

Paleo-Eskimo genetic legacy across North America

Pavel Flegontov^{1,2,3}, N. Ezgi Altınışık^{1,‡}, Piya Changmai^{1,‡}, Nadin Rohland⁴, Swapan Mallick^{4,5,6}, Deborah A. Bolnick⁷, Francesca Candilio^{8,9}, Olga Flegontova³, Choongwon Jeong¹⁰, Thomas K. Harper¹¹, Denise Keating⁸, Douglas J. Kennett¹¹, Alexander M. Kim⁴, Thiseas C. Lamnidis¹⁰, Iñigo Olalde⁴, Jennifer Raff¹², Robert A. Sattler¹³, Pontus Skoglund⁴, Edward J. Vajda¹⁴, Sergey Vasilyev¹⁵, Elizaveta Veselovskaya¹⁵, M. Geoffrey Hayes¹⁶, Dennis H. O'Rourke¹², Ron Pinhasi^{8,17}, Johannes Krause¹⁰, David Reich^{4,5,6,@}, Stephan Schiffels^{10,@}

¹ *Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava 71000, Czech Republic*

² *A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127051, Russia*

³ *Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice 37005, Czech Republic*

⁴ *Department of Genetics, Harvard Medical School, Boston, MA 02115, USA*

⁵ *Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA*

⁶ *Broad Institute of MIT and Harvard, Cambridge, MA 02412, USA*

⁷ *Department of Anthropology and Population Research Center, University of Texas at Austin, Austin, TX 78712, USA*

⁸ *Earth Institute and School of Archaeology, University College Dublin, Dublin 4, Ireland*

⁹ *Soprintendenza Archeologia belle arti e paesaggio per la città metropolitana di Cagliari e per le province di Oristano e Sud Sardegna, Cagliari 9124, Italy*

¹⁰ *Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena 07745, Germany*

¹¹ *Department of Anthropology and Institutes for Energy and the Environment, Pennsylvania State University, University Park, PA 16802, USA*

¹² *Department of Anthropology, University of Kansas, Lawrence, KS 66045, USA*

¹³ *Tanana Chiefs Conference, Fairbanks, AK 99701, USA*

¹⁴ *Department of Modern and Classical Languages, Western Washington University, Bellingham, WA 98225, USA*

¹⁵ *Institute of Ethnology and Anthropology, Russian Academy of Sciences, Moscow 119017, Russia*

¹⁶ *Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA*

¹⁷ *Department of Anthropology, University of Vienna, Vienna 1090, Austria*

[‡] The authors contributed equally

[@] Co-senior authors

Abstract

Paleo-Eskimos were the first people to settle vast regions of the American Arctic around 5,000 years ago, and were subsequently joined and largely displaced around 1,000 years ago by ancestors of the present-day Inuit and Yupik. The genetic relationship between Paleo-Eskimos and Native American populations remains

45 **uncertain. We analyze ancient and present-day genome-wide data from the**
46 **Americas and Siberia, including new data from Alaskan Iñupiat and West Siberian**
47 **populations, and the first genome-wide DNA from ancient Aleutian Islanders, ancient**
48 **northern Athabaskans, and a 4,250-year-old individual of the Chukotkan Ust'-Belaya**
49 **culture. Employing new methods based on rare allele and haplotype sharing as well**
50 **as established methods based on allele frequency correlations, we show that Paleo-**
51 **Eskimo ancestry is widespread among populations who speak Na-Dene and Eskimo-**
52 **Aleut languages. Using phylogenetic modelling with allele frequency correlations**
53 **and rare variation, we present a comprehensive model for the complex peopling of**
54 **North America.**

55
56 Current evidence suggests that present-day Native Americans descend from at least four
57 distinct streams of ancient migration from Asia¹⁻³. The largest ancestral contribution was
58 from populations that separated from the ancestors of present-day East Asian groups
59 ~23,000 calendar years before present (calBP), occupied Beringia for several thousand
60 years, and then moved into North and South America approximately 16,000 calBP². To be
61 consistent with the previous genetic literature we call this lineage "First Americans", while
62 acknowledging that indigenous scholars have suggested the term "First Peoples" as an
63 alternative. The deepest phylogenetic split in this group gave rise to one lineage that
64 contributed to northern North American groups (including speakers of Na-Dene,
65 Algonquian and Salishan languages), and to another lineage that is found in some North
66 Americans as well as all Native Americans from Mesoamerica southward^{1,2,4}. The 12,600
67 calBP ancient genome from an individual assigned to the Clovis culture belongs to the
68 southern lineage⁵. In addition, a separate source of Asian ancestry that has been called
69 "Population Y" contributed more to Native American groups in Amazonia than to other
70 Native Americans^{2,3}. A third stream of migration contributed up to ~50% of the ancestry of
71 the Inuit and Aleut peoples (Eskimo-Aleut speakers), but the Asian source population for
72 this stream remained unidentified¹. Of key importance for understanding the impact of
73 these different lines of ancestry are populations speaking Na-Dene languages, which
74 include the Tlingit, Eyak (recently extinct), and Northern and Southern Athabaskan
75 languages, spoken across much of Alaska and northwestern Canada, with additional
76 isolated Na-Dene languages spoken further south along the Pacific Coast and in
77 southwestern North America⁶. It has been argued¹ that Na-Dene-speaking populations
78 harbor ancestry from another distinct migration: ancient Paleo-Eskimos deriving from
79 Chukotka around 5,000 calBP and expanding throughout the American Arctic for more
80 than 4,000 years⁷⁻⁹. An alternative view is that Paleo-Eskimo-derived ancestry disappeared
81 entirely from temperate North America after the arrival of Thule Inuit, and the distinctive
82 ancestry in Na-Dene speakers might instead reflect admixture from Thule Inuit^{2,8,10}.

83
84 The archaeological record in the Arctic provides clear evidence for the spread of Paleo-
85 Eskimo culture, which spread across the Bering strait about 5,000 calBP^{9,11-13}, and
86 expanded across coastal Alaska, Arctic Canada and Greenland a few hundred years later.
87 Direct ancient DNA data has proven that the Paleo-Eskimo cultural spread was strongly
88 correlated with the spread of a new people^{7,8} that continuously occupied the American
89 Arctic for more than four millennia until ~700 calBP^{9,14,15}. A long-term cultural, and likely
90 linguistic and genetic, boundary was established upon their arrival, which separated
91 populations in the coastal Arctic tundra from indigenous Native American groups who
92 populated the interior forest zone and were plausibly ancestors of present-day Na-Dene
93 speakers¹⁶. Paleo-Eskimo archeological cultures are grouped under the Arctic Small Tool

94 tradition (ASTt), and include the Denbigh, Choris, Norton, and Ipiutak cultures in Alaska
95 and the Saqqaq, Independence, Pre-Dorset, and Dorset cultures in the Canadian Arctic and
96 Greenland⁹. The ASTt source has been argued to lie in the Syalakh-Bel'kachi-Ymyakhtakh
97 culture sequence of East Siberia, dated to 6,500 – 2,800 calBP^{17,18}. In this paper, we use the
98 genetic label “Paleo-Eskimo” to refer to the ancestry associated with ancient DNA from the
99 ASTt and “Neo-Eskimo” to refer to ancient DNA from the later Northern Maritime tradition.
100 While we recognize that some indigenous groups would prefer that the term “Eskimo” not
101 be used, we are not aware of an alternative term that all relevant groups prefer instead.
102 The terms “Paleo-Inuit” and “Thule Inuit” have been proposed as possible replacements for
103 “Paleo-Eskimo” and “Neo-Eskimo”, respectively¹⁹, but the use of “Inuit” in this context
104 might seem to imply that individuals from these ancient cultures are more closely related
105 to present-day Inuit than to present-day Yupik, whereas genetic data show that Yupik and
106 Inuit derive largely from the same ancestral populations (see below). Moreover, the term
107 “Thule” does not cover the whole spectrum of Northern Maritime cultures, being strongly
108 associated with the latest phase of this tradition. We therefore use the “Eskimo”
109 terminology here while acknowledging its imperfections.

110
111 Paleo-Eskimo dominance in the American Arctic ended about 1,350 – 1,150 calBP, when
112 the Thule culture became established in Alaska and rapidly spread eastwards after 750 –
113 650 calBP^{9,14,15}. This spread has been shown genetically to reflect the movement of people⁸.
114 The Thule Inuit had material culture links to hunter-gatherer societies in the Bering Strait
115 region (e.g., Old Bering Sea culture, starting about 2,200 calBP, and Birnik culture), who
116 depended on marine resources²⁰. More complex and diverse transportation technologies,
117 weaponry, and, most importantly, a food surplus created by whale hunting, contributed to
118 the success of these Neo-Eskimo cultures and to eventual disappearance of the Paleo-
119 culture with which it competed^{11,15,21}.

120
121 A 4,000-year-old Paleo-Eskimo from western Greenland, associated with the Saqqaq
122 culture, was the first ancient anatomically modern human to have his whole genome
123 sequenced, yielding a genome of 16x coverage⁷. Later work reported low-coverage data for
124 additional individuals affiliated with the Pre-Dorset, Dorset and Saqqaq cultures⁸. These
125 studies showed that Paleo-Eskimos were a genetically continuous population⁸ and are
126 most closely related, among present-day groups, to Chukotko-Kamchatkan-speaking
127 Chukchi and Koryaks who live in far eastern Siberia^{2,7,8}. The split time between the first
128 Saqqaq individual sequenced and the Chukchi was estimated at 6,400 – 4,400 calBP⁷,
129 consistent with archaeological data. Present-day speakers of Eskimo-Aleut languages and
130 ancient Neo-Eskimos represent another continuous population, related to Paleo-Eskimos
131 and Chukchi, but distinct⁸. No admixture was detected between Neo- and Paleo-Eskimos in
132 the Canadian Arctic and Greenland⁸, consistent with the lack of evidence for interactions
133 between their material cultures¹⁴. However, Raghavan *et al.*⁸ hypothesized early gene flow
134 from the Neo-Eskimo into the Paleo-Eskimo lineage in Beringia, and Raff *et al.*²² found
135 mitochondrial evidence for possible gene flow from Paleo-Eskimos into the ancestors of
136 contemporary Iñupiat from the North Slope of Alaska. It is important to recognize that
137 substantial coverage genome-wide data from Alaskan Paleo-Eskimo cultures, including
138 Choris and Norton, and from Chukotkan cultures possibly related to Paleo-Eskimos (the
139 Ust'-Belaya and Wrangel island sites) have never been reported.

140
141 In this study, we resolve the debate around the distinctive ancestry in Na-Dene and
142 determine the genetic origin of Neo-Eskimos and their relationships with Paleo-Eskimos

143 and Chukotko-Kamchatkan speakers. We present the first genomic data for ancient
144 Aleutians, ancient Northern Athabaskans, Chukotkan Neo- and Paleo-Eskimos, and
145 present-day Alaskan Iñupiat. We also present new genotyping data for West Siberian
146 populations (Enets, Kets, Nganasans, and Selkups). Analyzing these data in conjunction
147 with an extensive set of public sequencing and genotyping data, we demonstrate that the
148 population history of North America was shaped by two major admixture events between
149 Paleo-Eskimos and the First Americans, which gave rise to both the Neo-Eskimo and Na-
150 Dene populations.

151 **Results**

152 **Dataset**

153 We generated new genome-wide data from 11 ancient Aleutian Islanders that date from
154 2,320 to 140 calBP, three ancient Northern Athabaskans (McGrath, Upper Kuskokwim
155 River, Alaska, 790 – 640 calBP), two Neo-Eskimos of the Old Bering Sea culture (Uelen,
156 Chukotka, 1,970 – 830 calBP), and one individual of the Ust'-Belaya culture (Ust'-Belaya,
157 Chukotka, 4,410 – 4,100 calBP) (Table 1, Supplementary Table 1, Supplementary
158 Information sections 1 and 2). For each of these 17 individuals, we extracted bone powder
159 in a dedicated clean room, extracted DNA²³, and prepared a double-stranded library
160 treated with uracil-DNA glycosylase enzymes to greatly reduce the rate of characteristic
161 ancient DNA damage²⁴. We enriched the libraries for a targeted set of approximately 1.24
162 million single nucleotide polymorphisms (SNPs)²⁵. We assessed the authenticity of the
163 samples based on the rate of matching of sequences to the mitochondrial consensus, X
164 chromosome polymorphism in males, and cytosine-to-thymine mismatch to the human
165 reference genome in the terminal nucleotides of each read, which is a characteristic
166 signature of genuine ancient DNA (Table 1, Supplementary Information section 3). By itself,
167 this dataset increases the number of individuals from the American Arctic and from far
168 eastern Siberia with more than 1.0x coverage on analyzed positions by 11-fold (10 samples
169 in our study meet this threshold compared to only one that met this threshold previously⁷).
170 In addition to the newly reported ancient data, we report new SNP genotyping data for
171 present-day populations: 35 Alaskan Iñupiat (Inuit), 3 Enets, 19 Ket, 22 Nganasan, and 14
172 Selkup (Supplementary Table 2).

