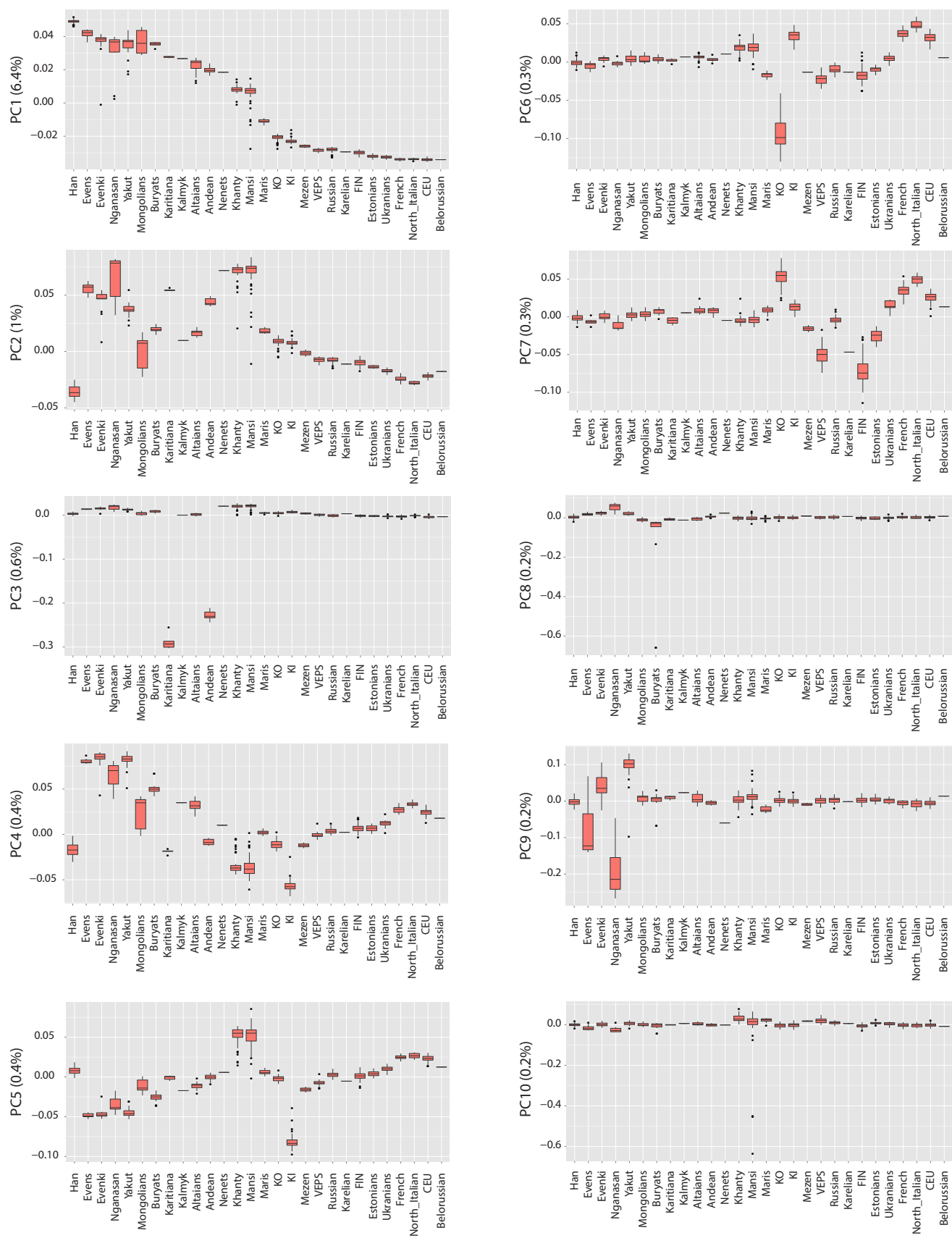


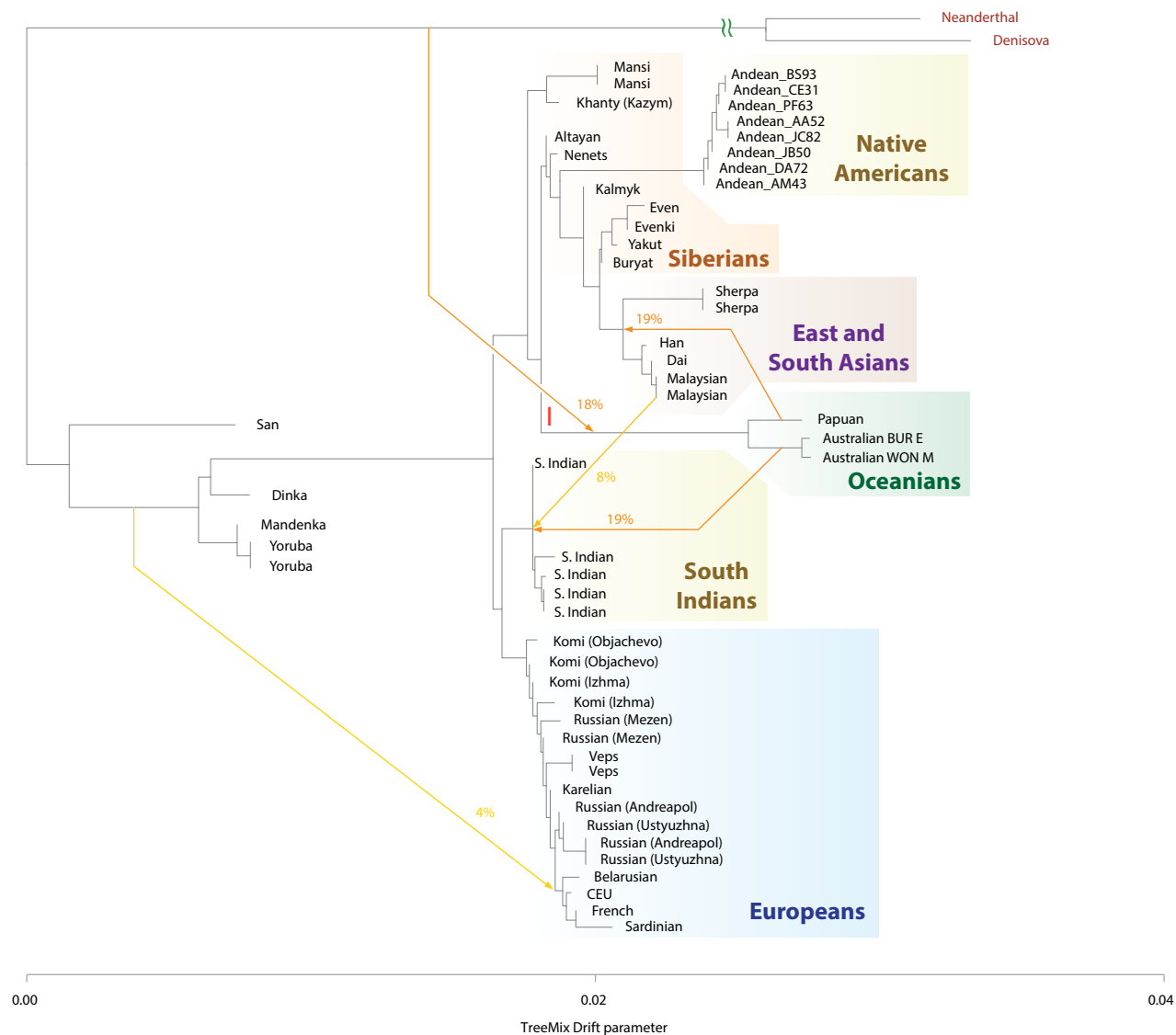
Linguistic classification of major Northeastern European, Siberian and East Asian populations. Tentative grouping and classification of Finno-Ugric and Altaic languages according to their linguistic similarities. Japonic and Koreanic language groups may be related to other Altaic languages, although this relationship is not yet definitively established (therefore indicated by a question mark). Language groups are represented by blue nodes, populations are represented by orange nodes.

FIGURE S1



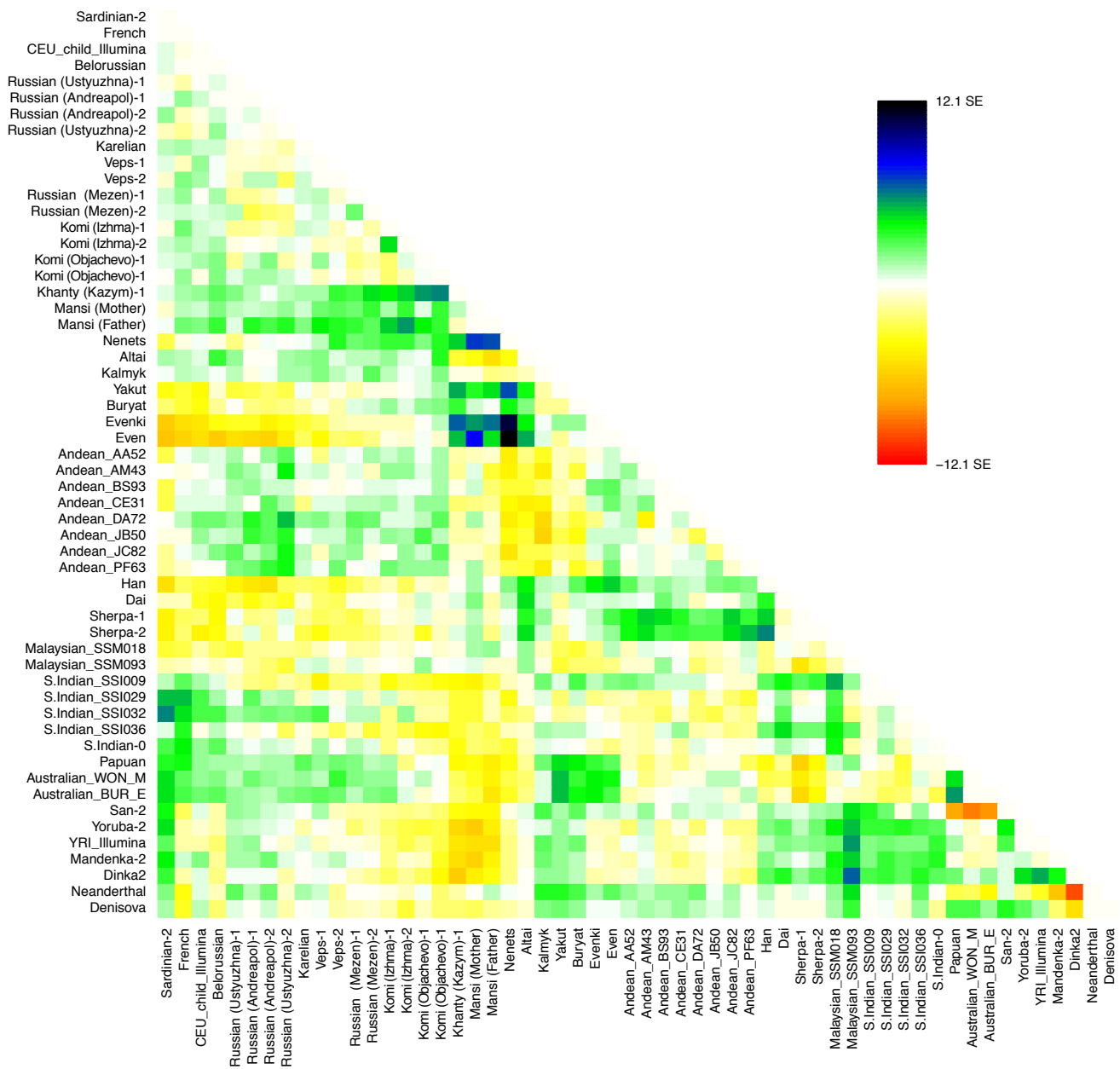
PC values across populations. Values of individual principle components PC1-PC10 shown across the same set of populations as in Figure 1b.

