

## Supplemental information

### **Recent Siberian ancestry in Na-Dene populations: tracing the second migration into America**

Pavel Flegontov, N. Ezgi Altınışık, Piya Changmai, Edward J. Vajda, Johannes Krause, Stephan Schiffels

#### **CONTENTS**

<b>Supplemental tables</b>	<b>2</b>
<b>Supplemental figures</b>	<b>3</b>
<b>Supplemental text 1. Rarecoal analysis</b>	<b>18</b>
<b>Supplemental text 2. Conflicting interpretations of Na-Dene population history</b>	<b>31</b>
<b>Supplemental text 3. Estimating the time depth of the Dene-Yeniseian linguistic connection</b>	<b>33</b>

**Suppl. Table 1.** List of populations and meta-populations used in this study: African (AFR), American Arctic (AARC), Siberian Arctic (SARC), European (EUR), Na-Dene (mostly Athabaskan, ATH), northern North American (NAM), South American (SAM), Southeast Asian (SEA), core Siberian (cSIB), Siberian with ancient North Eurasian ancestry (SIB+ANE). Sequencing data were taken from three sources: the Simons Genome Diversity Project (Mallick et al. 2016), Raghavan et al. (2015), and 1000 Genomes project (1000 Genomes Project Consortium 2015). The latter source was included into the African, European, and Southeast Asian meta-populations only. Two SNP array datasets were used: based on the HumanOrigins array and on Illumina arrays. HumanOrigins data were taken from Mathieson et al. (2015) or generated in this study for four Siberian populations (Enets, Kets, Nganasans, Selkups). Illumina data were taken from the following sources: Li et al. 2008, Behar et al. 2010, Rasmussen et al. 2010, Fedorova et al. 2013, Raghavan et al. 2014a, 2014b, 2015, Verdu et al. 2014, Kushniarevich et al. 2015. Notes:

\* Aleut\_Tlingit is a sample of Aleuts having Tlingit (Na-Dene) maternal ancestors (David Reich, personal communication). According to ADMIXTURE (Suppl. Fig. 1A) and fineSTRUCTURE clustering tree (Suppl. Fig. 3A) and PCA analyses (data not shown), this group is indistinguishable from other Aleuts.

\*\* Both populations belong to Yupik (Chukotkan Eskimos), but one of them is referred to Eskimo in the figures, preserving the population name used by Mathieson et al. 2015 and Mallick et al. 2016.

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

^The Middle Eastern (ME) and Oceanian (OCE) meta-populations were included into the HumanOrigins dataset, but were not used in most analyses.

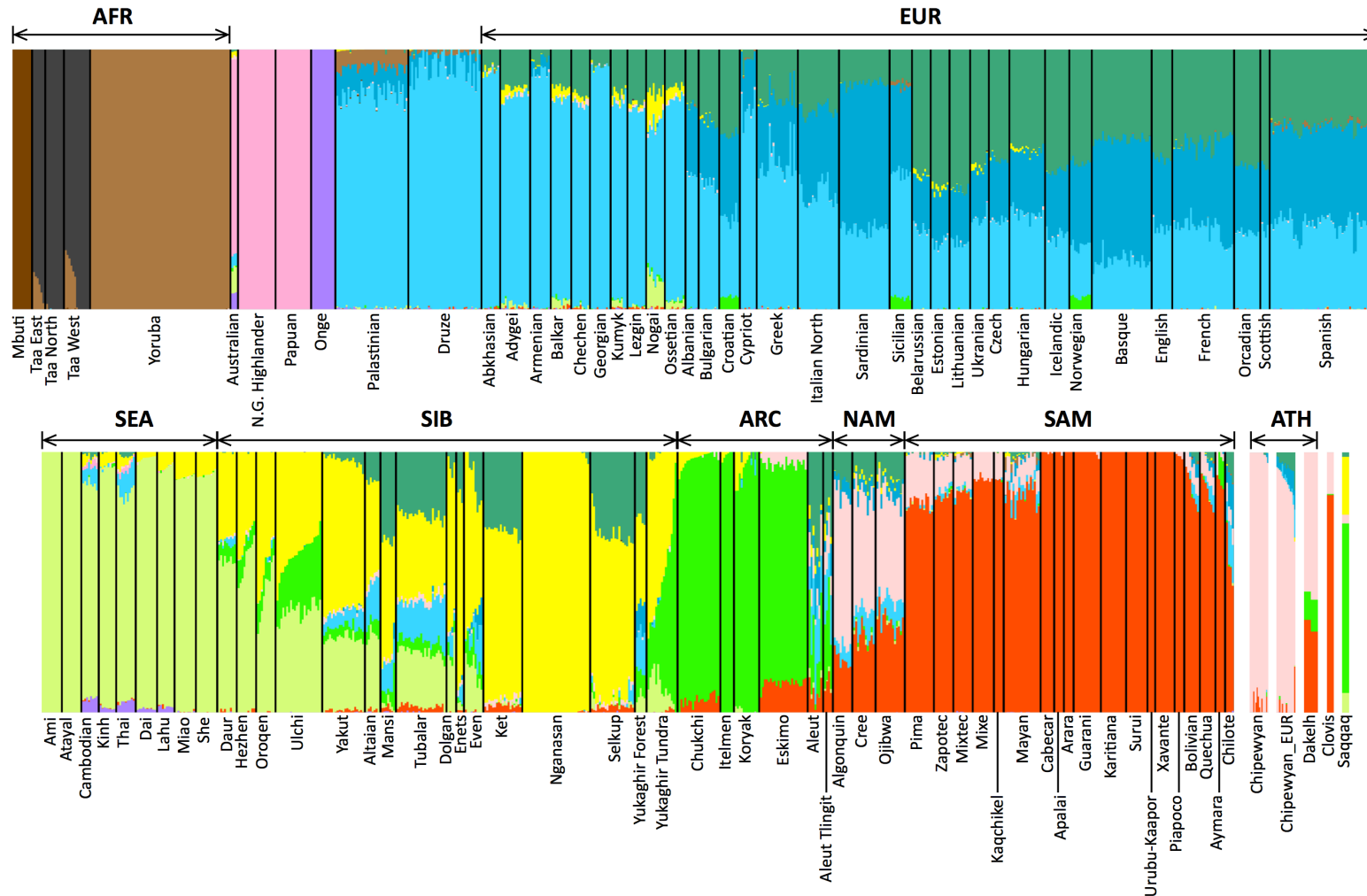
**Suppl. Table 2.** Details of datasets used in this study.

**Suppl. Table 3.** Information on newly genotyped Siberian individuals.

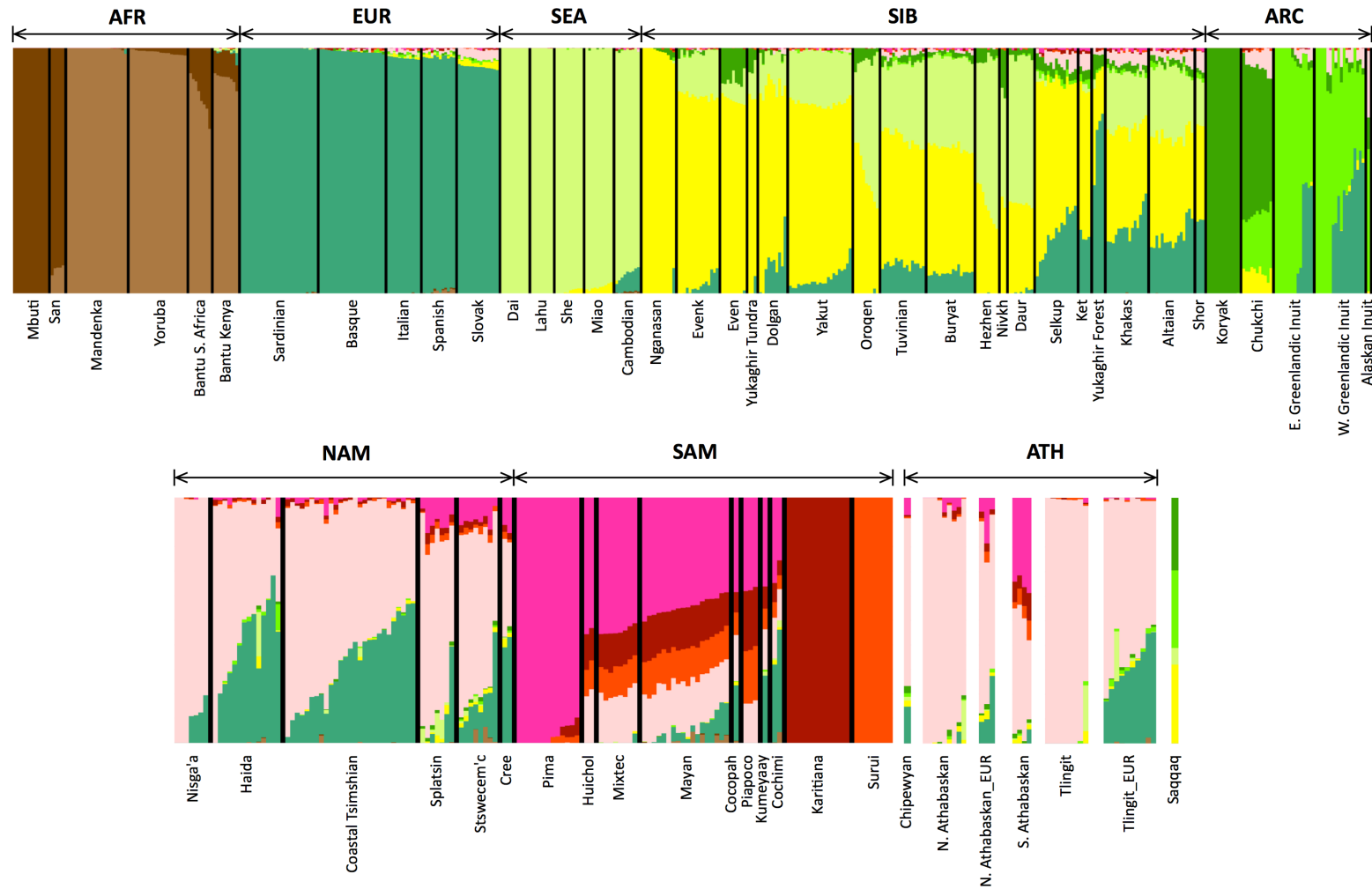
### *References (for suppl. tables)*

- 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
- Behar, D. M. *et al.* The genome-wide structure of the Jewish people. *Nature* **466**, 238–242 (2010).
- Fedorova, S. A. *et al.* Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol. Biol.* **13**, 127 (2013).
- Kushniarevich, A. *et al.* Genetic heritage of the Balto-Slavic speaking populations: A synthesis of autosomal, mitochondrial and Y-chromosomal data. *PLoS ONE* **10**, e0135820 (2015).
- Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
- Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* (2016), in press.
- Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
- Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345**, 1255832 (2014a).
- Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014b).
- Raghavan, M. *et al.* Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, 1–20 (2015).
- Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
- Verdu, P. *et al.* Patterns of admixture and population structure in native populations of northwest North America. *PLoS Genet.* **10**, e1004530 (2014).

**Suppl. Fig. 1. (A)** ADMIXTURE plot for the HumanOrigins SNP array dataset that models contributions from 13 hypothetical ancestral populations. One hundred iterations were calculated for each value of  $K$  from 5 to 20 (where  $K$  is the number of ancestral populations), and  $K=13$  was selected based on ten-fold cross-validation. Contributions from hypothetical ancestral populations are color-coded, and meta-populations used in this study are indicated above the plot: AFR, Africans; ARC, Arctic groups; ATH, Athabaskans; EUR, populations of Europe and the Caucasus; NAM, northern North Americans, excluding Na-Dene, Yupik and Inuit; SEA, Southeast Asians; SIB, Siberians, excluding populations of Chukotka and Kamchatka; SAM, native populations of South, Central America, Mexico and southern USA. Chipewyan individuals with European admixture are plotted in a separate bar, as well as the Dakelh individuals, the Saqqaq and Clovis ancient genomes.

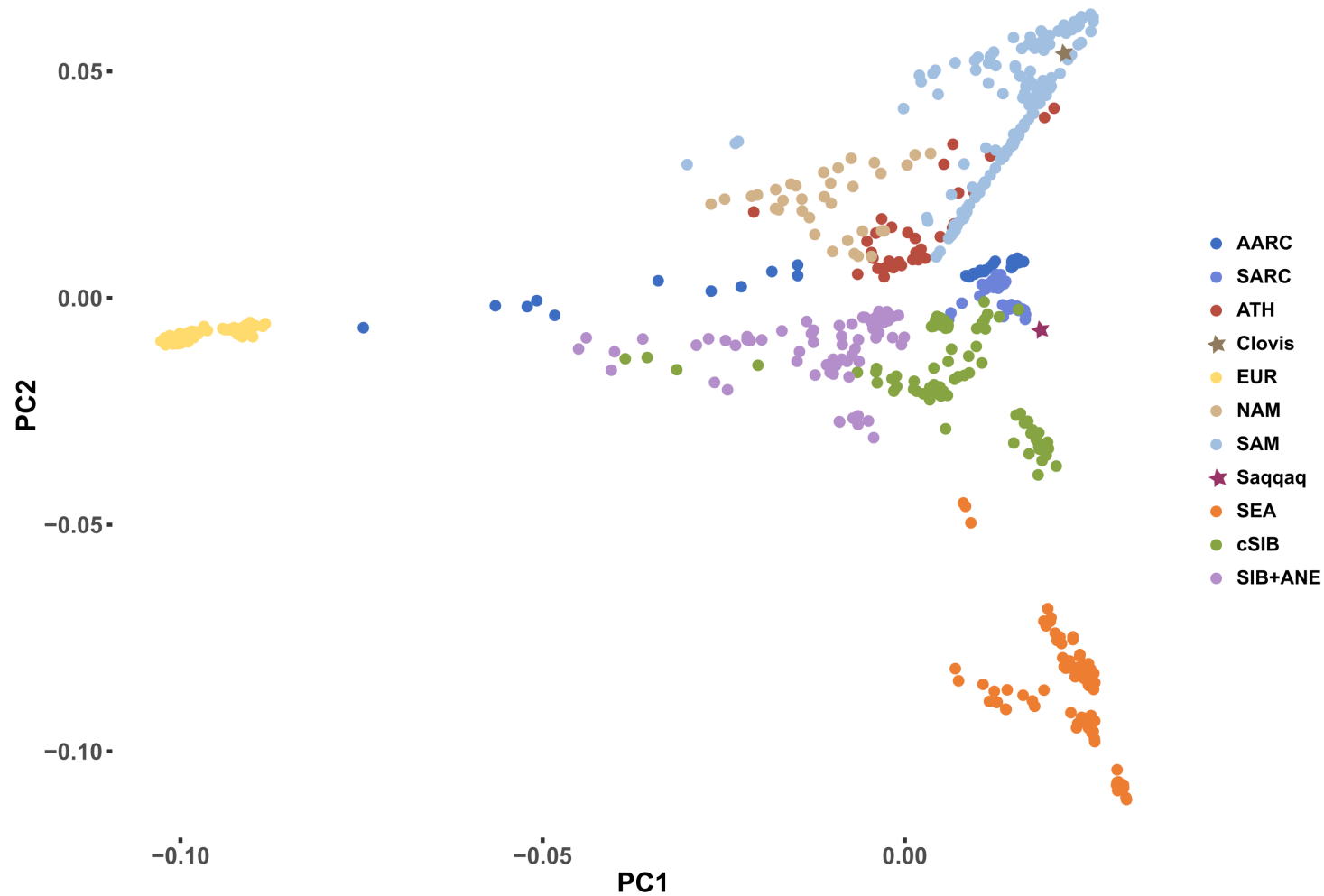


**(B)** ADMIXTURE plot for the Illumina SNP array dataset that models contributions from 11 hypothetical ancestral populations. One hundred iterations were calculated for each value of  $K$  from 5 to 20 (where  $K$  is the number of ancestral populations), and  $K=11$  was selected based on ten-fold cross-validation. Contributions from hypothetical ancestral populations are color-coded, and meta-populations used in this study are indicated above the plot: AFR, Africans; EUR, Europeans; SEA, Southeast Asians; SIB, Siberians, excluding populations of Chukotka and Kamchatka; SAM, native populations of South, Central America, Mexico and southern USA. Northern Athabaskan and Tlingit individuals with European admixture are plotted in separate bars, as well as the Saqqaq ancient genome.