173
174 We merged the newly reported ancient and modern data with previously published data to
175 create three main datasets covering Africa, Europe, Southeast Asia, Siberia, and the
176 Americas (Fig. 1, Supplementary Tables 3, 4). For most analyses, we combined groups into
177 meta-populations, as indicated in Fig. 1 and summarized in Supplementary Table 3. The
178 breakdown of groups into these meta-populations was guided by unsupervised clustering
179 using *ADMIXTURE* (Extended Data Fig. 1), *fineSTRUCTURE* (Extended Data Fig. 2), Principal
180 Component Analysis (PCA) (Fig. 2, Extended Data Fig. 3, Supplementary Information
181 section 4). For naming the Arctic meta-populations, we use names of recognized language
182 families: Na-Dene, Eskimo-Aleut, Chukotko-Kamchatkan. We chose these terms since
183 genetic and linguistic relationship patterns are highly congruent in this region (see below).
184

185 **Gradient of Paleo-Eskimo-related ancestry**

186 PCA applied to SNP array datasets (Fig. 2) reveals a striking linear cline with Paleo-
187 Eskimos (Saqqaq and Late Dorset) and some Chukotko-Kamchatkan speakers at one

188 extreme, then Chukchi, then contemporary Eskimo-Aleut speakers and ancient Neo-
189 Eskimos and Aleuts, then Na-Dene speakers, then northern North Americans, and finally
190 southern First Americans at the other extreme. The patterns were qualitatively identical
191 for the HumanOrigins and Illumina datasets, in analyses carried out with or without
192 transition polymorphisms (Fig. 2, Extended Data Fig. 3, Supplementary Information section
193 4). This qualitative pattern in PCA is driven by admixture, as we verified using the *qpWave*
194 method¹. *qpWave* relies on a large matrix of f_4 -statistics measuring allele sharing
195 correlation rates between all possible pairs of a set of outgroups and all possible pairs of a
196 set of test populations. A statistical test¹ can then be performed to determine whether
197 allele frequencies in the test populations can be explained by one, two, or more streams of
198 ancestry derived in different ways from the outgroups; this test gives a single P-value that
199 appropriately corrects for multiple hypothesis testing. We verified that all the individuals
200 on the PCA cline could be modeled as descended from two streams of ancestry relative to a
201 diverse set of Siberians, Southeast Asians, Europeans, and Africans. Since Chukotko-
202 Kamchatkan speakers are closely related to Paleo-Eskimos as shown here (Fig. 2, Extended
203 Data Figs. 2, 3) and in previous studies^{2,7}, we included them along with the American
204 groups as test populations. With this setup, a great majority of all possible population
205 quadruplets of the form (First American, Na-Dene, Eskimo-Aleut, Paleo-Eskimo) were
206 consistent with two streams of ancestry derived from the outgroups ($P > 0.05$), especially on
207 the datasets lacking transition polymorphisms in order to avoid possible confounding
208 effects due to ancient DNA degradation (Supplementary Information section 5).

209
210 Under the assumption that the populations at the extremes of the cline are descended
211 solely from one of the source populations, we can assign admixture proportions to all
212 populations in the middle of the cline using *qpAdm*, an extension of *qpWave*²⁶. Thus, we
213 attempted modelling diverse American populations as descended from both southern or
214 northern First Americans and Paleo-Eskimos. This analysis reveals a gradient of Paleo-
215 Eskimo-related ancestry proportions, with the relative values almost perfectly
216 proportional to the position along the PCA gradient (Fig. 2, Extended Data Fig. 3). The
217 *qpAdm* estimates of Paleo-Eskimo-related ancestry are as follows: southern First
218 Americans (by definition 0%), northern First Americans (3%), present-day Na-Dene (7-
219 22%), ancient Northern Athabaskans (23-38%, depending on the dataset), Eskimo-Aleuts
220 other than Yupik (30-68%), Yupik (71-76%), Chukotko-Kamchatkans (~100%), and Paleo-
221 Eskimos (by definition 100%) (Fig. 3, Extended Data Figs. 4, 5). Adding a Chukotko-
222 Kamchatkan-speaking population without recent American back-flow (Koryak) to the
223 outgroup dataset changed these results: three streams of ancestry generally fit the data in
224 the full datasets, but the picture was more ambiguous in the transition-free datasets, with
225 the HumanOrigins-based transition-free dataset still supporting the model with two
226 migration streams (Supplementary Information section 5). Nevertheless, admixture
227 proportions inferred by *qpAdm* remained largely unchanged (Extended Data Figs. 4, 5).

228
229 In summary, all indigenous populations of North America, Chukotka and Kamchatka are
230 consistent with deriving from two ancestry streams to the limits of our resolution, which
231 we term First American and proto-Paleo-Eskimo (PPE). This “distant perspective” treats
232 the region west of the Bering Strait (notably Chukotka and Kamchatka) as part of the
233 American radiation. Usage of a close outgroup within the PPE radiation (Koryak), as also
234 done in Reich *et al.*¹, yields a “close perspective” and models additional population
235 structure within the PPE radiation which explains the finding in that study of three streams
236 of ancestry connecting Asia to the Americas rather than the two streams of ancestry we

237 focus on here. We find that the PPE source population for Eskimo-Aleut speakers is a
238 distinct line of ancestry, different from Paleo-Eskimos *sensu stricto*, in that it is more closely
239 related to present-day Chukotko-Kamchatkan speakers (see the demographic modelling
240 results below). In contrast, the PPE source that contributed to Na-Dene is most closely
241 related to Paleo-Eskimos *sensu stricto*, as seen by a *qpWave* analysis on population triplets
242 (Na-Dene, First American, Paleo-Eskimo), which are generally consistent with two
243 migration streams on all datasets even with Koryak in the outgroups (Extended Data Figs.
244 4, 5). Below, we use methods based on autosomal haplotypes and rare variants to further
245 investigate whether Paleo-Eskimos¹ or Thule Inuit^{2,8} contributed the distinctive ancestry
246 which these analyses show were present in Na-Dene speakers.
247

248 **Source of distinct ancestry in Na-Dene**

249 To investigate Paleo-Eskimo ancestry in Native Americans in a hypothesis-free way, we
250 considered haplotypes shared with the ancient Saqqaq individual. As compared to allele
251 frequencies at unlinked loci, autosomal haplotypes in some cases have more power to
252 distinguish potential closely related sources of gene flow^{27,28}, such as Thule Inuit and Paleo-
253 Eskimos. Cumulative lengths of shared autosomal haplotypes were produced with
254 *ChromoPainter v.1* for all pairs of individuals²⁹. First, for each American individual, we
255 considered the length of haplotypes shared with Saqqaq (in cM), which we refer to as
256 Saqqaq haplotype sharing statistic or HSS. We also estimated haplotype sharing between
257 each American individual and African, European, Siberian, and Arctic (Chukotko-
258 Kamchatkan- and Eskimo-Aleut-speaking) individuals by averaging HSS across members of
259 a given meta-population. To correct for potential biases caused by sequence quality and
260 coverage, the Saqqaq HSS was divided by the African HSS for each group, and the resulting
261 statistic was termed relative HSS (Extended Data Fig. 6).
262

263 In both genome-wide genotyping datasets, most Native American individuals with the
264 highest relative Saqqaq HSSs belonged to the Na-Dene group. This enrichment cannot be
265 explained by either Arctic or European admixture in these individuals, as shown by the
266 poor correlation with Arctic and European relative HSSs (Extended Data Fig. 6). We note
267 that some correlation of the Saqqaq and Arctic HSSs is expected under any admixture
268 scenario since Saqqaq falls into the Arctic clade in trees based on haplotype sharing
269 patterns (Extended Data Fig. 2).
270

271 While the HumanOrigins dataset includes only two Northern Athabaskan-speaking groups
272 from Canada (Chipewyans and Dakelh) and only three other northern First American
273 groups (Algonquins, Cree, Ojibwa), the Illumina dataset includes six such populations in
274 addition to all extant major branches of the Na-Dene language family: four groups of
275 Northern Athabaskan speakers, one Southern Athabaskan group, and one Tlingit group. At
276 least one individual from each Na-Dene branch demonstrates a relative Saqqaq HSS
277 surpassing that of any Central or South American (Extended Data Fig. 6). The results were
278 very similar when using a genetically distant meta-population (African) and a much closer
279 one (Siberian) as normalizers (Supplementary Information section 6).
280

281 To interpret haplotype sharing in a more quantitative way, we analyzed putative admixture
282 events in Na-Dene speakers using *GLOBETROTTER*³⁰. To make a complex ancestry history
283 of Na-Dene amenable to *GLOBETROTTER* analysis, we pre-selected individuals based on
284 low European admixture and high Saqqaq HSS (selected individuals are marked in

285 Supplementary Information section 6). Consistent with our qualitative observations, Paleo-
286 Eskimos (represented by the Saqqaq individual) and First Americans were identified by
287 *GLOBETROTTER* as the most likely sources of ancestry for Na-Dene, with the Paleo-Eskimo
288 contribution ranging from 7% to 51%, depending on the dataset and *GLOBETROTTER* set-
289 up. Admixture dates were estimated as 2,202 – 479 calBP (Supplementary Information
290 section 7).

291
292 As an independent test, we analyzed rare genetic variants in the complete genome dataset.
293 Rare variants, with global frequency of less than 1%, have been shown to have more power
294 to resolve subtle relationships than common variants^{31,32}. We calculated rare allele sharing
295 statistics (RASS), which measure the number of rare variants (up to allele frequency 0.2%
296 in the entire dataset) an individual shares with reference meta-populations, in this case
297 Siberian and Arctic (Extended Data Fig. 7, Supplemental Information section 4). To
298 normalize coverage differences between individuals and dataset-specific variant calling
299 biases, we divided these statistics by allele sharing with Europeans or Africans (Extended
300 Data Fig. 7). On a two-dimensional plot combining Arctic and Siberian RASS, four meta-
301 population lines are visible: Siberian, First American, Chukotko-Kamchatkan, and Eskimo-
302 Aleut (Fig. 4). All four Northern Athabaskan (Dakelh and Chipewyan) individuals are
303 shifted on the Arctic axis by more than three standard error intervals from the First
304 American cluster. The Arctic/Siberian RASS ratios are almost identical in Athabaskans and
305 Saqqaq, but significantly different in present-day Eskimo-Aleut-speaking individuals and in
306 an ancient Aleut individual, for which we generated whole genome shotgun data of 2.7x
307 coverage (Table 1, Fig. 4). Allele sharing statistics behave linearly under recent admixture,
308 and we used linear combinations to calculate expected statistics for First American/Saqqaq
309 and First American/Eskimo-Aleut admixture. Notably, relative RASSs for both Dakelh
310 individuals match those of the simulated First American/Saqqaq admixture, but the
311 statistics for two Chipewyans are consistent with both admixture scenarios (Extended Data
312 Fig. 8). Similar results were obtained in an analysis without transitions (Extended Data Fig.
313 8).

314
315 Taken together, our results from several analyses show remarkable consistency: PCA,
316 haplotype and rare allele sharing, *GLOBETROTTER*, and *qpWave/qpAdm* suggest that
317 present-day Na-Dene speakers lacking post-Columbian admixture have roughly 10% to
318 25% Paleo-Eskimo ancestry. Our newly reported data from the three ancient individuals
319 from the Tochak McGrath site dated at ~800 calBP, found in a region currently inhabited
320 by Na-Dene speakers, are derived from the same combination of First American and Paleo-
321 Eskimo lineages as present-day Na-Dene, providing support for the hypothesis of local
322 population continuity, also supported by continuity in material culture¹⁶. However, the
323 ancient Tochak McGrath samples have a higher estimated proportion of Paleo-Eskimo
324 ancestry than any present-day Na-Dene speakers in our dataset, 25-40%, suggesting that
325 ongoing gene flow from neighboring First American populations has been reducing the
326 Paleo-Eskimo ancestry in Na-Dene. Paleo-Eskimo ancestry is likely present at a low level in
327 other northern First Americans (Extended Data Figs. 4-6) due to this bidirectional gene
328 flow. Two Dakelh individuals with genome sequencing data available⁸ have yielded
329 consistent results throughout all analyses (Figs. 2-4, Extended Data Figs. 4-8,
330 Supplementary Information section 6), and just a few of the 350 First American individuals
331 sampled exhibited a signal of Paleo-Eskimo ancestry that is comparable to that seen in Na-
332 Dene speakers (Extended Data Fig. 6b). These results suggest that the common ancestor of
333 all Na-Dene branches, now scattered from Arizona and New Mexico to Alaska, experienced

334 Paleo-Eskimo admixture. This scenario is in agreement with evidence from archaeology
335 (see Discussion), and below we further investigate it with explicit demographic modelling
336 based on the rare joint site frequency spectrum.

337

338 **No evidence for population turnover in the Aleutian Islands around 1,000 calBP**

339 Morphological disparities between human remains in the Aleutian Islands dated before
340 and after around 1,000 calBP, the time of the Thule expansion, were suggested by Hrdlička
341 as reflecting a population turnover³³. Archaeological evidence also suggests dramatic
342 material culture changes around this time including burial practices and other cultural
343 expressions³⁴, and these distinct cultures were termed Paleo- and Neo-Aleut. Mitochondrial
344 DNA analysis provided some evidence for population turnover via an increase in the
345 frequency of mitochondrial DNA haplogroup D2a1a at the expense of A2a after around
346 1,000 calBP³⁵. However, in the genome-wide ancient DNA that we report here, including 4
347 samples labeled as Paleo-Aleuts and 7 samples labeled as Neo-Aleuts, we find no evidence
348 for genetic differences among the two groups. This is evident from PCA and *ADMIXTURE*
349 analyses including 2 Paleo-Aleuts and 4 Neo-Aleuts with the highest number of genotyped
350 sites (Fig. 2, Extended Data Fig. 1), in allele frequency differentiation ($F_{ST} = 0.003 \pm 0.002$,
351 which is consistent with zero), and in tests for being derived from a homogeneous
352 ancestral population (all statistics of the form $D(\text{Outgroup, Test; Neo-Aleut, Paleo-Aleut})$
353 which measure whether a Test population shares more alleles with Neo-Aleuts and Paleo-
354 Aleuts are within 3 standard errors of zero). With the *qpWave* method, we also failed to
355 detect additional ancestry in Neo-Aleuts: both groupings were consistent with one stream
356 of ancestry with P-values ranging from 0.089 to 0.395, depending on the outgroups used.
357 We conclude from this that the Aleutian population largely remained continuous during
358 this transition, unlike the transition between Paleo-Eskimos and Neo-Eskimos in much of
359 the mainland (see further discussion in Supplemental Information section 8).

