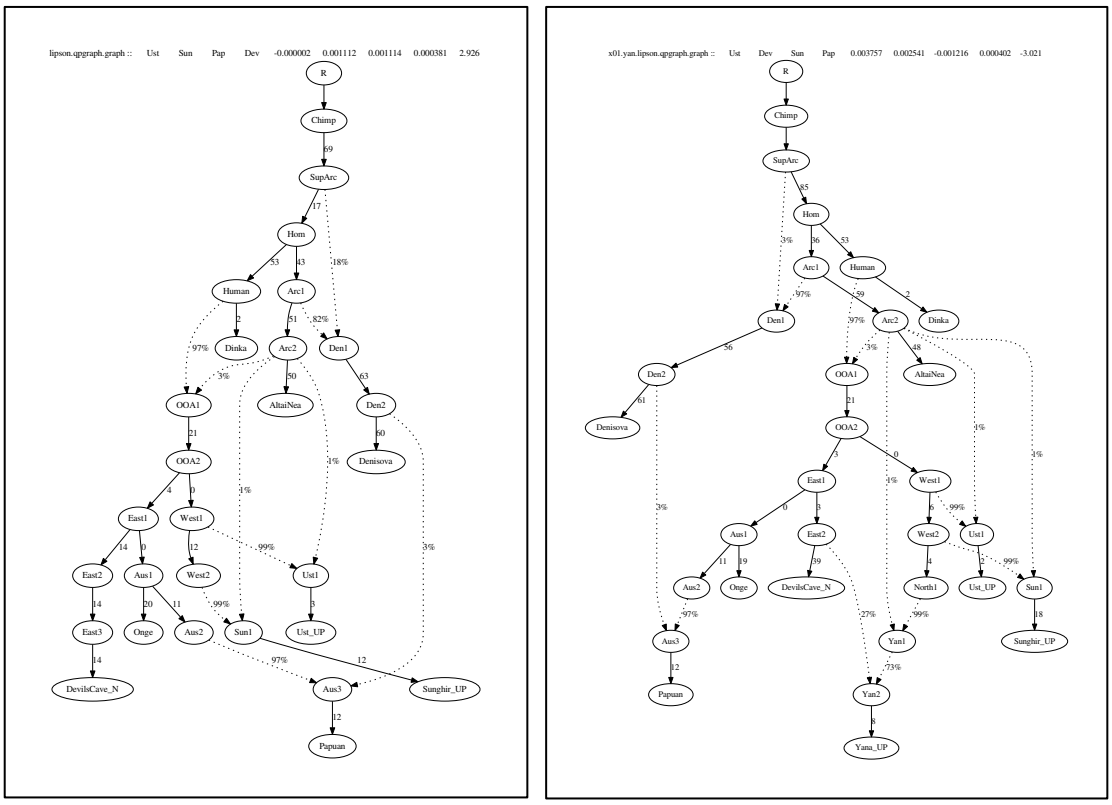


Figure S6.2 Admixture graphs. (left) Base model adopted from Lipson et al[4]. (right) Fit of Yana as a mixture of EEA and EWE.

<u>Relationship with Mal'ta and ANE</u>

We were interested to determine the relationship of Yana with the 24,000-year old Mal'ta individual, generally referred to as the type specimen for a hypothetical "Ancient North Eurasian" (ANE) population which forms one of the ancestral lineages of Native Americans. Outgroup-f_3 statistics suggested a closer affinity of Yana with Mal'ta than with any other early Eurasian individual (Table S6.1). We used f_4 statistics for all triplets involving Mal'ta, Yana and the remaining UP Eurasians in the respective panels to test whether Malt'a and Yana form part of the same lineage. Results for different analysis panels are shown in Tables S6.2 – S6.4. We find that tests are generally consistent with Mal'ta and Yana forming a clade with each other, albeit with a relatively weak signal for some comparisons, namely those involving samples with lower coverage data obtained with SNP capture (Vestonice, Ostuni). The relationship of GoyetQ116-1 with Yana or Mal'ta is unresolved, notable due to the fact that this individual also shares the signal of EEA affinity with Yana and Mal'ta[3] (Extended Data Fig. 3).

Table S6.2 f_4 statistics for Yana, Mal'ta and UP Eurasians (1240K)

X	f_4(Mbuti,Malta_UP;Yana_UP,X) f_4	Z	f_4(Mbuti,X;Malta_UP;Yana_UP) f_4	Z	f_4(Mbuti,Yana_UP;X,Malta_UP) f_4	Z
Ostuni_UP	-0.0020	-2.66	0.0006	0.80	0.0015	2.01
Vestonice_UP	-0.0012	-1.90	0.0001	0.19	0.0011	1.67
GoyetQ116_UP	-0.0011	-1.52	0.0007	0.99	0.0004	0.53
Kostenki_UP	-0.0025	-3.71	0.0006	1.01	0.0019	2.77
Sunghir_UP	-0.0023	-4.14	0.0004	0.74	0.0019	3.33
Tianyuan_UP	-0.0068	-9.83	0.0017	2.50	0.0051	7.70
Ust_UP	-0.0082	-11.39	0.0005	0.77	0.0077	10.49

Table S6.3 f_4 statistics for Yana, Mal'ta and UP Eurasians (2240K)

X	f_4(Mbuti,Malta_UP;Yana_UP,X) f_4	Z	f_4(Mbuti,X;Malta_UP;Yana_UP) f_4	Z	f_4(Mbuti,Yana_UP;X,Malta_UP) f_4	Z
Ostuni_UP	-0.0019	-2.77	0.0005	0.69	0.0015	2.26
Vestonice_UP	-0.0011	-1.85	0.0000	0.09	0.0010	1.79
Kostenki_UP	-0.0023	-3.69	0.0009	1.68	0.0014	2.21
Sunghir_UP	-0.0019	-3.99	0.0006	1.33	0.0013	2.61
Tianyuan_UP	-0.0063	-10.38	0.0017	2.89	0.0046	7.82
Ust_UP	-0.0071	-10.97	0.0006	0.99	0.0065	9.93

Table S6.4 f_4 statistics for Yana, Mal'ta and UP Eurasians (WGS transversions only)

X	f_4(Mbuti,Malta_UP;Yana_UP,X) f_4	Z	f_4(Mbuti,X;Malta_UP;Yana_UP) f_4	Z	f_4(Mbuti,Yana_UP;X,Malta_UP) f_4	Z
Kostenki_UP	-0.0007	-3.06	0.0003	1.33	0.0004	1.79
Sunghir_UP	-0.0006	-3.52	0.0001	0.86	0.0005	2.62
Ust_UP	-0.0022	-9.82	0.0001	0.57	0.0021	8.99

Using the admixture graph fit with Yana (Fig. S6.2) as a base, we find that Mal'ta can be successfully fitted as a clade with Yana in the "Ancient North Siberian" (ANS) lineage sharing the EWE-EEA mixture, consistent with the results from the f_4 statistics (Fig. S6.3).

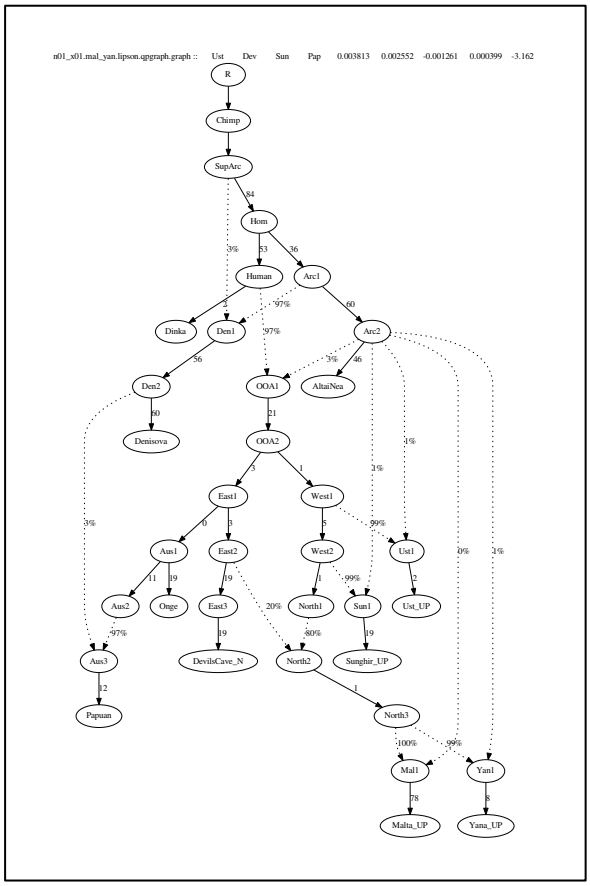


Figure S6.3 Admixture graph including Mal'ta.

Despite the clade relationship of Yana and Mal'ta with respect to early UP Eurasians, we find that many later populations share significantly more alleles with Mal'ta than with Yana, namely those previously identified as rich in "ANE" ancestry. This includes Native Americans (Table S6.5), which we successfully fitted as mixtures of ANS and EEA, with Mal'ta diverging more recently from the contributing ANS lineage (Fig. 2). Thus, while both Yana and Mal'ta can therefore attributed to the ANE ghost population, Mal'ta is more closely related to the unsampled populations contributing this ancestry.

Table S6.5 Top f_4 statistics f4(Mbuti,X;Mal'ta,Yana) (WGS transversions only)

P2	P3	P4	f_4	Z	N Sites
AfontovaGora_LP	Malta_UP	Yana_UP	-0.0050	-14.57	164,160
Ucraine_EN			-0.0024	-9.00	413,198
EasternHG_M			-0.0023	-10.52	1,952,173
Latvia_MN			-0.0022	-11.05	1,747,575
Okunevo_EBA			-0.0019	-9.78	983,460
Afanasievo_EBA			-0.0019	-11.38	2,354,824
Pima			-0.0018	-8.39	2,128,508
Alaska_LP			-0.0017	-8.51	2,318,158
Mixe			-0.0017	-10.09	2,371,907
Kennewick_LP			-0.0017	-7.36	1,464,112

West Eurasian hunter-gatherer structure

We next sought to further investigate the relationship of Yana and Mal'ta with later West Eurasian hunter-gatherers. Previous studies have documented at least three major EWE lineages:

- Western hunter-gatherers (WHG) from Western Europe, represented in this analysis by a Late Palaeolithic (LP) individual from Bichon, Switzerland[6] and a Mesolithic individual from Loschbour, Luxembourg[7]
- Eastern hunter-gatherers (EHG)[2] from Eastern Europe, represented by a Mesolithic individual from Karelia, Russia[8]
- Caucasus hunter-gatherers (CHG), represented by a LP individual from Satsurblia cave, Georgia[6]

We fitted those individuals sequentially on the base graph as above after removing the eastern out-of-Africa branch (East Asians, Papuans, Onge, Native Americans) to reduce the complexity of the base graph.

We began by fitting the LP CHG individual from Satsurblia Cave. Previous studies have documented the contribution of an early "basal Eurasian" lineage to CHG[6], as well as it's genetic affinity to ANE. Consequently, Fu et al.[8] proposed a model where CHG was fitted as a mixture of basal Eurasian and a lineage related to ANE. We successfully replicate this model fit using Sunghir[5] and Kostenki[9] as representatives for West Eurasian UP hunter-gatherers and Mal'ta as ANE (max(|Z| = 2.5) (Figure S6.4 left). However, fitting CHG in the same position on a base model that further includes Yana fails (max(|Z|) = 4.3) (Figure S6.4 right), with the worst outliers indicating a lack of shared drift genetic drift between Yana and Mal'ta.

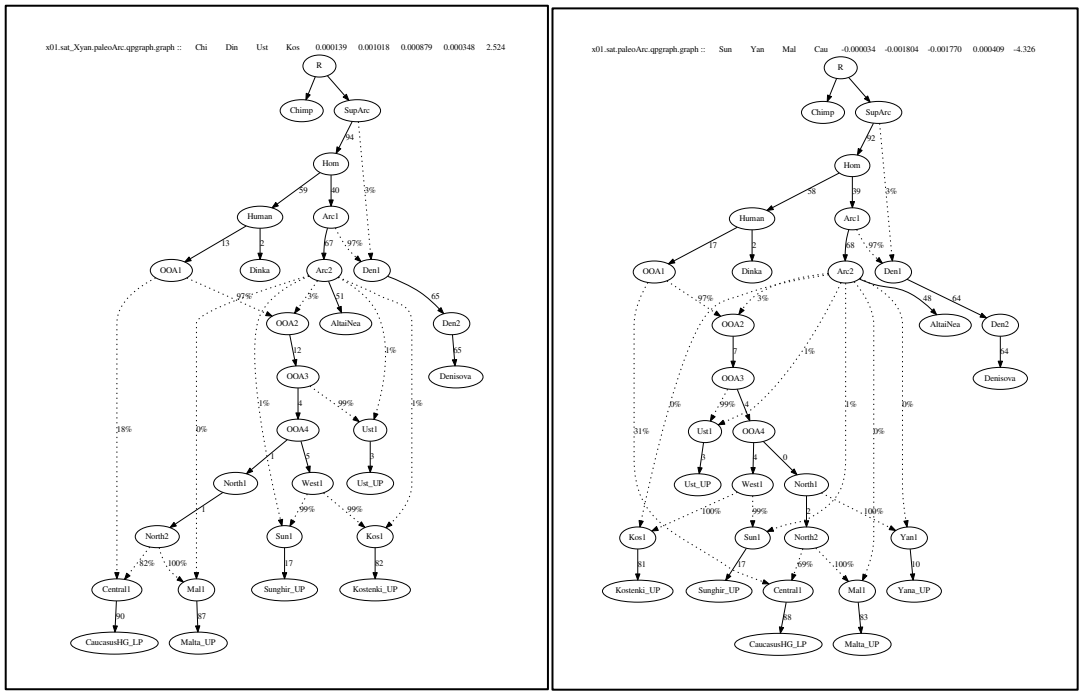


Figure S6.4 Admixture graphs fitting CHG (CaucasusHG_LP). (left) Successful fit on a base model excluding Yana. (right) Unsuccessful CHG fit onto a base model including Yana.

Using f_4 statistics, we confirmed on the other hand that CHG shares more alleles with Mal'ta than with Yana (f_4(Mbuti,CaucasusHG_LP;Yana_UP,Malta_UP), |Z| = 4.8). An alternative scenario to account for this increased sharing involves gene flow from a CHG-related group into Mal'ta, and a corresponding admixture graph including gene flow from a source diverging prior to the basal Eurasian mixture provided a good fit (max(|Z|) = 2.5) (Figure S6.5). Mal'ta was also successfully modelled as a 2-way mixture of Yana (72 ± 4%) and CHG (28 ± 4%) (p=0.22) using *qpAdm* with a set of 9 outgroups (Mbuti, Yoruba, Mota_N, Onge, Papuan, Ust_UP, Kostenki_UP, Sunghir_UP, Zagros_EN).

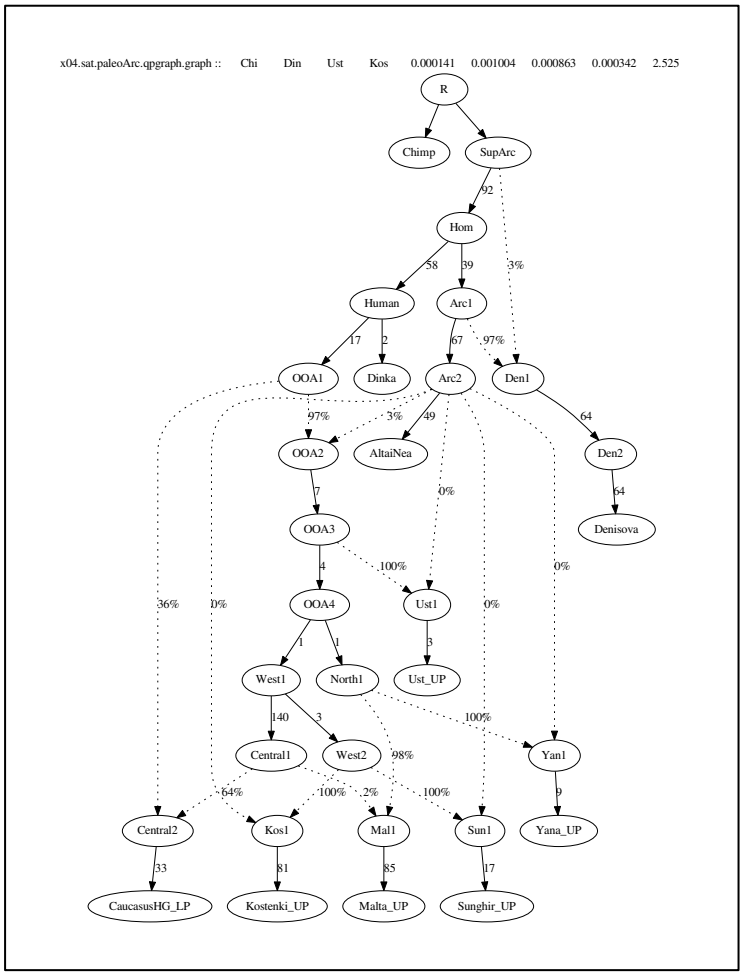


Figure S6.5 Admixture graph fitting Mal'ta with CHG-related admixture.

We next added the EHG individual from Karelia to the graph. Given its close affinity with Mal'ta, we began with a model of EHG forming a clade with Mal'ta. This model fit was unsuccessful (max(|Z|) = 4.8), with the worst *f*-statistics showing a lack of shared genetic drift of EHG with CHG (Fig. S6.6 left). Modelling a pulse of admixture from a CHG-related source into EHG improves the fit considerably (max(|Z| = 3.5; Fig. S6.6 right). Using *qpAdm* with a set of 10 outgroups (9 described above + Yana_UP), we estimate that EHG can be formed as a mixture of 18.8 ± 4% CHG and 81.2 ± 4% Mal'ta (p = 0.014).

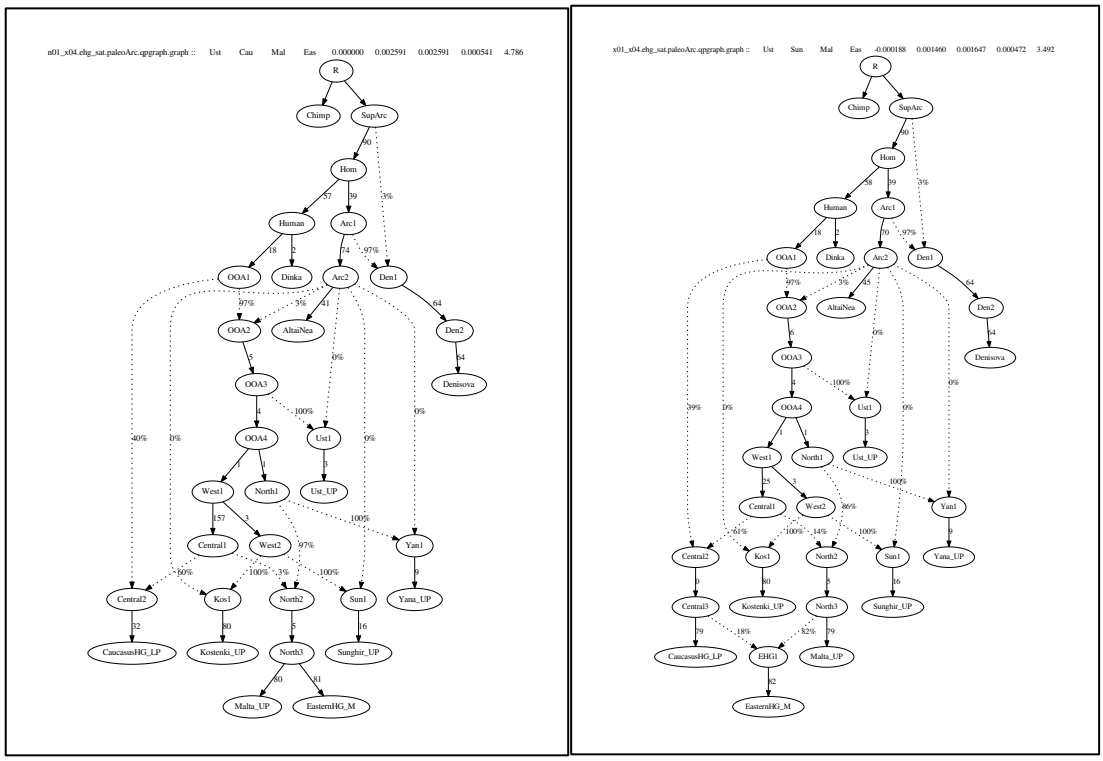


Figure S6.6 Admixture graphs fitting EHG (EasternHG_M). (left) Unsuccesful fit as sister clade to Mal'ta (Malta_UP). (right) An improved fit with EHG as a mixture of Mal'ta and CHG (CaucasusHG_LP).

The presence of CHG-related ancestry in EHG predicts that EHG should also contain "basal Eurasian" ancestry, albeit at a reduced amount compared to CHG. We used f_4 -statistics of the form f_4(Mbuti,Ust_UP;EasternHG_M,X) to test this hypothesis, as this statistic is expected to be positive in the presence of basal Eurasian ancestry that pre-dates the divergence of Ust'Ishim. We find a weakly positive signal for all test groups X (Table S6.6), suggesting that EHG indeed harbours basal Eurasian ancestry in a diluted form.

Table S6.6 f_4 statistics Mbuti,Ust_UP;EasternHG_M,X)

P2	P3	P4	f_4	Z	N Sites
Ust_UP	EasternHG_M	Kostenki_UP	0.0006	2.44	2,418,348
		Sunghir_UP	0.0003	1.89	2,805,748
		Bichon_LP	0.0004	2.12	1,822,240
		Loschbour_M	0.0004	2.04	2,681,732
		Yana_UP	0.0007	3.84	2,792,248
		Malta_UP	0.0007	3.27	1,937,098

While both admixture graph fitting and *qpAdm* modelling yielded reasonable fits for EHG as mixture of CHG and Mal'ta, the maximum Z-score from *qpGraph* (3.5) and tail probabilities from *qpAdm* (p = 0.014) indicated further possible improvements in their fits. In particular, the worst *f*-statistics in the admixture graph suggested a lack of shared drift between EHG and Sunghir, which shares ancestry with later WHG individuals[5]. Furthermore, previous studies have reported that simple tree models involving EHG, WHG and Mal'ta could not be fit, requiring that at least one of them would have to be admixed[2]. We therefore added the WHG individuals from Bichon and Loschbour to the graph as a possible third ancestral source for EHG. We found that this provided an excellent fit to the data, with only a single predicted *f*-statistic slightly exceeding a Z = 3 threshold (Fig. 2; Fig. S6.7).

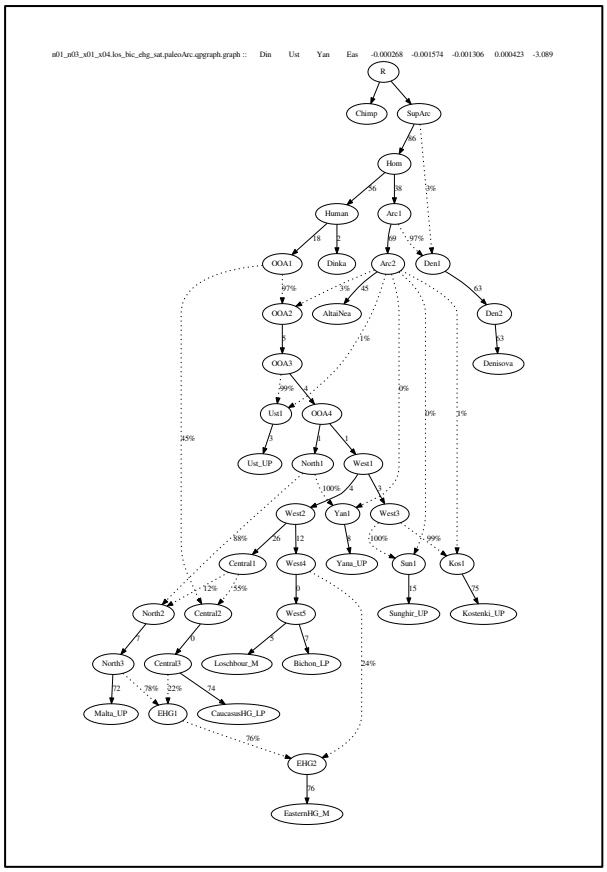


Figure S6.7 Final admixture graph fitting CHG, WHG and EHG

We also successfully modelled EHG as a three-way mixture of Mal'ta (44.5 ± 11%), CHG (21.1 ± 4%) and WHG (34.4 ± 11%) (p = 0.28) using the *qpAdm* framework, consistent with the admixture graph results. We note that in our admixture graph Bichon and Loschbour form a clade with each other and are hence equally related to EHG. In a recent study, Lazaridis et

al.[10] used Bichon as a reference group to document apparent EHG-related gene flow into WHG, which would be inconsistent with the symmetrical relationship we observed here. A notable difference is that Lazaridis et al. used a pooled set of three individuals to represent WHG, whereas we restricted our analysis here to Loschbour only. Indeed, repeating their *qpAdm* modelling with the three WHG individuals separately, we found that the EHG signal is driven entirely by the Eastern European hunter-gatherer individual from Koros, Hungary (KO1), which is estimated as (70.3 ± 15%) Bichon and (29.7 ± 15%) EHG (also observed in[11]) (Table S6.7). Models for the two other Western European individuals (Loschbour, La Brana) result in infeasible mixture proportions and can therefore be considered symmetric with respect to EHG. In summary, our modelling highlights the complex interplay between different lineages of EWE in the formation of Late Paleolithic and Mesolithic hunter-gatherers from Western Eurasia.

Table S6.7 *qpAdm* modelling of WHG individuals

testPop	refPop	alpha	se	nSnps	chisq	pval	pvalNested	flag
Loschbour_M	Bichon_LP	1.06	0.12	2,326,775	8.15	4.19E-01	6.00E-01	infeasible
	EasternHG_M	-0.06	0.12					infeasible
Brana_M	Bichon_LP	1.11	0.16	2,281,343	17.00	3.01E-02	4.01E-01	infeasible
	EasternHG_M	-0.11	0.16					infeasible
HungaryHG_M	Bichon_LP	0.70	0.15	2,049,264	8.95	3.47E-01	2.85E-02	
	EasternHG_M	0.30	0.15					

Relationship with Native Americans, Siberians and East Asians

Here, we consider the Yana individuals' relationship to present-day Asian and Siberian populations, as well as ancient and present-day Native American populations. We compare to a set of recent southern Native Americans (SAMER=Anzick-1, Surui, Karitiana, Mixe), northern present-day northern Native Americans and select Far East Siberians (NAMER+SIB1=Tlingit, Cree, Aleut, Chukchi), other present-day Siberians (SIB2=Sireniki Eskimos, Itelman, Even, Yakut) and present-day East and Southeast Asians (EAS=Ami, Dai, Han, Japanese).

We note major trends for these populations, relative to the UstIshim individual, the Tianyuan individual, early Europeans (EE=Kostenki14/GoyetQ116-1/Vestonice16), ancient North Eurasians (ANE=Malta1, AfontovaGora3), more recent Europeans (RE=ElMiron,

Villabruna, Loschbour, Bichon, LaBrana1, Hungarian.KO1) and the Karelia individual, who shows European ancestry, but also has a strong connection to the ANE.

Relative to the Ust'-Ishim individual, we find the same relationship across all Asian-affiliated populations (Tables S6.8-S6.11) – the Yana individuals are more closely related to Asian-affiliated populations than to the Ust'-Ishim individual. Our results suggest that the Yana individuals do not share a close relationship to the Ust'-Ishim individual. However, we do find that *D(Yana1, NAMER+SIB1; UstIshim, Mbuti)* may be greater than zero, as 2.3<Z<3.0, whereas we fail to reject *D(Yana2, NAMER+SIB1; UstIshim, Mbuti)=0* (0.6<Z<1.2, Table S.6.9). This potential connection between Yana1 and the Ust'-Ishim individual is stronger when we use transversions only (2.6<Z<3.7), but the high Z-score does not appear when we use the Chimp as an outgroup (1.0<Z<1.8). Recent European ancestry in the NAMER+SIB1 populations with a small amount of Basal Eurasian ancestry may explain the slight affinity between the Yana1 and Ust'-Ishim individuals, though it is less clear why we would not see this signal in both Yana individuals.

For the SAMER and NAMER+SIB1, we find relatively similar relationships between the Tianyuan individual, the Yana individuals, and these Americans and Siberians (D~0, Tables S6.8-S6.10). However, most SAMER share more alleles with the Tianyuan individual than the Yana individuals share (*D(Yana, SAMER; Tianyuan, Mbuti)<0*; for Mixe, Z=-1.8, Table S6.8), while the NAMER+SIB1 populations mostly do not show this feature (we fail to reject *D(Yana, Tianyuan; NAMER+SIB1, Mbuti)=0*, 0.5<Z<2.0,Table S6.9). The Yana individuals share more alleles with the Tlingit than with the Tianyuan individual (Z>3.2, Table S6.9), potentially due to recent European admixture into the Tlingit (Verdu et al. 2014). The SIB2 and EAS populations share the closest relationship to the Tianyuan individual, relative to the Yana individuals (Table S6.10-S6.11).

For the EE individuals, we find largely similar relationships across all four Asian-affiliated populations: the Yana individuals form a clade with the EE individuals, or *D(EE; SAMER/NAMER+SIB1/SIB2/EAS; Yana, Mbuti)>0* and *D(Yana, SAMER/NAMER+SIB1/SIB2/EAS; EE, Mbuti)>0* (Z>2.8, Tables S6.8-S6.11). This is consistent with our results using the ancient Asian individual, Tianyuan, in Table 1. We also find that the Yana individuals share more alleles with many (but not all) Asian-affiliated populations (*D(Yana, EE; SAMER/SIB2/EAS, Mbuti)>0*, Z>2.5). The instances where we fail

to reject zero are (1) when *D(Yana, Vestonice16/GoyetQ116-1; NAMER+SIB1, Mbuti)* (0.3<Z<2.2, Table 4), perhaps because the NAMER+SIB1 have their own European connections as we will see when comparing to the RE (Table S6.9), and (2) in some comparisons of *D(Yana, GoyetQ116-1; SIB2/EAS, Mbuti)* (1.9<Z<2.5, Table S6.10), perhaps related to the GoyetQ116-1 individual's connection to the Tianyuan individual[3].

For the ANE, we find that Malta1 and AfontovaGora3 differ some in their relationship to Native Americans and Siberians. The AfontovaGora3 individual forms a clade with the SAMER/NAMER+SIB1 (except the Chukchi where Z=-1.8/-2.3, Tables S6.8-S6.9). However, the SAMER/NAMER+SIB1 groups differ in their relationship to the Malta1 individual. The NAMER+SIB1 show a similar relationship to the Malta1 individual as the Yana individuals do (|Z|<1.4, Table S6.9), while the SAMER populations may show a weak connection to the Malta1 individual relative to the Yana individuals, or *D(Yana, SAMER/NAMER+SIB1; Malta1, Mbuti)* may be less than zero (-2.5<Z<-1.9, Table S6.8), while. The SIB2 populations are mostly similar to the SIB1, but with weaker, more sporadic results for both ANE individuals, suggesting a weaker connection to the ANE. In fact, for the Even/Yakut, we see that *D(Yana, Even/Yakut; Malta1, Mbuti)>0* (Table S6.10), similar to the EAS, who show no connection to the ANE (Table S6.11). We find in all cases that *D(ANE, AMER+SIB+EAS; Yana, Mbuti)>0*, suggesting that the Yana have a stronger connection to the ANE individuals than Asian-affiliated populations have.

The Karelia is an 8,000-year-old individual with stronger connections to Western Eurasians who also possesses ANE ancestry. We find very similar results for the Karelia individual as for the AfontovaGora3 individual, but with stronger connections in comparisons to the Chukchi, Even and Yakut, where connections were weaker for AfontovaGora3 (Tables S6.9-S6.10). It is unclear whether these stronger connections are due to a closer relationship to the Karelia individual, or whether they are due to coincidentally similar amounts of Asian, European or northern Eurasian ancestries in these Siberians and the Karelia individual. For all Native Americans and most Siberians, we observe a stronger connection to the ANE individuals (particularly AfontovaGora3) and the Karelia individual than to the Yana individuals.