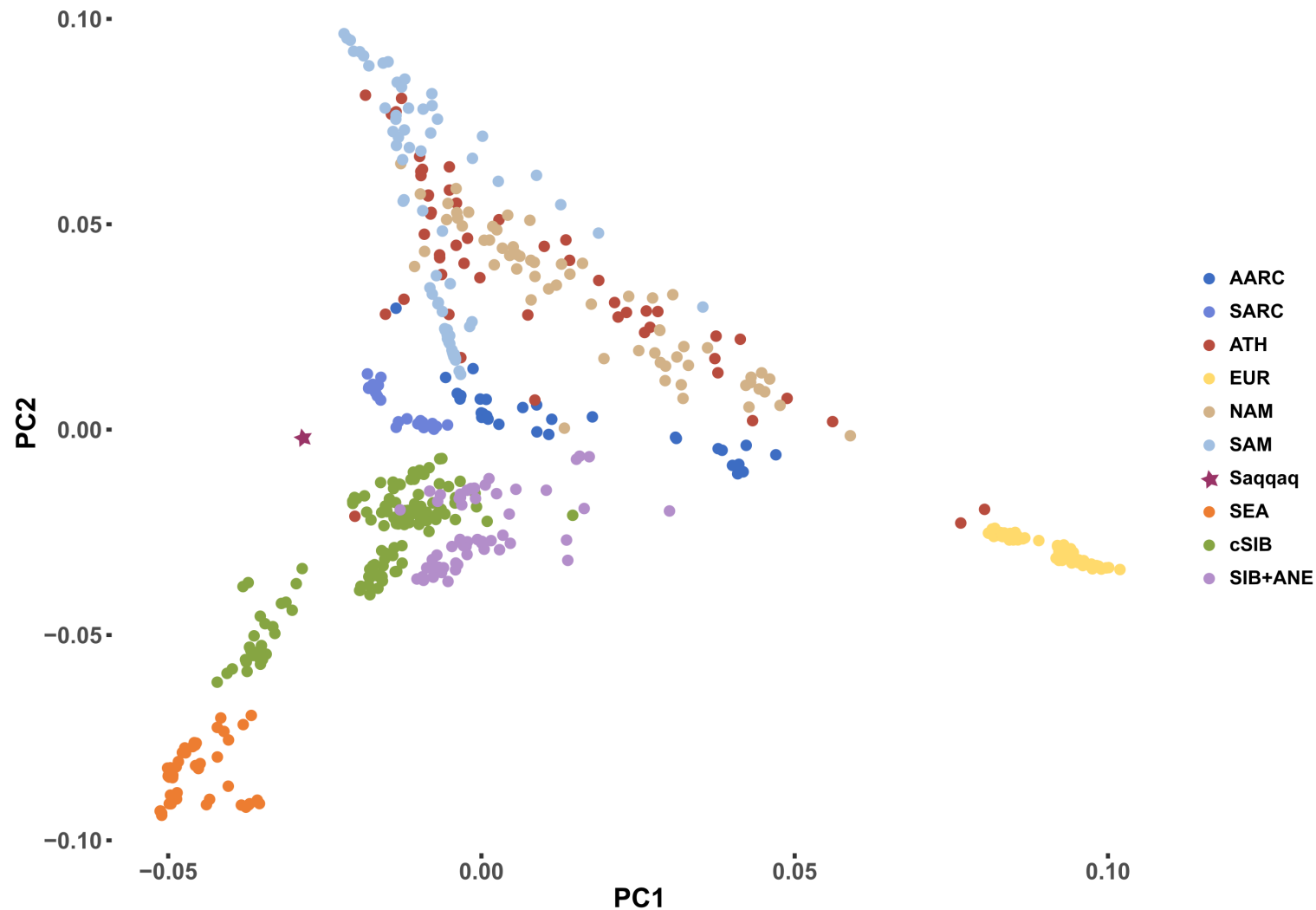




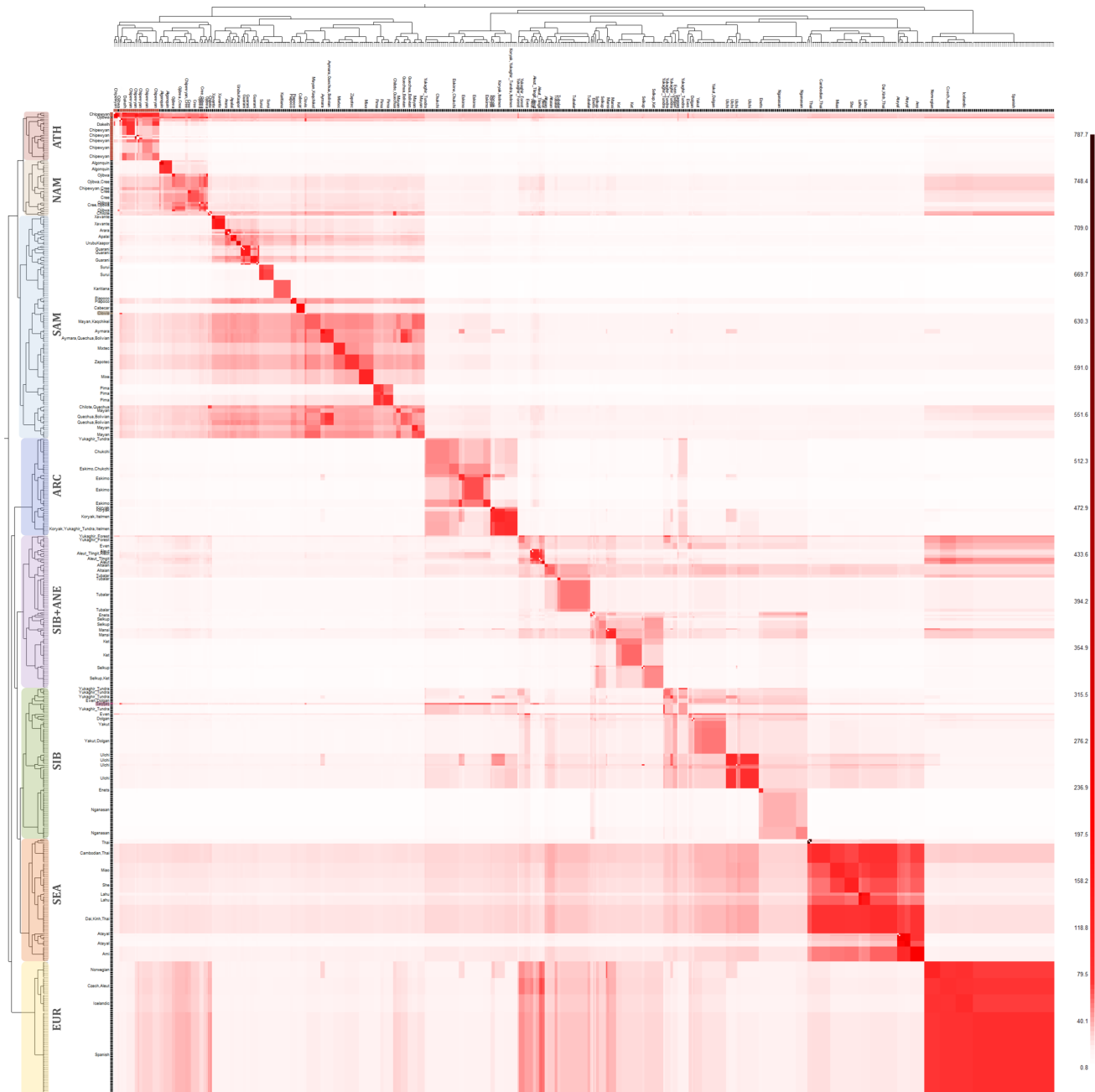
**Suppl. Fig. 2. (A)** A plot of two principal components (PC1 vs. PC2) generated by fineSTRUCTURE based on a coancestry matrix of shared haplotype counts. A reduced version of the HumanOrigins SNP array dataset was used (655 individuals and 58 populations, see Table 2), including only the following meta-populations most relevant for our study: American and Siberian Arctic groups (AARC, SARC), Athabaskans (ATH), Europeans (EUR), northern North Americans (NAM), other First Americans (SAM), Southeast Asians (SEA), Siberians (SIB), and the Saqqaq and Clovis ancient genomes (marked with asterisks in the plot). Siberians with extensive ancient North Eurasian ancestry, i.e. Altaians, Enets, Kets, Mansi, Selkups, Tubalars, and Forest Yukaghirs (Flegontov et al. 2016, Raghavan et al. 2014b) are marked as SIB+ANE, and the other Siberians as cSIB.



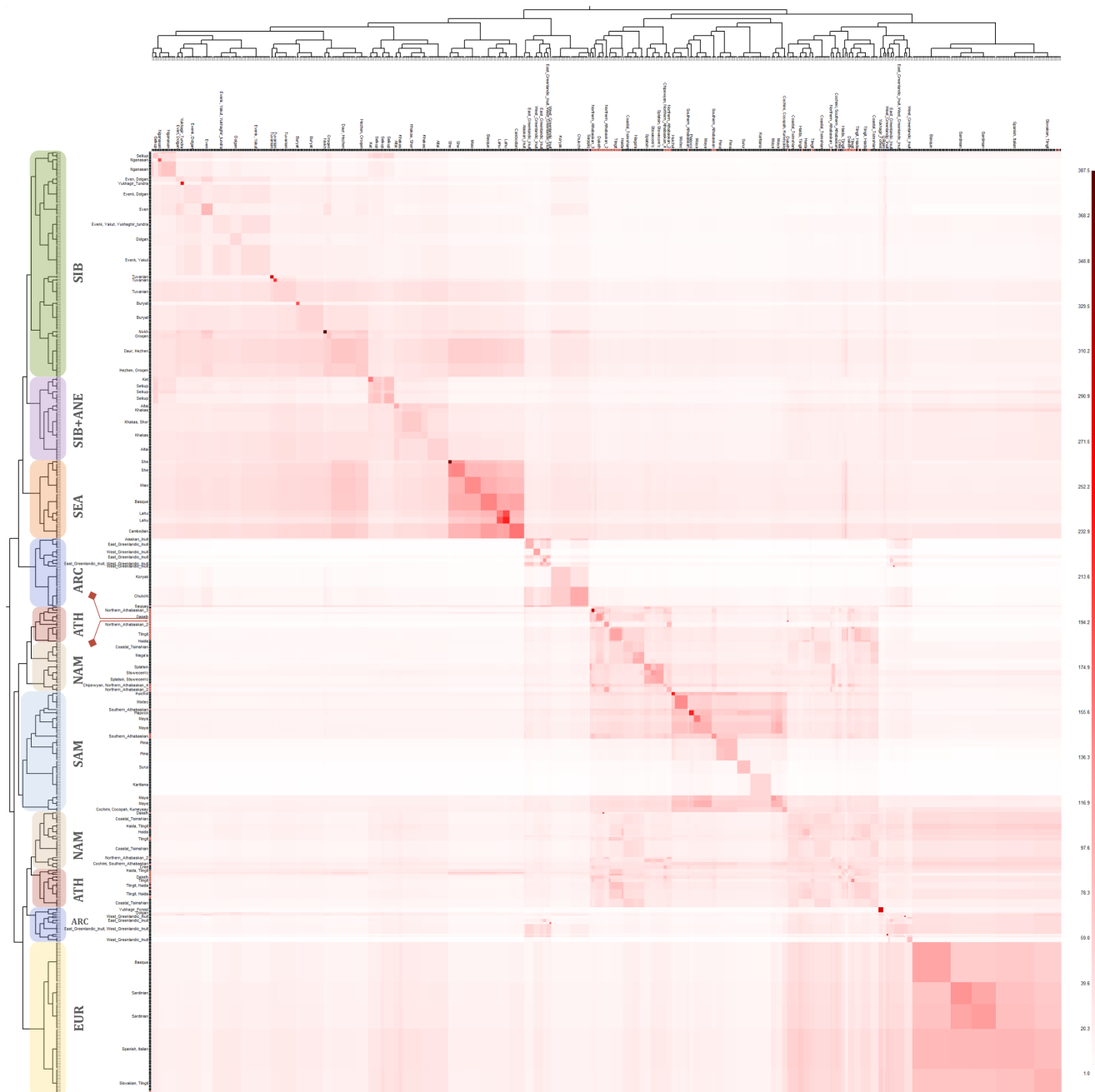
**(B)** A plot of two principal components (PC1 vs. PC2) generated by fineSTRUCTURE based on a coancestry matrix of shared haplotype counts. A reduced version of the Illumina SNP array dataset was used (567 individuals and 57 populations, see Table 2), including the following meta-populations most relevant for our study: American and Siberian Arctic groups (AARC, SARC), Na-Dene (ATH), Europeans (EUR), northern North Americans (NAM), other First Americans (SAM), Southeast Asians (SEA), Siberians (SIB), and the Saqqaq ancient genome (marked with an asterisk in the plot). Siberians with extensive ancient North Eurasian ancestry, i.e. Altaians, Kets, Khakases, Selkups, Shors, and Forest Yukaghirs (Flegontov et al. 2016, Raghavan et al. 2014b) are marked as SIB+ANE, and the other Siberians as cSIB.



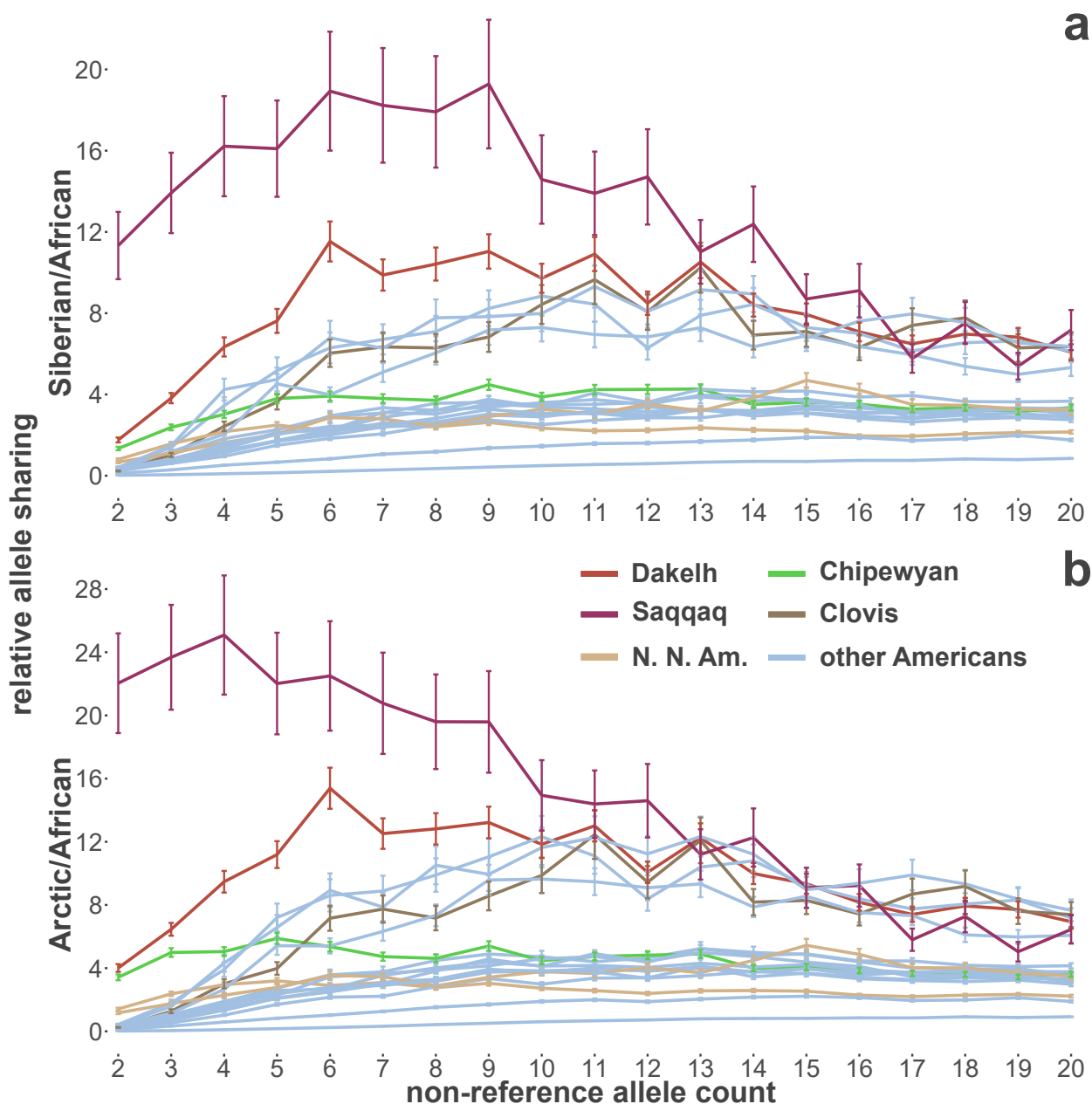
**Suppl. Fig. 3. (A)** A heat map of a coancestry matrix (based on counts of shared haplotypes), and a clustering tree of individuals computed by fineSTRUCTURE on this matrix. A reduced version of the HumanOrigins SNP array dataset was used (655 individuals and 58 populations, see Suppl. Table 2), including only the following meta-populations most relevant for our study: Arctic groups (ARC), Athabaskans (ATH), Europeans (EUR), northern North Americans (NAM), other First Americans (SAM), Southeast Asians (SEA), Siberians (SIB), and the Saqqaq and Clovis ancient genomes. Athabaskan individuals are marked in dark-red. Clades corresponding to meta-populations are highlighted in color on the right, as well as the names of the ancient genomes: Clovis within the South American clade, and Saqqaq within the Siberian clade. Siberians with extensive ancient North Eurasian ancestry, i.e. Altaians, Enets, Kets, Mansi, Selkups, Tubalars, and Forest Yukaghirs (Flegontov et al. 2016, Raghavan et al. 2014b) are marked as SIB+ANE. Most members of each clade belong to the meta-populations indicated, with few exceptions: some Aleuts fall into the Siberian+ANE clade; several Tundra Yukaghirs cluster in the Arctic clade; and one Ojibwa individual falls into the Athabaskan clade.