360

361 **A Paleo-Eskimo with West Siberian ancestry**

362 The Ust'-Belaya culture of interior Chukotka shows connections with both late Neolithic of
363 interior Siberia (e.g., Bel'kachi, Ymyakhtakh) as well as with Paleo-Eskimo cultures in the
364 Bering Strait region^{36,37}. We dated a single burial at the Ust'-Belaya site at the confluence of
365 the Belaya and Anadyr Rivers, from which we also generated genome-wide data
366 (Supplementary Information section 1), and obtained a date of 4,410 – 4,100 calBP
367 (Supplementary Information section 2). Our targeted enrichment approach generated
368 pseudo-haploid genotypes at 832,452 sites across this individual's genome (Table 1). The
369 position of this sample in the space of two principal components (PC1 and PC2) suggests
370 that it might have ancestry from both Paleo-Eskimos and western Siberian lineages (Fig. 2).
371 Indeed, *qpAdm* analysis demonstrates that the Ust'-Belaya individual can be modelled as
372 descended from Paleo-Eskimos (represented by Saqqaq) and West Siberians (represented
373 by Kets or other groups), with Siberian admixture proportions ranging from ~20 to ~50%,
374 depending on source populations and datasets (Extended Data Fig. 9). Models with East
375 Siberians instead of West Siberians (Extended Data Fig. 9) and/or Chukotko-Kamchatkan
376 speakers instead of Paleo-Eskimos (data not shown) were often inconsistent with two
377 sources of ancestry, or demonstrated negative admixture proportions or very wide error
378 intervals. In line with these findings, the Ust'-Belaya individual is closely related to both
379 Saqqaq and Kets, a West Siberian population, according to f_4 -statistics (Ust'-Belaya, Yoruba;
380 Saqqaq, any other population) and (Ket, Yoruba; Ust'-Belaya, any other population)
381 (Supplementary Table 5). Moreover, this individual has a high level of Mal'ta (ancient
382 North Eurasian, ANE) ancestry according to f_4 -statistics (Mal'ta, Yoruba; Ust'-Belaya,

383 Siberian population) (Supplementary Table 5): an expected result given that West
384 Siberians have substantial ANE ancestry³⁸. In summary, striking parallels in archaeological
385 and genetic results suggest that admixture between proto-Paleo-Eskimos and Siberian
386 lineages in Chukotka took place not long after they diverged (see the next section),
387 indicating that cultural contact between these groups at this time almost certainly occurred
388 as well. This result has implications for archaeology and historical linguistics, as discussed
389 below.

390
391 Two later individuals from Old Bering Sea culture burials at Uelen, Chukotka, dated at
392 1,970 – 1,590 calBP and 1,180 – 830 calBP (Supplementary Information section 2), were
393 also subjected to targeted enrichment and sequenced, producing pseudo-haploid
394 genotypes at 608,585 or 797,816 sites (Table 1). Their genetic signature was typical for
395 Neo-Eskimos according to all analyses (see, for example, Figs. 2 and 3). Thus, the older
396 individual represents the earliest Neo-Eskimo for which genetic data were ever reported.
397

398 Demographic modelling

399 To further interpret our findings, we built an explicit demographic model for the
400 populations analyzed here. We used *rarecoal*³¹ to estimate split times and population sizes,
401 as well as admixture events, in a population tree connecting Europeans, Southeast Asians,
402 Siberians, Chukotko-Kamchatkan, Eskimo-Aleut, and Northern Athabaskan speakers, and
403 southern First Americans. Sample sizes and additional details are provided in
404 Supplementary Information section 9. The model was derived in an iterative way: we
405 started off with fitting a model to three populations only (Europeans, Southeast Asians, and
406 First Americans), and then added one population at a time, re-estimating all previous and
407 new parameters (see details in Supplementary Information section 9). Admixture edges
408 were added when the model fit showed significant deviations for particular allele sharing
409 statistics. We explicitly corrected for Post-Columbian admixture in Chukotko-Kamchatkan
410 and Eskimo-Aleut speakers by adding admixture edges from Europeans into these groups,
411 with a fixed time at 200 calBP (not shown in Fig. 5a). The maximum likelihood parameter
412 estimates for this final model are shown in Supplementary Information section 9.
413

414 Our final model suggests that Arctic populations on both sides of the Bering strait, i.e.
415 Chukotko-Kamchatkan and Eskimo-Aleut speakers, form a clade that separated ~6,300
416 calBP from the ancestors of present-day Siberians further to the west. The Arctic clade
417 inherited an additional 18% ancestry from the Asian lineage ancestral to Native Americans.
418 We did not attempt to include the ancient Mal'ta genome³⁹, which would be a
419 representative of ANE, since *rarecoal* requires high-quality genomes for modelling
420 phylogenies. However, the admixture edge from a European sister group into the ancestor
421 of all American, Siberian, and Arctic groups, and a later admixture exclusive to the Native
422 American lineage (in line with the admixture graphs in Extended Data Fig. 10, see below)
423 most likely reflect ANE admixture.

424
425 At 4,000 calBP we infer a split between the ancestors of the Chukotko-Kamchatkan and
426 Eskimo-Aleut speakers, with the latter inheriting a substantial proportion of their ancestry
427 (33%) from an early mixture with a group distantly related to Athabaskan speakers. Finally,
428 we infer that Northern Athabaskans derive 21% of their ancestry from the common
429 ancestor of Chukotko-Kamchatkan and Eskimo-Aleut speakers, likely related to Paleo-
430 Eskimos (see below). We also tested alternative models without this last admixture edge,

431 and with admixture from Eskimo-Aleut or Chukotko-Kamchatkan speakers into
432 Athabaskans, but found substantially poorer fits (Supplementary Information section 9).
433 We note that all time and population size estimates depend linearly on estimates of the
434 human mutation rate (here taken as 0.44×10^{-9} per year^{40,41}), which has substantial
435 uncertainty. While this model was derived by fitting rare variants of allele frequency up to
436 1.8%, we also tested whether it fits F-statistics computed for common variants. Specifically,
437 we used *qpGraph*⁴² to test various topologies including the final one derived with *rarecoal*
438 (Extended Data Fig. 10, Supplementary Information section 9), and found that it was
439 indeed consistent with all possible F-statistics in all combinations of meta-populations (the
440 worst Z-score was 0.883).

441
442 With a robust maximum likelihood model connecting 7 extant groups, we included two
443 ancient genomes in the modelled tree, correcting for missing genotype calls due to limited
444 coverage in these individuals. We find that the Saqqaq genome⁷ most likely branches off
445 the tree very close to the admixture edge into Northern Athabaskan speakers (Fig. 5b). This
446 positioning of Saqqaq at the ancestral branch of the Arctic clade prior to the Chukotko-
447 Kamchatkan/Eskimo-Aleut split suggest that: 1) Paleo-Eskimos are closely related (but not
448 identical) to the founding population of Neo-Eskimos, and 2) Paleo-Eskimos contributed
449 substantially to Na-Dene speakers. We also mapped the ancient Aleutian Islander, for
450 which we generated whole genome data, onto our fitted tree. We find that this individual is
451 most closely related to present-day Eskimo-Aleut speakers, as seen by the maximum
452 likelihood split point on the ancestral branch of that population. This position confirms the
453 continuity with present-day Aleuts as seen in the PCA and other analyses.

454 Discussion

455 The new data and analyses presented in this work derive two key results on the genetic
456 legacy of Paleo-Eskimos. First, we show that Paleo-Eskimos were very closely related to the
457 Asian founder lineage that gave rise to Eskimo-Aleut speakers. Second, we show that Paleo-
458 Eskimos contributed substantially to the ancestry of Native Americans speaking Na-Dene
459 languages. These results add significantly to previous studies on these topics. Reich *et al.*¹
460 inferred that an unspecific Asian source contributed around 43% of the ancestry of
461 Eskimo-Aleut speakers and around 10% of the ancestry of Chipewyans, a Northern
462 Athabaskan-speaking population. Our analyses show that this Asian source is equivalent to
463 the ancestral population that we here term proto-Paleo-Eskimos. We show that within this
464 lineage, two sub-lineages formed that contributed to almost all Na-Dene speakers and to
465 Neo-Eskimos, respectively. According to a different study², Northern Athabaskan speakers
466 did not receive Paleo-Eskimo admixture, but admixture between Athabaskans and Eskimo-
467 Aleut speakers was proposed. While we observe substantial First American ancestry in
468 Eskimo-Aleut speakers, we find no evidence for gene flow from them into Athabaskans.
469 Instead, we propose that the observed genetic patterns can be explained by Paleo-Eskimo
470 ancestry in Athabaskans, as well as in other Na-Dene-speaking populations. Similarly, in a
471 third study⁸, admixture of unresolved direction between Saqqaq and ancestral Neo-
472 Eskimos was interpreted as most likely reflecting Neo-Eskimo admixture into Paleo-
473 Eskimos. Here we show that substantial proto-Paleo-Eskimo ancestry contributed to the
474 founder lineage of Eskimo-Aleut speakers, and think this explains the observed admixture,
475 as well as the presence of mitochondrial haplogroup D2a in the North Slope Iñupiat²².
476

477 Our results show that Paleo-Eskimo ancestry is a nearly perfect tracer-dye for speakers of
478 Na-Dene languages including the most divergent linguistically (Tlingit) and the most
479 geographically remote ones (Southern Athabaskans, Fig. 6a). It is plausible that Paleo-
480 Eskimos rather than Neo-Eskimos contributed to Na-Dene populations in light of
481 archaeological evidence. The arrival of Neo-Eskimos (the Birnirk and Thule cultures) into
482 western Alaska is dated to 1,350 – 1,150 calBP^{8,43}, but at that point Tlingit had probably
483 already come to occupy their current position in southeastern Alaska^{16,44}. It has been
484 hypothesized⁴⁵⁻⁴⁷ that the spread of Southern Athabaskan speakers from the Subarctic was
485 triggered by a massive volcanic ash fall 1,100 calBP⁴⁸ (Fig. 6a). If this hypothesis is correct,
486 both Tlingit and Apache would have had little opportunity to mix with newly arriving Neo-
487 Eskimos, which would explain why in our analysis, southern Athabaskan speakers and
488 Tlingit have the Paleo-Eskimo ancestry but not the Neo-Eskimo ancestry. In contrast,
489 Paleo-Eskimo peoples lived alongside Na-Dene ancestors for millennia, providing ample
490 opportunity for genetic interaction¹⁶. Although archaeological evidence for such interaction
491 across the coastal-interior cultural boundary remains sparse^{16,46}, our genetic analyses
492 demonstrate that substantial gene flow from Paleo-Eskimos took place (25-40% in ancient
493 Northern Athabaskans).

494
495 The time and place of the Eskimo-Aleut founder event remains uncertain. Under our
496 demographic model, divergence of the lineage leading to Eskimo-Aleut speakers was dated
497 at ~3,500 calBP, and involved gene flow from a northern First American population
498 distantly related to Athabaskans (Fig. 5). There is no clear archaeological evidence for a
499 First American back-migration to Chukotka^{16,49}, so this admixture event may have occurred
500 in North America. A parsimonious explanation is that the Asian ancestral population
501 contributing to Eskimo-Aleut speakers may have remained in Chukotka after splitting from
502 the Paleo-Eskimo lineage *sensu stricto*, and that members of this lineage later separated
503 from the ancestors of Chukotko-Kamchatkan speakers and crossed the Bering Strait (Fig.
504 6b). In turn, the First American ancestral lineage that contributed to Eskimo-Aleut
505 speakers was likely located in southwestern Alaska since the Alaskan Peninsula and Kodiak
506 Island have long been suggested as a source of influences shaping the Neo-Eskimo material
507 culture^{37,50}. The earliest maritime adaptations in Beringia and America are encountered in
508 this region associated with the Ocean Bay tradition (~6,800 – 4,500 calBP)^{51,52}. Early Paleo-
509 Eskimo people used marine resources on a seasonal basis only, depended for the most part
510 on hunting caribou and muskox, and lacked sophisticated hunting gear that allowed the
511 later Inuit to become specialized in whaling⁴³. It is conceivable that a transfer of cultural
512 traits and gene flow happened simultaneously.

513
514 Where did the First American and Paleo-Eskimo-related source populations meet? A
515 succession of western Alaskan cultures, namely the Old Whaling, Choris, Norton, and
516 Ipiutak (with the earliest dates around 3,100, 2,700, 2,500, and 1,700 calBP, respectively),
517 combined cultural influences from earlier local Paleo-Eskimo sources as well as sources in
518 Chukotka and southwestern Alaska^{37,53}. Parallels between these cultures and subsequent
519 Neo-Eskimos are notable³⁷. The Old Bering Sea culture, the earliest culture assigned
520 archaeologically and genetically to Neo-Eskimos⁸, has been dated to around 2,200 calBP
521 and later^{20,54}. An individual from Uelen with the Neo-Eskimo genetic signature was dated in
522 this study at ~1,800 calBP. Considering these dates, we provisionally suggest that the
523 admixture that happened early in the history of the Neo-Eskimos may have occurred in the
524 context of the Old Whaling, Choris, or Norton cultures (Fig. 6b), although other scenarios
525 cannot be ruled out without further ancient DNA sampling. It is possible that Paleo-

526 Eskimos *sensu stricto* may have also contributed to some lesser extent to the emergence of
527 Neo-Eskimo peoples (Fig. 6b).

528
529 The descendants of proto-Paleo-Eskimos speak widely different languages, belonging to
530 the Chukotko-Kamchatkan, Eskimo-Aleut, and Na-Dene families. Based on lexicostatistical
531 studies of languages surviving in the 20th century, the time depth of the former two families
532 is likely shallow, and the Na-Dene family is probably much older, on the order of 5,000
533 years (Supplementary Information section 10). Thus, the linguistic affiliation of Paleo-
534 Eskimos is unclear. A Siberian linguistic connection was proposed for the Na-Dene family
535 under the Dene-Yeniseian hypothesis^{55,56}. This hypothetical language macrofamily unites
536 Na-Dene languages and Ket, the only surviving remnant of the Yeniseian family, once
537 widespread in South and Central Siberia^{57,58}. Perhaps consistent with this hypothesis, one
538 ancient Chukotkan sample from the Ust'-Belaya culture that was first reported in this study
539 shows evidence of ancestry from both Paleo-Eskimos and a western Siberian group related
540 to Kets. This genetic evidence suggests that links across geographic distances such as that
541 between Kets and Paleo-Eskimos may have been possible. Although the Dene-Yeniseian
542 macrofamily is not universally accepted among historical linguists^{59,60}, and correlations
543 between linguistic and genetic histories are far from perfect, evidence of a genetic
544 connection between Siberian and Na-Dene populations mediated by Paleo-Eskimos
545 suggests that future research should further explore the genealogical relationships
546 between these language families, either the closest sister-groups⁵⁶ or those within a wider
547 clade⁶⁰.