For the RE individuals, we find that *D(Yana, SAMER; RE, Mbuti)>0* (Z>2.6, Table S6.8) and *D(RE, SAMER; Yana1, Mbuti)>0* (Z>2.3, Table S6.8), suggesting that the Yana and RE

individuals form a clade with each other, relative to southern Native Americans. However, we note that *D(RE, Anzick-1/Surui/Karitiana; Yana2, Mbuti)* is close to zero, with 1.0<Z<2.9 (Table S6.8), suggesting that these SAMER populations may share more alleles with the Yana2 individual. The NAMER+SIB1 show a very strong connection with most RE individuals, such that *D(Yana, RE; NAMER+SIB1, Mbuti)<0* (Z<-2.5 for almost all RE individuals, Table S6.9). We do not observe the connection to the RE in the SIB2 or EAS populations (Table S6.10-S6.11). These results suggest that the NAMER+SIB1 have a strong connection to recent individuals of European ancestry, perhaps due to recent gene flow.

With present-day East Asians (EAS), we find that the pattern of connections is similar to but slightly different from that observed for the Tianyuan individual in the Table 1. We also observe *D(Yana, Kostenki14/Vestonice16/ElMiron; EAS, Mbuti)>0* (Z>2.3, Table S6.11) and fail to reject *D(Yana, ANE; EAS, Mbuti)=0* (-0.9<Z<1.8, Table 6). However, *D(Yana, GoyetQ116-1; EAS, Mbuti)* may be slightly greater than zero (Z=1.4 and 1.5 for the Tianyuan individual in Table 1, but 1.9<Z<3.1 in Table S6.11), consistent with the EAS individuals showing a weaker or nonexistent connection with GoyetQ116-1[3]. We find similar results for transversions (2.1<Z<2.7) and using the Chimp as an outgroup (2.8<Z<3.3). We also mostly fail to reject *D(Yana, Bichon/Loschbour/ LaBrana1/Hungarian.KO1; EAS, Mbuti)=0* (Table S6.11), whereas *D(Yana, Bichon/Loschbour/LaBrana1/Hungarian.KO1; Tianyuan, Mbuti)>0*, consistent with these West Eurasians showing a connection to present-day East Asians, but not to the Tianyuan individual[3].

Again, we emphasize that the Yana individuals show a connection to Asians, except relative to West Eurasians that show evidence of some Asian ancestry.

Table S6.8 Z-scores for D-statistics comparing the Yana individuals to ancient Eurasians and southern Native Americans

X=Yana1	D(X, Y; Z, Mbuti)				D(X, Z; Y, Mbuti)				D(Y, Z; X, Mbuti)			
Y/Z	**Anzick-1**	**Surui**	**Karitiana**	**Mixe**	**Anzick-1**	**Surui**	**Karitiana**	**Mixe**	**Anzick-1**	**Surui**	**Karitiana**	**Mixe**
UstIshim	5.8	5.7	6.0	6.0	1.0	1.0	1.0	2.4	-4.7	-4.3	-4.5	-3.2
Tianyuan	-1.3	-2.5	-1.6	-1.4	-3.0	-4.1	-3.2	-1.8	-1.7	-1.8	-1.7	-0.4
Kostenki14	4.6	3.4	3.9	3.8	7.6	7.0	7.3	8.6	2.9	3.2	3.3	4.6
GoyetQ116-1	3.2	2.7	4.2	3.0	7.4	7.2	8.5	8.9	4.6	4.9	5.0	6.3
Vestonice16	3.6	2.9	4.2	2.8	7.3	6.5	7.9	8.1	3.6	3.7	3.9	5.4
Malta1	-7.6	-7.9	-9.1	-10.2	-2.5	-2.0	-2.6	-2.0	5.0	5.4	5.8	7.2
AfontovaGora3	-8.9	-10.9	-12.4	-12.9	-6.1	-6.9	-7.0	-6.5	3.2	3.2	3.8	4.8
ElMiron	1.4	2.4	2.3	1.7	4.6	5.8	5.9	6.4	3.2	3.4	3.7	4.8
Villabruna	2.4	2.0	2.3	2.2	3.7	3.5	3.9	5.1	1.3	1.4	1.7	3.0

Bichon	2.3	1.0	1.8	0.6	5.0	4.4	5.1	5.4	2.9	3.5	3.5	5.1
Karelia	-7.8	-9.6	-9.7	-10.7	-4.8	-5.2	-5.4	-4.5	3.2	3.7	3.7	5.7
Loschbour	0.5	-0.5	0.3	-0.4	3.8	3.3	3.9	4.9	3.1	3.8	3.6	5.2
LaBrana1	1.0	0.7	0.7	-0.1	3.9	3.9	4.0	4.8	2.8	3.1	3.3	4.9
Hungarian.KO1	1.3	0.7	0.6	-0.2	3.6	3.3	3.4	4.0	2.3	2.7	2.9	4.4
X=Yana2	**D(X, Y; Z, Mbuti)**				**D(X, Z; Y, Mbuti)**				**D(Y, Z; X, Mbuti)**			
Y/Z	**Anzick-1**	**Surui**	**Karitiana**	**Mixe**	**Anzick-1**	**Surui**	**Karitiana**	**Mixe**	**Anzick-1**	**Surui**	**Karitiana**	**Mixe**
UstIshim	6.3	5.7	6.3	6.4	-0.6	-0.6	-0.7	0.7	-6.5	-5.8	-6.3	-5.0
Tianyuan	-0.6	-2.0	-1.0	-1.0	-2.4	-3.5	-2.6	-1.2	-1.8	-1.6	-1.6	-0.2
Kostenki14	5.4	4.0	4.5	4.4	7.7	7.1	7.5	8.8	2.8	3.3	3.2	4.8
GoyetQ116-1	3.0	3.1	4.5	3.3	6.4	6.2	7.4	7.8	3.4	3.2	3.3	4.7
Vestonice16	4.1	3.1	5.0	3.4	7.2	6.3	7.8	7.9	3.2	3.3	3.2	4.6
Malta1	-7.8	-7.7	-8.8	-10.0	-2.5	-1.9	-2.6	-1.9	4.9	5.5	5.6	7.2
AfontovaGora3	-9.5	-10.9	-12.1	-12.6	-6.6	-7.3	-8.0	-7.5	3.1	3.1	3.4	4.7
ElMiron	1.7	3.1	3.3	2.1	3.7	4.7	4.8	5.3	2.0	2.0	1.9	3.4
Villabruna	2.9	2.4	3.0	2.6	4.3	3.9	4.4	5.5	1.6	1.6	1.6	3.2
Bichon	2.7	1.0	2.1	0.8	4.3	3.3	4.0	4.4	1.6	2.4	2.1	3.8
Karelia	-7.6	-9.1	-9.5	-10.1	-5.4	-5.8	-6.1	-5.2	2.3	3.1	2.9	4.9
Loschbour	0.8	-0.4	0.5	-0.2	3.1	2.6	3.3	4.1	2.1	2.9	2.6	4.2
LaBrana1	1.6	1.1	1.3	0.3	3.3	3.2	3.3	4.2	1.6	2.2	2.1	3.9
Hungarian.KO1	2.0	0.9	1.0	0.1	2.9	2.7	2.8	3.4	1.0	1.8	1.7	3.3

Table S6.9. Z-scores for D-statistics comparing the Yana individuals to ancient Eurasians and the Siberians Aleut, Chukchi and the northern Native Americans Tlingit and Cree

X=Yana1	**D(X, Y; Z, Mbuti)**				**D(X, Z; Y, Mbuti)**				**D(Y, Z; X, Mbuti)**			
Y/Z	**Tlingit**	**Cree**	**Aleut**	**Chukchi**	**Tlingit**	**Cree**	**Aleut**	**Chukchi**	**Tlingit**	**Cree**	**Aleut**	**Chukchi**
UstIshim	7.9	7.8	7.1	6.6	3.0	3.0	2.3	2.5	-3.8	-4.1	-3.8	-4.0
Tianyuan	3.3	0.5	1.3	2.0	2.2	-0.8	0.3	0.7	-0.8	-1.3	-0.9	-1.2
Kostenki14	2.5	2.6	3.2	3.2	7.0	6.9	7.6	7.1	4.7	4.2	4.4	3.7
GoyetQ116-1	0.9	1.6	1.8	1.0	7.2	7.2	7.7	6.9	6.7	5.9	6.4	6.0
Vestonice16	0.3	2.2	1.9	1.8	5.9	6.8	6.6	6.3	5.6	4.9	5.1	4.7
Malta1	-7.8	-8.8	-7.8	-5.1	0.7	-1.4	-0.1	1.2	7.5	6.8	7.0	5.8
AfontovaGora3	-9.3	-11.3	-8.6	-6.1	-3.3	-5.4	-3.2	-1.8	4.9	4.4	4.5	3.8
ElMiron	-3.7	-0.8	-1.3	-2.5	1.8	3.8	3.5	2.2	5.2	4.4	4.8	4.5
Villabruna	-7.9	-1.7	-3.9	-4.2	-4.0	0.6	-1.0	-2.0	3.1	2.2	2.6	2.1
Bichon	-8.7	-3.3	-5.5	-5.2	-2.8	1.5	-0.2	-0.9	5.4	4.6	5.1	4.3
Karelia	-13.0	-11.6	-10.7	-8.5	-6.3	-5.9	-5.0	-4.5	5.7	4.9	5.2	4.2
Loschbour	-10.3	-4.3	-6.2	-6.8	-3.8	0.8	-0.5	-2.3	5.7	4.9	5.3	4.5
LaBrana1	-7.6	-2.5	-4.4	-5.1	-1.8	1.8	0.5	-1.1	5.2	4.3	4.8	4.0
Hungarian.KO1	-9.8	-3.7	-6.1	-6.8	-4.6	0.3	-1.4	-3.2	4.5	3.7	4.5	3.5
X=Yana2	**D(X, Y; Z, Mbuti)**				**D(X, Z; Y, Mbuti)**				**D(Y, Z; X, Mbuti)**			
Y/Z	**Tlingit**	**Cree**	**Aleut**	**Chukchi**	**Tlingit**	**Cree**	**Aleut**	**Chukchi**	**Tlingit**	**Cree**	**Aleut**	**Chukchi**
UstIshim	7.5	7.8	5.9	5.6	1.2	1.2	0.6	0.9	-5.2	-5.5	-4.4	-4.4
Tianyuan	3.2	0.6	0.6	0.8	2.9	-0.1	1.0	1.3	0.0	-0.7	0.5	0.6
Kostenki14	2.5	2.8	2.5	2.2	7.2	7.1	7.8	7.1	5.3	4.6	5.8	5.5
GoyetQ116-1	0.1	1.0	0.7	-0.2	5.9	5.9	6.5	5.4	5.4	4.7	5.5	5.7
Vestonice16	-0.2	2.1	0.9	0.4	5.4	6.8	6.4	6.1	5.6	4.7	5.6	5.7
Malta1	-7.7	-9.2	-9.0	-6.8	0.9	-1.4	0.0	1.4	7.8	7.2	8.1	7.6
AfontovaGora3	-9.9	-12.4	-10.2	-8.5	-3.9	-6.0	-3.8	-2.3	5.5	5.2	5.8	5.8
ElMiron	-4.2	-1.0	-2.3	-3.7	0.7	2.7	2.4	1.0	4.4	3.6	4.4	4.4
Villabruna	-8.2	-1.8	-5.0	-5.8	-3.5	1.2	-0.4	-1.3	4.2	2.9	4.3	4.0
Bichon	-9.0	-3.4	-6.9	-6.8	-4.0	0.4	-1.3	-1.9	4.4	3.5	5.0	4.4
Karelia	-13.0	-11.3	-12.0	-10.2	-7.3	-6.9	-5.6	-5.2	5.3	4.4	5.9	5.5
Loschbour	-10.3	-4.2	-7.7	-8.4	-4.5	0.0	-1.2	-2.7	4.9	4.0	5.3	4.9
LaBrana1	-8.0	-2.6	-5.7	-6.6	-2.5	1.1	-0.2	-1.7	4.7	3.5	5.0	4.5
Hungarian.KO1	-9.9	-3.9	-7.4	-8.3	-5.8	-0.6	-2.3	-3.9	3.6	3.1	4.6	4.0

Table S6.10 Z-scores for D-statistics comparing the Yana individuals to ancient Eurasians and the Siberians: the Sireniki Eskimos, Itelman, Even and Yakut

X=Yana1	D(X, Y; Z, Mbuti)				D(X, Z; Y, Mbuti)				D(Y, Z; X, Mbuti)			
Y/Z	Esk_Sireniki	Itelman	Even	Yakut	Esk_Sireniki	Itelman	Even	Yakut	Esk_Sireniki	Itelman	Even	Yakut
UstIshim	5.9	4.7	6.0	5.0	1.6	1.0	1.9	2.1	-3.8	-3.3	-3.7	-2.6
Tianyuan	-2.6	-3.4	-3.5	-5.2	-3.7	-3.8	-4.4	-4.8	-1.2	-0.5	-0.9	0.2
Kostenki14	5.2	5.2	6.4	5.2	9.2	9.5	10.7	10.6	4.1	4.3	4.4	5.1
GoyetQ116-1	4.1	3.4	3.3	2.6	9.2	9.3	9.2	9.5	5.6	6.3	6.2	7.0
Vestonice16	4.9	4.3	4.3	4.1	9.1	8.9	9.3	9.8	4.5	5.0	5.0	5.7
Malta1	-6.2	-6.1	-1.8	-2.7	0.7	0.7	5.3	5.0	6.5	6.5	6.9	7.3
AfontovaGora3	-7.6	-6.1	-1.9	-2.4	-2.9	-1.2	2.4	2.8	3.9	4.5	4.0	4.8
ElMiron	3.3	1.8	2.6	1.7	7.3	6.5	7.6	7.3	4.2	4.7	4.7	5.3
Villabruna	3.4	3.2	3.2	2.0	5.5	5.5	5.8	5.7	2.3	2.8	2.6	3.7
Bichon	1.4	1.4	1.4	0.5	5.7	5.9	6.1	6.3	4.6	4.7	5.0	5.7
Karelia	-6.7	-5.5	-4.3	-3.6	-1.7	-0.8	1.2	2.6	4.7	4.7	5.2	5.8
Loschbour	1.1	0.6	0.0	-0.6	5.6	5.1	5.2	5.9	4.6	4.9	5.1	5.9
LaBrana1	1.1	0.8	1.5	0.8	5.3	5.0	6.2	6.4	4.2	4.3	4.6	5.3
Hungarian.KO1	1.2	0.5	0.2	-0.7	4.5	4.1	4.4	4.7	3.5	3.9	4.3	5.3
X=Yana2	D(X, Y; Z, Mbuti)				D(X, Z; Y, Mbuti)				D(Y, Z; X, Mbuti)			
Y/Z	Esk_Sireniki	Itelman	Even	Yakut	Esk_Sireniki	Itelman	Even	Yakut	Esk_Sireniki	Itelman	Even	Yakut
UstIshim	5.4	5.0	5.2	4.3	0.0	-0.6	0.1	0.4	-4.8	-5.1	-4.6	-3.5
Tianyuan	-3.4	-2.9	-4.5	-5.9	-3.1	-3.2	-3.7	-4.1	0.0	-0.4	0.6	1.6
Kostenki14	5.0	5.5	5.8	4.5	9.4	9.6	10.8	10.6	4.9	4.3	5.7	6.5
GoyetQ116-1	3.3	3.3	2.3	1.7	8.2	8.3	8.2	8.4	5.0	4.8	5.7	6.6
Vestonice16	3.9	4.6	3.5	3.2	8.6	8.8	9.2	9.5	4.9	4.5	5.9	6.3
Malta1	-7.2	-5.7	-3.0	-3.7	0.9	0.9	5.6	5.3	7.7	6.2	8.3	8.7
AfontovaGora3	-9.0	-6.4	-3.1	-3.3	-3.4	-1.7	2.0	2.3	4.9	4.7	5.2	5.7
ElMiron	2.7	2.4	1.8	0.7	6.1	5.5	6.3	6.1	3.7	3.3	4.6	5.2
Villabruna	2.8	3.4	2.4	1.2	6.0	6.0	6.1	6.0	3.4	3.2	4.1	5.1
Bichon	0.6	1.6	0.2	-0.5	4.6	4.7	5.2	5.2	4.1	3.5	5.0	5.8
Karelia	-7.6	-5.5	-5.5	-4.6	-2.3	-1.4	0.6	2.0	4.9	4.0	5.8	6.4
Loschbour	0.4	0.8	-1.2	-1.5	4.9	4.4	4.3	5.0	4.4	3.8	5.2	6.0
LaBrana1	0.5	1.3	0.7	0.0	4.4	4.4	5.4	5.7	4.1	3.4	5.0	5.8
Hungarian.KO1	0.5	0.8	-0.7	-1.4	4.1	3.5	3.7	4.0	3.5	2.8	4.4	5.3

Table S6.11 Z-scores for D-statistics comparing the Yana individuals to ancient Eurasians and present-day East and Southeast Asians

X=Yana1	D(X, Y; Z, Mbuti)				D(X, Z; Y, Mbuti)				D(Y, Z; X, Mbuti)			
Y/Z	Ami	Dai	Han	Japanese	Ami	Dai	Han	Japanese	Ami	Dai	Han	Japanese
UstIshim	2.6	3.6	3.5	4.1	-0.1	1.1	1.2	1.1	-2.5	-2.2	-2.2	-2.8
Tianyuan	-6.9	-7.9	-8.0	-6.6	-6.1	-6.2	-6.4	-6.1	0.4	1.3	1.2	0.4
Kostenki14	5.5	5.4	5.3	6.5	11.0	11.6	11.3	11.9	5.5	6.2	6.1	5.4
GoyetQ116-1	3.1	2.6	2.8	3.1	10.0	10.5	10.3	10.3	7.4	8.2	7.9	7.4
Vestonice16	5.7	5.4	5.1	5.7	11.2	12.3	11.8	11.3	6.0	7.0	6.8	6.1
Malta1	0.3	-0.4	-0.1	0.6	7.8	8.0	8.4	8.3	7.5	8.5	8.4	7.7
AfontovaGora3	0.9	0.7	1.0	1.8	5.9	6.9	6.9	6.9	5.0	6.1	5.8	5.1
ElMiron	3.1	3.4	3.2	3.0	8.6	10.3	9.5	8.8	5.2	6.7	6.4	5.9
Villabruna	4.5	4.7	4.3	4.6	8.1	9.0	8.7	8.2	3.8	4.6	4.5	3.8
Bichon	2.1	2.5	1.7	3.0	7.8	9.1	8.4	8.6	6.0	6.8	7.0	6.1

Karelia	-0.4	-0.3	-1.4	-0.4	6.1	7.1	5.8	5.9	6.4	6.9	6.9	6.2
Loschbour	1.5	2.4	0.9	2.1	7.8	9.3	7.9	8.3	6.1	6.9	6.9	6.1
LaBrana1	2.4	2.8	1.7	3.5	8.2	9.9	8.4	9.2	5.7	6.6	6.4	5.6
Hungarian.KO1	1.8	2.9	1.4	3.0	7.0	8.9	7.3	8.0	5.5	6.0	6.0	5.2
X=Yana2	**D(X, Y; Z, Mbuti)**				**D(X, Z; Y, Mbuti)**				**D(Y, Z; X, Mbuti)**			
Y/Z	**Ami**	**Dai**	**Han**	**Japanese**	**Ami**	**Dai**	**Han**	**Japanese**	**Ami**	**Dai**	**Han**	**Japanese**
UstIshim	2.0	3.1	3.0	3.3	-1.6	-0.6	-0.5	-0.6	-3.6	-3.4	-3.2	-3.6
Tianyuan	-7.9	-8.6	-8.9	-7.7	-5.4	-5.6	-5.7	-5.4	1.7	2.2	2.5	1.8
Kostenki14	5.2	5.4	4.9	5.9	10.7	11.6	11.4	11.7	6.5	7.2	7.3	6.7
GoyetQ116-1	2.5	2.3	1.9	2.4	9.1	9.5	9.6	9.2	6.6	7.2	7.4	6.8
Vestonice16	5.0	5.4	4.7	5.0	10.7	11.8	11.6	11.3	6.3	6.9	7.4	6.7
Malta1	-0.2	-0.9	-0.9	-0.3	8.1	8.5	8.9	8.5	8.2	9.3	9.4	8.6
AfontovaGora3	0.3	0.4	-0.1	0.8	5.4	6.5	6.5	6.4	5.3	6.4	6.7	5.9
ElMiron	2.4	3.3	2.6	2.3	7.5	8.9	8.5	7.7	4.9	5.8	6.1	5.6
Villabruna	3.8	4.5	3.7	3.7	8.4	9.6	9.0	8.5	5.0	5.6	6.0	5.3
Bichon	1.5	2.0	1.0	2.0	6.8	8.0	7.4	7.5	5.5	6.4	6.5	5.9
Karelia	-1.0	-0.9	-2.3	-1.7	5.4	6.3	5.3	5.2	6.2	7.1	7.3	6.7
Loschbour	0.9	1.8	0.1	1.1	6.9	8.5	7.1	7.2	5.9	6.7	6.8	6.2
LaBrana1	2.0	2.6	1.2	2.7	7.3	9.0	7.7	8.2	5.5	6.4	6.6	5.8
Hungarian.KO1	1.4	2.5	0.9	2.0	6.6	8.3	6.9	7.4	4.9	5.7	5.8	5.2

No differentiation in southern Native Americans relative to the Yana individuals?

We compared southern Native Americans to each other, using *D(SAMER, SAMER; Yana, Mbuti)* (Table S6.12). Because some South American populations (Surui, Karitiana and Chane) show a closer relationship to the Tianyuan individual and the Onge and Papuan[3,12], we wanted to test whether an asymmetric relationship would appear relative to the Yana individuals.

We do not find clear evidence for any closer or further relationship to South Americans. However, we note that while we fail to reject *D(Pima/Mixe, SAMER; Yana1, Mbuti)=0* as we do for other combinations, the Z-scores are much lower, such that -2.8<Z<-1.9, suggesting that all other southern Native Americans share more alleles with the Yana1 individual than the Mixe and Pima share. We do not find as strong a trend for the Yana2 individual, but we note that *D(Mixe, Anzick-1/Karitiana; Yana2, Mbuti)* might be less than zero, with Z=-2.7 and -2.8. While these results may hint at unique population ancestry in the Mixe and Pima, rather than in southern Native Americans, we do not find more significant results using transversions (-2.9<Z<0.4) or the chimp as the outgroup (-3.0<Z<0.2), so it is difficult to determine whether this pattern is significant.

Table S6.12 Z-scores for D-statistics comparing southern Native Americans to the Yana individuals

	D(P1, P2; Yana1, Mbuti)										D(P1, P2; Yana2, Mbuti)									
P1/P2	**Anzick-1**	**Pima**	**Mayan**	**Mixe**	**Zapotec**	**Piapoco**	**Karitiana**	**Surui**	**Quechua**	**Chane**	**Anzick-1**	**Pima**	**Mayan**	**Mixe**	**Zapotec**	**Piapoco**	**Karitiana**	**Surui**	**Quechua**	**Chane**
Anzick-1	nan	2.3	0.3	2.4	0.2	0.8	0.3	0.5	0.6	0.4	nan	1.9	1.7	**2.7**	1.2	2.3	0.5	0.8	1.5	1.3
Pima	-2.3	nan	**-2.6**	-0.1	**-2.7**	-1.9	-2.3	-2.0	-2.3	-2.0	-1.9	nan	-0.3	0.8	-0.9	0.5	-1.8	-1.1	-0.6	-0.4
Mayan	-0.3	**2.6**	nan	**2.7**	-0.1	0.6	0.0	0.2	0.3	0.1	-1.7	0.3	nan	1.2	-0.6	0.8	-1.5	-0.9	-0.4	-0.2
Mixe	-2.4	0.1	**-2.7**	nan	**-2.8**	-2.0	-2.5	-2.1	-2.4	-2.0	**-2.7**	-0.8	-1.2	nan	-1.8	-0.3	**-2.8**	-1.9	-1.7	-1.1
Zapotec	-0.2	**2.7**	0.1	**2.8**	nan	0.7	0.1	0.3	0.5	0.2	-1.2	0.9	0.6	1.8	nan	1.3	-0.9	-0.3	0.3	0.3
Piapoco	-0.8	1.9	-0.6	2.0	-0.7	nan	-0.6	-0.3	-0.3	-0.4	-2.3	-0.5	-0.8	0.3	-1.3	nan	-2.4	-1.6	-1.2	-0.8
Karitiana	-0.3	2.3	0.0	2.5	-0.1	0.6	nan	0.3	0.3	0.1	-0.5	1.8	1.5	**2.8**	0.9	2.4	nan	0.5	1.2	1.1
Surui	-0.5	2.0	-0.2	2.1	-0.3	0.3	-0.3	nan	0.1	-0.1	-0.8	1.1	0.9	1.9	0.3	1.6	-0.5	nan	0.6	0.6
Quechua	-0.6	2.3	-0.3	2.4	-0.5	0.3	-0.3	-0.1	nan	-0.1	-1.5	0.6	0.4	1.7	-0.3	1.2	-1.2	-0.6	nan	0.1
Chane	-0.4	2.0	-0.1	2.0	-0.2	0.4	-0.1	0.1	0.1	nan	-1.3	0.4	0.2	1.1	-0.3	0.8	-1.1	-0.6	-0.1	nan

qpAdm tests of admixture

We used *qpWave* and *qpAdm* to test whether Native Americans or Siberians can be described as a mixture of an Asian population (Han or Tianyuan) and a population represented by the Yana individuals (Table S6.13). We begin with an outgroup set of Mbuti, Ju_hoan_North and Yoruba from Africa, and the Ust'-Ishim, Kostenki14, Tianyuan and Papuan from Eurasia and Oceania (**Out**), and when using the Tianyuan individual as a source population, we drop the Tianyuan individual from the outgroup set (**Out-T**).

Using the East Asian Han as a source population, we generally find that the Yana individuals can be used just as readily as the Malta1, AfontovaGora3 (AG3) and Karelia individuals, as well as the Vestonice16 individual. Surprisingly, we find the highest mixture proportions for the two Yana individuals (00_f2 ranges from 0.18 to 0.32, Table S6.13), and the lowest for AG3 (00_f2 ranges from 0.04 to 0.08). This is unusual, as our D-statistics above suggest that Native Americans share a closer relationship to the Malta1, AG3 and Karelia than to the Yana individuals. We also find here that the Mixe do not always satisfy the condition of rank1>0.05, meaning in some instances, at least three ancestral components are needed to describe the Han, European-affiliated individual, and the Mixe. We do not observe this for the Surui, and this may also be related to the Mixe/Surui differences seen above and in Skoglund et al.[12] and Yang et al.[3].

Using the Tianyuan individual as a source population, we find similar patterns as for the Han, but smaller amounts of estimated admixture into Siberians and Native Americans. This is

consistent with the Asian ancestry found in these recent Siberians and Native Americans being more closely related to present-day East Asians than the Tianyuan individual[3]. Here, the Mixe can be described as a mixture of components related to the Tianyuan individual and ancient northern/western Eurasians, while the Surui cannot. Furthermore, the Siberians (particularly the Itelman) does not need to be described as a mixture of West and East Eurasian components, but rather only a component related to the Tianyuan individual. While we find these slight variations, overall, we observe that the Yana individual can be used interchangeably with ancient northern and western Eurasians in testing for admixture into the Americas.

We also considered admixture into the ancient Siberians: the Yana, Malta1 and AG3 individuals (Table S6.14). Here, to consider more ancient Eurasians as potential sources, we removed both the Tianyuan and Kostenki14 individuals from the outgroup set (**Out-TK**). We find that the ancient northern Eurasians and the Yana individuals can be described as a mixture of a population related to the Han and a population related to early West Eurasians, the Kostenki14, GoyetQ116-1 and Vestonice16 individuals. We also find that the Yana individuals show a lower mixture proportion from early West Eurasians (00_f2 ranges from 0.45-0.65, Table S6.14) than the Malta1 and AG3 individuals show (00_f2 ranges from 0.55-0.74, Table S6.14). This may be related to the unusual results for Table 8 as well.

Part of the difficulty may be related to complex scenarios of ancient structure in Siberia that are not easily reflected in a simple test of admixture. Overall, though, we observe that the Yana individuals behave relatively similarly to the Malta1 and AG3 individuals, with respect to early West Eurasians, but they do not come from the same population to which the Malta1 and AG3 individuals would later be related.

Table S6.13 *qpWave* and *qpAdm* results for admixture into Native Americans using the form (S1=Han, S2; Target), where conditions satisfying admixture are when rank1>0.05 and 01_pnest<0.05. 00_f2 is the mixture proportion from the S2 population into the Target population.

(Han, S2; Target), Out											
S2	**Target**	**rank1**	**00_f2**	**sd**	**01_pnest**	**S2**	**Target**	**rank1**	**00_f2**	**sd**	**01_pnest**
Yana1	**Mixe**	0.01	0.32	0.04	2.40E-21	**Yana1**	**Surui**	0.29	0.26	0.04	8.80E-13
Yana2	**Mixe**	0.004	0.31	0.04	4.50E-21	**Yana2**	**Surui**	0.44	0.26	0.04	3.40E-13
Malta1	**Mixe**	0.02	0.23	0.03	5.90E-21	**Malta1**	**Surui**	0.4	0.19	0.03	1.40E-12
AG3	**Mixe**	0.79	0.08	0.01	1.80E-18	**AG3**	**Surui**	0.18	0.06	0.01	9.30E-10
Karelia	**Mixe**	0.19	0.25	0.03	7.10E-22	**Karelia**	**Surui**	0.05	0.2	0.03	2.90E-11
Vestonice16	**Mixe**	0.05	0.13	0.02	2.70E-22	**Vestonice16**	**Surui**	0.24	0.1	0.02	4.10E-13
Yana1	**Itelman**	0.09	0.19	0.04	1.80E-07	**Yana1**	**Esk_Sireniki**	0.13	0.21	0.03	3.40E-12
Yana2	**Itelman**	0.04	0.18	0.04	4.40E-07	**Yana2**	**Esk_Sireniki**	0.18	0.21	0.03	1.80E-12
Malta1	**Itelman**	0.11	0.14	0.03	1.40E-07	**Malta1**	**Esk_Sireniki**	0.46	0.16	0.02	3.90E-13
AG3	**Itelman**	0.3	0.04	0.01	1.30E-06	**AG3**	**Esk_Sireniki**	0.97	0.05	0.01	2.90E-11
Karelia	**Itelman**	0.22	0.15	0.03	1.00E-07	**Karelia**	**Esk_Sireniki**	0.73	0.17	0.02	6.10E-13
Vestonice16	**Itelman**	0.1	0.08	0.02	6.00E-08	**Vestonice16**	**Esk_Sireniki**	0.17	0.09	0.01	2.10E-14
(Tianyuan, S2; Target), Out-T											
S2	**Target**	**rank1**	**00_f2**	**sd**	**01_pnest**	**S2**	**Target**	**rank1**	**00_f2**	**sd**	**01_pnest**
Yana1	**Mixe**	0.36	0.23	0.06	2.80E-04	**Yana1**	**Surui**	0.03	0.19	0.07	5.00E-03
Yana2	**Mixe**	0.28	0.22	0.06	3.80E-04	**Yana2**	**Surui**	0.03	0.18	0.06	4.20E-03
Malta1	**Mixe**	0.19	0.17	0.05	4.50E-04	**Malta1**	**Surui**	0.01	0.14	0.05	5.90E-03
AG3	**Mixe**	0.41	0.05	0.02	3.60E-03	**AG3**	**Surui**	0.02	0.03	0.02	1.10E-01
Karelia	**Mixe**	0.14	0.17	0.05	1.60E-03	**Karelia**	**Surui**	0	0.11	0.06	5.20E-02
Vestonice16	**Mixe**	0.32	0.1	0.02	6.90E-05	**Vestonice16**	**Surui**	0.01	0.08	0.03	2.90E-03
Yana1	**Itelman**	0.8	0.1	0.06	1.20E-01	**Yana1**	**Esk_Sireniki**	0.99	0.12	0.06	6.30E-02
Yana2	**Itelman**	0.67	0.09	0.06	2.00E-01	**Yana2**	**Esk_Sireniki**	0.99	0.11	0.06	6.40E-02
Malta1	**Itelman**	0.52	0.07	0.05	1.30E-01	**Malta1**	**Esk_Sireniki**	0.88	0.09	0.04	4.70E-02
AG3	**Itelman**	0.99	0.02	0.02	2.50E-01	**AG3**	**Esk_Sireniki**	0.67	0.02	0.02	1.30E-01
Karelia	**Itelman**	0.53	0.06	0.05	2.60E-01	**Karelia**	**Esk_Sireniki**	0.92	0.09	0.05	9.70E-02
Vestonice16	**Itelman**	0.84	0.05	0.03	7.70E-02	**Vestonice16**	**Esk_Sireniki**	0.95	0.06	0.02	1.90E-02

Table S6.14. *qpWave* and *qpAdm* results for admixture into ancient Siberians using the form (S1=Han, S2; Target) and the outgroup set Out-TK, where conditions satisfying admixture are when rank1>0.05 and 01_pnest<0.05. 00_f2 is the mixture proportion from the S2 population into the Target population.

