

Supplementary information

Paleo-Eskimo genetic legacy across North America

Pavel Flegontov, N. Ezgi Altınışık, Piya Changmai, Nadin Rohland, Swapan Mallick, Deborah A. Bolnick, Francesca Candilio, Olga Flegontova, Choongwon Jeong, Thomas K. Harper, Denise Keating, Douglas J. Kennett, Alexander M. Kim, Thisseas C. Lamnidis, Iñigo Olalde, Jennifer Raff, Robert A. Sattler, Pontus Skoglund, Edward J. Vajda, Sergey Vasilyev, Elizaveta Veselovskaya, M. Geoffrey Hayes, Dennis H. O'Rourke, Ron Pinhasi, Johannes Krause, David Reich, Stephan Schiffels

CONTENTS

Supplementary tables	2
SI 1. Description of samples and archaeological sites	4
SI 2. Radiocarbon dating	8
SI 3. Ancient DNA isolation and sequencing	11
SI 4. Principal component analysis	14
SI 5. Exhaustive analysis of ancestry streams in small population sets	15
SI 6. Haplotype sharing statistics	21
SI 7. Admixture inference with <i>GLOBETROTTER</i>	30
SI 8. Additional results on Aleutian population history	34
SI 9. Demographic modelling with <i>qpGraph</i> and <i>rarecoal</i>	37
SI 10. Overview of the Dene-Yeniseian linguistic hypothesis	54

Supplementary tables

Supplementary Table 1. Reservoir-adjusted radiocarbon calibrations and stable isotope data for ancient skeletal samples analyzed in this study.

Supplementary Table 2. Information on newly genotyped present-day individuals.

Supplementary Table 3. List of populations and meta-populations used in this study: Paleo-Eskimo (P-E), Eskimo-Aleut speakers and ancient Neo-Eskimos (E-A), Chukotko-Kamchatkan speakers (C-K), Na-Dene speakers (mostly Athabaskans, ATH), northern First Americans (NAM), southern First Americans (SAM), West Siberians (WSIB), East Siberians (ESIB), Southeast Asians (SEA), Europeans (EUR), Africans (AFR). Shotgun sequencing data were generated in this study for one ancient Aleut individual or taken from three published sources: the Simons Genome Diversity Project (Mallick et al. 2016), Raghavan et al. (2015), and 1000 Genomes project (1000 Genomes Project Consortium 2015). The latter source was included into the African, European, and Southeast Asian meta-populations only. Two SNP array datasets were used: based on the HumanOrigins array and on Illumina arrays. HumanOrigins data were taken from Mathieson et al. (2015) and Jeong et al. (submitted) or generated in this study for Alaskan Iñupiat and West Siberians (Enets, Kets, Nganasans, and Selkups). Illumina data were taken from the following sources: Li et al. 2008, Behar et al. 2010, Rasmussen et al. 2010, Fedorova et al. 2013, Raghavan et al. 2014a, 2014b, 2015, Verdu et al. 2014, Kushniarevich et al. 2015. Genome-wide targeted enrichment data were generated in this study for 17 ancient individuals (11 Aleuts, 3 Northern Athabaskans, 2 ancient Neo-Eskimos of the Old Bering Sea culture, and one Paleo-Eskimo-related individual of the Ust'-Belaya culture), and merged with both SNP array datasets. Before the merging step, 5 low-coverage ancient Aleut samples were removed, and one ancient Athabaskan sample was removed as a first-degree relative of another sample.

Notes:

* The Dakelh population was referred to as Athabaskan in Rasmussen et al. (2010) and as 'Northern Athabaskan 1' or simply Athabaskan in Raghavan et al. 2015.

** The Caucasian meta-population (CAU) was considered as a part of the European meta-population in the sequencing dataset only, and in the HumanOrigins dataset it was not used for most analyses except for ADMIXTURE.

*** The Middle Eastern (ME), South Asian (SAS), and Australo-Melanesian (OCE) meta-populations were included into the HumanOrigins dataset, but were not used for most analyses except for ADMIXTURE.

Supplementary Table 4. Details of datasets used in this study.

Notes:

* transitions were removed in this dataset version

** rare variants occurring in the dataset from 2 to 5 times

*** rare variants occurring in the dataset from 1 to 4 times

^ analyzed as 7 meta-populations and 2 ancient genomes mapped on the tree

Supplementary Table 5. f_4 -statistics calculated for the Ust'-Belaya ancient individual on the HumanOrigins and Illumina datasets, full versions or datasets without transition polymorphisms. The following statistics were calculated: $f_4(\text{Ust'-Belaya, Yoruba; Saqqaq, any other population})$; $f_4(\text{Ket, Yoruba; Ust'-Belaya, any other population})$; $f_4(\text{the Mal'ta individual MA-1, Yoruba; Ust'-Belaya, any East or West Siberian population})$. Statistics significantly different from 0 ($|Z\text{-score}| > 3$) are highlighted in green, and Z-scores between 2 and 3 or -2 and -3 are highlighted in yellow. The number of sites used for calculating the statistic is also shown.

References (for Supplementary tables)

1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

Behar, D. M. *et al.* The genome-wide structure of the Jewish people. *Nature* **466**, 238–242 (2010).
 Fedorova, S. A. *et al.* Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol. Biol.* **13**, 127 (2013).
 Jeong, C. *et al.* Characterizing the genetic history of admixture across inner Eurasia. submitted manuscript.
 Kushniarevich, A. *et al.* Genetic heritage of the Balto-Slavic speaking populations: A synthesis of autosomal, mitochondrial and Y-chromosomal data. *PLoS ONE* **10**, e0135820 (2015).
 Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
 Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
 Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
 Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345**, 1255832 (2014a).
 Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014b).
 Raghavan, M. *et al.* Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, 1–20 (2015).
 Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
 Verdu, P. *et al.* Patterns of admixture and population structure in native populations of northwest North America. *PLoS Genet.* **10**, e1004530 (2014).

Supplementary Information section 1 Description of samples and archaeological sites

1.1 Ancient Eastern Aleutian islanders

The skeletal samples from the eastern Aleutians were selected from curated collections at the Smithsonian Institution by M. Geoffrey Hayes, who was gloved, sleeved, and masked at all times to prevent self-contamination of the samples. All samples were small, fragmentary ribs free of pathological lesions and were immediately placed in sterile ziplock bags (Hayes 2002) for transport to the lab for analysis.

The remains were excavated or collected by Aleš Hrdlička in the late 1930s. The geographic locations of the material are burial caves on Shiprock Island (northeast of Umnak Island), and Kagamil, one of the sacred Islands of the Four Mountains, immediately west of Umnak (Table 1). The third site providing samples for molecular analysis is Chaluka, a deep midden site on Umnak adjacent to the contemporary village of Nikolski.

For the present study, the samples available for analysis included six individuals from Kagamil, with three osteologically determined to be female, one as probably female (later identified genetically as a male), and two as male. As reported by Brenner Coltrain *et al.* (2006), these six Kagamil samples exhibit a calibrated age range of 479 – 596 years before present (calBP). The single individual from Shiprock was identified as a male with an age of 749 calBP. Finally, four individuals from the Chaluka site at Nikolski (three males and one female according to genetic data) exhibited an age range of 702 – 2,305 calBP. In this study, the dates were recalibrated (Table 1, Supplementary Table 1) using a more appropriate marine reservoir correction as described in Supplementary Information section 2.

Based on cranial metrics, Hrdlička (1945) postulated that the mummified remains from the burial caves on Kagamil and Shiprock represented immediate ancestors of modern Aleut people who had replaced an earlier population of ‘Pre-’ or ‘Paleo-Aleuts’ about a millennium ago. He viewed the remains at Chaluka as representatives of this earlier occupation of the Islands.

Although Hrdlička (1945) considered the ‘Paleo-Aleuts’ to be older than ‘Neo-Aleuts’, with only the latter ancestral to modern Aleut people following a replacement event around 1,000 years ago, direct dating of the ancient remains (Brenner Coltrain, *et al.* 2006) clearly established that while all individuals recovered from Chaluka were ‘Paleo-Aleuts’ by Hrdlička’s cranial metric criteria, they coexisted with ‘Neo-Aleuts’ for several hundred years following the appearance of

the latter at about 1,000 calBP. Thus, the strict replacement model of Hrdlička's was untenable and the prehistory of peoples of the Aleutian chain, at least in the east, proved to be more complex than previously thought (Smith et al. 2009).

References (for this section)

- Brenner Coltrain, J. B., Hayes, M.G. & O'Rourke D.H. Hrdlička's Aleutian population-replacement hypothesis. A radiometric evaluation. *Curr. Anthropol.* **47**, 537–548 (2006).
- Hayes, M. G. Paleogenetic assessments of human migration and population replacement in North American Arctic prehistory. Doctoral Dissertation, University of Utah (2002).
- Hrdlička, A. *The Aleutian and Commander Islands and their inhabitants*. Philadelphia: Wistar Institute of Anatomy and Biology (1945).
- Smith, S. et al. (2009) Inferring population continuity versus replacement with aDNA: A cautionary tale from the Aleutian Islands. *Hum. Biol.* **81**, 19–38.

1.2 Ancient Northern Athabaskans

The ancient Athabaskan population in this study is derived from three individuals found intermingled in a non-burial context in the riparian zone of the upper Kuskokwim river. Tochak is the Athabaskan place-name for the area around the modern mixed ethnic community of McGrath, southwest Interior Alaska. Known as the Tochak McGrath Discovery, the three individuals were buried in overbank sediments that also feature unassociated buried organic bands with terrestrial and aquatic fauna, hearth matrix, flaked stone and bone artifacts. The human remains could not be linked stratigraphically to the surrounding cultural occupation features. The three individuals representing the Tochak family are successive generations of consanguinial relatives: 30-40 year-old male (MT-1), 19-20 year-old male (MT-2), and 2-3 year old female (MT-3) (Table 1). The genetic analysis indicates a father-son relationship for MT-1 and MT-2 and suggests grandfather-granddaughter and uncle-niece relationships. For this reason, only individuals MT-1 and MT-3 were selected for downstream genetic analyses. Nearly complete skeletal representation and articulation pattern of all three individuals in massive sand overbank deposits suggest that these individuals drowned together.

Soon after the time of discovery, a tripartite agreement was reached for scientific analysis between the McGrath Native Village Council (the federal recognized Alaska Native tribe), MTNT Ltd. (consortium of Alaska Native Claims Settlement Act village corporations) and Tanana Chiefs Conference (the regional non-profit consortium of 37 federally recognized Athabascan Tribes and Alaskan Native associations in the Yukon and Kuskokwim river basins in Interior Alaska). R.A. Sattler has facilitated community-based research, collaboration with academic institutions, Tribal consultation, public outreach and further data recovery at the Tochak discovery locale (Sattler et al. 2013).

References (for this section)

- Halffman, C. M., Sattler, R. & Clark, J. L. Bone collagen stable isotope analysis of three late Holocene humans from Interior Alaska. *Am. J. Phys. Anthropol.* **156** (S60), 157 (2015).
- Sattler, R. A. et al. Tochak McGrath discovery: Precontact human remains in the Upper Kuskokwim River region of interior Alaska. *Alaska J. Anthropol.* **11**, 185–186 (2013).

1.3 Early Neo-Eskimos (Old Bering Sea culture)

Two individuals buried at Uelen, Chukotka, associated with the Old Bering Sea culture were sequenced in this study (Table 1). The Uelen and related Ekven cemeteries of the Old Bering Sea (Drevneberingomorskaya) culture are located on the Chukotka Peninsula. The Uelen burial ground is separated by only 170 m from the present-day settlement Uelen on the coast of the Chukchi Sea. The site was discovered in 1955 by D. A. Sergeyev, and its further excavation was

carried out by the Institute of Ethnography of the Academy of Sciences of the USSR (details are reported in Levin & Sergeyev 1964, Dikov 1967, Arutyunov & Sergeyev 1969).

Both sites represent burial grounds of the Old Bering Sea culture of sea mammal hunters and fishers of the Arctic zone of Siberia and North America. This culture is related to others in the Bering Straits region that partially overlap in time (1700-1000 calBP): Okvik, Punuk, and Birnik, collectively (with the later related Thule tradition) termed the Northern Maritime tradition (Collins 1964). The Old Bering Sea stage is the earliest in development of this cultural tradition and is dated to between ~2300-1300 calBP (Arutyunov and Sergeyev 1975, Gerlach and Mason 1992) with evidence for continuity with the later Okvik, Punuk, and Birnik cultures (Arutyunov and Sergeyev 1969, Gerlach and Mason 1992).

Mortuary behavior at the Ekven burial ground (189 burials) is more variable in comparison with other cemeteries of this culture. The buried were laid not only in the extended position, but also in the curled position, and there is a considerable number of pair and group burials. Human remains from the Uelen and Ekven burial grounds provide an important source of data for the bioanthropology of Old Bering Sea culture individuals (Levin & Sergeyev 1964, Debets 1975). Odontological materials from the Ekven burial ground, and to a lesser extent from Uelen, are very similar to those of present-day Eskimos (especially from Alaska) (Zubov 1969).

Two samples analyzed in this study (Table 1, Supplementary Table 1) came from Uelen burials no. 13 (inventory no. 172) and 22 (inventory no. 163) (Arutyunov & Sergeyev 1969). Burial 13 was excavated in 1958. An adult female's skeleton lay on the side. Fragments of an infant skull and ribs were found near adult's skull. Grave goods were associated: two zoomorphic bone figurines beside the right hip and the pelvis; a bone scraper and a bone awl; a hoe and traces of pottery beside the right leg (Arutyunov & Sergeyev 1969). Burial 22 contains a male skeleton and was excavated in 1962. Both burials were radiocarbon dated and dates were calibrated to account for the marine reservoir effect in this study (Supplementary Information section 2, Supplementary Table 1): 1970-1590 calBP for burial 13 and 1180-830 calBP for burial 22.

References (for this section)

- Arutyunov, S. A. & Sergeyev, D. A. *Drevniye kul'tury aziatskih eskimosov: uelenskiy mogil'nik* [Ancient cultures of Asian Eskimos: the Uelen burial ground]. Moscow: Nauka (1969).
- Arutyunov, S. A. & Sergeyev, D. A. *Problemy etnicheskoy istorii Beringomorya: ekvenskiy mogil'nik* [Problems of the ethnic history of the Bering Sea region: the Ekven burial ground]. Moscow: Nauka (1975).
- Collins, H. B. The Arctic and Subarctic. *Prehistoric Man in the New World*, ed. Jennings, J. D. & Norbeck, E. Chicago: University of Chicago Press. 85-114 (1964).
- Debets, G. F. Paleoantropologicheskiye materialy iz drevneberingomorskih mogil'nikov Uelen i Ekven [Paleoanthropological materials from the Old Bering Sea burial grounds Uelen and Ekven]. *Problemy etnicheskoy istorii Beringomorya: ekvenskiy mogil'nik* [Problems of the ethnic history of the Bering Sea region: the Ekven burial ground], ed. Arutyunov, S. A. & Sergeyev, D. A. Moscow: Nauka. 198 (1975).
- Dikov, N. N. Uelenskiy mogil'nik po dannym raskopok 1956, 1958 i 1963 gg [The Uelen burial ground according to excavations in 1956, 1958 and 1963]. *Istoriya i kul'tura narodov Severa Dal'nego Vostoka* [History and culture of the peoples of the northern Far East], *Trudy SVKNII SO AN SSSR*, vol. 17. Moscow: Nauka (1967).
- Gerlach, C. & Mason, O. K. Calibrated radiocarbon dates and cultural interaction in the Western Arctic. *Arctic Anthropol.*, **29**, 54-81 (1992).
- Levin, M. G. & Sergeyev, D. A. Drevniye mogil'niki Chukotki i nekotorye aspekty eskimosskoy problemy [Ancient burial grounds in Chukotka and some aspects of the Eskimo problem]. *Doklady na VII Mezhdunarodnom kongresse antropologicheskikh i etnograficheskikh nauk*. Moscow (1964).
- Zubov, A. A. Antropologicheskii analiz cherepnykh seriy iz Ekvenskogo i Uelenskogo mogil'nikov [Anthropological analysis of cranial series from the Ekven and Uelen burial grounds]. *Drevniye kul'tury aziatskih eskimosov: uelenskiy mogil'nik* [Ancient cultures of Asian Eskimos: the Uelen burial ground], ed. Arutyunov, S. A. & Sergeyev, D. A. Moscow: Nauka (1969).

1.4 A Paleo-Eskimo-related culture at the Ust'-Belaya site

A single individual from the Ust'-Belaya site in interior Chukotka was investigated in this study (Table 1, Supplementary Table 1). The archaeological site is located at the confluence of the Belaya and Anadyr rivers on a hill towering over the present-day village Ust'-Belaya. The excavation was carried out by N. N. Dikov in 1958-1959. The Ust'-Belaya cemetery consisted of 15 low and flat stone barrows built at a flat top of the hill (Dikov 1961, 1979).

The details below are summarized from a previous report (Dikov 1961). Preservation of human remains is variable. In some barrows of the Ust'-Belaya cemetery human skeletons are not preserved, some corpses are fully or partially burned, and only in Barrow 8 four skeletons of incomplete preservation were discovered. They were positioned at different levels, one under another.

The first skeleton had the best preservation of all and was the only one sampled and dated in this study (Supplementary Information section 2, Supplementary Table 1). This male was buried deeper than the others and lay directly on permafrost, and thus possibly represents the most ancient burial within the barrow. The corpse lay on the back, with the head oriented to the west, with hands on the waist and with legs crossed. No stone tools were found in this burial. On the right side of the mandible a small bronze knife carefully wrapped in birch bark was found. The cranium was well preserved. It was dated with the radiocarbon method to 4410-4100 calBP. Dikov had tentatively dated the Ust'-Belaya burial ground to the end of the 2nd - beginning of the 1st millennium BCE (Dikov 1979), and sites of the Ust'-Belaya culture have been dated with the radiocarbon method to 4500-3000 calBP (Kir'iak 1993). Thus, the investigated skeleton is one of the oldest samples associated with this culture.

The second skeleton (possibly a woman) had the head oriented to the southwest. It is notable that under a crushed skull remains of a hair decoration were found (a set of small round beads located closely one to another). To the right of the skull, fragments of a clay pot that contained food were found. The top part of the skeleton and the skull were covered with ochre. Grave goods were associated: obsidian and chalcedony knives, scrapers and finely flaked arrow points very similar to those of the Yakut and Baikal late Neolithic and early Bronze Age periods.

Two other skeletons had skulls oriented to the northwest and to the southeast. Bone preservation was poor, and skulls were crushed. Beside them many stone knives, scrapers, arrow points were found, and under one of the skulls a small bronze knife was revealed.

Funeral ceremonies at the Ust'-Belaya burial ground reflect spiritual and material culture of interior Chukotkan tribes in the late Neolithic and Bronze Age. Customs of partial cremation, a ceremony of covering the dead with red ochre, decoration of a hairdress by beads, finds of bear canines, the form of stone points, a tip of a toggling harpoon found at the site, and bronze artifacts suggest extensive connections of the Ust'-Belaya population with inner Siberia (Dikov 1974, 1979).

References (for this section)

- Dikov N. N. O raskopkakh Ust'-Bel'skogo mogil'nika, po dannym 1958 goda. [About the excavations of the Ust'-Belaya burial ground in 1958]. *Zapiski Chukotskogo kraevedcheskogo muzeya* **2** (1961).
- Dikov N. N. Ust'-Bel'skaya kul'tura tsentral'noy Chukotki [The Ust'-Belaya culture of the central Chukotka]. *Ocherki istorii Chukotki: S drevneyshih vremen do nashih dney* [Sketches of the history of Chukotka: From ancient times to the present day]. Novosibirsk: Nauka 1974.
- Dikov, N. N. *Drevniye kul'tury Severo-Vostochnoy Azii: Aziya na styke s Amerikoy v drevnosti*. Moscow: Nauka (1979). Translated by Bland, R. L. as Early cultures of Northeastern Asia. Anchorage: US Department of the Interior, National Park Service, Shared Beringian Heritage Program (2004).
- Kir'iak, M. A. Ymyakhtakhskii kompleks so stoianki Bol'shoi El'gakhchan I (Bassein Omolona) [The Ymyakhtakh complex from the Bol'shoi El'gakhchan I site (Omolon Basin)]. *Kraevedcheskie Zapiski* **19**, 3-15 (1993).

1.5 Alaskan Inupiat

Inupiat samples in this study were collected, along with genealogical records and participant surveys, by M. Geoffrey Hayes and Jennifer A. Raff from the communities of Atkasuk, Anaktuvuk Pass, Utqiagvik (formerly known as Barrow), Kaktovik, Nuiqsut, Point Hope, Point Lay, and Wainwright between 2008-2010 as described in Raff et al. (2015). This project was begun at the suggestion of an Elder in Utqiagvik, to complement ancient DNA work on burial populations in the region, and was approved by Northwestern University's Institutional Review Board, after consultation with the Ukpeagvik Inupiat Corporation, the Native Village of Barrow, and Senior Advisory Council of Barrow (Elders). Of the 181 samples collected, 35 individuals who consented to have their DNA used for ancestry research were selected for inclusion in this present study to represent a diversity of mitochondrial haplogroups and geographic origins (reported in Raff et al. 2015) and to represent both sexes in as close to equal proportions as possible. During the outlier removal procedure described in the Methods section, 20 individuals with minimal admixture from outside populations were selected for downstream analyses. Study results are returned and discussed with communities prior to publication.

References (for this section)

Raff, J. A. *et al.* Mitochondrial diversity of Inupiat people from the Alaskan North Slope provides evidence for the origins of the Paleo- and Neo-Eskimo peoples. *Am. J. Phys. Anthropol.* **157**, 603–614 (2015).

Supplementary Information section 2

Radiocarbon dating

We report four new direct AMS ^{14}C bone dates from the Penn State Accelerator Mass Spectrometer laboratory (PSUAMS) and recalibrate 12 previously published radiocarbon dates from two other AMS radiocarbon laboratories (Arizona [AA] = 11; Beta Analytic [Beta] = 1, see Supplementary Table 1 and Fig. S2.1). Bone preparation and quality control methods for the AA and Beta samples are described elsewhere (Brenner Coltrain et al. 2006, Halffman et al. 2015).

2.1 Old Bering Sea and Ust'-Belaya culture samples

At PSUAMS, bone collagen for ^{14}C and stable isotope analyses and was extracted and purified using a modified Longin method with ultrafiltration (Kennett et al. 2017). Bones were initially cleaned of adhering sediment and the exposed surfaces were removed with an X-acto blade. Samples (200–400 mg) were demineralized for 24–36 h in 0.5N HCl at 5 °C followed by a brief (<1 h) alkali bath in 0.1N NaOH at room temperature to remove humates. The residue was rinsed to neutrality in multiple changes of Nanopure H_2O , and then gelatinized for 12 h at 60 °C in 0.01N HCl. Resulting gelatin was lyophilized and weighed to determine percent yield as a first evaluation of the degree of bone collagen preservation. Rehydrated gelatin solution was pipetted into pre-cleaned Centriprep (McClure et al. 2010) ultrafilters (retaining 430 kDa molecular weight gelatin) and centrifuged 3 times for 20 min, diluted with Nanopure H_2O and centrifuged 3 more times for 20 min to desalt the solution. Carbon and nitrogen concentrations and stable isotope ratios were measured at the Yale Analytical and Stable Isotope Center with a Costech elemental analyzer (ECS 4010) and Thermo DeltaPlus analyzer. Sample quality was evaluated by % crude gelatin yield, %C, %N and C/N ratios before AMS ^{14}C dating. C/N ratios for all four samples fell between 2.9 and 3.6, indicating good collagen preservation (Van Klinken 1999).

Collagen samples (~2.1 mg) were combusted for 3 h at 900 °C in vacuum-sealed quartz tubes with CuO and Ag wires. Sample CO_2 was reduced to graphite at 550 °C using H_2 and a Fe catalyst, with reaction water drawn off with $\text{Mg}(\text{ClO}_4)_2$ (Santos et al. 2004). Graphite samples were pressed into targets in Al cathodes and loaded on the target wheel for AMS analysis. The ^{14}C ages were corrected for mass-dependent fractionation with measured $\delta^{13}\text{C}$ values (Stuiver and Polach 1977) and compared with samples of Pleistocene whale bone (backgrounds, 48,000 ^{14}C BP), late Holocene bison bone (~1,850 ^{14}C BP), late AD 1800s cow bone and OX-2 oxalic acid standards for calibration.

2.2 Northern Athabaskan (Tochak McGrath) samples

Collagen removed from the femur of MT-1 (the eldest individual) yielded a radiocarbon age of 1170 ± 30 BP (AMS lab code Beta-337194). This age estimate provides an older limiting age on the time of death of the Tochak family. Isotopic analysis has determined relatively high carbon and nitrogen values on all three individuals that suggest a strong marine component to their diet (i.e., anadromous salmon) (Halffman et al. 2015). The isotopic values suggest that the radiocarbon age on human collagen may over-estimate the actual time of death, and the date was calibrated as described below. Given the direct age on MT-1 as a maximum limiting age, charcoal dates from matrix of two spatially separate hearths at the Tochak site provide a younger limiting age of around 350 years before present: 320 ± 30 BP (465-300 calBP; AMS lab code Beta-333837) and 380 ± 30 BP (505-320 calBP; AMS lab code Beta-343499).