548 **Methods**

549 **Ancient DNA sampling, extraction and sequencing**

550 In dedicated clean rooms at Harvard Medical School (the 11 Aleutian Islanders and 3
551 Tochak McGrath samples), and at University College Dublin (the 3 Chukotkan samples), we
552 prepared powder from human skeletal remains, as described previously²⁶. We extracted
553 DNA using the Dabney *et al.*²³ protocol, and prepared double-stranded barcoded libraries
554 that were treated by UDG to remove characteristic cytosine to thymine damage in ancient
555 DNA using the Rohland *et al.*²⁴ protocol. We enriched the libraries for a set of
556 approximately 1.24 million SNPs²⁵, and sequenced on a NextSeq instrument using 75 nt
557 paired-end reads, which we merged before mapping to the human reference genome
558 (requiring at least 15 base pairs of overlap) (Supplementary Information section 3). We
559 also carried out shotgun sequencing of one Aleutian Islander individual (Table 1). The
560 work with ancient individuals was conducted only after consultation with local
561 communities and authorities, and after formal permissions were granted. Results have
562 been communicated in person and in writing to descendant communities.

563 **Sampling present-day populations**

564 Sampling of the Alaskan Iñupiat population (35 individuals) was performed with informed
565 consent as described in Raff *et al.*²² (see also Supplementary Information section 1). Saliva
566 samples of four West Siberian ethnic groups (Enets, Kets, Nganasans, Selkups, 58
567 individuals in total) were collected and DNA extractions were performed as described in
568 Flegontov *et al.*³⁸ (see also Supplementary Table 2). Please see ethical approval statements
569 in the respective papers^{22,38}.

570 Dataset preparation

571 To analyze rare allele sharing patterns, we composed a set of sequencing data covering
572 Africa, Europe, Southeast Asia, Siberia, and the Americas: 1,207 individuals from 95
573 populations (Supplementary Table 3). We assembled the dataset using three published
574 sources: the Simons Genome Diversity Project⁶¹, Raghavan *et al.*², and the 1000 Genomes
575 Project³². We used variant calls generated in the respective publications, kept biallelic
576 autosomal SNPs only, and applied the following filtering procedure. We first generated
577 separate masks for the Raghavan *et al.* data and for the SGDP data, based on sites at which
578 at least 90% of all individuals in those data sets have non-missing genotype calls. We then
579 used the overlap of these two masks to generate the final mask for the joint data set. Within
580 this final mask, we treated the few missing genotypes as homozygous reference calls. This
581 was necessary, since in the *rarecoal* analysis we cannot handle missing data, and justified
582 since we are analyzing rare variation, for which a missing genotype is much more likely to
583 be homozygous reference than any other genotype. For the one ancient Aleut individual for
584 which shotgun data was generated, we called variants using a method tailored to rare
585 genetic variants shared with a reference set: at every position in our reference set with
586 allele count below 10, we checked reads overlapping that position. If at least two reads
587 supported the alternative allele, we called a heterozygous genotype. In all other cases, if at
588 least two reads cover a site we called a homozygous reference allele. This method results in
589 a large false negative rate, but relative sharing ratios with reference populations should be
590 relatively unbiased³¹. When analyzing the ancient Aleut together with modern data, we
591 restricted the analysis to regions in which the Aleut sample had non-missing genotypes (i.e.
592 had at least 2x coverage).

593
594 Additionally, we assembled two independent SNP datasets: see dataset compositions in
595 Supplementary Table 3 and filtration settings in Supplementary Table 4. Initially, we
596 obtained phased autosomal genotypes for large worldwide collections of Affymetrix
597 HumanOrigins or Illumina SNP array data (Supplementary Table 4), using *Shapelt v.2.20*
598 with default parameters and without a guidance haplotype panel⁶². Then we applied
599 missing rate thresholds for individuals (<50% or <51%) and SNPs (<5%) using *PLINK*
600 *v.1.90b3.36*⁶³. For *ADMIXTURE*, *PCA*, and *qpWave/qpAdm* analyses, more relaxed missing
601 rate thresholds for individuals were applied, 75% or 70% depending on the dataset
602 (Supplementary Table 4). This allowed us to include relevant ancient samples genotyped
603 using the targeted enrichment approach (Supplementary Table 1). For the *ADMIXTURE*
604 analysis, unlinked SNPs were selected using linkage disequilibrium filtering with *PLINK*
605 (Supplementary Table 4). Ten principal components (PC) were computed using *PLINK* on
606 unlinked SNPs, and weighted Euclidean distances defined as:

$$d(q,p) = \sqrt{\frac{\lambda_1(q_1 - p_1)^2}{\sum_1^n \lambda} + \frac{\lambda_2(q_2 - p_2)^2}{\sum_1^n \lambda} + \frac{\lambda_3(q_3 - p_3)^2}{\sum_1^n \lambda} + \dots + \frac{\lambda_n(q_n - p_n)^2}{\sum_1^n \lambda}}$$

607 were calculated among individuals within populations (q_n and p_n refer to PCs from 1 to 10
608 in a population, λ_n is the corresponding eigenvalue). We removed outliers manually
609 considering the weighted Euclidean distances and results of an unsupervised
610 *ADMIXTURE*⁶⁴ analysis (K=13). Populations having on average >5% of the Siberian
611 ancestral component according to *ADMIXTURE* analysis (Extended Data Fig. 1), e.g. Finns
612 and Russians, were excluded from the European and Southeast Asian meta-populations. In
613 the case of the Illumina SNP array dataset, Na-Dene populations were exempt from PCA
614 outlier removal and from removal of supposed relatives identified by Raghavan *et al.*². This
615 was done to preserve maximal diversity of Na-Dene and to ensure that both Dakelh

616 individuals with sequencing data available would be included; our analysis is designed to
617 be robust to the presence of European admixture. Finally, we selected relevant meta-
618 populations, generating datasets of 489-1161 individuals further analyzed with
619 *ADMIXTURE*⁶⁴, PCA as implemented in *PLINK v.1.90b3.36*⁶³, *qpWave/qpAdm*^{1,26},
620 *ChromoPainter v.1* and *fineSTRUCTURE*²⁹, *ChromoPainter v.2* and *GLOBETROTTER*³⁰
621 (Supplementary Tables 3 and 4). For the *qpWave/qpAdm* analyses^{1,26}, any American
622 individuals with >1% European, African, or Southeast Asian ancestry according to
623 *ADMIXTURE* (Extended Data Fig. 1) were removed, as well as Chukotkan and Kamchatkan
624 individuals with >1% European or African ancestry. Some additional Chipewyan and West
625 Greenlandic Inuit individuals were removed since “cryptic” European ancestry
626 undetectable with *ADMIXTURE* was revealed in them using *D* statistics (Yoruba or Dai,
627 Icelander; Chipewyan individual, Karitiana) and (Yoruba or Dai, Slovak; West Greenlandic
628 Inuit individual, Karitiana). Any individual with any of the two $|Z|$ -scores >3 was removed.
629 The dataset pruning procedure is illustrated on PCA plots presented in Fig. 2, Extended
630 Data Fig. 3, and Supplementary Information section 4.

631 **ADMIXTURE analysis**

632 The *ADMIXTURE* software⁶⁴ implements a model-based Bayesian approach that uses a
633 block-relaxation algorithm in order to compute a matrix of ancestral population fractions
634 in each individual (*Q*) and infer allele frequencies for each ancestral population (*P*). A given
635 dataset is usually modelled using various numbers of ancestral populations (*K*). We ran
636 *ADMIXTURE* on HumanOrigins-based and Illumina-based datasets of unlinked SNPs
637 (Supplementary Table 4) using 10 to 25 and 5 to 20 *K* values, respectively. One hundred
638 analysis iterations were generated with different random seeds. The best run was chosen
639 according to the highest likelihood. An optimal value of *K* was selected using 10-fold cross-
640 validation.

641 **Principal component analysis (PCA)**

642 PCA was performed using *PLINK v.1.90b3.36*⁶³ with default settings. No pruning of linked
643 SNPs was applied prior to this analysis (Supplementary Table 4), and almost identical
644 results were obtained on pruned datasets.

645 **Admixture modelling with qpWave and qpAdm**

646 We used the *qpWave* tool (a part of *AdmixTools*) to infer how many of streams of ancestry
647 relate a set of test populations to a set of outgroups¹. *qpWave* relies on a matrix of statistics
648 $f_4(\text{test}_1, \text{test}_i; \text{outgroup}_1, \text{outgroup}_x)$. Usually, a few test populations from a certain region
649 and a diverse worldwide set of outgroups (having no recent gene flow from the test region)
650 are co-analyzed^{3,26,65}, and a statistical test is performed to determine whether allele
651 frequencies in the test populations can be explained by one, two, or more streams of
652 ancestry derived from the outgroups. If a group of three populations, a triplet, is derived
653 from two ancestry streams according to a *qpWave* test, and any pair of the constituent
654 populations shows the same result, it follows that one of the populations can be modelled
655 as having ancestry from the other two using another tool, *qpAdm*, which makes the implicit
656 assumption that the two populations used as the sources have not undergone admixture²⁶.
657

658 The following sets of outgroup populations were used for analyses on the HumanOrigins
659 dataset: 1) “9 Asians”, 8 diverse Siberian populations (Nganasan, Tuvinian, Ulchi, Yakut,
660 Even, Ket, Selkup, Tubalar) and a Southeast Asian population (Dai); 2) “19 outgroups” from
661 five broad geographical regions: Mbuti, Taa, Yoruba (Africans), Nganasan, Tuvinian, Ulchi,
662 Yakut (East Siberians), Altaian, Ket, Selkup, Tubalar (West Siberians), Czech, English,

663 French, North Italian (Europeans), Dai, Miao, She, Thai (Southeast Asians); 3) “9 Asians +
664 Koryak”, 8 Siberian populations, Dai, and Koryak, a close outgroup for Americans that
665 should provide higher resolution. The following sets of outgroup populations were used for
666 analyses on the Illumina dataset: 1) “10 Asians”, 9 Siberian populations (Buryat, Dolgan,
667 Evenk, Nganasan, Tuvianian, Yakut, Altaian, Khakas, Selkup) and Dai; 2) “20 outgroups”:
668 Bantu (Kenya), Mandenka, Mbuti, Yoruba (Africans), Buryat, Evenk, Nganasan, Tuvianian,
669 Yakut (East Siberians), Altaian, Khakas, Selkup (West Siberians), Basque, Sardinian, Slovak,
670 Spanish (Europeans), Dai, Lahu, Miao, She (Southeast Asians); 3) “10 Asians + Koryak”, 9
671 Siberian populations, Dai, and Koryak. All possible triplets of the form (First American or
672 Na-Dene population; Eskimo-Aleut population; Paleo-Eskimo population) and (First
673 American or Na-Dene pop.; Eskimo-Aleut pop.; Chukotko-Kamchatkan pop.) and
674 quadruplets of the form (First American pop.; Na-Dene pop.; Eskimo-Aleut pop.; Paleo-
675 Eskimo pop.) were tested with *qpWave* on both the HumanOrigins and Illumina SNP array
676 datasets, with or without transition polymorphisms, and using three alternative outgroup
677 sets. Paleo-Eskimos were represented by the Saqqaq or Late Dorset individuals, or by these
678 two individuals combined. For admixture inference with *qpAdm*, all possible triplets of the
679 form (any American, Chukotkan or Kamchatkan pop.; Paleo-Eskimo pop.; Guarani,
680 Karitiana, or Mixe) were considered in the HumanOrigins dataset, and all possible triplets
681 of the form (any American, Chukotkan or Kamchatkan pop.; Paleo-Eskimo pop.; Karitiana,
682 Mixtec, Nisga’a, or Pima) were considered in the Illumina dataset. Paleo-Eskimos were
683 represented by the Saqqaq individual or by the Saqqaq and Late Dorset individuals
684 combined.

685 ***fineSTRUCTURE* clustering**

686 We used *fineSTRUCTURE* v.2.0.7 with default parameters to analyze the output of
687 *ChromoPainter* v.1²⁹. Clustering trees of individuals were generated by *fineSTRUCTURE*
688 based on counts of shared haplotypes²⁹, and two independent iterations of the clustering
689 algorithm were performed. The clustering trees and coancestry matrices were visualized
690 using *fineSTRUCTURE GUI* v.0.1.0²⁹.

691 **Haplotype sharing statistics**

692 The Haplotype Sharing Statistic (HSS_{AB}) is defined as the total genetic length of DNA (in cM)
693 that a given individual A shares with individual B_j under the model^{29,30}. HSS_{AB} was
694 computed in the all vs. all manner by *ChromoPainter* v.1²⁹ running with default parameters,
695 and in practice we summed up the length of DNA that individual A copied from individual
696 B_j and the length of DNA copied in the opposite direction (from B_j to A), i.e. we disregarded
697 the donor/recipient distinction introduced by the *ChromoPainter* software. For each
698 individual A (in practice an American individual), HSS_{AB} values were averaged across all
699 individuals of a reference population B (the Siberian or Arctic meta-population, or the
700 Saqqaq ancient genome⁷), and then normalized by the haplotype sharing statistic HSS_{AC} for
701 the European, African, or Siberian outgroup C . The resulting statistics HSS_{AB}/HSS_{AC} are
702 referred to as Siberian, Arctic, or Saqqaq relative haplotype sharing, and were visualized
703 for separate individuals. Similar statistics were calculated for Siberian and Arctic
704 individuals using the leave-one-out procedure. Relative HSSs for recently admixed
705 populations, with ancestry from population A and population B , were calculated in the
706 following way: $a \times HSS_{AC}/HSS_{AD} + b \times HSS_{BC}/HSS_{BD}$, where a and b are admixture proportions
707 being simulated.