S2	**Target**	**rank1**	**00_f2**	**sd**	**10_pnest**	**S2**	**Target**	**rank1**	**00_f2**	**sd**	**10_pnest**
Yana2	**Yana1**	0.12	NA	NA	NA	**Yana1**	**Yana2**	0.12	1.00	0.11	0.97
Malta1	**Yana1**	0.82	0.87	0.13	0.36	**Malta1**	**Yana2**	0.88	0.87	0.13	0.36
AG3	**Yana1**	0.03	0.61	0.42	0.51	**AG3**	**Yana2**	0.85	0.56	0.30	0.37
Karelia	**Yana1**	0.04	0.83	0.10	0.12	**Karelia**	**Yana2**	0.25	0.85	0.10	0.15
Tianyuan	**Yana1**	0.29	NA	NA	NA	**Tianyuan**	**Yana2**	0.55	NA	NA	NA
Kostenki14	**Yana1**	0.54	0.65	0.07	8.2E-05	**Kostenki14**	**Yana2**	0.90	0.64	0.07	1.9E-05
GoyetQ116-1	**Yana1**	0.27	0.56	0.10	4.7E-03	**GoyetQ116-1**	**Yana2**	0.20	0.57	0.11	0.01
Vestonice16	**Yana1**	0.68	0.45	0.07	6.4E-07	**Vestonice16**	**Yana2**	0.45	0.47	0.07	1.8E-06
Yana1	**Malta1**	0.82	NA	NA	NA	**Yana1**	**AG3**	0.03	NA	NA	NA
Yana2	**Malta1**	0.88	NA	NA	NA	**Yana2**	**AG3**	0.85	NA	NA	NA
AG3	**Malta1**	0.14	NA	NA	NA	**Malta1**	**AG3**	0.14	0.95	0.14	0.71
Karelia	**Malta1**	0.06	0.94	0.11	0.54	**Karelia**	**AG3**	0.95	0.85	0.10	0.14
Tianyuan	**Malta1**	0.30	NA	NA	NA	**Tianyuan**	**AG3**	0.41	NA	NA	NA
Kostenki14	**Malta1**	0.98	0.72	0.08	1.8E-03	**Kostenki14**	**AG3**	0.12	0.74	0.09	0.01
GoyetQ116-1	**Malta1**	0.80	0.69	0.13	0.08	**GoyetQ116-1**	**AG3**	0.01	0.71	0.11	0.01
Vestonice16	**Malta1**	0.75	0.55	0.08	1.1E-04	**Vestonice16**	**AG3**	0.10	0.65	0.08	6.3E-05

6.3 Ancient Paleosiberians (Kolyma1)

Affinities with Eurasians and Native Americans

Quantifying the genetic affinities of the Kolyma1 individual using outgroup-f_3 statistics revealed the highest amount of shared drift with both ancient and present-day groups from northeastern Siberia and the Americas (Table S6.15; Fig. S6.8).

Table S6.15 Top outgroup-f_3 statistics for Kolyma1 (HO 1240K panel)

f_3(Mbuti;Kolyma1,X)				
X	**Group**	**f_3**	**SE**	**N SNPs**
M0831	Magadan_BA	0.324	0.005	368,069
Koryak_WGS	Koryak_WGS	0.322	0.004	594,255
Saqqaq	Saqqaq_PE	0.317	0.005	211,251
Athabascan_WGS	Athabascan_WGS	0.316	0.004	579,988
Eskimo_Yupik_WGS	Eskimo_Yupik_WGS	0.316	0.004	457,272
M9984_2	Magadan_BA	0.315	0.005	431,735
NEO241	Ekven_IA	0.313	0.005	142,724
NEO251	Ekven_IA	0.313	0.005	378,584
NEO242	Ekven_IA	0.313	0.005	103,965
Anzick-1	Clovis_LP	0.311	0.005	376,228

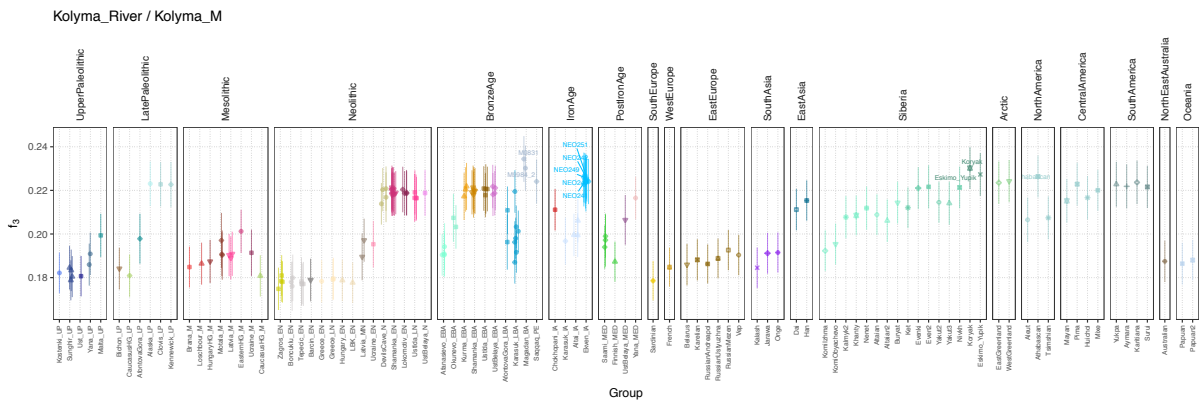


Figure S6.8. Outgroup-f_3 statistics for Kolyma1 individual (WGS panel, no transitions).

Using f_4 statistics, we further observed that Kolyma1 has increased affinity to Mal'ta compared to EEA, similarly to Native Americans who have been shown to derive from a mixture of ANS and EEA (Extended Data Table 2). Consequently, when attempting to fit Kolyma1 on the "eastern" admixture graph including ANS (Fig. S6.3), we found that it is best modelled as a mixture of ANS and EEA, with an estimated 28% contribution from ANS (max |Z| = 3.22) (Figure S6.9).

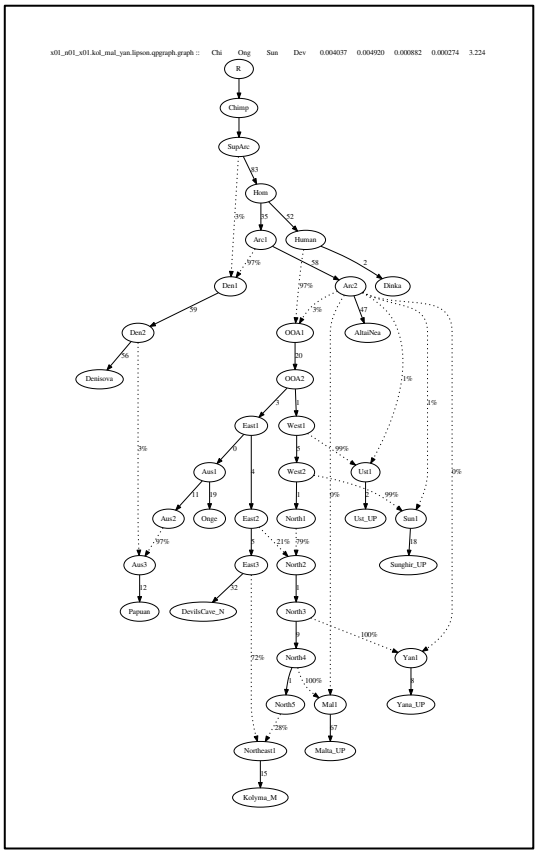


Figure S6.9 Admixture graph fitting of Kolyma1

We then added the ancient Beringians (USR1) to investigate the relationship between Kolyma1 and ancestral Native Americans. A fit of USR1 as a sister group of Kolyma1 deriving from the same mixture event resulted in poor fit (max |Z| = 6.11), with the worst *f*-statistics indicating a lack of shared drift between Kolyma1 and EEA (Fig. S6.10 left). Modelling USR1 as an independent ANS / EEA mixture on the other hand was successful (max |Z| = 3.41) with an increased ANS contribution compared to Kolyma1 (40% vs 27%) (Fig. S6.10 right). Finally, the 12.6 kya Clovis individual which had previously been shown to belong to the "southern branch" lineage of Native Americans[13] was fitted as sister group to USR1 (Fig. 2), consistent with a single founding population of Native Americans[14].

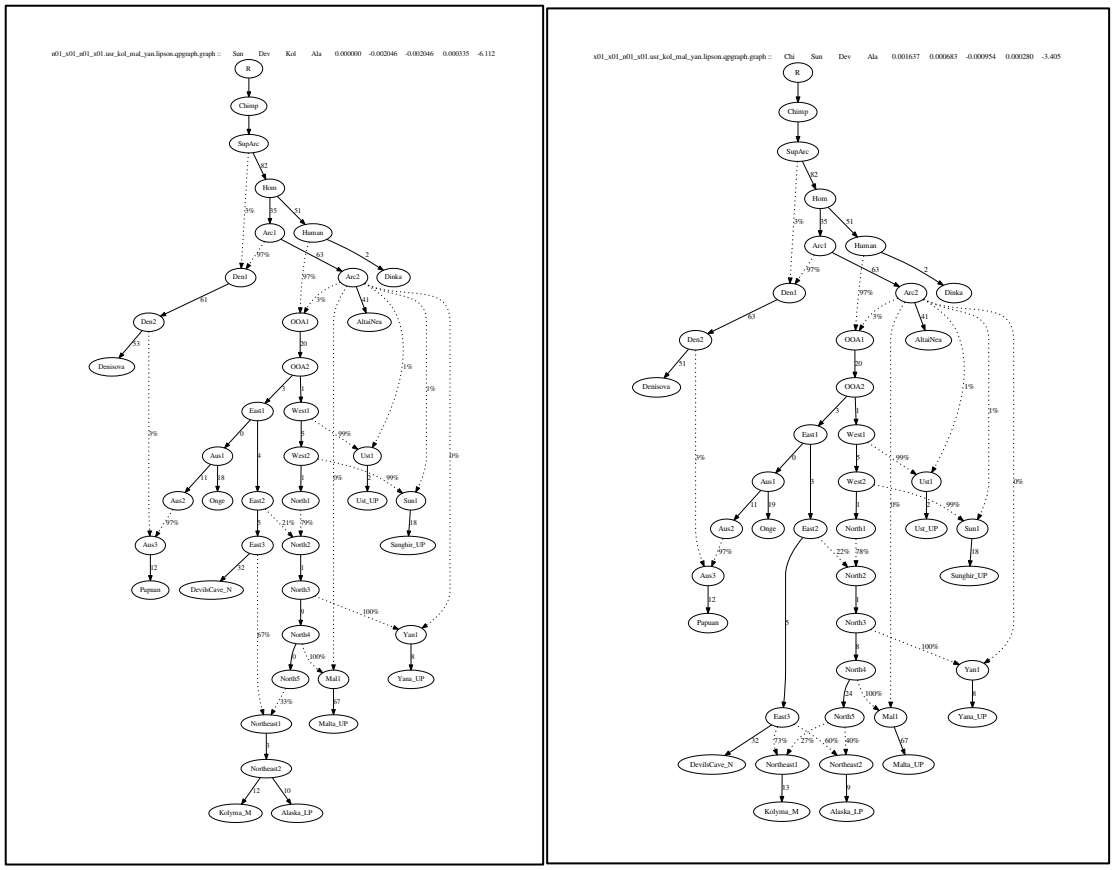


Figure S6.10 Admixture graph fitting of USR1 (left) failed fit as a clade with Kolyma1. (right) successful fit as independent ANS/EEA mixture.

Mixture proportions inferred using *qpAdm* using the set of ten outgroups described above were in line with the admixture graph results. We found that Kolyma1, USR1 and Clovis were successfully modelled as two-way mixtures of ANS (Mal'ta) and EEA (DC), with a higher ANS contribution to ancestral Native Americans than to Kolyma1 (Table S6.16).

Table S6.16 *qpAdm* modelling of Kolyma1, Saqqaq and ancient Native American individuals

testPop	refPop	alpha	se	nSnps	chisq	pval	pvalNested
Alaska_LP	Malta_UP	0.33	0.03	2,640,944	7.83	4.50E-01	1.14E-22
	DevilsCave_N	0.67	0.03				
Clovis_LP	Malta_UP	0.31	0.03	2,105,096	5.56	6.96E-01	3.38E-17
	DevilsCave_N	0.69	0.03				
Kolyma_M	Malta_UP	0.29	0.03	2,619,289	11.59	1.71E-01	4.47E-17
	DevilsCave_N	0.72	0.03				
Saqqaq_PE	Malta_UP	0.15	0.04	1,751,589	8.10	4.24E-01	2.48E-04
	DevilsCave_N	0.85	0.04				

The outgroup-f_3 statistics also showed a strong affinity of Kolyma1 with Paleoeskimos, represented by the 4kya Saqqaq individual from Greenland, so we further investigated their relationship. We found that Saqqaq can be modelled as a two-way mixture of ANS and EEA similarly to Native Americans and Kolyma1, albeit with the lowest amount of ANS ancestry among them (15 ± 4%; Table S6.16). Recently, it has been suggested that the Siberian ancestry observed in present-day Athabascan speakers link reflects gene flow from a Paleoeskimo source represented by Saqqaq[15], but another recent study found evidence for a ghost source population more closely related to Koryaks[14]. We observed a high level of shared genetic drift between Athabascans and Kolyma1, opening up the possibility that AP represented by Kolyma1 might be a good source for this Siberian gene flow. Using *qpAdm* and a set of twelve ougtgroups (Mbuti, Yoruba, Mota_N, Onge, Papuan, Ust_UP, Kostenki_UP, Sunghir_UP, Zagros_EN, Yana_UP, Alaska_LP, Karitiana), we confirm Siberian gene flow into Athabascans, which are formed as a 2-way mixture of Native Americans (Clovis_LP) and either Kolyma1 (31.1 ± 3%; %; p = 0.46) or Saqqaq (29.3 ± 3%; p = 0.31). We further tried to fit Athabascans onto an admixture graph fit including Native Americans, Kolyma1 and Saqqaq, and found that a Siberian gene flow source related to Kolyma1 provided a slightly better fit than one related to Saqqaq (max |Z| = 3.04 vs max |Z| = 3.20; Fig. S6.11), as well as the latter resulting in multiple branches with length zero along the admixture path. Together with the haplotype sharing results discussed in the main text (Extended Data Fig. 5) our results therefore suggest that the Siberian ancestry in Athabascans does not predominantly derive from Paleoeskimo-related gene flow.

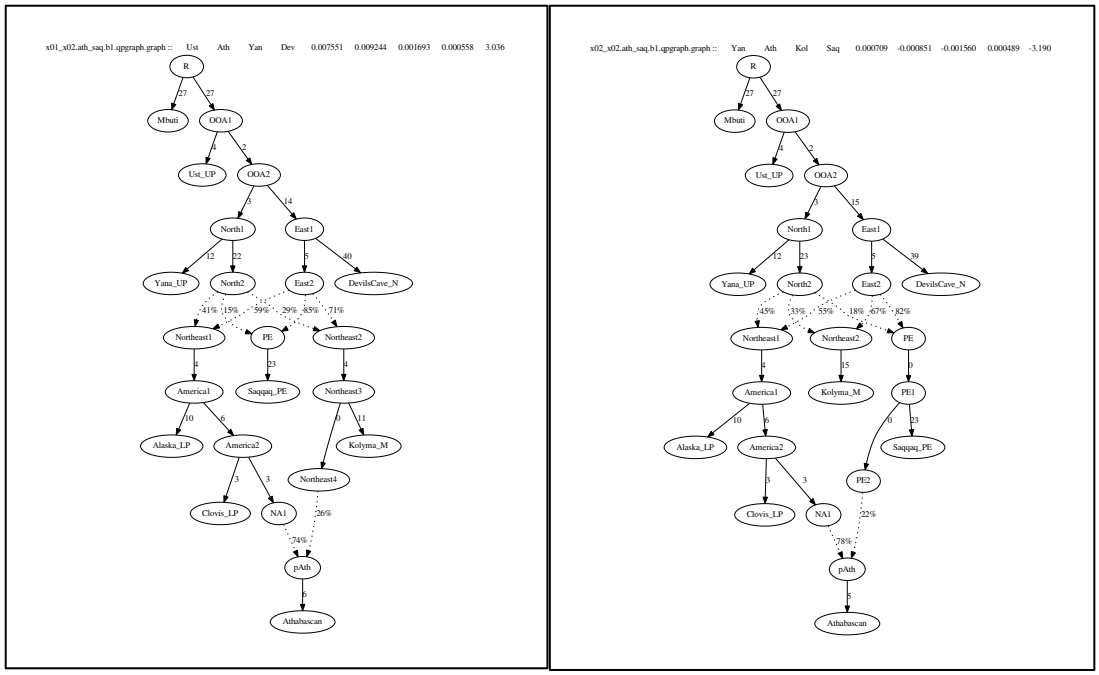


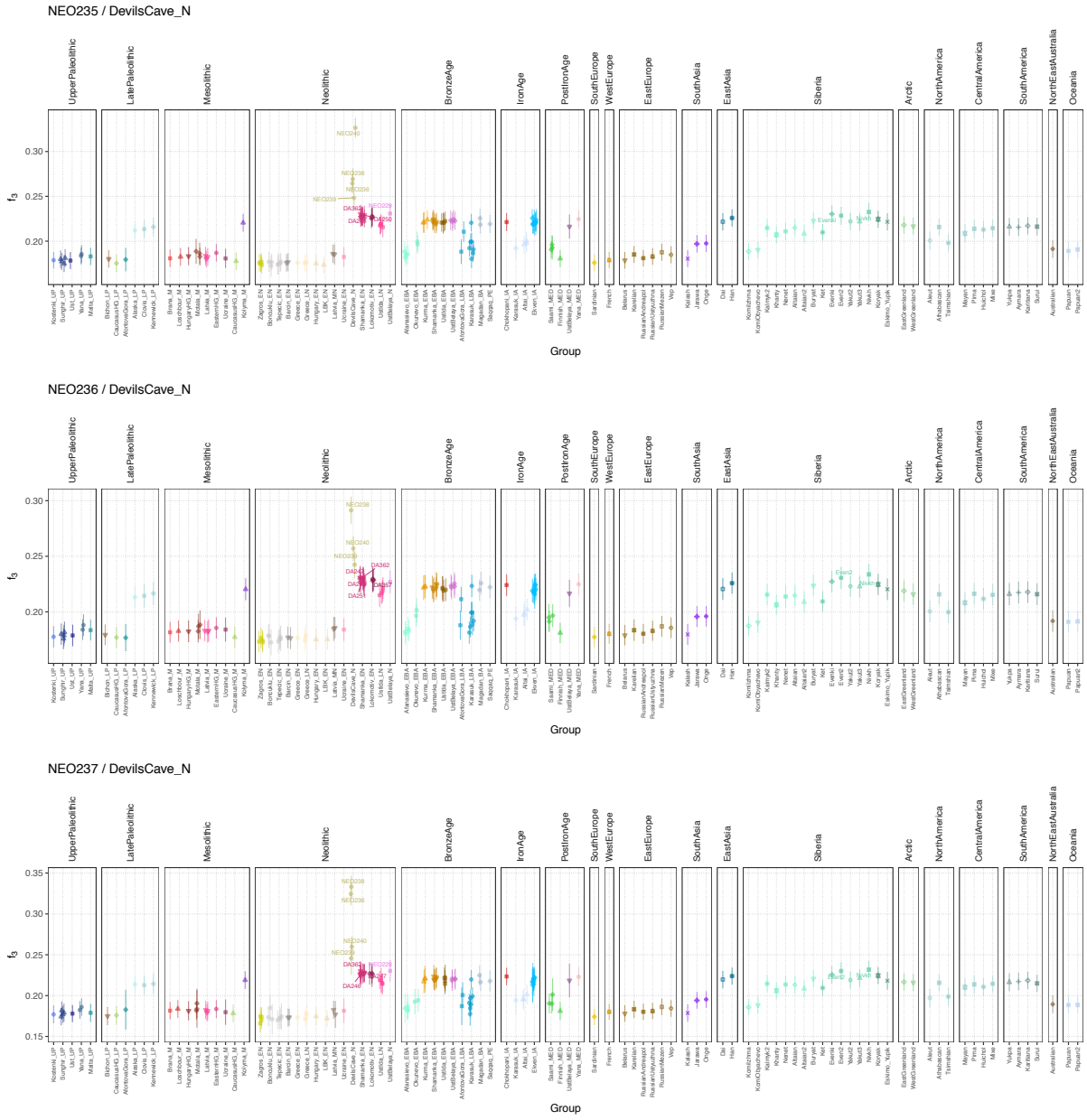
Figure S6.11 Admixture graph fitting of Athabascans (left) Siberian gene flow related to Kolyma1. (right) Siberian gene flow related to Saqqaq.

6.4 Devil's Gate Cave

As part of this study we generated new genomic data from six early Neolithic hunter-gatherer individuals from Devil's Gate Cave, a site in in Primorskoye, northern East Asia. Two individuals from the same site had been sequenced to low coverage in a previous study (labelled here DC1, DC2)[16], so we used outgroup-f_3 statistics to compare genetic affinities between those individuals and our newly generated data. In Table S6.17 we show the top f_3 statistics with modern populations for NEO240 (the individual with the highest coverage in our dataset) as well as the two previously published individuals. We find highly concordant results for the different individuals, with the highest amount of drift shared with Ulchi and other Tungusic-speaking populations from the region (Table S6.17; Fig. S6.12).

Table S6.17 Highest outgroup-f_3 statistics for Devil's Gate cave individuals with modern populations (HO 1240K panel)

f_3(Mbuti;NEO240,X)				f_3(Mbuti;DC1,X)				f_3(Mbuti;DC2,X)			
X	f_3	SE	N SNPs	X	f_3	SE	N SNPs	X	f_3	SE	N SNPs
Ulchi	0.302	0.004	207,623	Ulchi	0.299	0.006	19,354	Ulchi	0.313	0.010	7,775
Oroqen	0.296	0.004	206,769	Oroqen	0.294	0.006	19,282	Nganasan	0.307	0.010	7,731
Hezhen	0.295	0.004	206,523	Hezhen	0.293	0.006	19,265	Oroqen	0.306	0.010	7,750
Japanese	0.295	0.004	207,767	Japanese	0.293	0.006	19,361	Japanese	0.305	0.010	7,780
Daur	0.295	0.004	206,853	Korean	0.292	0.006	19,227	Hezhen	0.305	0.010	7,747
Korean	0.294	0.004	205,683	Nganasan	0.291	0.007	19,167	Xibo	0.304	0.010	7,747
Xibo	0.293	0.004	206,262	Daur	0.290	0.006	19,287	Yukagir_Tundra	0.303	0.010	7,758
Yukagir_Tundra	0.292	0.004	206,935	Eskimo_Chaplin	0.289	0.007	18,954	Daur	0.302	0.010	7,754
Nganasan	0.292	0.004	205,422	Eskimo_Sireniki	0.289	0.007	19,032	Eskimo_Chaplin	0.302	0.010	7,622
Han_NChina	0.291	0.004	206,859	Yukagir_Tundra	0.289	0.006	19,294	Han_NChina	0.301	0.010	7,752



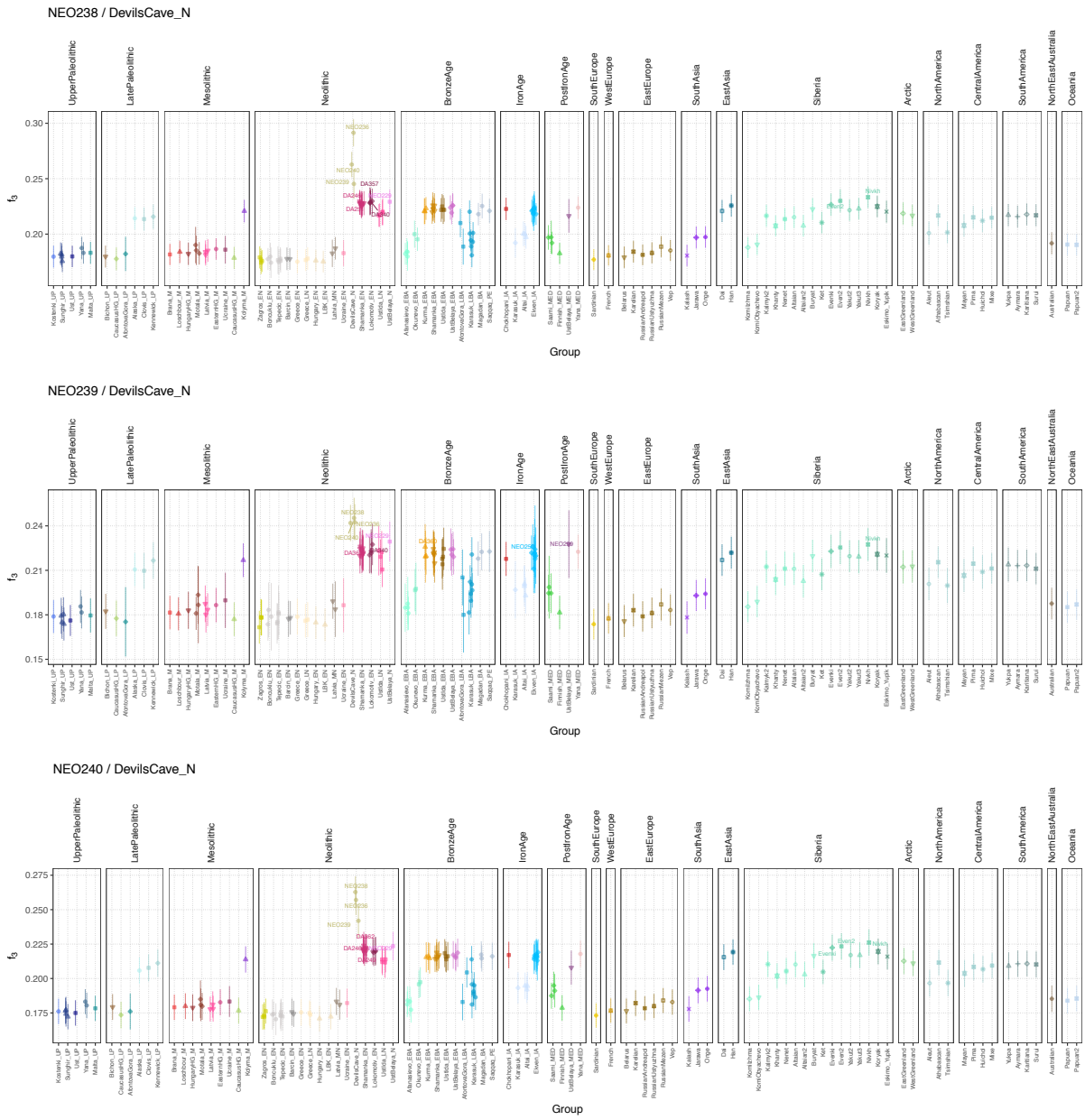


Figure S6.12. Outgroup-f_3 statistics for Devil's Gate cave individuals (WGS panel, no transitions).

We next investigated the genetic affinities between the two published individuals and our newly sequenced individuals. We found that both DC1 and DC2 shared most genetic drift with each other, followed by individual NEO237 from our dataset. The surprisingly high amount of shared drift among only those three individuals indicates the possibility that the genomic data for DC1 and DC2 originates from the same individual (NEO237 in our dataset), but we caution that these results are based on a limited number of SNPs due to the low coverage of the published individuals (Table S6.18).

Table S6.18 Highest outgroup-f_3 statistics for published Devil's Gate cave individuals with other ancient individuals (1240K panel)

f_3(Mbuti;DC1,X)				f_3(Mbuti;DC2,X)			
X	f_3	SE	N SNPs	X	f_3	SE	N SNPs
DC1	0.611	0.035	1,126	DC2	0.611	0.035	1,126
NEO237	0.598	0.022	2,883	NEO237	0.555	0.015	6,946
NEO238	0.490	0.014	6,949	NEO238	0.475	0.010	16,861
NEO236	0.470	0.011	14,270	NEO236	0.461	0.008	34,892
NEO240	0.394	0.011	9,453	NEO240	0.381	0.008	22,929

6.5 Four-way modelling of Holocene Siberians

We used *qpAdm* to investigate the genetic makeup of Holocene Siberians and their relationship to the earlier inhabitants of the region. For this analysis, we used a set of twelve outgroups (Mbuti, Yoruba, Mota_N, Onge, Papuan, Ust_UP, Kostenki_UP, Sunghir_UP, Zagros_EN, Yana_UP, Alaska_LP, Karitiana) which include two Native American groups, in order to be able to infer more recent Native American gene flow into the target populations. The target populations were modelled as all possible 2-way, 3-way and 4-way mixtures of the following source populations

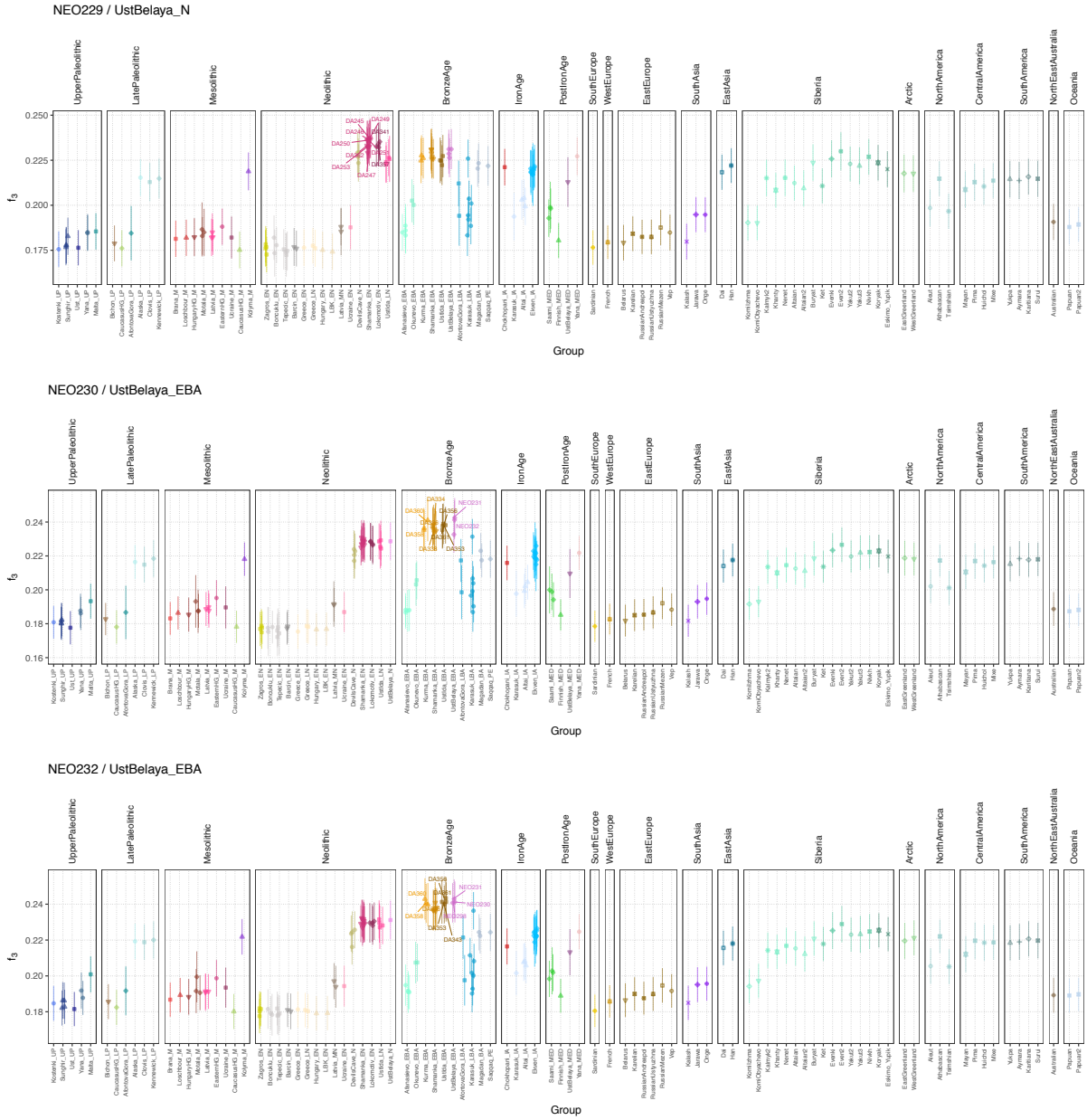
- Kolyma_M (Ancient Paleosiberian)
- DevilsCave_N (Early East Asian)
- Clovis_LP (Native American)
- EasternHG_M (Eastern hunter-gatherer)
- Afanasievo_EBA (Steppe ANE)
- Bichon_LP (Western hunter-gatherer)
- Barcin_EN (Anatolian Neolithic farmer)

Below we discuss some key results for various groups, results for all possible source population combinations are shown in Extended Data Fig. 7 and Supplementary Data Table 4.