2.3 Calibration of radiocarbon dates

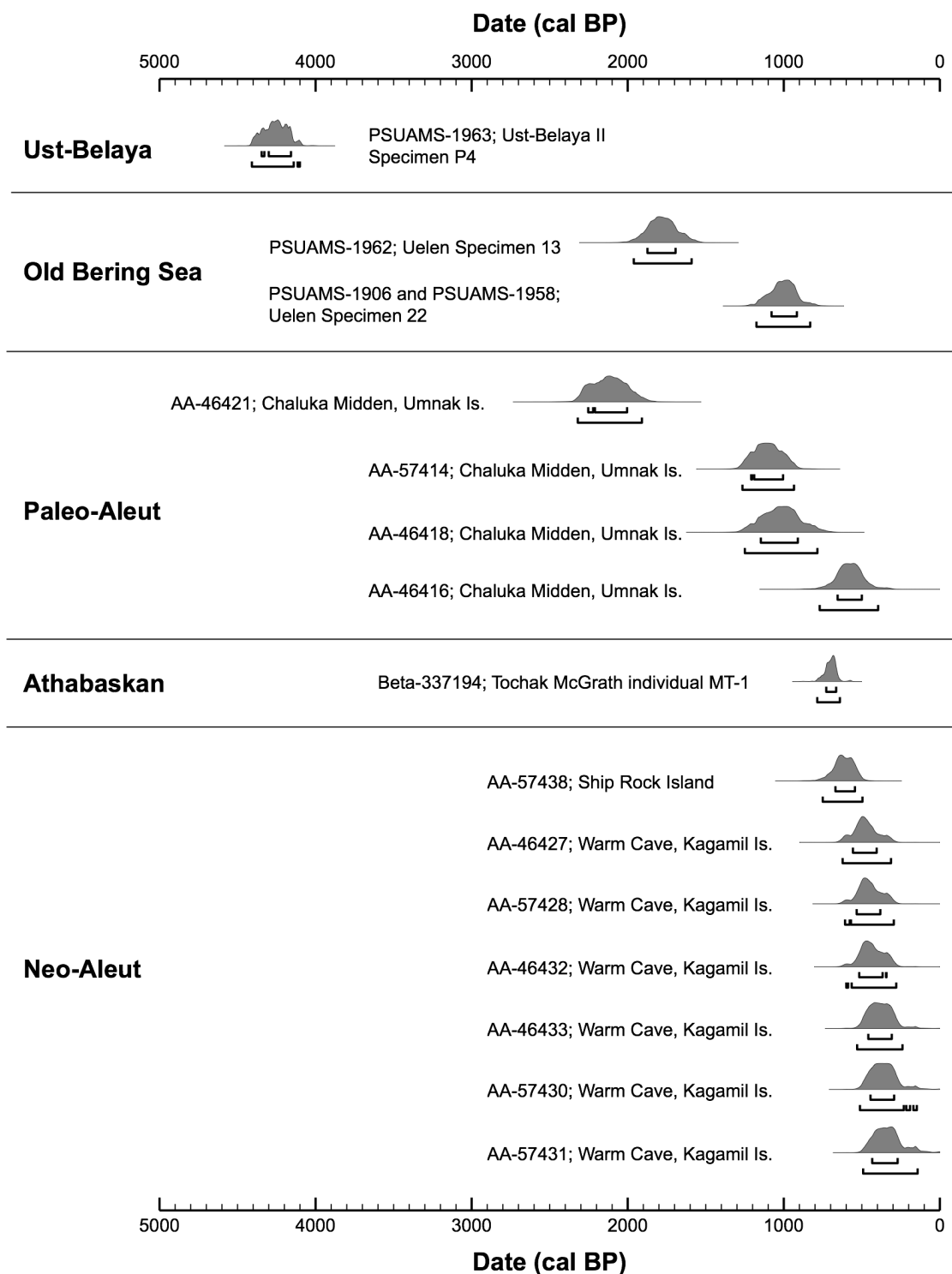
All ^{14}C ages were calibrated with OxCal version 4.2.3 (Bronk Ramsey 2013) using a mixed northern hemisphere terrestrial calibration curve (IntCal13) and the marine curve (Marine13; Reimer et al. 2013). Marine contribution was estimated using stable carbon and nitrogen isotopes, and we used a ΔR of 455 ± 81 (Misarti and Maschner 2015), which is based on an

average for this region (Reimer and Reimer 2001). A 50% marine contribution was applied to inland samples, where a diet with a moderately marine component is nonetheless indicated (with $\delta^{15}\text{N}$ values ranging from 13.7 to 15.2‰). For coastal samples, where $\delta^{15}\text{N}$ values ranged from 18.3 to 21.3‰, a 90% marine contribution was applied. The results are presented in Supplementary Table 1 and Fig. S2.1.

References (for this section)

- Brenner Coltrain, J., Hayes, M. G. & O'Rourke, D. H. Hrdlička's Aleutian population-replacement hypothesis. A radiometric evaluation. *Curr. Anthropol.* **47**, 537–548 (2006).
- Bronk Ramsey, C. OxCal 4.23 Online Manual https://c14.arch.ox.ac.uk/oxcalhelp/hlp_contents.html (2013).
- Kennett, D. J. *et al.* Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017).
- Halffman, C. M., Sattler, R. & Clark, J. L. Bone collagen stable isotope analysis of three late Holocene humans from Interior Alaska. *Am. J. Phys. Anthropol.* **156** (S60), 157 (2015).
- Misarti, N. & Maschner, H. D. G. The Paleo-Aleut to Neo-Aleut transition revisited. *J. Anthropol. Archaeol.* **37**, 67–84 (2015).
- Reimer, P. J. *et al.* Intcal13 and Marine13 radiocarbon age calibration curves 0–50,000 years Cal Bp. *Radiocarbon* **55**, 1869–1887 (2013).
- Reimer, P. J. & Reimer, R. W. A marine reservoir correction database and on-line interface. *Radiocarbon* **43**, 461–463 (2001).
- Santos, G. M. *et al.* Magnesium perchlorate as an alternative water trap in AMS graphite sample preparation: A report on sample preparation at KCCAMS at the University of California, Irvine. *Radiocarbon* **46**, 165–173 (2004).
- Stuiver, M. & Polach, H. A. Reporting of C-14 Data-Discussion. *Radiocarbon* **19**, 355–363 (1977).
- Van Klinken, G. J. Bone collagen quality indicators for palaeodietary and radiocarbon measurements. *J. Archaeol. Sci.* **26**, 687–695 (1999).

Figure S2.1. Plot of probability distributions for new AMS ^{14}C dates (PSUAMS results) and previously published regional radiocarbon data (Brenner Coltrain et al. 2006, Halffman et al. 2015) with marine reservoir correction ($\Delta R = 455 \pm 81$; Misarti and Maschner 2015).



Supplementary Information section 3 Ancient DNA isolation and sequencing

3.1 Ancient DNA isolation

DNA extraction, library preparation, target capture enrichment and Illumina sequencing was done at the Harvard Medical School in Boston (USA). Powder for DNA extractions was produced in Boston (14 Aleutian and Tochak McGrath Athabaskan samples) and at University College Dublin (3 Old Bering Sea and Ust'-Belaya samples); for details see Table S3.1 below. All work performed in Dublin and Boston was conducted in dedicated cleanroom facilities up to the setup of the library amplification step.

In Dublin, after surface cleaning by fine sandblasting the dentine area of roots and crowns was milled to obtain fine powder. All Old Bering Sea and Ust'-Belaya samples were processed this way. In Boston, rib bones from the Aleutian Islands were cleaned at the surface with a sanding disk and fine powder was collected for DNA extraction by drilling into the cleaned area. For the petrous parts (for the 3 Tochak McGrath Athabaskan samples), the cochlear region of the inner ear was extracted by sandblasting and subsequently milled into fine powder.

About 75 mg (+/- 9 mg) of powder was then used for DNA extraction following an established protocol by Dabney *et al.* (2013), with modifications as in Korlević *et al.* (2015): MinElute/Zymo funnel assembly was replaced by the funnel-column assembly from the Roche High Pure Viral Nucleic Acid Large Volume Kit. The final volume of DNA extract was 90 µl.

A double-stranded barcoded Illumina library was prepared for each sample using the 'partial UDG treatment' protocol (Rohland *et al.* 2015). For three of the libraries the settings were identical to the original publication, and for the remaining 14 libraries the following updated setting were used: 1) the elution volume after the MinElute cleanup of the ligation reaction was reduced from 20 µl to 16 µl; 2) the Fill-in reaction volume was reduced from 40 µl to 25 µl; 3) the ThermoPol buffer was replaced by the Isothermal amplification buffer; 4) *Bst* polymerase, large fragment (New England Biolabs), was replaced by *Bst* 2.0 Polymerase, large fragment (New England Biolabs); 5) PCR volume was reduced from 400 µl to 100 µl.

After cleanup of the amplified libraries, we performed a screening step: a capture enrichment targeting the mitochondrial genome and additional nuclear loci (manuscript in preparation) following the procedure described in Maricic *et al.* (2010). After a unique pair of indices was added to each enriched library, we then sequenced the enrichment product together with the original libraries (also after addition of a unique index pair to each library) – shotgun, on an Illumina NextSeq 500 instrument for 2x 76 cycles and 2x 7cycles.

Nuclear data were produced by enriching the original short libraries for 1.24 million SNP loci following the protocol by Fu *et al.* 2015 (SNP information in Haak *et al.* 2015, Mathieson *et al.* 2015). For all but 3 libraries this was done in a single capture enrichment reaction (1,240 thousand loci) for each library, while for the 3 libraries the capture reactions were split into two (390 thousand and 840 thousand loci). After the addition of a unique index pair to each library, sequencing was performed on an Illumina NextSeq 500 instrument for 2x 76 cycles and 2x 7cycles.

Sample I0719 (ancient Aleutian islander) was also shotgun sequenced on two NextSeq 500 flow-cells for 2x 76 cycles and 2x 7 cycles.

References (for this section)

- Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15758–15763 (2013).
- Fu, Q. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
- Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).

Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).

Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* **5**, e14004 (2010).

Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).

Rohland, N. *et al.* Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130624 (2015).

Table S3.1. DNA extraction, library preparation and nuclear targeted enrichment.

individual ID	sample type	powder produced in	powder used for extraction, mg	extraction protocol	extract used for library preparation, µl	library preparation	nuclear capture
I0712	bone (rib)	Boston	74	Dabney <i>et al.</i> 2013 [#]	30	Rohland <i>et al.</i> 2015	390k + 840k
I0719	bone (rib)	Boston	68	Dabney <i>et al.</i> 2013 [#]	30	Rohland <i>et al.</i> 2015	390k + 840k
I0721	bone (rib)	Boston	74	Dabney <i>et al.</i> 2013 [#]	30	Rohland <i>et al.</i> 2015	390k + 840k
I1118	bone (rib)	Boston	67	Dabney <i>et al.</i> 2013 [#]	30	*	1240k
I1123	bone (rib)	Boston	76	Dabney <i>et al.</i> 2013 [#]	30	*	1240k
I1124	bone (rib)	Boston	75	Dabney <i>et al.</i> 2013 [#]	30	*	1240k
I1125	bone (rib)	Boston	74	Dabney <i>et al.</i> 2013 [#]	30	*	1240k
I1126	bone (rib)	Boston	74	Dabney <i>et al.</i> 2013 [#]	30	*	1240k
I1127	bone (rib)	Boston	73	Dabney <i>et al.</i> 2013 [#]	30	*	1240k
I1128	bone (rib)	Boston	73	Dabney <i>et al.</i> 2013 [#]	30	*	1240k
I1129	bone (rib)	Boston	73	Dabney <i>et al.</i> 2013 [#]	30	*	1240k
I1524	molar	Dublin	68	Dabney <i>et al.</i> 2013 [#]	30	*	1240k
I1525	molar	Dublin	72	Dabney <i>et al.</i> 2013 [#]	30	*	1240k
I1526	molar	Dublin	71	Dabney <i>et al.</i> 2013 [#]	30	*	1240k
I5319	pars petrosa	Boston	83	Dabney <i>et al.</i> 2013 [#]	10	*	1240k
I5320	pars petrosa	Boston	75	Dabney <i>et al.</i> 2013 [#]	10	*	1240k
I5321	pars petrosa	Boston	66	Dabney <i>et al.</i> 2013 [#]	10	*	1240k

[#] the funnel-column assembly from the Roche kit as in Korlević *et al.* (2015), elution in 2x 45 µl.

* Rohland *et al.* (2015) with the following modifications: 1) the elution volume after the MinElute cleanup of the ligation reaction was reduced from 20 µl to 16 µl; 2) the Fill-in reaction volume was reduced from 40 µl to 25 µl; 3) the ThermoPol buffer was replaced by the Isothermal amplification buffer; 4) *Bst* polymerase, large fragment (New England Biolabs), was replaced by *Bst* 2.0 Polymerase, large fragment (New England Biolabs); 5) PCR volume was reduced from 400 µl to 100 µl.

3.2 Bioinformatic processing

Raw sequencing data was generated on an Illumina NextSeq 500 instrument. For libraries captured against the set of 1.24 million nuclear SNPs, sample-identifying sequences (barcodes) were trimmed. Adapters were stripped and read pairs with at least 15 bp overlap were merged into a single sequence (allowing for 1 mismatch) at least 30 bp in length, using a modified form of the SeqPrep tool (<https://github.com/jstjohn/SeqPrep>) which retains the highest quality base in the

overlap region. Autosomal sequences were aligned to the human reference genome hg19 (1000 genomes version, downloaded at http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/human_g1k_v37.fasta.gz) using *bwa* v.0.6.1 with the *samse* command (Li and Durbin 2009). Following alignment, clusters of duplicate reads were identified based on start and end position, and orientation; for each cluster of reads, the highest quality representative was used.

For libraries with mitochondrial DNA capture, the same procedure was used, except that the mitochondrial sequences were treated separately and aligned to the RSRS reference genome (Behar et al. 2012) rather than hg19.

References (for this section)

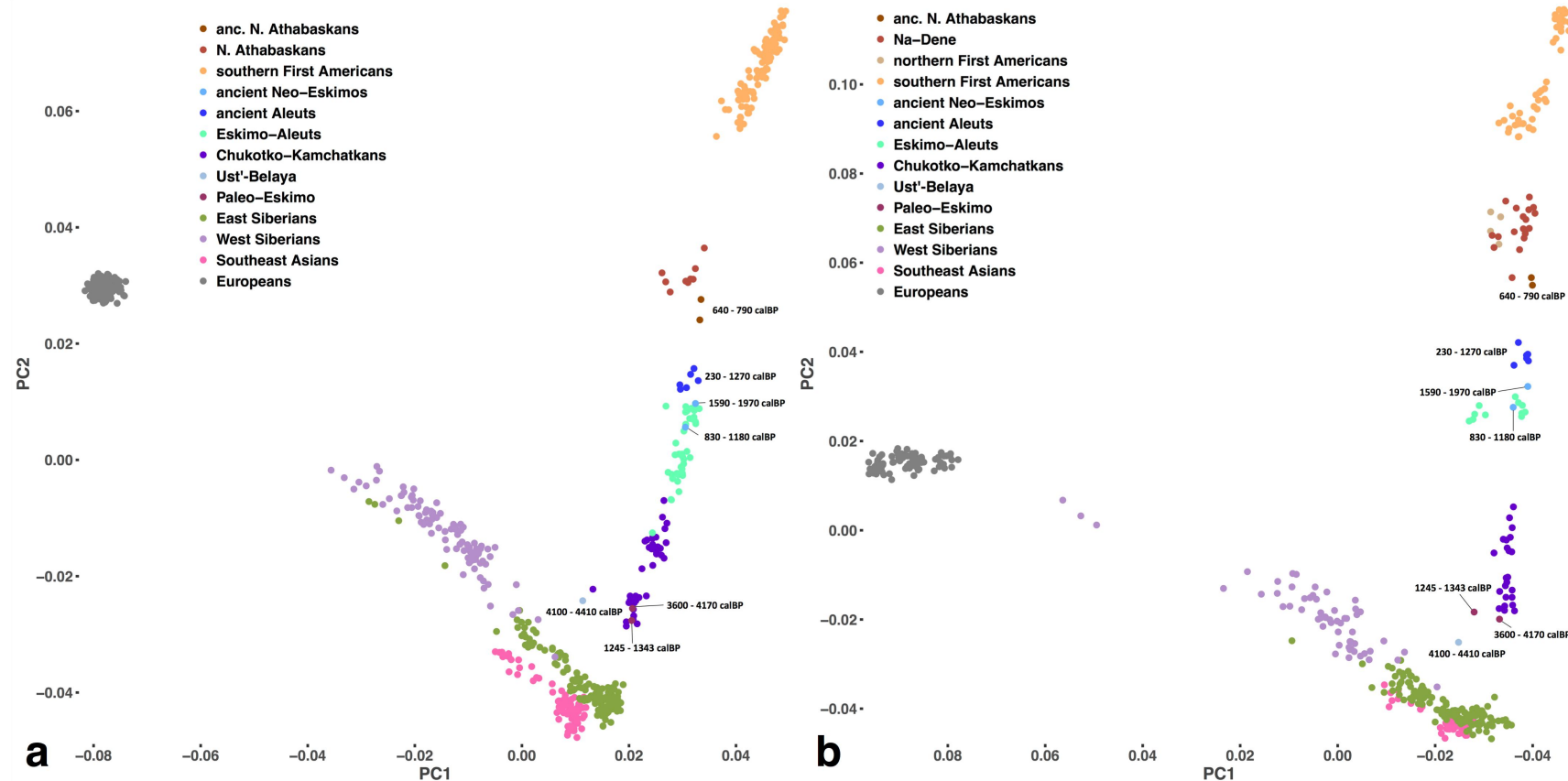
- Behar, D. M. *et al.* A 'Copernican' reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25**, 1754–60 (2009).

Supplementary Information section 4

Principal component analysis

To illustrate the effects of outlier removal procedures, we performed principal component analysis (PCA) on the datasets without transitions used for the *qpWave/apAdm* analyses (Fig. S4.1). Initial manual outlier removal was based on weighted Euclidean distances and *ADMIXTURE* profiles. Subsequently, First American and Eskimo-Aleut-speaking individuals having >1% European, African, or Southeast Asian ancestry according to *ADMIXTURE* were removed, as well as Chukotkan and Kamchatkan individuals with >1% European or African ancestry. PCA plots for original datasets prior to any outlier removal are shown in Fig. 2 and Extended Data Fig. 3.

Figure S4.1. PCA on the HumanOrigins (a) and Illumina (b) datasets without transitions used for the *qpWave/apAdm* analyses. Plots of two principal components (PC1 vs. PC2) are shown (linkage disequilibrium pruning was not applied). The following meta-populations most relevant for our study are plotted: ancient and present-day Eskimo-Aleut and Chukotko-Kamchatkan speakers, Paleo-Eskimos (the Saqqaq, Late Dorset, and Ust'-Belaya individuals), ancient Northern Athabaskans, present-day Na-Dene speakers, northern and southern First Americans, West and East Siberians, Southeast Asians, and Europeans. Calibrated radiocarbon dates in calBP are shown for ancient samples. For individuals, 95% confidence intervals are shown, and for populations, minimal and maximal dates among all confidence intervals of that population are shown.



Supplementary Information section 5

Exhaustive analysis of ancestry streams in small population sets

Prior to modelling Paleo-Eskimo ancestry in indigenous Americans using the *qpAdm* approach, we performed exhaustive analysis of ancestry streams, using *qpWave* applied to small sets of three or four American populations. Each quadruplet included four different meta-populations, and all combinations of populations were tested: southern First Americans (23 populations); Eskimo-Aleut speakers and ancient Aleutians and Neo-Eskimos (5 populations); Northern Athabaskans (3 populations, all present-day Athabaskan speakers were considered alternatively as one population); Paleo-Eskimos (Saqqaq, the low-coverage Late Dorset individual, or both individuals combined). For instance, there are 115 quadruplets composed of Late Dorset, Chipewyans, Eskimo-Aleut speakers/Neo-Eskimos and southern First Americans. Each triplet included three different meta-populations: southern First Americans; Eskimo-Aleut speakers and ancient Aleutians and Neo-Eskimos; Paleo-Eskimos or Chukotko-Kamchatkan speakers (3 populations).

Population counts listed above refer to the HumanOrigins dataset, and below the same information is given for the Illumina dataset. Each quadruplet included four different meta-populations, and all combinations of populations were tested: First Americans (10 populations), Eskimo-Aleut speakers and ancient Aleutians and Neo-Eskimos (5 populations), Na-Dene (6 populations, all present-day Na-Dene speakers were considered alternatively as one population), Paleo-Eskimos (Saqqaq, the low-coverage Late Dorset individual, or both individuals combined). Each triplet included three different meta-populations: southern (7 populations) or northern First Americans (3 populations), Eskimo-Aleut speakers and ancient Aleutians and Neo-Eskimos, Paleo-Eskimos or Chukotko-Kamchatkan speakers (2 populations).

A quadruplet's genetic data were tested for consistency with one to four streams of ancestry derived from the outgroups, and similar testing was done for triplets. Individuals with >1% European, African or Southeast Asian ancestry according to *ADMIXTURE* were removed prior to *qpWave* tests. The following outgroup sets were used for the HumanOrigins dataset. (1) 8 diverse Siberian populations (Nganasan, Tuvian, Ulchi, Yakut, Even, Ket, Selkup, Tubalar, min. population size 8 individuals) and a Southeast Asian population (Dai). (2) A set of 19 outgroups from five broad geographical regions: Mbuti, Taa, Yoruba (Africans); Nganasan, Tuvian, Ulchi, Yakut (East Siberians); Altaian, Ket, Selkup, Tubalar (West Siberians); Czech, English, French, North Italian (Europeans); Dai, Miao, She, Thai (Southeast Asians). (3) 8 Siberian populations, Dai, and a Chukotko-Kamchatkan-speaking population (Koryak), a close outgroup for Americans that should provide higher resolution. This set of outgroups is similar to that used by Reich et al. (2012). (4) If Chukotko-Kamchatkan speakers were included into the ingroup set, Koryaks in the outgroup set were replaced with Saqqaq.

The following outgroup sets were used for the Illumina dataset. (1) 9 diverse Siberian populations (Buryat, Dolgan, Evenk, Nganasan, Tuvian, Yakut, Altaian, Khakas, Selkup, min. population size 10 individuals) and a Southeast Asian population (Dai). (2) A set of 20 outgroups from five broad geographical regions: Bantu (Kenya), Mandenka, Mbuti, Yoruba (Africans); Buryat, Evenk, Nganasan, Tuvian, Yakut (East Siberians); Altaian, Khakas, Selkup (West Siberians); Basque, Sardinian, Slovak, Spanish (Europeans); Dai, Lahu, Miao, She (Southeast Asians). (3) 9 Siberian populations, Dai, and a Chukotko-Kamchatkan-speaking population (Koryak), a close outgroup for Americans that should provide higher resolution. (4) If Chukotko-Kamchatkan speakers were included into the ingroup set, Koryaks in the outgroup set were replaced with Saqqaq.

The results of this exhaustive *qpWave* analyses on thousands of population quadruplets and triplets are summarized for the HumanOrigins (Tables S5.1, S5.2) and Illumina (Table S5.4) datasets without transition polymorphisms, as well as for full dataset versions (Tables S5.3, S5.5). We also tested dropping Chukotkan Yupik populations from the Eskimo-Aleut-speaking group (Table S5.2). Both Yupik Chaplin/Sireniki and Yupik Naukan have Chukotko-Kamchatkan admixture (Extended Data Fig. 1a), and thus might increase the number of migration streams consistent with the data.

635 **Table S5.1.** Results of *qpWave* tests for quadruplets and triplets composed of American and Chukotkan/Kamchatkan populations on the
636 HumanOrigins dataset without transition polymorphisms. The bars show the number of quadruplets/triplets passing the test for 2, 3, or 4 ancestry
637 streams (p -value <0.05).

populaton count:				# quadruplets	outgroups	migration streams			outgroups	Yupik included migration streams			outgroups	migration streams		
23	5	4	3			2	3	4		2	3	4		2	3	4
southern First Americans	Eskimo-Aleuts	Chipewyan	Late Dorset	115	8 Siberians + Dai				Siberians, Southeast Asians, Europeans, Africans (19 pop.)				8 Siberians + Dai + Koryak			
			Saqqaq+Late Dorset	115												
			Saqqaq	115												
		Dakelh	Late Dorset	115												
			Saqqaq+Late Dorset	115												
			Saqqaq	115												
		present-day Na-Dene	Late Dorset	115												
			Saqqaq+Late Dorset	115												
			Saqqaq	115												
		ancient Athabaskan	Late Dorset	115												
			Saqqaq+Late Dorset	115												
			Saqqaq	115												
23	5		3	# triplets	8 Siberians + Dai				Siberians, Southeast Asians, Europeans, Africans (19 pop.)				8 Siberians + Dai + Koryak			
southern First Americans	Eskimo-Aleuts		Late Dorset	115												
			Saqqaq+Late Dorset	115												
			Saqqaq	115												
3	5		3	# triplets												
Na-Dene	Eskimo-Aleuts		Late Dorset	15												
			Saqqaq+Late Dorset	15												
			Saqqaq	15												
23	5		3	# triplets									8 Siberians + Dai + Saqqaq			
southern First Americans	Eskimo-Aleuts		Chukchi	115												
			Itelmen	115												
			Koryak	115												
3	5		3	# triplets												
Na-Dene	Eskimo-Aleuts		Chukchi	15												
			Itelmen	15												
			Koryak	15												

638

639 **Table S5.2.** This table is identical to Table S5.1, except that two Chukotkan Yupik populations were dropped from the Eskimo-Aleut-speaking group.
640 The bars show the number of quadruplets/triplets passing the test for 2, 3, or 4 ancestry streams (p -value <0.05).

populaton count:					outgroups	migration streams			outgroups	Yupik removed migration streams			outgroups	migration streams		
23	3	4	3	# quadruplets		2	3	4		2	3	4		2	3	4
southern First Americans	Eskimo-Aleuts	Chipewyan	Late Dorset	69	8 Siberians + Dai				Siberians, Southeast Asians, Europeans, Africans (19 pop.)				8 Siberians + Dai + Koryak			
			Saqqaq+Late Dorset	69												
			Saqqaq	69												
		Dakelh	Late Dorset	69												
			Saqqaq+Late Dorset	69												
			Saqqaq	69												
		present-day Na-Dene	Late Dorset	69												
			Saqqaq+Late Dorset	69												
			Saqqaq	69												
		ancient Athabaskan	Late Dorset	69												
			Saqqaq+Late Dorset	69												
			Saqqaq	69												
23	3	3	# triplets		8 Siberians + Dai				Siberians, Southeast Asians, Europeans, Africans (19 pop.)				8 Siberians + Dai + Koryak			
southern First Americans	Eskimo-Aleuts	Late Dorset	69													
		Saqqaq+Late Dorset	69													
		Saqqaq	69													
3	3	3	# triplets													
Na-Dene	Eskimo-Aleuts	Late Dorset	9													
		Saqqaq+Late Dorset	9													
		Saqqaq	9													
23	3	3	# triplets										8 Siberians + Dai + Saqqaq			
southern First Americans	Eskimo-Aleuts	Chukchi	69													
		Itelmen	69													
		Koryak	69													
3	3	3	# triplets													
Na-Dene	Eskimo-Aleuts	Chukchi	9													
		Itelmen	9													
		Koryak	9													

643 **Table S5.3.** The same results as in Table S5.1, but on the full HumanOrigins dataset including transitions. Both Yupik populations are included into
644 the Eskimo-Aleut-speaking group. The bars show the number of quadruplets/triplets passing the test for 2, 3, or 4 ancestry streams (p -value <0.05).
645

populaton count:				# quadruplets	outgroups	migration streams			outgroups	migration streams			outgroups	migration streams		
23	5	4	3			2	3	4		2	3	4		2	3	4
southern First Americans	Eskimo-Aleuts	Chipewyan	Late Dorset	115	8 Siberians + Dai				Siberians, Southeast Asians, Europeans, Africans (19 pop.)				8 Siberians + Dai + Koryak			
			Saqqaq+Late Dorset	115												
			Saqqaq	115												
		Dakelh	Late Dorset	115												
			Saqqaq+Late Dorset	115												
			Saqqaq	115												
		present-day Na-Dene	Late Dorset	115												
			Saqqaq+Late Dorset	115												
			Saqqaq	115												
		ancient Athabaskan	Late Dorset	115												
			Saqqaq+Late Dorset	115												
			Saqqaq	115												
23	5		3	# triplets	8 Siberians + Dai				Siberians, Southeast Asians, Europeans, Africans (19 pop.)				8 Siberians + Dai + Koryak			
southern First Americans	Eskimo-Aleuts		Late Dorset	115												
			Saqqaq+Late Dorset	115												
			Saqqaq	115												
3	5		3	# triplets												
Na-Dene	Eskimo-Aleuts		Late Dorset	15												
			Saqqaq+Late Dorset	15												
			Saqqaq	15												
23	5		3	# triplets									8 Siberians + Dai + Saqqaq			
southern First Americans	Eskimo-Aleuts		Chukchi	115												
			Itelmen	115												
			Koryak	115												
3	5		3	# triplets												
Na-Dene	Eskimo-Aleuts		Chukchi	15												
			Itelmen	15												
			Koryak	15												

