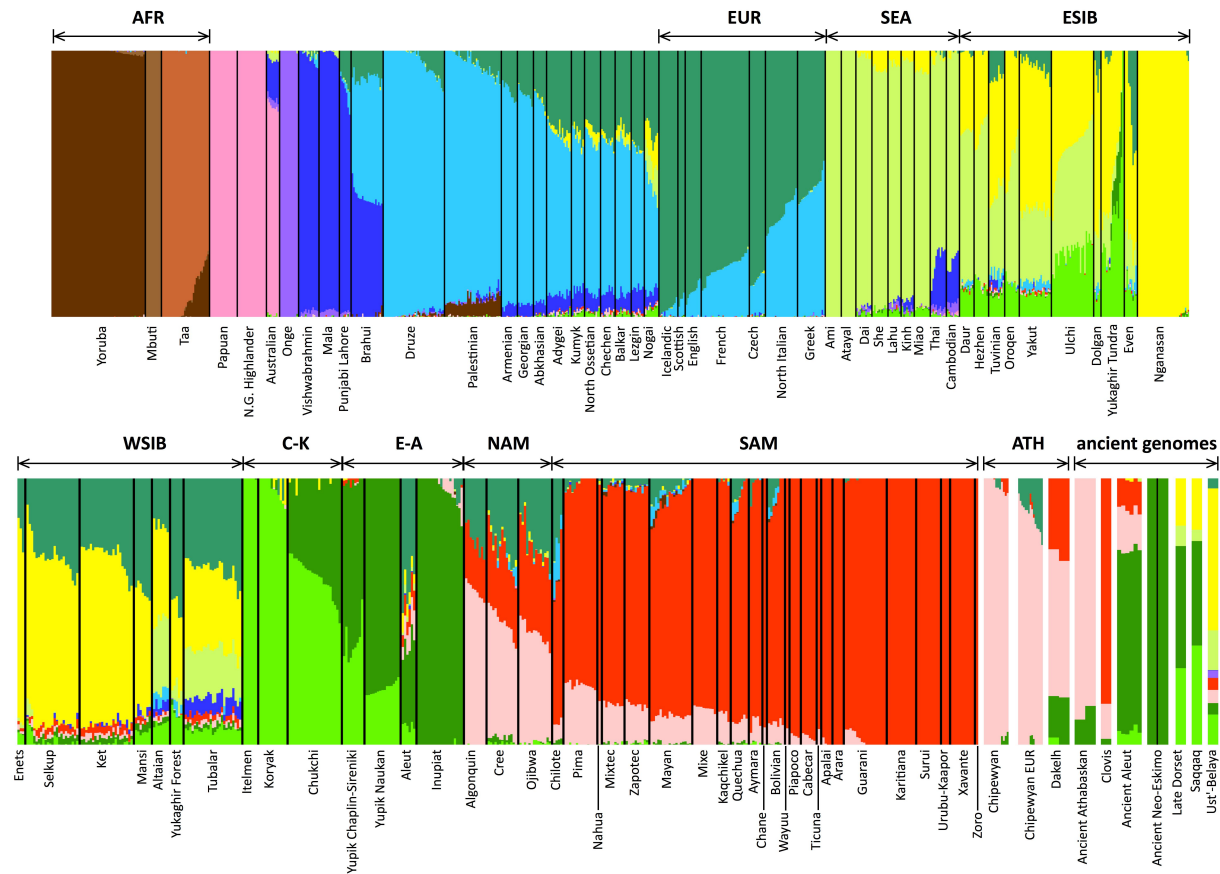


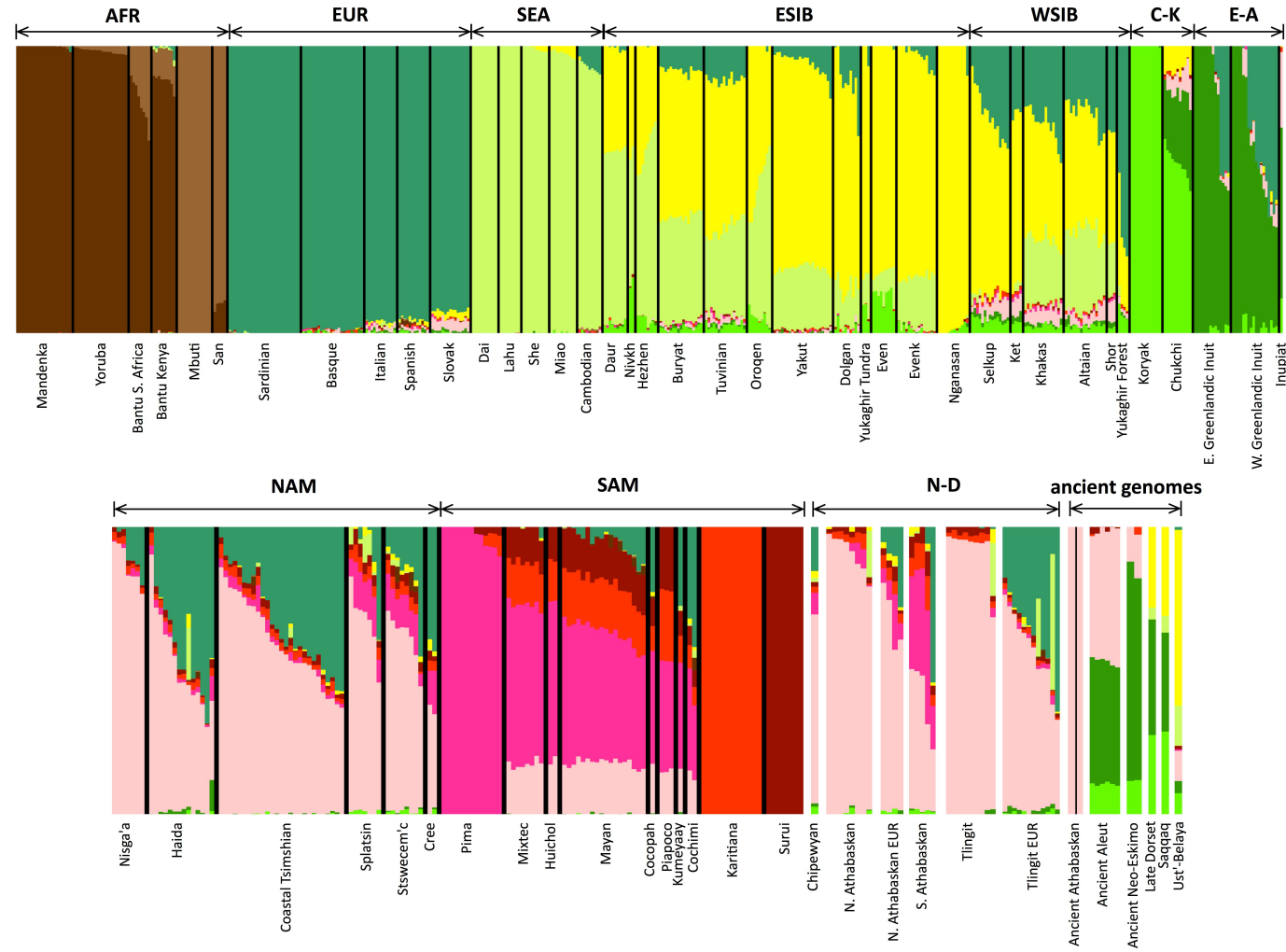
Extended Data figures

Extended Data Figure 1. ADMIXTURE plots for the HumanOrigins (a) and Illumina (b) SNP array datasets that models contributions from 14 or 11 hypothetical ancestral populations, respectively. One hundred iterations were calculated for each value of K from 5 to 20 (where K is the number of ancestral populations), and the optimal K values were 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; ESIB, East Siberians; WSIB, West Siberians; C-K, Chukotko-Kamchatkan speakers; E-A, Eskimo-Aleut speakers; NAM, northern First Americans; SAM, southern First Americans; ATH, Northern Athabaskan speakers; N-D, Na-Dene speakers. Chipewyan (a) or Northern Athabaskan and Tlingit (b) individuals with European admixture are plotted in separate bars, as well as ancient individuals: Clovis, Northern Athabaskans, Aleuts, Chukotkan Neo-Eskimos, Saqqaq, Late Dorset and Ust'-Belaya Paleo-Eskimos. Outliers, including individuals admixed with Europeans and East Asians, were not removed from Na-Dene-speaking populations in the Illumina dataset (b) to preserve their maximal diversity.