<u>Ust'Belaya</u>

Individuals at Ust'Belaya stem from three different time periods, each of which shows a different genetic makeup. The earliest individual is successfully modelled as a 2-way mixture of predominantly EEA ancestry, with only a small fraction of another ancestry component

from multiple possible sources. To note is that in all cases the 2-way model does not significantly improve the 1-way EEA fit (all nested p-values > 0.05). The four Bronze Age individuals exhibit substantial amount of AP ancestry in all fitting models, whereas the most recent individual resembles other Nesosiberians with ¾ EEA and ¼ West Eurasian ancestry.



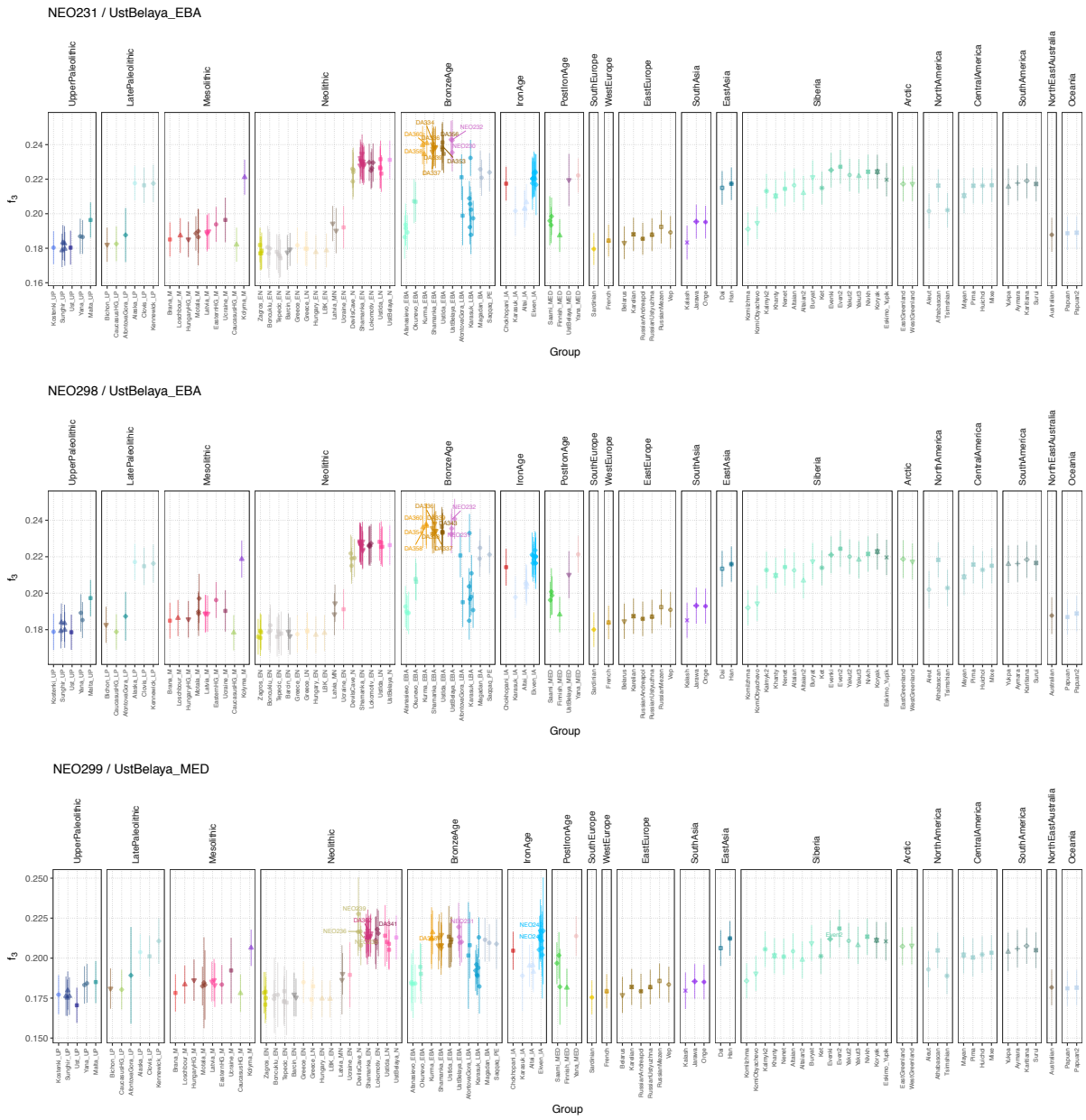


Figure S6.13. Outgroup-f_3 statistics for Ust'Belaya individuals (WGS panel, no transitions).

<u>Magadan</u>

The Bronze Age individuals from Ol'skaya (Magadan) closely resemble Kolyma1, with only a minor ancestry contribution from a West Eurasian source.

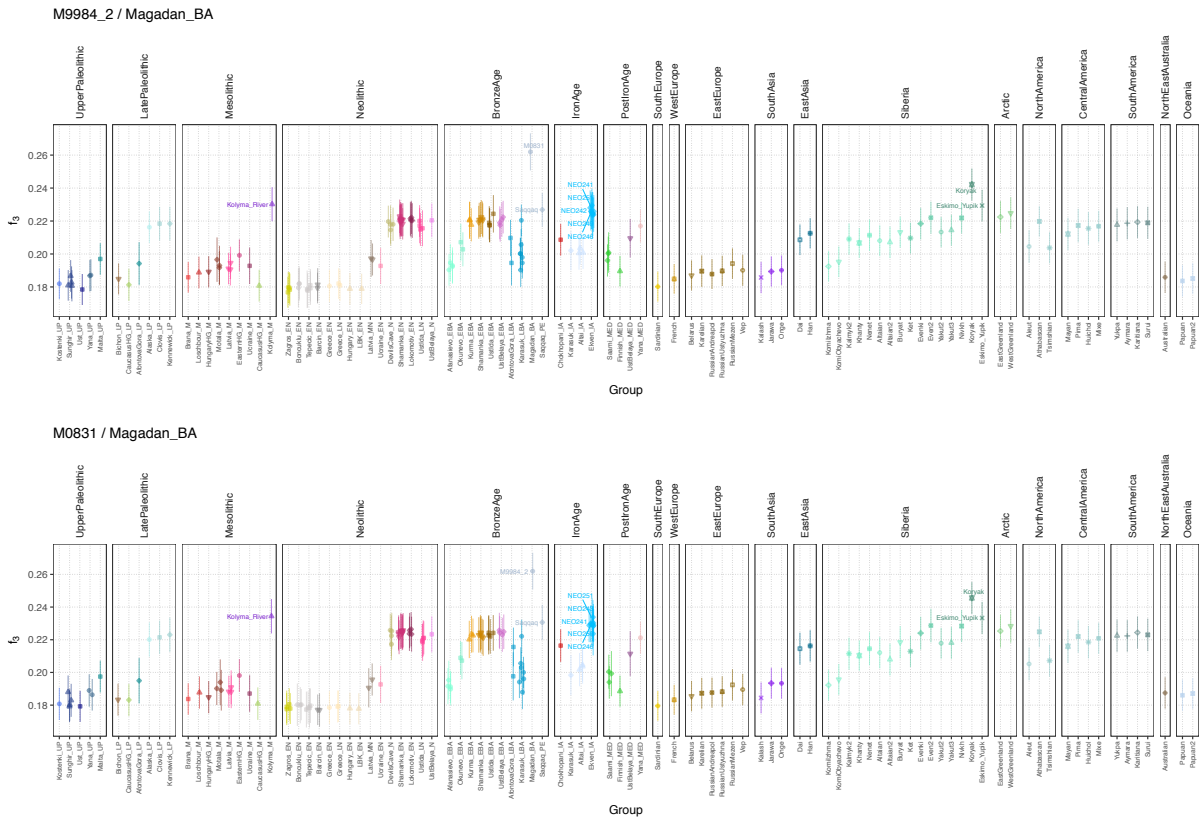


Figure S6.14. Outgroup-f_3 statistics for Magadan individuals (WGS panel, no transitions).

<u>Neoeskimos (Uelen, Ekven)</u>

For the Neoeskimo individuals from Uelen and Ekven cemeteries we only find a single possible 2-way fit, as a mixture of AP ancestry and Native American ancestry. These results are mirrored in present-day Eskimo populations, demonstrating that by ~2.7 kya their ancestral gene pool involving Native American gene flow had already been established.

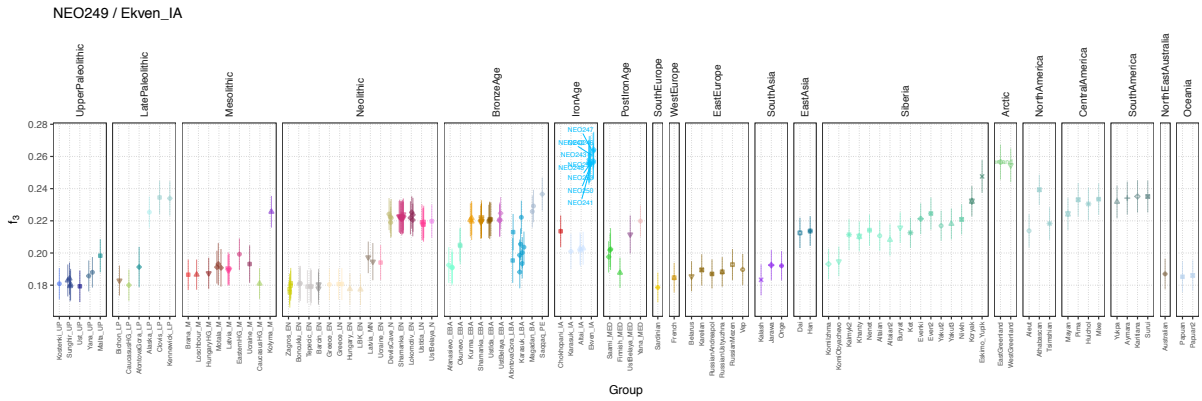


Figure S6.15. Outgroup-f_3 statistics for representative Ekven individual NEO249 (WGS panel, no transitions).

We further investigated the timing of this gene flow using admixture linkage disequilibrium (LD)[17]. Using Saqqaq and Anzick-1 as source populations and the four Ekven individuals with the highest coverage as target, we find a significant weighted LD curve with an estimated admixture date between 100 – 200 generations prior to their age, depending on the data set (Table S6.19).

Table S6.19 ALDER results for Ekven

Ref 1	Ref 2	Dataset	Amplitude			Decay		
			Estimate	SE	Z	Estimate	SE	Z
Saqqaq_LP	Clovis_LP	1240K	0.00207	0.00060	3.5	193.0	43.5	4.4
		WGS	0.00009	0.00002	4.4	104.4	20.7	5.0
Saqqaq_LP	Athabascan	1240K	0.00088	0.00050	1.8	130.1	58.5	2.2
		WGS	0.00005	0.00002	2.6	106.4	27.2	3.9

Neosiberians (Young Yana)

The medieval "Young Yana" individual shows a fit typical of Neosiberians, with ¾ EEA and ¼ West Eurasian ancestry. Many present-day Neosiberian populations from the region (e.g. Yakut, Evenki, Buryat) are modelled with similar mixtures, with a notable absence of AP ancestry in all possible 2-way fits.

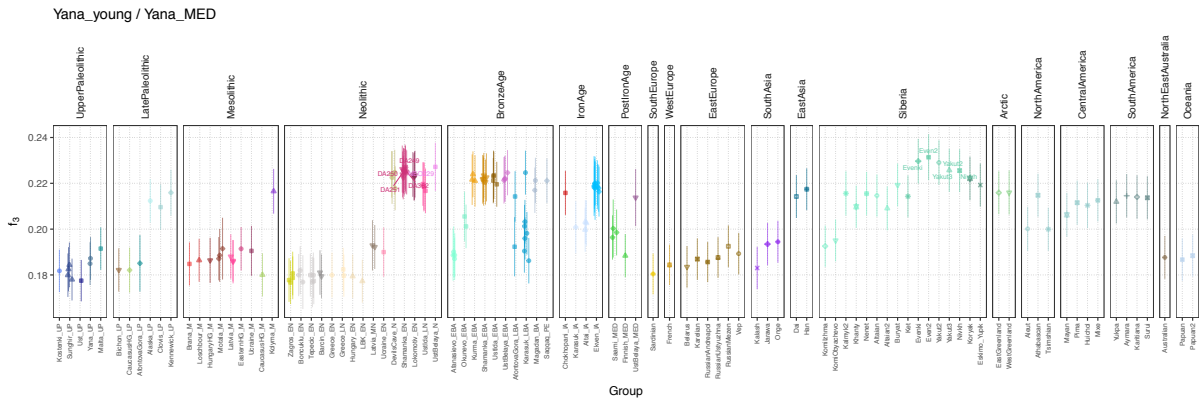


Figure S6.16. Outgroup-f_3 statistics for Young Yana individual (WGS panel, no transitions).

Ancient Saami and Finns

At the western end of our sample distribution, we find the Ancient Finn individual best modelled as a 3-way mixture of Neolithic Farmers, Western hunter-gatherers and Steppe herders, as has been observed for many West Eurasian population impacted by the Yamnaya expansions during the early Bronze Age. The ancient Saami in contrast are the only group in our study that can only be modelled as a 4-way mixture, involving Neolithic farmers, Eastern

hunter-gatherers, Steppe herders as well as EEA. Among all modern populations, all three ancient Saami individuals share most genetic drift with present-day Saami (Fig. S6.17).

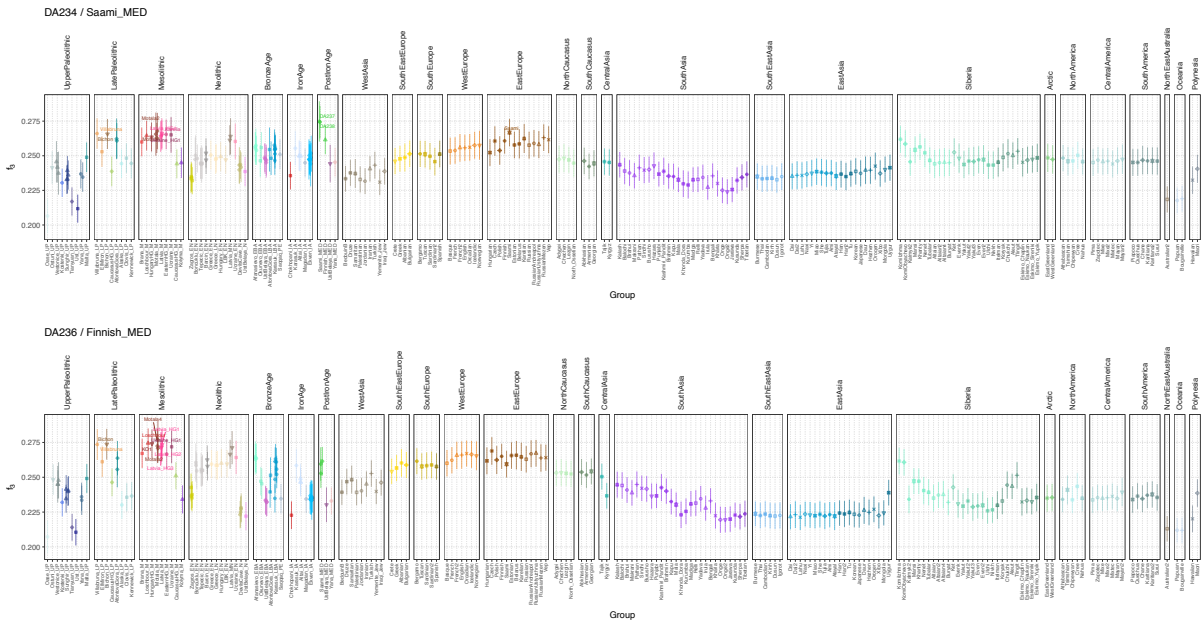


Figure S6.17. Outgroup-f_3 statistics for Saami and Finnish individuals (2240K panel).

Present-day populations

The distribution of ancestry components among the present-day population shows a striking pattern following the most recent migration wave of predominantly East Asian ancestry Neosiberians. Most Neosiberians in northeastern Siberia appear derived from this expansion, and as a consequence AP ancestry is today only found in groups at the extreme fringes of the region. Present-day populations with the highest amount of AP ancestry include the Koryaks, who also carry low amounts of Native American ancestry, possibly due to gene flow with neighbouring Eskimo populations.

6.6 Archaic admixture

To investigate patterns of archaic admixture in the newly sequenced study individuals we first carried out D-statistics of the form D(Chimp,Archaic;X,Yoruba). This is the "canonical" D-statistic for Neanderthal admixture measuring excess allele sharing between archaic hominins and a particular test individual X compared to Africans (represented by Yoruba). We find that all individuals show the signal of Neanderthal admixture typical of non-African populations, without any noticeable affinity to Denisovans beyond that due to its shared ancestry with Neanderthals (Fig. S6.8).

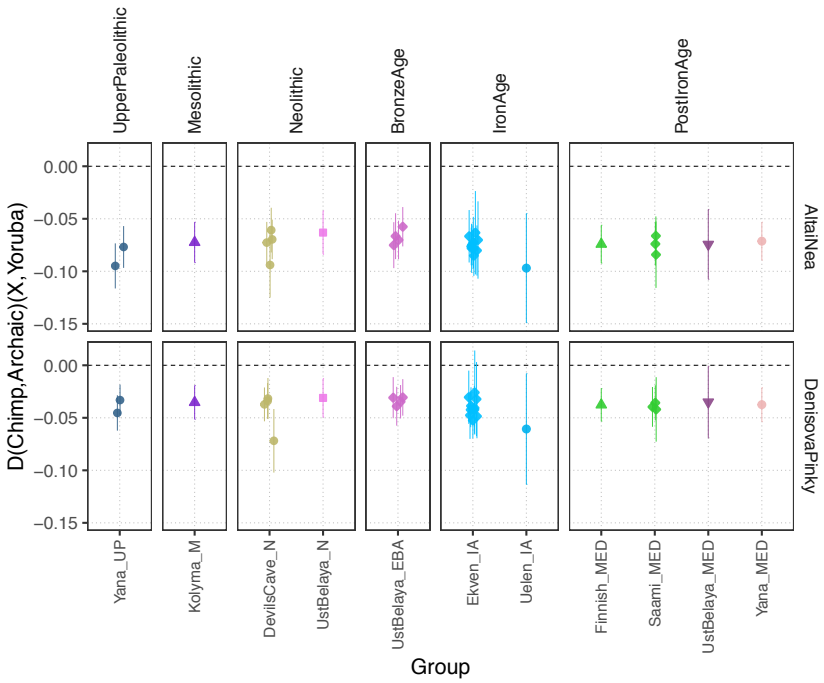


Figure S6.8 D-statistic results for archaic admixture in the study individuals. Error bars represent ± 3 standard errors, estimated using a block jackknife.

Archaic tract lengths

We next used a sliding window approach adopted from[9] to identify genomic tracts of putative archaic origin tracts for all high coverage ancient individuals. For this analysis we merged our genotype data of modern and ancient humans with the archaic human genotypes generated by Prüfer et al[18] (available at http://cdna.eva.mpg.de/neandertal/Vindija/VCF/). For a particular test individual, we identify runs of consecutive SNPs where the individual carries derived alleles of likely archaic origin, defined using the following criteria:

- At least one copy of the derived allele observed in one of the archaic individuals (Altai Neandertal, Vindija33.19, Denisova)
- Only ancestral alleles observed in all African populations from the 1000 Genomes Project Phase 3 dataset as well as all sub-Saharan African populations in our WGS dataset.

Archaic tracts were then defined as runs of putative archaic alleles, allowing a maximum gap size of up to 100 kb between consecutive SNP and requiring at least five SNPs within a called tract. The fraction of Neanderthal ancestry per individual was calculated as the ratio of the sum of the total length of admixture tracts over the total length of the genome passing all filters. Results in Fig. S6.9. show that Neanderthal ancestry in the UP individuals from Yana is organized in longer tracts, similar to the UP individual SIII from Sunghir with a comparable age, and consistent with being temporally closer to the admixture event.

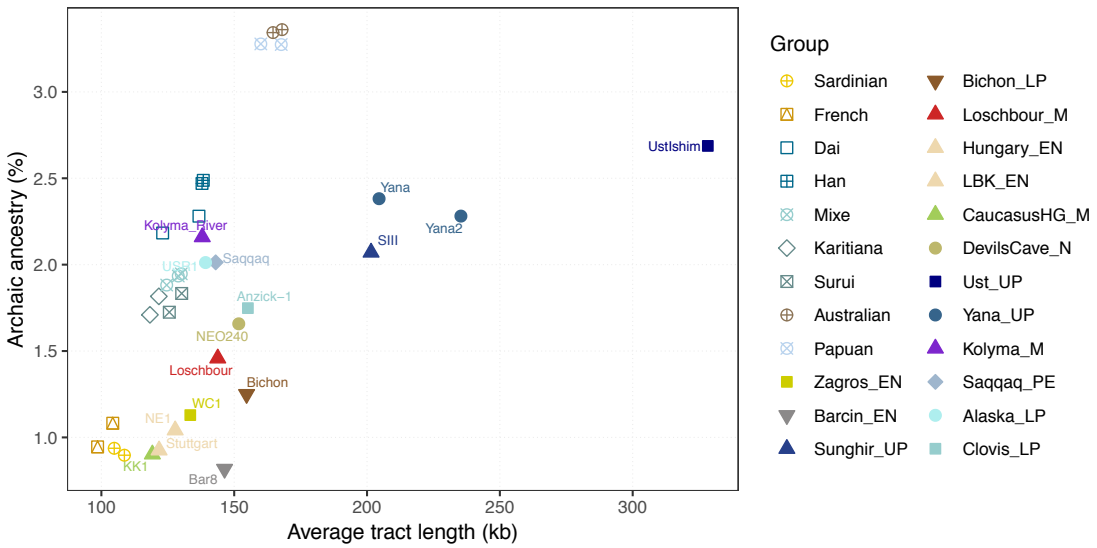


Figure S6.9 Distribution of archaic ancestry segments. Shown are the distributions of the average length and overall fraction of genomic segments of putative archaic origin in selected modern and high coverage ancient genomes.

References

1. Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192,** 1065–1093 (2012).
2. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522,** 207–211 (2015).
3. 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).
4. 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).
5. Sikora, M. *et al.* Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* **358,** 659–662 (2017).
6. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6,** 8912 (2015).
7. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,** 409–413 (2014).
8. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534,** 200–205 (2016).
9. Seguin-Orlando, A. *et al.* Genomic structure in Europeans dating back at least 36,200 years. *Science* **346,** 1113–1118 (2014).

10. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536,** 419–424 (2016).
11. Mathieson, I. *et al.* The Genomic History Of Southeastern Europe. *bioRxiv* 135616 (2017). doi:10.1101/135616
12. Skoglund, P. *et al.* Genetic evidence for two founding populations of the Americas. *Nature* **525,** 104–108 (2015).
13. Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506,** 225–229 (2014).
14. Moreno-Mayar, J. V. *et al.* Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* **553,** 203 (2018).
15. Flegontov, P. *et al.* Paleo-Eskimo genetic legacy across North America. *bioRxiv* 203018 (2017). doi:10.1101/203018
16. Siska, V. *et al.* Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Sci. Adv.* **3,** e1601877 (2017).
17. Loh, P.-R. *et al.* Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. *Genetics* **193,** 1233–1254 (2013).
18. Prüfer, K. *et al.* A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* eaao1887 (2017). doi:10.1126/science.aao1887

Supplementary Information 7 - Demographic history inferences

Vitor Sousa[1,2,3], Isabelle Dupanloup[1,2], Laurent Excoffier[1,2]

[1]CMPG, Institute of Ecology and Evolution, 3102 Berne, Switzerland

[2]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

[3]CE3C, Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal

Statistical framework

Likelihood inference based on the Site Frequency Spectrum

The parameters of alternative demographic scenarios were inferred using the site frequency spectrum (SFS) (Nielsen 2000; Adams, Hudson 2004) by approximating the likelihood of a given model with coalescent simulations (Nielsen 2000). All computations were done with an extension of the fastsimcoal2 simulation software (Excoffier et al. 2013).

We followed the procedure used in recent studies (de Manuel et al. 2016; Malaspinas et al. 2016) by performing the likelihood optimization in two steps: (1) optimize the full likelihood L_{full} (accounting for the monomorphic and polymorphic sites) for 15 cycles, and then (2) for the remaining 40 cycles optimize the likelihood L_{SFS} accounting only for the polymorphic sites. This strategy aims at maximizing the fit between the expected and the observed SFS. At the end of the run, a rescaling factor is computed as $RF = S_{obs} / S_{\exp}$, where S_{obs} and $S_{\exp}$ are the observed and expected numbers of polymorphic sites, respectively. The final maximum-likelihood parameters are then rescaled according to the *RF* such that the number of expected polymorphic sites equals the observed one: the effective population sizes and times of events (including the age of samples) are multiplied by *RF*, whereas admixture rates are left unchanged.

Despite the relatively high depth of coverage and high confidence in genotype calls, we masked out singletons when computing the likelihood because the current dataset contains three ancient modern humans (older than 8,000 years), and singletons are especially likely to occur due to sequencing errors with ancient DNA. We used a modified version of fastsimcoal2 (ver 2.6), which estimate parameters and perform the likelihood optimization by discarding singletons.

Demographic history of Siberia and the Americas

Data preparation and processing

We selected, for our analyses, the ancient individuals with the highest coverage, one individual from Yana RHS site (a representative of Ancestral North Eurasians, hereafter referred to as Yana), one individual from Duvanni Yar at the lower Kolyma River (a representative of Ancstral Paleosiberians, hereafter referred to as Kolyma) and one Northern Alaskan individual from the Upward Sun River site (USR1, a representative of Ancient Beringians, hereafter referred to as Alaskan, Moreno-Mayar et al. 2018). In our demographic models, we also included four present-day modern human representative of Europe (two Sardinians, HGDP01078 and HGDP01079), Siberia (two Even, Nlk3 and Nlk18), East Asia (two Han Chinese, HGDP00783 and HGDP00785) and America (two Karitiana, HGDP01012 and HGDP01018). These had been previously sequenced at high coverage (panel C from Simons Genome Diversity Project (Mallick et al. 2016)). We selected Even as a representative of present day Siberian (Neosiberian) populations because this group showed no evidence of recent admixture with other Eurasian populations. However, we discarded

individual Nlk1 from the three available Even genomes, as ancestry analyses suggested this individual was actually admixed.

To obtain a SFS reflecting historical demographic events, we focused on neutral autosomal SNPs, which we defined as those found outside genic regions, according to Ensembl version 71, April 2013 (Cunningham et al. 2015), and outside CpG islands, according to the UCSC platform (Rosenbloom et al. 2015). We discarded SNPs for which the depth coverage was lower than 10X in any of the individuals. The ancestral state of the SNPs was inferred using the ancestral hg19 genome provided by the 1000G consortium (Abecasis et al. 2012), which was inferred from the alignments of six primate genomes and released in the Ensembl Compara 59 database (Flicek et al. 2011). From our dataset of autosomal SNPs passing the above filters, we generated a dataset by considering concatenated autosomal segments (blocks) of 1Mb (not necessarily contiguous along the chromosome) considering only sites without missing data in any of the individuals. We identified 625 such blocks on the autosomes, with a total of 613,981,492 sites, of which 1,465,258 were SNPs. For the non-parametric block-bootstrap analysis used to infer parameter confidence intervals, we resampled these whole blocks with replacement to generate 100 sets of 625 blocks and hence bootstrap datasets have the same cumulative length of 625 Mb. From this dataset, we generated the multidimensional SFS for datasets with a different combination of populations with the Arlequin software ver 3.5.2.2 (Excoffier, Lischer 2010).

Modelling the relationship of Siberian and Native American populations

By analyzing genomes from Siberia sampled at different times (Yana at 31.6 kya, Kolyma at 9.8 kya and Even from present day), we could investigate in detail the population history of a geographic region that was strongly affected by glaciation cycles and which is of extreme importance to understand the historical movement of people across Eurasia and the colonization of the Americas. We started by modeling and inferring the relationships among Siberian and present day Eurasian populations. Then, conditional on the estimated demographic history of Siberian populations, we modeled and inferred their relationship with Native American populations, namely one ancient Alaskan ~11.5kya and present-day Karitiana. We modeled gene flow among populations as pulses of admixture between populations, characterized by two parameters, the admixture contribution proportion and the admixture time.

Population replacement history in Siberia

To investigate the relationship among Siberian populations we considered models with data from five populations (5D-SFS), namely three Siberian populations: Yana (31.6 kya), Kolyma (9.8 kya), and Even (present day); one European population (Sardinian), and one East Asian population (Han Chinese). The major aim of this model-based demographic analysis was to test whether the SFS data is consistent with: (i) a scenario of continuity, where younger Siberian samples would represent direct descendants from older samples, versus a (ii) scenario involving episodes of population replacements, where older populations would be replaced by new migrants, leaving no direct descendants in present day populations inhabiting that area. The full model we considered assumes that Siberian populations are differentially related to European and East Asian lineages (Figure S7.1). Going from past to present, the ancestral population of all Eurasians is allowed to experience a bottleneck (mimicking the effects of the out-of-Africa), followed by 3% contribution from Neanderthal (Malaspinas et al. 2016) and the split of European and East Asian lineages. We considered an extra 0.5% contribution of Neanderthal to the East Asian lineage (Malaspinas et al. 2016). The Siberian Yana population is assumed to be a mixture of both lineages, receiving a

contribution ω from Europe and 1-ω from East Asia at the time of split, whereas Kolyma is assumed to diverge from the ancestors of Even and Han Chinese. To test for replacement, "older" Siberian populations (Yana and Kolyma) can potentially contribute to "younger" Siberian populations, i.e. Yana can contribute to Kolyma (proportion β) and Even (proportion α), and Kolyma can contribute to Even (proportion τ). The estimates of the admixture proportions indicate whether the SFS data favour a scenario of continuity or replacement. To account for potential founder effects, we allowed for the possibility of bottlenecks associated with the population split events. We also allowed and modelled recent gene flow between Europe and East Asia (from present day until time of Han and Even split). Note that although we assumed a fixed topology with Even closer to Han Chinese, with their ancestors diverging from Kolyma, by letting the proportions α, β, and τ to vary up to 1.0, we allow for different topologies and relationships among populations, which would be supported by inferred α, β, and τ close to one.