FIGURE S2



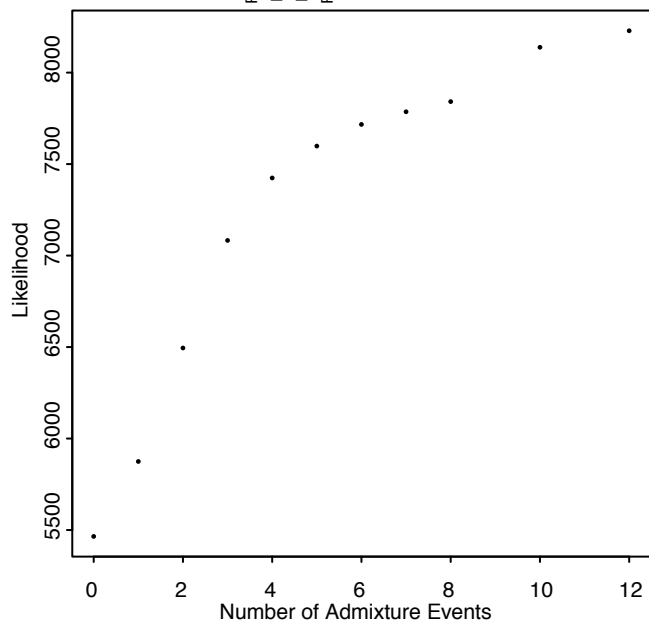
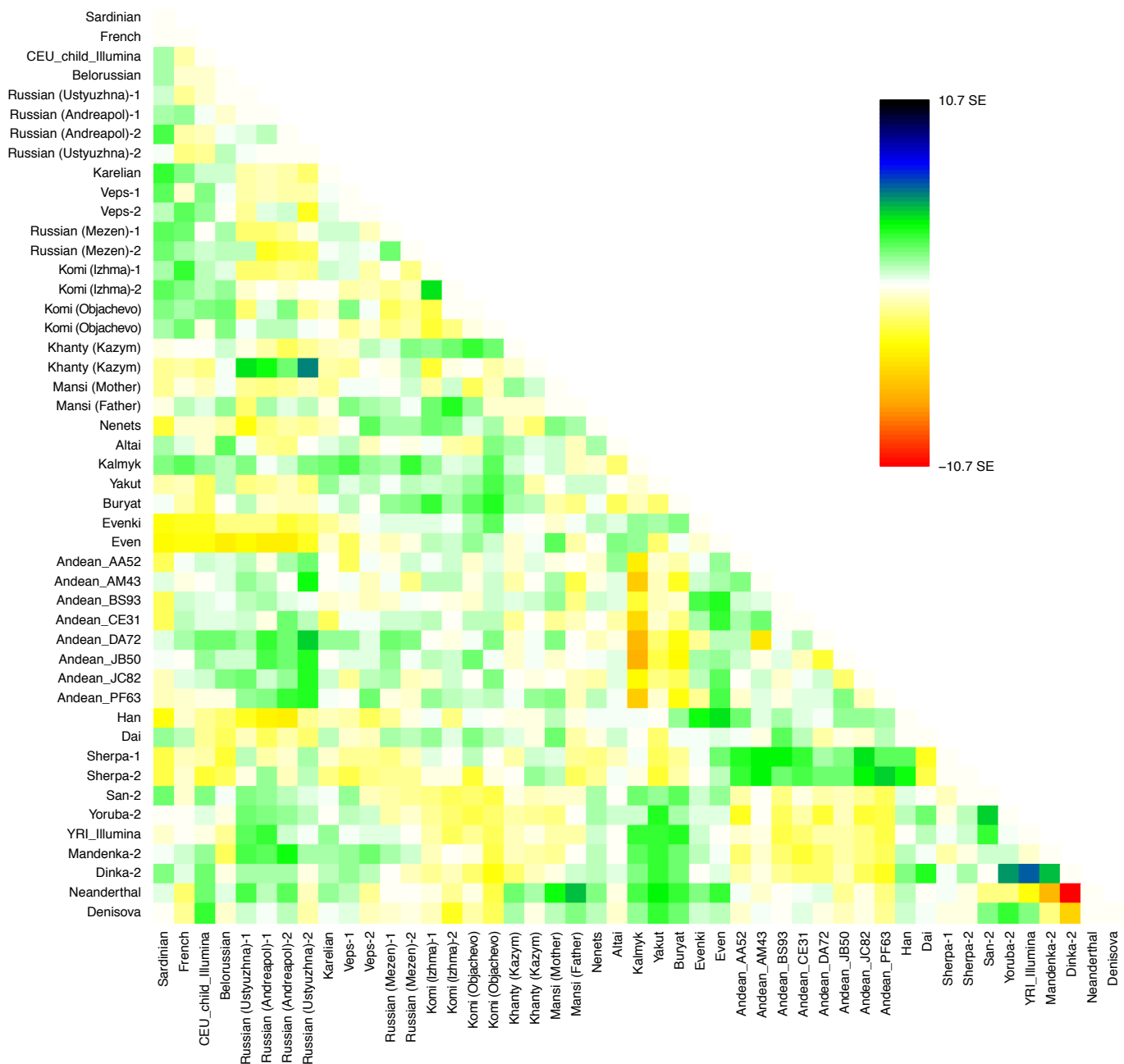
Genetic relationships of 56 people from across the world. Autosomal TreeMix admixture graph using whole-genome sequencing data from 54 individuals from 32 world populations. The tree is based on 25,589,077 polymorphic sites. The X-axis represents genetic drift, which is proportional to the effective population size N_e . Genomes of ancient hominids (Neanderthal and Denisova) are shown in red.

FIGURE S3



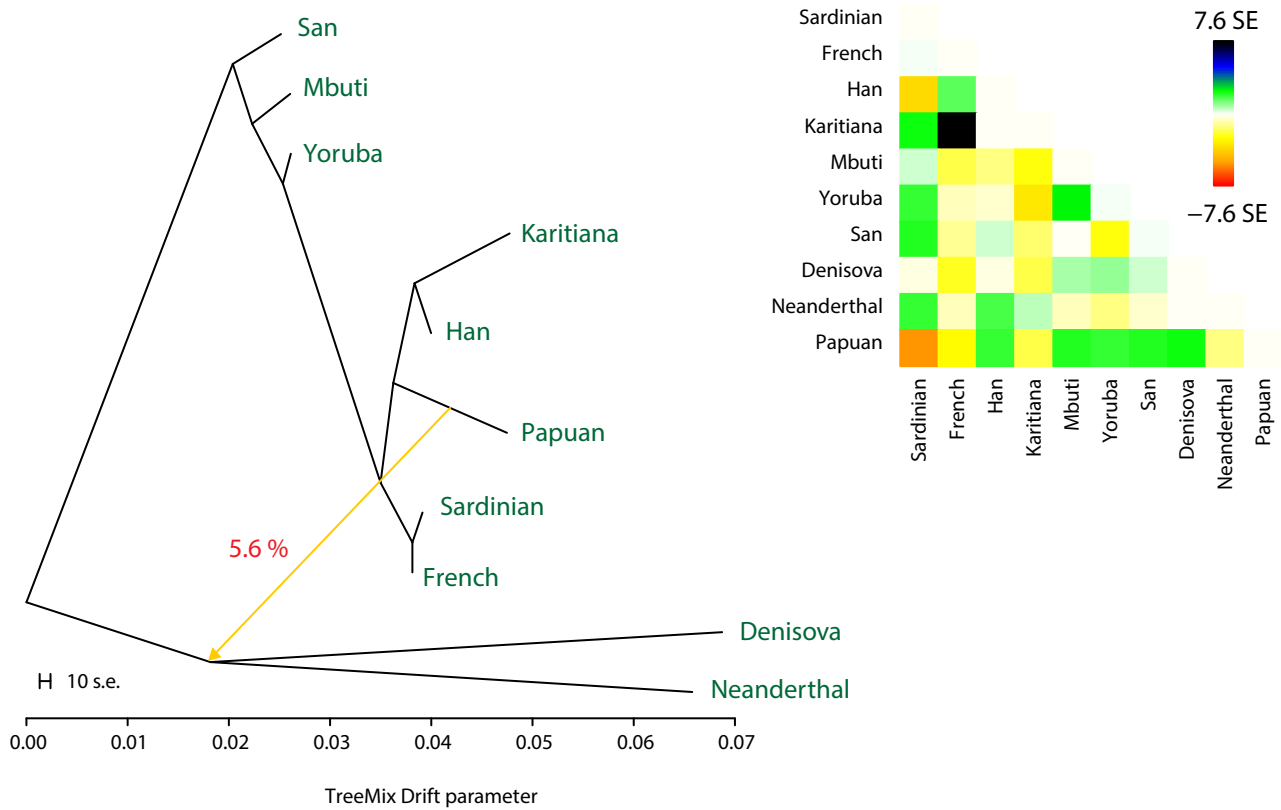
TreeMix residuals for the 56 whole-genome tree shown in Figure S3.

FIGURE S4



TTreeMix residuals from the European-Siberian autosomal tree shown in Figure 2. The likelihood plot (left) shows the improvement of data fit with a given number of graph admixture events.

FIGURE S5



Analysis of Neanderthal-related admixtures among Oceanians. The full autosomal TreeMix model was calculated using a limited set of eight individuals and two hominids. The most significant admixture was between a common ancestor of Denisova / Neanderthal and Papuan. The amount of admixture was estimated at 5.6% with standard error of 0.58%. Standard error was obtained using a jackknife estimate.