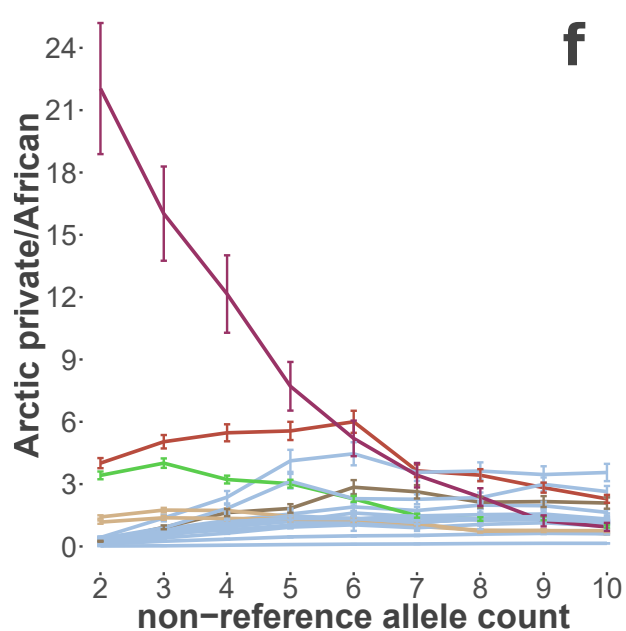
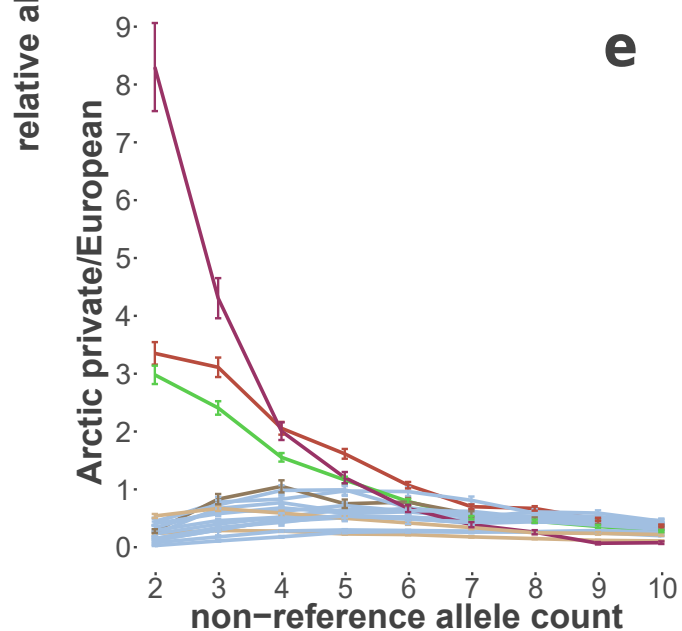
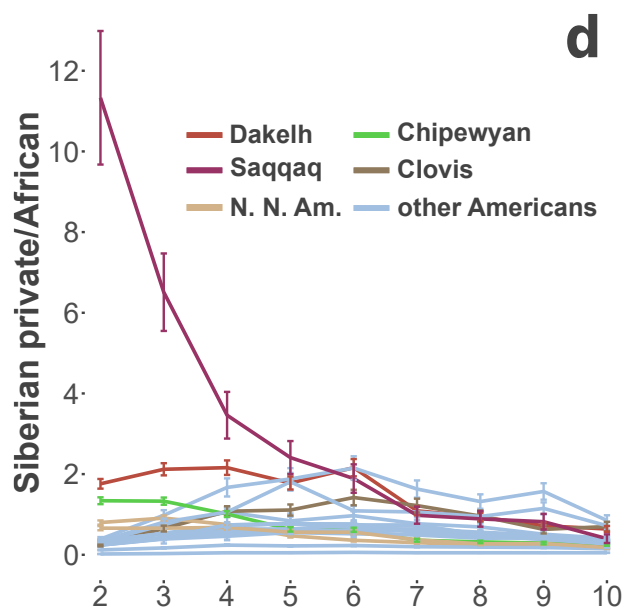
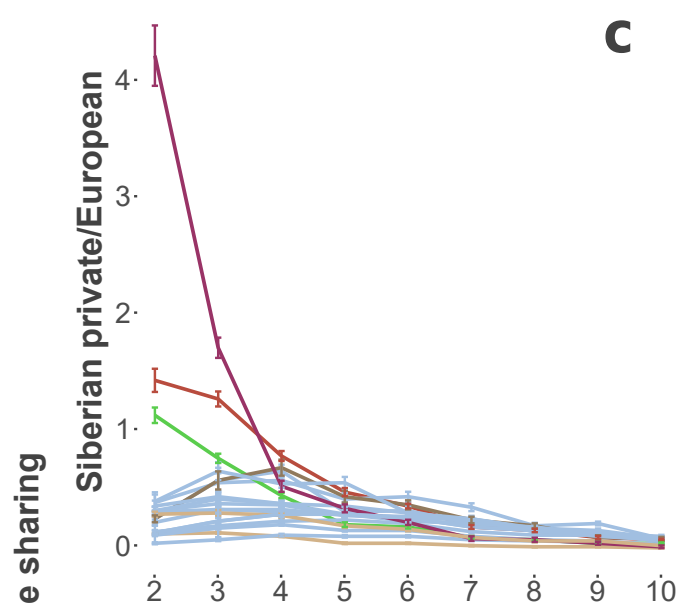


**(B)** A heat map of a coancestry matrix (based on counts of shared haplotypes), and a clustering tree of individuals computed by fineSTRUCTURE on this matrix. A reduced version of the Illumina SNP array dataset was used (567 individuals and 57 populations, see Suppl. Table 2), including only the following meta-populations most relevant for our study: Arctic groups (ARC), Na-Dene (ATH), Europeans (EUR), northern North Americans (NAM), other First Americans (SAM), Southeast Asians (SEA), Siberians (SIB), and the Saqqaq ancient genome. Na-Dene individuals are marked in dark-red, and two Dakelh with sequenced genomes are marked with callouts. Clades corresponding to meta-populations are highlighted in color on the right. Saqqaq falls within the Arctic clade. Siberians with extensive ancient North Eurasian ancestry, i.e. Altaians, Kets, Khakases, Selkups, Shors, and Forest Yukaghirs (Flegontov et al. 2016, Raghavan et al. 2014b) are marked as SIB+ANE. Notably, the Arctic, Na-Dene, and northern North American groups in this dataset are heterogeneous: a subset of each group clusters with Europeans, corresponding to individuals with a high proportion of European ADMIXTURE components (data not shown). Most members of each clade belong to the meta-populations indicated, with few exceptions: some Selkups in the core Siberian clade (SIB); all four Southern Athabaskans cluster with South American, reflecting their South American ADMIXTURE components (Suppl. Fig. 1B); one Haida individual clusters with Na-Dene; and five Northern Athabaskans cluster with northern North Americans.

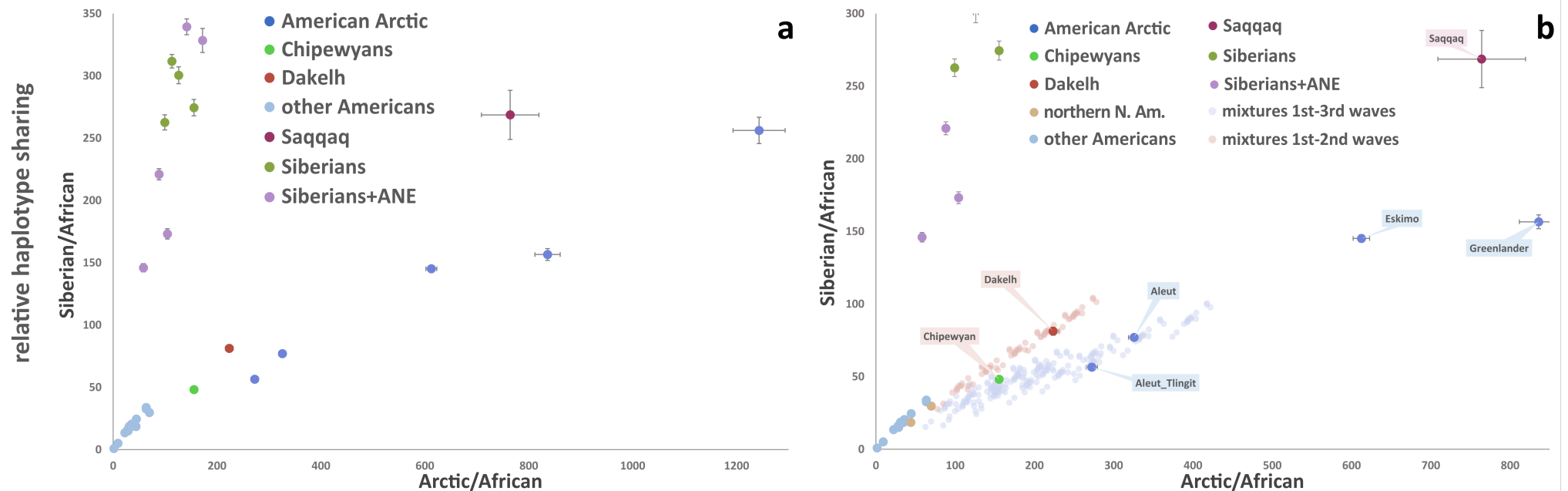


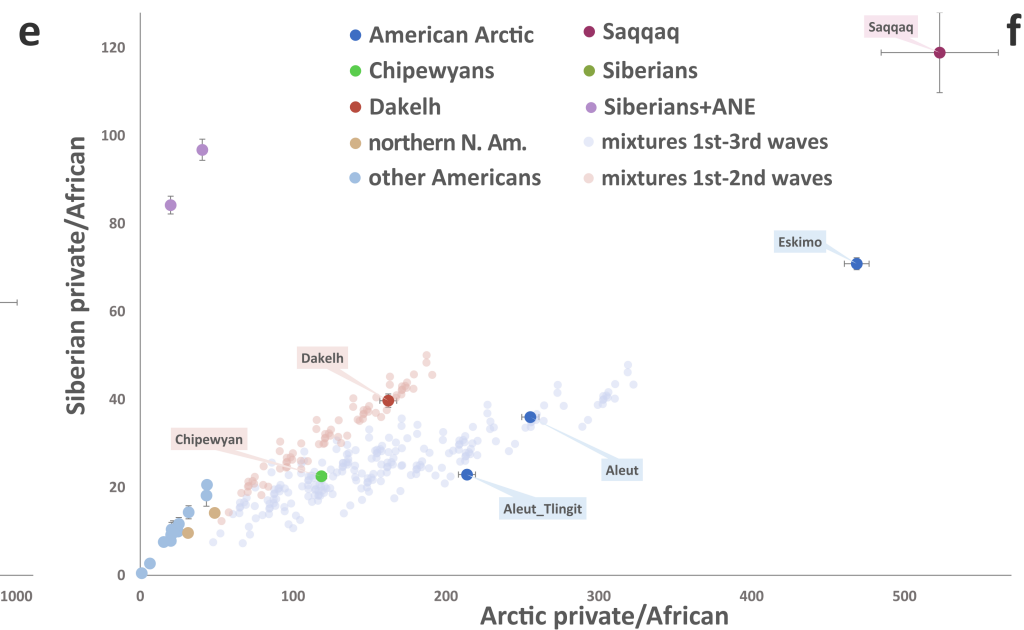
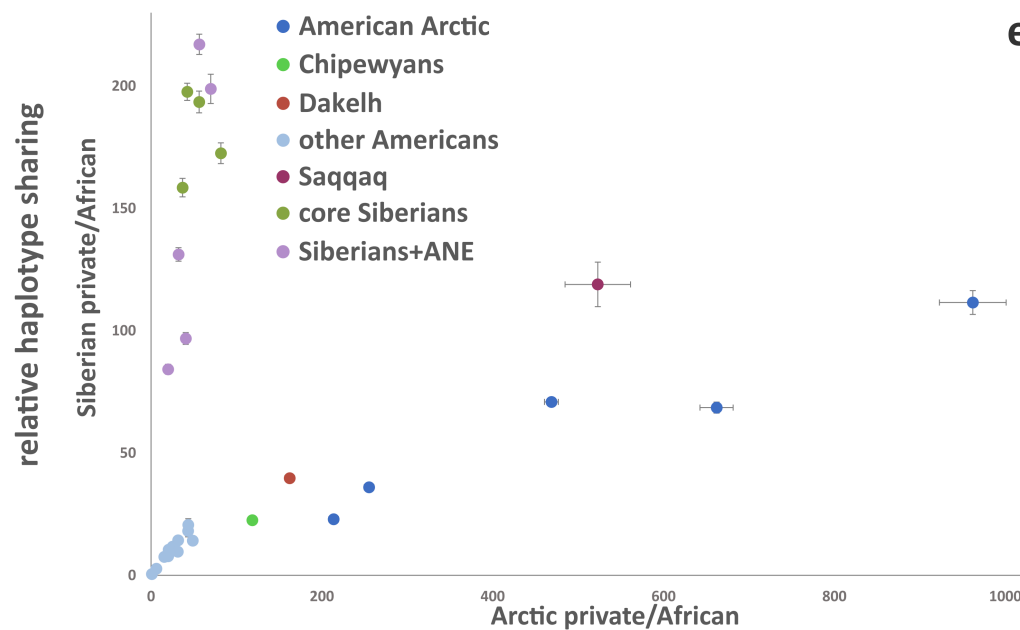
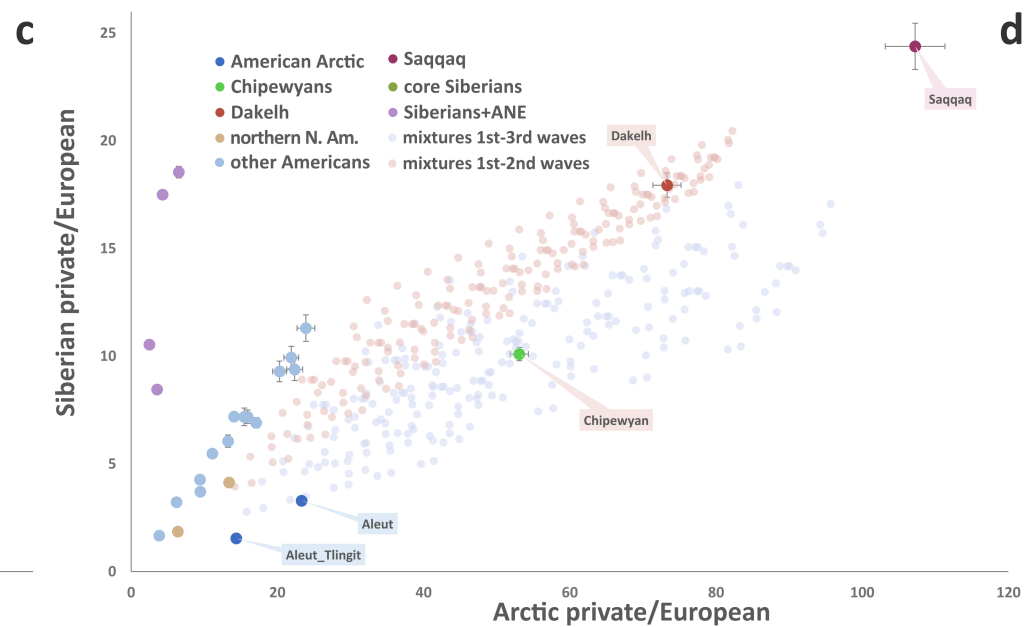
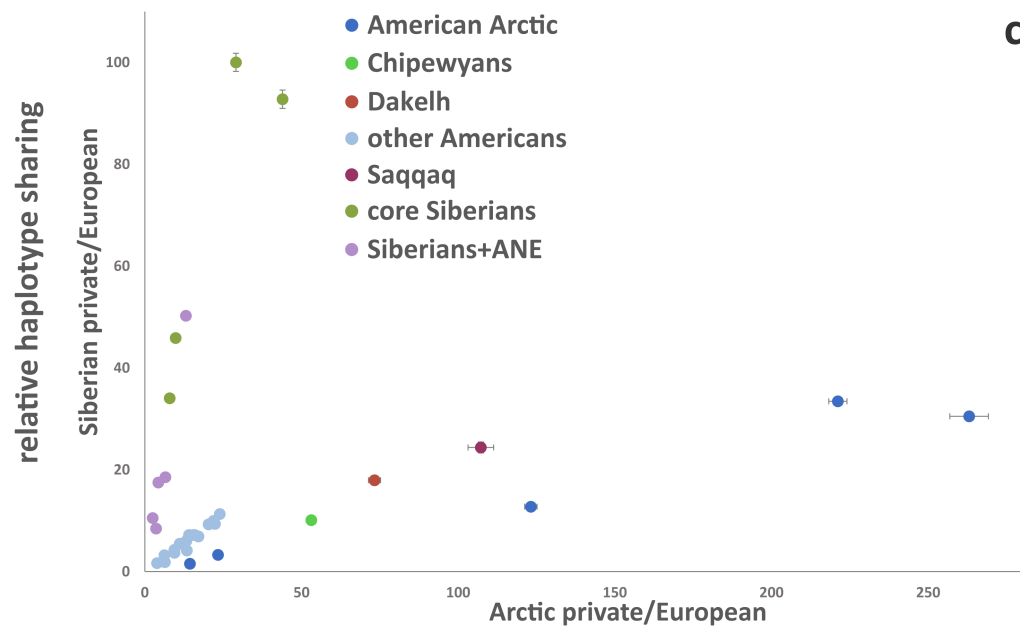
**Suppl. Fig. 4.** Normalized rare allele sharing counts calculated for each American population or ancient genome and the Siberian (**A, C, D**) or Arctic (panels **B, E, F**) meta-populations. All statistics were calculated separately for alleles of various frequency: occurring 2, 3, 4, ... and up to 20 times in the set of 2,412 chromosomes. To take care of variability in genome coverage across populations and of dataset-specific SNP calling biases, we normalized the counts of alleles shared by a given American population and the Arctic or Siberian meta-populations by similar counts of alleles shared with distant outgroups –Europeans (**C, E**) or Africans (**A, B, D, F**). We called a shared allele private, if it was present in an American population and Siberians or members of the Arctic group, but missing from all other meta-populations (we did not condition on the presence of this allele in other Americans). Plots for private alleles are shown in panels **C-F**. Saqqaq and Northern Athabaskans (Chipewyans and Dakelh) stand out from First American populations in all analyses shown here.