708 Dating admixture events using haplotype sharing statistics

709 We used *GLOBETROTTER*³⁰ to infer and date up to two admixture events in the history of
710 Na-Dene populations. To detect subtle signals of admixture between closely related source
711 populations, we followed the ‘regional’ analysis protocol of Hellenthal *et al.*³⁰ Using
712 *ChromoPainter v.2*³⁰, chromosomes of a target Na-Dene population were ‘painted’ as a
713 mosaic of haplotypes derived from donor populations or meta-populations: the Saqqaq
714 ancient genome, Chukotko-Kamchatkan groups, Eskimo-Aleuts, northern First Americans,
715 southern First Americans, West Siberians, East Siberians, Southeast Asians, and Europeans.
716 Target individuals were considered as haplotype recipients only, while other populations
717 or meta-populations were considered as both donors and recipients. That is different from
718 the *ChromoPainter v.1* approach, where all individuals were considered as donors and
719 recipients of haplotypes at the same time, and only self-copying was forbidden.

720
721 Painting samples for the target population and ‘copy vectors’ for other (meta)populations
722 called ‘surrogates’ served as an input of *GLOBETROTTER*, which was run according to
723 section 6 of the instruction manual of May 27, 2016. The following settings were used: no
724 standardizing by a “NULL” individual (null.ind 0); five iterations of admixture date and
725 proportion/source estimation (num.mixing.iterations 5); at each iteration, any surrogates
726 that contributed $\leq 0.1\%$ to the target population were removed (props.cutoff 0.001); the x-
727 axis of coancestry curves spanned the range from 0 to 50 cM (curve.range 1 50), with bins
728 of 0.1 cM (bin.width 0.1). Confidence intervals (95%) for admixture dates were calculated
729 based on 100 bootstrap replicates. Alternatively, when using separate populations as
730 haplotype donors, the setting ‘standardizing by a “NULL” individual’ was turned on to take
731 account for potential bottleneck effects. A generation time of 29 years was used in all
732 dating calculations².

733
734 The *GLOBETROTTER* software is able to date no more than two admixture events³⁰, and we
735 therefore had to reduce the complexity of original Na-Dene populations that likely
736 experienced more than two major waves of admixture. For that purpose, only a subset of
737 Na-Dene individuals was used for the *GLOBETROTTER* analysis: those with prior evidence
738 of elevated Paleo-Eskimo ancestry (Supplementary Information section 6) and with <10%
739 West Eurasian ancestry estimated with *ADMIXTURE* (Extended Data Fig. 1).

740 Rare allele sharing statistics

741 We define the Rare Allele Sharing Statistic ($RASS_{AB}$) as the average number of sites at which
742 an individual *A* shares a derived allele of frequency *k* with an individual from population *B*:

743

$$RASS_{A,B}(k) = \frac{1}{4n_B} \sum_i d_{A,i} d_{B,i} \delta_{D_i,k}$$

744

745 where n_B the number of individuals in population *B*, $d_{A,i}$ stands for the number of derived
746 alleles at site *i* in individual *A*, and the term $\delta_{x,y}$ equals 0 if the total count of derived alleles
747 in the dataset does not equal *k*, and is 1 otherwise. The sum across all sites *i* is normalized
748 by the product of population sizes multiplied by four to give the average number of shared
749 alleles between two randomly drawn haploid chromosome sets. Instead of counting
750 derived alleles, in practice we counted non-reference alleles, which should not make a
751 difference for low frequencies. To take care of variability in genome coverage across
752 individuals and of dataset-specific SNP calling biases, we calculated normalized (or relative)

753 RASS, dividing $RASS_{AB}$ by $RASS_{AC}$, where population C is a distant outgroup. Standard
754 deviation of $RASS_{AB}$ was calculated with a jackknife approach. Specifically, we re-estimated
755 RASS per one million base pairs for drop-one-out data sets excluding an entire
756 chromosome each time. We then used the weighted jackknife method⁶⁶ to estimate sample
757 variances across the drop-one-out data sets. The standard deviation of normalized RASS
758 was calculated using error propagation via partial derivatives:

759

$$\Delta RASS_{A,B}/RASS_{A,C}(k) = \sqrt{\left(\frac{\Delta RASS_{A,B}(k)}{RASS_{A,C}(k)}\right)^2 + \left(\frac{RASS_{A,B}(k)}{RASS_{A,C}(k)^2} \Delta RASS_{A,C}(k)\right)^2}$$

760
761 In practice, individual A was a present-day or ancient American, population B was
762 represented by Siberian or Arctic meta-populations or by the Saqqaq ancient genome⁷, and
763 population C – by Africans or Europeans (Supplementary Table 3). The resulting statistics
764 are referred to as relative Siberian, Arctic or Saqqaq allele sharing. Similar statistics were
765 calculated for Siberian and Arctic individuals using the leave-one-out procedure. The same
766 statistics were calculated on a dataset without transition polymorphisms. The ancient
767 Aleut and Saqqaq ancient genomes were not included into the Arctic reference meta-
768 population. Relative RASS for recent mixtures of individual A and individual B were
769 calculated in the following way: $a \times RASS_{AC}/RASS_{AD} + b \times RASS_{BC}/RASS_{BD}$, where a and b are
770 admixture proportions being simulated.

771 Demographic modelling

772 We used the *qpGraph* method⁴² to explore models that are consistent with F statistics and
773 arrived at a final model connecting 8 groups: Mbuti, French, Ami, Mixe, Even, Yupik Naukan,
774 Koryak and Chipewyan (discussed in Supplementary Information section 9). We then used
775 the *rarecoal* program³¹ (<https://github.com/stschiff/rarecoal>) to derive a timed admixture
776 graph for meta-populations (Fig. 5 and Supplementary Information section 9). We started
777 with a tree connecting Europeans, Southeast Asians, and southern First Americans into a
778 simple tree without admixture, and used “*rarecoal mcmc*” to infer maximum likelihood
779 branch population sizes and split times. We then iteratively added Core Siberians,
780 Chukotko-Kamchatkan, Eskimo-Aleut, and Northern Athabaskan speakers. After each
781 addition, we re-optimized the tree and inspected the fits of the model to the data. When we
782 saw a significant deviation between model and data for a particular pairwise allele sharing
783 probability, we added admixture edges (Supplementary Information section 9). After
784 *rarecoal*'s inference, we rescaled time and population size parameters to years and real
785 effective population size using a mutation rate of 1.25×10^{-8} per site per generation, and a
786 generation time of 29 years². We finally tested whether our final model (Fig. 5a) was
787 consistent also with F statistics using *qpGraph* (Supplementary Information section 9). In
788 order to map the two ancient genomes, Saqqaq and Aleut, we used “*rarecoal find*” to
789 explore a set of possible split points of the ancient lineage on the tree, distributed across all
790 branches and times. Here we restricted the analysis to variants between allele counts 2 and
791 4. We excluded singletons to reduce impact of false positive genotyping calls³¹.

792 References

- 793 1. Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370–
794 374 (2012).

- 795 2. Raghavan, M. *et al.* Genomic evidence for the Pleistocene and recent population
796 history of Native Americans. *Science* **349**, 1–20 (2015).
- 797 3. Skoglund, P. *et al.* Genetic evidence for two founding populations of the Americas.
798 *Nature* **525**, 104–108 (2015).
- 799 4. Lindo, J. *et al.* Ancient individuals from the North American Northwest Coast reveal
800 10,000 years of regional genetic continuity. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4093–
801 4098 (2017).
- 802 5. Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial
803 site in western Montana. *Nature* **506**, 225–229 (2014).
- 804 6. Krauss, M. Na-Dene. *Native languages of the Americas, vol. 1*, ed. Sebeok, T. A. New
805 York & London: Plenum Press. 283–358 (1976).
- 806 7. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo.
807 *Nature* **463**, 757–762 (2010).
- 808 8. Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345**,
809 1255832 (2014).
- 810 9. Friesen, T. M. Pan-Arctic population movements: the early Paleo-Inuit and Thule
811 Inuit migrations. *The Oxford Handbook of the Prehistoric Arctic*, ed. Friesen, T. M.,
812 Mason, O. K. New York: Oxford University Press. 673–692 (2016).
- 813 10. Szathmary, E. J. E. & Ossenberg, N. S. Are the biological differences between North
814 American Indians and Eskimos truly profound? *Curr. Anthropol.*, **19**, 673–701 (1978).
- 815 11. Park, R. W. The Dorset-Thule transition. *The Oxford Handbook of the Prehistoric*
816 *Arctic*, ed. Friesen, T. M., Mason, O. K. New York: Oxford University Press. 417–442
817 (2016).
- 818 12. Prentiss, A. M., Walsh, M. J., Foor, T. A. & Barnett, K. D. Cultural macroevolution
819 among high latitude hunter-gatherers: a phylogenetic study of the Arctic Small Tool
820 tradition. *J. Archaeol. Sci.* **59**, 64–79 (2015).
- 821 13. Tremayne, A. H. & Rasic, J. T. The Denbigh Flint Complex of Northern Alaska. *The*
822 *Oxford Handbook of the Prehistoric Arctic*, ed. Friesen, T. M., Mason, O. K. New York:
823 Oxford University Press. 303–322 (2016).
- 824 14. Friesen, T. M. Contemporaneity of Dorset and Thule cultures in the North American
825 Arctic: new radiocarbon dates from Victoria Island, Nunavut. *Curr. Anthropol.* **45**,
826 685–691 (2004).
- 827 15. Friesen, T. M. & Arnold, C. D. The timing of the Thule migration: new dates from the
828 western Canadian Arctic. *Am. Antiq.* **73**, 527–538 (2008).
- 829 16. Potter, B. A. Archaeological patterning in Northeast Asia and Northwest North
830 America: an examination of the Dene-Yeniseian hypothesis. *The Dene-Yeniseian*
831 *Connection*, ed. Kari, J., Potter, B. A. *Anthropological Papers of the University of Alaska:*
832 *New Series* **5**, 138–167 (2010).
- 833 17. Powers, W. R. & Jordan, R. H. Human biogeography and climate change in Siberia and
834 arctic North America in the fourth and fifth millennia B.P. *Philos. Trans. R. Soc. Lond.*
835 *A* **330**, 665–670 (1990).
- 836 18. Dumond, D. E. & Bland, R. L. Holocene prehistory of the northernmost North Pacific. *J.*
837 *World Prehist.* **9**, 401–451 (1995).
- 838 19. Friesen, T. M. On the naming of Arctic archaeological traditions: The case for Paleo-
839 Inuit. *Arctic* **68**, iii–iv (2015).
- 840 20. Mason, O. K. The Old Bering Sea florescence about Bering Strait. *The Oxford*
841 *Handbook of the Prehistoric Arctic*, ed. Friesen, T. M., Mason, O. K. New York: Oxford
842 University Press. 417–442 (2016).

- 843 21. Dumond, D. E. *The Eskimos and Aleuts (Vol. 180)*. London: Thames and Hudson
844 (1987).
- 845 22. Raff, J. A, Rzhetskaya, M., Tackney, J. & Hayes, M. G. Mitochondrial diversity of Iñupiat
846 people from the Alaskan North Slope provides evidence for the origins of the Paleo-
847 and Neo-Eskimo peoples. *Am. J. Phys. Anthropol.* **157**, 603–614 (2015).
- 848 23. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene
849 cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.*
850 **110**, 15758–15763 (2013).
- 851 24. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-
852 glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol.*
853 *Sci.* **370**, 20130624 (2015).
- 854 25. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal
855 ancestor. *Nature* **524**, 216–219 (2015).
- 856 26. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European
857 languages in Europe. *Nature* **522**, 207–211 (2015).
- 858 27. Leslie, S. *et al.* The fine-scale genetic structure of the British population. *Nature* **519**,
859 309–314 (2015).
- 860 28. Busby, G. B. *et al.* The role of recent admixture in forming the contemporary West
861 Eurasian genomic landscape. *Curr. Biol.* **25**, 2518–2526 (2015).
- 862 29. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure
863 using dense haplotype data. *PLoS Genet.* **8**, 11–17 (2012).
- 864 30. Hellenthal, G. *et al.* A genetic atlas of human admixture. *Science* **343**, 747–751 (2014).
- 865 31. Schiffels, S. *et al.* Iron Age and Anglo-Saxon genomes from East England reveal
866 British migration history. *Nat. Commun.* **7**, 10408 (2016).
- 867 32. 1000 Genomes Project Consortium. A global reference for human genetic variation.
868 *Nature* **526**, 68–74 (2015).
- 869 33. Hrdlička, A. *The Aleutian and Commander Islands and their inhabitants*. Philadelphia:
870 Wistar Institute of Anatomy and Biology (1945).
- 871 34. Brenner Coltrain, J., Hayes, M. G. & O'Rourke, D. H. Hrdlička's Aleutian
872 population-replacement hypothesis. A radiometric evaluation. *Curr. Anthropol.* **47**,
873 537–548 (2006).
- 874 35. Smith, S. E. *et al.* Inferring population continuity versus replacement with aDNA: a
875 cautionary tale from the Aleutian Islands. *Hum. Biol.* **81**, 407–426 (2009).
- 876 36. Dikov, N. N. *Drevnie kul'tury Severo-Vostochnoi Azii: Aziia na styke s Amerikoi v*
877 *drevnosti*. Moscow: Nauka (1979). Translated by Bland, R. L. as *Early cultures of*
878 *Northeastern Asia*. Anchorage: US Department of the Interior, National Park Service,
879 Shared Beringian Heritage Program (2004).
- 880 37. Dumond, D. E. Norton hunters and fisherfolk. *The Oxford Handbook of the Prehistoric*
881 *Arctic*, ed. Friesen, T. M., Mason, O. K. New York: Oxford University Press. 395–416
882 (2016).
- 883 38. Flegontov, P. *et al.* Genomic study of the Ket: A Paleo-Eskimo-related ethnic group
884 with significant ancient North Eurasian ancestry. *Sci. Rep.* **6**, 20768 (2016).
- 885 39. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of
886 Native Americans. *Nature* **505**, 87–91 (2014).
- 887 40. Scally, A. & Durbin, R. Revising the human mutation rate: implications for
888 understanding human evolution. *Nat. Rev. Genet.* **13**, 745–753 (2012).
- 889 41. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in
890 genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423
891 (2005).