FIGURES S6

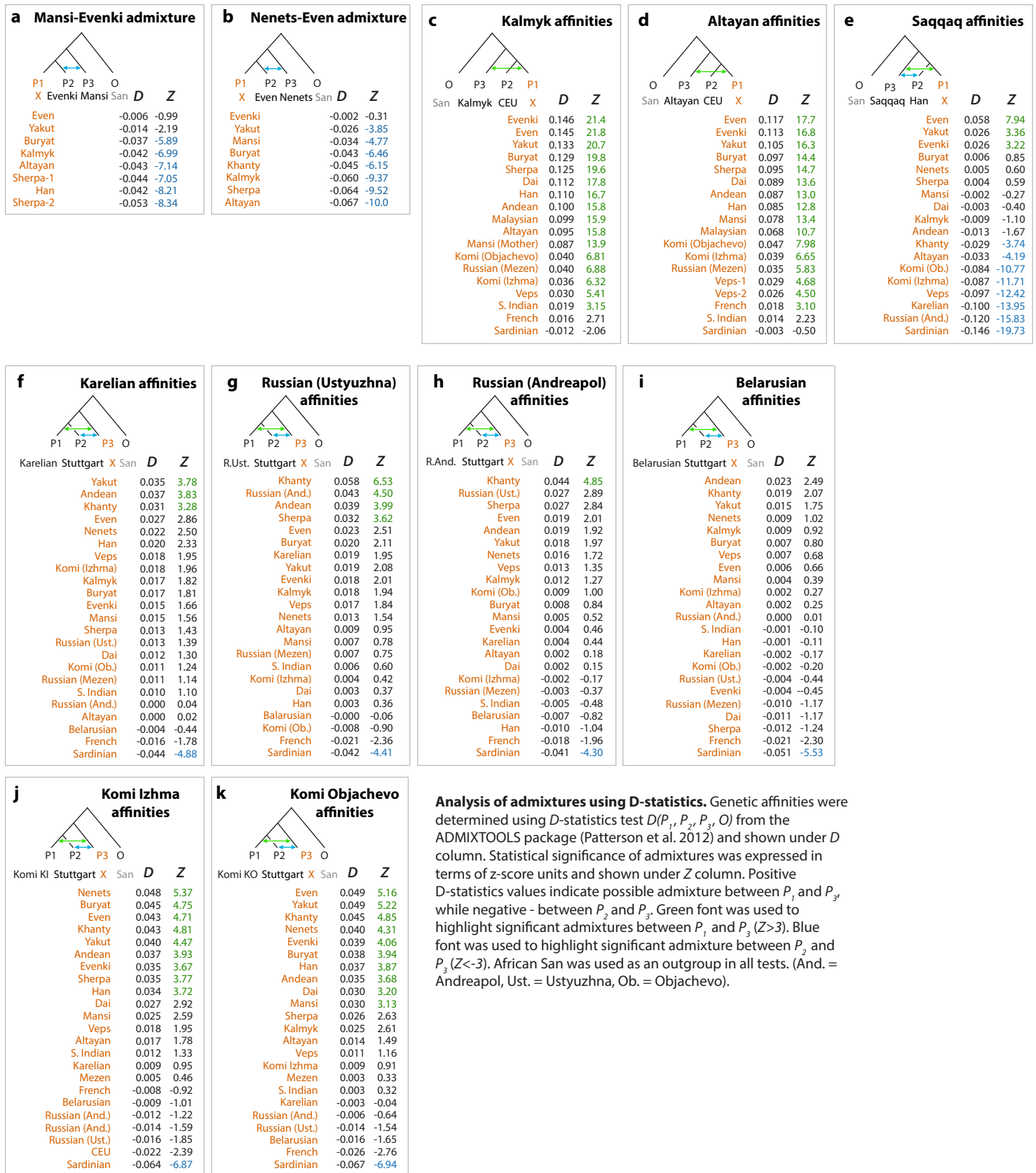
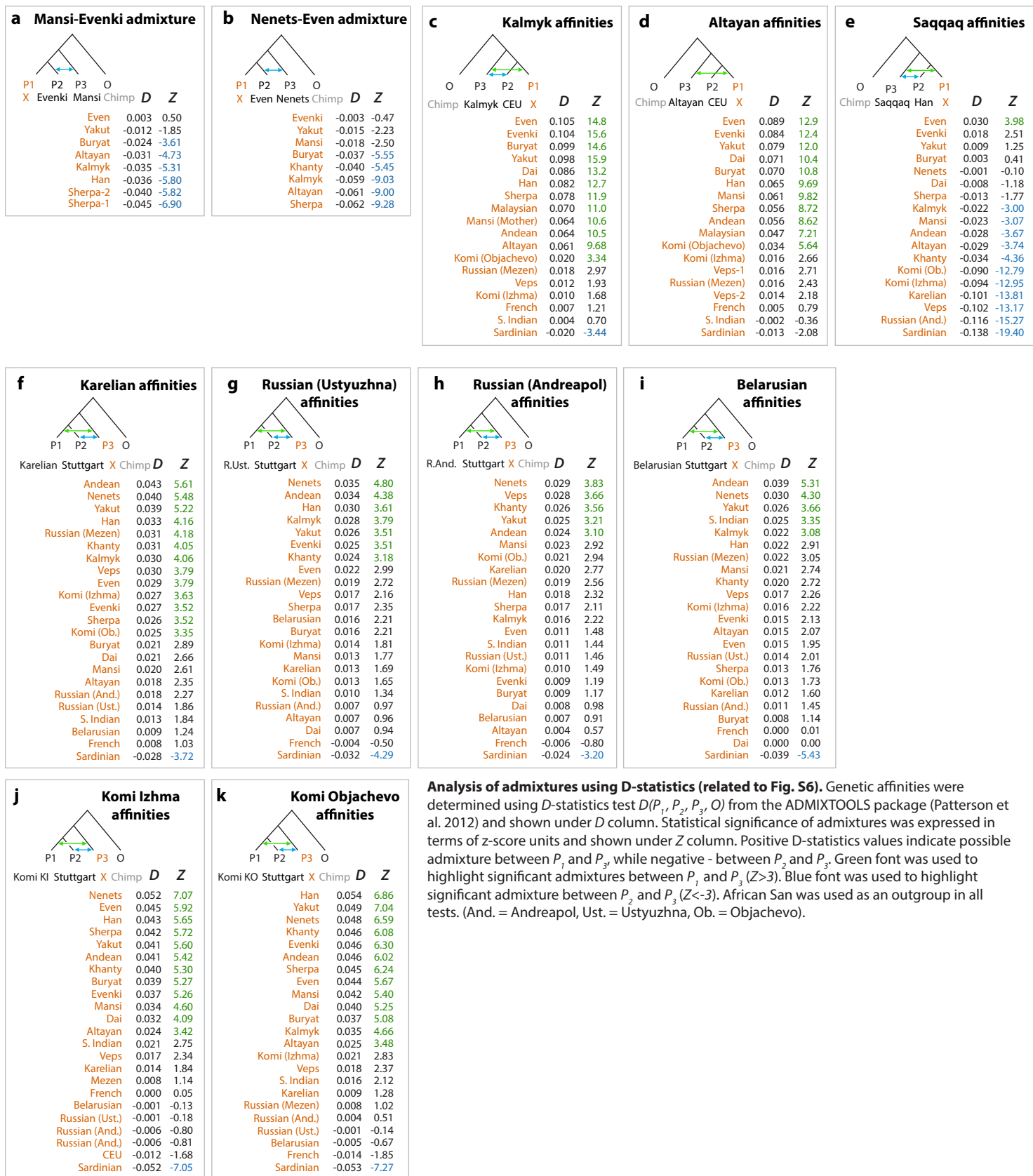
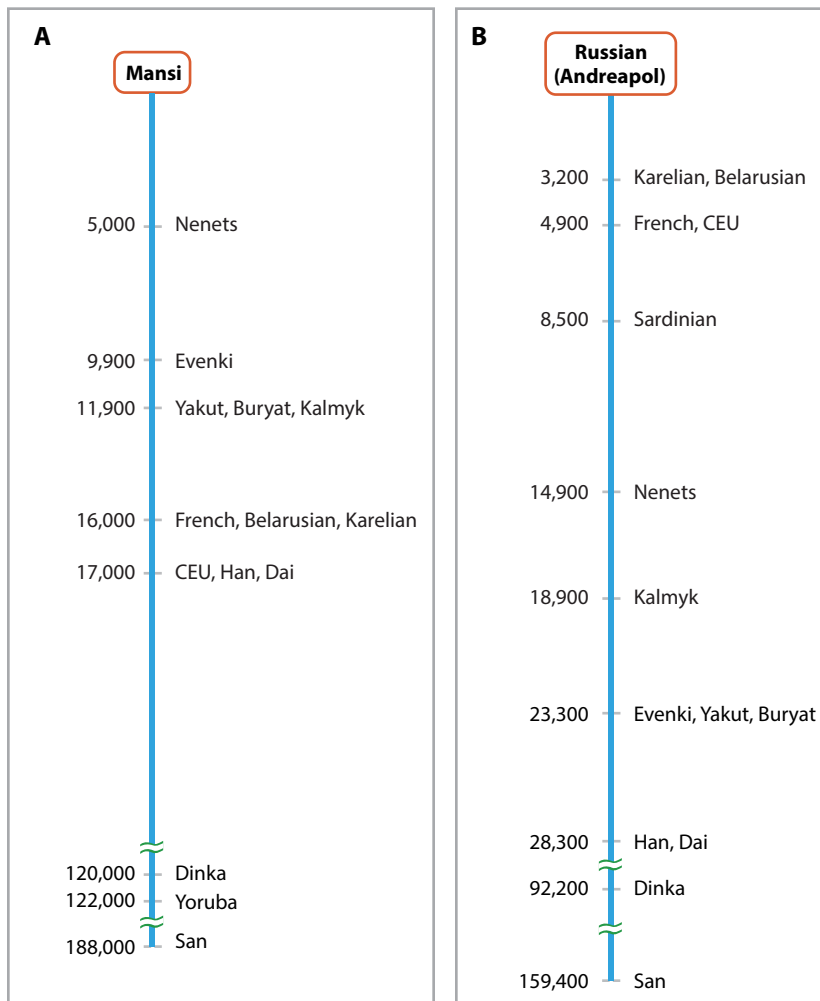


FIGURE S7



Analysis of admixtures using D-statistics (related to Fig. S6). Genetic affinities were determined using D-statistics test $D(P_1, P_2, P_3, O)$ from the ADMIXTOOLS package (Patterson et al. 2012) and shown under D column. Statistical significance of admixtures was expressed in terms of z-score units and shown under Z column. Positive D-statistics values indicate possible admixture between P_1 and P_2 , while negative - between P_2 and P_3 . Green font was used to highlight significant admixtures between P_1 and P_2 ($Z > 3$). Blue font was used to highlight significant admixture between P_2 and P_3 ($Z < -3$). African San was used as an outgroup in all tests. (And. = Andreapol, Ust. = Ustyuzhna, Ob. = Objachevo).

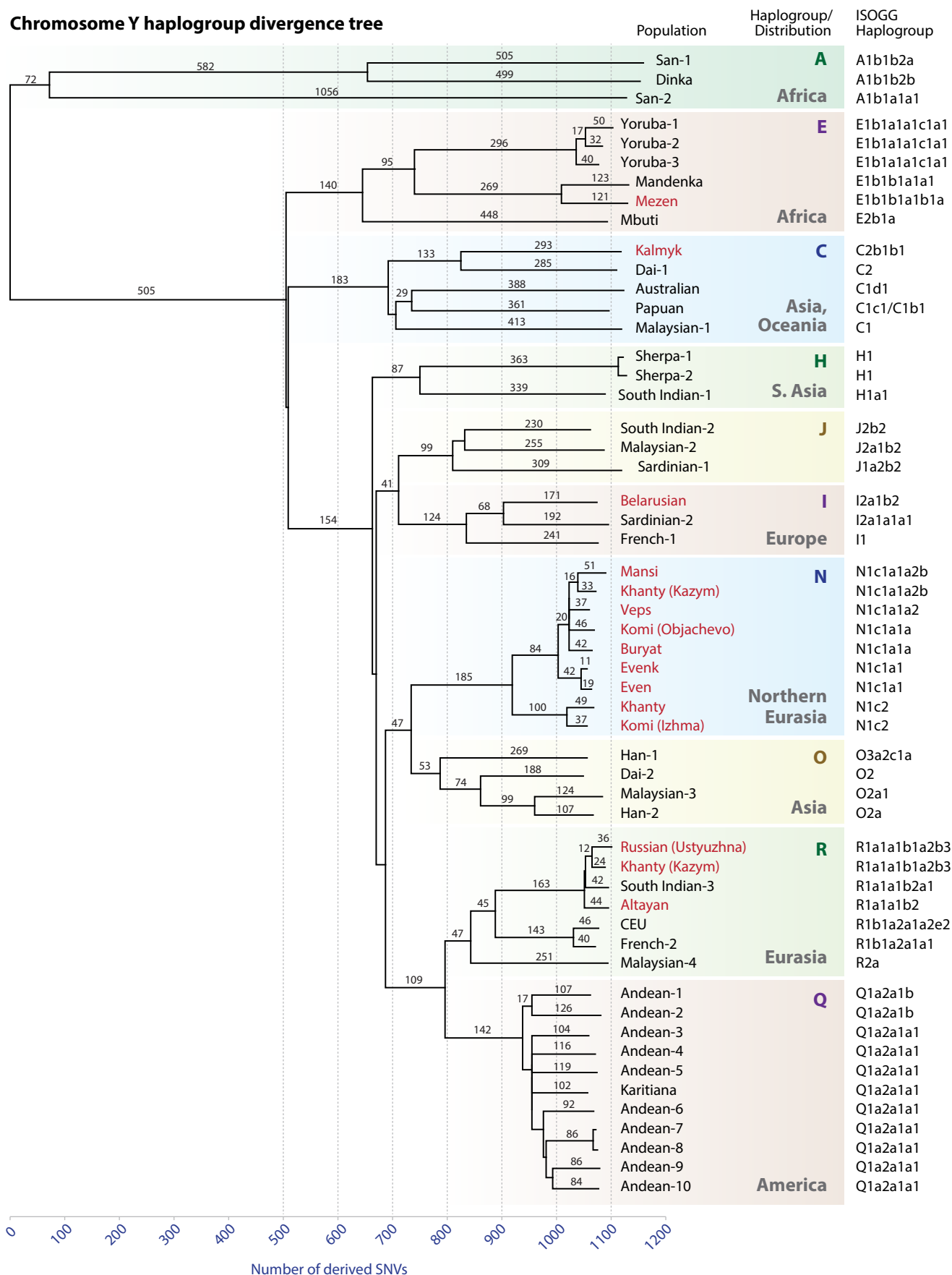
FIGURE S8



MSMC separation times estimates (8 haplotype analysis). MSMC analysis was performed using 2 individuals per population and median separation times are shown for Mansi and other populations (**A**) and for Russian Andreapol and other populations (**B**).