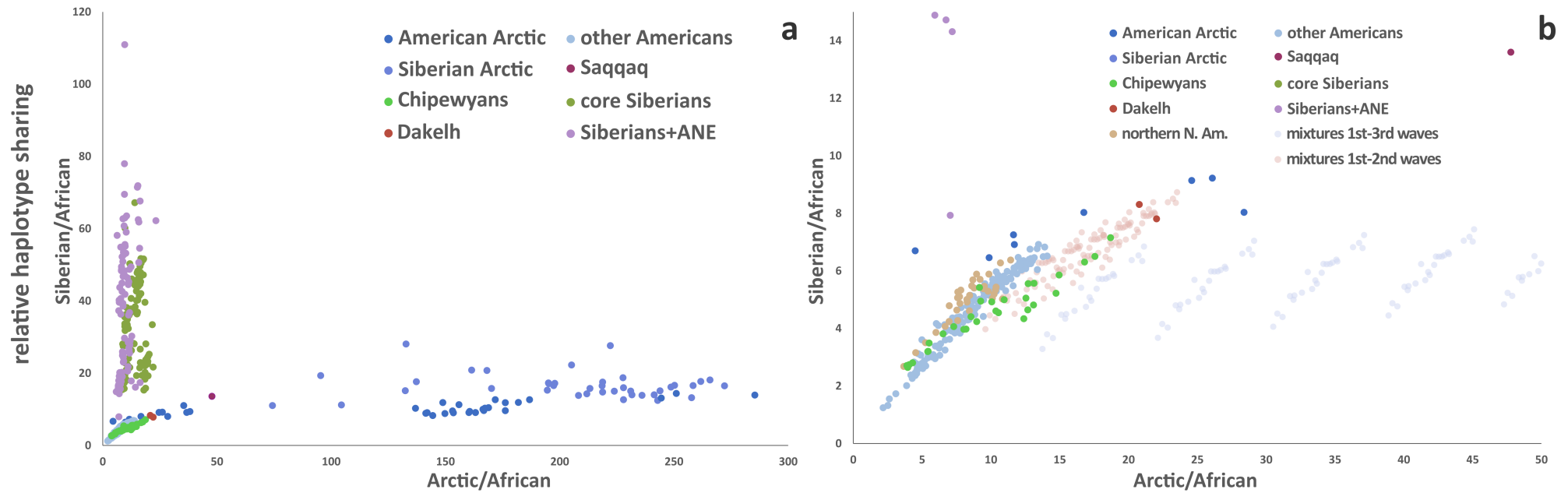
**Suppl. Fig. 5.** Two-dimensional plots of allele sharing counts normalized using the European (C, D) or African (A, B, E, F) meta-populations. Plots on the left (A, C, E) show statistics for all populations and standard deviations. Meta-populations are color-coded according to the legend. Among the Arctic group, only populations of the third wave are plotted (those residing in America or migrants having returned to Chukotka and Kamchatka). Plots on the right (B, D, F) are enlarged versions showing simulated mixtures of any modern First American population and the Saqqaq individual (from 10% to 30% in panels B and F, from 10% to 70% in panel D), and similar mixtures with any third-wave population, from 10% to 30% of Greenlander Inuit or Chukotkan Yupik (Eskimo) ancestry. We called a shared allele private, if it was present in an American population and Siberians or members of the Arctic group, but missing from all other meta-populations (we did not condition on the presence of this allele in other Americans). Plots for private alleles are shown in panels C-F.



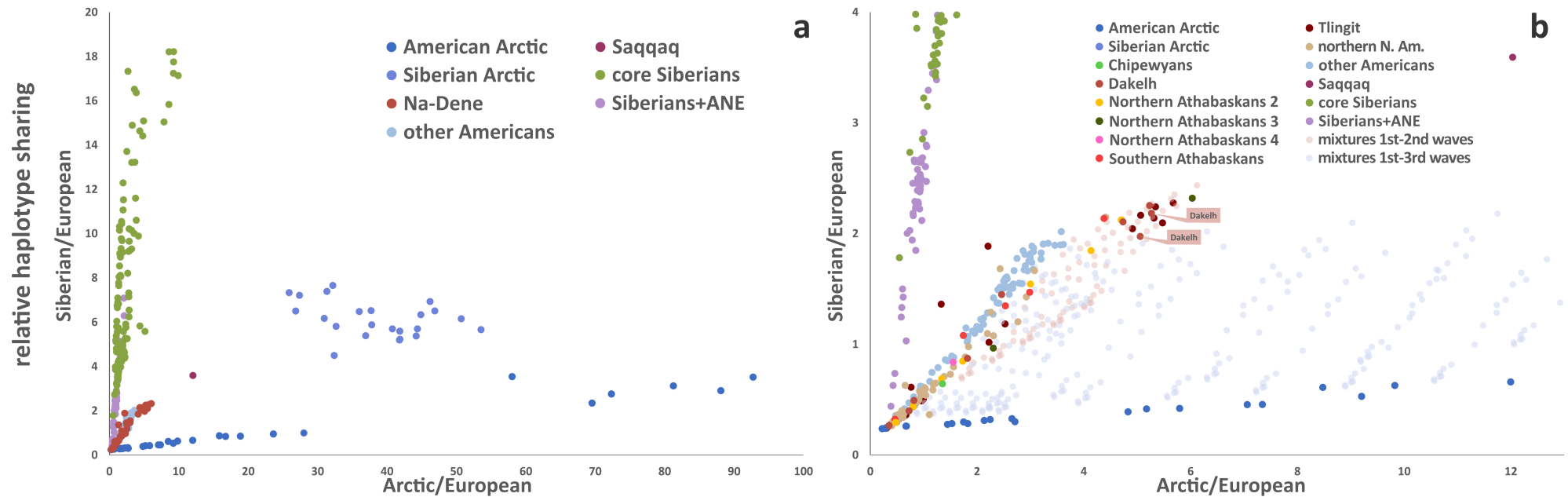




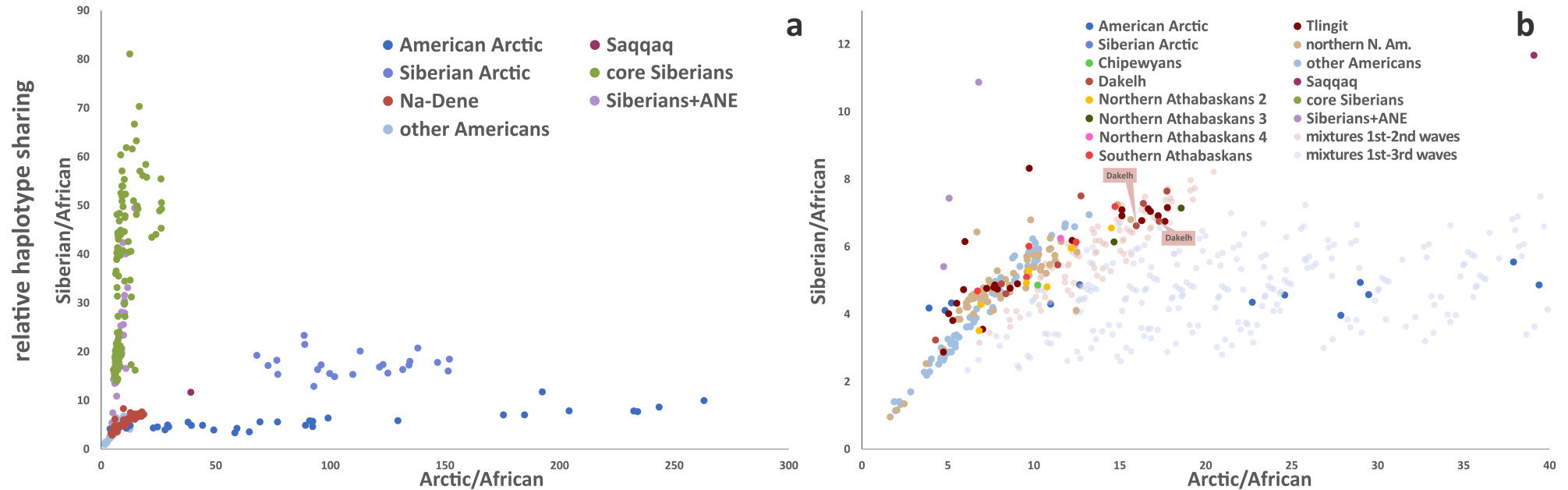
**Suppl. Fig. 6.** Two-dimensional plots of Siberian and Arctic haplotype sharing statistics normalized using the African meta-population and based on the HumanOrigins SNP array dataset. **(A)** A plot showing statistics for individuals of all relevant meta-populations (color-coded according to the legend). **(B)** An enlarged area of the plot showing statistics for American individuals and simulated mixtures of any modern First American population and the Saqqaq individual (from 10% to 30%), and similar mixtures with the Chukotkan Yupik population, referred to as Eskimo by Mathieson et al. (2015) (from 5% to 25% of Yupik ancestry). Average values of the statistics in populations were used to calculate the simulated statistics.



**Suppl. Fig. 7.** A two-dimensional plot of shared haplotype lengths normalized using the European meta-population and based on the Illumina SNP array dataset. The plot on the left (**A**) shows statistics for individuals of all relevant meta-populations (color-coded according to the legend). The plot on the right (**B**) is an enlarged version showing statistics for American individuals and simulated mixtures of any modern First American population and the Saqqaq individual (from 10% to 30%), and similar mixtures with any third-wave population. In the latter case, the following points fit the plot area: those with 5% to 30% of East or West Greenland Inuit ancestry, and with 5% to 20% of Alaskan Inuit ancestry. Average values of the statistics in populations were used to calculate the simulated statistics. Various Na-Dene populations are color-coded, and two Dakelh individuals with sequencing data, included into the HumanOrigins (Fig. 4, Suppl. Fig. 6) and rare allele datasets (Fig. 3, Suppl. Fig. 5), are marked with callouts.

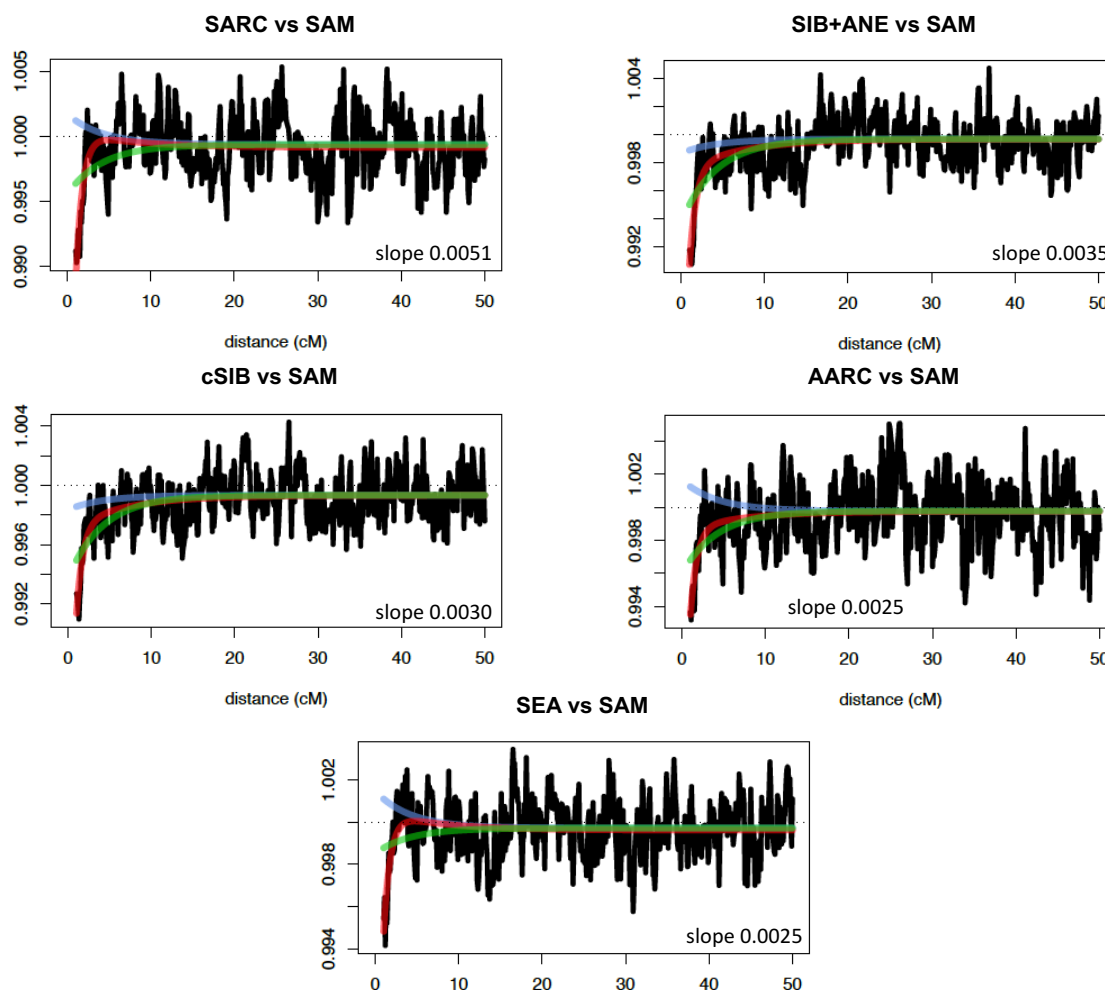


**Suppl. Fig. 8.** A two-dimensional plot of shared haplotype lengths normalized using the African meta-population and based on the Illumina SNP array dataset. The plot on the left (**A**) shows statistics for individuals of all relevant meta-populations (color-coded according to the legend). The plot on the right (**B**) is an enlarged version showing statistics for American individuals and simulated mixtures of any modern First American population and the Saqqaq individual (from 10% to 30%), and similar mixtures with any third-wave population. In the latter case, the following points fit the plot area: those with 5% to 30% of East or West Greenland Inuit ancestry, and with 5% to 15% of Alaskan Inuit ancestry. Average values of the statistics in populations were used to calculate the simulated statistics. Various Na-Dene populations are color-coded, and two Dakelh individuals with sequencing data, included into the HumanOrigins (Fig. 4, Suppl. Fig. 6) and rare allele datasets (Fig. 3, Suppl. Fig. 5), are marked with callouts.