The estimates of the contributions α, β and τ could be affected by several factors, such as different sampling times of populations, shared ancestral polymorphism, differential amount of drift in each population due to different effective sizes and bottlenecks, gene flow among Eurasian populations and Neanderthal admixture. To account for these effects, we considered explicitly in our models for: (i) different sampling times for Yana (31.6 kya, corresponding to 1090 generations ago, assuming 29y per generations (Fenner 2005)) and Kolyma (9.8 kya, corresponding to 3338 generations ago); (ii) different effective sizes for each population and lineage in the model; (iii) potential bottlenecks representing founder events associated with each population split events, as well as in the ancestor of all Eurasians, i.e. representing the out-of-Africa event (bottleneck intensity inferred by estimating the bottleneck effective size during a fixed duration of 10 generations); (iv) a period of continuous migration between Europe and East Asian since the divergence of East Asian from Even; and (v) differential Neanderthal contribution into Europe and East Asia. The Neanderthal contribution was modeled by considering an unsampled ("ghost") Neanderthal population contributing 3% into the ancestors of all Eurasian populations (time fixed as the mean between the out-of-Africa bottleneck and Eurasian divergence), and an extra 0.5% into the Asian lineage (time fixed as mean time between Eurasian divergence and Kolyma split). Neanderthal effective size and split times were fixed according to recent estimates based on genome-wide SFS (Malaspinas et al. 2016). The full model comprised 32 inferred parameters and 6 fixed parameters (Figure S7.1), including the effective sizes, times of events, admixture proportions and migration rate. The search range for the admixture proportions was set between 10^{-5} and 1.0, such that values very close to zero (e.g., below 10^{-4}) indicate replacement and values close to 1.0 indicate continuity.

Full versus nested models without admixture

We used two approaches to quantify if admixture contribution parameters were actually different from zero. First, we obtained likelihood profiles for each parameter, varying the admixture contribution parameter values conditional on fixing all the other parameters on the maximum likelihood estimates. For each admixture parameter we considered a grid of 60 points (20 points equally spaced between 10^{-9} and 0.1, and 40 points equally spaced between 0.1 and 1.0). For each point parameter value we simulated 20 expected SFS approximated with 7.5×10^5 coalescent simulations to obtain the distribution of likelihood values. This was also done for the Neanderthal admixture parameters, evaluating the likelihood at a grid of 60 equally spaced points between 10^{-9} and 0.1. Second, based on the likelihood profile results, we tested whether simpler replacement models fit the SFS data as well as the full model. Namely, we considered the following nested models: (i) no direct contribution from Yana into Even, setting α=0; (ii) no direct contribution from Kolyma into Even, setting τ=0; (iii) no

direct contribution from Yana into Even, and from Kolyma into Even, setting α=τ=0. In contrast with the likelihood profile analyses where all other parameters were fixed to the maximum likelihood estimates, here we re-run the estimation of all parameters under the nested models (same settings as for the full model) and compared the obtained maximum likelihood estimates across models. To investigate the effect of estimating all parameters conditional on a fixed Neanderthal contribution of 3% into ancestral of Eurasians and 0.5% extra contribution to Asian lineage, we re-run the estimation of all parameters for the best nested model with different fixed Neanderthal admixture values. We considered models with a Neanderthal contribution of 1%, 2% and 3% into ancestral of Eurasians, and of 0.10%, 0.25% and 0.50% extra contribution into the Asian lineage, respectively.

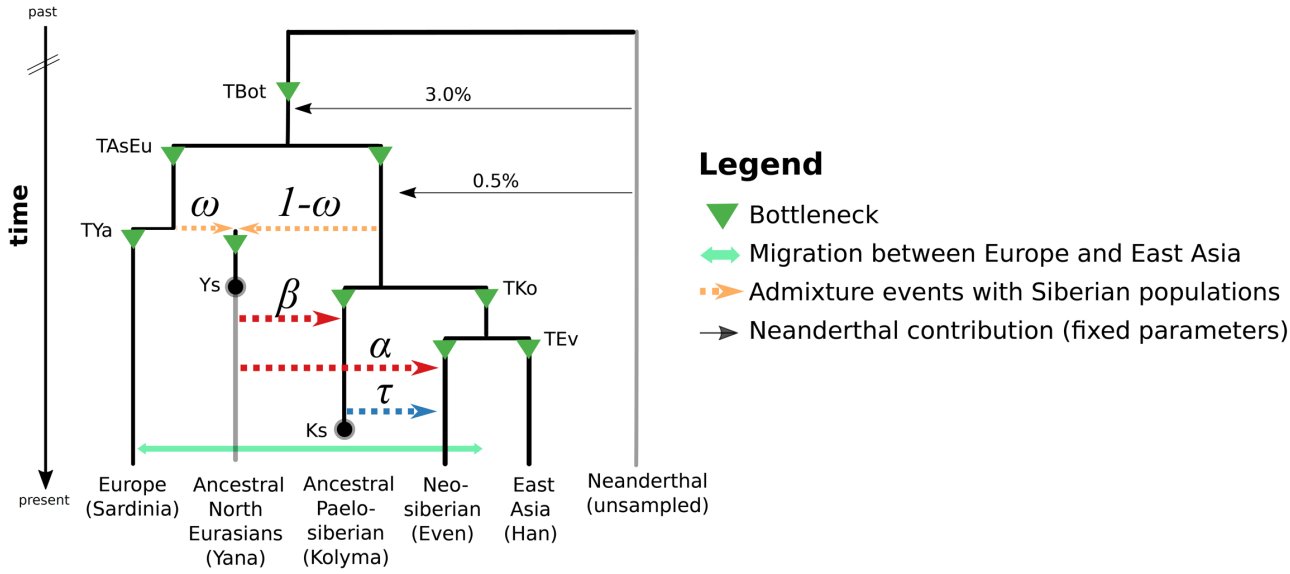


Figure S7.1. Schematic representation of the full model used to infer the relationship of present-day and ancient Siberian human samples with Eurasians, used to test for population replacement in Siberia (see text for details).

Reconstructing the colonization of the Americas

We have selected two Native American populations as representatives of different time points: (i) the ancient USR1 sample from Alaska estimated to be ~11.5ky old (Moreno-Mayar et al. 2018), and (ii) the present-day Karitiana sample. The choice of populations was done such as to minimize genotyping errors and the confounding effect of recent Eurasian gene flow, by selecting samples with high depth of coverage (>10x) and limited recent admixture with present-day Eurasian populations. Due to the high complexity of models including all seven populations, we have considered simpler models with six sampled populations (the five Eurasian samples analysed above and one Native American sample at a time). Furthermore, in order to optimize the inference procedure, we have reduced the number of estimated parameters by fixing all the parameters related to the Eurasian populations (e.g. effective sizes, times of split, admixture contributions between Siberian populations, gene flow between Sardinians and Han Chinese) to the maximum likelihood point estimates obtained in the previous section (Table S7.3). Therefore, 29 out of 44 parameters of the model with six sampled populations were fixed, and we estimated the remaining 15 parameters (Tables S7.3 and S7.4). We have thus based our inference on the multidimensional SFS from six populations (6D-SFS), including the five Eurasian populations analyzed in previous section and either the ancient Alaskan sample (one individual USR1 with 11.5 kya) or the present-day Karitiana sample (two individuals). The

major aim of these analyses was to answer the following questions: (i) What Siberian population along the time axis (Yana, Kolyma or Even) is more closely related to the Native American populations? (ii) Is there evidence of past historical admixture between Siberian and Native American populations?

To model the relationship of Eurasian and Native American populations, we have added a Native American population into the Siberian demographic model, inferring only the relevant parameters to answer the above questions. We assumed that the lineage leading to the Native American population diverged from the Asian ancestral population, and we implemented a bottleneck along this lineage to model a potential founder event associated with the colonization of the American continent (Figure S7.2). To test whether the SFS supports a scenario where Native Americans diverged from the ancestral population of Kolyma or from the population ancestral to Han and Even (Figure S7.2), we allowed the split time of Kolyma (TKo) to be younger or older than the split time of the ancestors of Han and Even (TEv). To detect and quantify past historical gene flow into the Native American population, we allowed for potential unidirectional contributions from all Siberian populations, as represented by admixture proportions γ (Yana into America), κ (Kolyma into America) and λ (Even into America) (Figure S7.2). Note that the inferred admixture proportions (γ, κ, λ) and their corresponding times also inform us on the gene pool composition of the ancestral population that colonized the Americas, as well as on the population tree topology. Values of admixture proportions close to zero would indicate that the populations diverged with no further gene flow and/or that the Native American ancestral population was not mixed with different lineages. Admixture proportions close to one would indicate that the Native American population is the direct descendant of one of the Siberian populations (e.g. estimates of $\gamma=1$ and $\kappa=\lambda=0$ would indicate that the Native American population descends directly from Yana, suggesting an alternative population tree topology). Intermediate values would indicate past gene flow and/or that the ancestral population that dispersed into the Americas was a mixture of different Siberian/East Asian lineages. We note that the tree topology and the admixture proportions are not fully identifiable, since alternative scenarios can lead to similar likelihood values. To illustrate this non-identifiability of parameters, consider a scenario of two populations that diverged 1000 generation ago. Under our model, this could be explained by a combination of split times, admixture proportions and admixture times. In this case the model with a recent split (1,000 generations ago) with no further admixture (0%) should have the maximum likelihood, but scenarios with older splits (e.g. 10,000 generations ago) followed by strong admixture (e.g. 100% admixture contribution 1,000 generations ago) could have likelihood values close to the maximum. To minimize this non-identifiability, we constrained the search range of the admixture proportions, such that the parameter search during the likelihood optimization was done on a log-uniform scale to ensure that the parameter search starts more often at proportions close to zero (Tables S7.6 and S7.7). We used recent estimates indicating a relatively small Ne for the Alaskan effective size (Moreno-Mayar et al. 2018) to define the search range values. We note that using a range of Ne similar to the one used for Han, Sardinian and Even led to lower likelihoods (not shown).

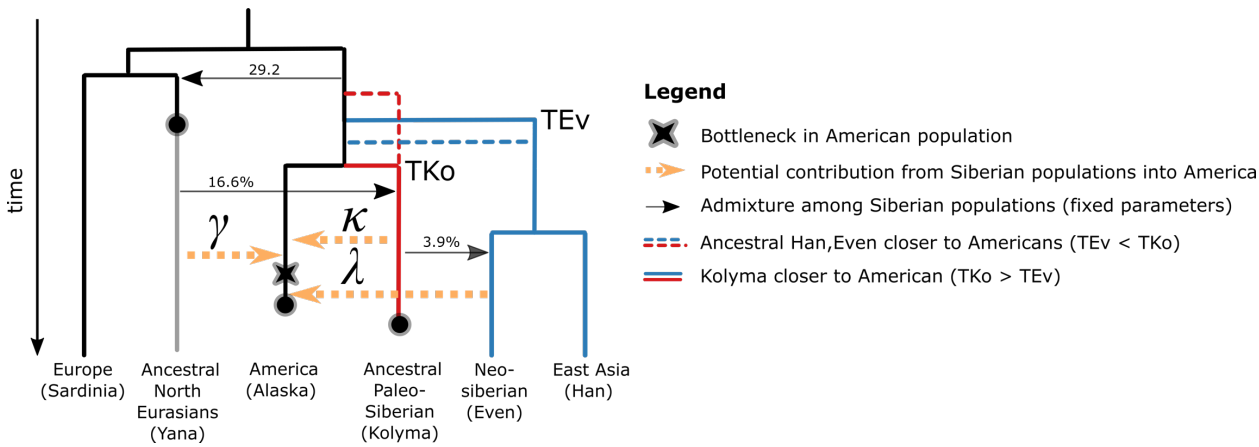


Figure S7.2. Schematic representation of the full model used to infer the relationship of Eurasian, Siberian and Native American population (see text for details). All parameters related to Eurasians were fixed to the maximum likelihood estimates obtained in a separate analysis. TEv: time of split of ancestral of Han,Even from Asian lineage. TKo: time of split of Kolyma from Asian lineage. Depending on the inferred relationship of TEv an TKo, we can obtain different topologies represented by the solid and dashed lines.

Full versus nested models without admixture

For the Native American history reconstruction, we used two approaches to quantify if admixture contribution parameters were actually different from zero. We obtained likelihood profiles for each admixture parameter using the same settings as for the Siberian replacement history. Second, we compared the likelihoods of nested models (setting some of the admixture parameters to zero) and the full model. Namely, we examined the following nested models: (i) no direct contribution from Yana into America, setting $\gamma=0$; (ii) no direct contribution from Kolyma into America, setting $\kappa=0$; (iii) no direct contribution from Even into America, setting $\lambda=0$; (iv) no direct contribution from Kolyma and Even into America, setting $\kappa=\lambda=0$; and (v) no admixture, setting $\kappa=\lambda=\gamma=0$. As in the previous section, we re-run the estimation of all parameters under the nested models using the same settings as for the full model. Given that our estimates suggested a similar contribution of Yana into Kolyma and Yana into Alaska at similar times, we tested a full model with a single pulse of admixture from Yana into ancestors of Kolyma and Alaska, i.e. assuming a single admixture event. We tested a full model and the same four nested models mentioned above.

Finding the maximum likelihood parameters

We estimated the set of parameters that maximize the likelihood by specifying the search ranges shown in Tables S7.2, S7.6, S7.7, for the models of Siberian and Native American demographic history, respectively. The expected SFS used to compute the likelihood for any given set of parameters was estimated from 750,000 coalescent simulations. For each model, we performed 100 optimizations runs starting from different initial conditions and selected the run leading to the highest likelihood to get the maximum likelihood parameter estimates. All divergence and admixture times were estimated assuming a constant mutation rate of 1.25e-8/gen/site (Scally, Durbin 2012) and a generation time of 29y per generations (Fenner 2005).

Non parametric bootstrap analysis

We estimated confidence intervals for the best model by estimating parameters from 100 bootstrap datasets. All the settings for parameter estimation were the same as those used for the analyses of the original dataset, except that due to computational constraints we only

performed five optimization runs per bootstrap dataset, starting from the maximum likelihood estimates obtained with the original dataset. The 95% confidence intervals for each parameter were computed based on the percentile method (interval $[Q_{0.025}, Q_{0.975}]$, where Q_a is the a percentile of the bootstrap distribution)(Davison, Hinkley 1997), as implemented in the R *boot* package.

Results for Siberian demographic history

The likelihood profiles and the comparison of the likelihood of nested models with the full model suggest a scenario of replacement with limited admixture (Figure S7.3, Table S7.1). The point estimates obtained under the full and nested models are shown in Table S7.2. The likelihood profiles of all admixture contributions have a peak far from zero, except for the admixture contribution α of Yana into Even (Figure S7.3). In agreement with this result, the likelihood of the nested model with α=0 is similar to that of the full model, suggesting that a model without a direct contribution from Yana into Even is compatible with the SFS data (Table S7.1). The point estimates and 95% Confidence intervals obtained for the model without direct contribution of Yana into Even are shown in Table S7.3. Estimates were obtained conditional on a fixed 3% admixture of Neanderthal into ancestral of Europeans and Asians, and an extra pulse of 0.5% into the Asian lineage (Malaspinas et al. 2016). Estimates are robust to varying the Neanderthal admixture proportions (Table S7.4). In particular, the estimates of the admixture of Yana into Kolyma, and of Kolyma into Even are similar across models with different Neanderthal contribution, and are all consistent with a scenario of replacement with limited admixture.

The representation of the model that best fits the data is shown in main Figure 3. The inferred time of events and the admixture contributions are consistent with a scenario with at least three waves and almost full replacement. We infer an initial split of the European and Asian lineages 43kya (95%CI: 33.4-48.6), closely followed by the split of Yana ~39kya from the European and Asian lineages, receiving ~29% (95%CI: 21.3-40.1) Asian contribution and therefore 71% European contribution. We infer a relatively old split ~27 kya (95%CI: 17.1-32.1) of the Kolyma lineage from the ancestral of Han and Even. Kolyma ancestors are likely to have migrated into the region of Siberia inhabited by Yana ~25.8kya (95%CI: 14.5-28.9), as this is the estimated time for admixture between Yana and Kolyma. Our estimates do not support full continuity between Yana and Kolyma populations, as Yana admixture contribution into Kolyma is low 16.6% (95% CI 7.5-22.2).

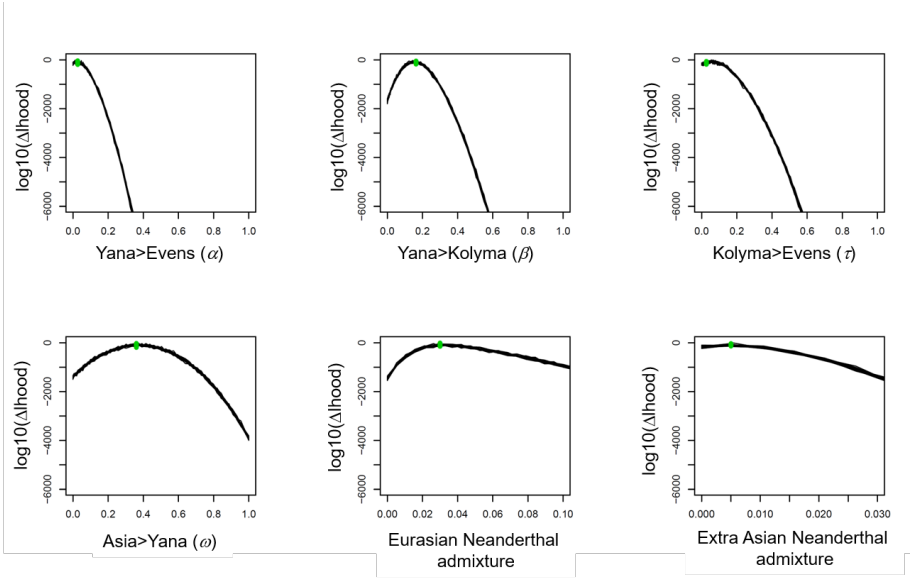


Figure S7.3. Likelihood profiles for each admixture parameter under the full model of Siberian demographic history (fixing all the other parameters to the maximum likelihood estimate). Each line corresponds to 1 of 20 runs for the difference between the estimated likelihood and the maximum likelihood across all runs and parameter values, using 10^6 coalescent simulations to approximate the expected SFS. The variance across runs is reduced, such the 20 lines are indistinguishable.

Our estimates are compatible with an almost full replacement of Yana by Kolyma, or in other words, a partial assimilation of a Yana descendant population by the Kolyma population. We infer a recent split of Han and Even ~19.6 kya (95% CI 11.9-22.0). The ancestors of Even are then likely to have migrated into Siberia to the region inhabited by Kolyma ~13kya (95%CI 10.4-18.3), and almost fully replaced them, receiving a small contribution from Kolyma (~3.9%, 95%CI 0.1-12.6). Note that the even though we obtain consistent point estimates across models (Table S7.2), the confidence intervals are wide, especially for the time of events (Table S7.3). This is probably because we have a limited sample size of two genomes (1 diploid individual) from Yana and Kolyma.

Table S7.1. Model comparison of full and nested models of Siberian demographic history. Maximum likelihood estimates suggest a scenario of replacement with no direct contribution from Yana into Even. #parameters corresponds to the number of parameters for each model. Estimated ML is the maximum likelihood estimate for each model in log10 units. Δlikelihood is given in log10 units and corresponds to the difference between the estimated likelihood and the maximum possible likelihood if there was an exact fit to the observed SFS, and hence the closer to zero the better. ΔmaxL is given in log10 units and corresponds to the difference between the likelihood of a given model and the best model, hence a value of zero indicates the best model.

Models of Siberian demographic history	#parameters	Estimated ML	Δlikelihood	ΔmaxL
Full model	32	-2,499,145	-2,763	0
Nested models:				
- no Yana>Even contribution (τ=0)	30	-2,499,157	-2,775	-12
- no Kolyma>Even contribution (α=0)	30	-2,499,226	-2,844	-81
- no Yana>Even and no Kolyma>Even contribution (τ=0,α=0)	28	-2,499,215	-2,833	-70

Table S7.2. Siberian demographic history, with inferred parameters for the full and nested models investigated. Search range of parameters are indicated and compared with the maximum likelihood estimates for each model. Parameter estimates of the best nested model indicated in bold. Ys – Yana sampling age (31.6 kya). Ks – Kolyma sampling age (9.8 kya). See Figure S7.1 for abbreviations of other parameters.

Parameters	**Search range**		**Full model**	**Nested models**		
Effective sizes (diploid)	**lower**	**upper**		$\alpha = 0$	$\tau=0$	$\alpha = \tau = 0$
Ne ancestral humans/Neanderthal	500	25,000	16,002	**15,560**	15,990	15,442
Ne ancestral modern humans	500	25,000	11,970	**12,142**	11,913	12,216
Ne ancestral non-Africans	500	25,000	9,298	**15,258**	3,405	11,102
Ne bottleneck Out of Africa	5	250 [b]	52	**59**	94	57
Ne ancestral Asian continent	500	25,000	5,238	**16,613**	15,967	2,992
Ne bottleneck Asian continent	5	250 [b]	61	**44**	46	116
Ne ancestral Europe	500	25,000	20,703	**18,454**	14,935	9,107
Ne Sardinians	500	25,000	10,982	**10,173**	11,541	11,273
Ne bottleneck Europe	5	250 [b]	202	**219**	212	150
Ne ancestral Han,Even	500	25,000	18,186	**20,668**	26,278	15,793
Ne Han Chinese	500	50,000	34,223	**38,232**	31,482	15,413
Ne bottleneck Han Chinese	5	250 [b]	183	**176**	213	191
Ne Even	50	25,000	19,248	**25,791**	28,444	26,687
Ne bottleneck Even	5	250 [b]	189	**216**	195	227
Ne Kolyma	50	25,000	2,415	**1,648**	11,398	25,136
Ne bottleneck Kolyma	5	250 [b]	27	**99**	23	19
Ne Yana	50	25,000	19,345	**4,155**	20,475	9,873
Ne ancestral Yana	50	25,000	1,246	**6,257**	14,162	16,774
Ne bottleneck Yana	5	250 [b]	126	**38**	40	36
Admixture proportions (%)						
Kolyma>Even (τ)	10^{-3}	100 [b,l]	2.7	**3.9**		
Yana>Kolyma (β)	10^{-3}	100 [b,l]	16.3	**16.6**	12.1	10.1
Yana>Even (α)	10^{-3}	100 [b,l]	2.7		1.9	
AsiaContinent>Yana (ω)	10^{-3}	100 [b,l]	36.1	**29.2**	28.3	58.9
Scaled migration rate (2Nm)						
Han > Sardinian	10^{-5}	5	3.6	**3.7**	3.2	3.4
Times of split (Kya)						
Bottleneck OoA (TBot)	37.7	116.0 [b]	42.7	**45.1**	51.9	42.0
Split Asia/Europe (TAsEu)	max(Ys, TKo)	TBot [b]	40.9	**43.1**	43.8	38.7
Split Yana (TYa)	Ys	TAsEu [b]	39.6	**38.7**	41.9	37.2
Split Kolyma (TKo)	11.0	34.8 [b]	24.4	**27.1**	30.3	22.4
Split Han,Even (TEv)	Ks	min(TYa, TKo)[b]	14.4	**19.6**	15.2	12.5
Times of admixture (Kya)						
Kolyma>Even	Ks	TAsEv[b]	13.2	**13.3**		
Yana>Kolyma	TEv	min(TKo, TYa)[b]	17.9	**25.8**	22.4	20.0
Yana>Even	0	min(TEv,TYa)[b]	10.0		1.3	

[b] – bounded upper search range, [b,l] – log-uniform search range with upper bound.

Table S7.3. Point estimates and 95% confidence intervals for the parameters of the best nested model for Siberian demographic history. Bootstrap confidence intervals were calculated according to the percentile method. Point estimates in bold correspond to the parameters inferred with the original data set. Times of divergence in years are obtained by assuming a generation time of 29 years and a mutation rate of 1.25e-8/gen/site.

Parameters of best nested model ($\alpha = 0$)		**95%CI interval**	
	Point estimate	**lower**	**upper**
Effective population sizes (diploid size)			
Ne ancestral humans/Neanderthal	**15,560**	15,261	16,822
Ne ancestral modern humans	**12,142**	11,211	12,536
Ne ancestral non-Africans	**15,258**	1,574	23,683
Ne bottleneck Out of Africa	**59**	48	182
Ne ancestral Asian continent	**16,613**	4,093	23,651
Ne bottleneck Asian continent	**44**	37	57
Ne ancestral Europe	**18,454**	5,154	24,818
Ne Sardinians	**10,173**	7,504	15,235
Ne bottleneck Europe	**219**	154	242
Ne ancestral Han,Even	**20,668**	4,897	26,221
Ne Han Chinese	**38,232**	18,060	50,882
Ne bottleneck Han Chinese	**176**	155	246
Ne Even	**25,791**	9,696	26,860
Ne bottleneck Even	**216**	148	241
Ne Kolyma	**1,648**	1,170	9,378
Ne bottleneck Kolyma	**99**	20	224
Ne Yana	**4,155**	2,280	23,344
Ne ancestral Yana	**6,257**	1,442	20,643
Ne bottleneck Yana	**38**	28	80
Admixture proportions (%)			
Kolyma>Even (τ)	**3.9**	0.1	12.6
Yana>Kolyma (β)	**16.6**	7.5	22.2
AsiaContinent>Yana (ω)	**29.2**	21.3	40.1
Scaled migration rate (2Nm)			
Han > Sardinian	**3.65**	0.00	8.76
Times of split (Kya)			
Bottleneck OoA (TBot)	**45.1**	37.5	54.9
Split Asia/Europe (TAsEu)	**43.1**	33.4	48.6
Split Yana (TYa)	**38.7**	32.2	45.8
Split Kolyma (TKo)	**27.1**	17.1	32.1
Split Han,Even (TEv)	**19.6**	11.9	22.0
Times of admixture (Kya)			
Kolyma>Even	**13.3**	10.4	18.3
Yana>Kolyma	**25.8**	14.5	28.9

Table S7.4. Effect of Neanderthal admixture in point estimates under the best nested model for Siberian demographic history. We re-run the parameter estimation varying the Neanderthal contribution into ancestral of Eurasians (adNeaEurasia) and the extra Neanderthal admixture pspecific to the Asian lineage (extraAdNeaAsi). By default we used adNeaEurasia=3% and extraAdNeaAsi=0.5% (Malaspinas et al. 2016). Times of divergence in years are obtained by assuming a generation time of 29 years and a mutation rate of 1.25e-8/gen/site.

Parameters of best nested model ($\alpha = 0$)			
adNeaEurasia=	**3%**	2%	1%
extraAdNeaAsi=	**0.5%**	0.25%	0.1%
Effective population sizes (diploid sizes)			
Ne ancestral humans/Neanderthal	**15,560**	16,802	19,193
Ne ancestral modern humans	**12,142**	11,430	11,154
Ne ancestral non-Africans	**15,258**	14,802	23,635
Ne bottleneck Out of Africa	**59**	89	122
Ne ancestral Asian continent	**16,613**	23,477	32,320
Ne bottleneck Asian continent	**44**	42	58
Ne ancestral Europe	**18,454**	12,734	19,222
Ne Sardinians	**10,173**	10,086	12,575
Ne bottleneck Europe	**219**	225	269
Ne ancestral Han,Even	**20,668**	24,874	30,163
Ne Han Chinese	**38,232**	46,051	34,579
Ne bottleneck Han Chinese	**176**	222	249
Ne Even	**25,791**	23,786	31,226
Ne bottleneck Even	**216**	204	252
Ne Kolyma	**1,648**	3,865	3,801
Ne bottleneck Kolyma	**99**	32	38
Ne Yana	**4,155**	4,503	11,187
Ne ancestral Yana	**6,257**	1,494	22,185
Ne bottleneck Yana	**38**	57	42
Admixture proportions (%)			
Kolyma>Even (τ)	**3.9**	5.6	7.3
Yana>Kolyma (β)	**16.6**	13.6	15.0
AsiaContinent>Yana (ω)	**29.2**	28.0	25.0
Scaled migration rate (2Nm)			
Han > Sardinian	**3.65**	2.4	2.2
Times of split (Kya)			
Bottleneck OoA (TBot)	**45.1**	41.7	57.1
Split Asia/Europe (TAsEu)	**43.1**	39.3	51.7
Split Yana (TYa)	**38.7**	38.3	44.6
Split Kolyma (TKo)	**27.1**	35.4	38.2
Split Han,Even (TEv)	**19.6**	17.2	21.1
Times of admixture (Kya)			
Kolyma>Even	**13.3**	12.3	18.6
Yana>Kolyma	**25.8**	21.2	27.3
Δlikelihood	**-2,775**	-2,774	-2,915

Results for demographic history of the Americas

The likelihood profiles and the comparison of the likelihoods of nested models with the full model suggest a non-zero contribution of Yana into the Americas but a limited contribution from Even into the Americas. This is consistently inferred for the two datasets with Alaska and Karitiana. Furthermore, for both Alaskan and Karitiana, estimates indicate that: (i) Kolyma is the closest related population to Native Americans, and (ii) that Yana contributed to Native Americans. Below we describe these results in detail, focusing especially in the relationship between Alaskan and Siberian populations.

Alaskan

We found that models without a direct contribution of Yana into Alaska have likelihoods lower than the full model by more than 50 log10 units (Table S7.5), suggesting that Yana contribution is an important event to fit the SFS data. When comparing the likelihood distributions of different nested models, we found that the models without Kolyma and Even contribution to Alaska have similar likelihoods to the full model (Figure S7.4a). This indicates that the direct contribution of Evens and Kolyma into Alaskan are not required to fit the SFS data. Furthermore, under the full model, the likelihood profile for the contributions of Kolyma and Even reach a maximum close to zero. These results suggest that the a model with $\lambda=\kappa=0$, i.e. only with a non-zero contribution of Yana into Alaska can explain the SFS data (Figure S7.5). Interestingly, even in the model without a direct Yana contribution, the best parameter estimates indicate ~14.5% admixture of indirect Yana contribution into Alaska. This is because the inferred times of admixture are the same for Kolyma into Alaska and Yana into Kolyma. Given the inferred admixture proportions of Kolyma into Alaska of 87.3% and of Yana into Kolyma of 16.6%, this means that Alaska receives 14.5% (0.873*0.166) of lineages from Yana through admixture with Kolyma. Taken together, we interpret the parameter estimates and the likelihood values of the nested models as evidence for a scenario with a direct Yana contribution into Alaska (Table S7.5 and S7.6).

Across all models, we infer that Alaska is more closely related to Kolyma than to Han and Even, i.e. TEv is older than TKo (Table S7.6). The parameter estimates under the best nested model with a direct Yana contribution further indicate a contribution of Yana into Alaska and Kolyma at similar times (Table S7.6, Figure S7.5). Given that we infer similar times and proportions for these two independent contributions, it raised the possibility that there was a single Yana contribution into the ancestors of Kolyma and Alaska. To test this possibility, we considered models with a single Yana contribution into ancestors of Alaska and Kolyma, and re-estimated the parameters and maximum likelihood across the full and nested models. Models with a single Yana admixture event into ancestors of Kolyma and Alaska reach slightly better likelihood values than the models with two independent Yana contributions, within 50 log10 units of its full model (Table S7.5). This indicates that a scenario with a single Yana admixture event can explain the data as well. Interestingly, as seen for the models with two independent Yana admixture events, the best nested model is the one with a direct Yana contribution (i.e, a nested model with $\lambda=\kappa=0$). This is supported by the maximum likelihood estimates across models (Figures S7.4b, Table S7.5), and the likelihood profiles obtained under the full model (Figure S7.5). Parameter estimates are similar to the ones under the model with two independent Yana contributions into Kolyma and Alaska, with overlapping confidence intervals (Table S7.8, Figure S7.6). Confidence intervals are wide, especially for the time of events. We note that this uncertainty is expected due to the limited sample sizes (1 ancient Alaskan, 1 Kolyma and 1 Yana diploid individuals). The conclusions that Kolyma is the closest related population to Alaska (TKo < TEv), and that Alaska received ~18% of Yana are robust to varying levels of Neanderthal admixture (Table S7.9).