648
649
650

Table S5.4. Results of *qpWave* tests for quadruplets and triplets composed of American and Chukotkan/Kamchatkan populations on the Illumina dataset without transition polymorphisms. The bars show the number of quadruplets/triplets passing the test for 2, 3, or 4 ancestry streams (p -value <0.05).

population count:				# quadruplets	outgroups	migration streams			outgroups	migration streams			outgroups	migration streams		
10	5	7	3			2	3	4		2	3	4		2	3	4
southern or northern First Americans	Eskimo-Aleuts	Dakelh	Late Dorset	50	9 Siberians + Dai				Siberians, Southeast Asians, Europeans, Africans (20 pop.)				9 Siberians + Dai + Koryak			
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
		Northern Athabaskan 2	Late Dorset	50												
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
		Northern Athabaskan 3	Late Dorset	50												
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
		Southern Athabaskan	Late Dorset	50												
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
		Tlingit	Late Dorset	50												
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
		present-day Na-Dene	Late Dorset	50												
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
		ancient Athabaskan	Late Dorset	50												
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
3	5		3	# triplets	9 Siberians + Dai				Siberians, Southeast Asians, Europeans, Africans (20 pop.)				9 Siberians + Dai + Koryak			
northern First Americans	Eskimo-Aleuts	Late Dorset	15													
		Saqqaq+Late Dorset	15													
		Saqqaq	15													
6	5		3	# triplets												
Na-Dene	Eskimo-Aleuts	Late Dorset	30													
		Saqqaq+Late Dorset	30													
		Saqqaq	30													
7	5		3	# triplets												
southern First Americans	Eskimo-Aleuts	Late Dorset	35													
		Saqqaq+Late Dorset	35													
		Saqqaq	35													
3	5		2	# triplets												
northern First Americans	Eskimo-Aleuts	Chukchi	15													
		Koryak	15													
		Koryak	15													
6	5		2	# triplets												
Na-Dene	Eskimo-Aleuts	Chukchi	30													
		Koryak	30													
		Koryak	30													
7	5		2	# triplets												
southern First Americans	Eskimo-Aleuts	Chukchi	35													
		Koryak	35													

652 **Table S5.5.** The same results as in Table S5.4, but on the full Illumina dataset including transitions. The bars show the number of
653 quadruplets/triplets passing the test for 2, 3, or 4 ancestry streams (p -value <0.05).
654

population count:				# quadruplets	outgroups	migration streams			outgroups	migration streams			outgroups	migration streams		
10	5	7	3			2	3	4		2	3	4		2	3	4
southern or northern First Americans	Eskimo-Aleuts	Dakelh	Late Dorset	50	9 Siberians + Dai				Siberians, Southeast Asians, Europeans, Africans (20 pop.)				9 Siberians + Dai + Koryak			
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
		Northern Athabaskan 2	Late Dorset	50												
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
		Northern Athabaskan 3	Late Dorset	50												
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
		Southern Athabaskan	Late Dorset	50												
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
		Tlingit	Late Dorset	50												
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
		present-day Na-Dene	Late Dorset	50												
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
		ancient Athabaskan	Late Dorset	50												
			Saqqaq+Late Dorset	50												
			Saqqaq	50												
3	5		3	# triplets	9 Siberians + Dai				Siberians, Southeast Asians, Europeans, Africans (20 pop.)				9 Siberians + Dai + Koryak			
northern First Americans	Eskimo-Aleuts	Late Dorset	15													
		Saqqaq+Late Dorset	15													
		Saqqaq	15													
6	5		3	# triplets												
Na-Dene	Eskimo-Aleuts	Late Dorset	30													
		Saqqaq+Late Dorset	30													
		Saqqaq	30													
7	5		3	# triplets												
southern First Americans	Eskimo-Aleuts	Late Dorset	35													
		Saqqaq+Late Dorset	35													
		Saqqaq	35													
3	5		2	# triplets									9 Siberians + Dai + Koryak			
northern First Americans	Eskimo-Aleuts	Chukchi	15													
		Koryak	15													
6	5		2	# triplets												
Na-Dene	Eskimo-Aleuts	Chukchi	30													
		Koryak	30													
7	5		2	# triplets												
southern First Americans	Eskimo-Aleuts	Chukchi	35													
		Koryak	35													

Supplementary Information section 6

Haplotype sharing statistics

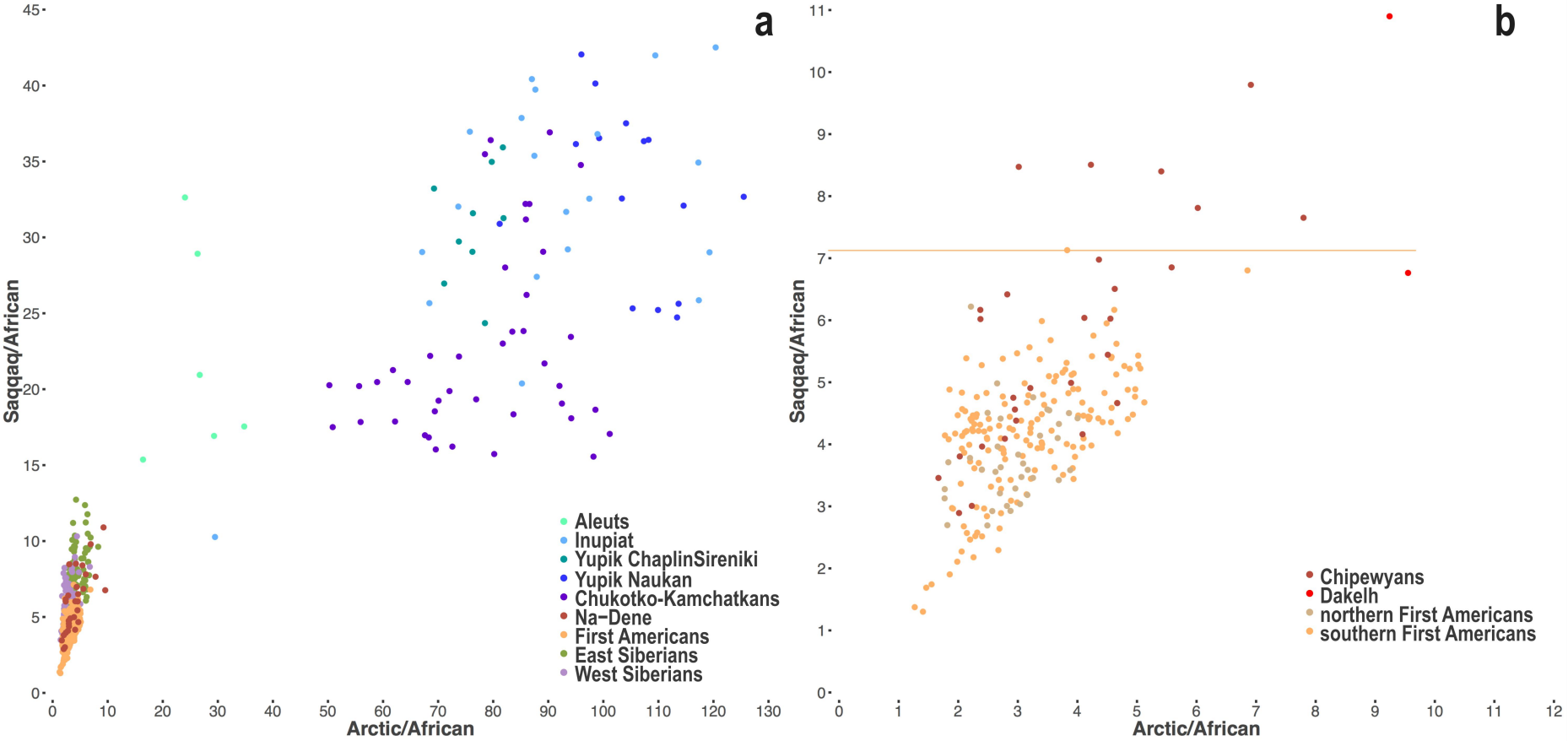
To investigate Paleo-Eskimo ancestry in First Americans in a hypothesis-free way, we considered haplotypes shared between First Americans and the ancient Saqqaq individual. Cumulative lengths of shared autosomal haplotypes were produced with *ChromoPainter v.1* for pairs of individuals, in the form of all vs. all “coancestry matrices” (Lawson et al. 2012). First, for each American individual we considered the length of haplotypes shared with Saqqaq (in cM), which we refer to as Saqqaq haplotype sharing statistic or HSS. We estimated haplotype sharing between an American individual and Africans, Europeans, Siberians, and Arctic (Chukotko-Kamchatkan- and Eskimo-Aleut-speaking) individuals by averaging HSS across members of a given meta-population. To take care of potential biases, the Saqqaq HSS was divided by the African HSS, and the resulting statistic was termed relative HSS. Alternatively, Siberian HSS was used as a normalizer. To visually assess correlation of haplotype sharing with Saqqaq and with closely related Chukotko-Kamchatkan- and Eskimo-Aleut-speaking populations collectively termed Arctic, we combined relative Saqqaq HSSs and relative Arctic HSS on two-dimensional plots. Both the HumanOrigins (Fig. S6.1) and the Illumina (Fig. S6.2) datasets with a more diverse collection of Na-Dene-speaking individuals were analyzed.

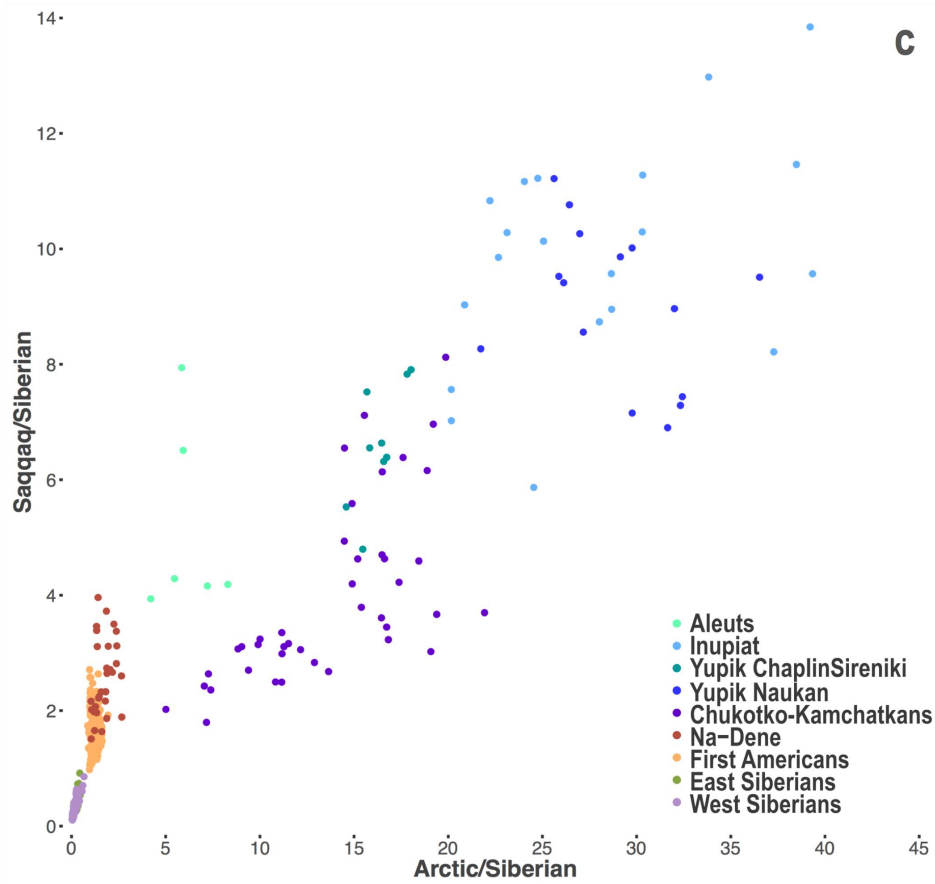
Since the ancient Saqqaq individual has demonstrable genetic affinities to both Arctic and Siberian meta-populations (Rasmussen et al. 2010, Raghavan et al. 2014, 2015, Flegontov et al. 2016, see also *ADMIXTURE* profiles in Extended Data Fig. 1), we also scrutinized relative Arctic and Siberian HSSs (Figs. S6.3, S6.4). We observe that each meta-population is scattered along a line on the Arctic vs. Siberian two-dimensional HSS plot, which reflects similar ratios of the Siberian and Arctic haplotype sharing among its members. The position of a population along the line depends on the presence of other ancestry components. For example, Aleuts, who have a high level of European admixture (Raghavan et al. 2014, 2015) (see also Extended Data Fig. 1), lie much closer to zero on both axes as compared to other Eskimo-Aleut-speaking groups (Fig. S6.3a,c). While First American individuals form a tight cluster, the Athabaskan-speaking Dakelh and some Chipewyans are shifted considerably towards the Saqqaq individual (Fig. S6.3a,c). Since haplotype sharing statistics behave linearly under recent admixture, we used linear combinations to calculate expected HSSs for mixtures of First Americans with Saqqaq or with Eskimo-Aleut-speaking populations. We find that HSSs for two Dakelh (Fig. S6.3b,d) and for several Northern Athabaskan, Southern Athabaskan, and Tlingit individuals (Fig. S6.4b,d) are inconsistent with a recent Inuit or Yupik admixture event, but consistent with Saqqaq admixture. However, these simple simulations do not rule out an ancient admixture event with a Neo-Eskimo group since subsequent drift in Siberians or Arctic groups could have skewed the HSSs.

References (for this section)

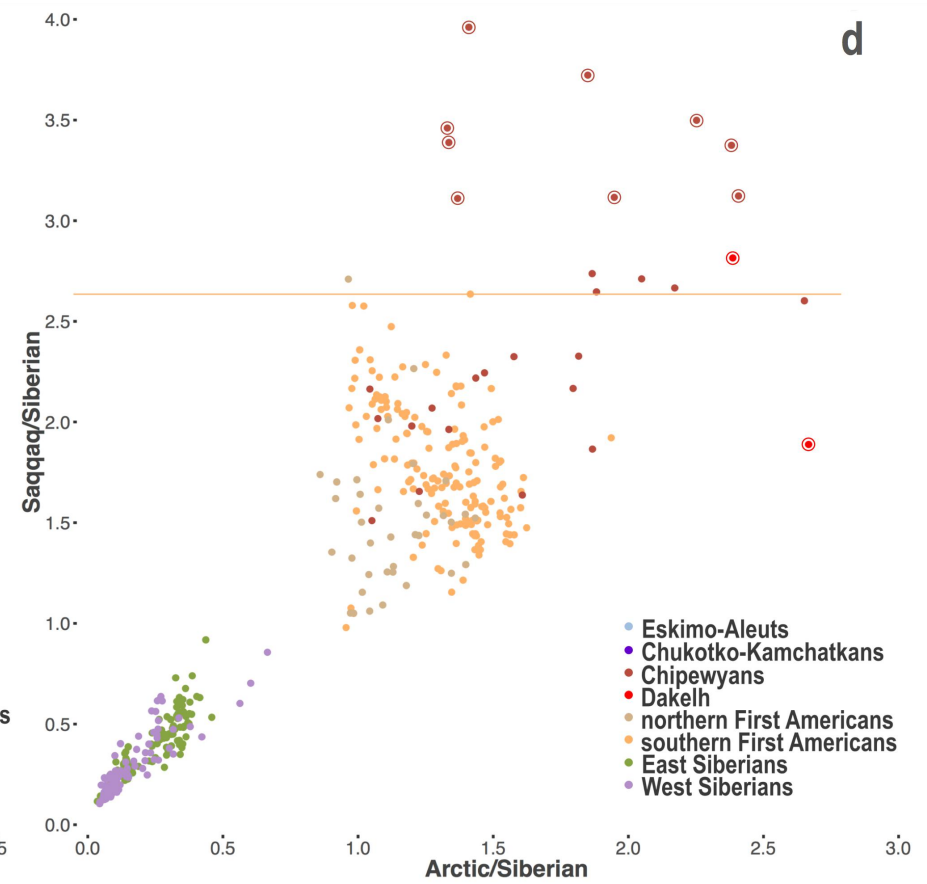
- Flegontov, P. *et al.* Genomic study of the Ket: A Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Sci. Rep.* **6**, 20768 (2016).
- Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, 11–17 (2012).
- Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345**, 1255832 (2014).
- Raghavan, M. *et al.* Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, 1–20 (2015).
- Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).

709 **Figure S6.1.** Two-dimensional plots of Arctic and Saqqaq haplotype sharing statistics normalized using the African (**a, b**) or Siberian (**c, d**) meta-
 710 populations and based on the HumanOrigins SNP array dataset. **a, c**, Plots showing statistics for individuals of all relevant populations and meta-
 711 populations (color-coded according to the legend). **b, d**, Enlarged plots showing statistics for First American individuals. The highest Saqqaq
 712 haplotype sharing statistics among southern First Americans is marked by the horizontal line. Northern Athabaskan-speaking individuals (outliers
 713 on the Arctic and/or Saqqaq axes) selected for the *GLOBETROTTER* analysis are marked with circles in panel **d**.
 714



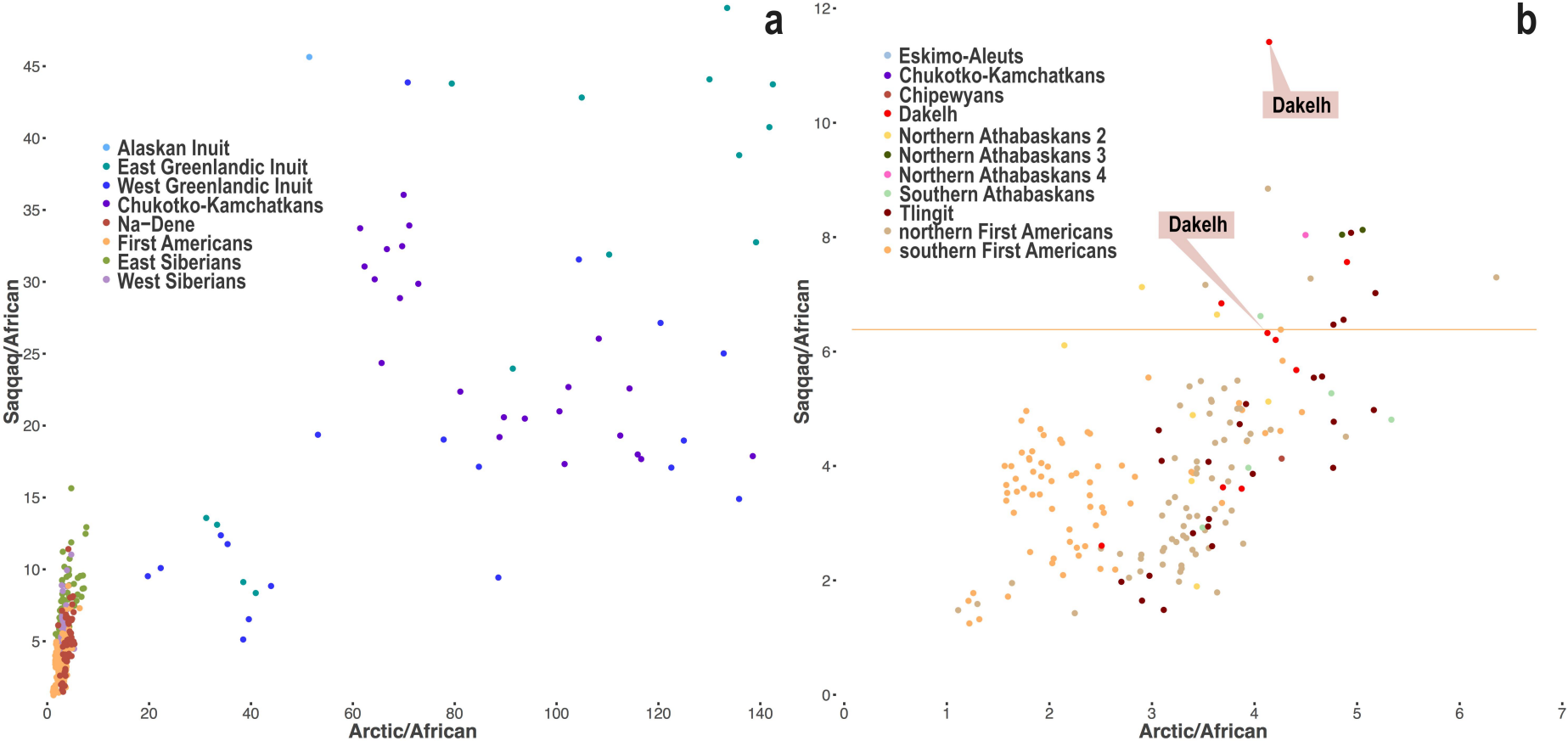


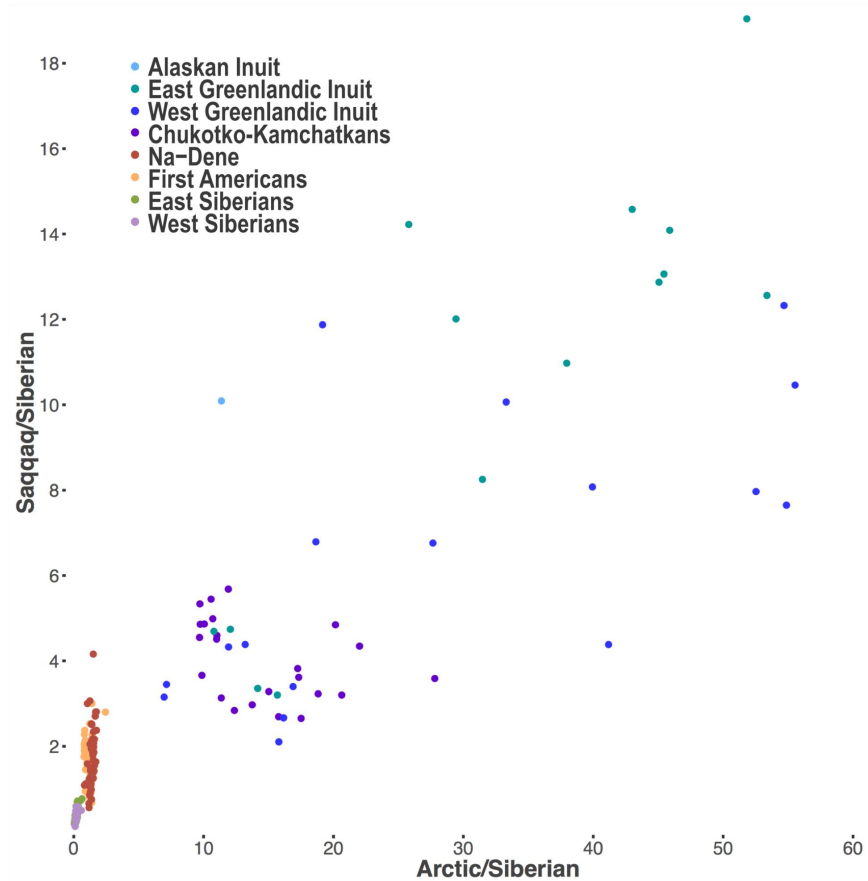
c



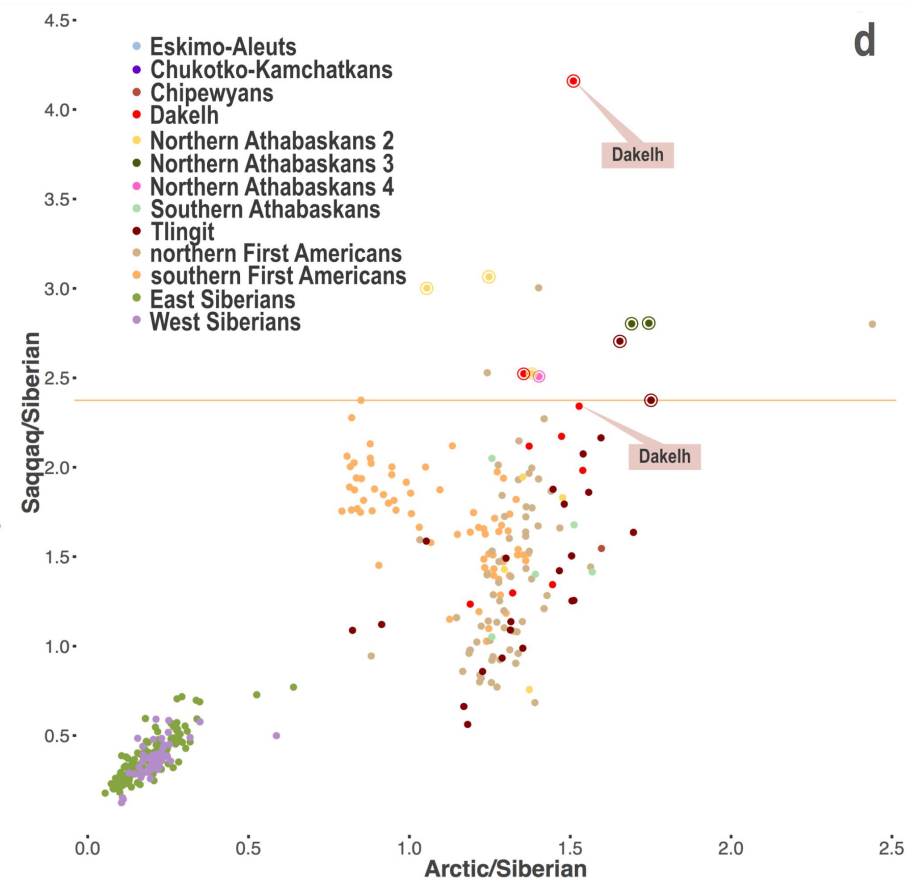
d

719 **Figure S6.2.** The same results as in Fig. S6.1, but on the Illumina SNP array dataset. Northern Athabaskan- and Tlingit-speaking individuals (outliers
 720 on the Arctic and/or Saqqaq axes) selected for the *GLOBETROTTER* analysis are marked with circles in panel **d**. Two Athabaskan-speaking Dakelh
 721 individuals with sequencing data, also included into the HumanOrigins and rare allele datasets, are marked with callouts.
 722



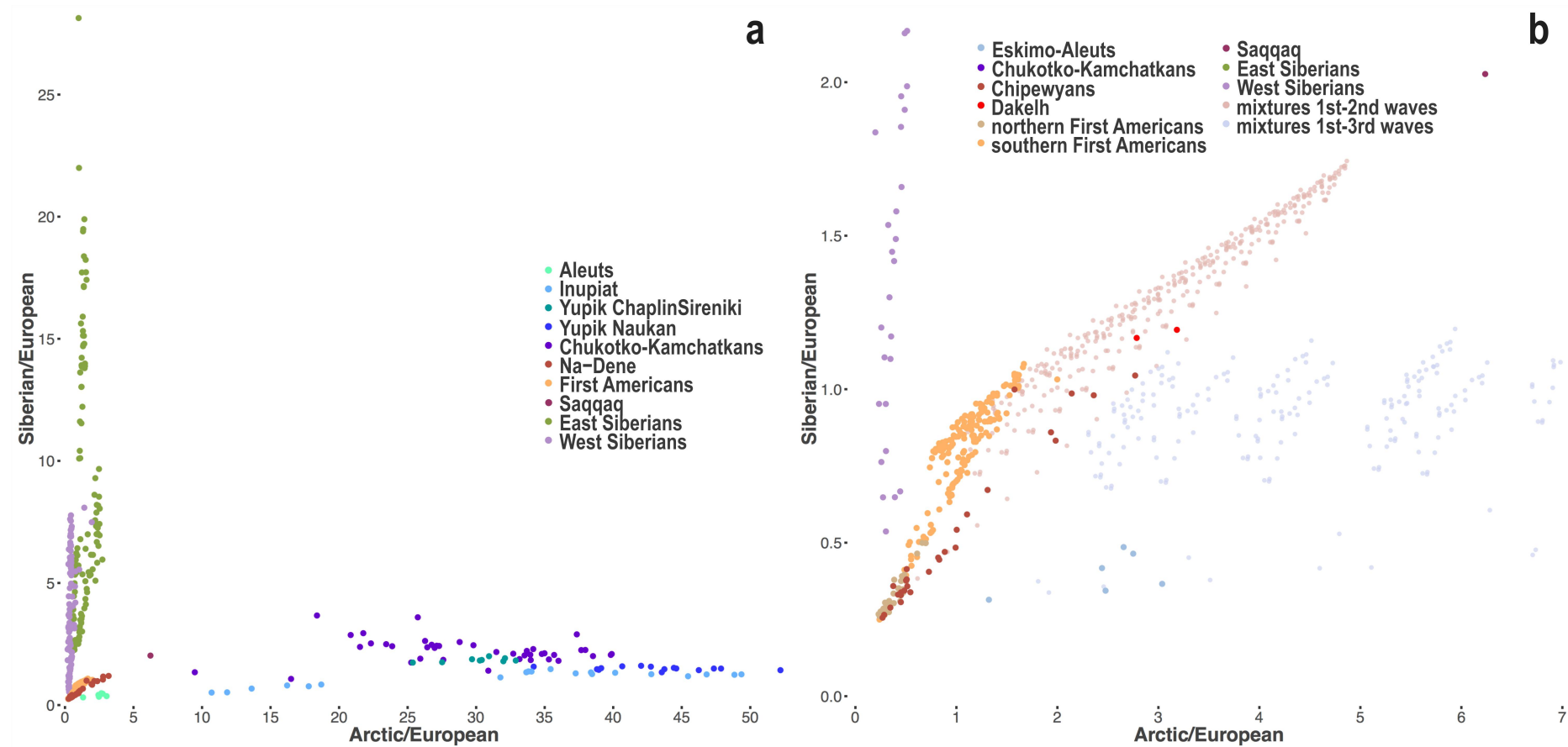


c



d

Figure S6.3. Two-dimensional plots of Arctic and Siberian haplotype sharing statistics normalized using the European (**a, b**) or African (**c, d**) meta-populations and based on the HumanOrigins SNP array dataset. **a, c**, Plots showing statistics for individuals of all relevant populations and meta-populations (color-coded according to the legend). **b, d**, Enlarged areas of the plots showing statistics for First American individuals and simulated mixtures of any present-day southern First American population and the Saqqaq individual (from 5% to 70%, with 5% increments), and similar mixtures with Eskimo-Aleut-speaking populations (>5% of Iñupiat or Yupik ancestry). Average values of the statistics in populations were used to calculate the simulated statistics.