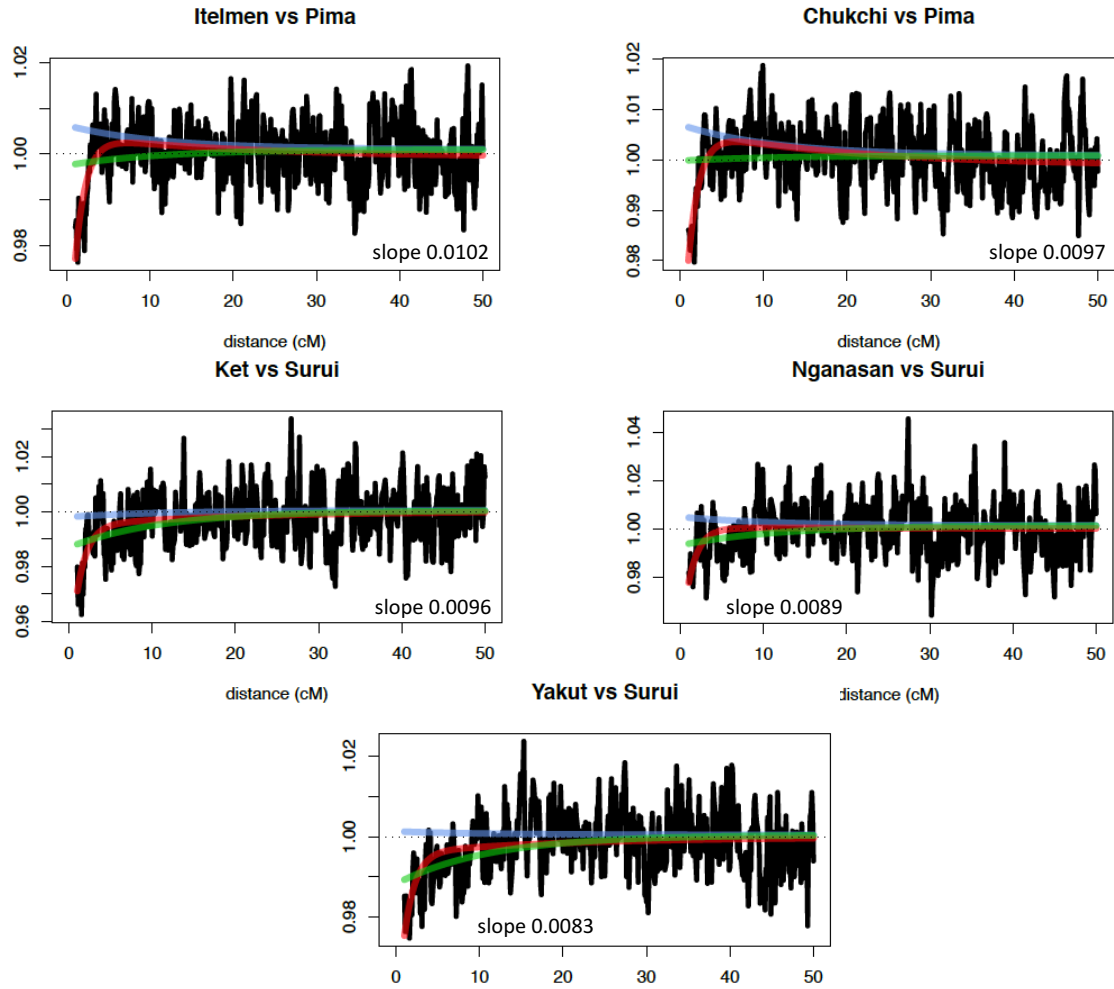


**Suppl. Fig. 9.** Coancestry curves: relative probability of jointly copying two genomic chunks from a pair of donors (y-axis) vs genetic distance in cM (x-axis). Original data are shown in black, and curves approximating two admixture events with different dates – in red, two events with a single date – in green, and one event – in blue. Slopes of the two-date curves were calculated in the range from 1 cM to 3 cM, and five curves with the highest positive slopes are shown for each set of haplotype donors analyzed with GLOBETROTTER: the HumanOrigins dataset with meta-populations (A) or populations (B) as haplotype donors; the Illumina dataset with meta-populations (C) or populations (D). Composition of target Na-Dene populations is shown in Table 2. The following meta-populations were used as donors: 1/ Siberian Arctic (abbreviated as SARC); 2/ American Arctic (AARC); 3/ Europe and the Caucasus (EUR); 4/ northern North Americans, excluding Na-Dene, Yupik and Inuit (NAM); 5/ Southeast Asians (SEA); 6/ native populations of South, Central America, Mexico and southern USA (SAM); 7/ Saqqaq; 8/ Siberians with extensive ancient North Eurasian ancestry (SIB+ANE); 9/ core Siberians (cSIB).

**A**



**B**



*References (for suppl. figures)*

- Flegontov, P. *et al.* Genomic study of the Ket: A Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Sci. Rep.* **6**, 20768 (2016).
- Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014b).

## Suppl. text 1. Rarecoal analysis

### *Stephan Schiffels*

#### Rarecoal

Rarecoal is a software that implements a fast algorithm to estimate the joint site frequency spectrum for rare alleles (Schiffels et al. 2016). Since the initial report in Schiffels et al. 2016, we have improved the software and added pulse-like admixture events as a new feature. The updated mathematical derivations of the model are included as a PDF document in the repository: <https://github.com/stschiff/rarecoal>.

#### Data

In the following analysis, we will use the abbreviations for meta-populations and samples shown in Suppl. Table 1.1.

Group	Populations <sup>1)</sup>	Number of samples
<b>SAM</b>	Aymara, Karitiana, Piapoco, Quechua, Surui, Yukpa	14
<b>NAM</b>	Cree, Tsimshian	3
<b>ATH</b>	Northern Athabaskan, Chipewyan	4
<b>AARC</b>	Aleut, Aleut_Tlingit, Chukotkan Yupik (Eskimo), Greenlander Inuit	14
<b>SIB</b>	Nivkh, Altaian, Buryat, Even, Ket, Mansi, Tubalar, Ulchi, Yakut	22
<b>SEA</b>	Ami, Atayal, Burmese, Cambodian, Dai, Kinh, Lahu, Miao, She, Thai	22
<b>EUR</b>	Basque, Bergamo, Bulgarian, Crete, Czech, English, Estonian, French, Greek, Hungarian, Norwegian, Orcadian, Polish, Sardinian, Spanish, Tuscan	33
<b>CLO</b>	Clovis	1
<b>SAQ</b>	Saqqaq	1

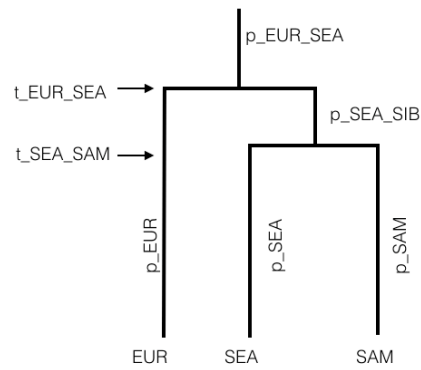
**Suppl. Table 1.1:** A table listing all samples and groups used in the Rarecoal analysis. Population groups are abbreviated as follows: SAM, South Americans; NAM, northern North Americans; ATH, Northern Athabaskans; AARC, people of the American Arctic meta-population belonging to the third migration wave; SIB, Siberians (excluding populations of Chukotka and Kamchatka); SEA, Southeast Asians; EUR, Europeans; CLO and SAQ, the Clovis and Saqqaq ancient Americans. Note: <sup>1)</sup> All individuals are from the two sources: Raghavan et al. 2015 and the Simons Genome Diversity Project data set (Mallick et al. 2016), as indicated in Suppl. Table 1.

#### Fitting a demographic model to six populations including Athabaskans

The final model shown in Figure 5 in the main text was derived in an iterative way: We start off with fitting a model to three populations only, and then add one population at a time, re-estimating all previous and new parameters. In all fits, we used the “rarecoal mcmc” program to estimate maximum likelihood values and confidence intervals for each parameter. We restricted analysis to variants of maximum allele count 4.

#### Fitting a model to Europeans, Southeast Asians and South Americans

We first fitted a simple three population tree with the topology shown in Suppl. Fig. 1.1 and a fixed population size in each internal and external branch.



**Suppl. Fig. 1.1:** A model connecting Europeans, Southeast Asians and Native South Americans.

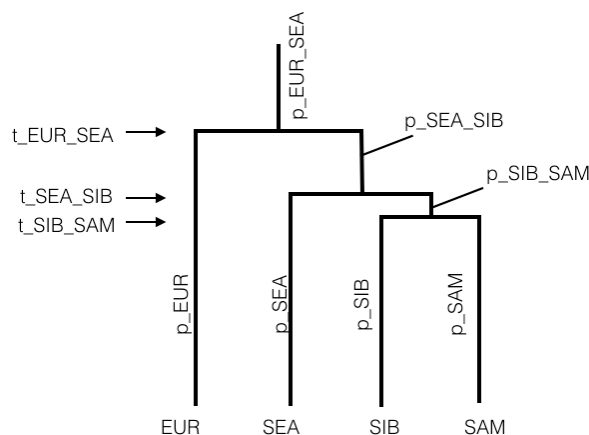
The inferred population size and split time are scaled to real time and size using a mutation rate of  $1.25 \times 10^{-8}$  (Scally and Durbin 2012) and a generation time of 29 years (Fenner 2005) (Suppl. Table 1.2).

Parameter	Scaled estimate
<b>p_EUR</b>	30,417
<b>p_SEA</b>	25,206
<b>p_SAM</b>	5,646
<b>t_SEA_SAM</b>	21,711y
<b>t_EUR_SEA</b>	28,125y
<b>p_SEA_SAM</b>	9,500
<b>p_EUR_SEA</b>	13,686

**Suppl. Table 1.2:** Parameter estimates for the three-population model.

### Adding Siberians

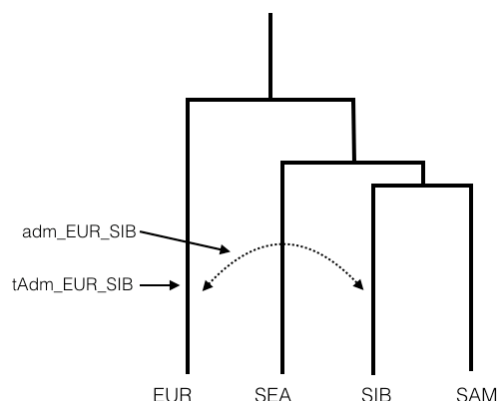
We next added Siberians onto the tree. We first used the “rarecoal find” program to find the most likely branch point of the Siberian meta-population and found that it most likely branches off the South American branch of the tree. We therefore fitted the model shown in Suppl. Fig. 1.2.



**Suppl. Fig. 1.2:** A four-population model including Siberians.

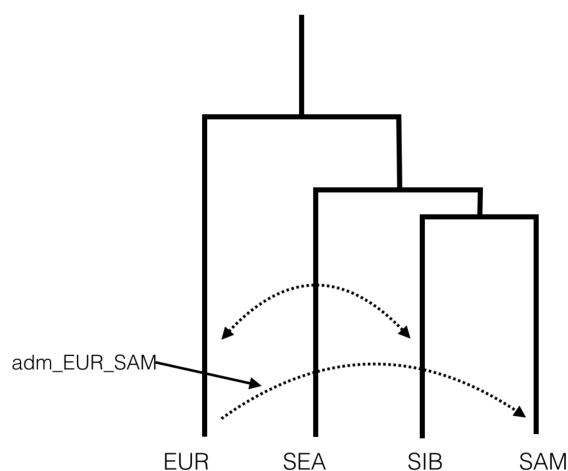
The fitted parameter estimates for this model are shown in Suppl. Table 1.3 (column named “no admixture”). We assessed the quality of this model by plotting the frequency

of shared and private variants (Suppl. Fig. 1.5, bars labeled “no admixture”). As can be seen by the large deviation between the no-admixture model and the real data, this is not a well-fitting model. Apart from various allele-sharing combinations being overestimated by the model, we see that the European/Siberian shared variants are underestimated by about 50%. We therefore tested a more complex model with admixture between EUR and SIB (Suppl. Fig. 1.3).



**Suppl. Fig. 1.3:** A four-population model with European/Siberian admixture.

Note that here admixture is bidirectional and symmetric for now, but we test asymmetric admixture further below. The parameter estimates for this and the previous model are shown in Suppl. Table 1.3. We infer 15% admixture between Europeans and Siberians, in both directions. The fits are much better (Suppl. Fig. 1.5, bars labeled “+AdmSIB\_EUR”), but we still see an underestimation of European/American sharing, so we augmented the model once more to allow for admixture of European ancestry in South Americans (Suppl. Fig. 1.4). Here we use unidirectional admixture because we can heuristically rule out Native American ancestry in Europeans. We also fix the time of this admixture event to reduce the number of parameters. We chose 500y as the admixture time point as the earliest possible time of European (post-Colombian) admixture in Native Americans. Note that the ancient north Eurasian (ANE) ancestry detected in all in Native Americans and shared with Europeans (Raghavan et al. 2014a) is too old to be detected with our approach relying on rare allele sharing (Schiffels et al. 2016).



**Suppl. Fig. 1.4:** A four-population model including European/American admixture.

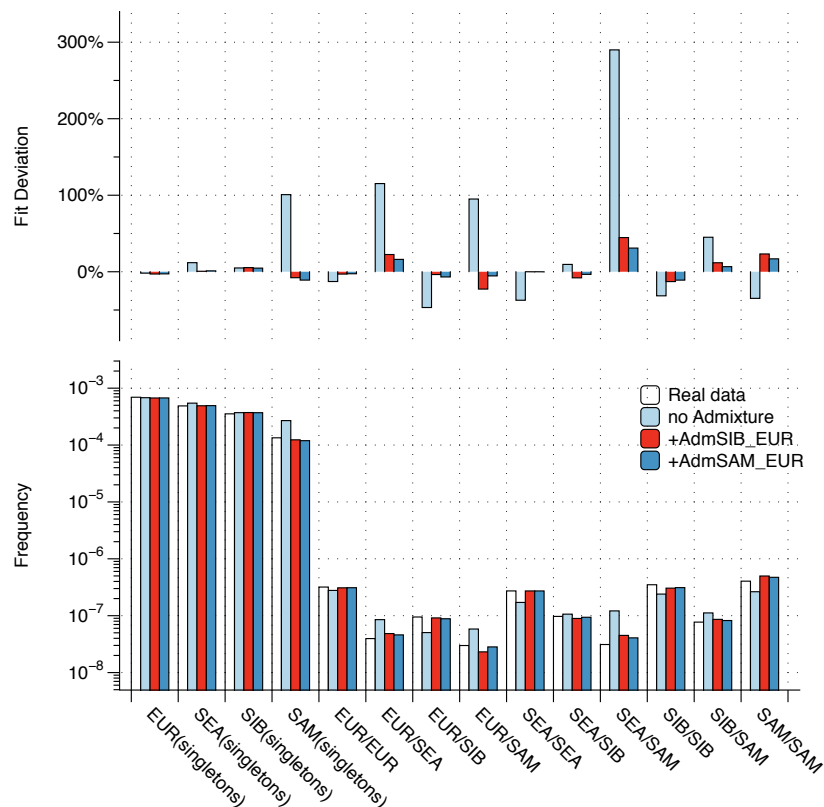
The estimated parameters for this model are shown in Suppl. Table 1.3.



Parameter	No admixture	+ SIB-EUR admixture	+SAM-EUR admixture
p_EUR	32,151	23,526	23,194
p_SEA	9,346,780	40,190	41,469
p_SIB	26,748	23,963	25,440
p_SAM	31,360	3,864	3,290
t_SIB_SAM	20,853y	20,140y	20,134y
t_SEA_SIB	21,358y	20,350y	20,345y
t_EUR_SEA	24,188y	38,568y	38,966y
p_SIB_SAM	1,513	137	132
p_SEA_SIB	3,918	10,277	10,124
p_EUR_SEA	12,850	9,087	8,971
tAdm_EUR_SIB		5,700y	2,284y
adm_EUR_SIB		15.6%	12.3%
adm_EUR_SAM			3.6%

**Suppl. Table 1.3:** Parameter estimates for the three different models for four populations that we tested.

As shown, we infer 3.6% European ancestry in Native Americans. The fits have improved further (Suppl. Fig. 1.5).