a

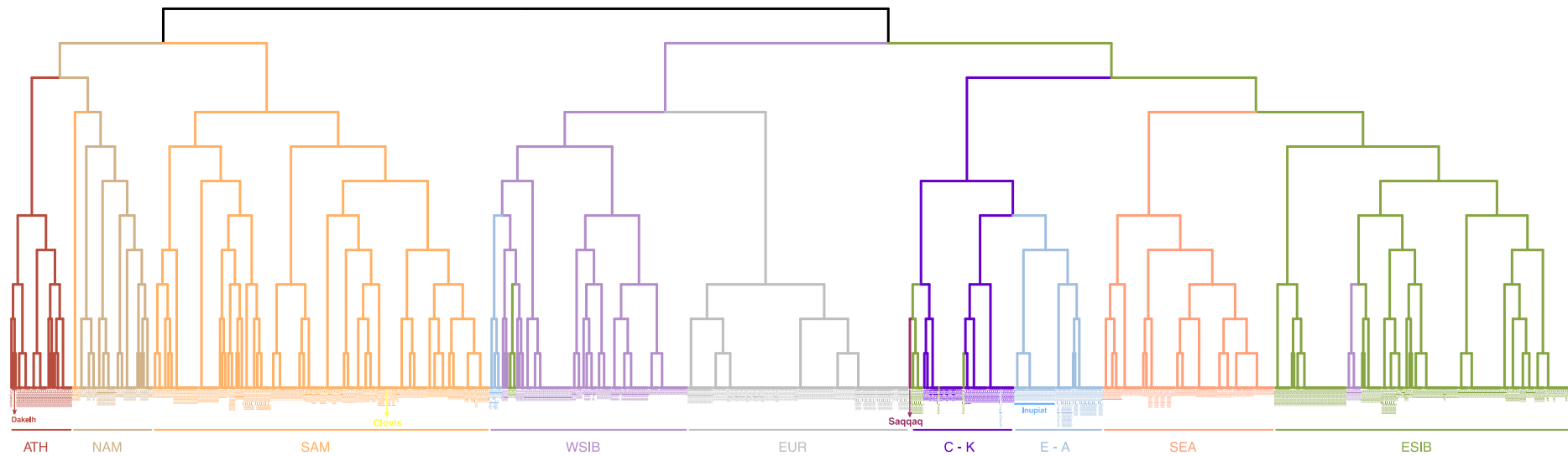


b

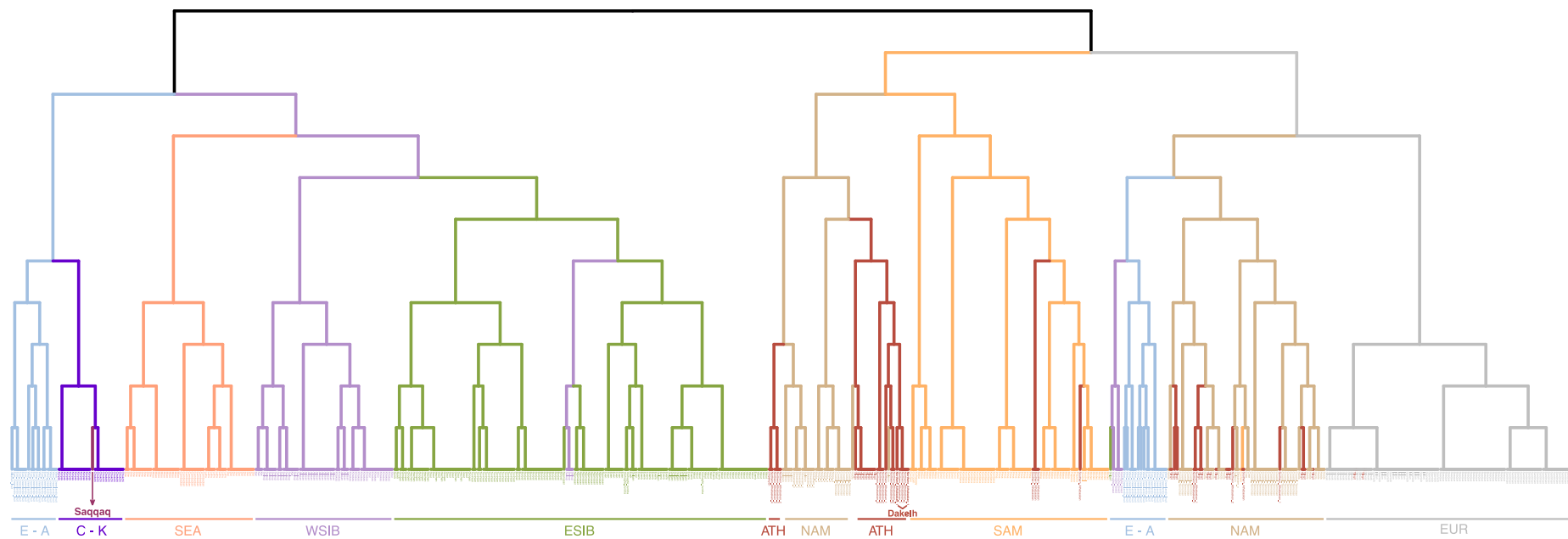


Extended Data Figure 2. Clustering trees of individuals computed by *fineSTRUCTURE* on coancestry matrices based on counts of shared haplotypes. Reduced versions of the HumanOrigins (a) an Illumina (b) SNP array datasets were used (Supplementary Table 4), including only the following meta-populations most relevant for our study: Eskimo-Aleut speakers (E-A), Chukotko-Kamchatkan speakers (C-K), Na-Dene speakers (ATH), northern First Americans (NAM), southern First Americans (SAM), West Siberians (WSIB), East Siberians (ESIB), Southeast Asians (SEA), Europeans (EUR). Meta-population affiliation is color-coded for individuals. Inupiat individuals genotyped in this study are marked with a blue line. The two Dakelh (Northern Athabaskan) individuals with sequenced genomes and the ancient individuals, Clovis within the southern First American clade and Saqqaq within the Chukotko-Kamchatkan clade, are also indicated. Most members of each clade belong to the meta-populations indicated, with few exceptions: **a**, Altaians fall into the ESIB clade, some Chilote fall into the NAM, and Aleuts fall into the WSIB clades (two latter cases might be explained by extensive European admixture in Chilote and in Aleuts (Extended Data Fig. 1a) which drives this clustering); **b**, some Selkups fall into the ESIB clade, all four Southern Athabaskan speakers cluster with South Americans, reflecting their substantial South American admixture (Extended Data Fig. 1b), one Haida individual clusters with Na-Dene speakers, and five Northern Athabaskan speakers cluster with other northern First Americans.