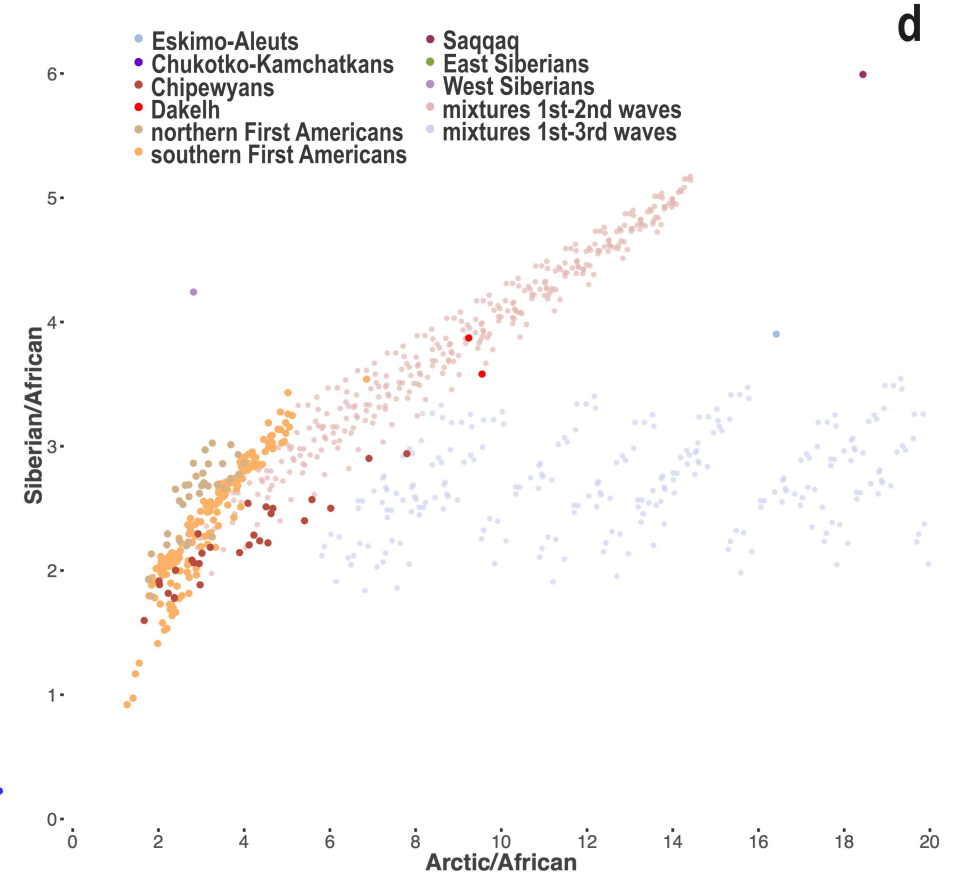
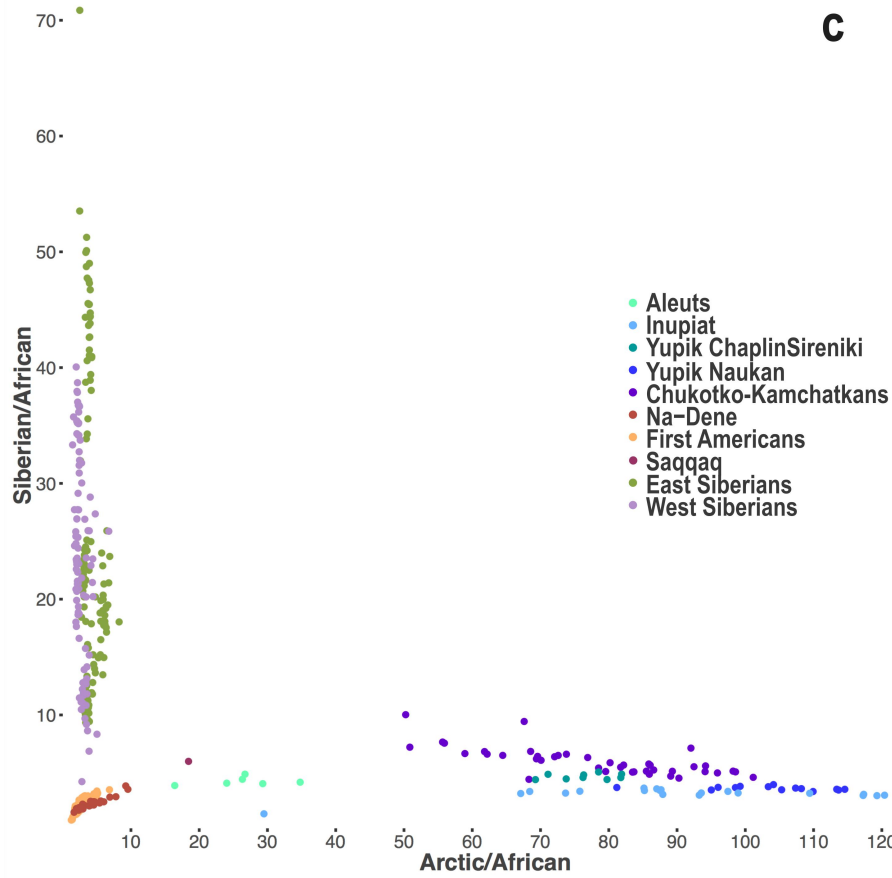
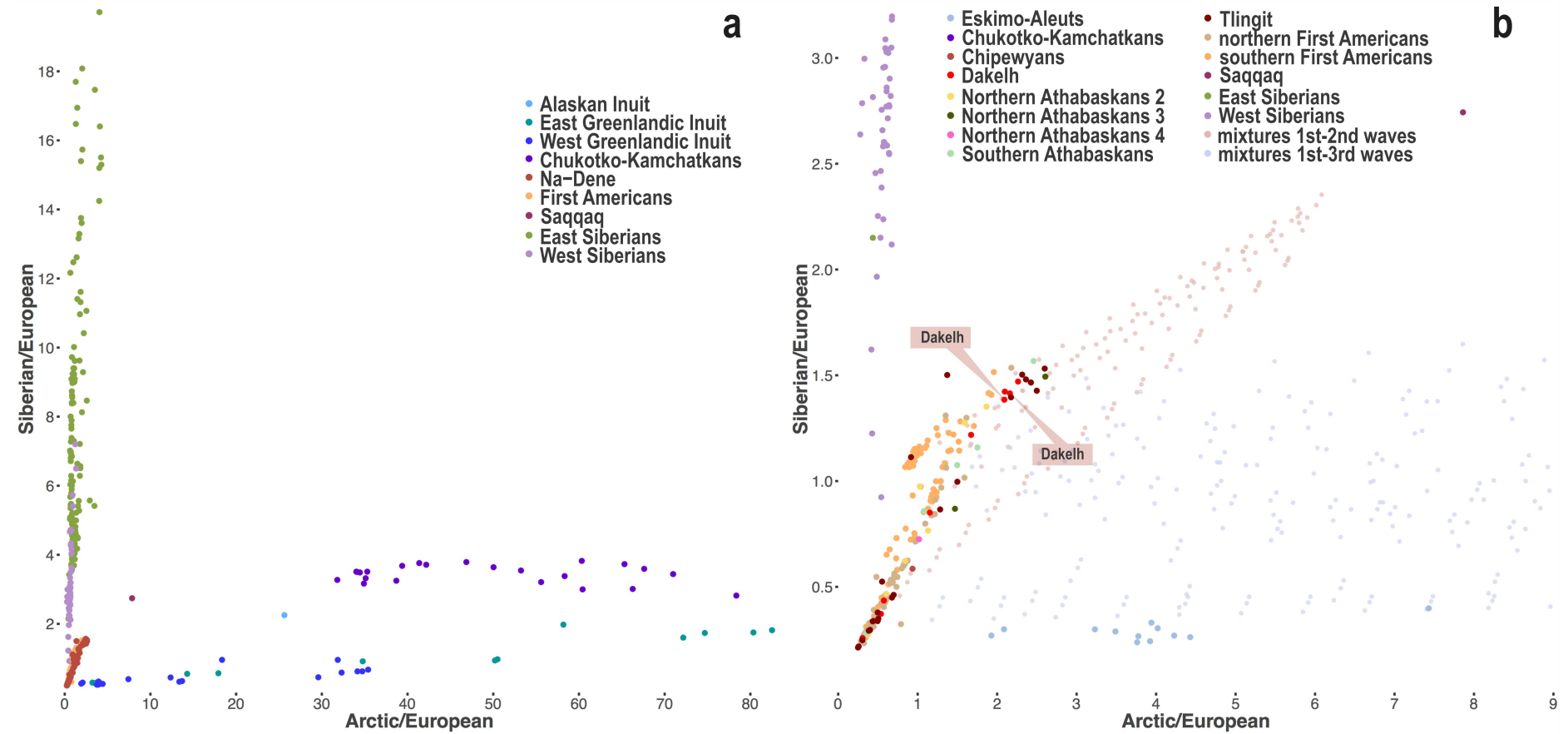
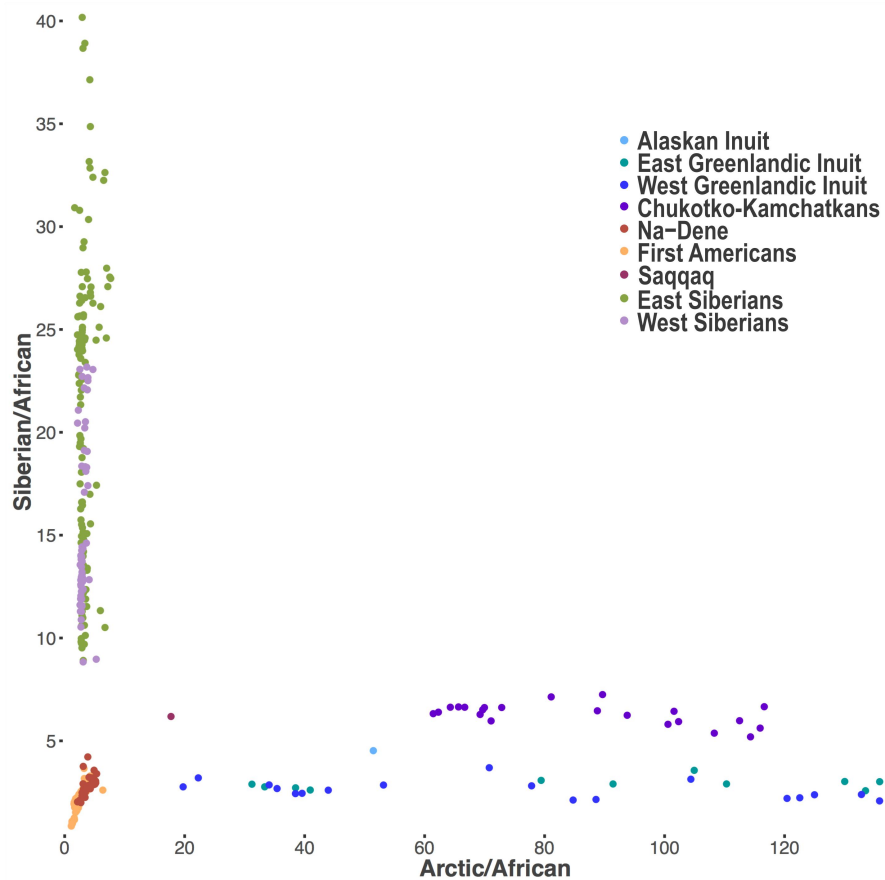
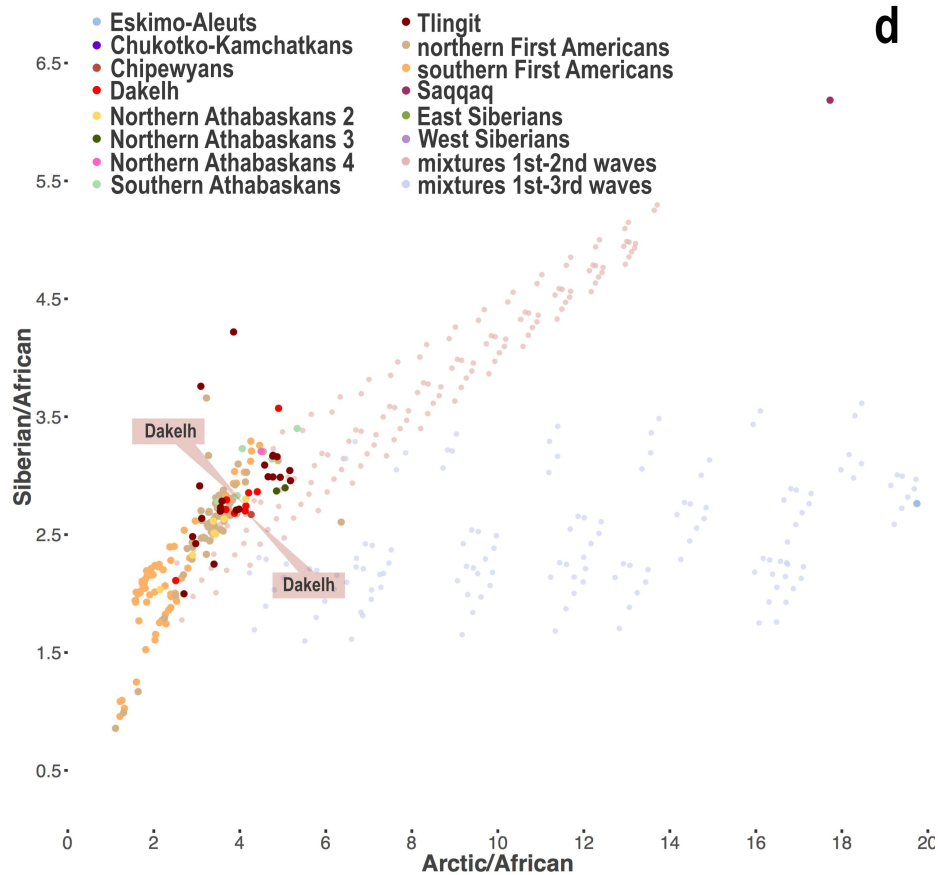


Figure S6.4. The same results as in Fig. S6.3, but on the Illumina SNP array dataset. For calculating simulated mixtures, the following Eskimo-Aleut-speaking populations were used: Alaskan Inuit, East or West Greenlandic Inuit. Various Na-Dene-speaking populations are color-coded, and two Athabaskan-speaking Dakelh individuals with sequencing data, also included into the HumanOrigins and rare allele datasets, are marked with callouts.





c



d

Supplementary Information section 7

Admixture inference with *GLOBETROTTER*

To interpret haplotype sharing in a more quantitative way, we analyzed putative admixture events in Na-Dene using *GLOBETROTTER* (Hellenthal et al. 2014). *GLOBETROTTER* operates on coancestry curves, generated from *ChromoPainter v.2* results (Hellenthal et al. 2014), finds the best proxies of admixture partners in a dataset, determines admixture ratios and dates up to two distinct admixture events. To make a complex mixture history of Na-Dene amenable to *GLOBETROTTER* analysis, we pre-selected individuals based on low European admixture and high Saqqaq HSS (selected individuals are marked on two-dimensional HSS plots in Figs. S6.1d and S6.2d). Meta-populations or separate populations were used as haplotype donors in the *ChromoPainter v.2* analyses. Substantiating our preliminary conclusions, Saqqaq and First Americans were revealed as most likely admixture partners for Na-Dene speakers, with the Saqqaq contribution ranging from 7% to 51%, depending on the dataset and *GLOBETROTTER* set-up. Admixture dates were estimated as follows: 479 – 1,534 calBP (95% confidence interval), if meta-populations were used as haplotype donors, and 1,073 – 2,202 calBP, if populations were used as haplotype donors (Table S7.1, Fig. S7.1).

Table S7.1. The table shows fit statistics for *GLOBETROTTER* coancestry curves, as well as inferred mixture partners, mixture proportions, dates and their 95% confidence intervals. The following abbreviations are used for meta-populations: Eskimo-Aleut speakers, E-A; northern First Americans, NAM; southern First Americans, SAM.

dataset		HumanOrigins	llumina	HumanOrigins	HumanOrigins
haplotype donors		9 meta-populations ^{a)}	9 meta-populations ^{a,b)}	67 populations ^{c)}	67 populations ^{c,d)}
target population		Northern Athabaskans (2 Dakelh, 9 Chipewyans) ^{e)}	2 Tlingit, 8 Northern Athabaskans ^{e)}	Northern Athabaskans (2 Dakelh, 9 Chipewyans) ^{e)}	Northern Athabaskans (2 Dakelh, 9 Chipewyans) ^{e)}
<i>p</i> -value for any admixture event		0	0	0.005	0.005
<i>GLOBETROTTER</i> conclusion		multiple dates	one-date multiway	uncertain	uncertain
coancestry curves	max. goodness-of-fit ^{f)}	0.987	0.503	0.908	0.695
	max. fit improvement for two-date curves ^{f)}	0.297	0.148	0.276	0.186
two dates, admixture event 1	inferred date, calBP	144	522	139	67
	95% confidence interval, calBP	92 – 178	315 – 898	29 – 249	29 – 153
	source 1	27% NAM	47% NAM	32% Cree (NAM)	36% Ojibwa (NAM)
	source 2	73% SAM	53% NAM	68% Nahua (SAM)	64% Nahua (SAM)
two dates, admixture event 2	inferred date, calBP	916	522	1,335	1,574
	95% confidence interval, calBP	479 – 1,534	N/A	739 – 3,487	1,073 – 2,202
	source 1	28% Saqqaq	7% Saqqaq	39% Iñupiat (E-A)	51% Saqqaq
	source 2	72% SAM	93% NAM	61% Cree (NAM)	49% Cree (NAM)

^{a)} The following non-overlapping meta-populations were used: 1/ the Saqqaq ancient genome and 2/ related Chukotko-Kamchatkan-speaking groups (abbreviated as C-K); 3/ Eskimo-Aleut speakers (Aleuts, Inuit, Iñupiat, Yupik, abbreviated as E-A); 4/ northern First Americans (NAM); 5/ southern First Americans (SAM); 6/ West Siberians (WSIB); 7/ East Siberians (ESIB); 8/ Southeast Asians (SEA); 9/ Europeans (EUR).

^{b)} Individuals with >15% West Eurasian admixture components were removed from the NAM meta-population.

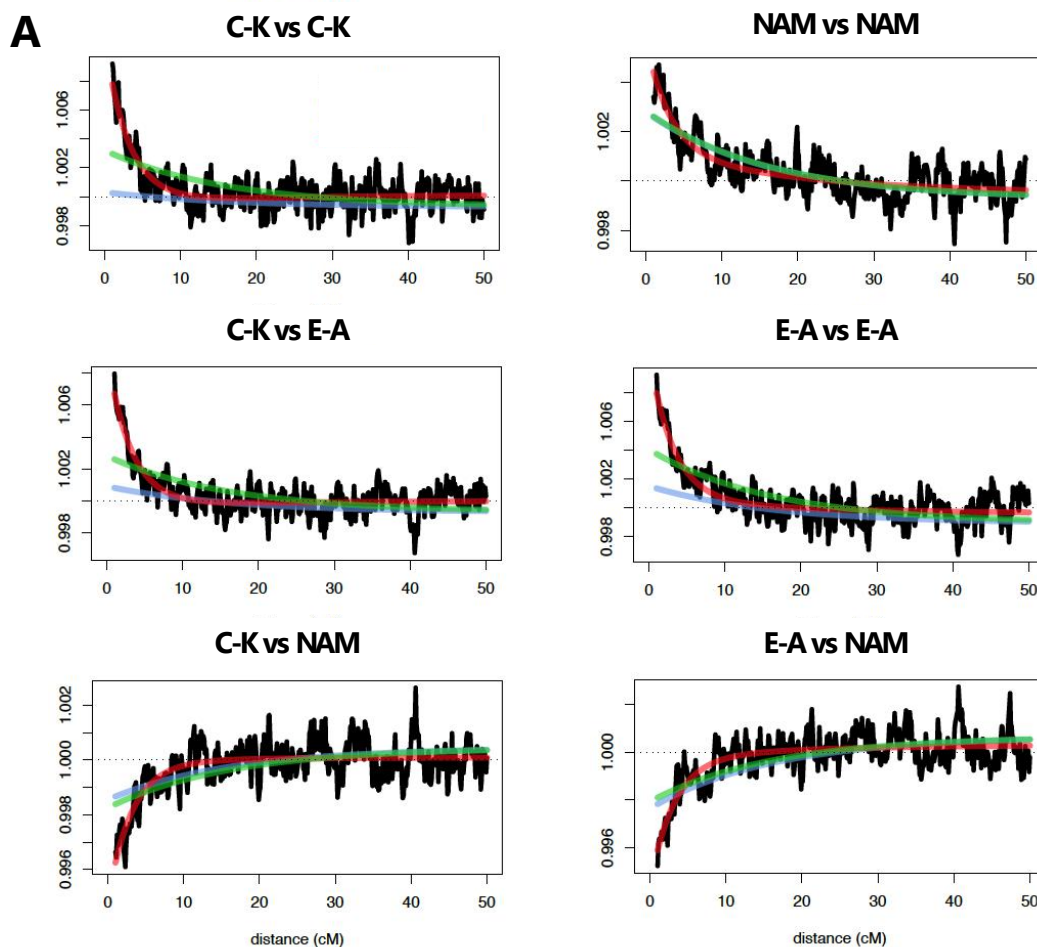
^{c)} Individuals with >15% West Eurasian admixture components were removed from NAM populations, and the remaining NAM individuals were merged into one population.

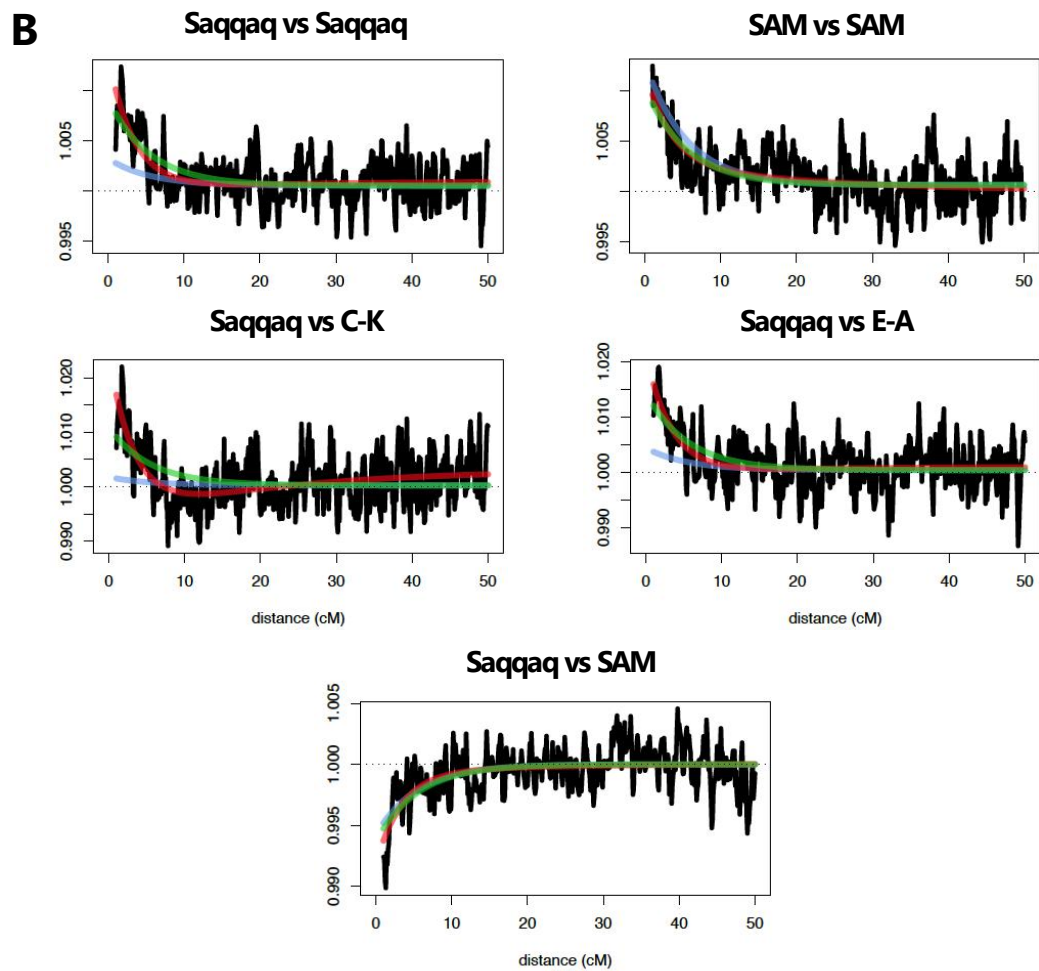
^{d)} Standardizing by a “null” individual was performed to test for consistency, as recommended by the *GLOBETROTTER* manual. This setting might be appropriate if the target population has undergone a bottleneck.