Karitiana

As seen for models with Alaskan, the maximum likelihood for each model indicates that the best nested model is the one with a direct Yana contribution and no Kolyma and Even direct contribution to Karitiana (i.e., the nested model with $\lambda=\kappa=0$, Table S7.5, Figure S7.4). However, it is worth noting that the point estimates obtained across models suggest that there are two equally likely solutions: (i) either an older split of Kolyma (30.6kya) followed by a large contribution of Kolyma into Karitiana (37.3%), which is supported by the full model; or (ii) a more recent split of Karitiana from Kolyma (<25kya) with no later contribution. Given that the latter scenario was consistently found for simpler models (with less parameters) that reached likelihood values close to the full model, we interpret this as evidence for a scenario without a direct contribution of Kolyma and Even into Karitiana. As expected if Alaskan and Karitiana shared a common ancestry, we obtained consistent estimates for the population tree topology and times of split whether we used the Alaskan or Karitiana populations to represent Native Americans. Indeed, most point estimates are similar, and the CI95% overlaps considerably for most parameters (Table S7.8, S7.10). This suggests that Alaskan and Karitiana share most of their demographic history. However, there are some differences, pointing to important differences. First, the contribution of Yana is larger for Alaskan (18.3%; CI95% 9.8-20.3) than for Karitiana (8.4%; CI95% 3.2-11.7) and with little overlap on the 95% confidence intervals. Second, in contrast to what we found for Alaska, for Karitiana, models with a single Yana contribution into the ancestors of Kolyma and Karitiana do not fit the data as well as the best model with two independent Yana contributions (Table S7.5). Nevertheless, rather than investigating the relationship between Alaska and Karitiana, our main goal was to elucidate the relationship between Siberian populations with Native Americans. Based on our demographic modeling results, our main conclusion is that the SFS data supports that Kolyma is the closest related population to Native Americans, and that Yana contributed to the ancestors of both the ancient sample from Alaska and the present-day sample from Karitiana.

Table S7.5. Maximum likelihood estimates for each model obtained during the parameter optimization for the datasets with Alaska and Karitiana. Best nested model indicated in bold. #param: number of parameters. Estimated ML: estimated maximum likelihood in log10 units. Δlhood corresponds to the difference between the estimated likelihood and the maximum possible likelihood if there was an exact fit to the observed SFS, and hence the closer to zero the better the fit. ΔML: Difference of log-likelihood of a given model in relation to the best model with the maximum inferred likelihood. For Karitiana we only tested the full model and the best nested model for models with a single pulse of admixture of Yana into ancestors of Kolyma and Karitiana.

		Alaska			**Karitiana**		
Models with two independent Yana admixture events	#param	Estimated ML	Δlhood	ΔML	Estimated ML	Δlhood	ΔML
Full model	15	-2,950,846	-7,279	-21	-3,201,579	-10,976	-1
Nested models:							
- no Even contribution (λ=0)	13	-2,950,841	-7,275	-17	-3,201,589	-10,987	-11
- no Kolyma contribution (κ=0)	13	-2,950,827	-7,261	-3	-3,201,609	-11,006	-31
- no Yana contribution (γ=0)	13	-2,950,909	-7,342	-84	-3,201,832	-11,229	-253
- no Even, no Kolyma contribution (λ=κ=0)	11	-2,950,849	-7,282	-24	-3,201,578	-10,975	0
- no contribution (λ=κ=γ=0)	9	-2,951,510	-7,943	-685	-3,201,810	-11,207	-232
Models with a single Yana admixture event							
Full model	14	-2,950,830	-7,263	-5	-3,201,623	-11,020	-45
Nested models:							
- no Even contribution (λ=0)	12	-2,950,824	-7,258	0	-		-
- no Kolyma contribution (κ=0)	12	-2,950,840	-7,274	-16	-		-
- no Yana contribution (γ=0)	12	-2,951,631	-8,064	-806	-		-
- no Even, no Kolyma contribution (λ=κ=0)	10	-2,950,833	-7,266	-8	-3,201,653	-11,050	-75
- no contribution (λ=κ=γ=0)	8	-2,951,754	-8,188	-930	-		-

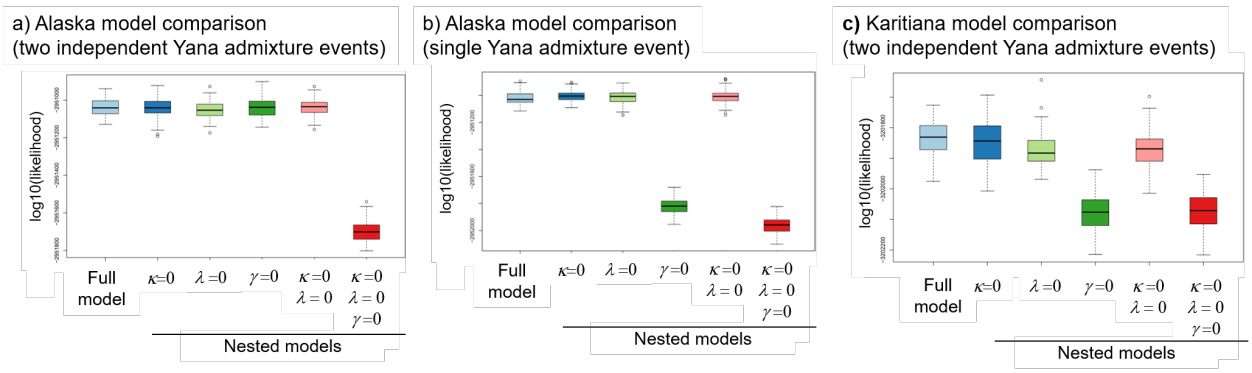


Figure S7.4. Maximum likelihood estimates of full and nested models suggest a direct contribution of Yana into Alaska and Karitiana. Likelihoods obtained for the full and nested models in log10 units, for the model with Alaska (a,b) and Karitiana (c). The distributions correspond to the likelihood values obtained by performing 100 simulations with the maximum likelihood point estimates for each model, approximating the SFS with 10^6 coalescent simulations. In all cases a nested model without Kolyma and Even admixture into Native Americans reached a likelihood similar to the full model, indicating that it can explain the SFS data as well as the full model.

Table S7.6. Point estimates for Alaskan demographic history for models with two independent Yana admixture events. Search range of parameters are indicated and compared with the maximum likelihood estimates for each model. Asian refers to ancestral of all Asian populations (Han, Even, Kolyma, Alaska, Karitiana). As - Alaskan sampling time (11.5Kya); TEv - time of split of (Han,Even) ancestor from Asian lineage); TKo - time of split of Kolyma from Asian lineage; Tadm – admixture times. min() – minimum function. The best nested model is shown in bold. Maximum likelihood estimates for each model are shown in Table S7.5.

	Search range		Full	Nested models				
	lower	upper	model	λ=0	κ=0	γ=0	κ=λ=0	κ=λ=γ=0
Effective sizes (diploid size)								
Ne ancestral Asian	500	25,000	16,449	6,689	10,537	16,979	**14,763**	11,065
Ne Bottleneck Asian	5	250[b]	46	52	50	46	**48**	50
Ne Alaskan	500	5,000[b]	1,623	1,527	1,847	1,457	**3,907**	2,892
Ne Ancestral Alaskan	500	5,000[b]	1,484	2,590	1,803	4,021	**1,786**	4,385
Ne Bottleneck	5	250[b,l]	88	71	73	104	**73**	78
Admixture contributions (%)								
Yana to Alaskan (γ)	10^{-5}	100[b,l]	15.76	16.11	17.33	-	**18.31**	-
Kolyma to Alaskan (κ)	10^{-5}	100[b,l]	0.20	0.00	-	87.31	**-**	-
Even to Alaskan (λ)	10^{-5}	100[b,l]	4.36	-	0.00	0.03	**-**	-
Time events (kya)								
Split Even (TEv)	19.7	38.3[b]	29.8	33.3	28.5	30.6	**30.3**	37.4
Split Kolyma (TKo)	13.3	38.3[b]	24.2	23.7	22.5	25.1	**24.2**	25.8
Bottleneck Alaska	As	min(TAdm)	12.2	15.2	15.3	18.3	**15.0**	19.4
Time of admixture (kya)								
Yana to Kolyma	Ks	TKo[b]	21.6	20.7	19.3	24.8	**20.2**	17.3
Yana to Alaska	As	min(TEv, TKo)[b]	13.2	17.0	16.8	-	**19.7**	-
Kolyma to Alaska	As	TKo[b]	23.0	23.1	-	24.8	**-**	-
Even to Alaska	As	TEv[b]	21.2	-	20.3	20.6	**-**	-

[b] – bounded upper search range, [b,l] – log-uniform search range with upper bound.

Table S7.7. Point estimates for Alaskan demographic history for models with a single Yana admixture event into ancestors of Kolyma and Alaska. Search range of parameters are indicated and compared with the maximum likelihood estimates for each model. Asian refers to ancestral of all Asian populations (Han, Even, Kolyma, Alaska, Karitiana). As - Alaskan sampling time (11.5Kya); TEv - time of split of (Han,Even) ancestor from Asian lineage); TKo - time of split of Kolyma from Asian lineage; Tadm – admixture times. min() – minimum function. The best nested model is shown in bold. Maximum likelihood estimates for each model are shown in Table S7.5.

	Search range		Full	Nested models				
	lower	upper	model	λ=0	κ=0	γ=0	κ=λ=0	κ=λ=γ=0
Effective sizes (diploid size)								
Ne ancestral Asian	500	25,000	13,778	9,102	16,536	24,627	**21,739**	22,289
Ne Bottleneck Asian	5	250[b]	47	49	47	59	**44**	50
Ne Alaskan	500	5,000[b]	833	4,124	1,282	3,895	**1,755**	1,130
Ne Ancestral Alaskan	500	5,000[b]	3,648	4,407	3,865	1,226	**4,303**	4,877
Ne Bottleneck Alaskan	5	250[b,l]	116	46	80	88	**67**	118
Admixture contributions (%)								
Yana to Alaska, Kolyma (γ)	10^{-5}	100[b,l]	11.94	13.75	13.50	-	**15.09**	-
Kolyma to Alaskan (κ)	10^{-5}	100[b,l]	0.02	0.00	-	61.41	**-**	-
Even to Alaskan (λ)	10^{-5}	100[b,l]	0.00	-	0.16	0.08	**-**	-
Time events (kya)								
Split Even (TEv)	19.7	38.3[b]	32.5	32.8	30.5	23.0	**31.9**	35.1
Split Kolyma (TKo)	13.3	38.3[b]	21.9	21.3	21.1	24.7	**21.5**	21.5
Bottleneck Alaska	As	min(TAdm)[b]	15.7	14.0	17.7	15.3	**18.4**	17.3
Time of admixture (kya)								
Yana to anc. Alaska, Kolyma	TKo	38.3[b]	23.1	29.7	29.1	-	**29.1**	-
Kolyma to Alaska	As	TKo[b]	17.0	17.2	-	23.0	**-**	-
Even to Alaska	As	TEv[b]	19.2	-	18.6	16.9	**-**	-

[b] – bounded upper search range, [b,l] – log-uniform search range with upper bound.

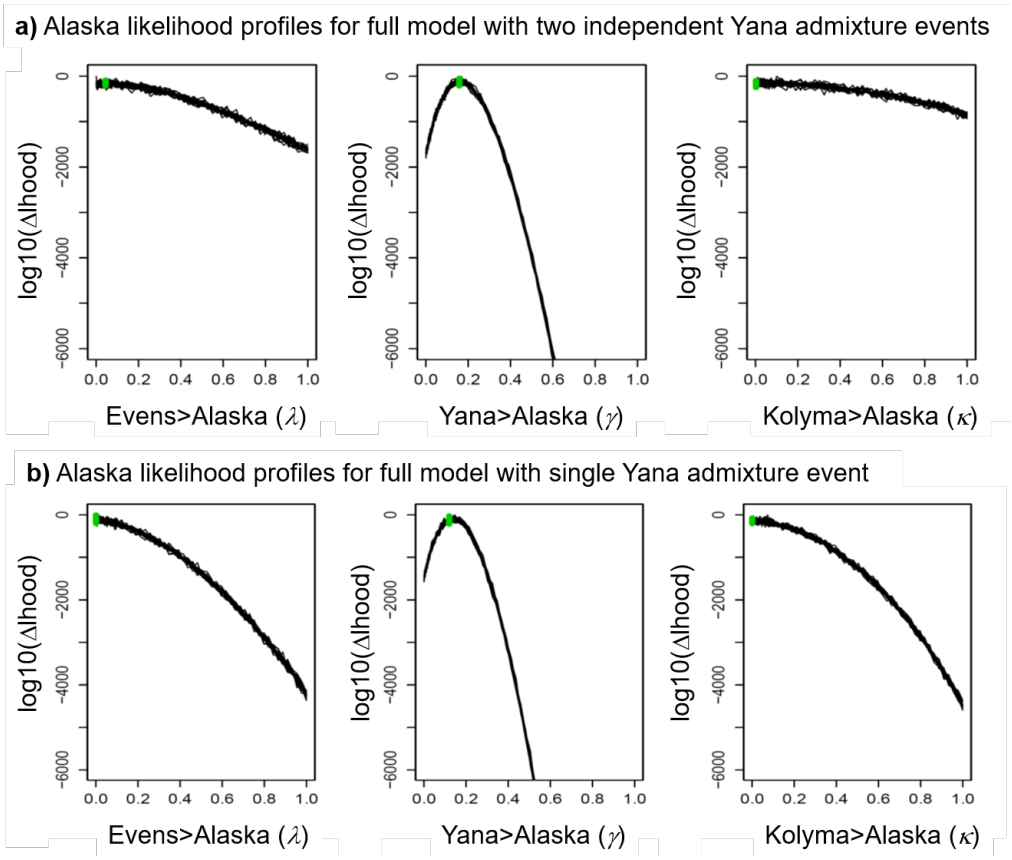


Figure S7.5. Likelihood profiles for each admixture parameter under the full model (fixing all the other parameters to the maximum likelihood estimate) for the model with Alaska. Each line corresponds to 1 of 20 runs for the difference between the estimated likelihood and the maximum likelihood across all runs and parameter values, using 10^6 coalescent simulations to approximate the expected SFS.

Table S7.8 Point estimates and 95% confidence intervals for the parameters of the best models obtained for the Alaskan demographic history. Confidence intervals were calculated according to the percentile method. Point estimates correspond to the parameters inferred with the original data set. Times of divergence in years are obtained assuming a generation time of 29 years and a mutation rate of 1.25e-8/gen/site.

	Alaska Two independent Yana admixture events			**Alaska Single Yana admixture with ancestor Kolyma, Alaska**		
	Point estimate	95% CI		Point estimate	95% CI	
Effective sizes						
Ne ancestral Asian	**14,763**	2,719	22,794	**21,739**	3,647	24,287
Ne Bottleneck Asian	**48**	43	72	**44**	43	67
Ne Alaska	**3,907**	954	4,309	**1,755**	788	4,356
Ne Ancestral Alaska	**1,786**	994	4,063	**4,303**	2,413	4,810
Ne Bottleneck Alaska	**73**	43	136	**67**	38	130
Admixture contributions (%)						
Yana to Alaska (γ)	**18.31**	9.84	20.33	**-**	-	-
Yana to Kolyma (fixed)	**16.6**	-	-	**-**	-	-
Yana to ancestor (Alaska, Kolyma)	**-**	-	-	**15.09**	11.43	17.72
Time events (kya)						
Split Even (TEv)	**30.3**	26.8	36.4	**31.9**	26.1	35.3
Split Kolyma (TKo)	**24.2**	20.9	27.9	**21.5**	18.7	24.0
Bottleneck Alaska	**15.0**	12.5	19.5	**18.4**	13.6	20.9
Time admixture (kya)						
Yana to Kolyma	**20.2**	15.5	23.7	**-**	-	-
Yana to Alaska	**19.7**	13.3	23.5	**-**	-	-
Yana to anc. (Alaska, Kolyma)	**-**	-	-	**29.1**	21.6	32.2

Table S7.9. Effect of Neanderthal admixture in point estimates under the best nested models for Alaskan demographic history. We re-run the parameter estimation varying the Neanderthal contribution into ancestral of Eurasians (adNeaEurasia) and the extra Neanderthal admixture pspecific to the Asian lineage (adNeaAsi). By default we used adNeaEurasia=3% and adNeaAsi=0.5% (Malaspinas et al. 2016). Times of divergence in years are obtained by assuming a generation time of 29 years and a mutation rate of 1.25e-8/gen/site.

	Alaska			**Alaska**		
adNeaEurasia=	**3%**	2%	1%	**3%**	2%	1%
adNeaAsi=	**0.5%**	0.25%	0.1%	**0.5%**	0.25%	0.1%
Effective sizes						
Ne ancestral Asian	14,763	9,814	4,618	21,739	4,686	3,193
Ne Bottleneck Asian	48	52	82	44	60	84
Ne Alaska	3,907	1,716	3,023	1,755	1,051	2,752
Ne Ancestral Alaska	1,786	1,336	1,447	4,303	4,197	3,924
Ne Bottleneck Alaska	73	95	103	67	67	53
Admixture contributions (%)						
Yana to Alaska (γ)	18.31	17.34	14.00	-	-	-
Yana to ancestor (Alaska, Kolyma)	-	-	-	15.09	14.95	14.65
Time events (kya)						
Split Even (TEv)	30.3	28.8	28.4	31.9	32.7	32.4
Split Kolyma (TKo)	24.2	23.7	23.6	21.5	20.3	21.1
Bottleneck Alaska	15.0	13.4	13.9	18.4	15.9	18.1
Time admixture (kya)						
Yana to Kolyma	20.2	18.6	20.4	-	-	-
Yana to Alaska	19.7	14.2	18.2	-	-	-
Yana to anc. (Alaska, Kolyma)	-	-	-	29.1	30.9	28.3
Δlikelihood	-7,282	-7,397	-7,886	-7,266	-7,359	-7,872

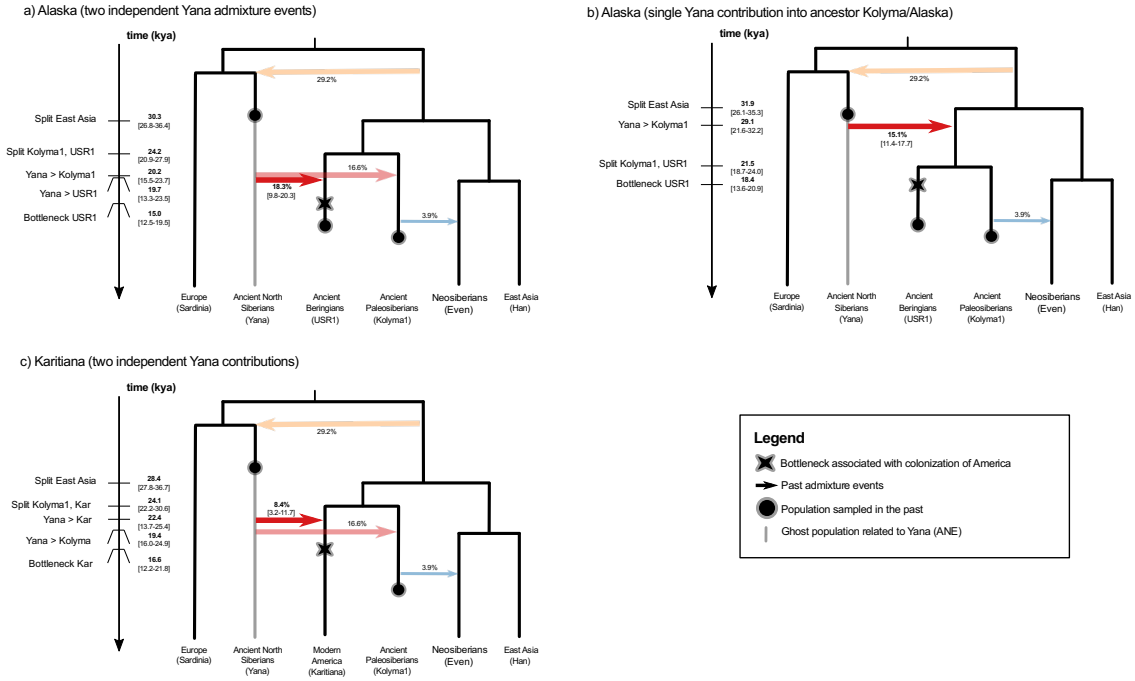


Figure S7.6. Inferred demographic history for Native American populations. Schematic representation of the parameter estimates obtained for the models with Native American populations. Point estimates are shown in bold, and 95% confidence intervals are shown within square brackets. Sampling times are represented by filled circles, the bottleneck by a filled star and admixture events by arrows. Parameters related with Native American populations were inferred by adding a Native American population to the Siberian replacement model, fixing all the parameters related with Eurasian and Siberian populations (see text for details). Times of events are shown in kya on the left, and the admixture estimates in percentage are shown above the arrows. Bold corresponds to the point estimates and the 95% confidence intervals obtained with a non-bootstrap approach are shown within square brackets. Times of divergence in years are obtained assuming a generation time of 29 years and a mutation rate of 1.25e-8/gen/site.

Table S7.10. Karitiana demographic history, with inferred parameters for the full and nested models considered with two independent Yana admixture events and a single Yana admixture. Kar - Karitiana. Kol - Kolyma. anc. – ancestors. Search range of parameters were the same as for the Alaska demographic history (see Tables S7.6 and S7.7). Maximum likelihood estimates are shown in Table S7.5.

	Two independent Yana admixture events						Single Yana admixture	
	Full model	Nested models					Full model	Nested
		λ=0	κ=0	γ=0	κ=λ=0	κ=λ=γ=0		κ=λ=0
Effective sizes (diploid dize)								
Ne ancestral Asian	9,413	11,129	19,394	10,811	6,309	15,485	16,874	19,076
Ne Bottleneck Karitiana	17	50	49	47	63	50	48	48
Ne Karitiana	2,899	1,442	3,233	3,342	4,305	3,302	3,716	2,786
Ne Ancestral Karitiana	3,526	1,494	3,857	4,565	1,779	3,868	4,429	3,347
Ne Bottleneck Karitiana	24	128	23	26	25	26	19	24
Admixture contributions (%)								
Yana to Karitiana (γ)	8.20	8.91	6.76	-	8.37	-	-	-
Yana to anc. (Kar,Kol)	-	-	-	-	-	-	12.76	9.95
Kolyma to Karitiana (κ)	37.27	0.00	-	0.10	-	-	0.10	-
Even to Karitiana (λ)	0.00	-	1.56	0.01	-	-	32.94	-
Time events (kya)								
Split Even (TEv)	30.6	30.5	33.2	36.2	28.4	35.5	33.4	28.8
Split Kolyma (TKo)	30.6	27.4	25.3	27.9	24.1	28.0	22.2	22.8
Bottleneck Karitiana	14.2	16.7	15.1	15.8	16.6	14.4	12.9	16.4
Time admixture (kya)								
Yana to Kolyma	21.6	17.0	21.5	19.7	19.4	17.3	-	-
Yana to Karitiana	16.8	22.4	21.1	-	22.4	-	-	-
Yana to anc. (Kar,Kol)	-	-	-	-	-	-	25.8	23.4
Kolyma to Karitiana	26.2	23.7	-	23.6	-	-	16.5	-
Even to Karitiana	20.1	-	16.0	21.9	-	-	20.2	-

[b] – bounded upper search range, [b,l] – log-uniform search range with upper bound.

Table S7.11 Point estimates and 95% confidence intervals for the parameters of the best models obtained for the Karitiana model with two independent Yana admixture events. Confidence intervals were calculated according to the percentile method. Point estimates correspond to the parameters inferred with the original data set. Times of divergence in years are obtained assuming a generation time of 29 years and a mutation rate of 1.25e-8/gen/site.

	Karitiana		
	Point estimate	95% CI	
Effective sizes			
Ne ancestral Asian	**6,309**	2,212	20,664
Ne Bottleneck Asian	**63**	45	118
Ne Native American	**4,305**	1,749	4,725
Ne Ancestral Native American	**1,779**	1,641	4,677
Ne Bottleneck Native American	**25**	20	59
Admixture contributions (%)			
Yana to Native American (γ)	**8.37**	3.19	11.65
Time events (kya)			
Split Even (TEv)	**28.4**	27.8	36.7
Split Kolyma (TKo)	**24.1**	22.2	30.6
Bottleneck Karitiana	**16.6**	12.2	21.8
Time of admixture (kya)			
Yana to Karitiana	**22.4**	13.7	25.4
Yana to Kolyma	**19.4**	16.0	24.9

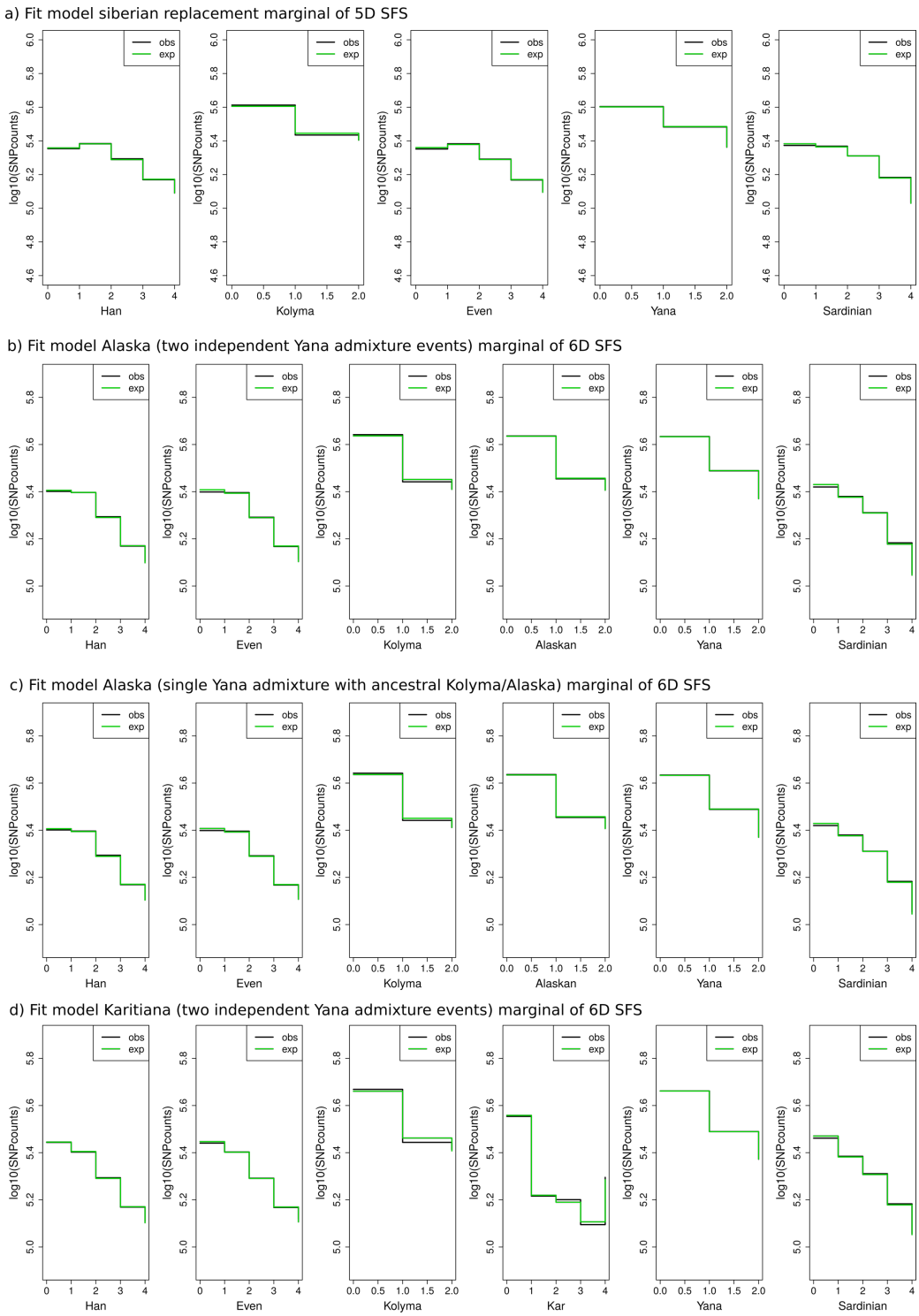


Figure S7.7 Comparison of the marginal observed and marginal expected SFS for the best model for the dataset for (a) Siberian replacement, (b) Alaska (model with two independent Yana contributions), (c) Alaska (model with single Yana contribution into ancestor Kolyma and Alaska) and (d) Karitiana. For each row marginal 1D-SFS is show for each sampled population in the corresponding model. The x-axis shows the derived allele frequencies (allele counts) and the y-axis shows the number of SNPs with a given frequency (in log10 scale). The expected SFS was obtained as the average of 100 simulated SFSs (each approximated with 106 coalescent simulations), according to the maximum-likelihood parameter estimates obtained with the original dataset for the corresponding best model.

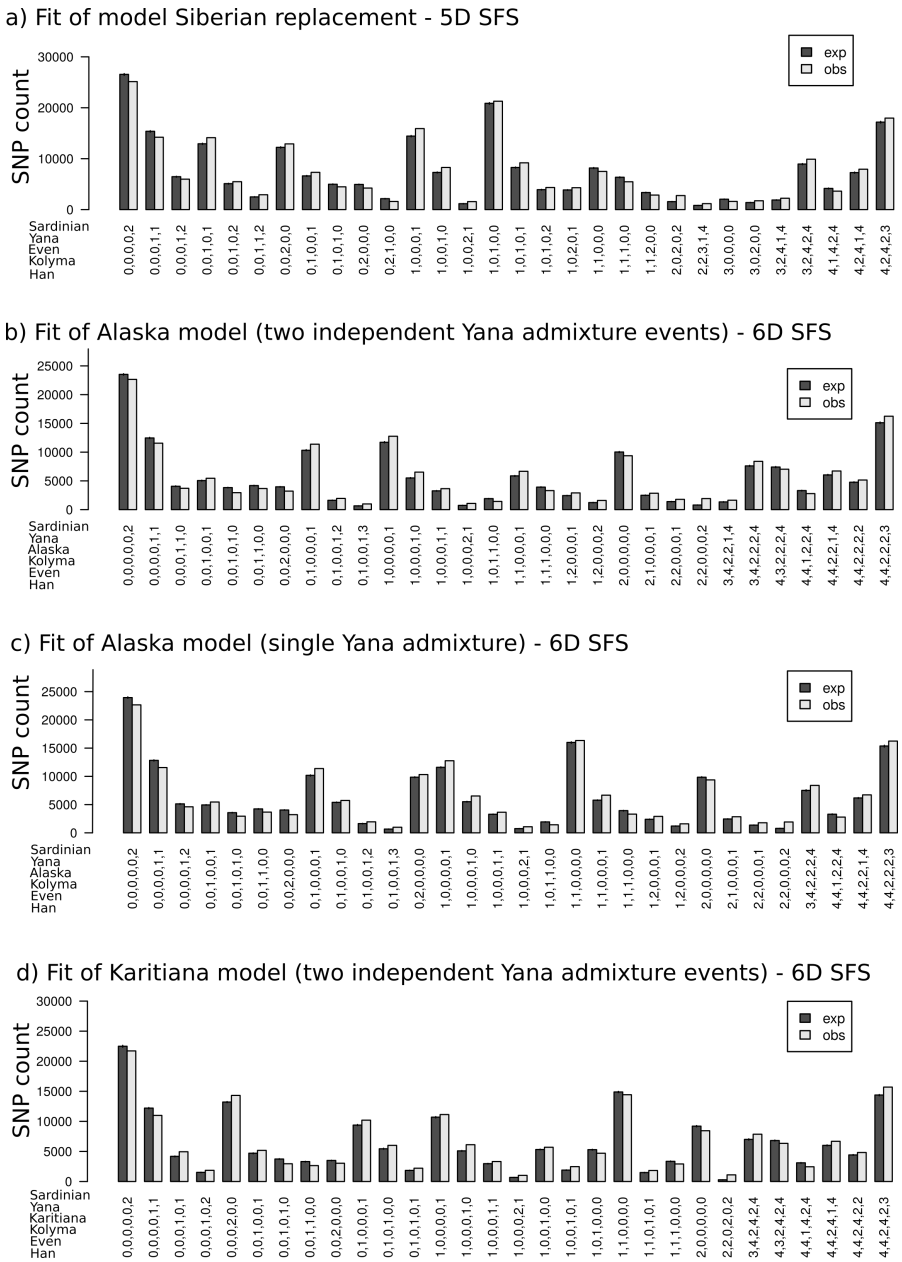


Figure S7.8 Comparison of the multidimensional joint observed and expected SFS for the 30 entries showing the worst fit (a) for Siberian replacement model in the multidimensional 5D joint SFS (out of the 1125 entries); (b) Alaskan model in the multidimensional 6D joint SFS (out of 3375 entries) for model with two independent Yana contributions into Kolyma and Alaska; (c) Alaskan model in the multidimensional 6D joint SFS (out of 3375 entries) for model with single Yana contributions into ancestor of Kolyma and Alaska; and (d) Karitiana model in the multidimensional 6D joint SFS (out of 5625 entries) for the model with two independent Yana contributions into Kolyma and Karitiana. We selected the 30 entries with the largest difference between the expected and observed SFS (i.e. larger $|(m_i\text{Log10}(p_i))- (m_i\text{Log10}(m_i/L)|$, where m_i is the observed counts at the i-th entry, p_i is the expected SFS at the i-th entry and L is the total number of polymorphic sites). Each column corresponds to one entry of the SFS, coded from bottom to top as (a) *h,k,e,y,s*; (b) *h,e,k,a,y,s*; (c) *h,e,k,a,y,s* and (d) *h,e,k,ka,y,s*, where these entries correspond to the frequency of the derived allele in Han (*h*), Kolyma (*k*), Even (*e*) , Alaskan (*a*), Karitiana (*ka*), Yana (*y*) and Sardianian (*s*). Expected SFS were obtained as the average of 100 simulated SFSs (approximated with 106 coalescent simulations), according to the parameter estimates obtained under the best model for the Siberian populations original dataset. Error bars correspond to the 0.01 and 0.99 quantiles of the 100 simulated SFSs.