- 892 42. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093
893 (2012).
- 894 43. Hoffecker J. F. *A Prehistory of the North: human settlement of the higher latitudes.*
895 Rutgers University Press (2004).
- 896 44. Moss, M. L., Erlandson, J. M. & Stuckenrath, R. The antiquity of Tlingit settlement on
897 Admiralty Island, southeast Alaska. *Am. Antiq.* **54**, 534–543 (1989).
- 898 45. Ives, J. W. Alberta, Athapaskans and Apachean origins. *Archaeology in Alberta, A view*
899 *from the new millennium*, Brink, J. W., Dormaar, J. F. Medicine Hat, Alberta: The
900 Archaeological Society of Alberta. 256–289 (2003).
- 901 46. Ives, J. W. Dene-Yeniseian, migration, and prehistory. *The Dene-Yeniseian Connection*,
902 ed. Kari, J., Potter, B. A. *Anthropological Papers of the University of Alaska: New Series*
903 **5**, 324–334 (2010).
- 904 47. Matson, R. G. & Magne, M. P. R. *Athapaskan migrations: the archaeology of Eagle Lake,*
905 *British Columbia.* Tuscon: University of Arizona Press. (2007).
- 906 48. Jensen, B. J. L. *et al.* Transatlantic distribution of the Alaskan White River Ash.
907 *Geology* **42**, 875–878 (2014).
- 908 49. Hoffecker, J. F. & Elias, S. A. *Human Ecology of Beringia.* New York: Columbia
909 University Press (2007)
- 910 50. Ackerman, R. E. Early maritime traditions in the Bering, Chukchi, and East Siberian
911 seas. *Arctic Anthropol.* **35**, 247–262 (1998).
- 912 51. Fitzhugh, B. The origins and development of Arctic maritime adaptations in the
913 Subarctic and Arctic Pacific. *The Oxford Handbook of the Prehistoric Arctic*, ed. Friesen,
914 T. M., Mason, O. K. New York: Oxford University Press. 253–278 (2016).
- 915 52. Steffian, A., Saltonstall, P. & Yarborough, L. F. Maritime economies of the central Gulf
916 of Alaska after 4000 B.P. *The Oxford Handbook of the Prehistoric Arctic*, ed. Friesen, T.
917 M., Mason, O. K. New York: Oxford University Press. 303–322 (2016).
- 918 53. Darwent, C. M. & Darwent, J. The enigmatic Choris and Old Whaling cultures of the
919 Western Arctic. *The Oxford Handbook of the Prehistoric Arctic*, ed. Friesen, T. M.,
920 Mason, O. K. New York: Oxford University Press. 371–394 (2016).
- 921 54. Bronshtein, M. M., Dneprovsky, K. A. & Savintsky, A. B. Ancient Eskimo cultures of
922 Chukotka. *The Oxford Handbook of the Prehistoric Arctic*, ed. Friesen, T. M., Mason, O.
923 K. New York: Oxford University Press. 469–488 (2016).
- 924 55. Ruhlen, M. The origin of the Na-Dene. *Proc. Natl. Acad. Sci. USA* **95**, 13994–13996
925 (1998).
- 926 56. Vajda, E. J. Siberian link with Na-Dene languages. *The Dene-Yeniseian Connection*, ed.
927 Kari, J., Potter, B. A. *Anthropological Papers of the University of Alaska: New Series* **5**,
928 33–99 (2010).
- 929 57. Dul'zon, A. P. Byloe rasselenie Ketov po dannym toponimiki [The former settlement
930 of the Kets according to the facts of toponymy]. *Voprosy Geografii* **68**, 50–84 (1962).
- 931 58. Vajda, E. J. Loanwords in Ket. *The Typology of Loanwords*, ed. Haspelmath, M.,
932 Tadmoo, U. Oxford: Oxford University Press, 125–139 (2009).
- 933 59. Campbell, L. Review of 'The Dene-Yeniseian Connection', ed. by James Kari and Ben A.
934 Potter. *Int. J. Am. Linguistics* **77**, 445–451 (2011).
- 935 60. Starostin, G. Dene-Yeniseian: a critical assessment. *J. Language Relationship* **8**,
936 117–138 (2012).

937

938

Additional references for Methods and figure legends:

939

- 940 61. Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse
941 populations. *Nature* **538**, 201–206 (2016).
- 942 62. O'Connell, J. *et al.* A general approach for haplotype phasing across the full spectrum
943 of relatedness. *PLoS Genet.* **10**, e1004234 (2014).
- 944 63. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based
945 linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 946 64. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in
947 unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- 948 65. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East.
949 *Nature* **536**, 419–424 (2016).
- 950 66. Busing, F. M. T. A., Meijer, E. & van der Leeden, R. Delete-m jackknife for unequal m.
951 *Stat. Comput.* **9**, 3–8 (1999).
- 952 67. Kari, J. The concept of geolinguistic conservatism in Na-Dene prehistory. *The Dene-*
953 *Yeniseian Connection*, ed. Kari, J., Potter, B. A. *Anthropological Papers of the University*
954 *of Alaska: New Series* **5**, 194–222 (2010).
- 955 68. Ives, J. W., Froese, G. F., Janetski, J. C., Brock, F. & Ramsey, C. B. A high resolution
956 chronology for Steward's Promontory culture collections, Promontory Point, Utah.
957 *Am. Antiq.* **79**, 616–637 (2014).
- 958

959 **Supplementary Information** is available in the online version of the paper

960

961 **Acknowledgements**

962 We gratefully acknowledge the Aleut Corporation, the Aleut/Pribilof Island Association,
963 and the Chaluka Corporation for granting permissions to conduct genetic analyses on the
964 eastern Aleutian remains to help elucidate the population history of the region. We also
965 thank the staff at the Smithsonian Institution's National Museum of Natural History for
966 facilitating the sample collection. Sample collection and the initial molecular, isotopic and
967 AMS ¹⁴C dating of the samples described here were funded by National Science Foundation
968 Office of Polar Program grants OPP-9726126, OPP-9974623, and OPP-0327641, by the
969 Natural Sciences and Engineering Research Council of Canada, and the Wenner-Gren
970 Foundation for Anthropological Research (#6364). We are also grateful to the McGrath
971 Native Village Council and MTNT Ltd. for granting permissions to conduct genetic analyses
972 on the Tochak McGrath remains, and to Jamie Clark, who performed biological age
973 estimates on these remains. We thank the research participants in Alaska who donated
974 samples for genome-wide analysis. We are grateful to all researchers that shared their data:
975 Maanasa Raghavan, Simon Rasmussen, Eske Willerslev, and Joan Brenner Coltrain. We also
976 acknowledge valuable advice on American archaeology from Ben A. Potter, John W. Ives
977 and T. Max Friesen. We thank Justin Tackney, Lauren Norman, and Kim TallBear for
978 comments on earlier drafts of this paper. P.F. and E.A. were supported by the Institutional
979 Development Program of the University of Ostrava and by EU structural funding
980 Operational Programme Research and Development for Innovation, project No.
981 CZ.1.05/2.1.00/19.0388. P.C. was supported by the grant no. 0924/2016/ŠaS from the
982 Statutory City of Ostrava and by the grant no. 01211/2016/RRC 'Strengthening
983 international cooperation in science, research and education' from the Moravian-Silesian
984 Region. D.R. was funded by NSF HOMINID grant BCS-1032255, NIH (NIGMS) grant
985 GM100233, and is an Investigator of the Howard Hughes Medical Institute. D.A.B. was
986 supported by a Norman Hackerman Advanced Research Program (NHARP) grant from the

987 Texas Higher Education Coordinating Board (THECB). AMS ¹⁴C work at Pennsylvania State
988 University was funded by the NSF Archaeometry program award BCS-1460369 to D.J.K.
989

990 **Author contributions**

991 S.S., P.F., and D.R. supervised the study. A.M.K., R.A.S., S.V., E.V., D.H.O'R., R.P., and D.R.
992 assembled the collection of archaeological samples. D.A.B., O.F., J.R., M.G.H., and J.K.
993 assembled the sample collection from present-day populations. T.K.H. and D.J.K. were
994 responsible for radiocarbon dating and calibration. N.R., F.C., and D.K. performed
995 laboratory work and supervised ancient DNA sequencing. P.F., N.E.A., P.C., S.M., C.J., T.C.L.,
996 I.O., P.S., and S.S., analyzed genetic data. E.J.V. wrote the supplemental section on linguistics.
997 P.F., D.R., and S.S. wrote the manuscript with additional input from all other co-authors.
998

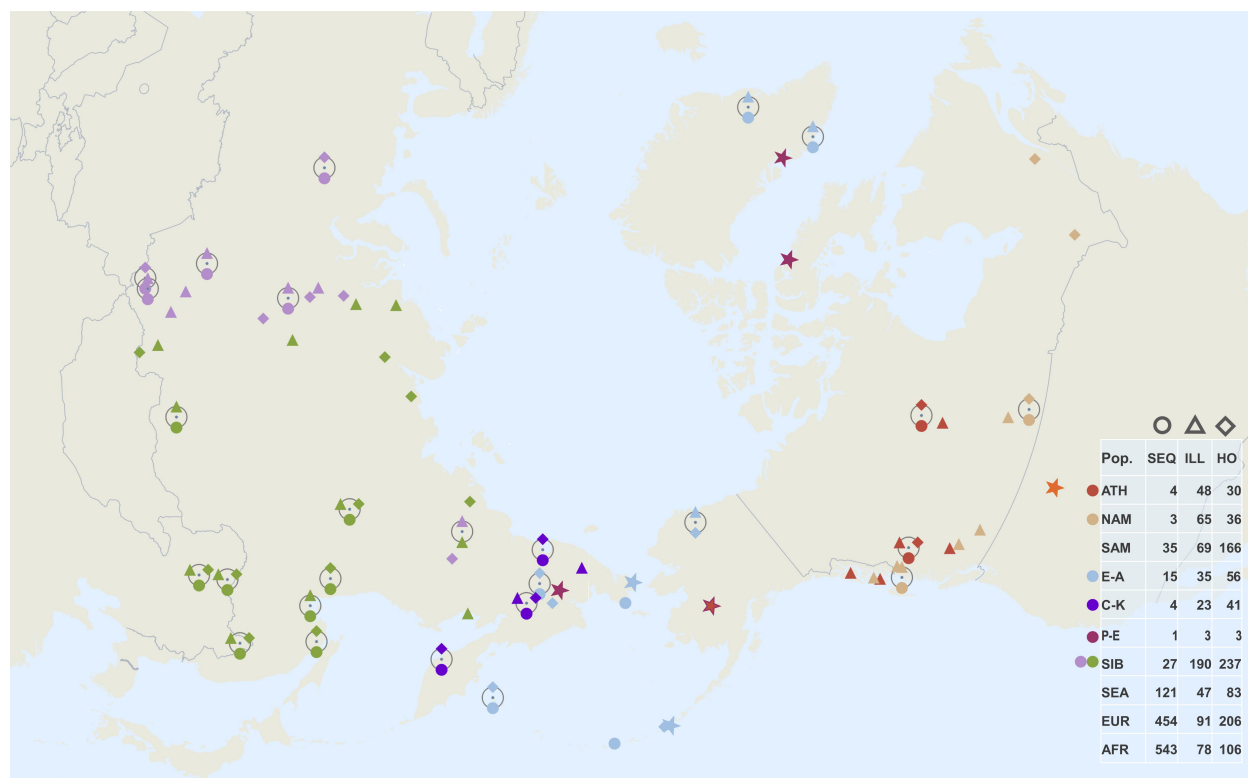
999 **Author information**

1000 Raw sequence data (bam files) from the 17 newly reported ancient individuals is available
1001 from the European Nucleotide Archive. The accession number for the sequence data
1002 reported in this paper is (to be provided prior to publication). The genotype data for the
1003 Iñupiat are obtained through informed consents that are not consistent with public posting
1004 of the data, analyses of phenotypic traits, or commercial use of the data. In order to protect
1005 the privacy of participants and ensure that their wishes with respect to data usage are
1006 followed, researchers wishing to use data from the Iñupiat samples should contact Geoffrey
1007 Hayes (ghayes@northwestern.edu) and Deborah Bolnick
1008 (deborah.bolnick@austin.utexas.edu), who can then arrange to share the data with
1009 researchers who can formally affirm that they will abide by these conditions. The newly
1010 reported SNP genotyping data for West Siberians (Enets, Ket, Nganasan, Selkup) is
1011 available to researchers who send a signed letter to J.K., P.F., and D.R. containing the
1012 following text: "(a) I will not distribute the data outside my collaboration; (b) I will not post
1013 the data publicly; (c) I will make no attempt to connect the genetic data to personal
1014 identifiers for the samples; (d) I will not use the data for commercial purposes." The
1015 programming code used in this study is available at <https://github.com/stschiff/rarecoal>.
1016 The authors declare no conflicting financial interests. Correspondence and requests for
1017 materials should be addressed to S.S. (schiffels@shh.mpg.de), P.F. (pavel.flegontov@osu.cz),
1018 and D.R. (reich@genetics.med.harvard.edu).
1019

1020 Figures

1021

1022 **Figure 1.** Geographic locations of Siberian and North American populations used in this
1023 study. Three main datasets are as follows (Supplementary Tables 3, 4): 1) a set based on
1024 the Affymetrix Human Origins genotyping array, including diploid genotypes for the
1025 ancient Saqqaq⁷ and Clovis⁵ individuals, together with SNP capture data from six ancient
1026 Aleuts who had the highest coverage, two unrelated ancient Athabaskans, two ancient
1027 Chukotkan Neo-Eskimos, and the Ust'-Belaya Chukotkan Paleo-Eskimo (Table 1); 2) a set
1028 based on various Illumina arrays, including Saqqaq and the other ancient samples, and 3) a
1029 whole genome data set of 1,207 individuals from 95 populations, including the Clovis,
1030 Saqqaq, and one ancient Aleut individual for which we generated a complete genome with
1031 2.7x coverage. The dataset composition, i.e. number of individuals in each meta-population,
1032 is shown in the table on the right. Locations of samples with whole genome sequencing
1033 data (SEQ) are shown with circles, and those of Illumina (ILL) and HumanOrigins (HO)
1034 array samples with triangles and diamonds, respectively. Meta-populations are color-
1035 coded in a similar way throughout all figures and designated as follows: Na-Dene speakers
1036 (abbreviated as ATH), other northern First Americans (NAM), southern First Americans
1037 (SAM), Eskimo-Aleut speakers (E-A), Chukotko-Kamchatkan speakers (C-K), Paleo-Eskimos
1038 (P-E), West and East Siberians (WSIB and ESIB), Southeast Asians (SEA), Europeans (EUR),
1039 and Africans (AFR). Locations of the Saqqaq, Clovis and other ancient samples are shown
1040 with asterisks colored to reflect their meta-population affiliation.
1041