^{e)} To make admixture history of the target population less complex and amenable to *GLOBETROTTER* analysis, only Na-Dene-speaking individuals with prior evidence of elevated Paleo-Eskimo ancestry (Figs. S6.1d, S6.2d) and with <10% West Eurasian ancestry estimated with *ADMIXTURE* (Extended Data Fig. 1) were used.

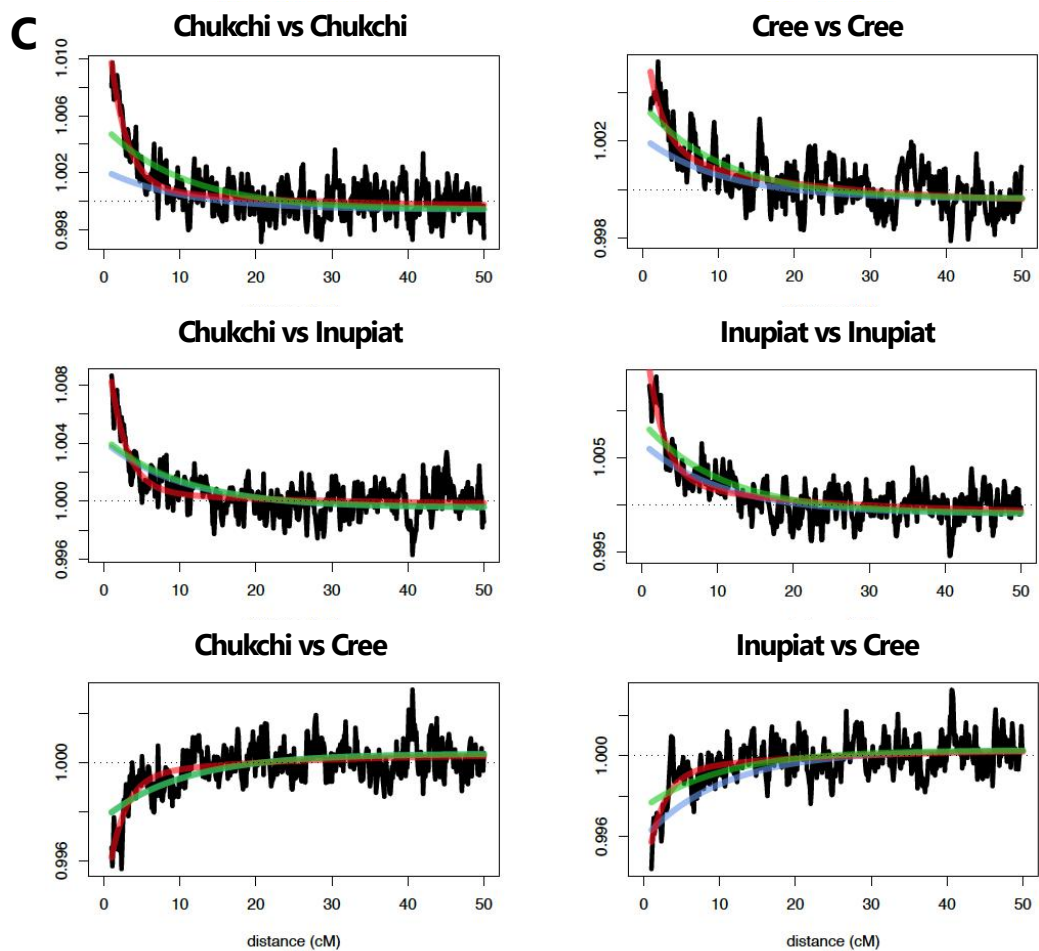
^{f)} A maximal fit value across all curves is shown: two-date curves were considered if the overall conclusion was “multiple dates”, and one-date curves were considered in other cases. Most relevant coancestry curves illustrating the inferred admixture events are shown in Fig. S7.1.

Figure S7.1. Coancestry curves: relative probability of jointly copying two genomic chunks from a pair of donors (y-axis) vs. genetic distance between the chunks in cM (x-axis). Several representative curves are shown for each model: those with the best fit and involving admixture partners inferred with *GLOBETROTTER* or their closest proxies. Only curves reflecting the older Paleo-Eskimo/First American admixture event are shown. Here is a list of *GLOBETROTTER* set-ups we explored: Northern Athabaskan speakers with meta-populations (a) or populations (c) as haplotype donors on the HumanOrigins dataset; Na-Dene speakers with meta-populations as haplotype donors on the Illumina dataset (b). Results under an alternative setting (normalization by a ‘null individual’) are also shown for populations as haplotype donors (d). Original data are shown in black, and curves approximating two admixture events with different dates – in red, two events with a single date – in green, and one event – in blue. Composition of target Na-Dene populations is given in Table S7.1, and Figs. S6.1d, S6.2d. The following meta-populations were used as haplotype donors: 1/ Saqqaq and 2/ related Chukotko-Kamchatkan speakers (abbreviated as C-K); 3/ Eskimo-Aleut speakers (E-A); 4/ northern First Americans (NAM); 5/ southern First Americans (SAM); 6/ West Siberians (WSIB); 7/ East Siberians (ESIB), 8/ Southeast Asians (SEA); 9/ Europeans (EUR).

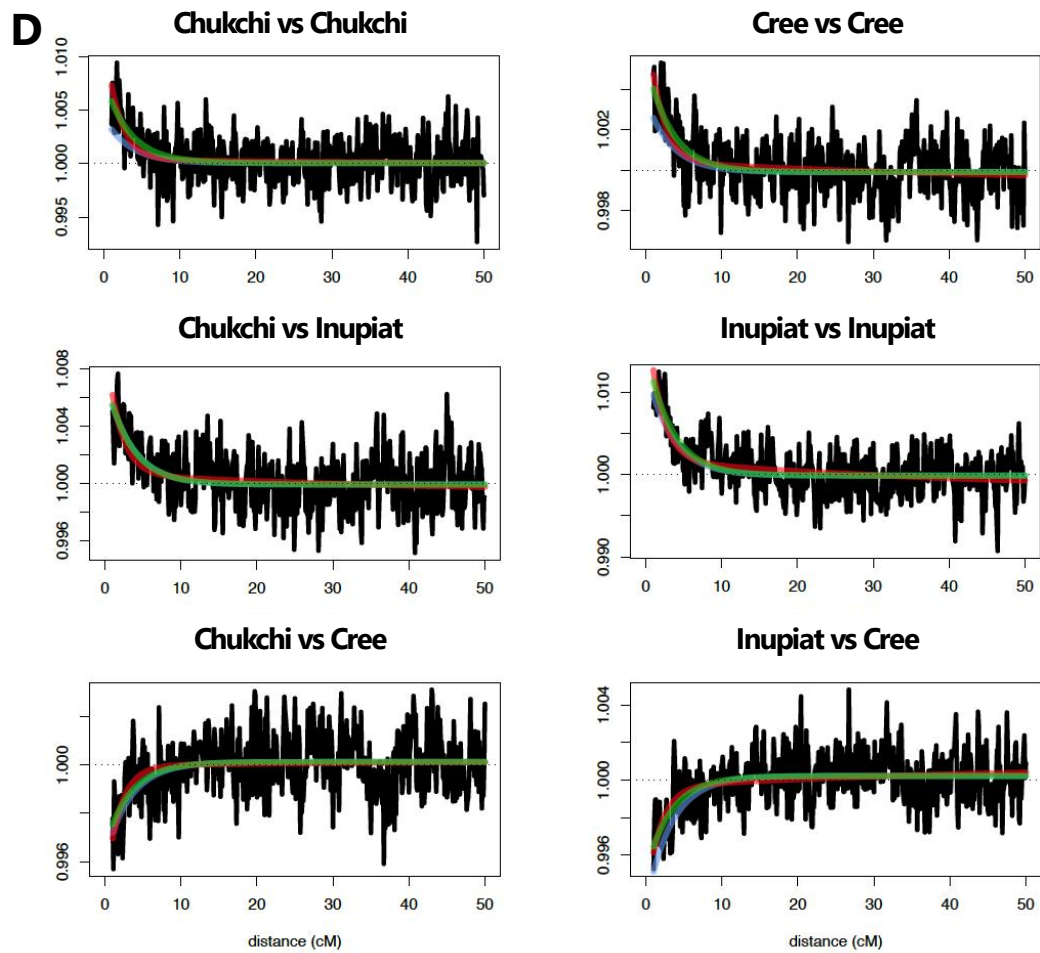




805



806



807

Supplementary Information section 8

Additional results on Aleutian population history

Another controversial chapter of the American Arctic prehistory concerns Aleuts (Balter 2012). The Aleutian Islands were settled much earlier than the American Arctic, about 9,000 calBP (Hatfield 2010), and discontinuities in the Aleutian archaeological record were observed at ~4,500 calBP (Knecht and Davis 2001, Hatfield 2010) and at ~800 – 900 calBP (Brenner Coltrain et al. 2006, Hatfield 2010). The first discontinuity was associated with Paleo-Eskimo, and the latter with Neo-Eskimo influence, although the extent of technological interactions, and the role of genetic continuity vs. population replacement is debated (Brenner Coltrain et al. 2006, Smith et al. 2009, Misarti and Maschner et al. 2015). Three burial sites in the eastern Aleutian Islands received most attention so far: the Chaluka midden site on the Umnak island was associated with an early population (3,600 – 300 calBP) with a dolichocranic morphology, inhumation burials (Hrdlička 1945, Brenner Coltrain et al. 2006) and a predominance of mtDNA haplogroup A2a (Smith et al. 2009). Other sites, at the Kagamil and Ship Rock islands, were associated with a later population (800 – 900 calBP and later), a brachycranial morphology, mummification burials (Hrdlička 1945, Brenner Coltrain et al. 2006) and a predominance of mtDNA haplogroup D2a (Smith et al. 2009). The former population has been historically termed Paleo-Aleut, and the latter Neo-Aleut.

We carried out a small-scale sampling of ancient genomes from all three sites (Table 1, Supplementary Table 1). Radiocarbon dates obtained for these individuals in a previous study (Brenner Coltrain et al. 2006) were recalibrated using a more appropriate marine reservoir correction (Misarti and Maschner 2015) resulting in the following dates: 2,320 – 390 calBP for Paleo-Aleuts and 760 – 140 calBP for Neo-Aleuts (Supplementary Table 1, Supplementary Information section 2). Among 11 ancient Aleuts subjected to in-solution target enrichment of more than 1.2 million SNPs using a protocol by Fu *et al.* (2015), 4 Neo-Aleuts and 2 Paleo-Aleuts passed the 70% missing rate cut-offs that we applied in order to permit high-density SNP analyses and were incorporated into both the HumanOrigins and Illumina SNP array datasets (Supplementary Table 3). In addition, one Paleo-Aleut individual dated to 770 – 390 calBP (IDs I0719 and 378620, the latter used by Brenner Coltrain et al. 2006) was sequenced with the shotgun approach at 2.7x coverage. Due to low coverage of both the enrichment and shotgun data, only pseudo-haploid SNP calls were generated for ancient Aleuts, hence these samples were used for *qpWave/qpAdm*, PCA, *ADMIXTURE*, and rare allele sharing analyses only.

Analyzing these data, we found that four Neo-Aleut samples dated to 760 – 230 calBP and two Paleo-Aleut samples dated to 1270 – 930 and 770 – 390 calBP are indistinguishable. In particular, in both the HumanOrigins and Illumina datasets, the Paleo- and Neo-Aleuts were indistinguishable according to PCA (Fig. 2, Extended Data Fig. 3, Supplementary Information section 4) and *ADMIXTURE* patterns (Extended Data Fig. 1), showing that the Neo-Aleuts arose directly from the Paleo-Aleuts and contradicting suggestions – based on morphology (Hrdlička 1945) and mitochondrial DNA haplogroup frequency changes (Smith et al. 2009) – that the transition between Paleo- and Neo-Aleuts was driven by new migration into the islands from the outside. Pooling the six ancient Aleuts together for *qpWave/qpAdm* analyses (Supplementary Information section 5), we find that both groups have a strong Neo-Eskimo genetic affiliation, and in this respect are similar to present-day Aleuts. In addition, the single Paleo-Aleut genome that we generated was placed into the Eskimo-Aleut branch with high certainty using *rarecoal* (Fig. 5c), but we lack resolution to determine its population affinity within this branch in more detail. Further method development to include not just rare alleles but the full site frequency spectrum might yield further insights into the structure of these closely related populations.

We also used this first data from the Aleutian Islands prior to European colonization to test a claim by Raghavan *et al.* (2015) of a genetic affinity between Papuans and Aleuts. The original study attempted to account for the substantial amounts of recent European ancestry in the present-day Aleutian individuals analyzed by identifying and excluding segments of the genomes that could be reliably called as European in ancestry. However, the accuracy and false negative rate of exclusion of genomic segments with affinity is not unbiased, and so could have affected the original reported signal that had a significance level of $Z=2$ to $Z=3$. We thus used *D-*

statistics to test whether there was evidence of an excess affinity to Papuans in the ancient Aleuts, using a variety of subsets of the data, but find no evidence of an excess affinity to Papuans ($Z < 2$). These results suggest that an excess affinity to Australo-Melanesians is exclusively found in South America and primarily observed in Amazonian populations (Skoglund et al. 2015).

References (for this section)

- Balter, M. The peopling of the Aleutians. *Science* **335**, 158–161 (2012).
- Brenner Coltrain, J. B., Hayes, M.G. & O'Rourke D.H. Hrdlička's Aleutian population-replacement hypothesis. A radiometric evaluation. *Curr. Anthropol.* **47**, 537–548 (2006).
- Fu, Q. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
- Hatfield, V. L. Material culture across the Aleutian archipelago. *Hum. Biol.* **82**, 525–556 (2010).
- Hrdlička, A. *The Aleutian and Commander Islands and their inhabitants*. Philadelphia: Wistar Institute of Anatomy and Biology (1945).
- Knecht, R. A. & Davis, R. S. A prehistoric sequence for the eastern Aleutians. *Archaeology in the Aleut zone of Alaska: Some recent research*, ed. Dumond, D. *University of Oregon Anthropological Papers* **58**, 269–288 (2001).
- Misarti, N. & Maschner, H. D. G. The Paleo-Aleut to Neo-Aleut transition revisited. *J. Anthropol. Archaeol.* **37**, 67–84 (2015).
- Raghavan, M. et al. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, 1–20 (2015).
- Skoglund, P. et al. Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015).
- Smith, S. et al. (2009) Inferring population continuity versus replacement with aDNA: A cautionary tale from the Aleutian Islands. *Hum. Biol.* **81**, 19–38.

Table S8.1. Aleutian ancient DNA shows no evidence of Papuan-related gene flow hypothesized by Raghavan *et al.* (2015) on the basis of present-day European-admixed Aleuts. The following D -statistics were calculated: $D(A, B; X, Y)$, where A=Yoruba or Dai; B=Papuan, Australian, or Onge; X=Mixe; Y=Neo-Aleut, Paleo-Aleut, Ancient Aleuts combined, or Surui. Z-scores are color-coded: $Z > 3$ in red, and $2 < Z < 3$ in yellow.

				pop Y							
				ancient Aleuts				ancient Aleuts			
				Neo-Aleut	Paleo-Aleut	combined	Surui	Neo-Aleut	Paleo-Aleut	combined	Surui
dataset	treatment	pop A, pop B	pop X	Z-scores				informative SNPs			
HumanOrigins (Lazaridis et al. 2014)	normal	Yoruba, Australian	Mixe	0.7	1.63	1.43	1.67	272,013	275,100	306,158	314,186
		Yoruba, Papuan		0.65	1.46	1.31	2.88	274,210	277,200	308,606	316,685
		Yoruba, Onge		0.85	1.32	1.41	3.87	273,679	276,757	308,026	316,118
		Dai, Australian		-1.34	-0.91	-1.21	1.08	262,660	265,765	295,150	302,843
		Dai, Papuan		-1.51	-1.28	-1.52	2.29	267,002	270,010	300,081	307,977
		Dai, Onge		-1.54	-1.78	-1.79	3.26	265,435	268,527	298,351	306,289
	no transitions	Yoruba, Australian		-0.18	0.76	0.06	0.85	50,109	51,071	56,733	58,428
		Yoruba, Papuan		-0.34	1.18	-0.03	2.11	50,487	51,441	57,158	58,866
		Yoruba, Onge		0.15	1.29	0.83	3	50,415	51,366	57,073	58,770
		Dai, Australian		-1.53	-1.29	-1.69	0.72	48,383	49,290	54,706	56,306
		Dai, Papuan		-1.74	-1.03	-1.93	2.05	49,161	50,067	55,601	57,250
		Dai, Onge		-1.43	-1.11	-1.24	3.08	48,864	49,762	55,263	56,918
genomes (Mallick et al. 2016)	normal	Yoruba.DG, Australian.DG	2.01	2.68	2.37	3.65	433,909	405,920	494,182	506,885	
		Yoruba.DG, Papuan.DG	1.56	1.9	1.86	4.06	459,513	428,834	523,154	536,078	
		Yoruba.DG, Onge.DG	2	2.01	2.25	3.27	432,997	405,045	493,166	506,158	
		Dai.DG, Australian.DG	-1.28	-0.92	-1.24	2.13	442,560	413,584	503,073	516,854	
		Dai.DG, Papuan.DG	-1.96	-2.02	-2.09	2.68	460,213	429,322	523,311	536,979	
		Dai.DG, Onge.DG	-1.35	-1.65	-1.45	1.59	441,305	412,384	501,610	515,733	
	no transitions	Yoruba.DG, Australian.DG	0.8	1.3	0.95	1.62	84,512	79,527	97,011	101,213	
		Yoruba.DG, Papuan.DG	0.76	1.58	1.05	2.66	89,441	83,985	102,642	107,037	
		Yoruba.DG, Onge.DG	1.65	2.81	2.59	2.05	84,355	79,368	96,826	101,115	
		Dai.DG, Australian.DG	-1.75	-2.11	-2.2	1.06	86,131	80,979	98,740	103,131	
		Dai.DG, Papuan.DG	-1.94	-2.13	-2.36	2.18	89,599	84,130	102,768	107,298	
		Dai.DG, Onge.DG	-0.88	-0.56	-0.6	1.38	85,861	80,703	98,441	102,861	

Supplementary Information section 9

Demographic modelling with *qpGraph* and *rarecoal*

Exploring possible admixture graphs with *qpGraph*

To investigate the phylogenetic relationship between relevant populations for this study, we tested models which fit observed f_4 -statistics using autosomal markers present in the 1240K capture panel¹. The f_4 -statistic measures a correlation between allele frequency differences of two pairs of groups, and thus provides a test for consistency of the data with a tree-like population relationship². Given a graph topology, the algorithm implemented in the *qpGraph* program infers branch lengths and mixture proportions that minimize the difference between the observed and expected f_4 -statistics.

For the analysis, we used whole genome sequence data from the Simons Genome Diversity Project³ and added additional 35 genomes published by Raghavan et al.⁴. Genotype calls for the autosomal part of the 1240K panel were extracted, and markers with > 10% missing rate were removed, leaving 1,150,639 SNPs for the analysis.

We first performed a comprehensive search for tree topologies fitting the data. For this, we selected the following populations: Mbuti, French, Ami, Mixe, Even, Yupik Naukan, Koryak and Chipewyan to represent each of the 7 meta-populations modelled in *rarecoal* further below (AFR, EUR, SEA, SAM, SIB, E-A, C-K and ATH, see Table S9.1). To perform an extensive search for possible population relationships, we began with a simple tree of three populations (Mbuti, (French, Ami)) and iteratively added one population to the tree. More specifically, the added population was modeled either as a sister branch of an existing one or as a mixture of two branches. We tested all branches and branch pairs. A total of 2,932 models were tested this way, and at the end of our search, we found 14 graphs that fit all observed f_4 -statistics within two standard error intervals (Fig. S9.1).

Although this extensive search resulted in a large number of well-fitted graphs with different topologies, many graphs share several key features. The most important feature is that Mixe, Even, Koryak, Chipewyan and Yupik Naukan are all modeled as a mixture of western and eastern Eurasian branches, and none of them forms a sister branch with each other: i.e. at least one additional gene flow is required to add each population to the graph (Fig. S9.1). For example, Koryak cannot be a sister group to Even because of its excessive affinity to Mixe. Also, Chipewyan cannot be modeled as a sister group of Mixe and requires a gene flow from a Siberian source, e.g. either Koryak- or Even-related branch. Finally, Yupik Naukan is well modeled as a mixture of Koryak- and Chipewyan-related branches, or as a mixture of Koryak- and Mixe-related ones (Fig. S9.1). The two models with the best fit $|(f_{4,observed} - f_{4,expected})| = 1.295 SE$ are very similar (Fig. S9.1a-b). First, Mixe receives two gene flows from the French-related branch, one of which is shared with the Siberian clade. Second, Koryak does not share the second gene flow with Mixe. Last, the branching order of the Arctic groups under this model is (PPE_{ATH}, (PPE_{C-K}, PPE_{E-A})). The abbreviations PPE_{ATH}, PPE_{C-K}, and PPE_{E-A} denote the sources of proto-Paleo-Eskimo-related ancestry in Chukotko-Kamchatkan, Athabaskan, and Eskimo-Aleut speakers, respectively. The only difference between those two models is the placement of Yupik, whose Native American component can be modelled as branching off the Chipewyan branch either before (Fig. S9.1a) or after (Fig. S9.1b) the PPE admixture in Chipewyans.

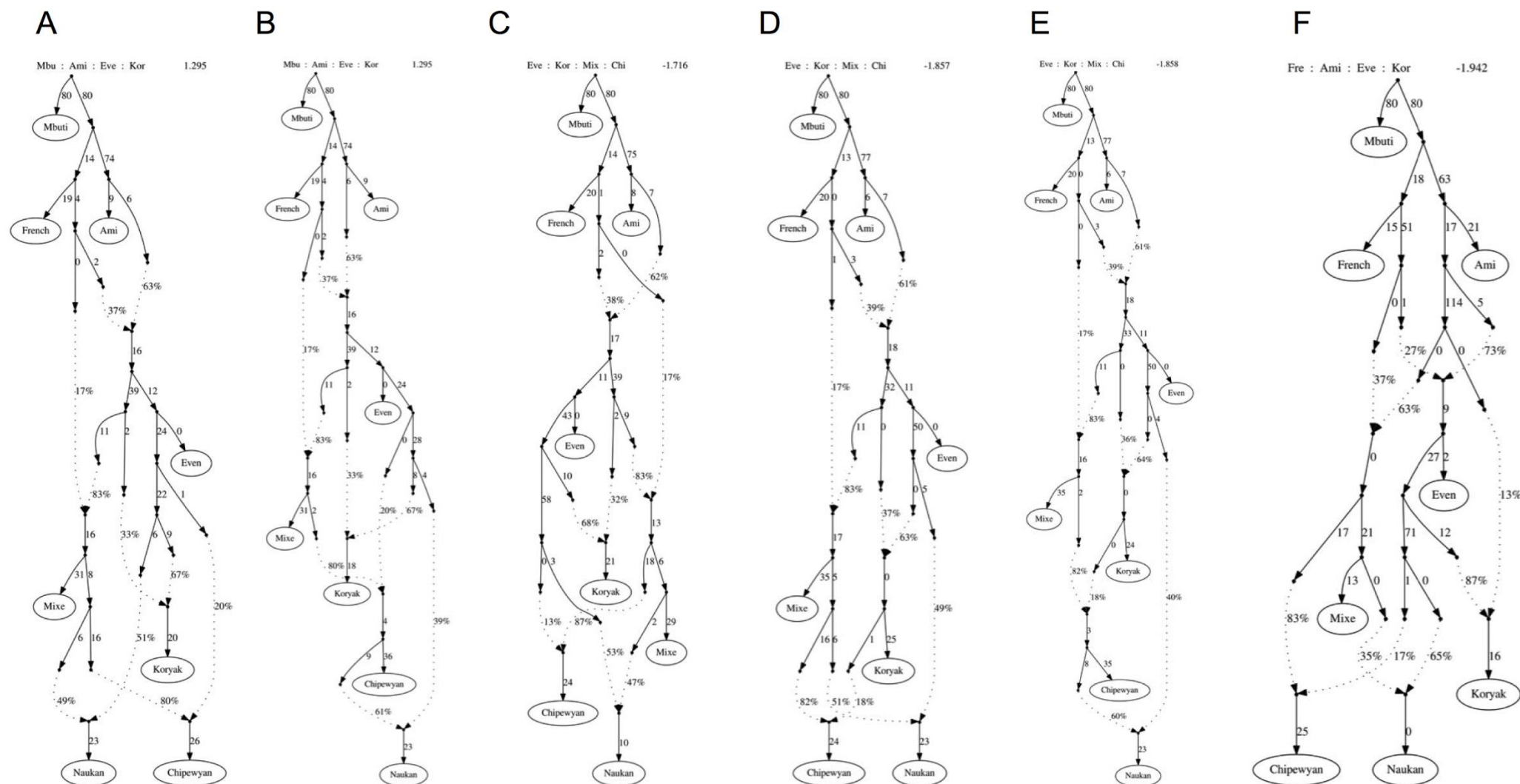


Figure S9.1. Representative models from an extensive search of population graphs. In most cases, Mixe receives two gene flows from a French-related branch (ANE). Yupik Naukan is modeled as a mixture of proto-Paleo-Eskimos (PPE) and a Chipewyan-related branch, either before (a, c, d, f) or after (b, e) the PPE gene flow into Chipewyan. Chipewyan is modeled as a mixture of a Native American lineage and either PPE (a-c, f) or Koryak (d, e).

Demographic modelling with *rarecoal*

Rarecoal is a software that implements a fast algorithm to estimate the joint site frequency spectrum for rare alleles⁵. Since the initial report we have improved the software and added pulse-like admixture events and more explicit treatment of missing data in single samples as new features. The updated mathematical derivations of the model are included as a PDF document in the repository: <https://github.com/stschiff/rarecoal>. We also built in a regularization for population size changes, which penalizes large changes of the population size and helps to avoid overfitting.

Data

In the following analysis, we will use the abbreviations for meta-populations and samples shown in Table S9.1.

Fitting a model connecting Europeans, Southeast Asians and First Americans

We started off with fitting a model to three populations only, and then added one population at a time, re-estimating all previous and new parameters. In all fits, we used the “rarecoal mcmc” program to estimate maximum likelihood values for each parameter. We restricted analysis to variants of maximum allele count 4. We first fitted a simple three population tree with a topology (EUR, (SEA, SAM)) (Fig. S9.2).