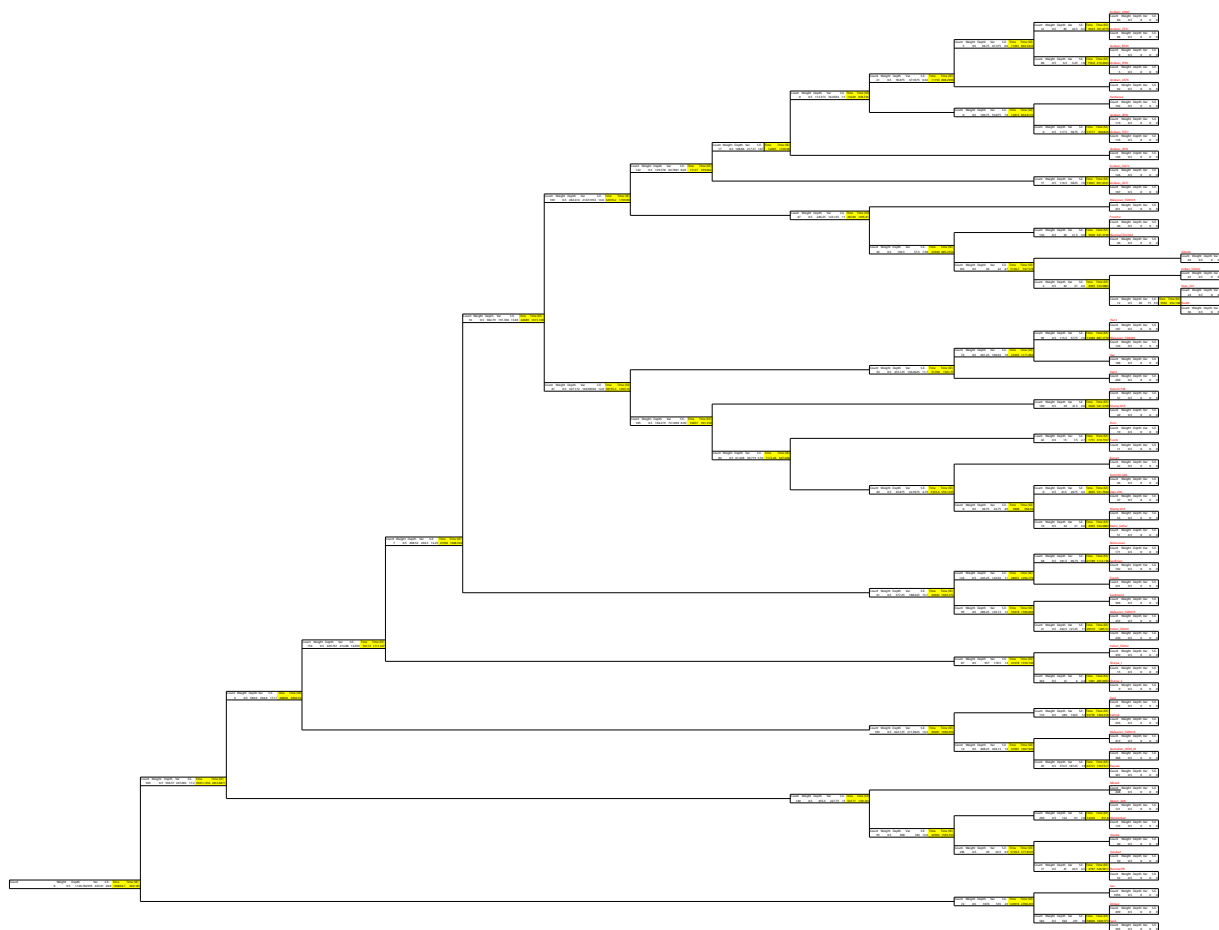
FIGURE S9

Chromosome Y haplogroup divergence tree



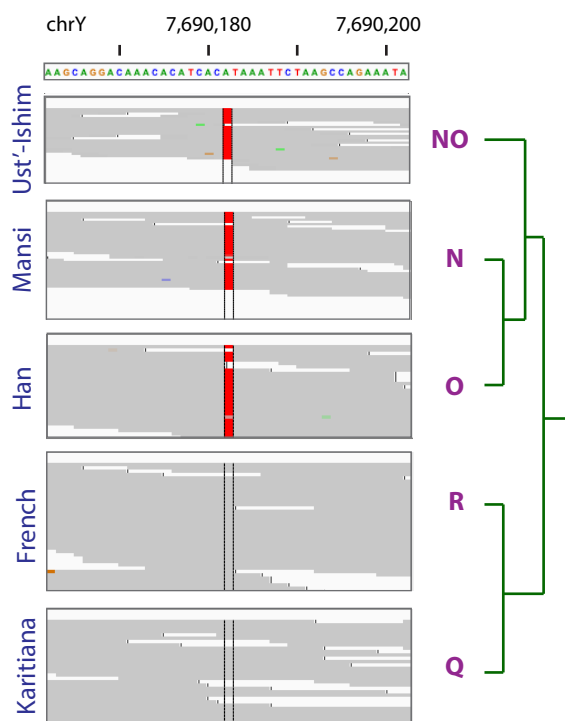
Phylogenetic view of Chromosome Y haplotypes and their divergence. The X-axis represents inferred number of mutations and branch lengths (shown by numbers) are proportional to the number of mutations inferred to have occurred along each branch. Tree topology was inferred with MEGA and SNV counts were determined using custom methods.

FIGURE S10



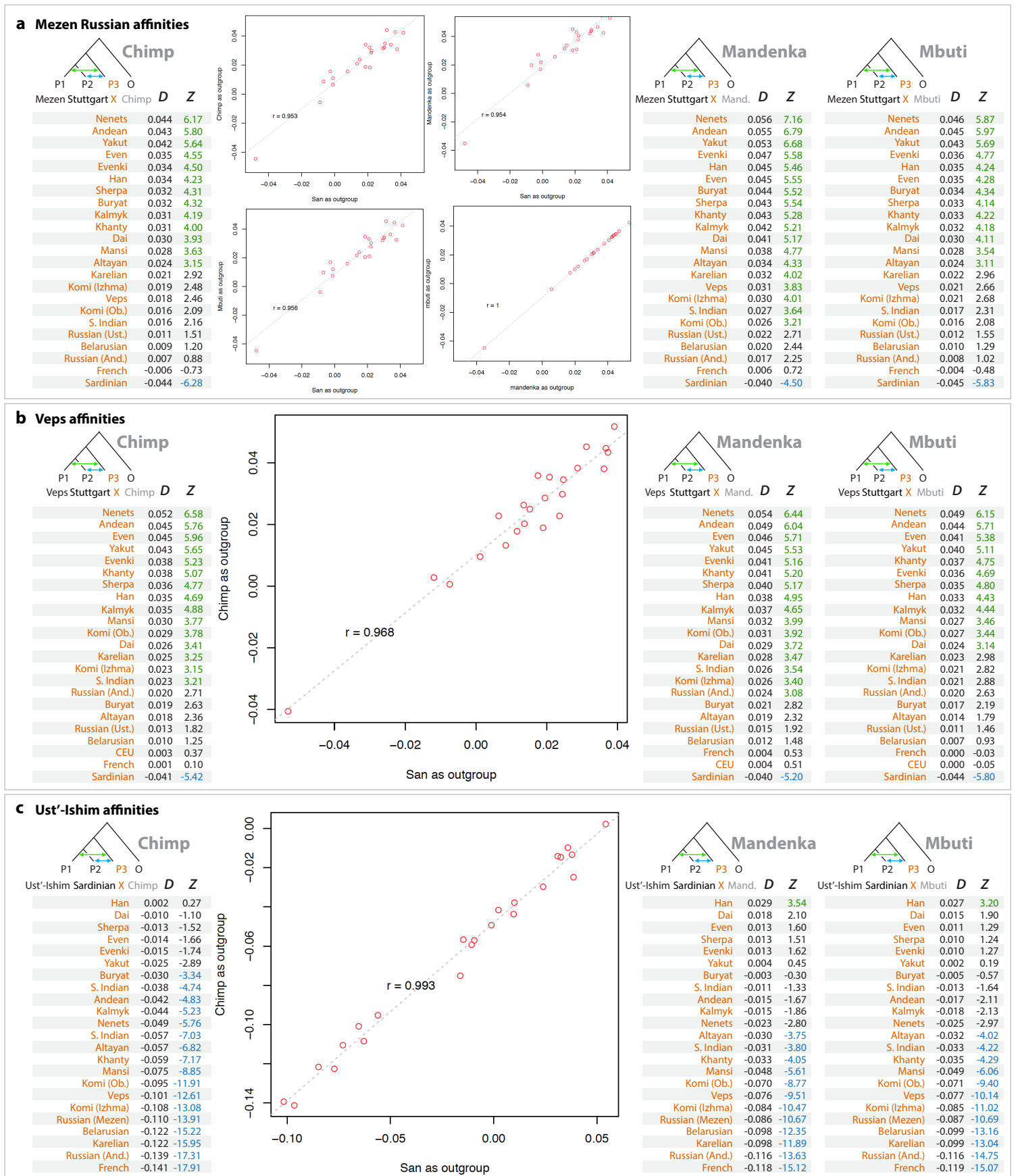
Detailed Y-DNA tree with error estimates. Terminal nodes (leaves) are represented by individual Y-DNA haplotypes. Sample names are shown with red font. Count represents a number of unique SNPs that are assigned to branch of the tree. If the branch is terminal, the count represents the number of SNPs specific to the individual's haplotype. If the branch is not terminal, the count represents the number of SNPs that are shared by two daughter subclades and are absent from individuals in other clades. Depth is calculated in SNP units and represents the averaged number of SNPs from the two subclades given their equal contribution (*Weight=0.5*). Var is a variance estimate of Depth based on Poisson mutation model from which standard error (S.E.) is calculated. Time is calculated from Depth value given a neutral rate of mutation; it represents the estimated split time of two daughter clades given a Poisson model and is expressed in years. Time (SE) is standard deviation of split time expressed in years.

FIGURE S11



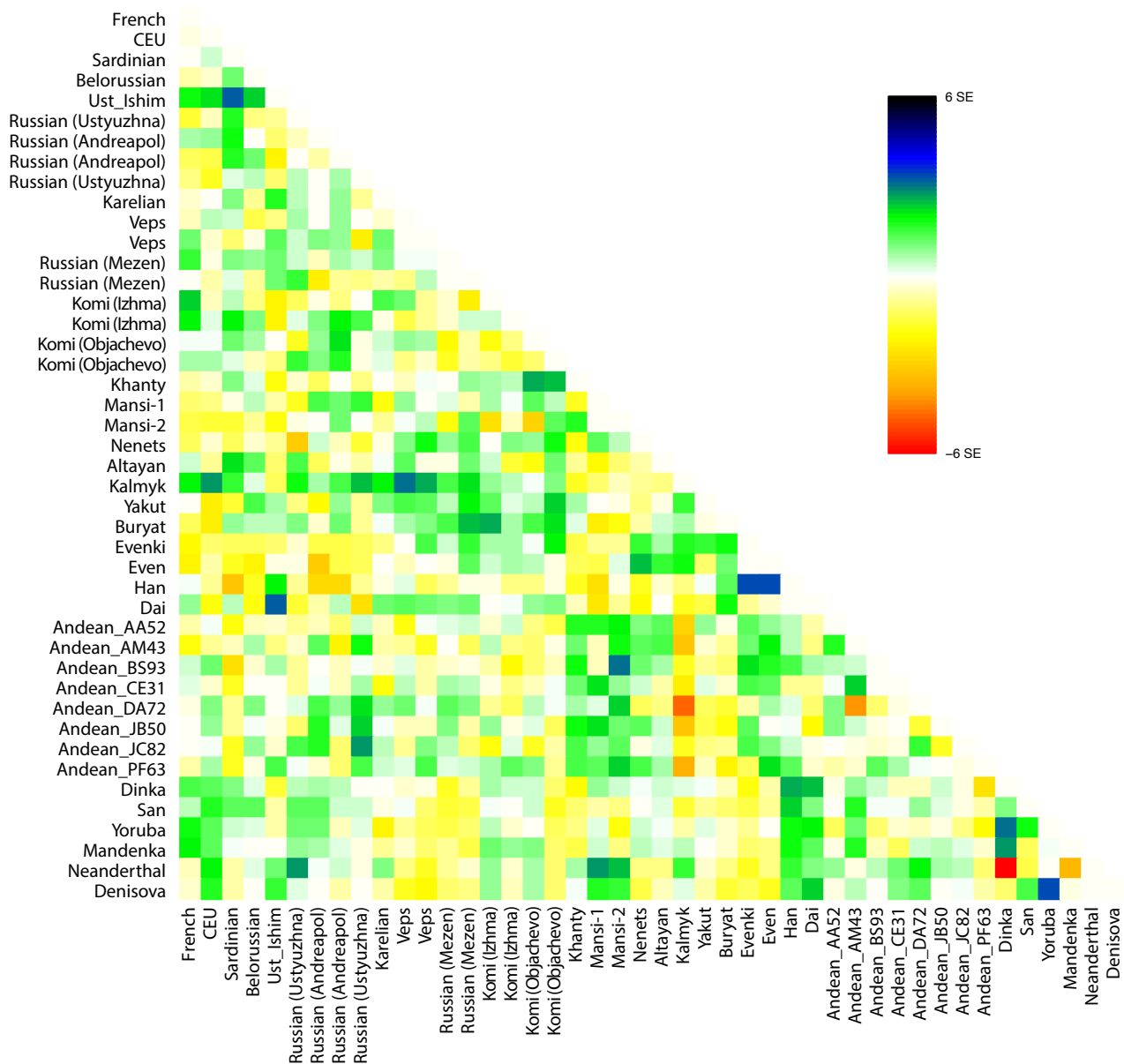
Ust'-Ishim is a member of NO Y-DNA clade. IGV browser (Robinson et al. 2011) was used to display raw sequencing reads aligned to NO-specific SNP (hg19 coordinate chrY: 7,690,182). Ust'-Ishim shares a T allele (indicated by red color) with Mansi (haplogroup N) and Han (haplogroup O), but not with French (haplogroup R) or Karitiana (haplogroup Q). Each horizontal grey bar represents a single sequencing read. Read bases differing from the reference genome are shown as green (for A), blue (for C), orange (for G) or red (for T). Other matching bases are shown with grey.