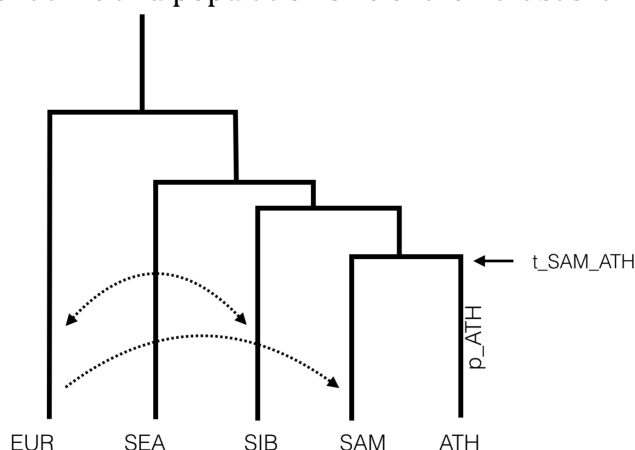


**Suppl. Fig. 1.5:** Fits comparing allele sharing frequencies within and between populations, and frequencies of singletons. Frequencies from real data (white bars) are compared to predictions from different models.

We were satisfied with the quality of the fits for the final four-population model and use this model to add Northern Athabaskans.

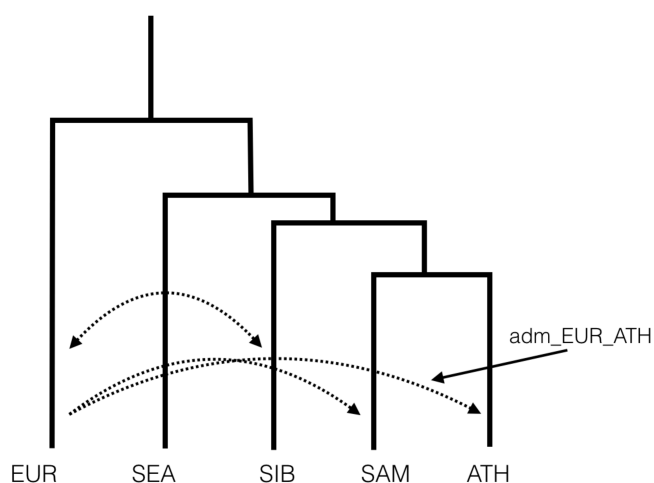
### Adding Athabaskans

We next added the Northern Athabaskan group to the model. We again used “rarecoal find” to estimate the most likely branch point of Athabaskans onto the best four-population model above. We found that the most likely branch to merge is the South American branch. We therefore tested the model shown in Suppl. Fig. 1.6, with two more parameters for the split time and population size of the Athabaskan branch.



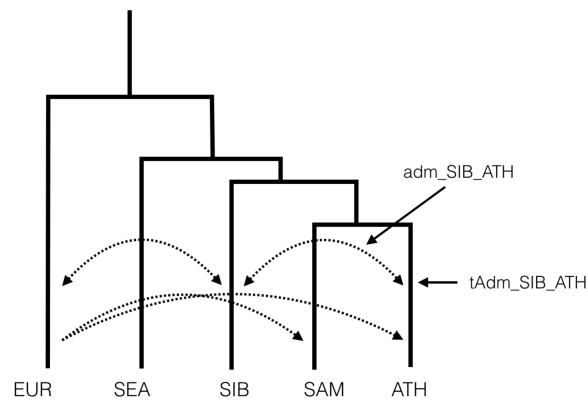
**Suppl. Fig. 1.6:** A five-population model including Athabaskans.

The parameter estimates for this model are shown in Suppl. Table 1.4 (column named “no ATH adm.”). As shown, we estimate the time of the Athabaskan/South American split to be at ~12,800 years ago. The fits of this model reveal an underestimation of Athabaskan/European sharing (Suppl. Fig. 1.9). We therefore added a unidirectional admixture edge at 500y from Europeans into Athabaskans (Suppl. Fig. 1.7).



**Suppl. Fig. 1.7:** A model with European/Athabaskan admixture.

Parameter estimates for this model (Suppl. Table 1.4, column “+ATH-EUR admixture”) reveal about 3% European admixture in Athabaskans. Fits for this model reveal an underestimation of allele sharing between Siberians and Athabaskans (Suppl. Fig. 1.9), so we added one more bidirectional admixture edge between Siberians and Athabaskans (Suppl. Fig. 1.8).



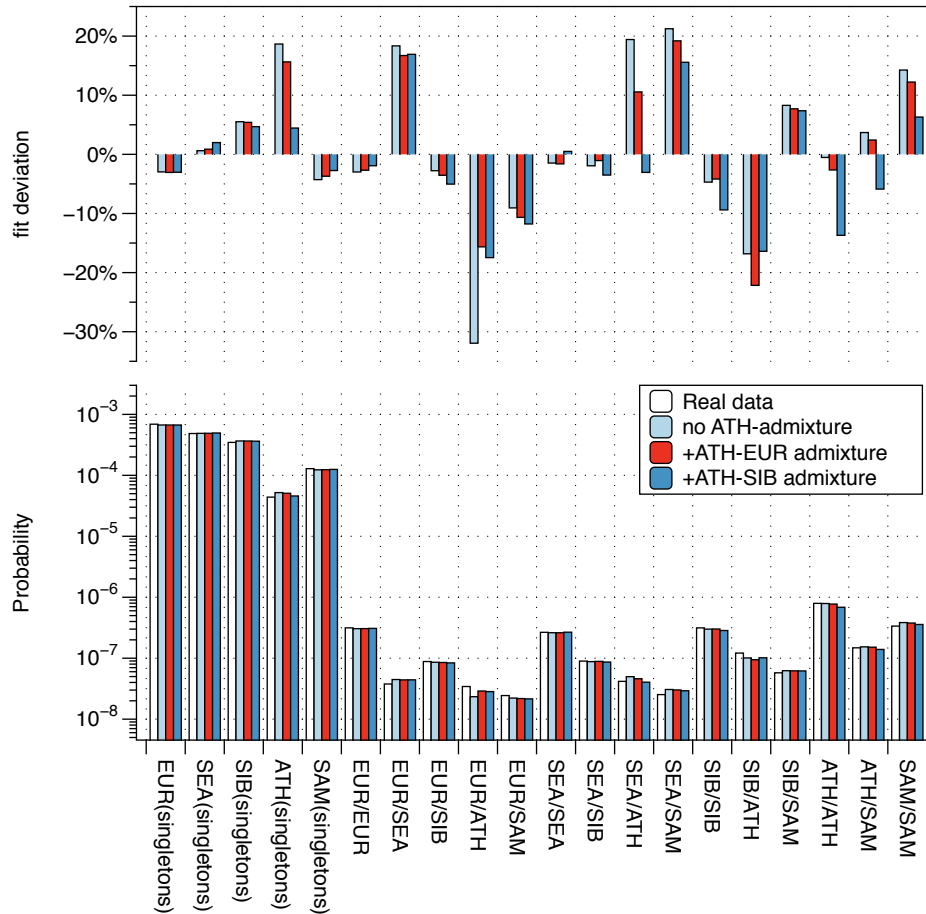
**Suppl. Fig. 1.8:** Adding Siberian/Athabaskan admixture to the five-population model.

Parameter estimates for this model reveal ~5% admixture between Siberians and Athabaskans in both directions (Suppl. Table 1.4, column “+ATH-SIB admixture”), with the time of admixture about 4,400 years ago.

Parameter	No ATH adm.	+ATH-EUR admixture	+ATH-SIB admixture
<b>p_EUR</b>	23,283	23,131	22,987
<b>p_SEA</b>	41,715	42,121	41,947
<b>p_SIB</b>	25,089	25,450	23,974
<b>p_ATH</b>	3,728	3,290	2,808
<b>p_SAM</b>	4,216	4,374	4,745
<b>t_SAM_ATH</b>	12,835y	12,476y	11,483y
<b>t_SIB_SAM</b>	20,139y	20,213y	20,218y
<b>t_SEA_SIB</b>	20,369y	20,393y	20,378y
<b>t_EUR_SEA</b>	38,352y	38,435y	39,034y
<b>p_SAM_ATH</b>	2,202	2,156	2,176
<b>p_SIB_SAM</b>	132	130	134
<b>p_SEA_SIB</b>	9,861	9,702	10,115
<b>p_EUR_SEA</b>	8,928	8,926	8,919
<b>tAdm_EUR_SIB</b>	2,624y	1,810y	2,569y
<b>adm_EUR_SIB</b>	12.6%	11.9%	12.7%
<b>adm_EUR_SAM</b>	3.0%	3.0%	2.8%
<b>adm_EUR_ATH</b>		2.9%	2.4%
<b>tAdm_SIB_ATH</b>			4,419y
<b>adm_SIB_ATH</b>			5.4%

**Suppl. Table 1.4:** Parameter estimates for the five-population models including Athabaskans.

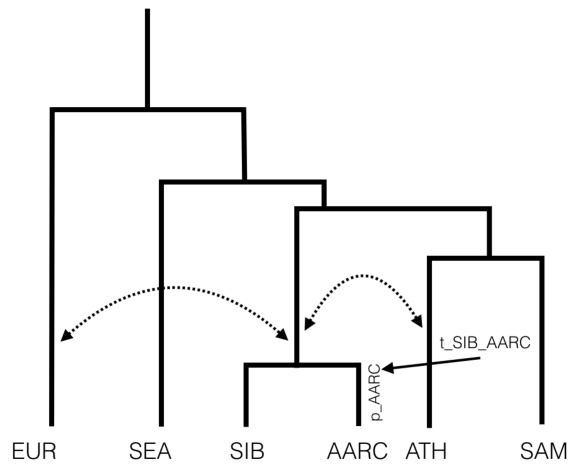
The final model fits are good (Suppl. Fig. 1.9), and we use this model to add American Arctic populations.



**Suppl. Fig. 1.9:** Model fits for the five-population models tested.

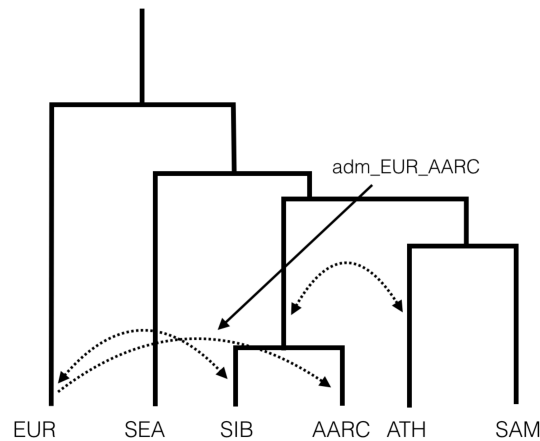
### Adding American Arctic populations

As a last population in our model, we add the American Arctic meta-population, whose ancestors migrated to America about 1000 years ago (Raghavan et al. 2014a). We again used “rarecoal find” to find the most likely branch point of the American Arctic branch onto the best five-population model. We find that the most likely branch point is the Siberian branch, but after the European/Siberian admixture event. We therefore first fitted the model shown in Suppl. Fig. 1.10.



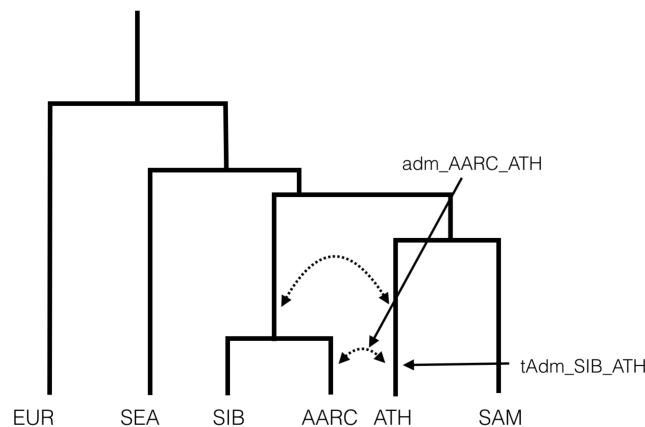
**Suppl. Fig. 1.10:** A six-population model with Arctic populations. Admixture edges from Europe into ATH and SAM are not shown for clarity, but are still part of the model.

Parameter estimates for this and other six-population models tested are shown in Suppl. Table 1.5, with fits shown in Suppl. Fig. 1.14. In this model, the European/Siberian admixture point was estimated to be suspiciously close to the Siberian/American Arctic split point. Therefore, we next tested a model with European admixture separately into the Siberian (bidirectional) and American Arctic groups (unidirectional at 500y) (Suppl. Fig. 1.11) and found that it yields substantially higher likelihood than the model in Suppl. Fig. 1.10. Parameter estimates for this model are also shown in Suppl. Table 1.5 (column “+EUR->AARC adm.”).



**Suppl. Fig. 1.11:** A six-population model with separate admixture into the Siberian and American Arctic branches. Admixture edges from Europe into Athabaskans and South Americans are not shown for clarity, but are still part of the model.

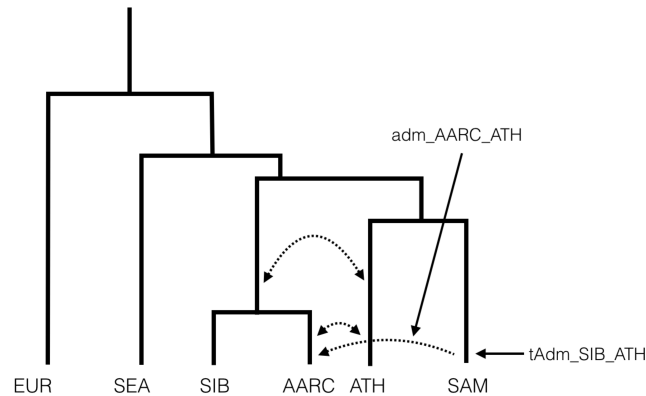
We find that the American Arctic population is ~17% admixed with Europeans. The fits of this model (Suppl. Fig. 1.14) reveal an underestimation of Arctic/Athabaskan sharing, so we added a bidirectional admixture edge between these two populations (Suppl. Fig. 1.12).



**Suppl. Fig. 1.12:** Adding admixture between the American Arctic group and Athabaskans with two more parameters.

We infer ~14% admixture between the American Arctic group and Athabaskans in both directions (Suppl. Table 1.5, column “+AARC<->ATH”). The fits of this model show an underestimation of alleles shared between the American Arctic group and South Americans, so the final model tested contains a unidirectional admixture edge from

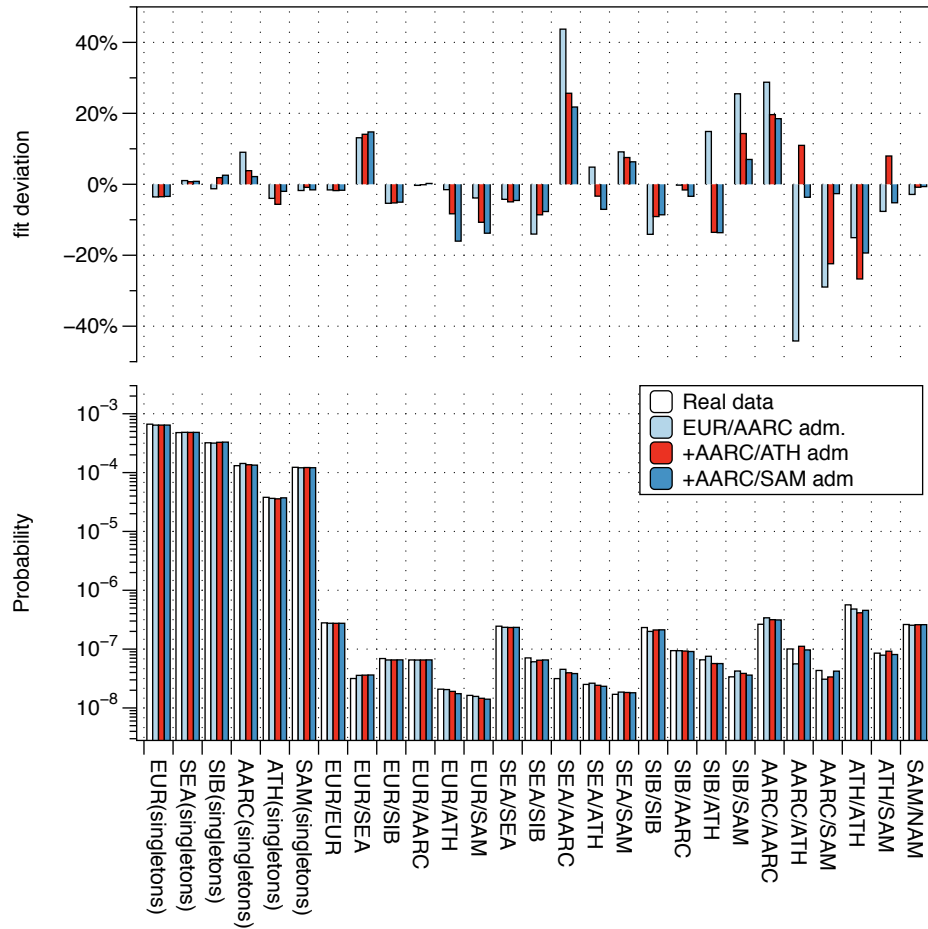
South Americans into the American Arctic branch (Suppl. Fig. 1.13). From this model we infer ~4% South American admixture in the American Arctic group.