Abb.	Full Name	Populations ¹⁾	Nr of samples
EUR	Europeans	Basque, Bergamo, Bulgarian, Crete, Czech, English, Estonian, French, Greek, Hungarian, Norwegian, Orcadian, Polish, Sardinian, Spanish, Tuscan	33
SEA	Southeast Asians	Ami, Atayal, Burmese, Cambodian, Dai, Kinh, Lahu, Miao, She, Thai	22
SIB	Core Siberians	Nivkh, Altaian, Buryat, Even, Ket, Mansi, Tubalar, Ulchi, Yakut	22
C-K	Chukotko-Kamchatkan speakers	Chukchi, Itelmen, Koryak	4
E-A	Eskimo-Aleut speakers	Aleut, East and West Greenlandic Inuit	7
ATH	Northern Athabaskan speakers	Dakelh, Chipewyan	4
SAM	Central/South Americans	Aymara, Mixe, Mixtec, Piapoco, Quechua, Yukpa, Zapotec	14

Table S9.1. A table listing all present-day samples and groups used in the *Rarecoal* analysis. Data is from the two sources: Raghavan *et al.*⁴ and the Simons Genome Diversity Project data set³ as indicated in Suppl. Table 2.

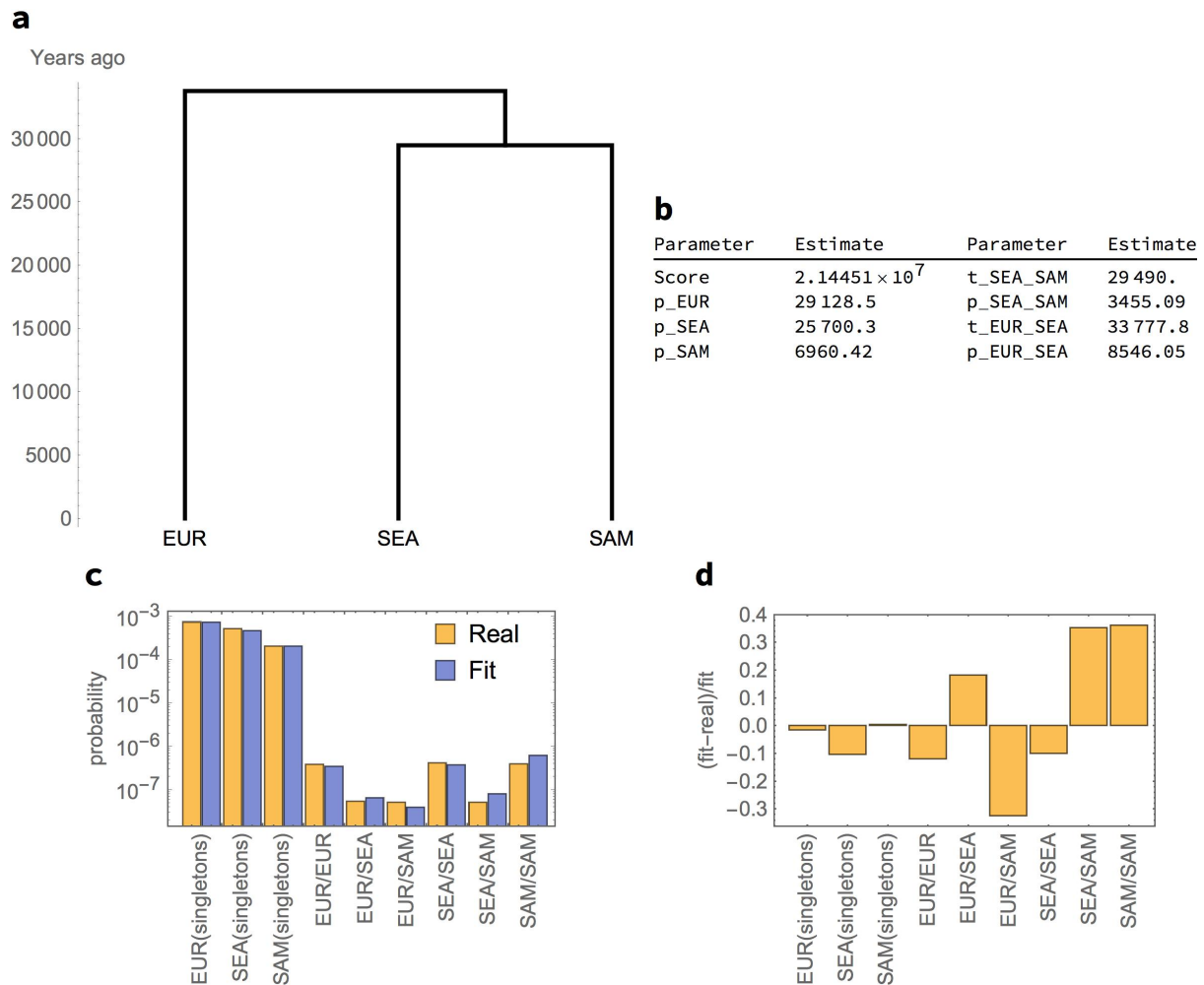


Figure S9.2. A model connecting Europeans, Southeast Asians and Native Central/South Americans.

Parameter names (Fig. S9.2b) starting with “p” denote population sizes, and those starting with “t” denote split times. By “Score” (Suppl. Fig 8.2b) we denote the negative log likelihood: the lower this number, the better the fit. The inferred population size and split time are scaled to real time and size using a mutation rate⁶ of 1.25×10^{-8} and a generation time of 29 years⁷.

Inspecting the fits between the model and the data (Fig. S9.2c) and the relative deviations between the two (Fig. S9.2d), we observe the largest underestimation of the true allele sharing ratios in variants shared between EUR and SAM. As we know from previous work⁸ and our *qpGraph* modeling (Fig. S9.1), Native Americans have ancient North Eurasian (ANE) admixture. We therefore added ANE as a separate (ghost) branch splitting off Europeans and added an admixture edge into the American branch. We also added an admixture edge at 250 years ago (ya) from EUR into SAM to reflect potential post-Columbian admixture and refitted all parameters.

Indeed, the resulting fit with ANE and European admixture (Fig. S9.3) is much better, with fit deviations being substantially smaller, around 10%. The inferred ANE admixture proportion is about 12% (parameters beginning with “adm” denote admixture proportions), which is substantially smaller than the estimate from *qpGraph* (around 40% overall), but this is plausibly due to the selection of European populations in the data set used here (Table S9.1), which includes populations with low ANE ancestry. Note that post-Columbian admixture is inferred to be quasi zero, so we removed that edge from further models.

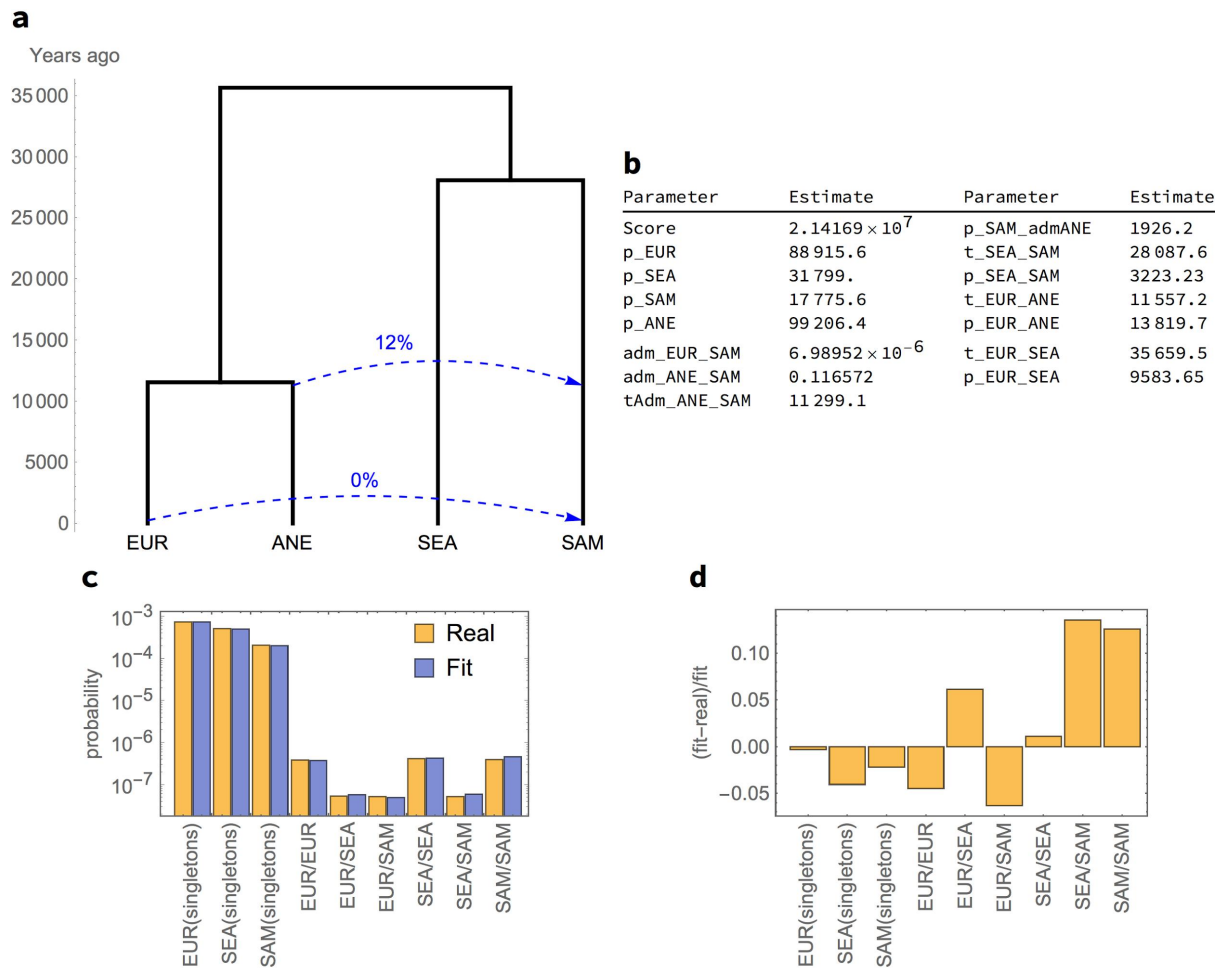


Figure S9.3. A model fit with post-Columbian admixture in Native Americans. Note that the ANE branch is a pure ghost branch, i.e. it is fully unsampled.

Adding Core Siberians onto the tree

We next added “Core Siberians”. As we know from the best-fitting *qpGraph* models (Extended Data Fig. 10a, Fig. S9.1), Native Americans and Siberians require two independent ANE admixture events, one into the ancestor of Native Americans and Siberians, and one separately into the Native American branch. We tested this model, adding also one more explicit admixture from EUR into SIB to reflect potentially more recent European admixture into Siberians as described previously⁹ (Fig. S9.4).

The largest fit deviation in this model (Fig. S9.4d) was an overestimation of the SEA/SAM allele sharing. We suspect this could be explained by SEA->SIB admixture, which would make it possible for the model to push back the SEA/SAM split time and hence reduce the allele sharing between these two populations. We revisit this point further below after adding another Siberian group (Chukotko-Kamchatkan speakers) with potentially less SEA admixture.

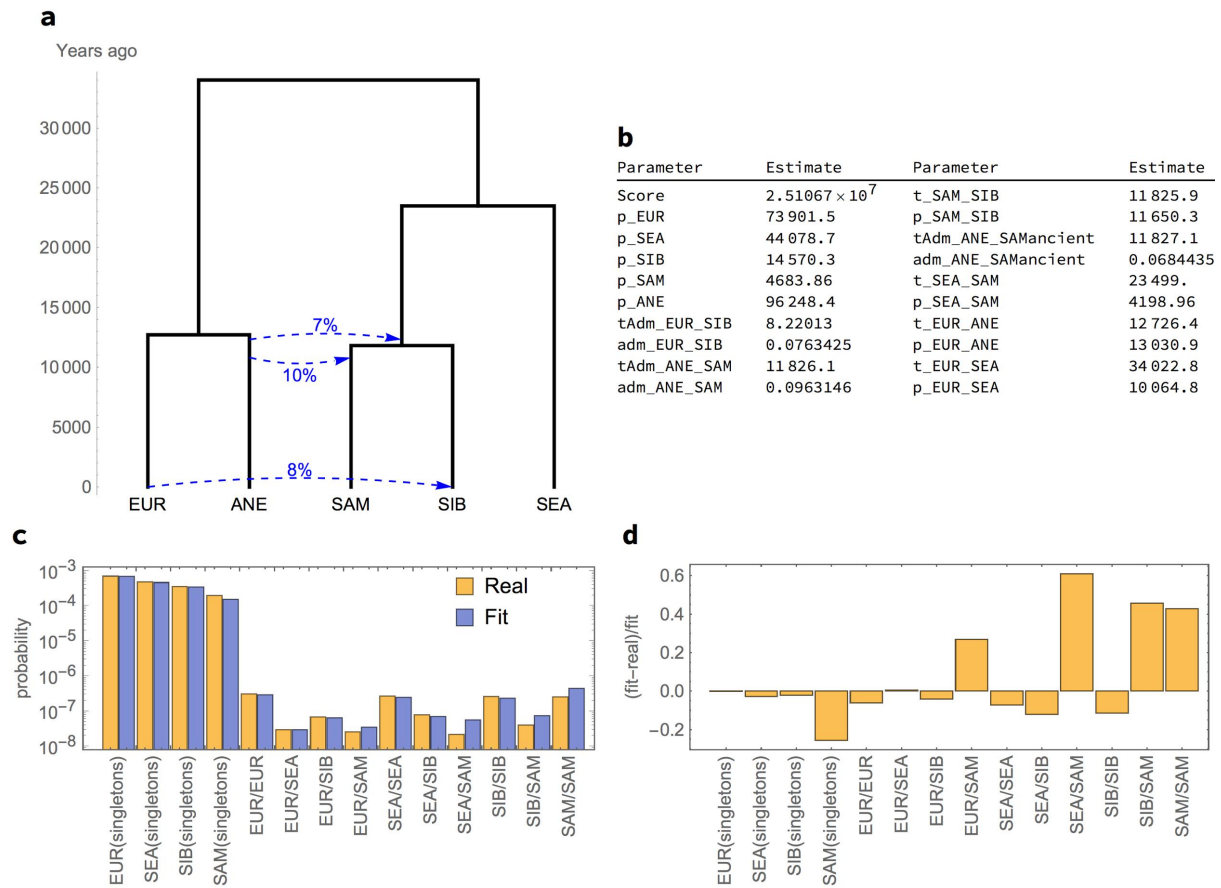


Figure S9.4. A model with Siberians added without further admixture onto the basal Central/South American branch.

Adding Chukotko-Kamchatkan speakers

Next, we added another Siberian meta-population, much closer to the Bering Strait, consisting of Chukotko-Kamchatkan-speaking populations (C-K). We started with a tree that adds the C-K meta-population onto the Siberian branch (Fig. S9.5).

As already seen in the previous model without C-K (Fig. S9.4), this model overestimates SEA/SAM and SIB/SAM sharing. This is likely caused by an affinity between C-K and SAM, as seen in the PCA (Fig. 2 and Extended Data Fig. 3) and *qpGraph* results (Extended Data Fig. 10a, Fig. S9.1), which both suggest that C-K are closer to SAM than expected under a simple tree. Indeed, the best-fitting *qpGraph* model (Extended Data Fig. 10a, Fig. S9.1) contains an admixture edge from the Asian ancestors of Native Americans (before ANE admixture) into the ancestors of C-K. We added this admixture edge as a ghost population splitting off the ancestors of SAM before ANE admixture and contributing to the ancestors of C-K right after their split from SIB. In addition, to mitigate the SEA/SAM overestimation, we suspect there might be SEA admixture in SIB, so we added an additional admixture edge there. The resulting model (Fig. S9.6) gives a better fit to the data and in particular removes the SIB/SAM overestimation in allele sharing and mitigates the SEA/SAM overestimation.

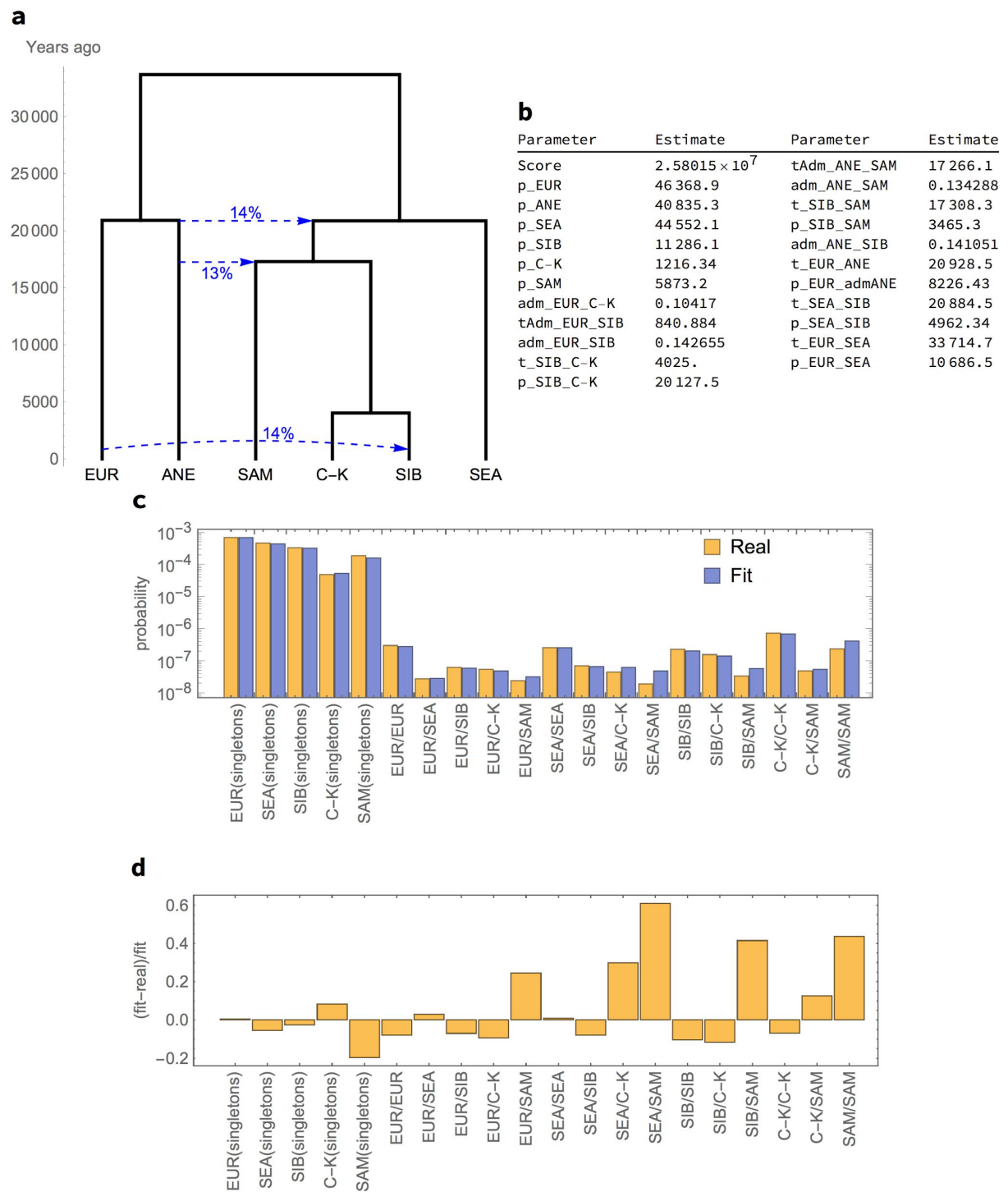


Figure S9.5. Adding Chukotko-Kamchatkan speakers onto the tree without further admixture. For clarity, post-colonial admixture from EUR into C-K is not shown.

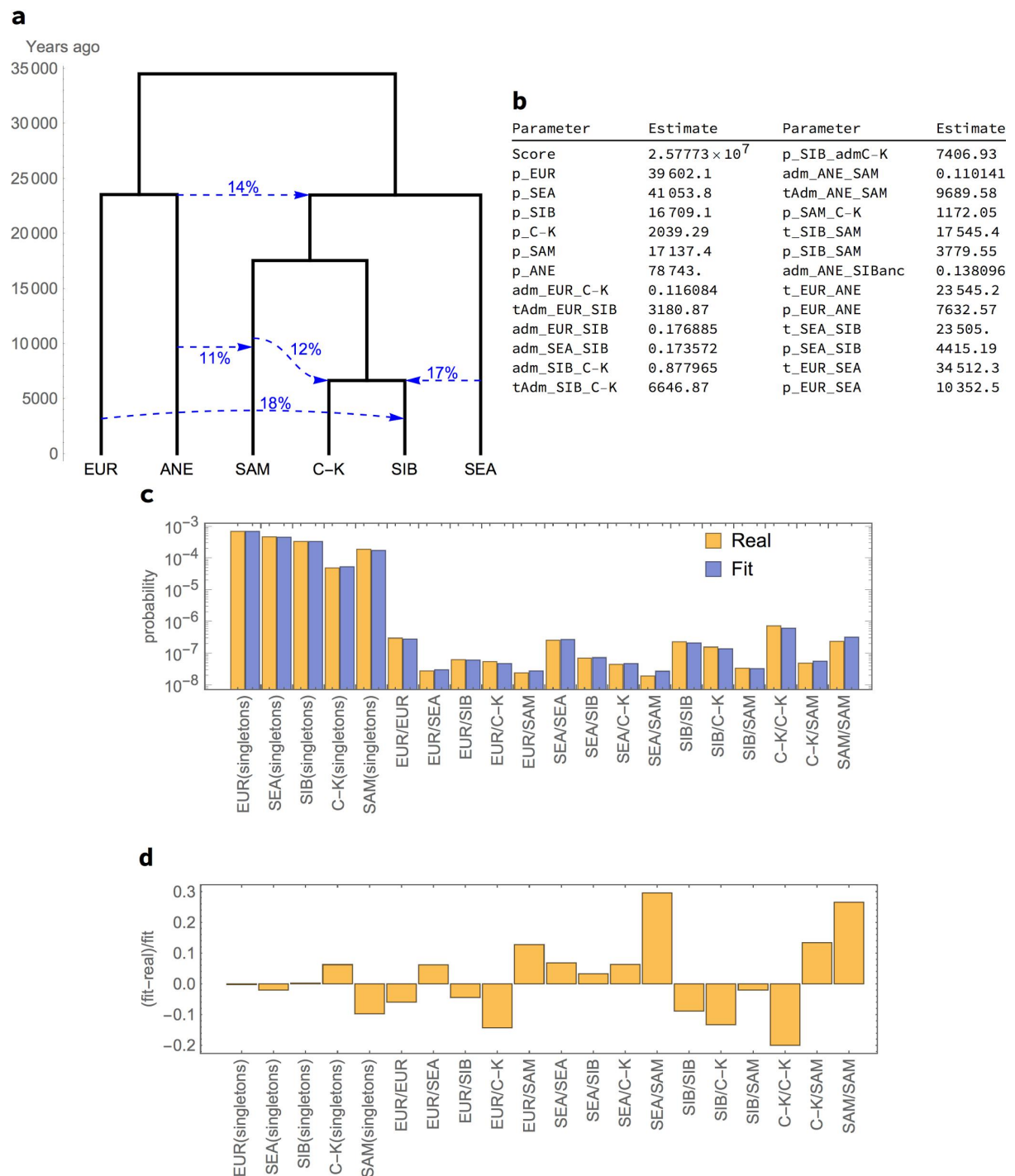


Figure S9.6. Adding admixture from SAM ancestors into C-K and from SEA into SIB.

Adding Athabaskan speakers

We next added Athabaskan speakers (ATH) onto the model as a sister clade to the SAM branch (Fig. S9.7).

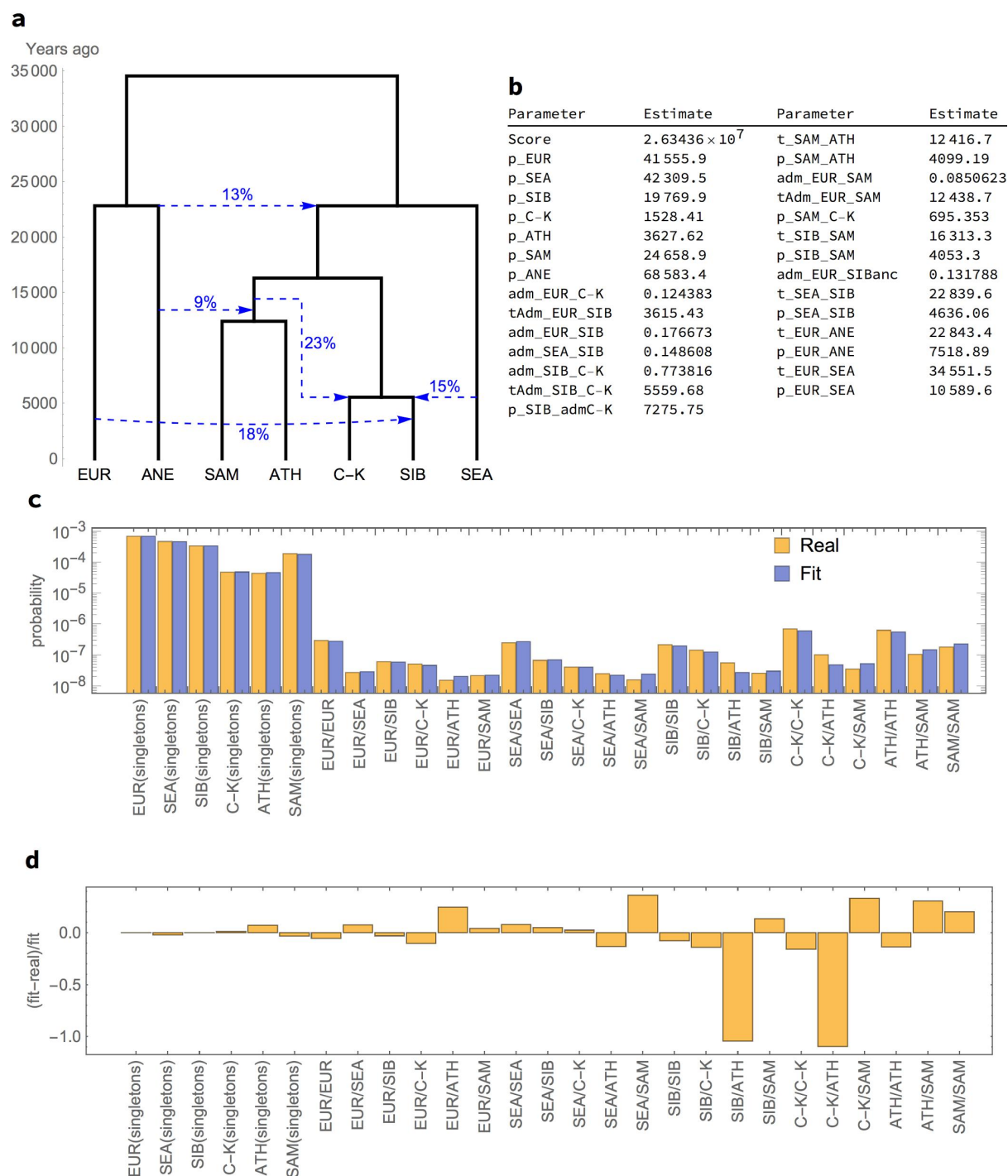


Figure S9.7. Adding ATH as a sister clade to SAM without further admixture. For clarity, colonial admixture in C-K is not shown.