Assessing the fit of the SFS

To visualize how well our model could reproduce the observed data, we compared the marginal distribution of the observed and expected SFS (Figures S7.7). Note that we performed the estimates by discarding the singletons, and hence the marginal likelihood is computed without those entries. Overall, we have a very good fit of the expected to the observed marginal SFS, suggesting that our model and the corresponding parameter estimates capture relevant aspects of the data. We also looked in more details at the joint SFS to find the entries that were not well explained by our model. Overall, even the worst fitted entries are relatively well predicted/fitted under our model (Figure S7.8).

Simulation study

We performed a simulation study to assess whether it is possible to infer demographic parameters of complex models from SFS data from datasets similar to ours, i.e. with sample sizes ranging from 1 to 4 individuals per population and discarding the singletons. We simulated SFS data according to the best model obtained for the Siberian demographic history, since it was the case with more inferred parameters. We simulated 100 datasets according to the point estimates obtained under the best model, which supported a replacement scenario with limited admixture. The observed SFS was obtained from 625 blocks with 1Mb, but due to missing data there were less sites. To mimic the observed data, keeping a similar number of sites and SNPs, while being conservative by simulating linked sites, we generated 600 blocks with 1.02 Mb with the same sample sizes as in observed data (rather than 625 blocks of 0.98 Mb). Recombination rate was assumed to be constant and 80% of the mutation rate, i.e. 1.0×10^{-8} per generation per pair of adjacent sites. This resulted in datasets with 613,981,800 sites and a number of SNPs ranging from 1,304,579 to 1,335,662, similar to the 613,981,492 sites and 1,319,809 SNPs in the observed SFS. We also tested whether we have power to detect continuity by analysing data simulated according to a continuity scenario, where Kolyma would descend from Yana, and Even would descend from Kolyma. We have thus simulated 100 datasets keeping the same parameters, but fixing the admixture proportions of Yana into Kolyma, and of Kolyma into Even to 90%. In sum, we have analysed datasets simulated according to two scenarios: (i) replacement and (ii) continuity. To find the maximum likelihood estimates for each simulated dataset, we analysed them with exactly the same settings as done for the observed data. The exception was that, due to computational constraints, we performed 20 independent runs rather than 100. Importantly, as done for the observed data to avoid sequencing errors, we ignored the singletons in the likelihood computations.

Although this is not an extensive simulation study, results indicate that the mean and median of the parameter estimates are very close to the true parameter values (Tables S7.12 and S7.13), suggesting that we have power to infer these parameters. Importantly, results indicate that with the SFS data we can infer the admixture proportions reasonably well (Table S7.12, Figure S7.9), allowing us to distinguish between a scenario of replacement from a scenario of continuity. Despite the higher uncertainty, the mean estimates of the time of events (Table S7.12) and effective sizes (Table S7.13) were also close to the correct values.

The results of simulations under the replacement scenario can also be used to obtain the 95% confidence intervals according to a parametric bootstrap approach. As expected due to the small sample sizes, most parameter estimates were associated with a wide range of values around the mean and median, comparable to the 95% CI obtained with the non-parametric bootstrap approach (Table S7.14).

Table S7.12 Results of simulation study for admixture proportions, migration and time of events. Comparison of point estimates (mean and median) and range of values (lower, 2.5%, and upper, 97.5%, quantiles) across simulations with the true value of parameters, for a scenario of replacement (best nested model for Siberian

demography), and a scenario of continuity (admixture proportions of 90%). The asterisk (*) indicates values similar to replacement scenario.

	Replacement scenario				Continuity scenario			
	True	Mean	Lower	Upper	True	Mean	Lower	Upper
Admixture proportions (%)								
Kolyma>Even (τ)	**3.9**	4.6	0.0	10.1	**90.0**	88.8	85.3	92.2
Yana>Kolyma (β)	**16.6**	15.4	8.6	20.6	**90.0**	89.5	78.0	97.2
AsiaContinent>Yana (ω)	**29.2**	34.1	24.2	56.6	*	40.9	23.5	78.3
Scaled migration rate (2Nm)								
Han > Sardinian	**3.7**	1.062	0.001	2.734	*	1.020	0.000	3.295
Times of split (Kya)								
Bottleneck OoA (TBot)	**45.1**	48.0	38.4	63.4	*	46.8	38.0	74.6
Split Asia/Europe (TAsEu)	**43.1**	41.4	32.8	53.6	*	39.6	33.2	49.8
Split Yana (TYa)	**38.7**	38.5	31.9	50.9	*	36.7	31.6	45.0
Split Kolyma (TKo)	**27.1**	24.5	12.3	33.4	*	24.7	12.4	33.7
Split Han,Even (TEv)	**19.6**	17.3	11.1	28.3	*	16.9	10.3	30.1
Times of admixture (Kya)								
Kolyma>Even	**13.3**	13.6	10.1	20.4	*	11.9	9.9	15.7
Yana>Kolyma	**25.8**	21.2	12.0	29.7	*	20.8	11.5	32.2

Table S7.13 Results of simulation study for effective sizes. Comparison of point estimates (mean and median) and range of values (lower, 2.5%, and upper, 97.5%, quantiles) across simulations with the true value of parameters, for a scenario of replacement (best nested model for Siberian demography), and a scenario of continuity (admixture proportions of 90%). The asterisk (*) indicates values similar to replacement scenario.

		Replacement scenario			Continuity scenario		
Effective sizes (diploid)	true	mean	lower	upper	mean	lower	upper
Ne ancestral humans/Neanderthal	**15,560**	15,851	15,176	16,506	15,873	15,334	16,397
Ne ancestral modern humans	**12,142**	11,807	11,129	12,650	11,743	11,163	12,608
Ne ancestral non-Africans	**15,258**	10,243	2,043	21,490	10,772	1,998	25,101
Ne bottleneck Out of Africa	**59**	79	49	185	81	50	196
Ne ancestral Asian continent	**16,613**	12,971	1,101	26,388	12,427	2,163	26,697
Ne bottleneck Asian continent	**44**	49	37	115	61	36	174
Ne ancestral Europe	**18,454**	13,303	3,336	24,722	14,030	4,356	23,587
Ne Sardinians	**10,173**	11,119	7,866	17,086	11,400	7,555	20,416
Ne bottleneck Europe	**219**	184	121	238	178	88	238
Ne ancestral Han,Even	**20,668**	15,772	5,296	26,889	13,485	3,457	23,784
Ne Han Chinese	**38,232**	29,803	10,771	47,680	25,311	6,067	48,872
Ne bottleneck Han Chinese	**176**	183	120	224	169	68	244
Ne Even	**25,791**	21,542	11,972	36,239	18,057	9,824	29,275
Ne bottleneck Even	**216**	196	146	243	129	43	211
Ne Kolyma	**1,648**	1,934	296	8,230	1,358	263	3,084
Ne bottleneck Kolyma	**99**	97	23	213	122	21	215
Ne Yana	**4,155**	13,435	4,071	23,417	10,358	2,011	22,596
Ne ancestral Yana	**6,257**	7,193	225	19,383	5,558	207	21,200
Ne bottleneck Yana	**38**	64	30	206	68	29	210

Table S7.14 Comparison of confidence intervals obtained with a parametric and non-parametric bootstrap approach, for the best nested model for the Siberian demographic history.

		Non-parametric bootstrap 95%CI interval		Parametric bootstrap 95%CI interval	
	Point estimate	lower	upper	lower	Upper

Effective population sizes (diploid size)					
Ne ancestral humans/Neanderthal	**15,560**	15,261	16,822	15,176	16,506
Ne ancestral modern humans	**12,142**	11,211	12,536	11,129	12,650
Ne ancestral non-Africans	**15,258**	1,574	23,683	2,043	21,490
Ne bottleneck Out of Africa	**59**	48	182	49	185
Ne ancestral Asian continent	**16,613**	4,093	23,651	1,101	26,388
Ne bottleneck Asian continent	**44**	37	57	37	115
Ne ancestral Europe	**18,454**	5,154	24,818	3,336	24,722
Ne Sardinians	**10,173**	7,504	15,235	7,866	17,086
Ne bottleneck Europe	**219**	154	242	121	238
Ne ancestral Han,Even	**20,668**	4,897	26,221	5,296	26,889
Ne Han Chinese	**38,232**	18,060	50,882	10,771	47,680
Ne bottleneck Han Chinese	**176**	155	246	120	224
Ne Even	**25,791**	9,696	26,860	11,972	36,239
Ne bottleneck Even	**216**	148	241	146	243
Ne Kolyma	**1,648**	1,170	9,378	296	8,230
Ne bottleneck Kolyma	**99**	20	224	23	213
Ne Yana	**4,155**	2,280	23,344	4,071	23,417
Ne ancestral Yana	**6,257**	1,442	20,643	225	19,383
Ne bottleneck Yana	**38**	28	80	30	206
Admixture proportions (%)					
Kolyma>Even (τ)	**3.9**	0.1	12.6	0.0	10.1
Yana>Kolyma (β)	**16.6**	7.5	22.2	8.6	20.6
AsiaContinent>Yana (ω)	**29.2**	21.3	40.1	24.2	56.6
Scaled migration rate (2Nm)					
Han > Sardinian	**3.65**	0.00	8.76	0.001	2.734
Times of split (Kya)					
Bottleneck OoA (TBot)	**45.1**	37.5	54.9	38.4	63.4
Split Asia/Europe (TAsEu)	**43.1**	33.4	48.6	32.8	53.6
Split Yana (TYa)	**38.7**	32.2	45.8	31.9	50.9
Split Kolyma (TKo)	**27.1**	17.1	32.1	12.3	33.4
Split Han,Even (TEv)	**19.6**	11.9	22.0	11.1	28.3
Times of admixture (Kya)					
Kolyma>Even	**13.3**	10.4	18.3	10.1	20.4
Yana>Kolyma	**25.8**	14.5	28.9	12.0	29.7

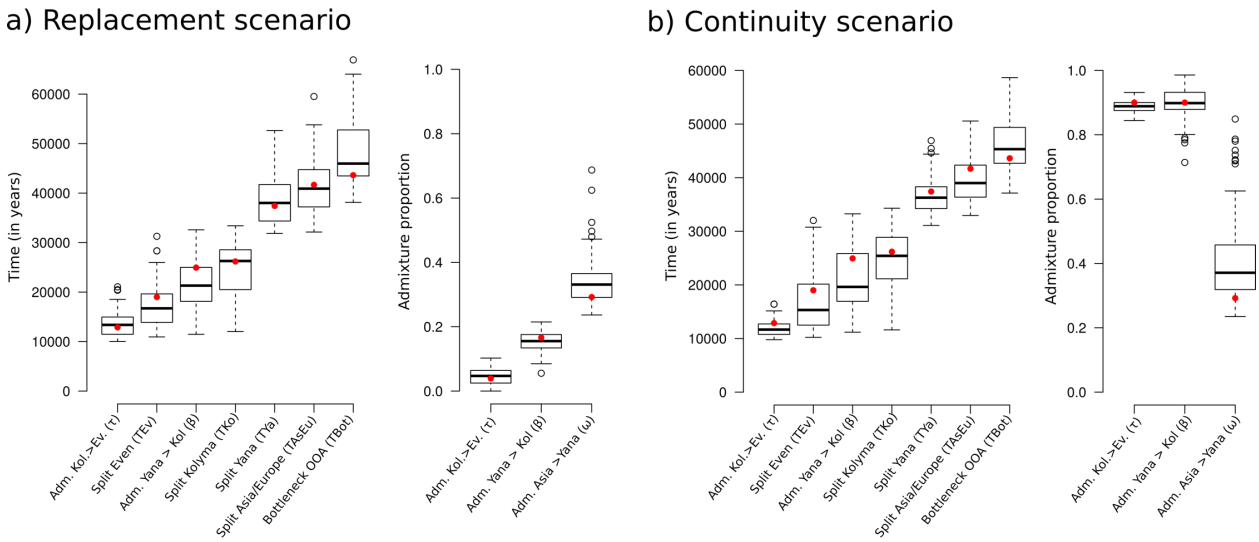


Figure S7.9 Results of simulation study. Estimates of times of admixture and admixture proportions for (a) the replacement scenario (best model for Siberian demographic history), and (b) continuity scenario. Boxplots represent the estimates obtained from analysis of 100 datasets with fastsimcoal with the same settings as with observed data (i.e. discarding singletons, see text for details). Red points indicate the true parameter values.

Discussion

Even though the models considered here are still a simplification of the likely complex demographic history of this geographic region, we have built parameter-rich specific models to test among alternative hypotheses, accounting for the potential confounding factors described above (e.g. different sampling times, population specific effective sizes, bottlenecks, Neanderthal admixture). Despite the inherent uncertainties of our model-based analyses reflected by some relatively wide confidence intervals, our composite likelihood approach has allowed us to distinguish among alternative hypotheses. In sum, our results indicate that:

(i) Yana diverged from European and Asian lineages ~38kya (95%CI 32-46) soon after their divergence ~43kya (95%CI 33-49), probably from European ancestors, but that they still show ~29% (95%CI 21-40) of Asian ancestry (Table S7.3, main Figure 3).

(ii) There were at least three waves of migration into Siberia, with almost complete replacement of Yana by Kolyma, and of Kolyma by Even (Table S7.3, main Figure 3).

(iii) Kolyma is the population most closely related to the Native American populations. This is indicated by the population tree topology supported, indicating that the Kolyma split from Alaskan/Karitiana (point estimates ranging from 21.5 to 24.2 kya), is younger than the split of the (Han, Even) lineage (point estimates ranging from 28.4 to 31.9 kya).

(iv) The population that colonized the Americas was likely related to the Kolyma, but also carried a Yana contribution of up to 11.7% for the Karitiana dataset, and up to 20.3% for the Alaskan dataset. In contrast, we do not find evidence of a significant contribution of the ancestors of present day Siberian Even into Alaska and Karitiana.

These results are in agreement with other recent population genomic studies based on present-day populations (Skoglund, Reich 2016; Nielsen et al. 2017), but add important insights, which were difficult to infer previously, mostly because the most closely related population to Native Americans are two ancient samples (Kolyma and Yana). Importantly,

our estimates indicate that Kolyma is the closest relative to Native American populations, and that Yana contributed up to 20% to Native Americans (Tables S7.6-S7.11, Figure S7.6).

We note that our estimates are associated with relatively wide confidence intervals, both for the modelling of the Siberian replacement population history and for the colonization of the Americas (Tables S7.3, S7.8, S7.11). These are likely due to the reduced sample sizes, especially of the ancient samples (1 Alaskan, 1 Kolyma and 1 Yana), and due to computational and statistical challenges of dealing with parameter-rich complex models, especially to find the overall combination of parameters that maximizes the likelihood. Nevertheless, the SFS-data together with the demographic modelling approach allowed us to test alternative hypothesis and reach the above conclusions that point to the following scenario. Furthermore, our simulation study indicates that it is possible to estimate admixture proportions and distinguish between models of replacement and continuity (Tables S7.12, S7.13, and Figure S7.9).

A possible scenario

Based on the point estimates obtained for the best models of the three datasets (Siberian replacement model, Alaskan and Karitiana) an overall picture emerges, suggesting: ~**43kya:** divergence of Western Europe and East Asia lineages; ~**39kya**: divergence of Yana from Europe lineage (with ~30% contribution of East Asian lineage); ~**30kya:** divergence of ancestors of (Han, Even) from ancestors of (Kolyma, Karitiana, Alaskan); ~**25kya:** divergence of Kolyma from ancestors of (Karitiana/Alaskan); ~**20kya:** Admixture of Yana with Kolyma and Native American lineages, probably associated with the movement of a Kolyma related lineage into Siberia, resulting in a contribution from Yana into Kolyma of ~17%, which is similar to the levels found in Alaskan (~18%), but higher than estimated for Karitiana (~8%); ~**20kya**: divergence of Han Chinese and Even; ~**15-18kya:** bottleneck in Alaskan and Karitiana lineage, probably resulting from founder events associated with the colonization of America; ~**13kya**: admixture of Even with Kolyma (~4%) probably associated with the movement of Even ancestors into Siberia.

It is noteworthy that our point estimates suggest a higher Yana admixture contribution into Alaska than into Karitiana. Furthermore, for the Alaska dataset we infer that a model with a single Yana contribution into the ancestors of Kolyma and Alaska is as likely as the model with divergence of Kolyma and Alaskan lineages previous to the admixture with Yana (i.e. with two independent Yana admixture events), but for Karitiana we find support for the two independent Yana contributions (Table S7.5). The model with two independent admixture events is in agreement with D-statistic and other admixture estimates results (Supplementary Information 7), and hence we favor that model. Nevertheless, the different results obtained for Karitiana and Alaska raise the possibility that Karitiana and Alaskans diverged from Kolyma at different times and have received different contributions from Yana, implying different waves in north east Siberia, and possibly into the Americas. However, we note that the 95% CI for the Yana contribution overlap very slightly for Karitiana and Alaska, and that a value of 10% admixture is compatible with both datasets, suggesting that it is still possible that Alaska and Karitiana have similar levels of admixture with Yana. Second, there are some alternative explanations for these differences. Assuming that all Native Americans descend from a single ancestral population, this difference could be mediated by the maintenance of gene flow over Beringia during some time. If the ancestors of Karitiana would had quickly moved south, splitting from Alaska soon after entering the Americas, the maintenance of gene flow over Beringia would mostly affect the Alaskan populations and increase the Yana contribution on those populations. Such period of gene flow between Native Americans and Siberia is supported by other analyses (Supplementary Information 7). Another explanation is

that there were two separate admixture events after the split of Alaskan and Karitiana lineages. This would be possible if the single ancestral Native American founder population was highly structured, or if there were two waves of migration into the Americas of two independent lineages (the Alaskan and Karitiana) following different routes. The fact that we infer very similar times of admixture, as well as times and intensity of bottlenecks in both populations, suggest that such two waves scenario is less likely, unless the two waves would have taken place very close in time. Finally, and still under a scenario where Alaskan and Karitiana descend from a unique founder population that experienced a single pulse of admixture with Yana, there are combinations of demographic and selective processes that could affect admixture proportions (e.g. (Harris, Nielsen 2016)). Because of these, the Yana admixture component could have been lost in Karitiana (e.g. similar to selection on Neanderthal alleles(Harris, Nielsen 2016)), or the Yana component could have been favored by selection in the Alaskan population. More data would be required to be able to distinguish among these models based on the SFS.

Although the timing of these events depends on two scaling parameters, the mutation rate and the generation time (assumed to be 1.25e-8/gen/site (Scally, Durbin 2012) and 29y per generations (Fenner 2005), respectively), we contextualize below the inferred demographic events in relation to geological and climatic changes during late Pleistocene in the area:

(i) Kolyma and Native Americans split from other Asian lineages just before the last glacial maximum (LGM) ~26.5kya. Then, after the LGM, they would have moved up North in Siberia, and met Yana descendants inhabiting those areas independently, admixing with various degrees with them, but to a large extent replacing them.

(ii) Even moved up north in Siberia after the younger Dryas during which they diverged from Han. This movement was associated with admixture with Kolyma ancestors in Siberia, but the progressively replaced the latter.

(iii) Native American ancestors had a bottleneck 15-18 kya probably when entering the Americas, after admixture with Yana descendants. This coincides with the period where crossing the Beringia was likely possible because sea level was not high enough to block the passage and the glaciers were melted up to a point that they were no longer blocking the migration corridor.

References

Abecasis, GR, A Auton, LD Brooks, MA DePristo, RM Durbin, RE Handsaker, HM Kang, GT Marth, GA McVean. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491:56-65.

Adams, AM, RR Hudson. 2004. Maximum-likelihood estimation of demographic parameters using the frequency spectrum of unlinked single-nucleotide polymorphisms. Genetics 168:1699-1712.

Cunningham, F, MR Amode, D Barrell, et al. 2015. Ensembl 2015. Nucleic Acids Res 43:D662-669.

Davison, AC, DV Hinkley. 1997. Bootstrap methods and their application: Cambridge University Press.

de Manuel, M, M Kuhlwilm, P Frandsen, VC Sousa, T Desai, J Prado-Martinez, J Hernandez-Rodriguez, I Dupanloup, O Lao, P Hallast. 2016. Chimpanzee genomic diversity reveals ancient admixture with bonobos. Science 354:477-481.

Excoffier, L, I Dupanloup, E Huerta-Sanchez, VC Sousa, M Foll. 2013. Robust demographic inference from genomic and SNP data. PLoS Genet 9:e1003905.

Excoffier, L, HE Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources 10:564-567.
Fenner, JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. American Journal of Physical Anthropology 128:415-423.
Flicek, P, MR Amode, D Barrell, et al. 2011. Ensembl 2011. Nucleic Acids Res 39:D800-806.
Harris, K, R Nielsen. 2016. The genetic cost of Neanderthal introgression. Genetics 203:881-891.
Malaspinas, A-S et al. 2016. A genomic history of Aboriginal Australia. Nature.
Mallick, S, H Li, M Lipson, I Mathieson, M Gymrek, F Racimo, M Zhao, N Chennagiri, S Nordenfelt, A Tandon. 2016. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538:201-206.
Moreno-Mayar, JV et al. 2018 Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. Nature 553: 203-207.
Nielsen, R. 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics 154:931-942.
Nielsen, R, JM Akey, M Jakobsson, JK Pritchard, S Tishkoff, E Willerslev. 2017. Tracing the peopling of the world through genomics. Nature 541:302-310.
Rosenbloom, KR, J Armstrong, GP Barber, et al. 2015. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res 43:D670-681.
Scally, A, R Durbin. 2012. Revising the human mutation rate: implications for understanding human evolution. Nature Reviews Genetics 13:745-753.
Skoglund, P, D Reich. 2016. A genomic view of the peopling of the Americas. Current opinion in genetics & development 41:27-35.

Supplementary Information 8 - Ensemble Species Distribution Models of Asian Palaeolithic modern humans

Climate Envelope Models, CEMs, have been used successfully to reconstruct the relationships between species and the climate conditions they inhabited across past time periods, (Nogués-Bravo et al. 2008; Lorenzen et al. 2011) including human species (Franklin et al. 2015; Giampoudakis et al. 2017). They provide predictions, spatially explicit hypothesis, grounded in strong theoretical understandings of the relation between species and the environment. We assess the temporal changes in the spatial distribution of the climatic suitability for Anatomically Modern Humans across central-north Asia between 48-12ka (1ka= thousand years ago) every one or two thousand years. We then identify what regions had the most suitable climatic conditions from 48 to 12ka (Fig x)

We collated a geo-referenced database of modern human fossil and archeological dated remains (INQUA Palaeolithic database v. 21a; Hamilton and Buchanan 2010; Pitulko et al. 2016) for central north Asia. We removed: records without radiocarbon dating lab codes or geographical coordinates, duplicated records from the same material, records associated with material that are ambiguously dated or potentially contaminated. The final dataset includes 936 modern human occurrences across all time intervals, spanning from 46ka to 12ka. Radiocarbon dates were calibrated to calendar years BP, using the online version of OxCal v4.3 software, and estimated the average values (mean calibrated range) of the 95.4% distribution (2 standard errors). To reduce the effect of the uneven distribution of the archaeological record across central-north Asia, all palaeoclimatic data were gridded to a 1x1 degree resolution, and all occurrences within a grid cell were aggregated to a single occurrence (see Fig. x).

Palaeoclimatic conditions were simulated under the HadCM3 (Hadley Centre Coupled Model, version 3) Atmospheric– Ocean General Circulation Model (AOGCM) (Singarayer et al., 2011). The simulations have time periods of 1000 years between 22 and 11 ka and of 2000 years before 22 ka, resulting in 23 intervals between 46 and 12 ka. We retrieved seasonal total precipitation and average temperature for each time step: spring, summer, autumn and winter. To avoid collinearity among climatic variables, we performed a Principal Component Analysis (PCA) and a Pearson's correlation estimate between all climatic variables per time period. We finally selected the three seasonal variables that maximized the climatic signal information and

minimized the collinearity: Autumn total precipitation, Summer average temperature and Autumn average temperature.

We used an ensemble of seven different algorithms to characterise the climatic niche of modern humans using Maxent, General Linear Models (GLM), Generalised Additive Models (GAM), Gradient Boosting Machines (GBM), Artificial Neural Networks (ANN), Multivariate Adaptive Regression Splines (MARS) and Random Forest (RF)). All analyses were done using the package "biomod2" (Thuiller et al, 2016). We built pseudo-absences, as our database only register human presences, to run the seven algorithms. Moreover, due to the stochastic nature of the random sampling of pseudo-absences we built several sets of pseudo-absences for each time period: five sets with 300 pseudo absences per set for each time period.

We validated the accuracy of the climatic suitability predictions using cross-validation within each time periods. We split modern human occurrence data for each time period, using 80% of the data to calibrate the models and 20% to validate them. This protocol was repeated five times for each data set (the human occurrences per time interval and each of the 5 datasets of pseudoabsences we sampled - see above). We validated each model using two metrics (True Skill Statistic (TSS), and receiver operating characteristic curve (ROC curve). Finally, we performed one round of modelling with all data within each time period. Based on these settings we built 210 models per time period (7 algorithms x 5 pseudo-absence datasets x 6 cross-validations). Different algorithms produce different results impacting the potential distribution estimates. To avoid arbitrary choice of modelling algorithm and to better represent the information across all models we performed ensemble models by weighted mean (WA), which preselects the best models based on specific criteria, combing the outputs of the 210 models for each time interval. We used the TSS validation methods and a threshold of 0.8 of accuracy to select only the best model outputs and define the weights of each model (i.e. only models with validation scores of TSS>0.8 were used to build the ensemble models). Finally, we projected the ensemble models output to the study area (from 60 degrees longitude east and from 30 degrees latitude north) (Figs x).

To identify what regions had 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. We finally identify the climatic potential across central-north Asia along all time periods (Fig x). We grouped the climatic potential in three categories. Highly suitable

areas were those with a >70% of the time periods holding high climatic suitability: within the higher quartile of suitability. Highly unsuitable areas were those with a 70% of the time periods within the lower three quartiles of suitability. Areas that shifted between periods of high climatic suitability and periods of lower suitability were also identified.

References

Franklin, Janet & Potts, Alastair & C. Fisher, Erich & Cowling, Richard & W. Marean, Curtis. (2015). Paleodistribution modeling in archaeology and paleoanthropology. Quaternary Science Reviews. 110. 1-15. 10.1016/j.quascirev.2014.12.015.

Giampoudakis, K., Marske, K.A., Borregaard, M.K., Ugan, A., Singarayer, J.S., Valdes, P.J., Rahbek, C. and Nogués-Bravo, D. (2017). Niche dynamics of Palaeolithic modern humans during the settlement of the Palaearctic. **Global Ecology and Biogeography** 26: 359–370

Hamilton, M.J. & Buchanan, B. (2010) Archaeological support for the three-stage expansion of modern humans across northeastern Eurasia and into the Americas. PLoS One, 5, e12472

Lorenzen, E. D., Nogues-Bravo, D., Orlando, L., Weinstock, J., Binladen, J., Marske, K. A., Ugan, A., Borregaard, M. K., Gilbert, M. T. P., Nielsen, R., Ho, S. Y. W., Goebel, T., Graf, K. E., Byers, D., Stenderup, J. T., Rasmussen, M., Campos, P. F., Leonard, J. A., Koepfli, K., Froese, D., Zazula, G., Stafford, J. T. W., Aaris-sorensen, K., Batra, P., Haywood, A. M., Singarayer, J. S., Valdes, P. J., Boeskorov, G., Burns, J. A., Davydov, S. P., Haile, J., Jenkins, D. L., Kosintsev, P., Kuznetsova, T., Lai, X., Martin, L. D., Mcdonald, H. G., Mol, D., Meldgaard, M., Munch, K., Stephan, E., Sablin, M., Sommer, R. S., Sipko, T., Scott, E., Suchard, M. A., Tikhonov, A., Willerslev, R., Wayne, R. K., Cooper, A., Hofreiter, M., Sher, A., Shapiro, B., Rahbek, C. and Willerslev, E. (2011). *Species-specific responses of Late Quaternary megafauna to climate and humans.* **Nature** 479: 359-364

Nogués-Bravo D, Rodríguez J, Hortal J, Batra P, Araújo MB (2008) Climate Change, Humans, and the Extinction of the Woolly Mammoth. PLoS Biol 6(4): e79

Pitulko et al. (2016) Early human presence in the Arctic: Evidence from 45,000-year-old mammoth remains. Science 351, 260-263.

Singarayer, J.S., Valdes, P.J., Friedlingstein, P., Nelson, S. & Beerling, D.J. (2011) Late Holocene methane rise caused by orbitally controlled increase in tropical sources. Nature, 470, 82–85.

Thuiller, W., Georges, D., Engler, R., 2016. biomod2: Ensemble platform for species distribution modeling. In. R package version 3.1.3/r539

Supplementary Information 9 - The Formation of the Siberian Linguistic Landscape

Michaël Peyrot (Leiden University) and Guus Kroonen (Universities of Leiden & Copenhagen)

Linguistic diversity, with approximately seven thousand languages spoken worldwide, is one of the most striking expressions of human cultural diversity as it evolved through the millennia. While the growth of human language diversity for the largest part took place in the deep, irretrievable past, humanity's shared linguistic heritage offers unique insights into the movements and contacts of linguistically distinct populations in the Holocene. In addition, the recent ancient DNA revolution opens hitherto unforeseen opportunities for cross-testing existing and formulating new hypotheses on prehistoric human mobility, including the dispersals of prehistoric language communities. Here, a concise linguistic survey is offered of the human genomic history of Siberia as outlined in the main paper.

The linguistic makeup of Siberia

Even before the Russian expansion, language diversity in northern Eurasia was relatively low, with rather few languages and language families as compared to hotbeds of diversity such as the Caucasus region, Central Africa, India or Papua New Guinea (Nichols 1992). In Siberia, extensive hunter-gatherer and hunter-herder subsistence strategies in the tundra and taiga ecozones require comparatively large foraging grounds and can sustain only limited population sizes (cf. Comrie 1981: 57). This stands in marked contrast to the grasslands directly to the south of Siberia, where the linguistic landscape has been shaped by more populous horse-riding groups whose expansions were driven by mobile steppe economies (Damgaard et al. 2018a,b). In both areas, however, several large expansions took place successively, each new wave replacing or marginalizing the preceding. The more recent expansions into and throughout Siberia can be traced back to the steppe.

One important question on Siberia's linguistic past is how the linguistic map developed in prehistoric times. Apart from Indo-European, represented by Russian, the largest language families in present-day Siberia are Uralic, Turkic, Mongolic and Tungusic. These languages have all been expansive at various periods in time, and can be referred to as "Neosiberian" languages (e.g. Vajda 2009: 434; on Uralic, see below). The remaining smaller and often moribund families Yeniseian, Yukaghir, Nivkh and Chukotko-Kamchatkan are thought to be vestiges of an older Siberian linguistic stratum shaped by a non-food producing cultural layer.