**Suppl. Fig. 1.13:** Adding admixture from South Americans into the American Arctic group.

Parameter	No AARC adm.	+EUR->AARC adm.	+AARC<->ATH	+AARC<->SAM
p_EUR	22,015	23,046	23,424	23,491
p_SEA	42,458	43,271	43,404	43,068
p_SIB	7,231	12,432	14,431	15,247
p_AARC	1,981	1,861	1,398	1,347
p_ATH	2,027	2,040	2,256	2,288
p_SAM	6,164	6,135	6,285	5,414
t_SIB_AARC	2,482y	4,200y	4,256y	4,335y
t_SAM_ATH	9,254y	9,415y	9,460y	11,233y
t_SIB_SAM	20,247y	20,238y	20,290y	20,213y
t_SEA_SIB	20,423y	20,378y	20,412y	20,407y
t_EUR_SEA	39,791y	39,410y	38,060y	37,958y
p_SIB_AARC	44,902	37,818	30,082	29,078
p_SAM_ATH	2,253	2,159	1,956	1,603
p_SIB_SAM	187	161	162	162
p_SEA_SIB	9,619	9,668	9,613	9,666
p_EUR_SEA	8,841	8,696	8,953	8,972
tAdm_EUR_SIB	2,492y	4,066y	2,462y	2,246y
adm_EUR_SIB	17.4%	16.1%	13.0%	12.6%
adm_EUR_SAM	1.8%	2.0%	2.4%	1.7%
adm_EUR_ATH	0.6%	0.8%	0.0%	0.0%
tAdm_SIB_ATH	5,239y	5,067y	7,397y	6,816y
adm_SIB_ATH	25.5%	22.5%	10.5%	9.0%
adm_EUR_AARC		16.8%	21.9%	22.4%
tAdm_AARC_ATH			493y	483y
adm_AARC_ATH			13.9%	10.7%
tAdm_AARC_SAM				482y
adm_AARC_SAM				3.7%

**Suppl. Table 1.5:** All parameter estimates for the six-population models tested.



**Suppl. Fig. 1.14:** Fits of the six-population models tested.

### Final model: Asymmetric migration

In the models above we have constrained the Siberian/European, Siberian/Athabaskan and American Arctic/Athabaskan admixtures to be bidirectional and symmetric. We now relax this constraint and try to estimate a separate admixture rate in each direction. We still enforce both admixtures to occur at the same time to reduce the overall number of parameters. This final model is shown in Figure 5, and the parameter estimates for this final model are shown in Suppl. Table 1.6 (parameter estimates and confidence intervals are also shown in Table 1 of the main text). Note that admixture parameters follow the rule “adm\_FROM\_TO” forward in time. As can be seen, allowing asymmetric admixture rates reveals ~23% admixture from Siberians into Northern Athabaskans, and only ~7% in the opposite direction.

### Estimating Siberian admixture in other northern North Americans

To assess whether the Siberian admixture inferred in Athabaskans is also present in other northern North Americans, we tested the final model with asymmetric migration shown above on a data set where Athabaskans are replaced with non-Na-Dene speaking northern North Americans (NAM, see Data section above). On this data, the estimated model parameters are shown in Suppl. Table 1.6. As can be seen, we still estimate ~10% Siberian admixture into northern North Americans (compare with 23% from Siberians into Athabaskans). However, the time of this admixture event (~600y) is extremely recent, and moreover after the European admixture event into Siberians. We think that

this may reflect recent admixture between Athabaskans and other Northern Americans. In any case, the signal is weaker and too recent to reflect the same historical admixture event that is seen in the Athabaskans.

Parameter	Maximum Likelihood estimate (NAM)	Maximum Likelihood estimate (ATH)
p_EUR	25,714	25,101
p_SEA	44,620	44,242
p_SIB	10,303	13,568
p_AARC	780	1,173
p_NAM	5,280	1,851
p_SAM	5,664	6,552
t_SIB_AARC	2,580y	4,126y
t_SAM_NAM	11,792y	9,744y
t_SIB_SAM	20,271y	20,290y
t_SEA_SIB	20,374y	20,402y
t_EUR_SEA	35,588y	36,095y
p_SIB_AARC	28,612	27,469
p_SAM_NAM	1,610	1,762
p_SIB_SAM	141	147
p_SEA_SIB	8,753	9,012
p_EUR_SEA	9,341	9,315
tAdm_EUR_SIB	1,753y	2,327y
adm_EUR_SIB	16.6%	16.1%
adm_SIB_EUR	6.7%	8.0%
adm_EUR_SAM	3.1%	2.7%
adm_EUR_NAM	29.2%	0.7%
tAdm_SIB_NAM	595y	6,940y
adm_SIB_NAM	9.6%	22.9%
adm_NAM_SIB	1.6%	6.8%
adm_EUR_AARC	18.8%	25.0%
tAdm_AARC_NAM	5y	490y
adm_AARC_NAM	4.8%	7.6%
adm_NAM_AARC	17.4%	11.5%
tAdm_AARC_SAM	2,570y	488y
adm_SAM_AARC	21.9%	7.6%

**Suppl. Table 1.6:** Comparing a model using the northern North American meta-population instead of Athabaskans.

### Testing the model with simulations

To test the robustness of our estimates we simulated the final six-population model with the Athabaskans under the full coalescent with recombination. We simulated 25 chromosomes, each with 100 Mb, and realistic mutation- and recombination rates ( $1.25 \times 10^{-8}$  and  $1 \times 10^{-8}$ , respectively). Using a scaling coefficient of  $N_0 = 20,000$ , the command line for the scrm simulator (Staab et al. 2015) is:



```

scrm 218 25 -t 100000.000000 -r 80000.000000 100000000 -l
100000 -I 6 66 44 44 28 8 28 -en 0.000000 1 1.254710 -en
0.000000 2 2.217343 -en 0.000000 3 0.672266 -en 0.000000 4
5.884996e-2 -en 0.000000 5 9.236993e-2 -en 0.000000 6 0.329472
-eps 2.071931e-4 4 6 0.925696 -eps 2.122413e-4 5 4 0.925150 -
eps 2.122413e-4 4 5 0.883895 -eps 2.155000e-4 6 1 0.973245 -
eps 2.155000e-4 5 1 0.993140 -eps 2.155000e-4 4 1 0.750078 -
eps 9.907748e-4 3 1 0.839261 -eps 9.907748e-4 1 3 0.919703 -ej
1.774119e-3 4 3 -en 1.774119e-3 3 1.381041 -eps 2.936409e-3 5
3 0.770043 -eps 2.936409e-3 3 5 0.932048 -ej 4.186964e-3 5 6 -
en 4.186964e-3 6 8.825812e-2 -ej 8.729929e-3 6 3 -en
8.729929e-3 3 7.386311e-3 -ej 8.797595e-3 3 2 -en 8.797595e-3
2 0.451562 -ej 1.557382e-2 2 1 -en 1.557382e-2 1 0.465810

```

Note that “rarecoal simCommand” generates this command line from a model template and a model estimate (see the Rarecoal documentation for details). The above command line simulates 109 diploid samples (218) haplotypes, with exactly the same distribution of haplotypes on populations as in the real data.

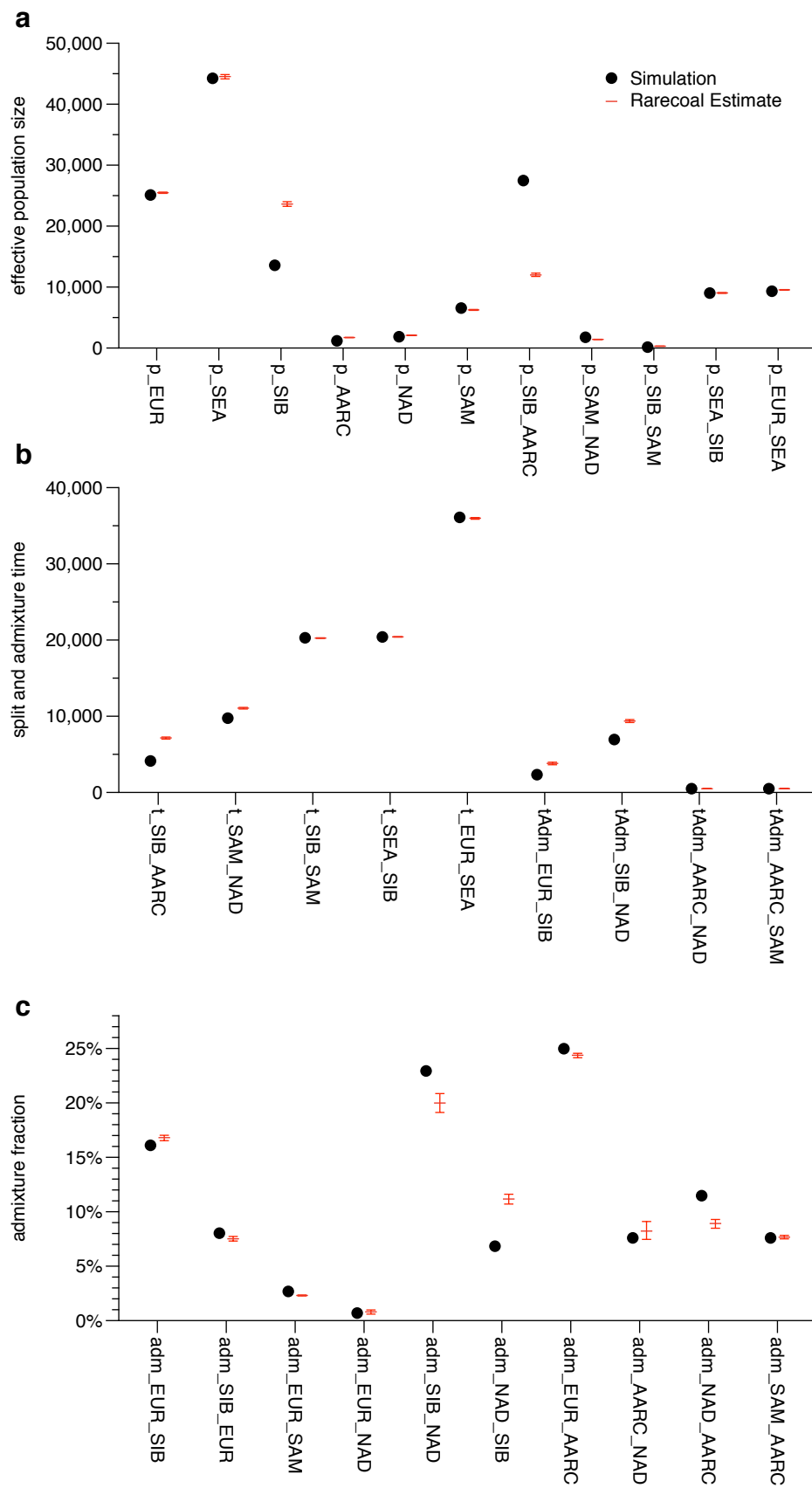
We then estimated parameters from the simulated data using Rarecoal and checked whether the inferred parameters match the simulated parameters. The results are summarized in Suppl. Fig. 1.15. As can be seen, most parameters are estimated very accurately, in particular all time estimates of splits and admixture events. Substantial deviation between simulated and estimated parameter is seen in the population size estimate of the Siberian branch, as well as the ancestral branch of Siberians and the American Arctic group.

### Mapping Saqqaq and Clovis onto the model

Finally, we used “rarecoal find” to test all points on the tree (before the age of a sample) for the likelihood of that sample’s branch to merge onto the tree. In this analysis we restricted the evaluation of the likelihood to variants between allele counts 2 and 4. Singletons are excluded because they can be enriched for false positives in ancient samples (this is similar to the analysis in Schiffels et al. 2016). The results are shown in Fig. 5B, C.

### References (for this section)

- Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128**, 415–423 (2005).
- Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* (2016), in press.
- Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345**, 1255832 (2014a).
- Raghavan, M. *et al.* Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, 1–20 (2015).
- Scally, A. & Durbin, R. Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* **13**, 745–753 (2012).
- Schiffels, S. *et al.* Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nat. Commun.* **7**, 10408 (2016).
- Staab, P. R. *et al.* scrm: efficiently simulating long sequences using the approximated coalescent with recombination. *Bioinformatics* **31**, 1680–1682 (2015).



**Suppl. Fig. 1.15:** Comparison of estimated vs. simulated parameters.

## Suppl. text 2. Conflicting interpretations of Na-Dene population history

The study by Raghavan et al. (2015) has reported no evidence of the Siberian-Dakelh gene flow after 12,000 YBP, using *diCal2.0* and other methods. On the other hand, a likely bidirectional gene flow between Dakelh and Neo-Eskimo groups was supported by ADMIXTURE, TreeMix, *D*-statistics employing both whole genome and SNP array genotype data, and by outgroup  $f_3$ -statistics using whole genome data (Raghavan et al. 2015). Neo-Eskimo ancestry was suggested for northern North American groups Haida and Nisga'a based on the *D*-statistic of the form (First American population, test population; Greenlander Inuit, Yoruba) on unlinked SNP array data, but similar statistics for Na-Dene (Chipewyans and Southern Athabaskans) as a 'test population' were not significant, with  $|Z\text{-scores}| < 3$  (Raghavan et al. 2015). We argue that significantly negative values of genome-based *D*-statistic of the form (Dakelh, South American population; outgroup, American Arctic population) (Raghavan et al. 2015) cannot be interpreted in a straightforward way as a sign of Neo-Eskimo gene flow. Neo-Eskimo, Siberian and Paleo-Eskimo populations form a clade that separated much later than South Americans and Na-Dene (according to this study and Raghavan et al. 2015), thus differentiating the Neo- vs. Paleo-Eskimo sources of gene flow requires additional analyses. A similar genome-based *D*-statistic (Dakelh, South American or American Arctic population; outgroup, Saqqaq) suggested that Saqqaq is closer to Neo-Eskimos vs. Dakelh and equidistant from Dakelh and South Americans (Raghavan et al. 2014a), which is possibly compatible with a low level of Paleo-Eskimo ancestry in Athabaskans. Paleo-Eskimo admixture in Na-Dene was also not revealed by ADMIXTURE and TreeMix (Raghavan et al. 2015), but interpreting results of these methods is often complicated by population-specific drift, population size biases and admixture from unsampled "ghost" populations (Falush et al. 2016).

Let us consider the following example. TreeMix analyses by Raghavan and co-authors (2014a) consistently revealed admixture edges between the 4,000-year-old Saqqaq genome (belonging to the Siberian clade) and the basal node of Neo-Eskimos, and that was interpreted as a sign of Neo-Eskimo admixture in Paleo-Eskimos, pre-dating their entry into America. However, in other trees reconstructed with TreeMix, Saqqaq branched with Neo-Eskimos, and in that case a migration edge appeared between Saqqaq and Siberians (Raghavan et al. 2014a). Thus, TreeMix, similarly to ADMIXTURE (Rasmussen et al. 2010, Raghavan et al. 2014a, Flegontov et al. 2016), modeled the Saqqaq genome as a mixture of Siberian and Arctic gene pools, and fitted this scenario by two possible topologies of branches and migration edges. In our view, Saqqaq is unlikely to be a mixture of the Siberian and Arctic groups, which under our model split around 4,000-4,200 YBP (Fig. 5), after the Paleo-Eskimo migration into America. Application of ADMIXTURE to Paleolithic (Lazaridis et al. 2014, Raghavan et al. 2014b) and even to less ancient genomes is problematic: "Indeed, if an ancient sample is put into a dataset of modern individuals, the ancient sample is typically represented as an admixture of the modern populations, which can happen even if the individual sample is older than the split date of the modern populations and thus cannot be admixed. A similar effect can happen when a source population is put into a dataset with two or more drifted sink populations. The source can be represented as a mix, even though there is no mixture within its history." (Falush et al. 2016)

Notably, a *D*-statistic (Karitiana, Dakelh; outgroup, Asian population) based on genome

sequencing data including both common and rare variants, has shown that not only Arctic populations, but also Dai and Ket are significantly closer to Athabaskans as compared to Karitiana ( $|Z\text{-score}| > 3$ ) (Raghavan et al. 2015). Remarkably, only Kets, but not other populations representing major genetically distinct Siberian sub-groups (Altaians, Buryats, Nivkhs, and Yakuts), showed a statistically significant  $D$ -statistic. This minor observation agrees with the results of the present study and with various analyses emphasizing a close relationship of Kets and Saqqaq (Flegontov et al. 2016).