FIGURE S12



Analysis of admixtures using D-statistics (related to Figure 5a,b,c). Each panel shows comparison of D-statistics using different outgroups (Chimp, San, Mandenka, Mbuti). Scatter plots show high correlations of D-statistics between tested outgroups, suggesting that San and Chimp provide comparable results and therefore San can be used as an outgroup in our analyses. Genetic affinities were determined using D-statistics test $D(P_1, P_2, P_3, O)$ from the ADMIXTOOLS package (Patterson et al. 2012) and shown under D column. Statistical significance of admixtures was expressed in terms of z-score units and shown under Z column. Positive D-statistics values indicate possible admixture between P_1 and P_3 , while negative - between P_2 and P_3 . Green font was used to highlight significant admixtures between P_1 and P_3 ($Z > 3$). Blue font was used to highlight significant admixture between P_2 and P_3 ($Z < -3$). (And. = Andreapol, Ust. = Ustyuzhna, Ob. = Objachevo).

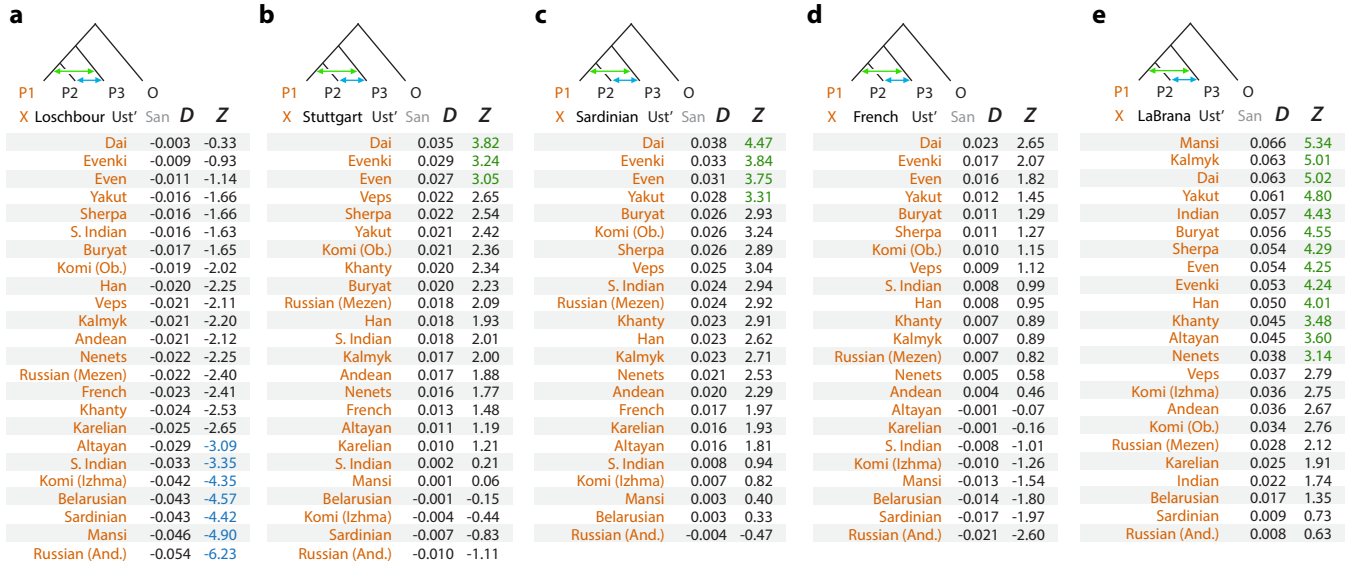
FIGURE S13



TreeMix residuals for Ust'-Ishim tree shown in Figure 6.

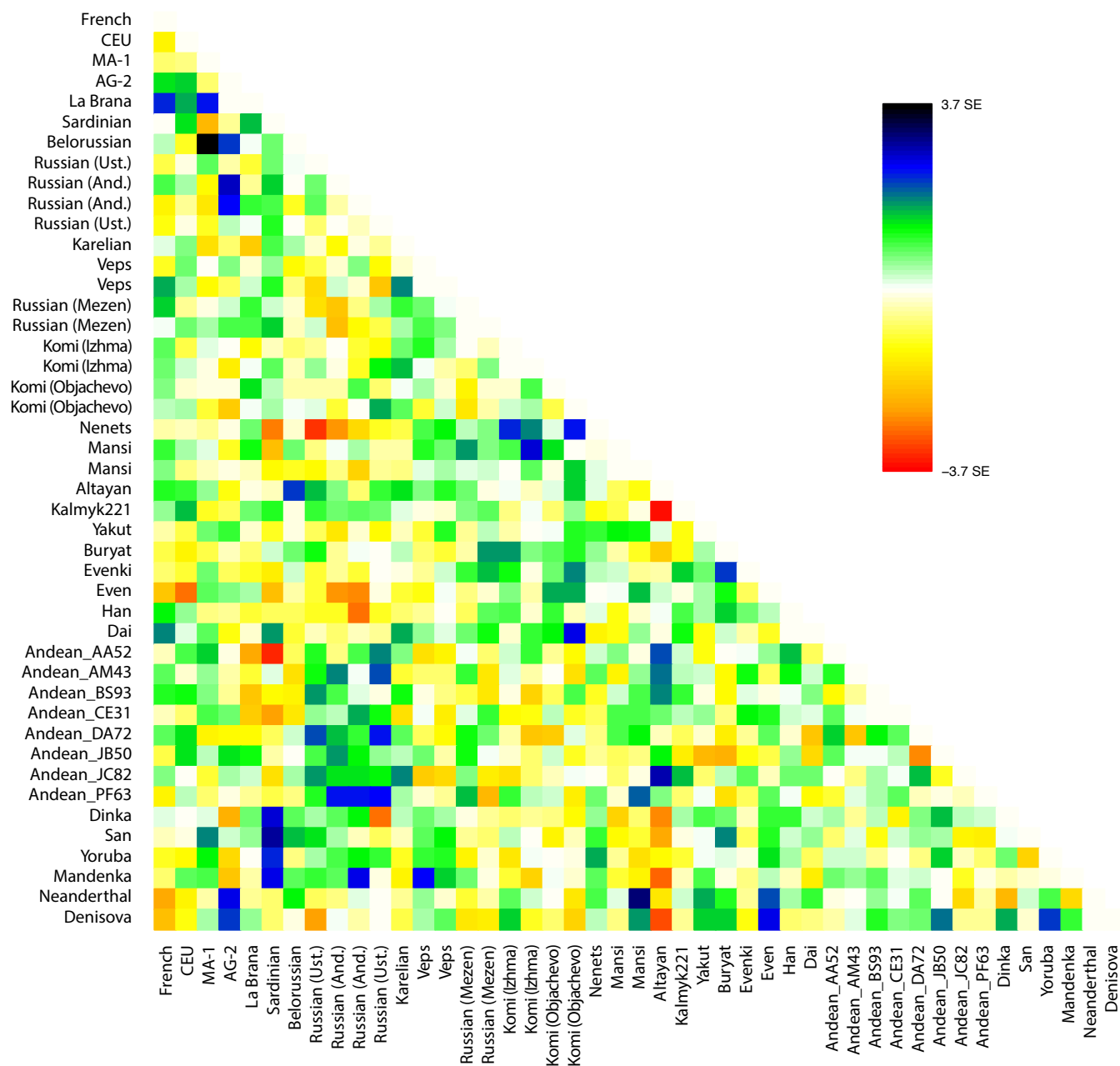
FIGURE S14

Ust'-Ishim affinities



Analysis of Ust'-Ishim affinities using D-statistics. Genetic affinities were determined using D-statistics test $D(P_1, P_2, P_3, O)$ from the ADMIXTOOLS package (Patterson et al. 2012) and shown under D column. Statistical significance of admixtures was expressed in terms of z-score units and shown under Z column. Positive D-statistics values indicate possible admixture between P_1 and P_3 , while negative - between P_2 and P_3 . Green font was used to highlight significant admixtures between P_1 and P_3 ($Z > 3$). Blue font was used to highlight significant admixture between P_2 and P_3 ($Z < -3$). African San was used as an outgroup in all tests. Alternative D-statistics using Chimp as an outgroup are shown in Supplementary Figure 11. (And. = Andreapol, Ust. = Ustyuzhna, Ob. = Objachevo). The results shown in panels **a-d** evaluate Ust'-Ishim's affinities using four different European individuals: Loschbour, Stuttgart, Sardinian, French. Results from Stuttgart and Sardinian (**b, c, e**) support the affinity between Ust'-Ishim and Dai/Evenki/Even/Mansi/Kalmyk. Results from Loschbour (**a**) do not show statistically significant admixture between Ust'-Ishim and Eastern Asians / Eastern Siberians, which may be due to existing admixture between Loschbour and Ust'-Ishim. Indeed, our additional analysis of TreeMix models using 10 ancient genomes show that Ust'-Ishim clades with ancient Hungarians NE1 and BR2 (Gamba et al. 2014), which are close to Loschbour, in addition to having admixture with Han (data not shown). Nonetheless, Dai/Evenki/Even are still the closest populations to Ust'-Ishim in this analysis. Panel (**d**) shows Dai/Evenki/Even having affinities with Ust'-Ishim, but the p-value does not reach the level of statistical significance ($Z \geq 3$).

FIGURE S15



TreeMix residuals for MA-1/AG-2/La Brana tree shown in Figure 7.

FIGURE S16

a Mal'ta boy (MA-1) affinities

Chimp			Mandenka			Mbuti		
P1	P2	P3	P1	P2	P3	P1	P2	P3
X	Han	MA-1	X	Han	MA-1	X	Han	MA-1
		Chimp			Mand.			Mbuti
		D			D			D
		Z			Z			Z
Andean	0.067	7.89	Andean	0.072	7.85	Russian (Mezen)	0.066	7.90
Russian (Mezen)	0.065	7.72	Russian (Mezen)	0.066	7.66	Andean	0.065	7.12
Karelian	0.059	7.15	Khanty	0.062	7.06	Karelian	0.062	7.43
Russian (Ust.)	0.059	7.06	Russian (Ust.)	0.061	7.14	Komi (Ob.)	0.059	7.05
Komi (Ob.)	0.057	7.05	Mansi	0.057	6.34	Russian (Ust.)	0.056	6.70
Khanty	0.054	6.46	Komi (Ob.)	0.057	6.11	Khanty	0.056	6.52
Veps	0.054	6.19	Karelian	0.055	6.54	French	0.054	6.64
Belarusian	0.052	6.37	Komi (Izhma)	0.051	6.02	Mansi	0.053	5.97
French	0.048	5.76	Veps	0.049	5.46	Veps	0.050	5.58
Nenets	0.047	5.53	Belarusian	0.045	5.13	Belarusian	0.049	5.90
Russian (And.)	0.047	5.90	French	0.045	5.03	Nenets	0.049	5.42
Komi (Izhma)	0.044	5.60	Nenets	0.043	4.80	Komi (Izhma)	0.046	5.51
Mansi	0.042	5.00	Russian (And.)	0.041	4.65	Russian (And.)	0.043	5.26
Yakut	0.030	3.70	Evenki	0.030	3.29	Yakut	0.033	3.93
Evenki	0.027	3.14	Yakut	0.028	3.23	Evenki	0.028	3.32
Altayan	0.027	3.19	Even	0.027	3.05	Altayan	0.022	2.56
Kalmyk	0.024	2.93	Kalmyk	0.025	3.07	Even	0.021	2.51
Even	0.017	2.02	Altayan	0.022	2.48	Kalmyk	0.018	2.25
Buryat	0.016	1.84	Buryat	0.019	2.23	Buryat	0.016	1.88
Sardinian	0.015	1.73	Sherpa	0.016	1.83	Sardinian	0.011	1.31
Dai	0.013	1.57	Sardinian	0.012	1.35	Dai	0.010	1.22
Sherpa	0.011	1.26	Dai	0.011	1.26	Sherpa	0.008	0.85