Together with the predominantly non-Siberian Eskimo-Aleut family, they are often referred to as the "Paleosiberian" languages (Comrie 1981: 10, 238–239; on Yukaghir, see below). None of either the larger or the smaller families have so far been convincingly shown to be related to each other, and can therefore be characterized as so-called *language isolates*.

Another major question is how the Siberian landscape is connected with the languages of the Americas, particularly Eskimo-Aleut and Na-Dene, and in Europe, with the western branches of Uralic.

To both of these questions, the 34 ancient genomes from across Eurasia reported and analyzed in the main text are of considerable importance. Together they reveal a complex population history of several admixture and replacement events throughout the prehistory of Siberia, with evidence for at least three human population waves into the region. While the earliest, Ancient North Siberian migration as represented by the ~31.6 ky old Yana RHS individual lies beyond the reach of existing methods of linguistic reconstruction, the subsequent Paleo- and Neosiberian population waves are of major importance to understanding the formation of the linguistic map of the circumpolar region.

Turkic, Mongolic and Tungusic

Today, the largest Siberian language families by number of speakers are Turkic, Mongolic and Tungusic. Of these, Mongolic has a comparatively restricted range limited to the south, the only Siberian member of this family being Buryat, spoken around Lake Baikal. In contrast, Turkic and Tungusic did spread over large areas in Siberia in relatively recent times. Turkic Yakut, with the two varieties Dolgan and Sakha, has been influenced by Mongol probably around Lake Baikal before it moved northwards only in the 13th century CE (Janhunen 1996: 162; Fortescue 1998: 195; Okladnikov 1964: 89; Pakendorf et al. 2006; Kałużyński 1962). The Siberian Tungusic languages Evenki and Even likewise have their origin in the south. They show little linguistic diversification, a clear indication of a relatively recent dispersal probably from the Middle Amur Basin starting in the late 1st millennium CE (Janhunen 1996: 169, 172). Since the center of Tungusic linguistic diversity is in Manchuria, where all its branches are attested, the linguistic homeland of this family must have been located there.

Most modern Siberian speakers of Neosiberian languages genetically fall on an East-West cline between Europeans and Early East Asians. Taking Even speakers as representatives, the Neosiberian turnover from the south, which largely replaced Ancient Paleosiberian ancestry, can be associated with the northward spread of Tungusic and probably also Turkic and Mongolic. However, the expansions of Tungusic as well as Turkic and Mongolic are too

recent to be associable with the earliest waves of Neosiberian ancestry, dated later than ~11 kya, but discernible in the Baikal region from at least 6 kya onwards. Therefore, this phase of the Neosiberian population turnover must initially have transmitted other languages or language families into Siberia, including possibly Uralic and Yukaghir.

Uralic

The Uralic family is not strictly speaking a Siberian language family, as most of its languages, including Finnish, Saami, Estonian and Hungarian, are spoken in Europe, but Samoyedic and the easternmost branch of Finno-Ugric, Ob-Ugric, are spoken in central and western Siberia, respectively. Traditionally, the Uralic family is divided into two main branches, Samoyedic and Finno-Ugric. Samoyedic consists of Selkup, Enets, Nenets and Nganasan, as well as the extinct Kamas and Mator. Finno-Ugric can be subdivided into the Saami, Finnic, Permic, Mari, Mordvinic, and Ugric sub-branches.

The time and origin of the Uralic dispersal are debated. In paleo-economic terms, the Uralic family represents a non-food producing layer, which positions it between Turkic and Tungusic and Yeniseian. For Samoyedic a homeland is commonly assumed in south Siberia, and the original area of settlement of Finno-Ugric was probably near the Southern Ural. The primary homeland of Uralic, however, has been placed both east and west of the Urals, as well as in south Siberia: While a West Eurasian origin has been supposed on the basis of, for instance, supposed contacts with Proto-Indo-European (Rédei 1988), an East Eurasian origin is supported by typological similarities of Uralic with Turkic, Mongolic and Tungusic (Janhunen 2001, 2009; Nichols and Rhodes 2018), and perhaps even by folkloristic evidence collected from East Eurasian and North American groups (Napolskikh 2012).

The documentation of a Central Finnish Iron Age individual that genetically resembles modern Saami populations and also contains East Asian ancestry is relevant to the question on the origin and dating of the Uralic expansion.

The Proto-Saami language evolved in southern Finland and Karelia in the Early Iron Age, an area now host to Finnish and the closely related Karelian, but with Saami toponyms showing that the latter two languages are intrusive here (Saarikivi 2004). Saami-speaking populations are thought to have retreated to Lapland during the Middle Iron Age (300–800 AD), where it diverged into the modern Saami dialects. Genetically, the northward retreat of the Saami language correlates with the documented decrease of Saami ancestry in Southern Finland between the Iron Age and the modern period (cf. Lamnidis et al. 2018).

On the way to Lapland, the Saami replaced at least two linguistically obscure groups. This can be inferred from 1) an influx of non-Uralic loanwords into Proto-Saami in the Finnish Lakeland area, and 2) an influx of non-Uralic, non-Germanic words into Saami dialects in Lapland (Aikio 2012). Both of these borrowing events imply contact with non-Saami-speaking groups, e.g. non-Uralic-speaking hunter-gatherers that may have left a genetic and linguistic footprint on modern Saami populations.

The linguistic prehistory of Finland thus does not allow for a straightforward interpretation of the genetic data. The detection of East Asian ancestry in the genetically Saami individual is indicative of a population movement from the east (cf. Lamnidis et al. 2018, Rootsi et al. 2007), one that given the affinities with the ~7.6 ky old individuals from the Devil's Gate Cave may have been a western extension of the Neosiberian turnover. However, it remains unclear whether this gene flow should be associated with the arrival of Uralic speakers, thus providing further support for a Uralic homeland in Eastern Eurasia, or with an earlier immigration of pre-Uralic, so-called "Paleo-Lakelandic" groups.

Yukaghir

When Russia conquered northeastern Siberia in the mid 17th century, the Yukaghirs, numbering probably about 5,000 people, occupied a large area stretching from the Lena River in the west to the Anadyr Basin in the east and the arctic coast in the north to the upper Yana, Indigirka, and Kolyma rivers in the south (Willerslev 2007: 3–5). Today, the Yukaghir language area and number of speakers have been severely reduced, and the two remaining varieties Kolyma and Tundra Yukaghir are together spoken by less than a 100 people.

Compared to the recently expansive Turkic and Tungusic peoples, the Yukaghirs evidently represent a more ancient layer of the Siberian linguistic landscape. Although Janhunen rightly points out that "the linguistic affiliations of the populations absorbed by the Ewenki and Ewen remain largely unknown" (1996: 170–171), it is likely that they include speakers of Yukaghir. Fortescue hypothesizes that many Tungus are in fact "tungusized" Yukaghirs (1998: 195; see also Janhunen 1996: 98).

The presence of Yukaghir speakers in the area may itself be the result of an earlier expansion. According to Kuzmin (2014), their arrival can be associated with the Ymyyakhtakh culture (2200–1300 BC), whose ultimate origins seem to be around Lake Baikal. There is linguistic evidence, too, that suggests a southern origin. While the hypothesis of a deeper-level genealogical relationship with the Uralic language family with a Proto-Uralo-Yukaghir homeland in southern Siberia (Fortescue 1998: 193; Pakendorf 2007: 18; Nikolaeva 2006)

remains controversial, it is widely accepted that the two show certain similarities. Aikio (2014), one of the critics of the hypothetical Uralo-Yukaghir macrofamily, acknowledges that the branches have lexical correspondences, but rather explains these as loanwords between Samoyedic and Yukaghir. The implied linguistic contacts between Samoyedic and Yukaghir would then have taken place in Central or even Southern Central Siberia in view of the established contacts of Proto-Samoyedic with early Turkic (Janhunen 1998: 477). Thus, in both scenarios, Proto-Yukaghir speakers are postulated to have spread to northeast Siberia from the south.

As argued in the main text, Paleosiberian ancestry, which was common throughout Siberia, became largely restricted to the northeast due to a major population turnover of Neosiberians coming from the south. Since the Turkic and Tungusic expansions are demonstrably too recent to have been part of the earliest phases of this Neosiberian wave, Yukaghir could be a better candidate for a language to have been dispersed with these Neosiberian newcomers from the south. This scenario is consistent with the fact that the ~ 760 year old "Young Yana" individual falls along the Neosiberian cline, as the Yana site is located within the historical Yukaghir language area.

Chukotko-Kamchatkan

Another language family of Siberia's northeast is Chukotko-Kamchatkan, consisting of Itelmen on the one hand, and the closely related Koryak and Chukchi on the other. Alyutor and Kerek, now extinct, can be considered dialects of the latter (Comrie 1981: 240). The family was formerly also spoken on Kamchatka, but is now restricted to the extreme northeast.

Chukotko-Kamchatkan is commonly held not to be demonstrably related to any other language family, but several proposals exist. Under the Uralo-Siberian hypothesis (Fortescue 1998; Rask 1818), Chukotko-Kamchatkan is claimed to be related to the Uralic and Yukaghir language families, which can potentially be identified as early manifestations of the Neosiberian linguistic landscape. This suggests that the languages were adopted from or even introduced by early or late Neosiberian groups. Under another hypothesis (Fortescue 2011), Chukotko-Kamchatkan is held to be related to the both linguistically and genetically Paleosiberian Nivkh.

This linguistic issue finds a parallel in a question on the original subsistence strategy of Chukotko-Kamchatkan speakers: while the Chukchi and Koryak practice reindeer-herding, a typically late Neosiberian food production mode, the Itelmen have a more Paleosiberian lifestyle as foragers (Vajda 2009: 434). The question therefore is whether the Itelmen retain

the original Paleosiberian lifestyle, or abandoned a typically Neosiberian pastoral economy in favor of a non-food producing economy.

The present study documents a close genetic affinity of present-day Koryaks, Itelmen and Chukchis with the 9.8 ky old Kolyma individual representing Paleosiberian ancestry. While this genetic outcome is inconsistent with Fortescue's initial Uralo-Siberian hypothesis, which connects Chukotko-Kamchatkan to Uralic and Yukaghir, two presumably Neosiberian groups, it does not preclude a shared population history of prehistoric Chukotko-Kamchatkan speakers and ancestors of Nivkh as under his revised hypothesis. The implied hypothetical Nivkh-Chukotko-Kamchatkan language community would then have been part of the Paleosiberian linguistic landscape formed by populations that fall along the Paleosiberian genetic cline.

Thus, the Chukotko-Kamchatkan Koryaks show close affinity to the 9.8 ky old Kolyma individual. At the same time, the Paleosiberian population represented by the Kolyma individual can be modeled as one of the constituent groups of the Athabaskans of North America. This suggests that the Chukotko-Kamchatkan and Athabaskan languages may be related. However, it is commonly held that the Na-Dene languages, to which Athabaskan belongs, are not related to Chukotko-Kamchatkan, but possibly rather to Yeniseian, a language family strongly associated with Paleosiberian populations (see further below). One possible explanation is that Chukotko-Kamchatkan is actually a Neosiberian language adopted by a Paleosiberian population through language shift (cf. also Fortescue 1998: 210–213). Alternatively, it may have been the language of Paleosiberian groups closely resembling the Kolyma individual, but coming from the south, which fits a scenario in which the Chukotko-Kamchatkan languages share a linguistic past with Nivkh.

Eskimo-Aleut

The Eskimo-Aleut family is widely held to have evolved on the central coast of Alaska, in particular on the Seward Peninsula, a center of dialectal diversity (Berge 2016). The most important division within this family is the split between Proto-Eskimo (2.5 kya) and Proto-Aleut (1 kya), thought to have taken place around 4 kya (Fortescue 1998: 188; Berge 2016).

The Eskimo subclade is in turn divided into Inuit, spoken in Alaska, Canada, and Greenland, and Yupik, spoken in western Alaska and easternmost Siberia. There appears to be consensus among linguists and archaeologists that the Yupik varieties spoken in Siberia are the result of a back-migration from Alaska. The status of the now extinct Sirenik language, until recently spoken in the Chukotka Peninsula, is more uncertain; it has been analyzed as an

offshoot from Yupik, but claims exist that it constitutes an independent third branch within the Eskimo clade that back-migrated to Siberia at an earlier stage or never moved to America (Fortescue 1998: 191; Berge 2016). The early Neoeskimo presence in Northeast Siberia around 5 kya is demonstrated, at any rate, by the sampled Ekven and Uelen individuals from the Old Bering Sea culture, and Eskimo loanwords in North Tungusic may show that Eskimo-Aleut was once more widespread in Eastern Siberia than previously assumed (Vovin 2015).

The more basal Eskimo-Aleut proto-language is estimated to have been spoken some 5–6 kya. Although various hypotheses exist (e.g. Fortescue 1998), it is not demonstrably related to any other language family in North America or Eurasia. Linguists supporting the Dene-Yeniseian hypothesis, under which the Yeniseian languages of Siberia are related to the Na-Dene languages of North America (see also below), do not include Eskimo-Aleut, which is seen as a more recent intrusion into North America between 6 and 4 kya. Genetically, modern Inuit can be modeled as resulting from an admixture of Paleosiberians (as represented by the 9.8 ky old Kolyma individual) and Native Americans that occurred prior to 2.7 kya. The Eskimo-Aleut family thus likely represents the latest known prehistoric wave of migrants from the Old to the New World.

Yeniseian

Today, the only surviving Yeniseian language is the seriously endangered Ket language in north central Siberia. Related languages such as Assan, Arin, Yugh and Pumpokol were previously spoken over a much larger area directly to the south, but they went extinct in the most recent centuries. The wider distribution of the historically recorded Yeniseian languages in this area is further confirmed by hydronyms, river names that can be identified as Yeniseian (see Fig. S9.1). Since these hydronyms extend far beyond historically Yeniseian-speaking areas, into northern Mongolia and east of Lake Baikal, and west perhaps even as far as the Kama Basin west of the Urals (Vajda forth.; Maloletko 2002; Werner 2002: 34–69), the Yeniseian family must have occupied an even more extensive Siberian territory in prehistory. An additional clue that Yeniseian speakers formed a part of the aborigine Siberians exists in Evenki folklore, where these are referred to as "Čaŋyt", a word for which a Yeniseian etymology is at hand (Janhunen 1996: 171; Menges 1965).

Until the Bronze Age, Paleosiberian ancestry was widespread in Siberia between the Ural Mountains in the west and the Bering Sea in the east. Since present-day Kets harbor a considerable Paleosiberian genetic component, it is reasonable to assume that an older stage of

Yeniseian was part of the Paleosiberian linguistic landscape just as Yeniseian speakers were part of the Paleosiberian genetic landscape.

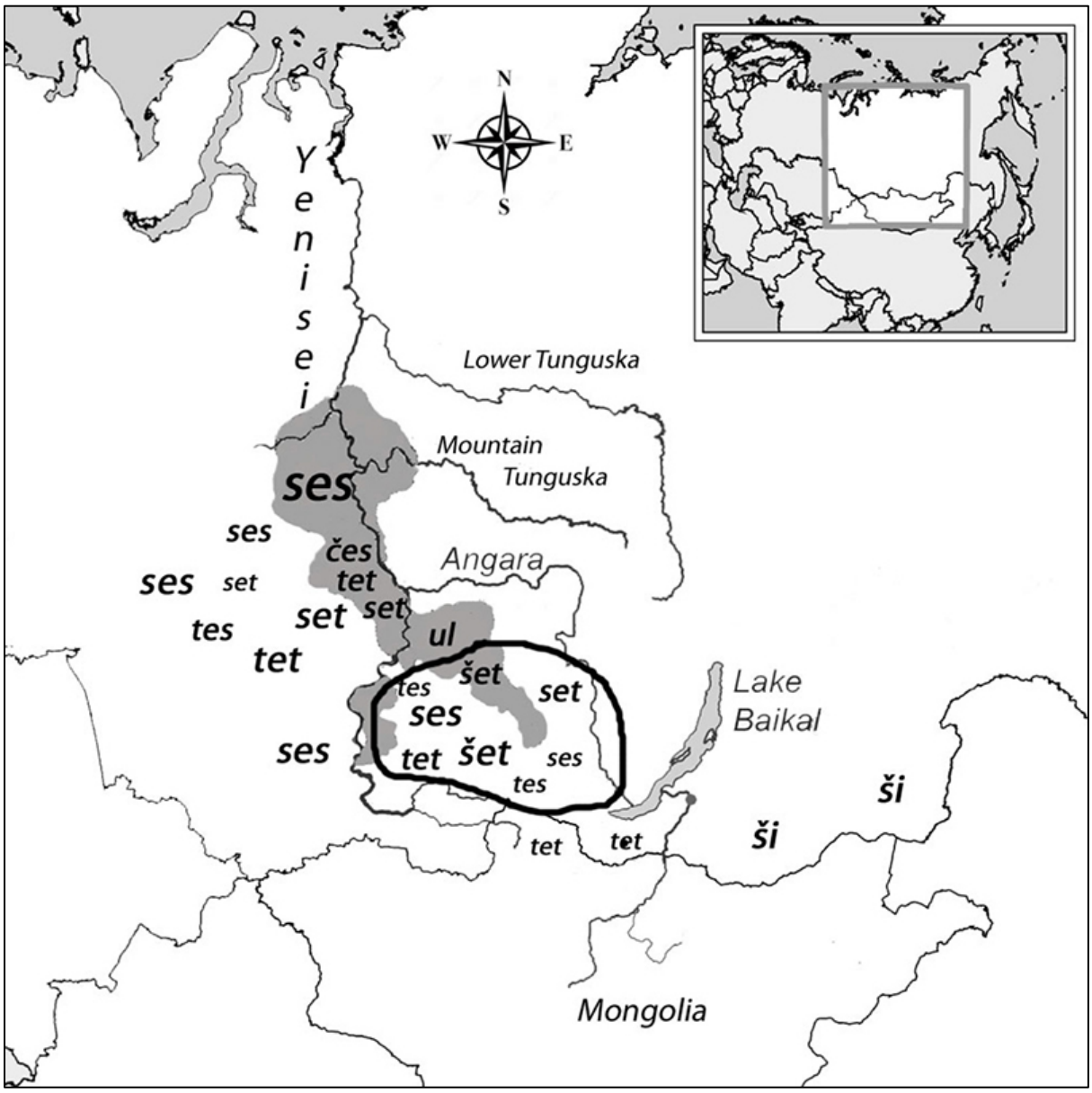


Figure S9.1 Map of a Yeniseian hydronym and probable location of Yeniseian homeland (Vajda 2016).

The Dene-Yeniseian Hypothesis

The genetic affinity between Paleosiberians as represented by the ~9.8 ky old Kolyma individual and present-day Athabascan speakers observed in the sampled and analyzed individuals (see main text; see also Flegontov et al. 2016) has a linguistic parallel in the Dene-Yeniseian hypothesis, which sees the Siberian Yeniseian languages as related to the Na-Dene languages of North America. Although this hypothesis concerns a macrofamily whose existence is not generally accepted, the important progress made by Vajda (2010; 2018; for a

sceptic view, cf. e.g. Campbell 2011) has become accepted by a number of specialists. Within Vajda's framework, the Na-Dene family consists of Athabascan, Tlingit and Eyak, and excludes Haida. If the Dene-Yeniseian languages indeed belong to a single family, then the age of the proto-language that it presupposes approaches the limits of linguistic reconstruction, but it revolutionarily maps the Dene-Yeniseian proto-language onto the spread of Paleosiberian ancestry to North America. However, the closest living representatives of the analyzed Kolyma individual, which closely resembles the population that contributed genetically to the Athabaskan groups through a migration postdating the initial peopling of the Americas, are not Yeniseian-speaking Kets, but Chukotko-Kamchatkan Koryak speakers. This does not necessarily mean that the Dene-Yeniseian Hypothesis is incorrect: Proto-Dene-Yeniseian and early Chukotko-Kamchatkan speaking groups may have had the same Paleosiberian genetic profile, or the Northeast Siberian distribution of the Chukotko-Kamchatkan languages may be the result of a prehistoric linguistic expansion from the south with little or no genetic impact on the local population.

Linguistic macrofamilies

The Ancient North Siberians represented by the ~31.6 ky old Upper-Paleolithic Yana individual by far predates the Neosiberian and Paleosiberian populations discussed so far. While linguistic interpretations of the Neo- and Paleosiberian demographic waves are in part feasible, the Ancient North Siberian timeframe lies well beyond the scope of the linguistic historical-comparative method, which is limited to approximately eight to ten thousand years (Nichols 1992: 2–3; Trask 1996: 377; Thomason 1999: 20; Comrie 2002). Beyond this estimated time limit, linguistic change will typically have obscured the evidence required to demonstrate genealogical relationships between languages and language families.

The different scopes of historical linguistics and ancient genomics are also manifest in the chronological layers of the Neo- and Paleosiberians. It is clear that the languages demand a much more diversified and fine-grained classification than is currently possible on the basis of the genetic distinction between Neosiberian and Paleosiberian groups. At the same time, an obvious question is whether the associated linguistic families ultimately derive from larger proto-families in the deep linguistic past. Since, however, the evidence by which linguistic relationships can be established is progressively lost with time, the status of all of these *macrofamilies* is fundamentally more uncertain than those of well-established primary proto-languages such as Indo-European or Uralic. Thus, although deeper genealogical links among

the Neosiberian and among the Paleosiberian languages are conceivable *a priori*, none has been proved to date.

The Uralo-Yukaghir and Dene-Yeniseian macrofamilies have both been briefly mentioned above. Another important macrofamily that has been proposed is Altaic, consisting of Turkic, Mongolic and Tungusic, and sometimes including Korean and Japonic as well. Typological, lexical and grammatical affinities and correspondences between these languages are widely acknowledged, and taken as proof of an Altaic proto-language by proponents, but attributed to chance and intensive prehistoric language contact by opponents (for a proponent, cf. e.g. Starostin 2016).

More controversial are larger linguistic units. The most relevant here is the Uralo-Siberian family proposed by Fortescue 1998, which comprises Uralic, Yukaghir, Chukotko-Kamchatkan and Eskimo-Aleut. Recently Fortescue made an alternative proposal connecting Chukotko-Kamchatkan rather with Nivkh (2011). Fortescue's Uralo-Siberian hypothesis is accepted by Kortlandt (2004), who connects Nivkh as well. On a still deeper level, Kortlandt sees Uralo-Siberian related to Indo-European and to Altaic, according to him consisting of Turkic, Mongolic, Tungusic, Korean and Japonic (2010a, 2010b). A similar so-called macro-family, "Eurasiatic", has been posited by Greenberg (2000). The latter two, most sizable macrofamilies take together all the known Neosiberian languages, but additionally include Chukotko-Kamchatkan and Eskimo-Aleut, which are usually classified as Paleosiberian, and Indo-European.

Discussion

In this linguistic supplement, an outline has been given of some of the overall trends in the evolution of the Siberian linguistic map against the background of the genetic data presented in the main paper. While the Ancient North Siberian population lies beyond the scope of the linguistic comparative-historical method, the observed genetic population shifts from Paleosiberians to Neosiberians is at least partly mirrored by the linguistic evolution of Siberia.

In general, the Neosiberian turnover appears not just to correspond to a replacement of languages families, but also to a wider, cross-linguistic shift in language typology. This shift is most clearly visible in the contrast between the Paleosiberian Yeniseian languages on the one hand and the Neosiberian-associated Uralic, Tungusic or Turkic on the other, specifically in a transition from prefixing to suffixing morphology and the diffusion of phonologies with vowel harmony (Fortescue 1998: 78–81).

The Yeniseian languages, now represented only by moribund Ket, were once distributed across west and central Siberia, and must have been a major component of the Paleosiberian linguistic landscape. An even wider prehistoric geographic range is presupposed by the hypothesized linguistic relationship of this family with the Na-Dene languages in the New World. However, the languages belonging to this layer were largely replaced by Neosiberian groups in later periods. While comparatively early waves of Neosiberian ancestry may have spread along with the expansions of the Uralic and Yukaghir, the most recent waves can with reasonable certainty be identified with Turkic, Mongolic and Tungusic groups, as well as a typological shift across North Eurasia. Only the Siberian extreme northeast remains as a refuge area of a more ancient Paleosiberian genetic and linguistic situation, as is implied by the close genetic affinity of modern Chukotko-Kamchatkan speakers with the 9.8 ky old Kolyma individual.

The genetic prehistory of Northeast Siberia itself is shown to have been highly dynamic, both during and after the Last Glacial Maximum, as it served as the corridor for multiple population waves from Siberia to North America, and at least one in the opposite direction. Although the initial peopling of the Americas occurred in a period that lies beyond the scope of the existing linguistic methods, the shared Paleosiberian ancestry in modern Kets and Athabaskan speakers finds a potential parallel in the linguistic Dene-Yeniseian hypothesis, and the backflow to Siberia around ~5 kya can be associated with a westward expansion of Inuit languages. All these findings offer important new perspectives on the linguistic links across the North Pacific, and although many questions on the deep linguistic past of this as well as the wider Siberia region remain, it is clear that the field of ancient genomics promises to provide answers that are difficult to obtain with comparative-linguistic methods alone.

References

Aikio, A. 2012. An essay on Saami ethnolinguistic prehistory. In: R. Grünthal and P. Kallio (eds.), *A Linguistic Map of Prehistoric Northern Europe.* Helsinki: Suomalais-Ugrilainen Seura, 63–117.

Aikio, A. 2014. The Uralic-Yukaghir lexical correspondences: genetic inheritance, language contact or chance resemblance? *Finnisch-Ugrische Forschungen* 62:7–76.

Berge, A. 2016. Eskimo-Aleut. In: *Oxford Research Encyclopedia of Linguistics.* Retrieved 15 Feb. 2018, from http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-9.

Campbell, L. 2011. Review of Vajda 2010. *International Journal of American Linguistics* 77:445–457.

Comrie, B. 1981. *The languages of the Soviet Union*. Cambridge University Press: Cambridge Language Surveys.

Comrie, B. 2002. Is there a single time depth cut-off point? In: C. Renfrew, A. McMahon and L. Trask (eds.), *Time depth in historical linguistics. Vol. 1.* Oxford: The McDonald Institute for Archaeological Research, 33–43.

Damgaard P. de B. et al. 2018a. 137 ancient human genomes from across the Eurasian steppes. Nature 557: 369–374. doi.org/10.1038/s41586-018-0094-2

Damgaard P. de B. et al. 2018b. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. Science magazine 360: 1422. doi.org/10.1126/science.aar7711

Flegontov, P. et al. 2016. Genomic study of the Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Scientific Reports* 6:20768. doi.org/10.1038/srep20768

Fortescue, M. 1998. *Language Relations across Bering Strait: Reappraising the Archaeological and Linguistic Evidence.* London: Cassell Academic.

Fortescue, M. 2011. The relationship of Nivkh to Chukotko-Kamchatkan revisited. *Lingua* 121:1359–1376.

Greenberg, J.H. 2000. *Indo-European and its closest relatives: the Eurasiatic language family. Vol. 1: Grammar.* Stanford CA: Stanford University.

Janhunen, J. 1996. *Manchuria, an Ethnic History.* Helsinki: Suomalais-Ugrilainen Seura.

Janhunen, J. 1998. Samoyedic. In: Daniel Abondolo (ed.), *The Uralic languages.* London: Routledge, 457–479.

Janhunen, J. 2001. Indo-Uralic and Ural-Altaic: On the diachronic implications of areal typology. In: Chr. Carpelan e.a. (eds.), *Early contacts between Uralic and Indo-European: Linguistic and archaeological considerations.* Helsinki: Suomalais-Ugrilainen Seura, 207–220.

Janhunen, J. 2009. Proto-Uralic—what, where and when? In: J. Ylikoski (ed.), *The Quasquicentennial of the Finno-Ugrian Society.* Helsinki: Suomalais-Ugrilainen Seura, 57–78.

Kałużyński, S. 1962. *Mongolische Elemente in der jakutischen Sprache.* Warszawa: Państwowe Wydawnictwo Naukowe.

Kortlandt, F.H.H. 2004. Nivkh as a Uralo-Siberian language. In: A. Hyllested e.a. (eds.), *Per Aspera ad Asteriscos.* Innsbruck: Institut für Sprachwissenschaft, 285–289.

Kortlandt, F.H.H. 2010a. Indo-Uralic and Altaic. *Studies in Germanic, Indo-European and Indo-Uralic.* Amsterdam: Rodopi, 415–418.

Kortlandt, F.H.H. 2010b. Indo-Uralic and Altaic revisited. In: L. Johanson and M.I. Robbeets (eds.), *Transeurasian verbal morphology in a comparative perspective: genealogy, contact, chance.* Wiesbaden: Harrassowitz, 153–164.

Kuzmin, Y.V. 2014. Northern and northeastern Asia: archaeology. In: P. Bellwood and I. Ness (eds.), *The Global Prehistory of Human Migration.* Chichester: Wiley-Blackwell, 191–196.

Lamnidis, Th.C. et al. 2018. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *bioRxiv.* dx.doi.org/10.1101/285437

Maloletko, A.M. 2002. *Drevnie narody Sibiri. Ètničeskij sostav po dannym toponimiki. Tom. II. Kety.* 2nd edition. Tomsk: Tomsk University.

Menges, K.H. 1965. Die tungusischen Čaŋyt. *Central Asiatic Journal* 10:81–86.

Napolskikh, V.V. 2012. Earth-Diver Myth (A812) in northern Eurasia and North America: twenty years later. In: Frog, A.-L. Siikala and E. Stepanova (eds.), *Mythic Discourses. Studies in Uralic Traditions.* Helsinki: Finnish Literature Society, 120–140.

Nichols, J. 1992. *Linguistic diversity in space and time.* Chicago: University of Chicago.

Nichols, J. and R.A. Rhodes. 2018. Vectors of language spread at the central steppe periphery: Finno-Ugric as catalyst language. In: G. Kroonen and R. Iversen (eds.), *Digging for words.* Oxford: BAR, 58–68.

Nikolaeva, I.A. 2006. *A historical dictionary of Yukaghir.* Berlin: Mouton de Gruyter.

Okladnikov, A.P. 1964. Ancient population of Siberia and its culture. In: M.G. Levin and L.P. Potapov (eds.), *The Peoples of Siberia.* Chicago: University of Chicago, 13–98.

Pakendorf, B. et al. 2006. Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Human Genetics* 120:334–353. doi.org/10.1007/s00439-006-0213-2

Pakendorf, B. 2007. *Contact in the prehistory of the Sakha (Yakuts): Linguistic and genetic perspectives.* Utrecht: LOT.

Rask, R. 1818. *Undersögelse om det gamle nordiske eller islandske sprogs oprindelse*. Kjöbenhavn: Gyldendal.

Rédei, K. 1988. *Die ältesten indogermanischen Lehnwörter der uralischen Sprachen.* D. Sinor (ed.), *The Uralic languages*. Leiden: Brill, Wien: 638–664.

Rootsi, S. et al. 2007. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *European Journal of Human Genetics* 15:204–211.

Saarikivi, J. 2004. Über das saamische Substratnamengut in Nordrussland und Finnland. *Finnisch-Ugrische Forschungen* 58:187–214.

Starostin, G. 2016. Altaic languages. In: *Oxford Research Encyclopedia of Linguistics.* http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-35

Thomason, S. 1999. Speakers' choices in language change. *Studies in the Linguistic Sciences* 29(2):19–43.

Trask, L.R. 1996. *Historical linguistics.* London: Arnold.

Vajda, E.J. 2009. The languages of Siberia. *Language and Linguistics Compass* 3:424–440.

Vajda, E.J. 2010. A Siberian link with Na-Dene languages. *Anthropological Papers of the University of Alaska* 5 New Series:33–99.

Vajda, E.J. 2018. Dene-Yeniseian. Progress and unanswered questions. *Diachronica* 35:277–295. doi.org/10.1075/dia.18001.vaj

Vajda, E.J. forth. Yeniseian substrates and typological accommodation in central Siberia.

Vovin, A. 2015. Eskimo loanwords in northern Tungusic. *Iran and the Caucasus* 19:87–95.

Werner, H. 2002. *Vergleichendes Wörterbuch der Jenissej-Sprachen. Band 3: Onomastik.* Wiesbaden: Harrassowitz.

Willerslev, R. 2007. *Soul Hunters: Hunting, Animism and Personhood among the Siberian Yukaghirs.* Berkeley, CA: University of California.