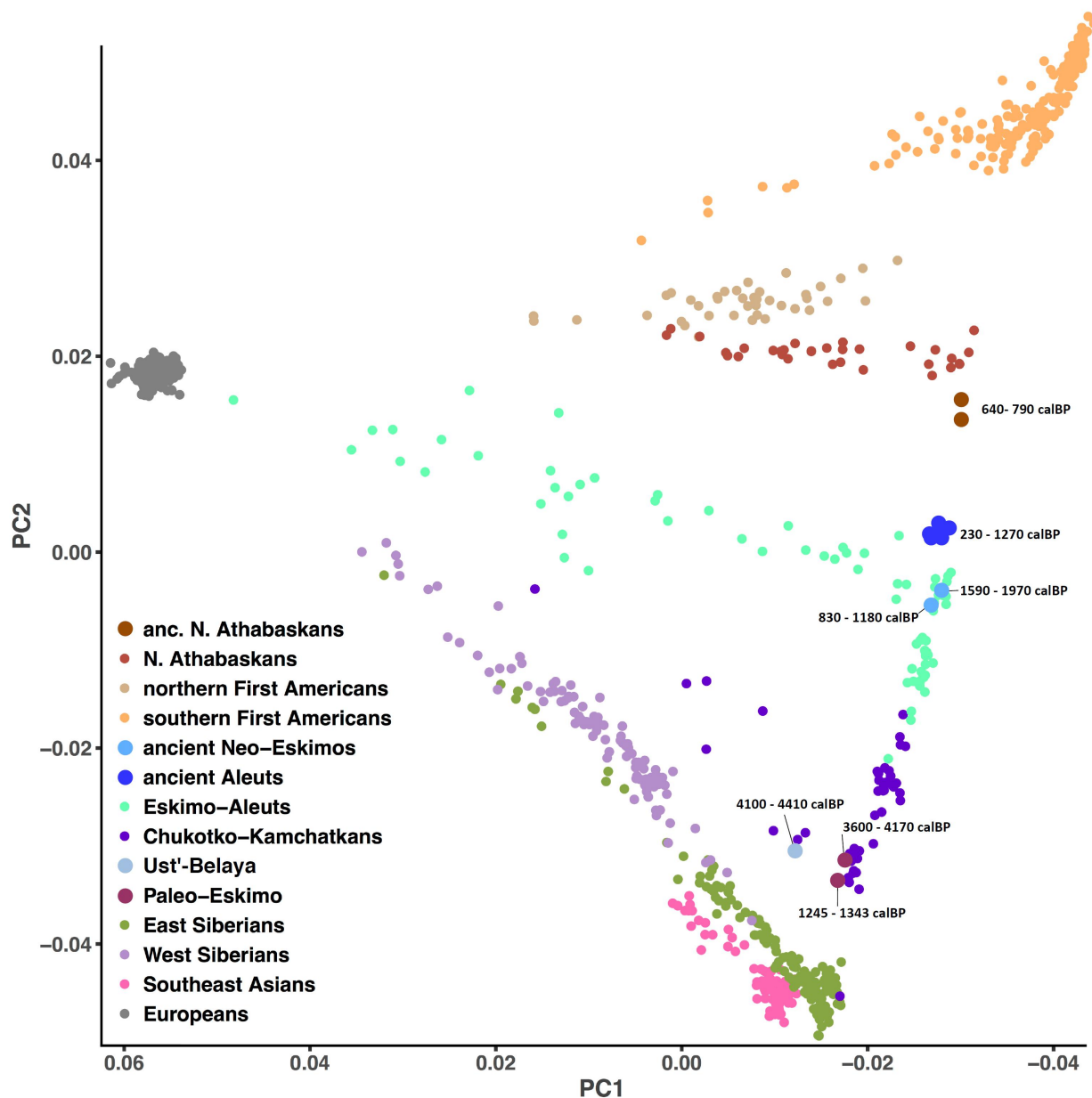


1042

1043

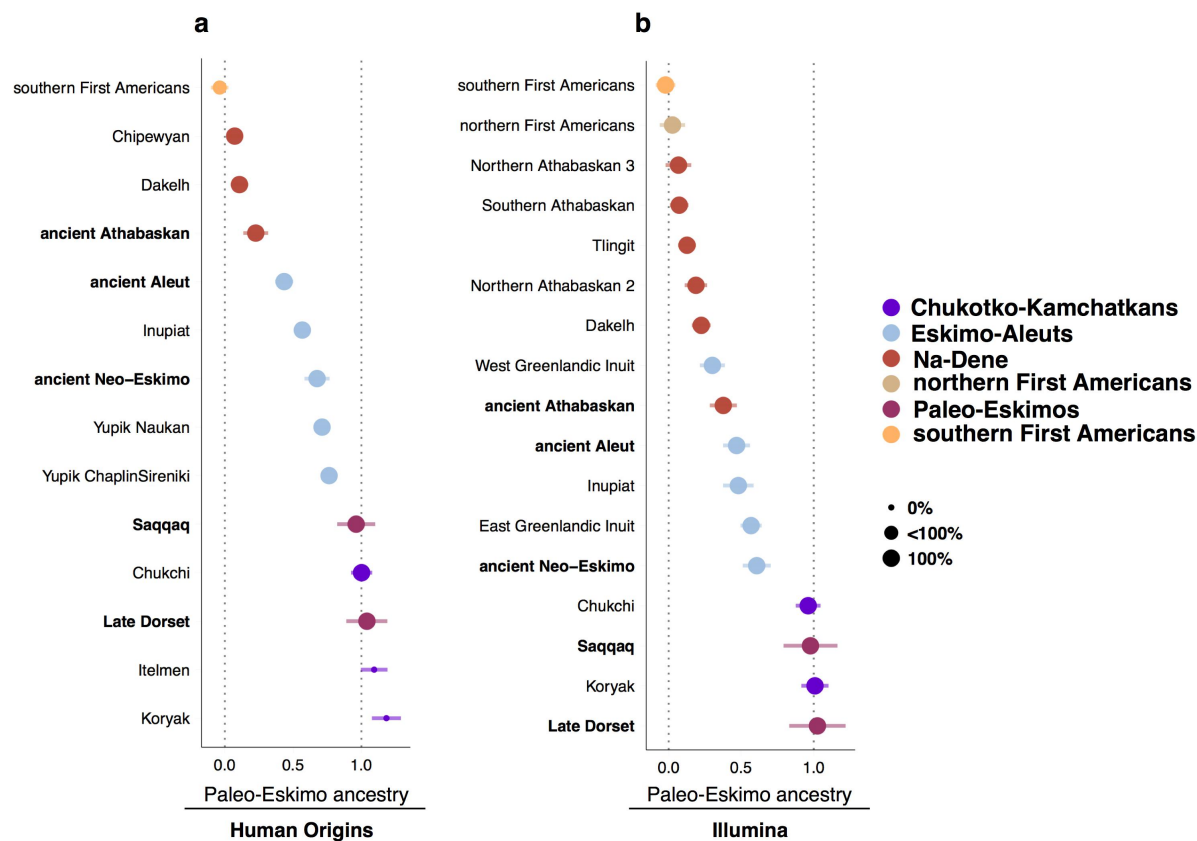
1044

1045 **Figure 2.** Principal component analysis (PCA) on the HumanOrigins datasets prior to any
1046 outlier removal. A plot of two principal components (PC1 vs. PC2) is presented. Calibrated
1047 radiocarbon dates in calBP are shown for ancient samples (large circles). For individuals,
1048 95% confidence intervals are shown, and for populations, minimal and maximal dates
1049 among all confidence intervals of that population are shown. A similar plot for the Illumina
1050 dataset is displayed in Extended Data Fig. 3, and plots for the datasets used for
1051 *qpWave/qpAdm* analyses are shown in Supplementary Information section 4. In those
1052 datasets, First American, Chukotko-Kamchatkan-speaking and Eskimo-Aleut-speaking
1053 individuals having >1% European, African, or Southeast Asian ancestry according to
1054 *ADMIXTURE* were removed.
1055



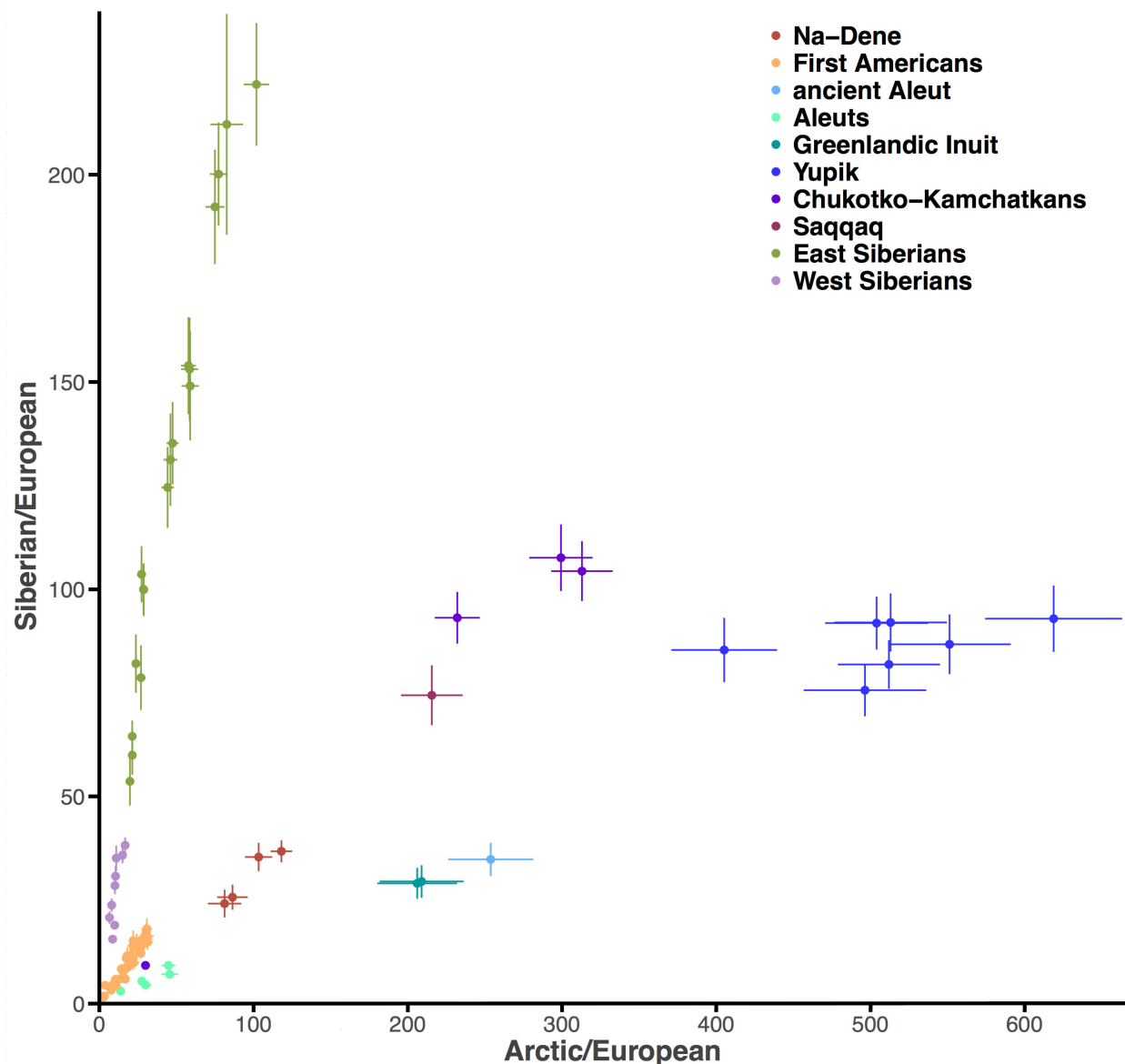
1056
1057
1058

1059 **Figure 3.** A gradient of Paleo-Eskimo ancestry in America revealed using the *qpAdm*
 1060 approach. American, Chukotkan, and Kamchatkan populations were modelled as
 1061 descended from both First American and Paleo-Eskimo sources on the HumanOrigins (a)
 1062 and Illumina (b) datasets without transition polymorphisms. First, population triplets were
 1063 tested with *qpWave* for consistency with two or three streams of ancestry derived from
 1064 outgroups. Second, *qpAdm* was used to infer admixture proportions in present-day or
 1065 ancient (in bold) target populations. Saqqaq was considered as a Paleo-Eskimo source for
 1066 all populations apart from Saqqaq itself, for which Late Dorset was used as a source, and
 1067 alternative First American sources were selected among the largest populations with little
 1068 or no detectable admixture: Mixe, Guarani, or Karitiana for the HumanOrigins dataset;
 1069 Nisga'a, Mixtec, Pima, or Karitiana for the Illumina dataset. Admixture proportions and
 1070 their standard errors were averaged across triplets including these different First
 1071 American sources, or across many alternative target populations in the case of southern
 1072 and northern First Americans. Meta-populations are color-coded according to the legend
 1073 on the right. Proportion of population triplets consistent with two migration streams is
 1074 coded by the circle size: small (0%), medium (>0% and <100%), and large (100%). The
 1075 following sets of outgroups were used: 8 diverse Siberian populations (Nganasan, Tuvinian,
 1076 Ulchi, Yakut, Even, Ket, Selkup, Tubalar) and a Southeast Asian population (Dai) on the
 1077 HumanOrigins dataset; 9 Siberian populations (Buryat, Dolgan, Evenk, Nganasan, Tuvinian,
 1078 Yakut, Altaian, Khakas, Selkup) and Dai on the Illumina dataset. See results for other
 1079 outgroup sets and for datasets including transitions in Extended Data Figs. 4 and 5.
 1080