Afontova Gora (AG-2) affinities

Chimp			Mandenka			Mbuti		
P1	P2	P3	P1	P2	P3	P1	P2	P3
X	Han	AG-2	X	Han	AG-2	X	Han	AG-2
		Chimp			Mand.			Mbuti
		D			D			D
		Z			Z			Z
Belarusian	0.083	8.13	Belarusian	0.078	6.95	Belarusian	0.095	9.02
Russian (Mezen)	0.077	6.93	Russian (Mezen)	0.076	6.69	Mansi	0.090	8.12
Karelian	0.076	6.82	Komi (Ob.)	0.072	6.11	Komi (Ob.)	0.088	7.56
Komi (Ob.)	0.076	6.71	Mansi	0.071	6.15	Karelian	0.087	8.11
Mansi	0.074	6.46	Russian (Ust.)	0.070	6.11	Russian (Mezen)	0.083	7.68
Russian (And.)	0.073	6.63	Komi (Izhma)	0.066	5.69	Russian (Ust.)	0.078	7.07
French	0.073	6.81	Russian (And.)	0.063	5.58	Komi (Izhma)	0.077	6.93
Andean	0.072	6.06	Andean	0.062	5.21	Russian (And.)	0.076	6.97
Veps	0.070	6.42	Karelian	0.059	4.92	French	0.072	6.79
Russian (Ust.)	0.068	6.09	Khanty	0.058	4.95	Andean	0.071	5.96
Komi (Izhma)	0.060	5.51	French	0.057	5.11	Khanty	0.071	6.27
Nenets	0.060	5.19	Veps	0.054	4.84	Veps	0.070	6.47
Khanty	0.057	4.97	Altayan	0.044	3.75	Netets	0.063	5.71
Altayan	0.055	4.83	Nenets	0.043	3.66	Altayan	0.051	4.41
Sardinian	0.050	4.58	Sardinian	0.034	2.97	Kalmyk	0.048	4.11
Evenki	0.036	3.26	Kalmyk	0.029	2.47	Sardinian	0.046	4.15
Kalmyk	0.033	2.97	Even	0.021	1.83	Buryat	0.036	3.35
Even	0.030	2.67	Yakut	0.019	1.62	Yakut	0.035	3.08
Buryat	0.027	2.44	Buryat	0.019	1.64	Evenki	0.034	3.05
Yakut	0.027	2.33	Evenki	0.016	1.32	Even	0.027	2.39
Dai	0.015	1.29	Dai	-0.004	-0.34	Dai	0.011	0.92
Sherpa	0.009	0.80	Sherpa	-0.004	-0.33	Sherpa	0.007	0.65

b Ire8 affinities

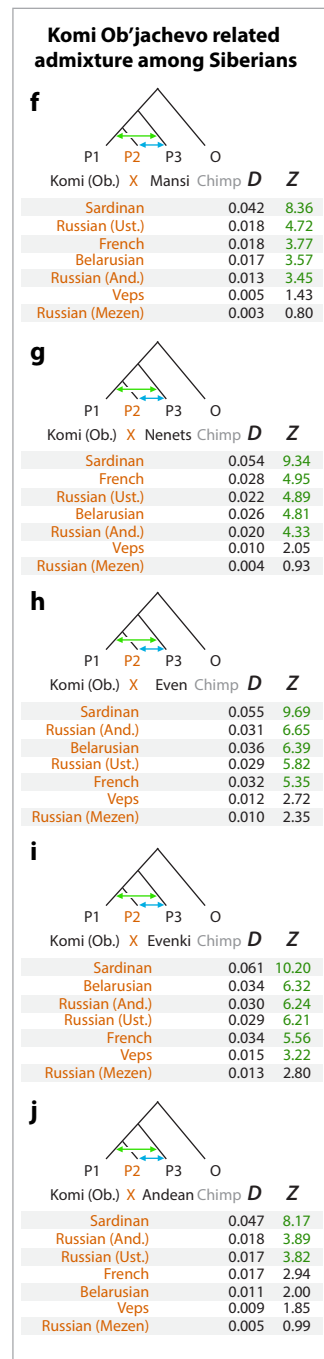
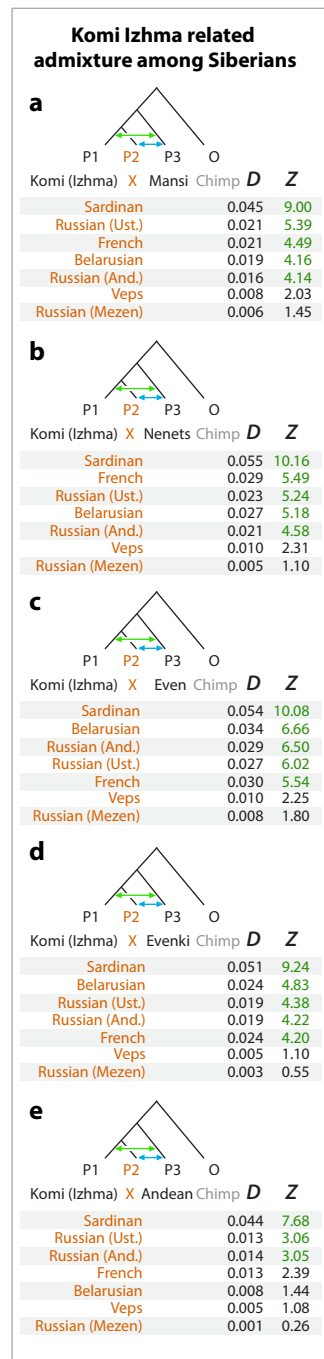
Chimp			Mandenka			Mbuti		
P1	P2	P3	P1	P2	P3	P1	P2	P3
X	Han	Ire8	X	Han	Ire8	X	Han	Ire8
		Chimp			Mand.			Mbuti
		D			D			D
		Z			Z			Z
French	0.133	3.93	Mansi	0.128	3.89	Mansi	0.169	5.34
Russian (Mezen)	0.125	3.89	Russian (Mezen)	0.109	3.45	Russian (Mezen)	0.136	4.45
Belarusian	0.107	3.21	Belarusian	0.100	3.01	French	0.135	4.12
Karelian	0.099	3.13	Karelian	0.094	2.99	Karelian	0.118	3.98
Russian (Ust.)	0.095	2.90	Veps	0.088	2.81	Belarusian	0.105	3.45
Veps	0.091	2.83	French	0.084	2.75	Veps	0.105	3.36
Russian (And.)	0.090	2.79	Russian (Ust.)	0.083	2.55	Russian (Ust.)	0.096	3.16
Khanty	0.079	2.34	Komi (Ob.)	0.070	2.10	Russian (And.)	0.091	2.93
Sardinian	0.070	2.23	Sardinian	0.068	2.34	Sardinian	0.081	2.67
Mansi	0.067	2.08	Khanty	0.055	1.73	Komi (Ob.)	0.080	2.39
Buryat	0.065	1.94	Russian (And.)	0.047	1.57	Khanty	0.075	2.36
Komi (Izhma)	0.057	1.70	Komi (Izhma)	0.046	1.40	Nenets	0.057	1.67
Andean	0.054	1.60	Andean	0.043	1.30	Andean	0.049	1.57
Komi (Ob.)	0.033	0.99	Altayan	0.032	0.94	Komi (Izhma)	0.043	1.34
Nenets	0.024	0.65	Evenki	0.026	0.77	Sherpa	0.035	1.01
Altayan	0.022	0.62	Kalmyk	0.025	0.76	Buryat	0.026	0.82
Evenki	0.018	0.50	Yakut	0.018	0.52	Altayan	0.019	0.60
Kalmyk	0.017	0.50	Nenets	0.014	0.42	Kalmyk	0.019	0.56
Yakut	0.005	0.13	Buryat	0.013	0.38	Yakut	0.017	0.52
Sherpa	-0.016	-0.47	Dai	-0.001	-0.03	Evenki	0.012	0.38
Dai	-0.025	-0.71	Even	-0.003	-0.10	Even	-0.016	-0.48
Even	-0.046	-1.25	Sherpa	-0.021	-0.61	Dai	-0.017	-0.50

Ajv52 affinities

Chimp			Mandenka			Mbuti		
P1	P2	P3	P1	P2	P3	P1	P2	P3
X	Han	Ajv52	X	Han	Ajv52	X	Han	Ajv52
		Chimp			Mand.			Mbuti
		D			D			D
		Z			Z			Z
Belarusian	0.135	6.08	Russian (Mezen)	0.109	4.99	Mansi	0.123	5.04
Karelian	0.124	5.58	Mansi	0.107	4.39	Komi (Ob.)	0.112	5.20
French	0.124	5.36	Karelian	0.093	4.17	Belarusian	0.108	5.25
Veps	0.122	5.55	Russian (Ust.)	0.090	4.06	Karelian	0.104	4.92
Russian (Ust.)	0.119	4.95	French	0.090	4.15	Russian (Ust.)	0.098	4.74
Russian (Mezen)	0.110	4.78	Komi (Izhma)	0.088	3.99	Russian (And.)	0.098	4.64
Sardinian	0.103	4.63	Belarusian	0.087	4.06	Veps	0.097	4.44
Komi (Izhma)	0.097	4.22	Veps	0.084	3.74	Russian (Mezen)	0.090	4.31
Mansi	0.097	4.06	Komi (Ob.)	0.078	3.62	French	0.087	4.25
Russian (And.)	0.086	3.90	Russian (And.)	0.076	3.47	Nenets	0.086	3.73
Komi (Ob.)	0.073	3.14	Sardinian	0.070	3.22	Sardinian	0.071	3.38
Andean	0.068	2.82	Nenets	0.059	2.65	Komi (Izhma)	0.071	3.16
Khanty	0.065	2.97	Khanty	0.041	1.88	Khanty	0.050	2.38
Nenets	0.065	2.74	Andean	0.025	1.05	Yakut	0.027	1.13
Dai	0.052	2.30	Buryat	0.022	0.95	Even	0.018	0.80
Altayan	0.046	1.98	Sherpa	0.021	0.89	Andean	0.016	0.71
Sherpa	0.041	1.65	Yakut	0.018	0.78	Evenki	0.016	0.69
Yakut	0.041	1.73	Kalmyk	0.015	0.64	Kalmyk	0.016	0.68
Evenki	0.033	1.47	Evenki	0.006	0.25	Altayan	0.014	0.66
Buryat	0.027	1.14	Altayan	-0.001	-0.06	Buryat	0.013	0.57
Even	0.024	1.01	Even	-0.001	-0.06	Sherpa	0.010	0.45
Kalmyk	0.008	0.33	Dai	-0.006	-0.25	Dai	0.005	0.23

Analysis of admixtures using D-statistics (related to Figure 5d,e). Each panel shows comparison of D-statistics using different outgroups (Chimp, San, Mandenka, Mbuti). Genetic affinities were determined using D-statistics test $D(P_1, P_2, P_3, O)$ as defined previously (Durand et al. 2011). The affinities of a variable population X (orange color) relative to other populations were calculated using ADMIXTOOLS (Patterson et al. 2012) and shown under D column. Statistical significance of admixture is expressed as a z-score and shown under Z column. Green font was used to highlight significant admixtures between P_1 and P_3 ($Z > 3$). Blue font was used to highlight significant admixture between P_2 and P_3 ($Z < -3$). (And. = Andreapol, Ust. = Ustyuzhna, Ob. = Objachevo).