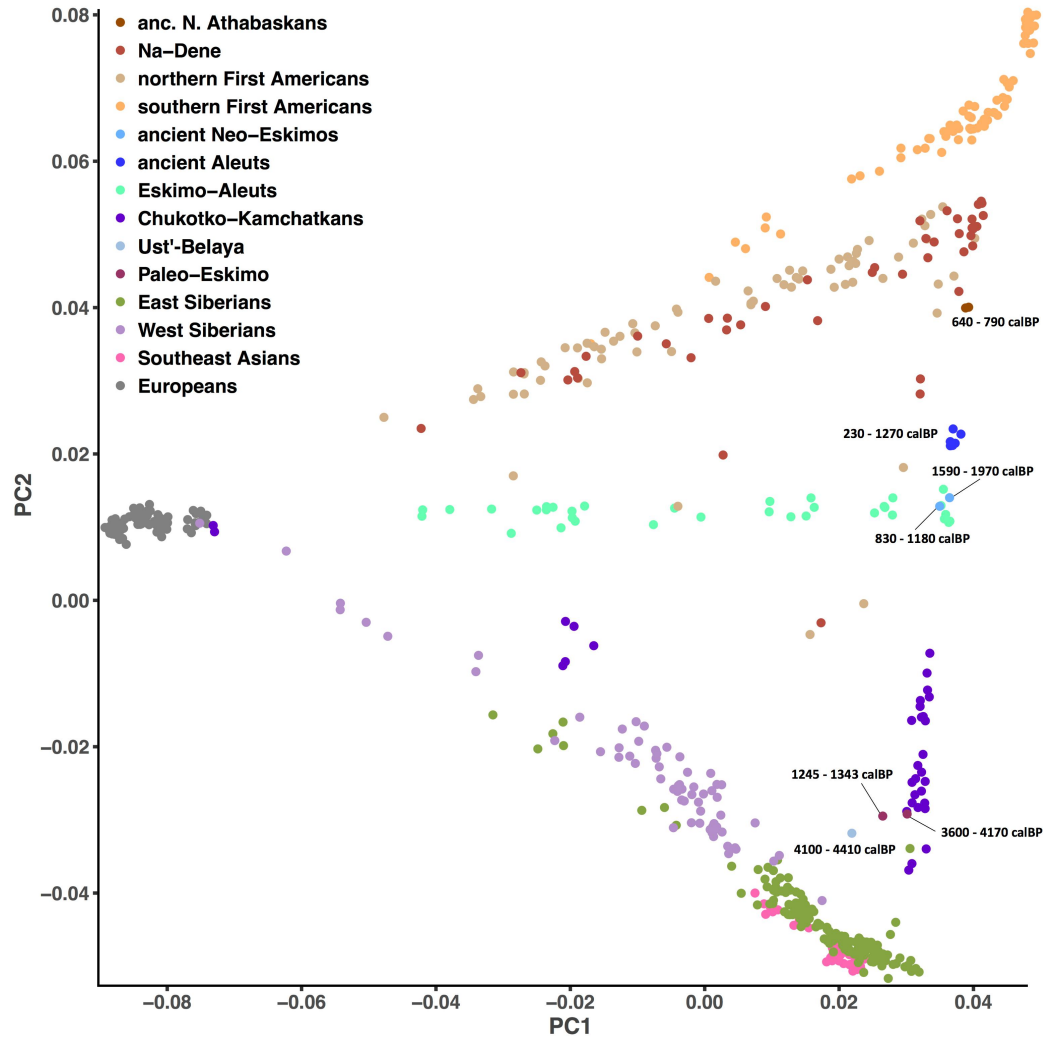
a



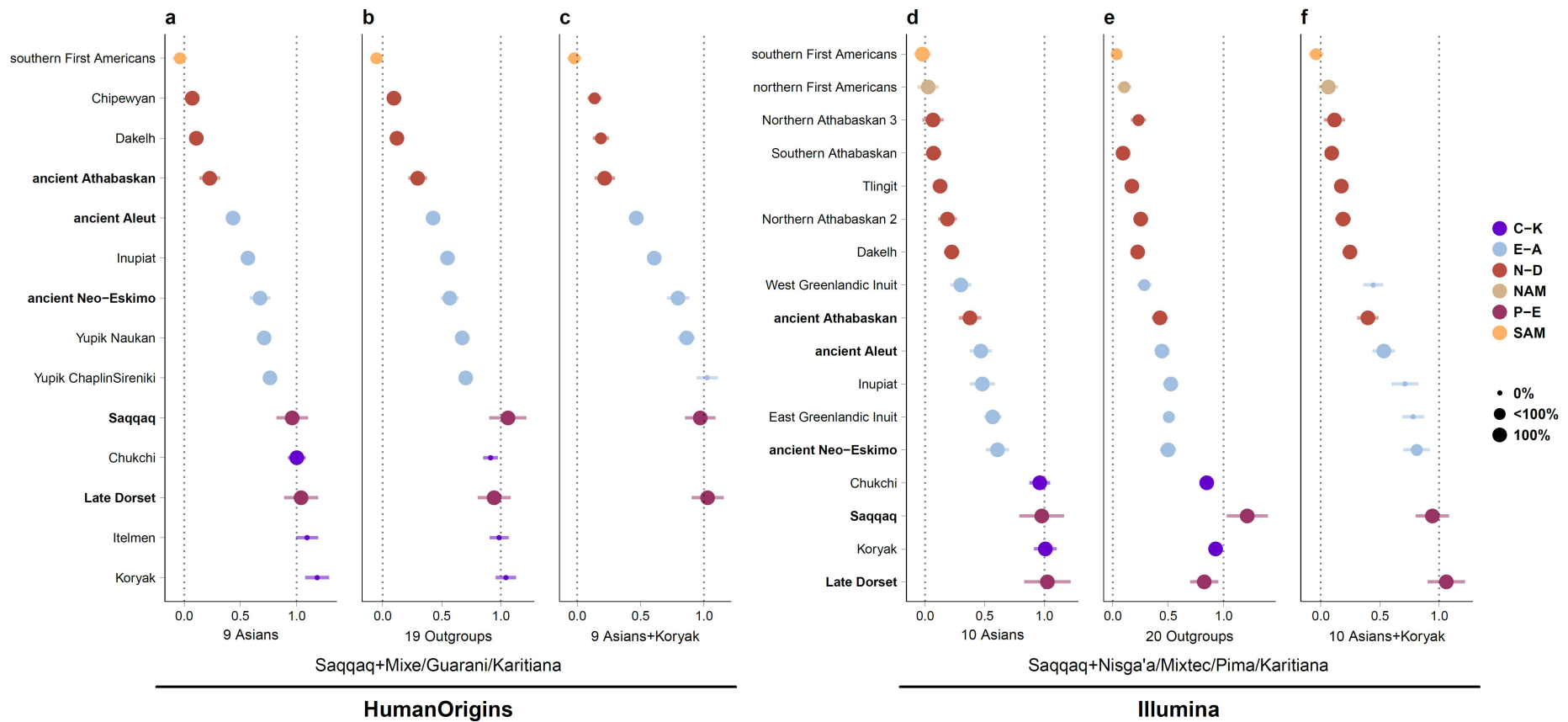
b



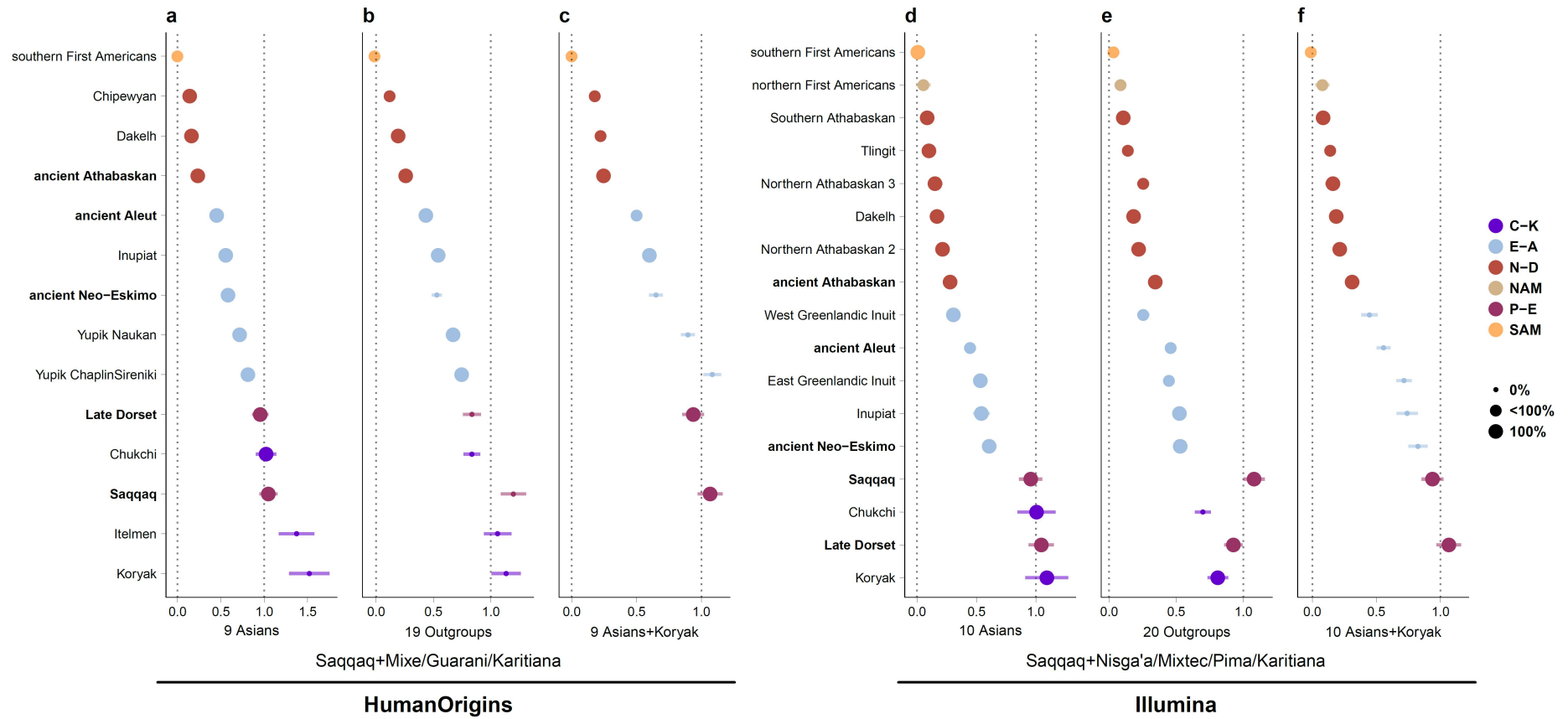
Extended Data Figure 3. Principal component analysis (PCA) on the Illumina dataset. A plot of two principal components (PC1 vs. PC2) calculated by *PLINK2* is shown (linkage disequilibrium pruning was not applied). The following meta-populations most relevant for our study are plotted: ancient Neo-Eskimos and present-day Eskimo-Aleut speakers, Chukotko-Kamchatkan speakers, Paleo-Eskimos (the Saqqaq, Late Dorset, and Ust'-Belaya individuals), ancient Northern Athabaskans, present-day Na-Dene speakers, northern and southern First Americans, West and East Siberians, Southeast Asians, and Europeans. Calibrated radiocarbon dates in calBP are shown for ancient samples. For individuals, 95% confidence intervals are shown, and for populations, minimal and maximal dates among all confidence intervals of that population are shown.



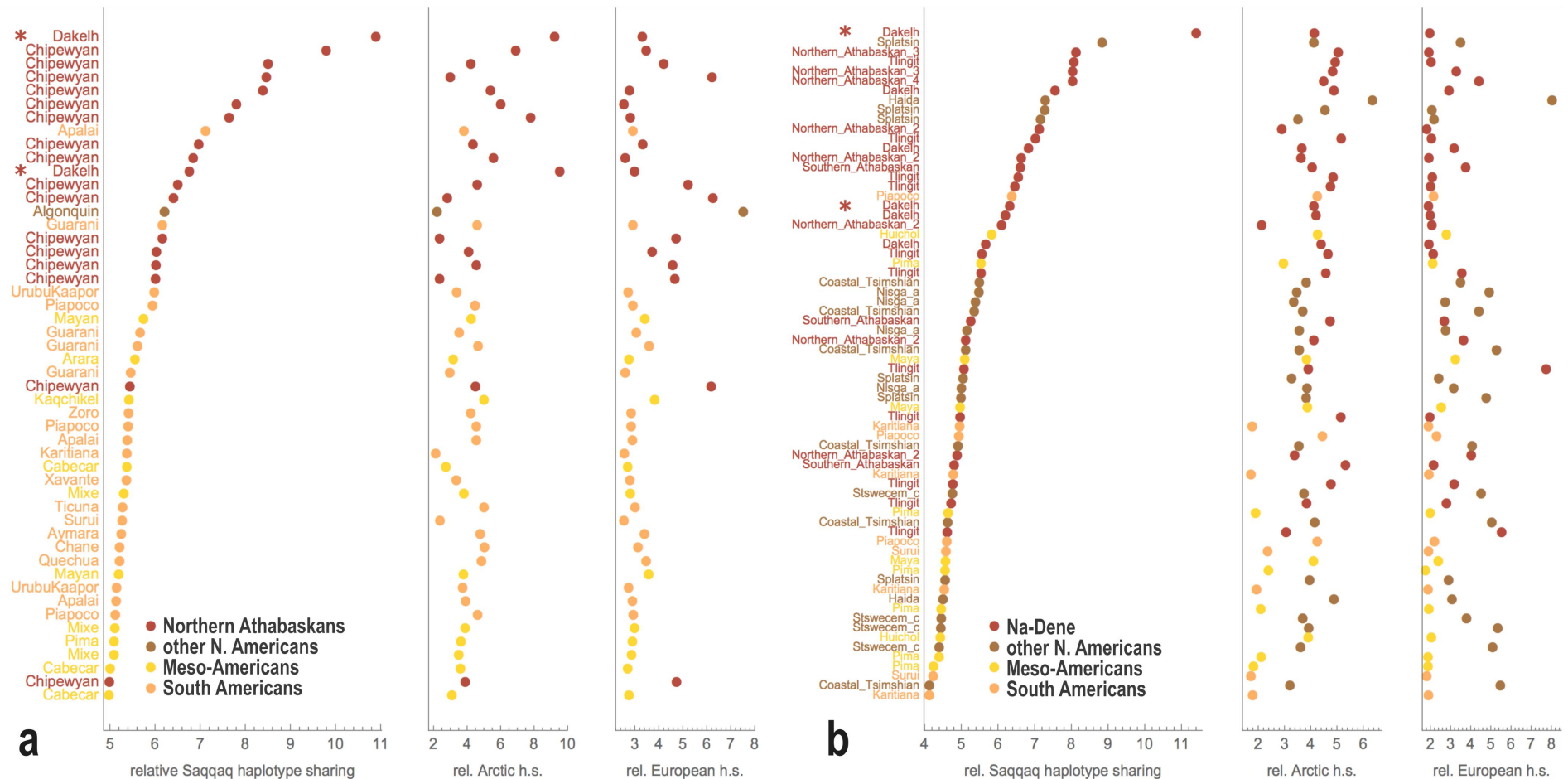
Extended Data Figure 4. Modelling American, Chukotkan, and Kamchatkan populations as mixtures of First American and Paleo-Eskimo sources on the HumanOrigins (a-c) and Illumina (d-f) datasets without transitions. First, population triplets were tested with *qpWave* for consistency with two or three streams of ancestry derived from outgroups. Alternative outgroup sets are indicated below the plots and described in detail in Methods. Second, *qpAdm* was used to infer admixture proportions in present-day or ancient target populations (in bold). Saqqaq was considered as a Paleo-Eskimo source for all populations apart from Saqqaq itself, for which Late Dorset was used as a source, and alternative First American sources were as follows: Mixe, Guarani, or Karitiana for the HumanOrigins dataset; Nisga'a, Mixtec, Pima, or Karitiana for the Illumina dataset. Admixture proportions and their standard errors were averaged across triplets including different First American sources, or across many alternative target populations in the case of southern and northern First Americans. Meta-populations are color-coded and abbreviated as follows: C-K, Chukotko-Kamchatkan speakers; E-A, Eskimo-Aleut speakers and ancient Neo-Eskimos; N-D, Na-Dene speakers; NAM, northern First Americans; SAM, southern First Americans. Proportion of triplets consistent with two migration streams is coded by the circle size: small (0%), medium (>0% and <100%), and large (100%).