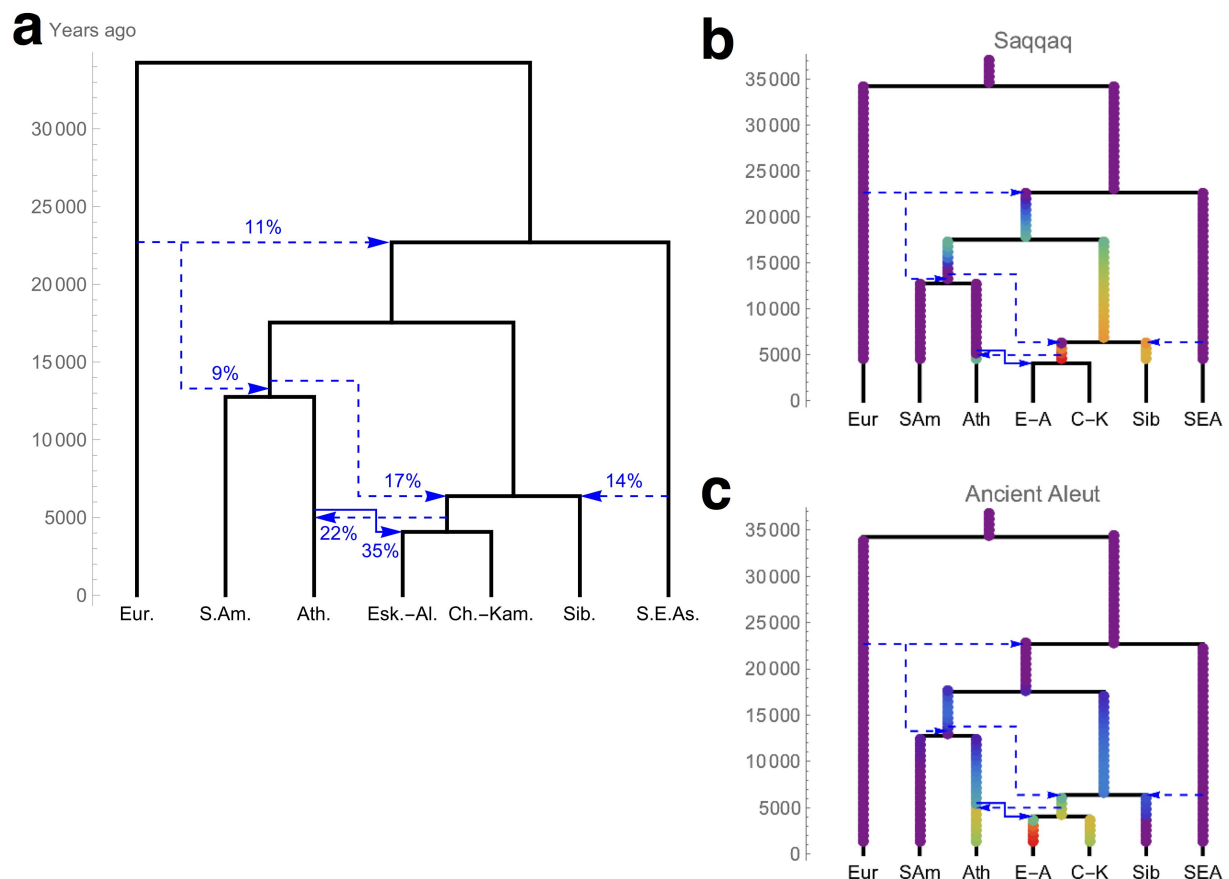
1081
 1082
 1083

1084 **Figure 4.** Relative rare allele sharing statistics calculated for each present-day or ancient
1085 American individual and the Arctic and Siberian meta-populations. Non-reference alleles
1086 occurring from 2 to 5 times in the dataset of 1,207 diploid genomes contributed to the
1087 statistics. To take care of variability in genome coverage across populations and of dataset-
1088 specific SNP calling biases, we normalized the counts of alleles shared by a given American
1089 individual and the Arctic or Siberian meta-populations by similar counts of alleles shared
1090 with Europeans. Standard deviations were calculated using a jackknife approach with
1091 chromosomes used as resampling blocks, and single standard error intervals are plotted.
1092 Plots on the dataset without transitions are shown in Extended Data Fig. 8c,d.
1093



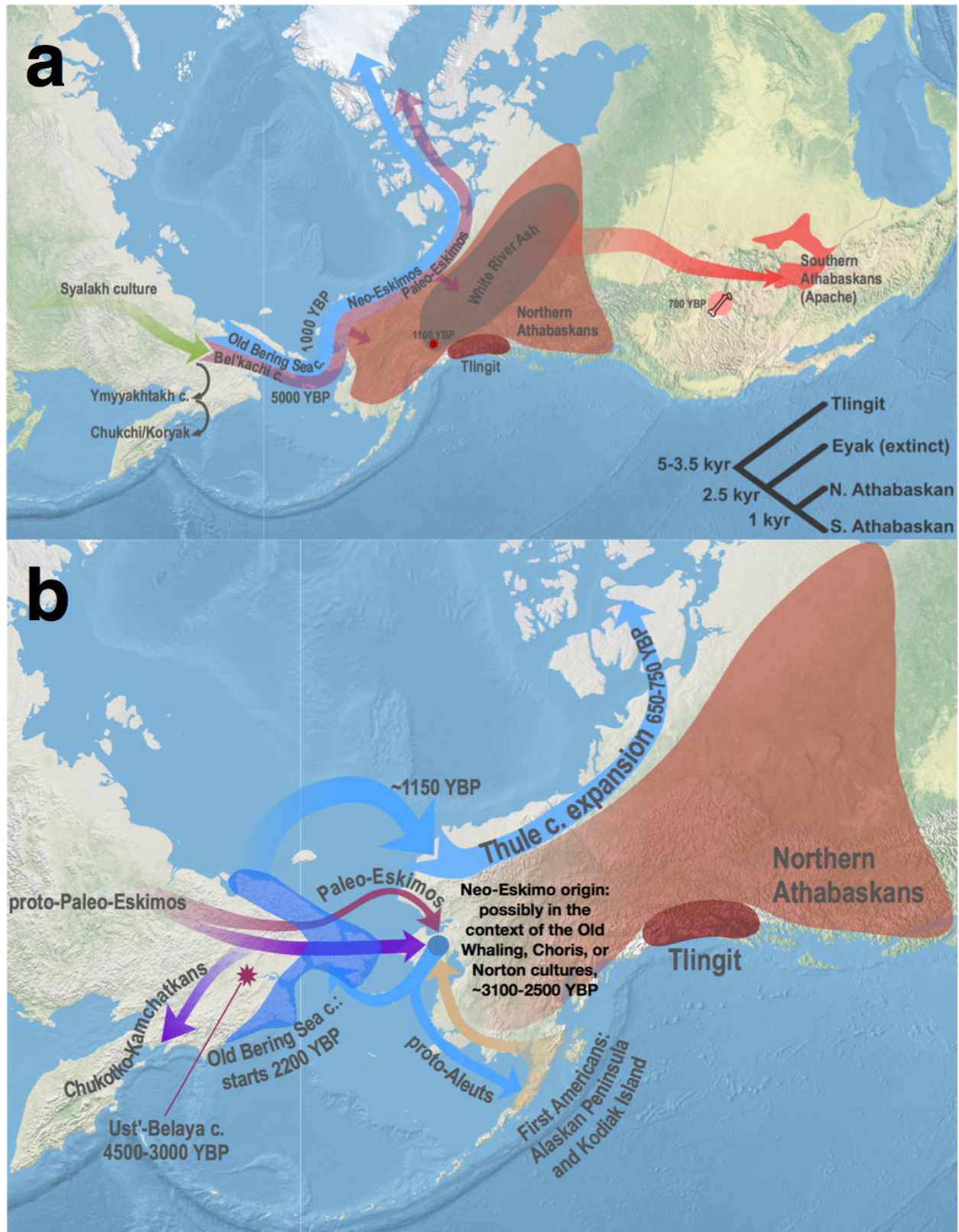
1094
1095
1096

1097 **Figure 5. a**, A demographic model based on 106 individuals from 7 meta-populations,
 1098 estimated using *rarecoal*, which is a fit to the data. Dashed arrows indicate salient
 1099 admixture events. For a figure showing all admixture events and for a complete list of
 1100 parameter estimates see Supplementary Information section 9. Admixture graphs with the
 1101 same topology are presented in Extended Data Fig. 10 and in Supplementary Information
 1102 section 9. In the case of European admixture in the Siberian/American and American
 1103 clades, the arrows indicate a ghost population that split off the European branch around
 1104 23,000 calBP and most likely corresponds to Ancient North Eurasians³⁹. A similar ghost
 1105 population is modelled splitting from the ancestors of Athabaskans and admixing into the
 1106 branch representing Eskimo-Aleut speakers. We also added admixture edges at 200 calBP
 1107 from Europeans into some extant groups (Eskimo-Aleut and Chukotko-Kamchatkan
 1108 speakers), modelling Post-Columbian admixture. These are not shown for clarity. A more
 1109 ancient European admixture event in the Siberian clade dated at ~4,000 calBP
 1110 (Supplementary Information section 9) is not shown either. **b** and **c**, Most likely branching
 1111 points for the Saqqaq and ancient Aleut sample for which we generated complete genome
 1112 data. Colored points indicate relative log likelihood with respect to the best fitting model.
 1113 Only branching prior to the radiocarbon dates of the samples was allowed.
 1114



1115
 1116
 1117

1118 **Figure 6.** An overview of North American and Chukotkan population history illustrating
1119 the history of major Na-Dene groups (**a**) and our model for the emergence and spread of
1120 Eskimo-Aleut speakers (**b**). Approximate earliest dates in calBP are indicated for
1121 archaeological or ethnic areas and for migrations. Due to space constraints, some migration
1122 paths are drawn to indicate just general directions, but not actual routes of population
1123 spread. **a**, The Paleo-Eskimo/Na-Dene gene flow we provisionally mapped across the
1124 coastal-interior boundary separating the ASTt and Northern Archaic cultures in Alaska,
1125 where the highest diversity of Na-Dene languages is found (for that reason Alaska was
1126 proposed as a homeland of the Na-Dene language family⁶⁷). The gene flow might take place
1127 further east along the same boundary. In addition, this panel shows the succession of
1128 archaeological cultures on the Siberian side of the Bering Strait, following the split with
1129 Paleo-Eskimos and culminating with present-day Chukchi, Itelmens, and Koryaks. A
1130 cladogram of the Na-Dene language family in the bottom right-hand corner shows the
1131 current consensus view of language relationships and summarizes published linguistic
1132 dating results (see further details in Supplementary Information section 10) The Mount
1133 Churchill volcanic eruption that deposited the precisely dated White River Ash⁴⁸ and
1134 possibly triggered the departure of Apachean ancestors^{45-47,68} is also shown. While early
1135 stages of the Apachean southward migration remain undated, their appearance at the
1136 Promontory Caves (Utah) has been dated at 700 – 660 calBP⁶⁹. **b**, A model of population
1137 history for Eskimo-Aleut speakers combining genetic and archaeological evidence; see
1138 Discussion for details. The Ust'-Belaya site in Chukotka is shown with an asterisk.
1139



1142

Table 1. Summary of genome-wide data from 17 newly reported samples

ID1	ID2	Skeletal element	Date, calBP (95% CI)	Label	Location	Country	Latitude	Longitude	Sex	mtDNA	Y haplo-group	Data type	Coverage	SNPs
I1526	barrow 8, skeleton 4	molar	4410 – 4100	Ust'-Belaya	Ust'-Belaya II, Chukotka	Russia	65.48	173.29	M	C4a1a3	Q1a2a	Capture	4.523	832,452
I1525	burial 13, inventory no. 172	molar	1970 – 1590	Old Bering Sea	Uelen, Chukotka	Russia	66.19	-169.9	F	A2a		Capture	1.031	608,585
I1524	burial 22, inventory no. 163	molar	1180 – 830				66.17	-170.75	M	A2a	Q1a2a1a1	Capture	2.618	797,816
I5319	MT_1		790 – 640						M	A2a1	Q1a2a1a1	Capture	5.731	865,897
I5320	MT_2	pars petrosa	790 – 640 (from MT_1)	Athabaskan	Tochak McGrath, Upper Kuskokwim River, Alaska	USA	62.95	-155.59	M	A2+(64)	Q1a2a1a1	Capture	5.038	839,833
I5321	MT_3		790 – 640 (from MT_1)						F	A2+(64)		Capture	4.851	827,885
I0721	378628		2320 – 1900						M	D2a1a	CT	Capture	0.025	28,805
I0712	378623		1270 – 930	Paleo-Aleut	Chaluka Midden, Umnak Island, Aleutian Islands	USA	52.99	-168.82	F	D2a1a		Capture	0.431	395,958
I1126	378622		1250 – 780						M	D2a1a	Q	Capture	0.092	103,481
I0719	378620		770 – 390						M	D2a1a	Q1a2a	Capture Shotgun	3.432 2.700	927,083 290,049
I1125	378544		760 – 490		Ship Rock Island, Aleutian Islands	USA	53.37	-167.83	M	D2a1a	Q1a2a1a1	Capture	1.335	640,629
I1127	377814	bone (rib)	630 – 310						F	D2a1a		Capture	1.354	662,276
I1128	377915		610 – 290						F	D2a1a		Capture	0.433	347,752
I1129	377917		600 – 270	Neo-Aleut	Kagamil Island Warm Cave, Aleutian Islands	USA	52.99	-169.71	M	D2a1a	Q1a2	Capture	0.687	495,889
I1118	377811		530 – 230						F	D2a1a		Capture	3.551	759,975
I1123	377918		520 – 140						M	A2a	Q1a	Capture	0.128	136,906
I1124	377919		500 – 140						M	D2a1a	Q1a2a	Capture	0.156	164,481

1143

1144

Notes: Genetic analysis indicates that I5319 and I5320 are a father-son pair. We produced both 1.24 million SNP capture and shotgun sequencing data for I0719

1145