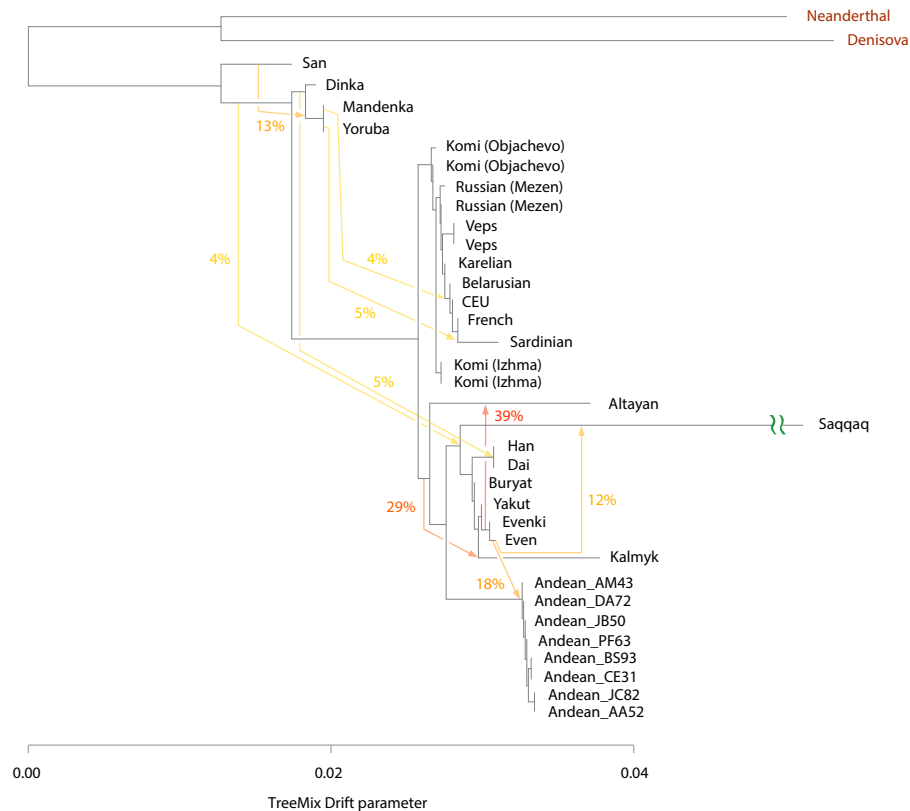
FIGURE S17



Analysis of recent European-related admixtures among Siberians. Genetic affinities were determined using D -statistics test $D(P_1, P_2, P_3, O)$ from the ADMIXTOOLS package (Patterson et al. 2012) and shown under D column. Statistical significance of admixtures was expressed in terms of z-score units and shown under Z column. Positive D -statistics values indicate possible admixture between P_1 and P_3 , while negative - between P_2 and P_3 . Green font was used to highlight significant admixtures between P_1 and P_3 ($Z > 3$). Blue font was used to highlight significant admixture between P_2 and P_3 ($Z < -3$). (And. = Andreapol, Ust. = Ustyuzhna, Ob. = Objachevo).

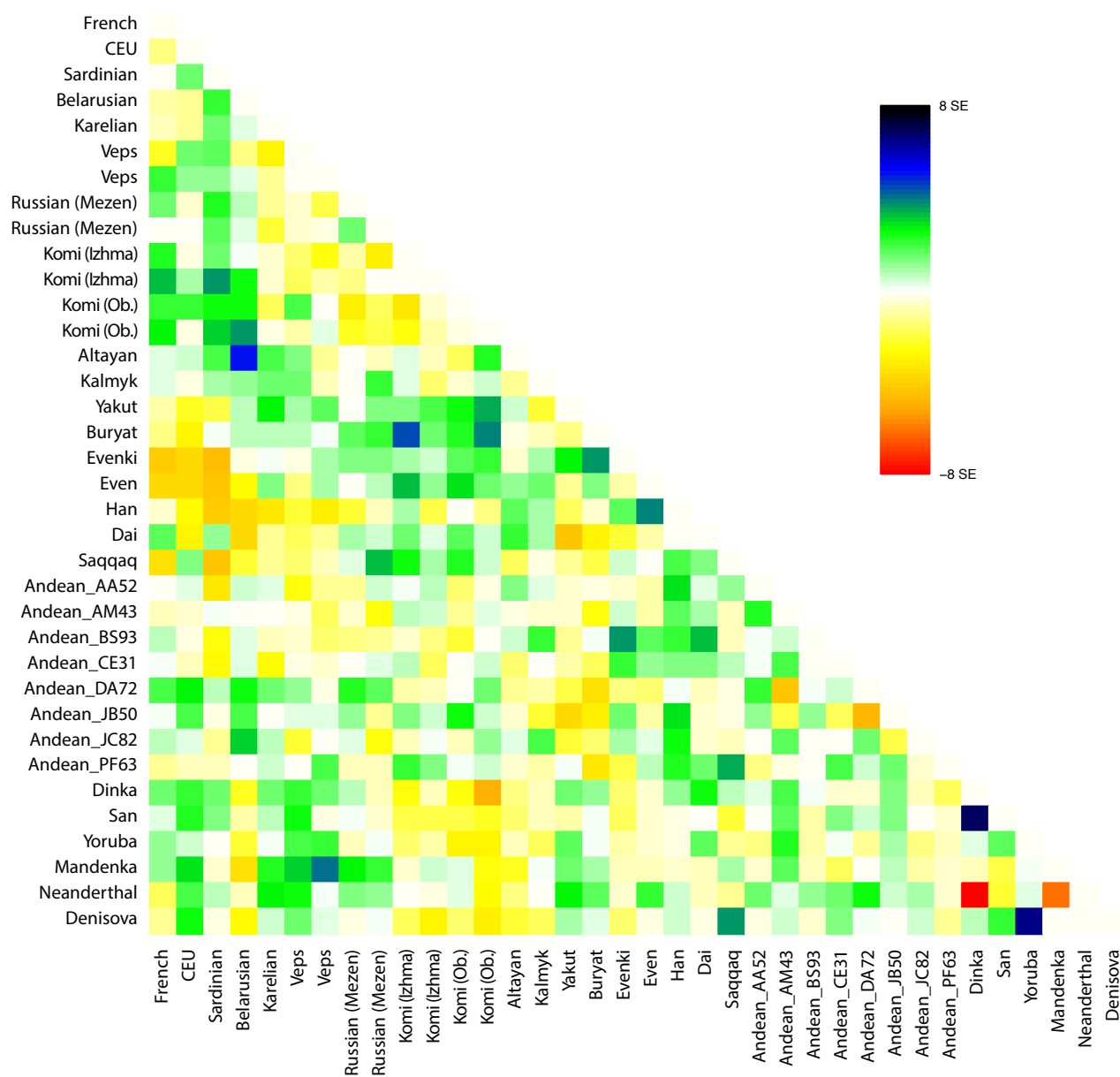
According to TreeMix models, Komi Objachevo and Komi Izhma are the two closest European populations to Mansi. To explore whether this could be explained by a recent European admixture into Mansi, we analyzed Komi-related admixtures among Siberians Mansi, Nenets, Even, Evenki and Native American Andean. Panels **a-e** and TreeMix models show that Mezen Russians and Veps are the two closest populations to Komi Izhma. However, Komi Izhma does not show statistical admixture signals with Siberians or Andean after they diverged from Veps and Mezen Russians, indicating that Siberian admixtures predated the divergence between Komi and Veps. Panels **f-j** show that Mezen Russians and Veps are the two closest populations to Komi Ob'jachevo (other than Komi Izhma). Furthermore, Komi Ob'jachevo is not statistically admixed with Siberians or Andean after their divergence from Mezen Russians or Veps. These observations argue against a recent admixture between Komi and Mansi. Instead, Komi-Mansi admixture likely occurred before the divergence between Komi, Veps and Mezen Russians.

FIGURE S18



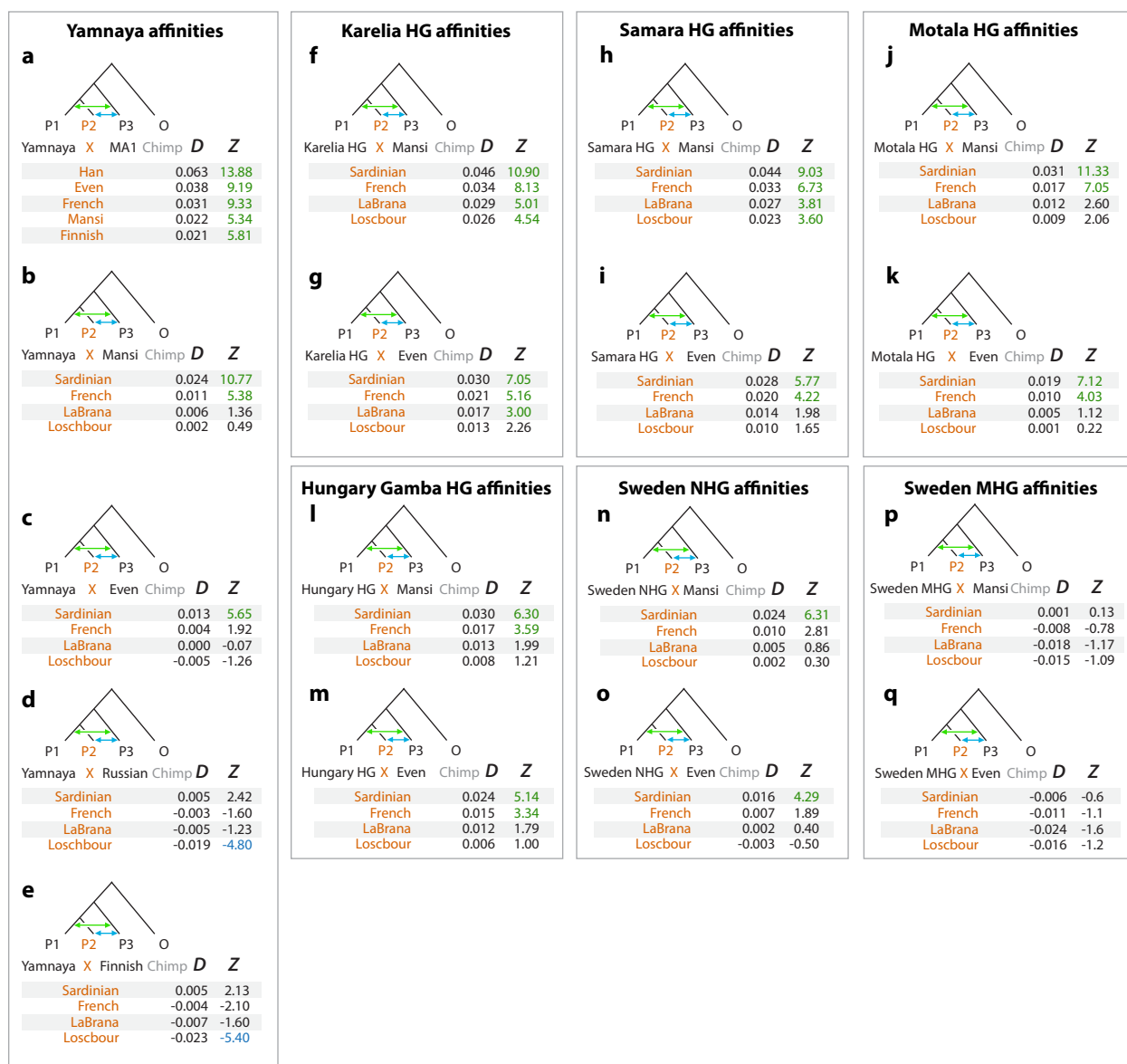
Genetic relationships of 4,000-year-old Greenland Saqqaq. TreeMix admixture graph that includes modern-day Eurasians and Saqqaq. Residual plot is shown in Supplementary Figure 20.

FIGURE S19



TreeMix residuals for the Saqqaq tree shown in Figure S19.

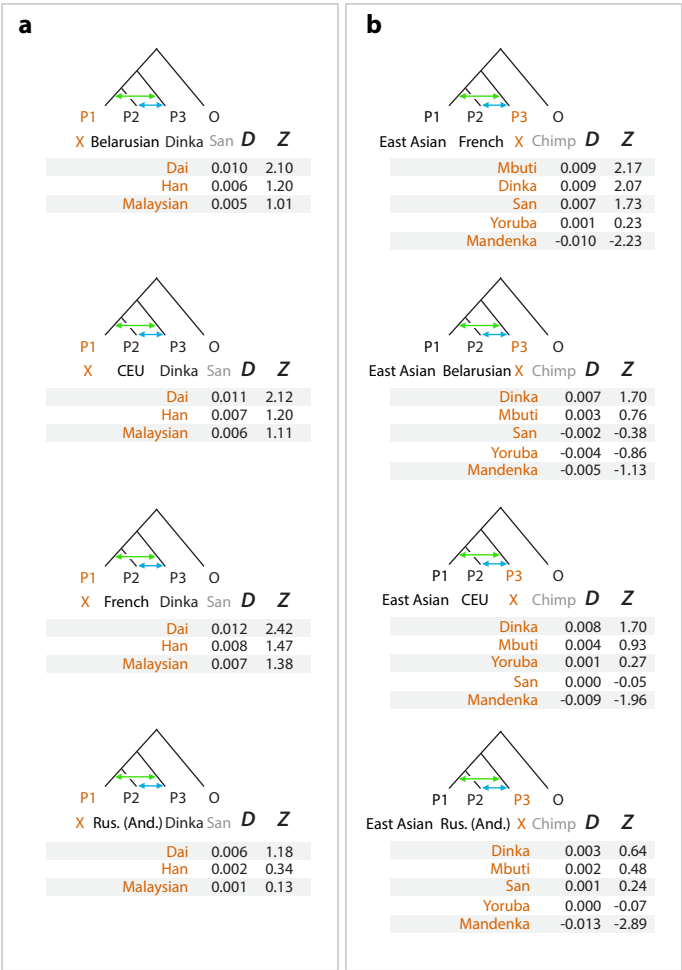
FIGURE S20



Analysis of Yamnaya and hunter-gatherer affinities using D-statistics. Genetic affinities were determined using D-statistics test $D(P_1, P_2, P_3, O)$ from the ADMIXTOOLS package (Patterson et al. 2012) and shown under D column. Statistical significance of admixtures was expressed in terms of z-score units and shown under Z column. Positive D-statistics values indicate possible admixture between P_1 and P_3 , while negative - between P_2 and P_3 . Green font was used to highlight significant admixtures between P_1 and P_3 ($Z > 3$). Blue font was used to highlight significant admixture between P_2 and P_3 ($Z < -3$). Panel (a) shows that Yamnaya has greater genetic affinity to ANE (represented by MA-1) than Mansi or Finnish. (b, c) Show that Mansi and to lesser degree Even have strong affinities with Yamnaya, unlike Russian (d) or Finnish (e). Panels (f-q) show D-statistics for detecting Mansi-related admixtures among Eastern European hunter gatherers from Karelia (f, g), Samara (h, i), Motala (j, k), Hungary Gamba (l, m), Sweden NHG (n, o) and Sweden MHG (p, q). Karelia, Samara and Motala hunter gatherers have the most affinity with Mansi, and weaker affinity with Even. Hungary Gamba HG and Sweden NHG show weak, but significant affinities with Mansi, while Sweden MHG does not have statistical admixture signals with Mansi or Even.