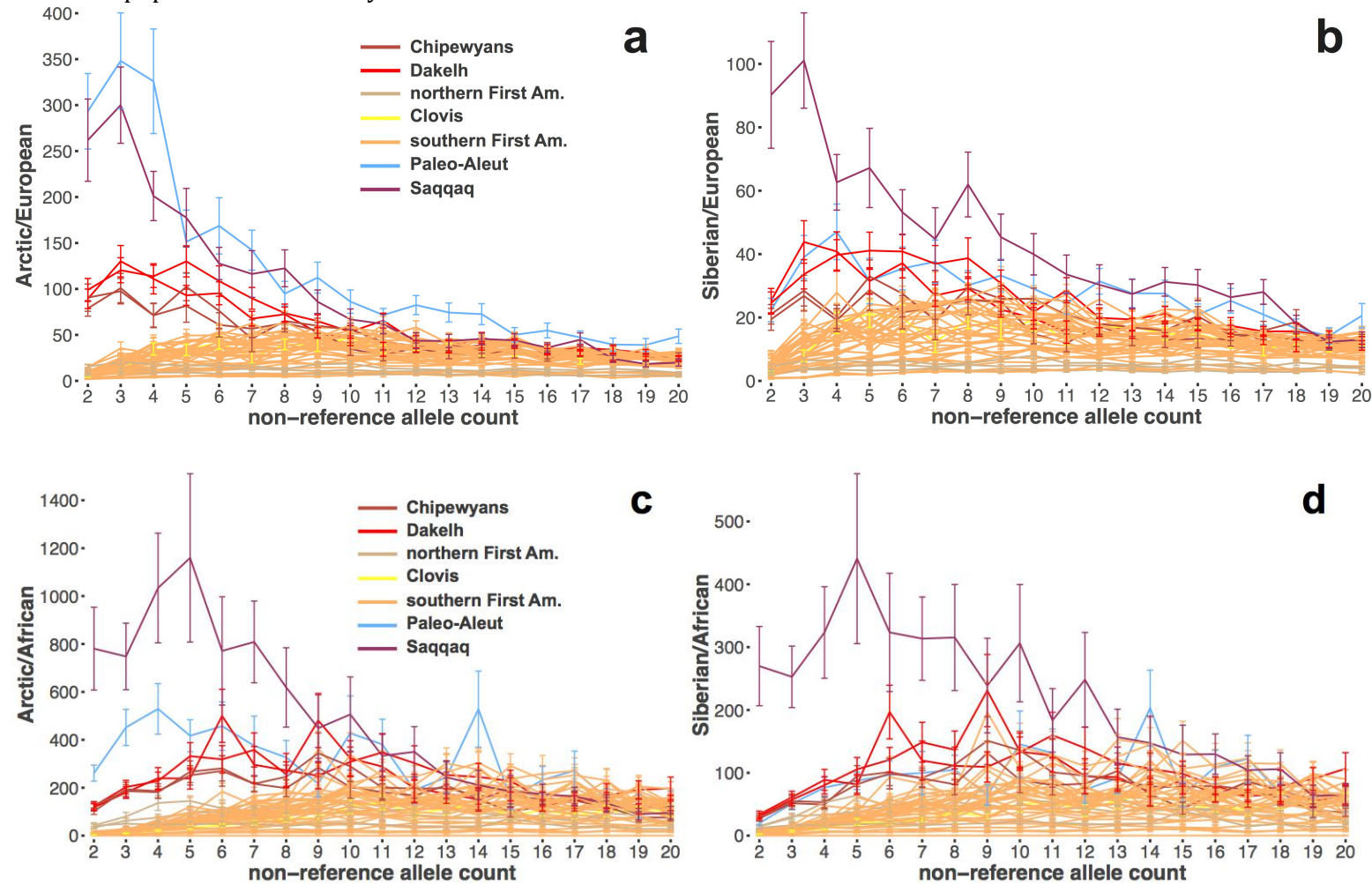
Extended Data Figure 5. The same results as in Extended Data Fig. 4, but on the full HumanOrigins (a-c) and Illumina (d-f) datasets including transitions.