The naïve model of adding ATH as a sister clade to SAM without further admixture (Fig. S9.7) resulted in a massive underestimation of ATH/SIB and ATH/C-K allele sharing, suggesting additional admixture in ATH. We therefore added an admixture edge from C-K into ATH with the admixture time as a free parameter (Fig. S9.8). Indeed, this admixture edge completely mitigated the underestimation of the ATH/SIB and ATH/C-K allele sharing.

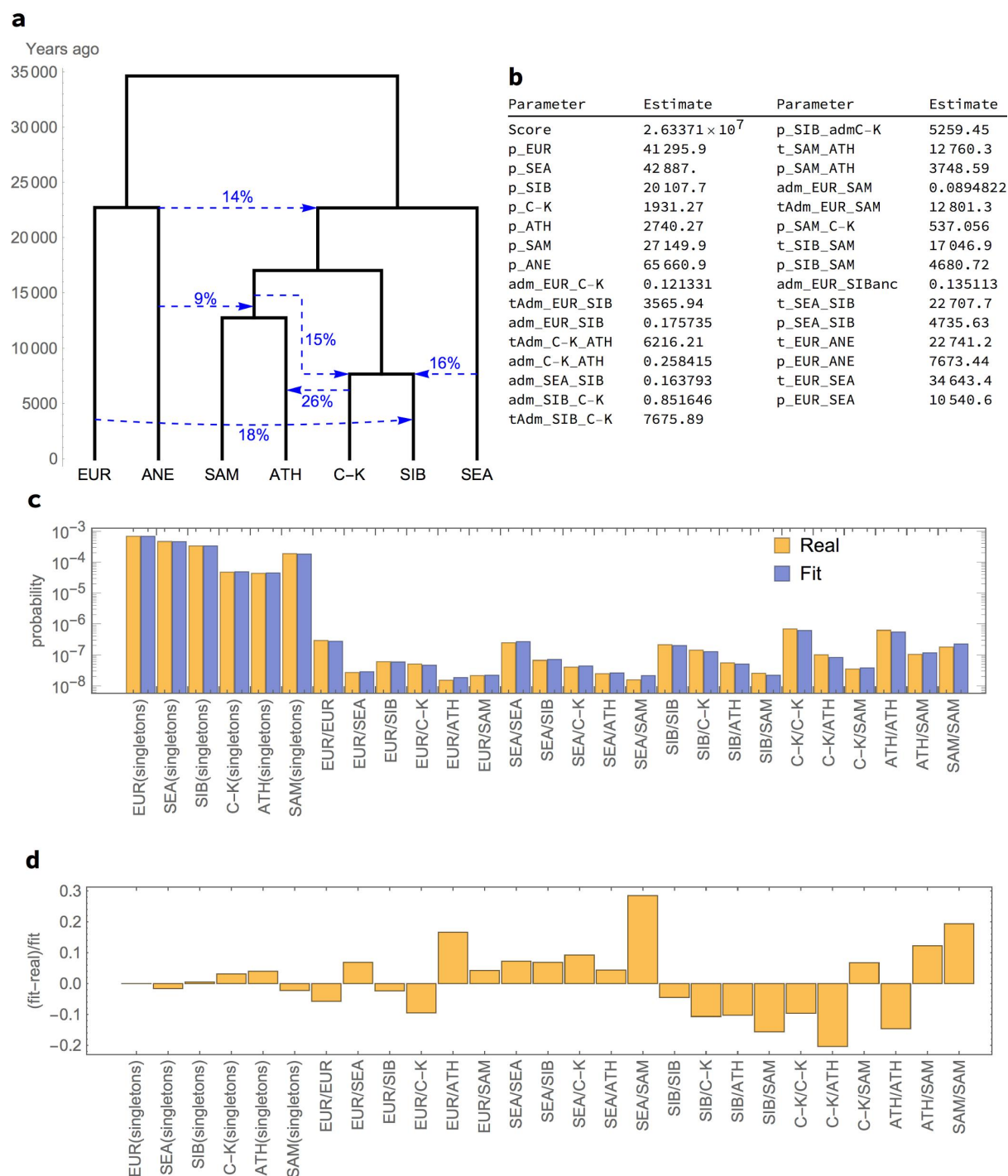


Figure S9.8. Adding C-K→ATH admixture. For clarity, colonial admixture in C-K is not shown.

Adding Eskimo-Aleut speakers to the model

We next added Eskimo-Aleut speakers (E-A). As we know from the *qpGraph* analysis (Extended Data Fig. 10a, Fig. S9.1), Eskimo-Aleut speakers can be modeled as admixed between a North American population and ancestors of Chukotko-Kamchatkan speakers. Furthermore, as known from the PCA, present-day E-A have substantial colonial European admixture. To model the northern North American ancestry in E-A, we therefore added a ghost contribution from Athabaskan ancestors into E-A, before the C-K admixture in ATH. We also added a 250ya European admixture (Fig. S9.9) into E-A to reflect colonial admixture, which was indeed inferred to be as high as 34%.

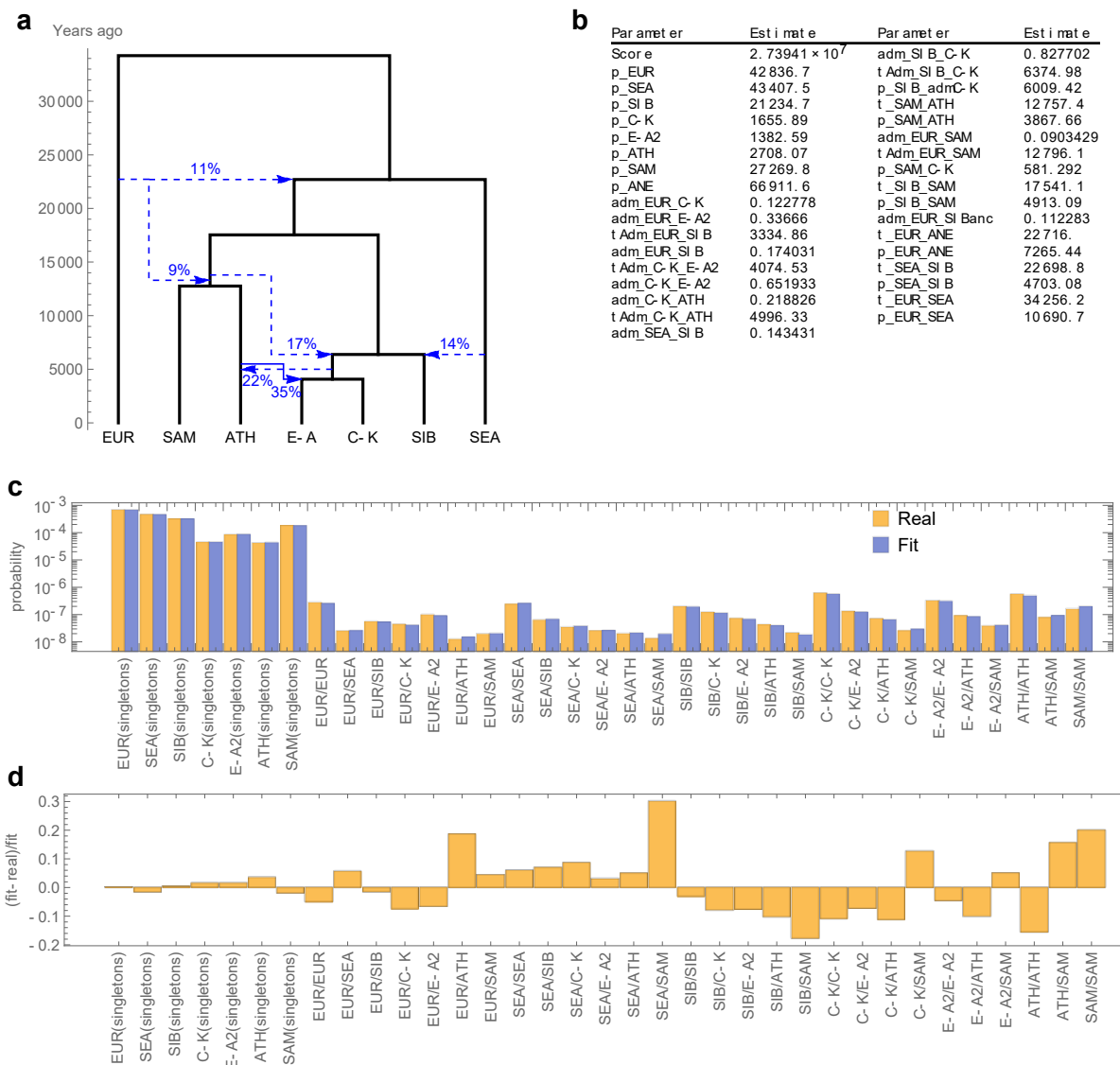


Figure S9.9. Adding Eskimo-Aleut speakers (E-A) as a sister group to C-K with ancestry from ATH ancestors and with colonial European admixture (not shown). For clarity, EUR→SIB admixture and colonial admixture in C-K and E-A are also not shown.

This is the final model that we also report in Fig. 5. The inferred parameters for this model are summarised in Table S9.2

	Parameter	Maximum Likelihood estimate
effective population sizes	EUR	43,000
	SEA	43,000
	SIB	21,000
	C-K	1,700
	E-A	1,400
	ATH	2,700
	SAM	27,000
	ANE	67,000

	EUR-SEA...	11,000
	SEA-SIB...	4,700
	SAM...-SIB...	4,900
	SAM-ATH	3,900
	SIB-C-K...	6,000
split times, ya	EUR-SEA...	34,000
	SEA-SIB...	23,000
	SAM...-SIB...	18,000
	SAM-ATH	13,000
	SIB-C-K...	6,400
	C-K-E-A	4,100
admixture proportions and dates in ya	EUR -> SIB	17%
	date EUR -> SIB	3,300
	EUR -> E-A	34%
	EUR -> C-K	13%
	ANE -> SAM	9%
	ANE -> SIB...	11%
	date C-K... -> ATH	5,000
	ATH -> C-K...	17%
	C-K -> ATH	22%
	ATH -> E-A	35%
	SEA -> SIB	14%

Table S9.2. Maximum likelihood estimates for all parameters of the final model (Fig. S9.9).

To test the robustness of the admixture edge into Athabaskans, we explored two alternative models, one with the admixture into Athabaskans originating in E-A (Fig. S9.10), and one where the admixture originates on the C-K branch *after the split* from E-A (Fig. S9.11).

In both cases, the fit is worse, as seen by the higher “Score” (negative log likelihood). In case of the model with admixture from E-A (Fig. S9.10), we also see that the SIB/ATH allele sharing is underestimated substantially. In case of the model with more recent C-K admixture into ATH, the model optimization puts the time of admixture close to the C-K/E-A split time, which results in a relatively similar fit as our best model (Fig. S9.9), but the likelihood is lower (higher Score).

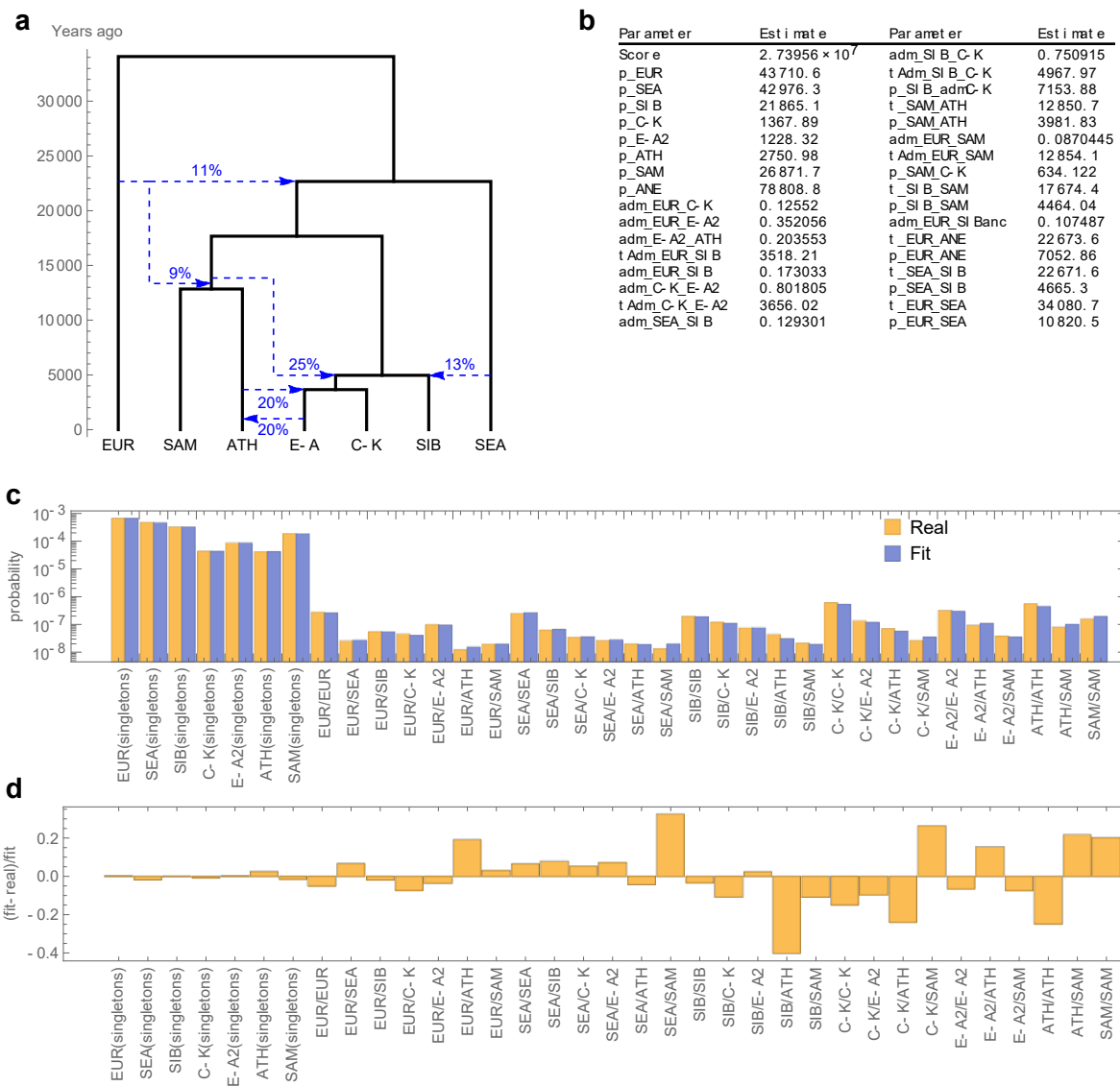


Figure S9.10. Alternative model which explains Siberian ancestry in Athabaskans by admixture from Eskimo-Aleut speakers.

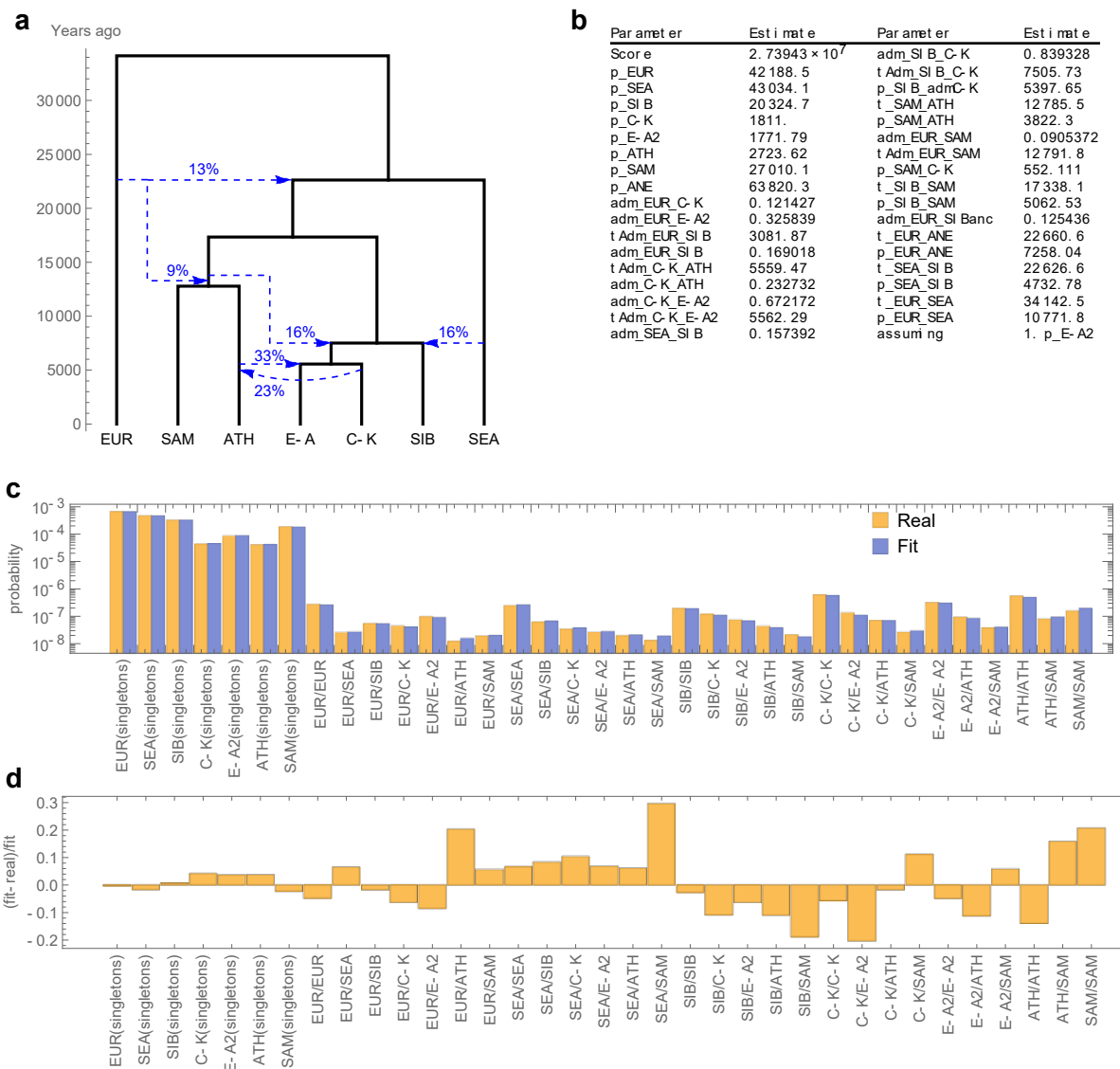


Figure S9.11. Alternative model which explains Siberian ancestry in Athabaskans by recent admixture from Chukotko-Kamchatkan speakers.

Testing the final model with *qpGraph*

We tested the final topology found by our *rarecoal* analysis (Fig. S9.9) to see if it fits the observed f_4 -statistics. Using the same meta-populations as in the *rarecoal* analysis (Table S9.1), we confirmed that the *rarecoal* topology fits the observed f_4 -statistics well ($\max |(f_{4,observed} - f_{4,expected})| = 0.883 SE$) (Fig. S9.12). Mixture proportion estimates are non-zero for all edges, with broadly similar weights as those of the *rarecoal* model (Fig. S9.9), also suggesting that they are necessary to explain the relationship between the meta-populations.

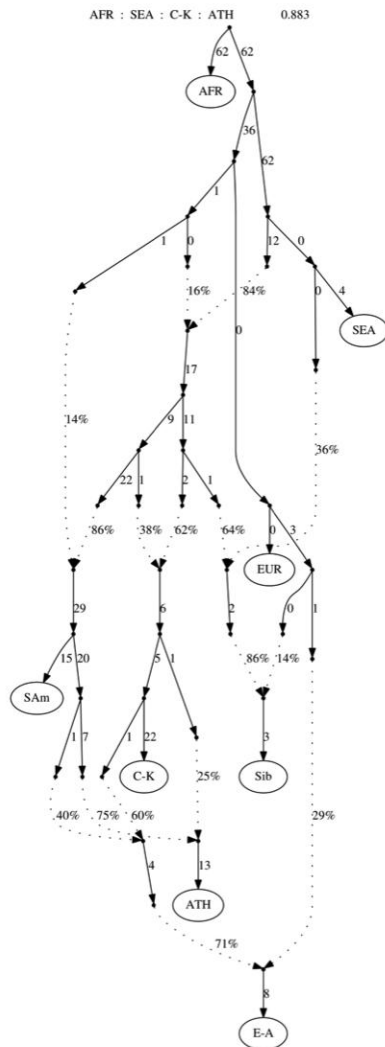


Figure S9.12. The *qpGraph*-based estimation of the *rarecoal* topology. All 7 meta-populations fit the graph without large difference between estimated and observed f_4 -statistics. The most extreme deviation in Z-score (0.883) is presented at the top. All admixture edges are inferred to have non-zero weight, suggesting that they are required to explain data.

We also tested if the *qpGraph* approach can provide a fine-scale resolution for population relationships around the Bering Strait. For this, we tested various topologies with minor modifications from that in Fig. S9.12. Specifically, we asked the following two questions: i) did the First American population contributing to E-A branch before or after the PPE contribution into the ATH lineage, as discussed in the main text? and ii) what is the branching order of PPE-related ancestry in C-K, ATH and E-A (denoted by PPE_{C-K} , PPE_{ATH} and PPE_{E-A})? All six topologies (Fig. S9.13) that we tested provided a good fit to the data ($\max |(f_{4,observed} - f_{4,expected})| = 1.362 SE$), suggesting that this part of the model is not resolved in fine detail. However, models assuming the branching orders of either (PPE_{E-A} , (PPE_{ATH} , PPE_{C-K})) or (PPE_{C-K} , (PPE_{ATH} , PPE_{E-A})) resulted in the shared branch between PPE_{ATH} and PPE_{C-K} having zero length, effectively

modeling a trifurcation (Fig. S8.13c,e,f). Therefore, the branching order (PPE_{ATH} , (PPE_{C-K} , PPE_{E-A})) is most likely.

References (for this section)

1. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
2. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
3. Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
4. Raghavan, M. *et al.* POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884–aab3884 (2015).
5. Schiffels, S. *et al.* Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nat. Commun.* **7**, 10408 (2016).
6. Scally, A. & Durbin, R. Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* **13**, 745–753 (2012).
7. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423 (2005).
8. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
9. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).

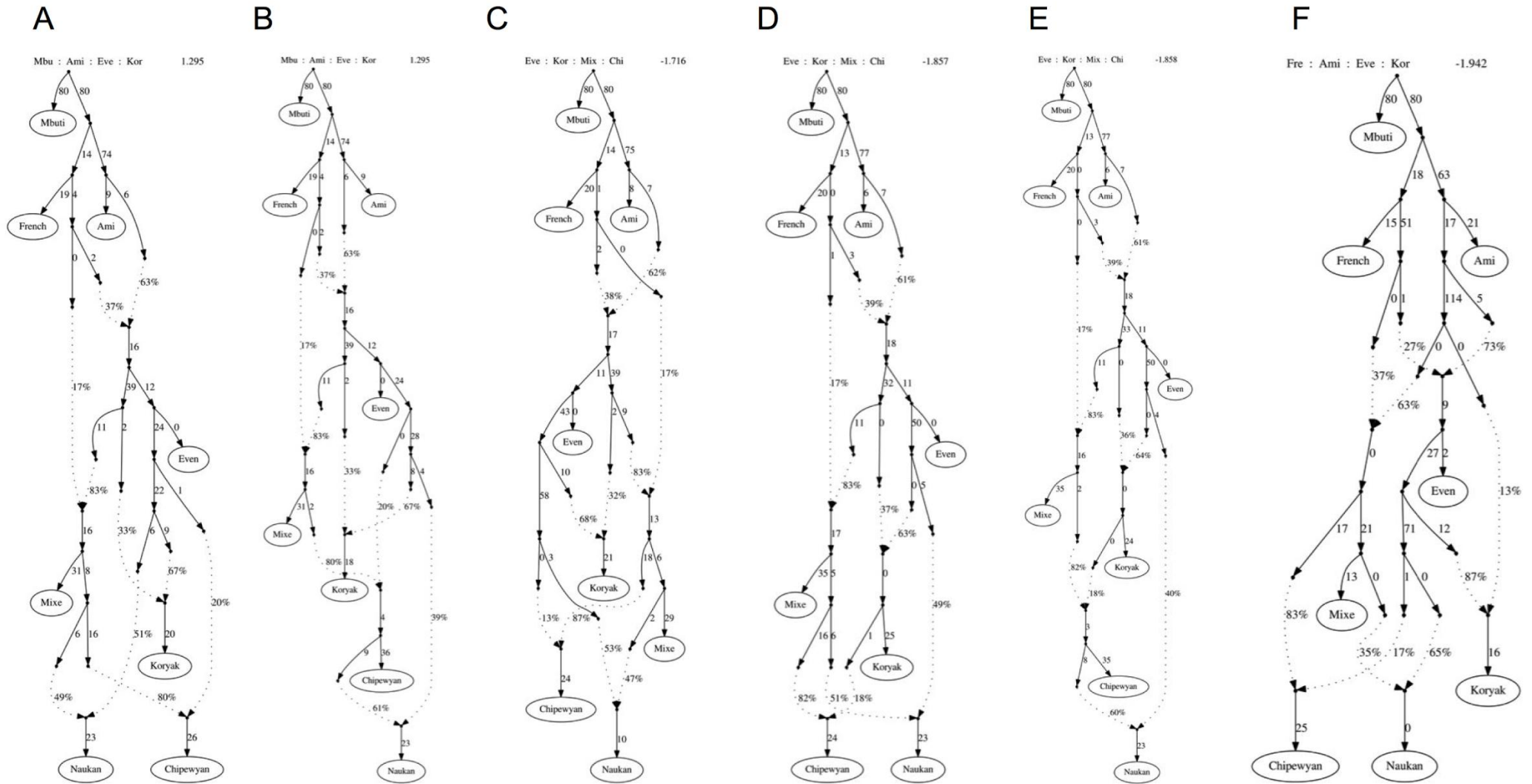


Figure S9.13. Testing additional topologies regarding the relationships of SAM, ATH, C-K and E-A. **a, c, e**, ATH and E-A share a gene flow from the PPE-related source; (**B, D, F**) ATH and E-A do not share the PPE-related gene flow; **a, b**, topology (PPE_{ATH}, (PPE_{C-K}, PPE_{E-A})); **c, d**, topology (PPE_{C-K}, (PPE_{ATH}, PPE_{E-A})); **e, f**, topology (PPE_{E-A}, (PPE_{C-K}, PPE_{ATH})). The most extreme deviation in Z-score is presented at the top.

Supplementary Information section 10

Overview of the Dene-Yeniseian linguistic hypothesis

by Edward J. Vajda

The Dene-Yeniseian language hypothesis is considered here in light of the demonstrated Paleo-Eskimo genetic admixture in Tlingit, Eyak and Athabaskan speakers dated to ~5,000 calBP and shared more distantly with Siberians at a time depth of ~6,500 calBP (Table S9.2). The timing of this genetic link and its likely archaeological parallels provides the first evidence from beyond linguistics that realistically supports the Dene-Yeniseian language hypothesis. Given that Paleo-Eskimo-related ancestry is likewise shared by speakers of languages belonging to the Eskimo-Aleut and Chukotko-Kamchatkan families, the Paleo-Eskimo genetic legacy could instead be associated with the linguistic origins of either of these families rather than with Dene-Yeniseian. However, because the accumulated Dene-Yeniseian and internal Na-Dene comparative linguistic evidence correlates so plausibly with the coalescence dates of the Paleo-Eskimo genetic loci shared by populations speaking precisely these languages, it is useful to elaborate further on the potential significance of these results for locating the Dene-Yeniseian language family in space and time – questions left without clear answers in Kari and Potter (2010).