#### *References (for this section)*

- Falush, D., van Dorp, L., Lawson, D. A tutorial on how (not) to over-interpret STRUCTURE/ADMIXTURE bar plots. *bioRxiv* doi: <http://dx.doi.org/10.1101/066431> (2016).
- Flegontov, P. *et al.* Genomic study of the Ket: A Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry. *Sci. Rep.* **6**, 20768 (2016).
- Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
- Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345**, 1255832 (2014a).
- Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014b).
- Raghavan, M. *et al.* Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, 1–20 (2015).
- Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).

### **Suppl. text 3. Estimating the time depth of the Dene-Yeniseian linguistic connection**

***Edward J. Vajda***

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

The timing of the Dene-Yeniseian language split could shed important light on Native American as well as North Asian prehistory. In attempting to reconcile the apparent closeness of Yeniseian and Na-Dene grammatical homologies with the much greater genetic distance between Ket and Na-Dene speakers, the various papers in Kari & Potter (2010) offered three possible scenarios for the Dene-Yeniseian connection: 1) a Late Pleistocene separation connected with the Paleo-Indian migrations into the Americas, with an extraordinary slow rate of linguistic change; 2) a separation involving a back migration of Yeniseians from Beringia; and 3) an Early to Mid-Holocene separation connected with the entrance into Alaska around 4,800 YBP by the population that later developed the Arctic Small Tool tradition (ASTt). The first two scenarios can now be excluded.

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

between languages (lexicostatistics) from the use of such data to assign precise dates for prehistoric language splits based on an assumed regular rate of linguistic change (glottochronology), which in fact does not exist across languages or even in a single language over time. While rejecting glottochronology, McMahon & McMahon (2005:204) support the value of gathering and comparing lexicostatistic data, which then can sometimes be useful for purposes of dating when combined with facts from other disciplines such as archaeology and genetics. Several types of evidence can potentially be combined with evidence of shared vocabulary and grammatical homologies to help narrow the range of plausible separation dates between related languages. For Dene-Yeniseian, all of them suggest a split roughly between 10,000 and 8,000  $\pm$ 500 YBP. The shallower end is favored by the detailed morphological homologies shared by the two families (Nichols 2010). The deeper end, which is suggested by the more meager number of shared lexical cognates, would still be far too shallow to match a connection with the earliest Paleo-Indian migrations during the Late Pleistocene. However, this range does provide a realistic temporal parallel for the migration of ASTt ancestors from North Asia into the Americas about 4,800 YBP. If this population consisted of Pre-Proto-Na-Dene speakers, then the split with their Yeniseian-speaking cousins in south-central Siberia would necessarily have been earlier.

Most previous calculations by historical linguists place the timeline for the internal diversification of Na-Dene languages within 4,800 YBP. The Na-Dene family contains the widespread Athabaskan (Dene) languages, which together are equally related to the recently extinct Eyak language of coastal Alaska. All Athabaskan languages, whether spoken in Alaska, Canada, California, or Arizona, share over 70% cognates in basic vocabulary, the number becoming higher if the list includes words associated with northern boreal lifestyle, such as ‘birch’, ‘wolverine’, etc. Krauss (1976:330) showed that all Athabaskan languages share 33% of basic vocabulary from the 100-word Swadesh List with Eyak. Athabaskan-Eyak, in turn, is clearly more distantly related to the Tlingit dialect cluster spoken in the Alaskan Panhandle and parts of interior Yukon Territory (Heggarty & Renfrew 2014:1236). Using a variety of lexicostatistic methods and reliable data, Krauss (1976:333) estimated a time depth for Proto-Athabaskan of 2400  $\pm$ 500 years and for Athabaskan-Eyak of 3400  $\pm$ 500 years. Estimates for the earlier breakup of Tlingit and Athabaskan-Eyak range from 5000 years (Swadesh 1958) to as shallow as 3500 years (Kaufman & Golla 2000), with an estimate of 4,500 years by Krauss (1980:11-13). The deeper dates would be favored by the known conservatism of Na-Dene languages and also by the fact that the phylogenetic relationship between Athabaskan-Eyak-Tlingit (Na-Dene) was universally accepted only in the past decade, despite being suspected for over a century (Campbell 2011). The late acceptance date derives mainly from the fact that before Leer (2010), the evidence for Athabaskan-Eyak-Tlingit in the form of shared finite verb structure significantly outweighed the expected parallel lexical evidence, making it unclear whether language mixing rather than genetic inheritance was involved in the historical similarities between these languages.

The relatedness between Athabaskan languages, despite their far flung geography, is close enough that it has never been in doubt (Campbell 1997), though no subgrouping beyond the geographic one separating Pacific Coast (Hupa, Tolowa, etc.) and Southwestern Athabaskan (Navajo and Apache) from Northern Athabaskan (the remaining languages in Canada and Alaska) has yet been demonstrated. This suggests a rapid spread from a common source, most likely somewhere in Northwestern Canada

near the current border between British Columbia and Alaska or in adjacent parts of Interior Alaska. Another support for a recent dispersal is the high rate of mutual intelligibility between geographically distant Athabaskan languages (Krauss 1976). Some scholars posit a time depth for Proto Athabaskan as shallow as 2,000 YBP (Kaufman & Golla 2000), though a date closer to 3,000 is more likely given the resistance to borrowing observed with all of these languages. A time depth of at least 2,500 years for Athabaskan, following the estimate in Krauss (1976), would concur with the westward spread of the Taltheilei Culture beginning 2,750 YBP, which has been associated with the spread of Athabaskan speakers (Potter 2010, Kari 2010).

The interior Alaskan and northwestern Canadian portions of the Athabaskan range show no clear archaeological evidence of prehistoric population replacement during the past 8000 years (Potter 2010, Kari 2010). For this reason, Kari (2010) posits that the Athabaskans have lived in interior northwestern North America for at least that span of time. Kari cites the near complete absence of substrate place names in the Northern Athabaskan areas as evidence for their ancient occupation of these areas. However, the Navajo and Apache areas of the American Southwest likewise have virtually no toponymic substrate from the languages previously spoken there, yet the Athabaskan presences there dates no farther back than 1,200 YBP. This reflects a strong Athabaskan avoidance of borrowing place names rather than ancient occupancy. In any event, such a degree of linguistic conservatism, whereby geographically distant languages maintain mutual intelligibility over a span of 8,000 years, would be unique and unprecedented. After adjusting for the conservatism of Na-Dene languages, retention rates for vocabulary and grammatical structures would appear to support a time depth of 5000  $\pm$  500 years for the ancestral Athabaskan-Eyak-Tlingit language (i.e., Proto-Na-Dene). This concurs well with the possibility that the language ancestral to Na-Dene could have been introduced around 4,800 YBP into Alaska by North Asian immigrants associated with the later development and spread of the ASTt. Also probably connected with these “Paleo-Eskimos” is the spread of other elements of North Asian material culture and folklore (Alekseenko 1995; Berezkin 2015) to the Na-Dene, including bow and arrow technology, thought to have been introduced into California 1,500 years ago by the ancestors of the Hupa and other Pacific Coast Athabaskans (Golla 2011:245).

Like the Athabaskan family, Yeniseian languages are obviously related genealogically. Ket and its now extinct relatives (Yugh, Kott, Assan, Arin, and Pumpokol) were recognized as closely related more than 150 years ago (Vajda 2001). Studies of substrate toponyms (Vajda, in press/2) show that the known Yeniseian daughter branches (excepting the Ket-Yugh sub-branch) had already diversified by 2,000 YBP, when Turkic and Uralic-speaking pastoralists started displacing them in most of their southern and western territory, acquiring Ket-related river names and other substrate linguistic elements in the process. If the main sub-branching existed 2,000 years ago, the family is clearly older. The high rate of shared cognates in basic vocabulary (over 70%) between Ket and Kott, which belong to different primary branches of the family, suggest that Proto-Yeniseian must be at least 2,500 to 3,000 years, if not older, which would roughly match the more plausible estimates of time depth for Athabaskan. It is possible to reconstruct Proto-Yeniseian vocabulary (Starostin 1995) and many aspects of grammatical structure (Vajda 2013; Vajda, in press/1) with a high degree of confidence. If Para-Yeniseian linguistic relatives once existed in other parts of North Asia, the influx of pastoral tribes from the south must have obliterated them during the past 3,000

years, leaving no observable traces. Taking into account the probability of language extinction, the breakup of the earliest Proto-Yeniseian language, one predating the form reconstructable on the basis of Ket and Kott, could conceivably have begun much earlier than 3,000 YBP.

All Na-Dene languages share innovations demonstrating their equidistance from Yeniseian, whose split from the language ancestral to Na-Dene must be significantly older than Proto-Na-Dene itself. To cite one particularly vivid example, Pre-Proto-Na-Dene restructured three of its inherited Dene-Yeniseian verb prefixes into the so-called classifier complex, for which the family is well known. All three component prefixes have cognates in Yeniseian but did not develop the characteristic function of transitivity increase and decrease found of all Na-Dene languages (Vajda, in press/1). Contrary to Holton and Sicoli (2014), there is no linguistic or genetic evidence indicating a back migration into Asia of Yeniseian speakers from Beringia after Na-Dene had already begun to diversify.

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

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



back migration from Beringia. Rather, the connection of Dene-Yeniseian with the ASTt migration, first suggested explicitly by Dumond (2010), appears increasingly plausible.

### *References (for this section)*

- Alekseenko, E. A. K izucheniju mifologicheskikh paralelej medvezh'emu kul'tu ketov [Mythological parallels to the Ket Bear Cult]. *Sistemnye issledovanija vzaimosvjazi drevnikh kul'turr Sibiri i Severnoj Ameriki. Vypusk 2: Dukhovnaja kul'tura*. St. Petersburg: RAN, 22-46. (1995).
- Berezkin, Yuri. 2015. Sibirskij fol'klor i proiskhozhdenie na-dene [Siberian folklore and Na-Dene origins]. *Arkheologija, ètnografija i antropologija Evrazii* 43.1: 122-134.
- Campbell, Lyle. *American Indian languages: the historical linguistics of Native America*. Oxford: Oxford University Press (1997).
- Campbell, Lyle. Review of "The Dene-Yeniseian Connection". *International Journal of American Linguistics* 77.3, 445-451 (2011).
- Campbell, Lyle. *Historical linguistics: an introduction* (3<sup>rd</sup> edition). Cambridge, Mass.: MIT Press (2013).
- Dul'zon, A. P. Ketskie toponimy Zapadnoy Sibiri [Ket toponyms of Western Siberia]. *Uchenye Zapisky Tomskogo Gosudarstvennogo Pedagogicheskogo Instituta [Scholarly Proceedings of Tomsk State Pedagogical Institute]* 18, 91-111 (1959).
- Dul'zon, A. P. Byloe rasselenie ketov po dannym toponimiki [The former settlement of the Kets according to the facts of toponymy]. *Voprosy Geografii* 68, 50-84 (1962).
- Dumond, Don. 2010. "The Dene arrival in Alaska." *The Dene-Yeniseian connection*, ed. Kari, J., Potter, B. A. *Anthropological Papers of the University of Alaska: New Series* 5, 335-346 (2010).
- Golla, Victor. *California Indian languages*. Berkeley, Los Angeles, London: University of California Press (2011).
- Heggarty, Paul, & Colin Renfrew. *The Americas: languages. Cambridge World Prehistory*. Cambridge: Cambridge University Press, 1326-1353 (2014).
- Holton Gary, & Mark Sicoli. 2014. Linguistic phylogenies support back-migration from Beringia to Asia. *PLoS ONE* 9.3: e91722. doi:10.1371/journal.pone.0091722
- Kari, Jim. 2010. "The concept of geolinguistic conservatism in Na-Dene prehistory." *The Dene-Yeniseian connection*, ed. Kari, J., Potter, B. A. *Anthropological Papers of the University of Alaska: New Series* 5, 194-222.
- Kari, Jim, & Ben Potter (eds.). 2010. *The Dene-Yeniseian connection. Anthropological papers of the University of Alaska: New Series* 5. Fairbanks, AK: ANLC (2010).
- Kaufman, Terry, & Victor Golla. Language groupings in the New World: their reliability and usability in cross-disciplinary studies." *America past, America present: genes and languages in the Americas and beyond*, ed. Renfrew, C. Cambridge: Macdonald Institute for Archaeological Research, 47-57 (2000).
- Krauss, Michael. Na-Dene. *Native languages of the Americas*, vol. 1, ed. Sebeok, T. A. New York & London: Plenum Press, 283-358 (1976).
- Krauss, Michael. *Alaska native languages: past, present and future*. Fairbanks, AK: ANLC (1980).
- Krauss, Michael. Athabaskan tone. *Athabaskan prosody*, ed. Hargus, S., Rice, K, Amsterdam & New York: John Benjamins, 55-136 (2005).
- Leer, Jeff. *Comparative Athabaskan lexicon*, scan of handwritten materials available online at: [www.uaf.edu/anla/collections/ca/cal/ \(2006\)](http://www.uaf.edu/anla/collections/ca/cal/ (2006)).
- Leer, Jeff. The palatal series in Athabaskan-Eyak-Tlingit with an overview of the basic sound correspondences. *The Dene-Yeniseian connection*, ed. Kari, J., Potter, B. A. *Anthropological Papers of the University of Alaska: New Series* 5, 168-193 (2010).
- McMahon, April, Robert McMahon. *Language classification by numbers*. Oxford: Oxford University Press (2005).
- Nichols, Johanna. Proving Dene-Yeniseian genealogical relatedness." *The Dene-Yeniseian connection*, ed. Kari, J., Potter, B. A. *Anthropological Papers of the University of Alaska: New Series* 5, 299-309 (2010).
- Ruhlen, M. The origin of the Na-Dene. *Proc. Natl. Acad. Sci. USA* 95, 13994-13996 (1998).
- Starostin, George. Dene-Yeniseian and Dene-Caucasian: pronouns and other thoughts. *Working papers in Athabaskan languages* 8. Fairbanks, AK: ANLC, 107-117 (2010).
- Starostin, Sergei A. Sravnitel'nyj slovar' enisejskikh jazykov [A comparative vocabulary of Yeniseian languages]. *Ketskij sbornik* vol. 4, Moscow: Vostochnaja Literatura, 176-315 (1995).
- Swadesh, Morris. Some new glottochronological dates for Amerindian linguistic groupings. *Proceedings of the 32<sup>nd</sup> International Congress of Americanists*, 670-674 (1958).
- Trombetti, A. *Elementi di glottologia*. Bologna: Nicola Zanichelli. pp. 486, 511 (1923).

- Vajda, E. J. Yeniseian peoples and languages: A history of their study with an annotated bibliography and a source guide. Surrey, England: Curzon Press (2001).
- Vajda, E. J. Loanwords in Ket. *The Typology of Loanwords*, ed. Haspelmath, M., Tadmor, U. Oxford: Oxford University Press, 125–139 (2009).
- Vajda, E. J. Siberian link with Na-Dene languages. *The Dene-Yeniseian Connection*, ed. Kari, J., Potter, B. A. *Anthropological Papers of the University of Alaska: New Series* 5, 33–99 (2010a).
- Vajda, E. J. Yeniseian, Na-Dene, and historical linguistics. *The Dene-Yeniseian Connection*, ed. Kari, J., Potter, B. A. *Anthropological Papers of the University of Alaska: New Series* 5, 100–118 (2010b).
- Vajda, Edward. Vestigial possessive morphology in Na-Dene and Yeniseian. *Working Papers in Athabaskan (Dene) Languages 2012*. (Alaska Native Language Center Working Papers No. 11). Fairbanks: ANLC, 79–91 (2013).
- Vajda, Edward. 2016. Dene-Yeniseian. *Oxford Research Encyclopedia of Linguistics*. Oxford Online (2016).
- Vajda, Edward. "Patterns of innovation and retention in templatic polysynthesis." *Handbook of Polysynthesis*, ed. Fortescue, M., Mithun, M., Evans, N.. Oxford: Oxford University Press (in press/1).
- Vajda, Edward. Yeniseian and Athapaskan toponyms. *Arctic Anthropology*. (in press/2).
- Werner, Heinrich. Zur jenesische-indianischen Urverwandtschaft. Wiesbaden: Harrassowitz (2004).
- Werner, Heinrich. Die Jenesische-Sprachen des 18. Jahrhunderts [Yeniseian languages of the 18th century]. Wiesbaden: Harrassowitz (2005).