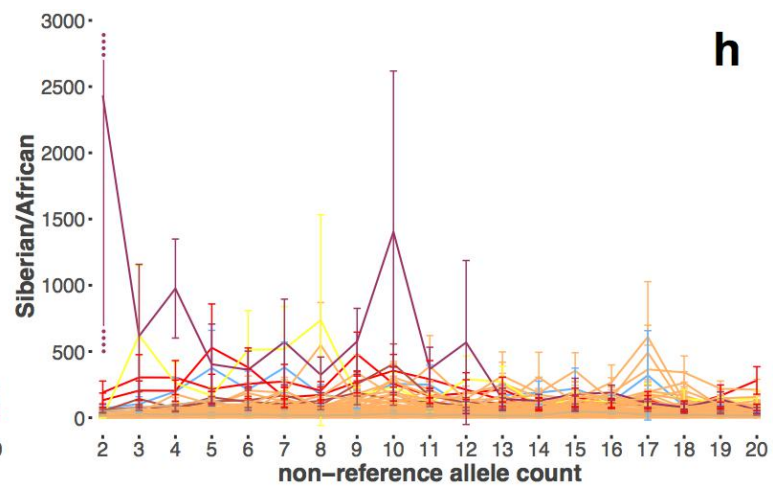
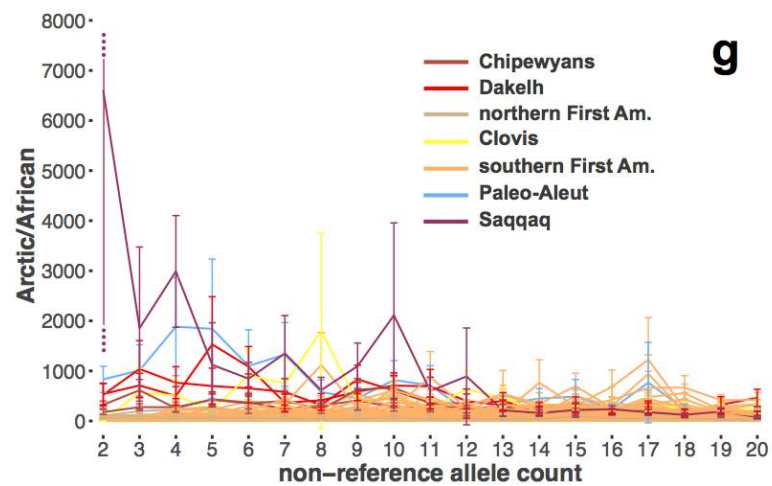
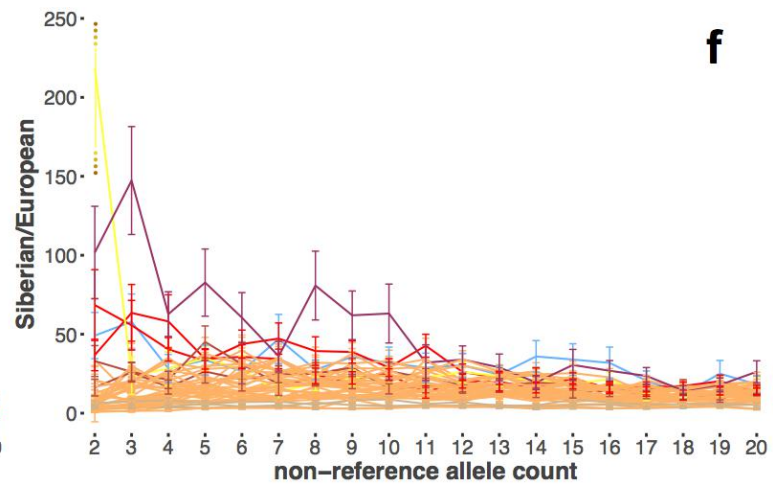
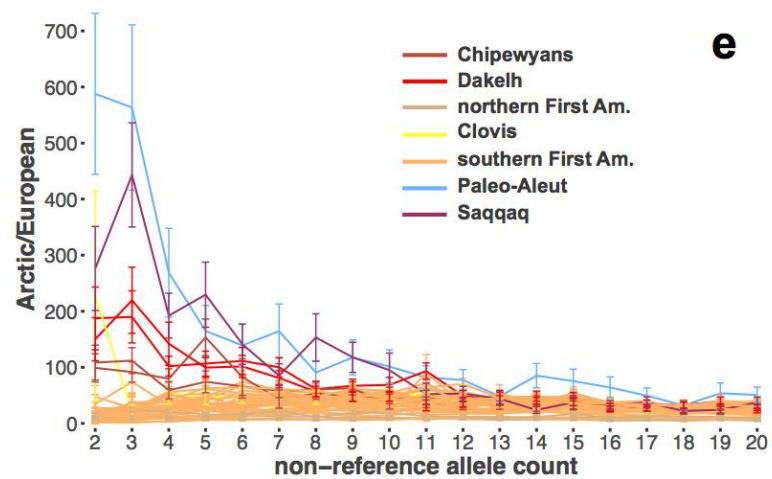


Extended Data Figure 6. Relative Saqqaq, Arctic, and European haplotype sharing statistics (HSS) for American individuals, normalized using the African meta-population. Both Eskimo-Aleut- and Chukotko-Kamchatkan-speaking groups contributed to the Arctic HSS. The same statistics and statistics with other normalizers are shown in the form of two-dimensional plots in Supplemental Information section 6. Two Dakelh (Northern Athabaskan) individuals with sequencing data, included also into the HumanOrigins (a) and Illumina (b) datasets, are marked with asterisks. The plots based on both datasets demonstrate that Na-Dene speakers have the highest relative Saqqaq HSS. One Haida and three Spltasin individuals also demonstrate outlying Saqqaq HSSs (b), however these individuals stand in contrast to a majority of non-Na-Dene northern First Americans, and Paleo-Eskimo ancestry in these individuals may be explained by recent admixture with Na-Dene speakers living in close proximity (Verdu et al. 2014). The Haida outlier demonstrates a maximal Arctic HSS among all First Americans, and its Arctic ancestry has contributed to its elevated Saqqaq HSS. Saqqaq, Arctic and European statistics are largely uncorrelated in First Americans: Pearson's correlation coefficients for Saqqaq vs. Arctic relative HSSs are 0.56 among all First Americans and 0.64 among northern First Americans on the Illumina dataset, and 0.66 and 0.72, respectively, on the HumanOrigins dataset.