The Dene-Yeniseian hypothesis claims that the Ket language spoken near the Yenisei River in a remote area of Central Siberia is related to the widespread Na-Dene language family in North America. Na-Dene comprises Tlingit and the recently extinct Eyak in Alaska, along with over thirty Athabaskan languages spoken from the western North American Subarctic to pockets in California (Hupa), Oregon (Tolowa) and the American Southwest (Navajo, Apache) (Krauss 1976). The severely endangered Ket is the sole survivor of Siberia's once widespread Yeniseian language family, whose ancient presence in the region predates the expansion of reindeer breeders and other pastoralists in North and Inner Asia (Dul'zon 1959, 1962, Vajda 2001, 2009, Werner 2005). Dene-Yeniseian as a linguistic hypothesis dates back to at least 1923, when Italian linguist Alfredo Trombetti linked Athabaskan and Tlingit with Ket on the basis of a few similar-sounding words (Trombetti 1923). In the past two decades new evidence supporting the connection has been published in the form of shared morphological systems and lexical cognates showing interlocking sound correspondences (Ruhlen 1998, Vajda 2001, Werner 2004, Vajda 2010a, 2010b). However, Dene-Yeniseian cannot be accepted as a proven language family until the evidence of lexical and morphological correspondences between Yeniseian and Na-Dene is significantly expanded and tested by further critical analysis. It will also be essential to determine the potential relationship between Yeniseian and Old World languages and families such as Sino-Tibetan, North Caucasian, and the Burushaski isolate of northern Pakistan – all of which have been proposed at various times in the past as relatives of Yeniseian, and sometimes also of Na-Dene (G. Starostin 2010). While parallel research from genetics, archaeology and folklore studies cannot prove a language connection (only comparative linguistic analysis can accomplish that), interdisciplinary studies of human prehistory can demonstrate in important ways the plausibility or implausibility of such a connection.

The timing of the Dene-Yeniseian language split could shed important light on Native American as well as North Asian prehistory. In attempting to reconcile the apparent closeness of Yeniseian and Na-Dene grammatical homologies with the much greater genetic distance between Ket and Na-Dene speakers, the various papers in Kari & Potter (2010) offered three possible scenarios for the Dene-Yeniseian connection: 1) a Late Pleistocene separation connected with the Paleo-Indian migrations into the Americas, with an extraordinary slow rate of linguistic change; 2) a separation involving a back migration of Yeniseians from Beringia; and 3) an Early to Mid-Holocene separation connected with the entrance into Alaska around 5,000 calBP by the population that later developed the Arctic Small Tool tradition (ASTt). The first two scenarios can now be excluded (see below).

In contrast to the ability of archaeologists to carbon-date their finds, or geneticists to calibrate the time separating two related populations, there is no universally accepted method to reliably and precisely compute the time of separation of languages known to be genealogically related. All proposed methods of dating prehistoric language splits have been criticized (Campbell 2013:447-492). McMahon & McMahon (2005: 177-204) distinguish between methods of

establishing relatedness or degrees of relatedness between languages (lexicostatistics) from the use of such data to assign precise dates for prehistoric language splits based on an assumed regular rate of linguistic change (glottochronology), which in fact does not exist across languages or even in a single language over time. While rejecting glottochronology, McMahon & McMahon (2005:204) support the value of gathering and comparing lexicostatistic data, which then can sometimes be useful for purposes of dating when combined with facts from other disciplines such as archaeology and genetics. Several types of evidence can potentially be combined with evidence of shared vocabulary and grammatical homologies to help narrow the range of plausible separation dates between related languages. For Dene-Yeniseian, all of them suggest a split roughly between 10,000 and 8,000±500 calBP. The shallower end is favored by the detailed morphological homologies shared by the two families (Nichols 2010). The deeper end, which is suggested by the more meager number of shared lexical cognates, would still be far too shallow to match a connection with the earliest Paleo-Indian migrations during the Late Pleistocene. However, this range does provide a realistic temporal parallel for the migration of ASTt ancestors from North Asia into the Americas about 5,000 calBP. If this population consisted of Pre-Proto-Na-Dene speakers, then the split with their Yeniseian-speaking cousins in south-central Siberia would necessarily have been earlier.

Most previous calculations by historical linguists place the timeline for the internal diversification of Na-Dene languages between 6,000 and 3,500 calBP. The Na-Dene family contains the widespread Athabaskan (Dene) languages, which together are equally related to the recently extinct Eyak language of coastal Alaska. All Athabaskan languages, whether spoken in Alaska, Canada, California, or Arizona, share over 70% cognates in basic vocabulary, the number becoming higher if the list includes words associated with northern boreal lifestyle, such as 'birch', 'wolverine', etc. Krauss (1976:330) showed that all Athabaskan languages share 33% of basic vocabulary from the 100-word Swadesh List with Eyak. Athabaskan-Eyak, in turn, is clearly more distantly related to the Tlingit dialect cluster spoken in the Alaskan Panhandle and parts of interior Yukon Territory (Heggarty & Renfrew 2014:1236). Using a variety of lexicostatistic methods and reliable data, Krauss (1976:333) estimated a time depth for Proto-Athabaskan of 2,400±500 years and for Athabaskan-Eyak of 3,400±500 years. Estimates for the earlier breakup of Tlingit and Athabaskan-Eyak range from 6,000 (Mülenbernd & Rama 2017) or 5,000 years (Swadesh 1958) to as shallow as 3,500 years (Kaufman & Golla 2000), with an estimate of 4,500 years by Krauss (1980:11-13). The deeper dates would be favored by the known conservatism of Na-Dene languages and also by the fact that the phylogenetic relationship between Athabaskan-Eyak-Tlingit (Na-Dene) was universally accepted only in the past decade, despite being suspected for over a century (Campbell 2011). The late acceptance date derives mainly from the fact that before Leer (2010), the evidence for Athabaskan-Eyak-Tlingit in the form of shared finite verb structure significantly outweighed the expected parallel lexical evidence, making it unclear whether language mixing rather than genetic inheritance was involved in the historical similarities between these languages.

The relatedness between Athabaskan languages, despite their far-flung geography, is close enough that it has never been in doubt (Campbell 1997), though no subgrouping beyond the geographic one separating Pacific Coast (Hupa, Tolowa, etc.) and Southwestern Athabaskan (Navajo and Apache) from Northern Athabaskan (the remaining languages in Canada and Alaska) has yet been demonstrated. This suggests a rapid spread from a common source, most likely somewhere in Northwestern Canada near the current border between British Columbia and Alaska or in adjacent parts of Interior Alaska. Another support for a recent dispersal is the high rate of mutual intelligibility between geographically distant Athabaskan languages (Krauss 1976). Some scholars posit a time depth for Proto-Athabaskan as shallow as 2,000 calBP (Kaufman & Golla 2000), though a date closer to 3,000 is more likely given the resistance to borrowing observed with all of these languages. A time depth of at least 2,500 years for Athabaskan, following the estimate in Krauss (1976), would concur with the westward spread of the Taltheilei Culture beginning 2,750 calBP, which has been associated with the spread of Athabaskan speakers (Potter 2010, Kari 2010).

The interior Alaskan and northwestern Canadian portions of the Athabaskan range show no clear archaeological evidence of prehistoric population replacement during the past 8000 years (Potter 2010, Kari 2010). For this reason, Kari (2010) posits that the Athabaskans have lived in

interior northwestern North America for at least that span of time. Kari cites the near complete absence of substrate place names in the Northern Athabaskan areas as evidence for their ancient occupation of these areas. However, the Navajo and Apache areas of the American Southwest likewise have virtually no toponymic substrate from the languages previously spoken there, yet the Athabaskan presences there dates no farther back than 1,200 calBP. This reflects a strong Athabaskan avoidance of borrowing place names rather than ancient occupancy. In any event, such a degree of linguistic conservatism, whereby geographically distant languages maintain mutual intelligibility over a span of 8,000 years, would be unique and unprecedented. After adjusting for the conservatism of Na-Dene languages, retention rates for vocabulary and grammatical structures would appear to support a time depth of 5,000±500 years for the ancestral Athabaskan-Eyak-Tlingit language (i.e., Proto-Na-Dene). This concurs well with the possibility that the language ancestral to Na-Dene could have been introduced around 5,000 calBP into Alaska by North Asian immigrants associated with the later development and spread of the ASTt. Also probably connected with these “Paleo-Eskimos” is the spread of other elements of North Asian material culture and folklore (Alekseenko 1995; Berezkin 2015) to the Na-Dene, including bow and arrow technology, thought to have been introduced into California 1,500 years ago by the ancestors of the Hupa and other Pacific Coast Athabaskans (Golla 2011:245).

Like the Athabaskan family, Yeniseian languages are obviously related genealogically. Ket and its now extinct relatives (Yugh, Kott, Assan, Arin, and Pumpokol) were recognized as closely related more than 150 years ago (Vajda 2001). Studies of substrate toponyms (Vajda, in press) show that the known Yeniseian daughter branches (excepting the Ket-Yugh sub-branch) had already diversified by 2,000 calBP, when Turkic and Uralic-speaking pastoralists started displacing them in most of their southern and western territory, acquiring Ket-related river names and other substrate linguistic elements in the process. If the main sub-branching existed 2,000 years ago, the family is clearly older. The high rate of shared cognates in basic vocabulary (over 70%) between Ket and Kott, which belong to different primary branches of the family, suggest that Proto-Yeniseian must be at least 2,500 to 3,000 years, if not older, which would roughly match the more plausible estimates of time depth for Athabaskan. It is possible to reconstruct Proto-Yeniseian vocabulary (Starostin 1995) and many aspects of grammatical structure (Vajda 2013; Vajda 2017) with a high degree of confidence. If Para-Yeniseian linguistic relatives once existed in other parts of North Asia, the influx of pastoral tribes from the south must have obliterated them during the past 3,000 years, leaving no observable traces. Taking into account the probability of language extinction, the breakup of the earliest Proto-Yeniseian language, one predating the form reconstructable on the basis of Ket and Kott, could conceivably have begun much earlier than 3,000 calBP.

All Na-Dene languages share innovations demonstrating their equidistance from Yeniseian, whose split from the language ancestral to Na-Dene must be significantly older than Proto-Na-Dene itself. To cite one particularly vivid example, Pre-Proto-Na-Dene restructured three of its inherited Dene-Yeniseian verb prefixes into the so-called classifier complex, for which the family is well known. All three component prefixes have cognates in Yeniseian but did not develop the characteristic function of transitivity increase and decrease found of all Na-Dene languages (Vajda 2017). Contrary to Holton and Sicoli (2014), there is no linguistic or genetic evidence indicating a back migration into Asia of Yeniseian speakers from Beringia after Na-Dene had already begun to diversify.

The evidence supporting Dene-Yeniseian so far appears asymmetrically stronger in the realm of shared morphology than in the lexicon (Nichols 2010). The number and specificity of homologies in verb structure on their own would seem to preclude a separation earlier than the Mid-Holocene. Given the low number of lexical cognates, the time depth of Dene-Yeniseian may be twice that of Na-Dene. So far, the number of proposed Dene-Yeniseian cognates, even if all of them are valid, is less than half the number shared between Tlingit and Athabaskan-Eyak. If the Dene-Yeniseian linguistic link is fully demonstrable, however, substantially more abundant evidence of lexical cognates should be expected to emerge as the sound correspondences shared between the two families are fully worked out, favoring a shallower time depth range in line with the morphological evidence. This would repeat the historiography of Athabaskan-Eyak-Tlingit comparative linguistic studies, whereby the family’s striking parallels in verb

morphology were successfully identified well in advance of the accumulation of a large enough body of cognates in basic vocabulary to support a full range of systematic sound correspondences between Tlingit and Athabaskan-Eyak and fully demonstrate the Na-Dene family.

Though linguistic science can only rarely offer precise dates for prehistoric language splits, few linguists would claim it is not possible to distinguish a split that occurred two or three thousand years ago from one that is at least eight or ten thousand years old. The evidence that can be brought to bear on the possible time depth of the lexical and grammatical homologies shared by Yeniseian and Na-Dene all point roughly to an Early to Mid-Holocene dispersal of 10,000 to 8,000±500 calBP as a plausible time depth for the breakup of Dene-Yeniseian. A separation date significantly earlier than 10,000 calBP would be incompatible with generally accepted facts about language change, while a date significantly more recent than 8,000 calBP is contradicted by the fact that Na-Dene itself shows evidence of internal diversification that likely began at least 4,500 calBP. Both the grammatical and lexical comparative data indicate that the Dene-Yeniseian connection is significantly deeper than Proto-Na-Dene but still detectable using the Comparative Method. The accumulated linguistic and genetic evidence preclude the possibility that the Dene-Yeniseian connection dates back to the original peopling of the Americas from a common Beringian population, or that the Yeniseians derive from a recent back migration from Alaska across Bering Strait. Rather, the connection of Dene-Yeniseian with the ASTt migration, first suggested explicitly by Dumond (2010), appears increasingly plausible.

However, the language(s) of a prehistoric population can never be identified based on DNA studies alone, and pairing genetic and linguistic data to hypothesize about the language of the founding ASTt population yields four additional possibilities. The ASTt / Paleo-Eskimo people could have spoken a language that disappeared leaving no living descendants. It is also possible that the Paleo-Eskimos spoke Proto-Eskimo-Aleut and were responsible for introducing that family into the Americas five millennia ago. Eskimo-Aleut consists of a branch containing the closely related Eskimoan languages (Yup'ik, Iñiupiaq, etc.), probably separated at a depth of less than 2,500 years, and a more divergent Aleut branch. Krauss (1980:7) roughly estimates the split between Eskimoan and Aleut at about 4,000 calBP, which, even with the inexactness of linguistic time depth estimations, would still roughly fit the scenario that the original Paleo-Eskimo founding population in fact spoke Proto-Eskimo-Aleut (Fortescue 2017). Less likely is that the Eskimo-Aleut family represents an earlier Native American substrate rather than a language brought from Asia after 5,000 calBP. This is contradicted by the many typological, areal, and possibly deep genetic affinities between Eskimo-Aleut, Uralic, Yukaghir and other North Asian families that have long been noted by linguists (Fortescue 1998, 2017). A final logical possibility is that the ASTt population, which also shows a close genetic link to present-day Chukchi and Koryak peoples in the Russian Far East, could have spoken a language belonging to the Chukotko-Kamchatkan family, but which subsequently disappeared in North America, leaving living relatives only on the Asian side of Bering Strait. Within Chukotko-Kamchatkan, the Itelmen branch is quite divergent from the family's other branch, which contains Chukchi and Koryak – languages so similar that they could almost be regarded as dialects of a single language (Comrie 1981: 240). Estimating the age of this family as a whole, however, is hindered by the probability that Itelmen's divergence derives in part from mixing with an unidentified substrate language in Kamchatka (Fortescue 1998: 210-213). The same could be argued for estimating the Aleut split with Eskimoan, as Aleut also shows possible signs of substrate admixture or at least of rapid phonological and morphological change (Fortescue 1998: 35-37), which could make the split appear older than it actually is. Chukotko-Kamchatkan and Eskimo-Aleut are both regarded as first-order families, not relatable to one another using the Comparative Method. Therefore, it is implausible that the bearers of the ASTt culture spoke a proto-language ancestral to more than one of the three families (Na-Dene, Chukotko-Kamchatkan or Eskimo-Aleut) spoken by present-day populations near Bering Strait that display substantial Paleo-Eskimo genetic inheritance. A fully convincing demonstration of the Dene-Yeniseian linguistic hypothesis, however, would favor the scenario whereby Paleo-Eskimos brought a language directly ancestral to Proto-Na-Dene into Alaska. The genetic link through Paleo-Eskimos between present-day Siberians (including Kets) and the population ancestral to Na-Dene speaking peoples appears to be the only physical connection between the two groups that falls within a time depth known to be recoverable by the Comparative Method.

Tables S10.1 and S10.2 below summarize the most plausible prehistoric scenario for the existence of a Dene-Yeniseian language link involving the Paleo-Eskimo arrival into Alaska 5,000 calBP from an earlier source in the Syalakh Culture (6,500 to 5,200 calBP) spreading eastward from Siberia.

Table S10.1. Chronology of Dene-Yeniseian linguistic diversification

- ~6,500 calBP – breakup of the Dene-Yeniseian proto-language in central-eastern Siberia (based on coalescence date of Paleo-Eskimo ancestry shared between contemporary Siberians and Na-Dene-speaking populations, see Table S9.2); speakers of the language ancestral to Proto-Yeniseian remained in Siberia, where diversification of the known Yeniseian daughter languages is unlikely to predate 4,000 calBP (based on lexicostatistic estimates).
- ~5,000 YPB – language ancestral to Proto-Na Dene brought into Alaska by Paleo-Eskimos (indexed by archaeological data).
- after 5,000 calBP – split between Tlingit and Athabaskan-Eyak (indexed by coalescence date of Paleo-Eskimo genetic ancestry shared by contemporary Na-Dene peoples, see Table S9.2).
- ~3,400 to 3,000 calBP – split between Eyak and Athabaskan (based on lexicostatistic estimates).
- ~2,700 to 2,200 calBP – beginning of diversification and spread of Athabaskan languages (based on lexicostatistic estimates).

Table S10.2. Hypothesized linguistic affiliations of Middle to Late Holocene cultural expansions in northeastern Asia

- Syalakh** Culture (6,500 to 5,200 calBP) involved the spread of *Dene-Yeniseian languages* across northeastern Asia and into Alaska by 5,000 calBP. The core Syalakh people were plausibly speakers of Proto-Dene-Yeniseian; however, the bearers of this widespread culture in its later stages, and especially of the **Bel'kachi** Culture (5,200 to 3,100 calBP) that succeeded it, plausibly spoke other languages. A single language family is unlikely to have held sway across such a vast region as northeastern Asia over thousands of years of hunter-gatherer prehistory.
- Ymyakhtakh** Culture (4,100 to 3,300 calBP), which developed on the basis of the Syalakh and Bel'kachi traditions, involved population movements that may have brought the language ancestral to the *Eskimo-Aleut family* across Bering Strait into the North American Arctic.
- Tokarev** Culture (3,800 to 1,400 calBP) on the shore of the Sea of Okhotsk is connected with speakers of the *Chukotko-Kamchatkan family* (Fortescue 2011).
- Ust-Mil** Culture (3,500 to 2,600 calBP) is connected with the spread of *Yukaghiric languages* across northeastern Siberia.
- Tungusic** pastoral expansions in the past two millennia from a homeland near the Amur river in Manchuria near the Khingan mountains (Pevnov 2012) erased most of the former linguistic diversity in eastern Siberia, including any surviving Dene-Yeniseian daughter branches in this area.

Despite the shared Paleo-Eskimo genetic component in their speakers, the Dene-Yeniseian, Eskimo-Aleut, and Chukotko-Kamchatkan language families are not relatable using the Comparative Method. Various deep connections have been proposed between Eskimo-Aleut, Uralic, and sometimes Yukaghiric and other Eurasian families (Fortescue 1998; see Campbell and Poser 2008 for a critique); however, even if any of these hypotheses are valid, the linguistic unity in question would greatly predate the spread of Middle Holocene cultures as well as the coalescence dates of the Paleo-Eskimo genetic ancestry shared by their speakers.

References (for this section)

- Alekseenko, E. A. K izucheniju mifologicheskikh paralelej medvezh'emu kul'tu ketov [Mythological parallels to the Ket Bear Cult]. *Sistemnye Issledovaniya Vzaimosvjazi Drevnikh Kul'tur Sibiri i Severnoj Ameriki. Vypusk 2: Dukhovnaja Kul'tura*. St. Petersburg: RAN, 22-46. (1995).
- Berezkin, Y. Sibirskij fol'klor i proiskhozhdenie na-dene [Siberian folklore and Na-Dene origins]. *Arkheologija, Ėtnografija i Antropologija Evrazii* **43.1**: 122-134. (2015)
- Campbell, L. *American Indian Languages: The Historical Linguistics of Native America*. Oxford: Oxford University Press (1997).
- Campbell, L. Review of "The Dene-Yeniseian Connection". *International Journal of American Linguistics* **77.3**, 445-451 (2011).
- Campbell, L. *Historical Linguistics: An Introduction* (3rd edition). Cambridge, Mass.: MIT Press (2013).
- Campbell, L. & Poser, W. *Language Classification: History and Method*. Cambridge: Cambridge University Press. (2008).
- Comrie, B. *The Languages of the Soviet Union*. Cambridge: Cambridge University Press. (1981).
- Dul'zon, A. P. Ketskie toponimy Zapadnoy Sibiri [Ket toponyms of Western Siberia]. *Uchenye Zapiski Tomskogo Gosudarstvennogo Pedagogicheskogo Instituta [Scholarly Proceedings of Tomsk State Pedagogical Institute]* **18**, 91-111 (1959).
- Dul'zon, A. P. Byloe rasselenie ketov po dannym toponimiki [The former settlement of the Kets according to the facts of toponymy]. *Voprosy Geografii* **68**, 50-84 (1962).
- Dumond, D. The Dene arrival in Alaska. *The Dene-Yeniseian Connection*, ed. Kari, J., Potter, B. *Anthropological Papers of the University of Alaska: New Series* **5**, 335-346 (2010).
- Fortescue, M. *Language Relations Across Bering Strait: Reappraising the Archaeological and Linguistic Evidence*. London & New York: Cassell (1998).
- Fortescue, M. The relationship of Nivkh to Chukotko-Kamchatkan revisited. *Lingua* **121**: 1359-1376 (2011)
- Fortescue, M. Correlating Palaeo-Siberian language populations: Recent advances in the Uralo-Siberian Hypothesis. *Man in India*: **97.1**, 47-68 (2017).
- Golla, V. *California Indian Languages*. Berkeley, Los Angeles, London: University of California Press (2011).
- Heggarty, P. & Renfrew, C. The Americas: languages. *Cambridge World Prehistory*. Cambridge: Cambridge University Press, 1326-1353 (2014).
- Holton G. & Sicoli, M. 2014. Linguistic phylogenies support back-migration from Beringia to Asia. *PLoS ONE* **9.3**: e91722. doi:10.1371/journal.pone.0091722
- Kari, J. The concept of geolinguistic conservatism in Na-Dene prehistory. *The Dene-Yeniseian Connection*, ed. Kari, J., Potter, B. *Anthropological Papers of the University of Alaska: New Series* **5**, 194-222. (2010).
- Kari, J. & Potter, B. (Eds.). *The Dene-Yeniseian Connection. Anthropological Papers of the University of Alaska: New Series* **5**. Fairbanks, AK: ANLC (2010).
- Kaufman, T., & Golla, V. Language groupings in the New World: their reliability and usability in cross-disciplinary studies. *America Past, America Present: Genes and Languages in the Americas and Beyond*, ed. Renfrew, C. Cambridge: Macdonald Institute for Archaeological Research, 47-57 (2000).
- Krauss, M. Na-Dene. *Native Languages of the Americas*, vol. 1, ed. Sebeok, T. A. New York & London: Plenum Press, 283-358 (1976).
- Krauss, M. *Alaska Native Languages: Past, Present and Future*. Fairbanks, AK: ANLC (1980).
- Krauss, M. Athabaskan tone. *Athabaskan Prosody*, ed. Hargus, S., Rice, K, Amsterdam & New York: John Benjamins, 55-136 (2005).
- Leer, J. *Comparative Athabaskan Lexicon*: www.uaf.edu/anla/collections/ca/cal/ (2006).
- Leer, J. The palatal series in Athabaskan-Eyak-Tlingit with an overview of the basic sound correspondences. *The Dene-Yeniseian Connection*, ed. Kari, J., Potter, B. *Anthropological Papers of the University of Alaska: New Series* **5**, 168-193 (2010).
- McMahon, A. & McMahon, R. *Language Classification by Numbers*. Oxford: Oxford University Press (2005).
- Mühlenbernd, R., & Rama, T. What phoneme networks tell us about the age of language families. *Journal of Language Evolution* **2**, 67-76. (2017).
- Nichols, J. Proving Dene-Yeniseian genealogical relatedness. *The Dene-Yeniseian Connection*, ed. Kari, J., Potter, B. *Anthropological Papers of the University of Alaska: New Series* **5**, 299-309 (2010).
- Pevnov, A. M. The problem of the localization of the Tungus-Manchu homeland. *Recent advances in Tungusic linguistics*, ed. A. Malchukov, Whaley, L. Wiesbaden: Harrassowitz, 17-40. (2012).
- Ruhlen, M. The origin of the Na-Dene. *Proc. Natl. Acad. Sci. USA* **95**, 13994-13996 (1998).
- Starostin, G. Dene-Yeniseian and Dene-Caucasian: pronouns and other thoughts. *Working Papers in Athabaskan Languages* **8**. Fairbanks, AK: ANLC, 107-117 (2010).
- Starostin, S. A. Sravnitel'nyj slovar' enisejskikh jazykov [A comparative vocabulary of Yeniseian languages]. *Ketskij Sbornik* vol. 4, Moscow: Vostochnaja Literatura, 176-315 (1995).
- Swadesh, M. Some new glottochronological dates for Amerindian linguistic groupings. *Proceedings of the 32nd International Congress of Americanists* 670-674 (1958).
- Trombetti, A. *Elementi di Glottologia*. Bologna: Nicola Zanichelli. pp. 486, 511 (1923).

- 1482 Vajda, E. *Yeniseian Peoples and Languages: A History of their Study with an Annotated Bibliography and a*
1483 *Source Guide*. Surrey, England: Curzon Press (2001).
- 1484 Vajda, E. Loanwords in Ket. *The Typology of Loanwords*, ed. Haspelmath, M., Tadmor, U. Oxford: Oxford
1485 University Press, 125–139 (2009).
- 1486 Vajda, E. Siberian link with Na-Dene languages. *The Dene-Yeniseian Connection*, ed. Kari, J., Potter, B.
1487 *Anthropological Papers of the University of Alaska: New Series* 5, 33–99 (2010a).
- 1488 Vajda E. Yeniseian, Na-Dene, and historical linguistics. *The Dene-Yeniseian Connection*, ed. Kari, J., Potter, B.
1489 *Anthropological Papers of the University of Alaska: New Series* 5, 100–118 (2010b).
- 1490 Vajda, E. Vestigial possessive morphology in Na-Dene and Yeniseian. *Working Papers in Athabaskan (Dene)*
1491 *Languages 2012*. (Alaska Native Language Center Working Papers No. 11). Fairbanks: ANLC, 79-91
1492 (2013).
- 1493 Vajda, E. Dene-Yeniseian. *Oxford Research Encyclopedia of Linguistics*. Oxford Online (2016).
- 1494 Vajda, E. Patterns of innovation and retention in templatic polysynthesis. *Handbook of Polysynthesis*, ed.
1495 Fortescue, M, Mithun, M., Evans, N. Oxford: Oxford University Press, 363-391 (2017).
- 1496 Vajda, E. Yeniseian and Athapaskan toponyms. *Arctic Anthropology*. (in press).
- 1497 Werner, H. *Zur jenseits-indianischen Urverwandtschaft [Yeniseian and Native American Relatedness]*.
1498 Wiesbaden: Harrassowitz (2004).
- 1499 Werner, H. *Die Jenseits-Sprachen des 18. Jahrhunderts [Yeniseian Languages of the 18th Century]*.
1500 Wiesbaden: Harrassowitz (2005).
- 1501