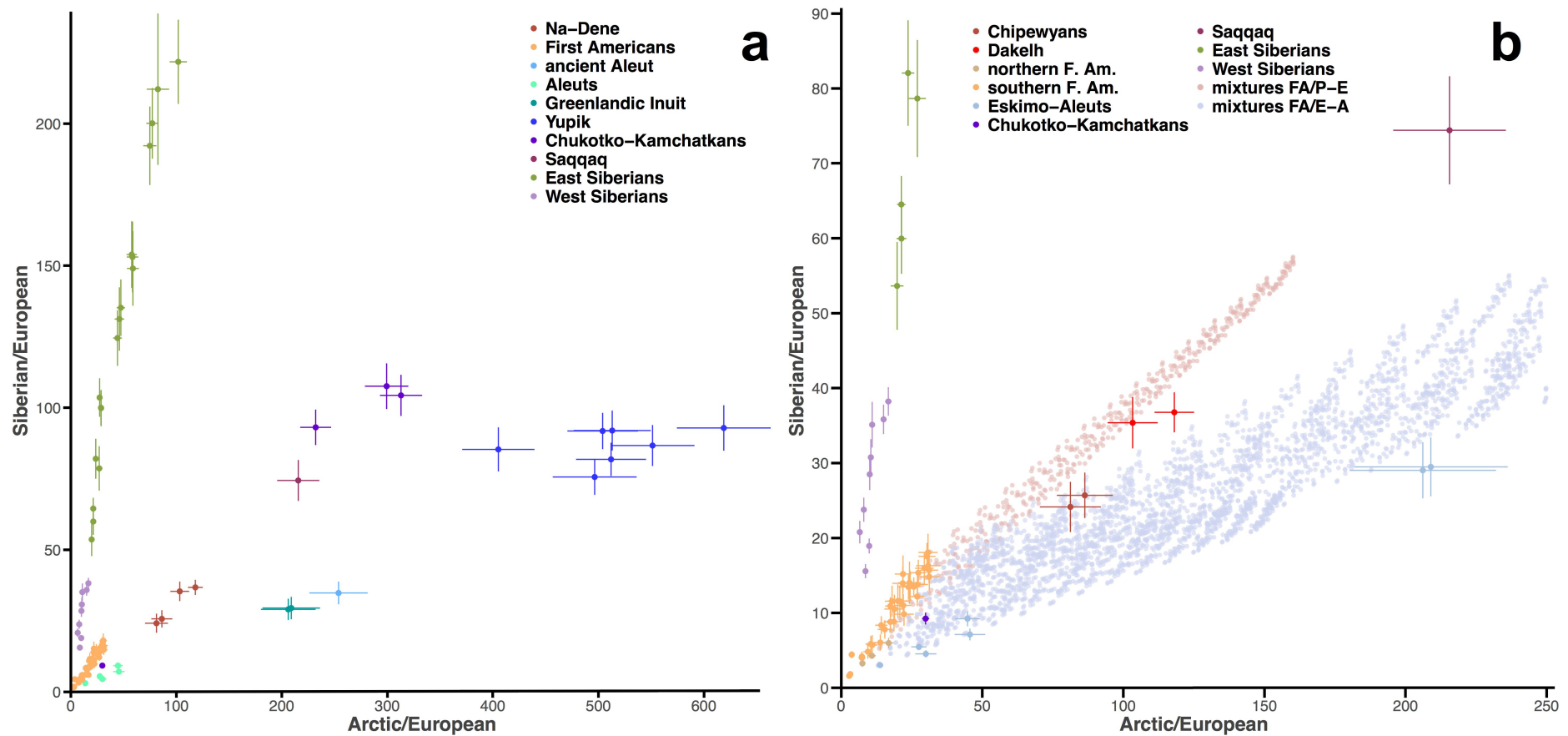


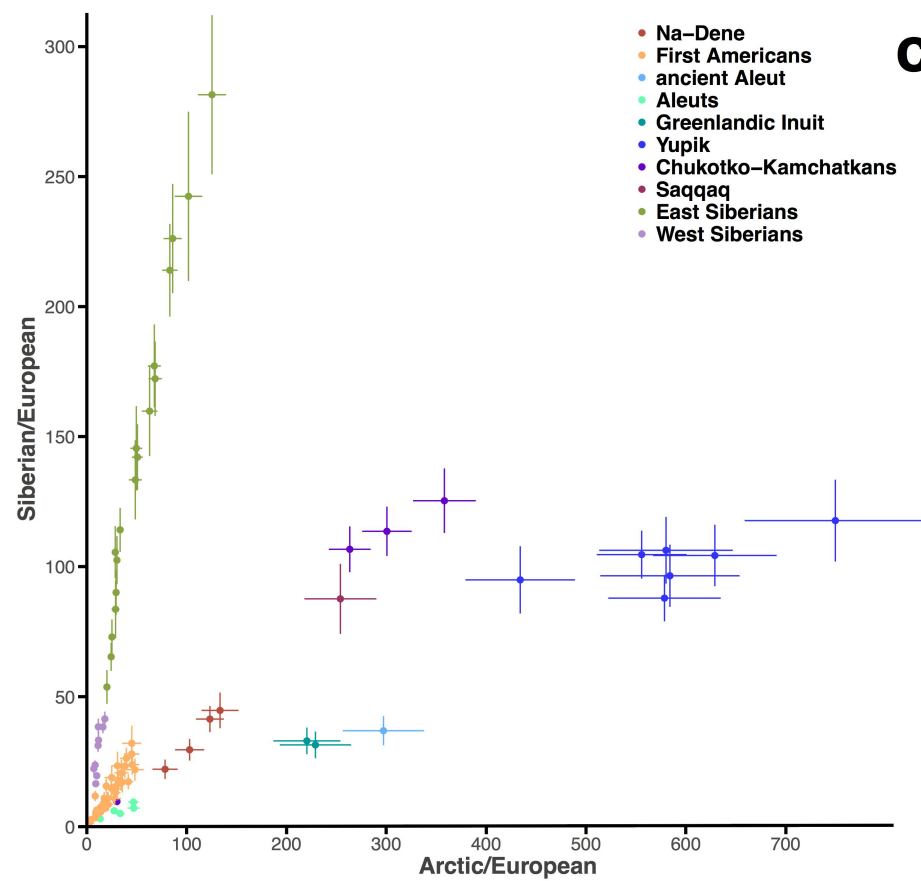
Extended Data Figure 7. Normalized rare allele sharing statistics calculated for each present-day or ancient American individual and the Arctic (panels **a**, **c**, **e**, **g**) or Siberian (**b**, **d**, **f**, **h**) meta-populations. Statistics for the dataset without transitions are shown in panels **e** – **h**. All statistics were calculated separately for alleles of various frequency: occurring 2, 3, 4, ... and up to 20 times in the set of 2,414 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 individual and the Arctic or Siberian meta-populations by similar counts of alleles shared with distant outgroups – Europeans (**a**, **b**, **e**, **f**) or Africans (**c**, **d**, **g**, **h**). Standard deviations were calculated using a jackknife approach with chromosomes used as resampling blocks, and single standard error intervals are plotted. Few extremely wide error intervals were observed on the smaller transition-free dataset: for convenience, they are shown with dotted lines in panels **f** – **h**. Saqqaq, ancient Aleut, and Northern Athabaskan speakers (Chipewyans and Dakelh) stand out among American populations in all analyses shown here.



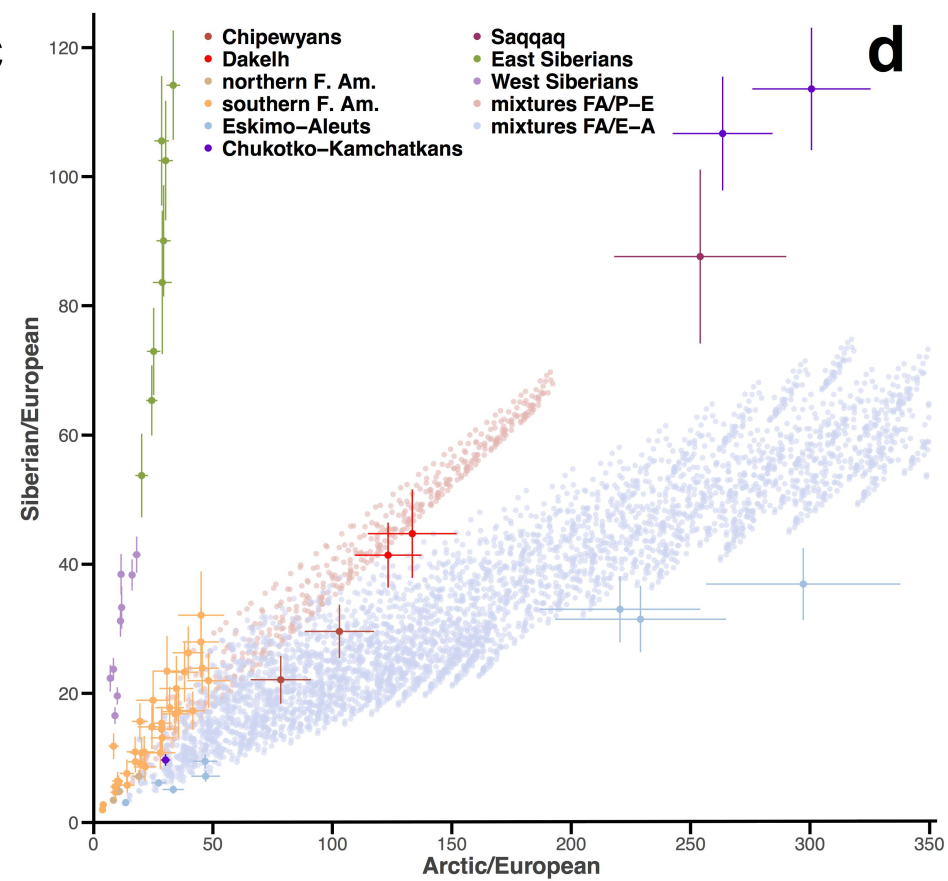


Extended Data Figure 8. Two-dimensional plots of Arctic and Siberian rare allele sharing statistics normalized using the European meta-population. Statistics for the dataset without transitions are shown in panels **c**, **d**. Non-reference alleles occurring from 2 to 5 times in the dataset contributed to the statistics. Standard deviations were calculated using a jackknife approach with chromosomes used as resampling blocks, and single standard error intervals are plotted. Populations and meta-populations are color-coded according to the legend. Plots on the left (**a**, **c**) show statistics for all individuals. Plots on the right (**b**, **d**) are enlarged versions showing simulated mixtures of any present-day southern First American individual and the Saqqaq individual (from 5% to 70% Saqqaq ancestry, with 5% increments), and similar mixtures with present-day Eskimo-Aleut-speaking individuals: >5% Yupik or Greenlandic Inuit ancestry.



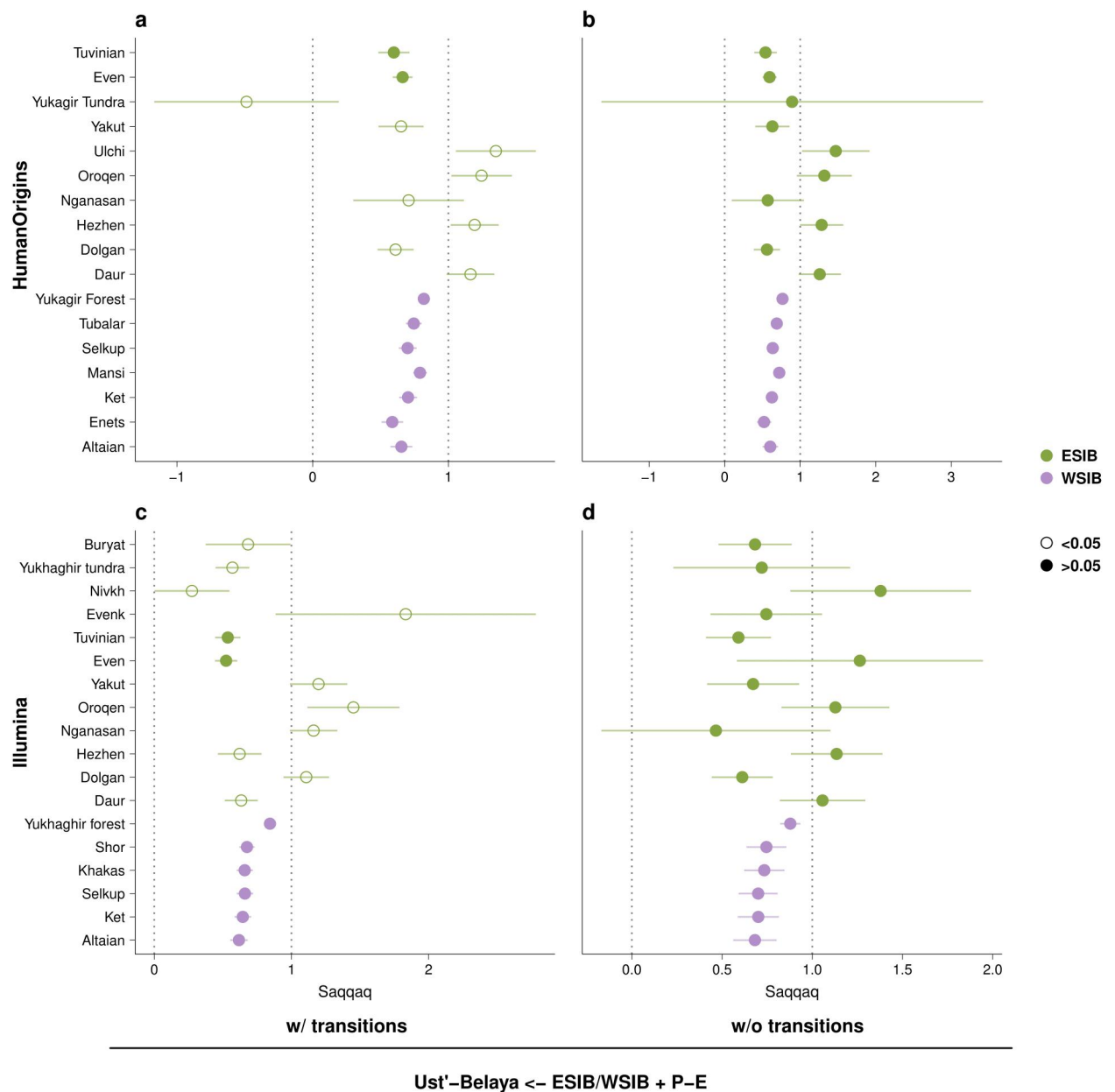


C



d

Extended Data Figure 9. Modelling the Ust'-Belaya individual as a mixture of Paleo-Eskimo (Saqqaq) and various Siberian sources on the HumanOrigins (a, b) and Illumina (c, d) datasets with (a, c) or without (b, d) transition polymorphisms. First, population triplets were tested with *qpWave* for consistency with two or three streams of ancestry derived from outgroups. Second, *qpAdm* was applied to infer admixture proportions and their standard errors. Models with East Siberian (ESIB) source populations were often inconsistent with two ancestry streams (empty circles), demonstrated negative admixture proportions or very wide error intervals. On the contrary, models with West Siberians (WSIB) demonstrated much more consistent results on all four dataset versions, suggesting that West Siberians are closely related to the true admixture partner. Since Siberians were used as source populations, the outgroup sets were adjusted accordingly. The following set of 19 outgroups from five broad geographical regions was used for the HumanOrigins dataset: Mbuti, Taa, Yoruba (Africans); Australian, New Guinea Highlander, Onge, Papuan (Melanesians); Czech, English, French, North Italian (Europeans); Guarani, Karitiana, Mixe, Surui (southern First Americans); Dai, Miao, She, Thai (Southeast Asians). The following set of 16 outgroups was used for the Illumina dataset lacking Melanesian populations: Bantu (Kenya), Mandenka, Mbuti, Yoruba (Africans); Basque, Sardinian, Slovak, Spanish (Europeans); Karitiana, Nisga'a, Pima, Surui (First Americans); Dai, Lahu, Miao, She (Southeast Asians).



Extended Data Figure 10. Admixture graphs connecting populations (a) or meta-populations (b) and matching the topology of the *rarecoal* model (Fig. 5a, Supplemental Information section 9). Meta-populations used to construct graph B were exactly the same as those of the *rarecoal* model, and populations in graph A were drawn from the same set of meta-populations. The population-based graph A lacks admixture edges reflecting post-colonial European admixture in Eskimo-Aleuts and Chukotko-Kamchatkans. In contrast, heterogeneous meta-populations used for graph B and for the *rarecoal* model include individuals with recent European admixture, and edges were added to reflect that. Putative ancient North Eurasian, present-day European, Siberian, and proto-Paleo-Eskimo lineages and gene flows are color-coded according to the legend. The following populations and meta-populations were used: Mbuti (the African meta-population, AFR); Ami (Southeast Asian, SEA); French (European, EUR); Even (Siberian, SIB); Koryak (Chukotko-Kamchatkan speakers, CHK); Yupik Naukan alternatively named Eskimo Naukan (Eskimo-Aleut speakers, ESK); Mixe (First American, FAM); Chipewyan (Athabaskan speakers, ATH). Z-scores corresponding to the most extreme f_4 -statistics deviating from expected values are shown beside the models.