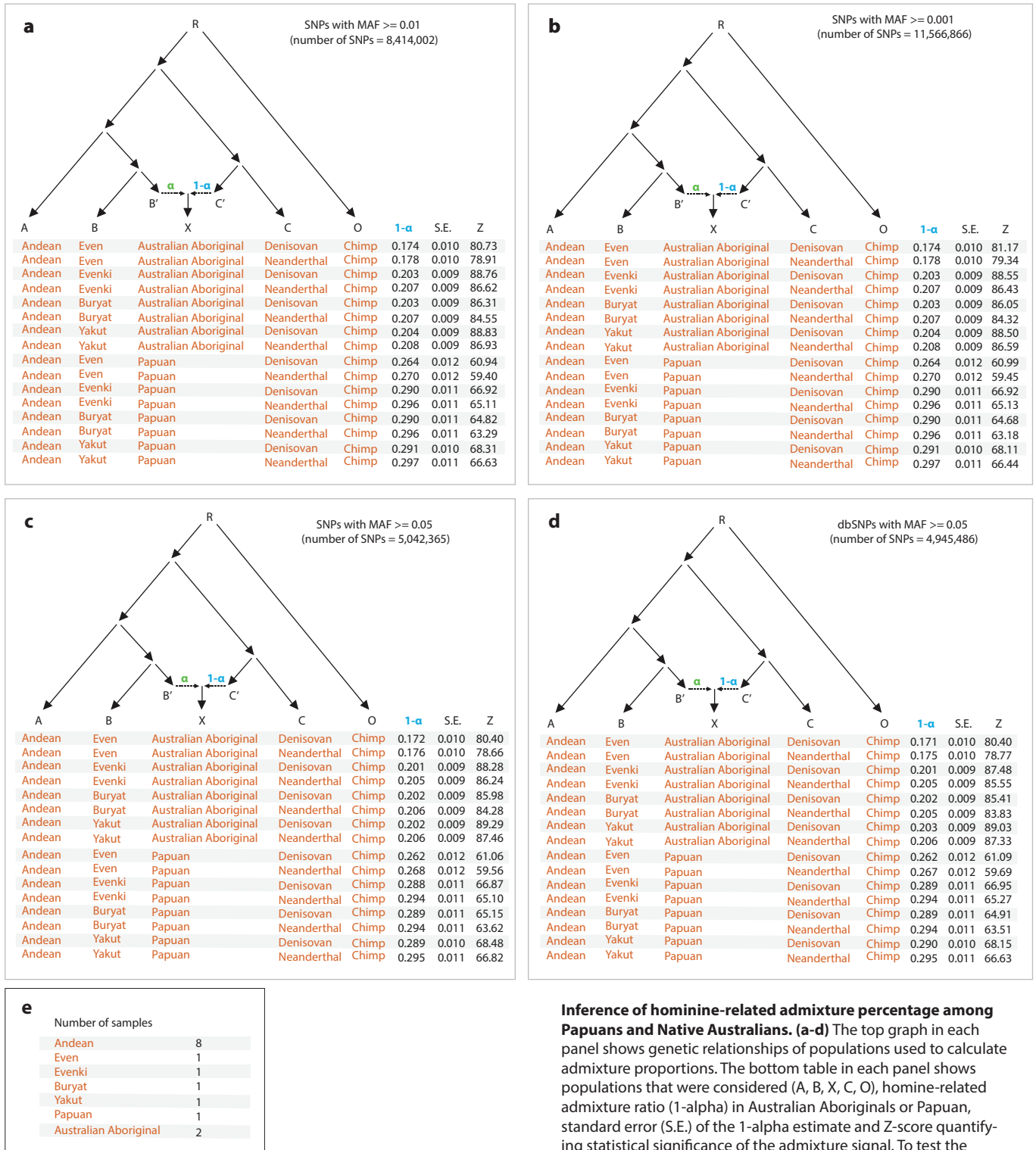
FIGURE S21

Genetic Affinities between East Asians and Africans



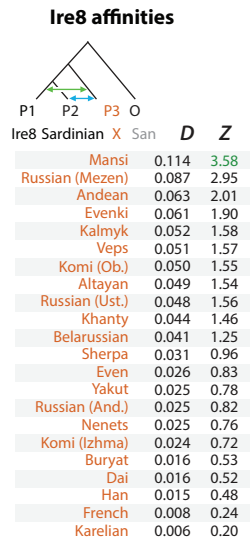
Analysis of African-related admixtures among East Asians using D-statistics. Genetic affinities were determined using D -statistics test $D(P_1, P_2, P_3, O)$ from the ADMIXTOOLS package (Patterson et al. 2012) and shown under D column. Statistical significance of admixtures was expressed in terms of z-score units and shown under Z column. Positive D values indicate greater affinity between P_1 and P_3 than between P_2 and P_3 . Negative D values indicate greater affinity between P_2 and P_3 than between P_1 and P_3 . Panel (a) shows that Dai individual has the greatest affinity to Dinka compared to other East Asians Han and Malaysian. Multiple individuals representing Europeans were investigated (Belarusian, CEU, French, Russian Andreapol). Panel (b) shows that Dinka is the population that has the highest admixture with East Asians (Han, Dai, Malaysian combined) compared to other African populations analyzed (Dinka, Mbuti, Yoruba, San, Mandenka). Mandenka is the least admixed with East Asians and instead shows admixture signals with Europeans. Admixtures should be interpreted with caution as none of the tests reach statistical significance level of $Z > 3$.

FIGURE S22



Inference of hominine-related admixture percentage among Papuans and Native Australians. (a-d) The top graph in each panel shows genetic relationships of populations used to calculate admixture proportions. The bottom table in each panel shows populations that were considered (A, B, X, C, O), homine-related admixture ratio (1-alpha) in Australian Aboriginals or Papuan, standard error (S.E.) of the 1-alpha estimate and Z-score quantifying statistical significance of the admixture signal. To test the robustness of the admixture percentage estimate, we considered various subsets of SNPs: (a) SNPs with MAF > 0.01 , (b) SNPs with MAF > 0.001 , (c) SNPs with MAF > 0.05 , (d) dbSNPs with MAF > 0.05 . The F4-ratio test $\alpha(A, B, X, C, O) = f_4(A, O; X, C) / f_4(A, O; B, C)$ was defined as in Patterson et al. 2012 and calculated using ADMIXTOOLS. Panel (e) shows the number of individuals in each population used for calculation of the test. Eight Andean Native American individuals were combined under "Andean" label in all calculations. Two Australian Aboriginals were combined under "Australian Aboriginal" label in all calculations. The F4-ratio test shows that Australian Aboriginals have 17-21% of hominine-related ancestry, while Papuan has 27-30% of hominine-related ancestry.

FIGURE S23



Ire8-Mansi admixture. Genetic affinities were determined using *D*-statistics test from the ADMIXTOOLS package (Patterson et al. 2012). The affinities of a variable population X (orange color) relative to other populations are reflected by the *D*-statistics score (*D* column). Statistical significance of admixture is expressed as a z-score and shown under *Z* column. Green font was used to highlight significant admixtures between P_1 and P_3 ($Z > 3$). Blue font was used to highlight significant admixture between P_2 and P_3 ($Z < -3$). (And. = Andreapol, Ust. = Ustyuzhna, Ob. = Objachevo).

Figure S24

Table S1 | List of 892 samples included in the PCA plot

Population	Number of sample	Reference
Altayan	1	This study
	3	Raghavan et al. 2014
	13	Fedorova et al. 2013
Andean Highlander	10	Zhou et al. 2013
Belarusian	1	This study
Buryat	1	This study
	15	Fedorova et al. 2013
CEU	101	1000 Genomes (Omni 2.5 array data)
Estonian	14	Raghavan et al. 2014
Evenki	1	This study
	13	Fedorova et al. 2013
Even	1	This study
	8	Fedorova et al. 2013
Finn	100	1000 Genomes (Omni 2.5 array data)
French	28	Jun et al. 2008
Han	43	Jun et al. 2008
	142	1000 Genomes (Omni 2.5 array data)
Kalmyk	1	This study
Karelian	1	This study
Karitiana	5	Jun et al. 2008
Khanty	41	This study
Komi Izhma	2	This study
	73	Khrunin et al. 2013
Komi Objachevo	2	This study
	67	Khrunin et al. 2013
Mansi	48	This study
Mari	14	Raghavan et al. 2014
Mongolian	9	Fedorova et al. 2013
	10	Jun et al. 2008
Nenets	1	This study
Nganasan	11	Fedorova et al. 2013
North Italian	13	Jun et al. 2008
Russian	6	This study
	25	Jun et al. 2008
Ukranian	20	Yunusbayev et al. 2012
Veps	2	This study
	20	Khrunin et al. 2013
Yakut	1	This study
	25	Jun et al. 2008

[1] Raghavan et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature. 2014.

[2] Fedorova et al. Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. BMC Evol Biol. 2013.

[3] The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012.

[4] Li et al. Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. Science. 2008.

[5] Khrunin et al. A genome-wide analysis of populations from European Russia reveals a new pole of genetic diversity in northern Europe. PLoS One. 2013.

[6] Yunusbayev et al. The Caucasus as an Asymmetric Semipermeable Barrier to Ancient Human Migrations. Mol Biol Evol. 2012.

[7] Zhou et al. Whole-Genome Sequencing Uncovers the Genetic Basis of Chronic Mountain Sickness in Andean Highlanders. Am J Hum Genet. 2013.

Table S2 | **List of previously published genomes used in this study**

Geographic region	Ethnos	Samples	Sequence coverage	Reference
Africa	San	1	36.9	Prüfer et al. 2014
	Yoruba	2	38.6, 41.6	Prüfer et al. 2014, Illumina
	Mandenka	1	35.3	Prüfer et al. 2014
	Dinka	1	36.2	Prüfer et al. 2014
	Mbuti	1	35.3	Prüfer et al. 2014
Europe	CEPH European	1	27.5	Illumina
	Sardinian	1	36.9	Prüfer et al. 2014
	French	1	40.4	Prüfer et al. 2014
Asia	Han	1	33.4	Prüfer et al. 2014
	Dai	1	35.1	Prüfer et al. 2014
	Malaysian	2	36.2, 43.0	Wong et al. 2013
	South Indian	5	33.7, 35.3, 35.3, 39.0, 40.6	Wong et al. 2014
	Sherpa	2	37.4, 42.0	Jeong et al. 2014
Oceania	Australian BUR, E	1	40.7	Prüfer et al. 2014
	Australian WON, M	1	39.1	Prüfer et al. 2014
	Papuan	1	41.0	Prüfer et al. 2014
America	Karitiana	1	34.6	Prüfer et al. 2014
	Andean Highlander	8	36.5, 37.5, 38.0, 38.9, 39.7, 40.0, 40.2, 42.0	Zhou et al. 2013

[1] Prüfer et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014.

[2] <http://www.illumina.com/platinumgenomes/>

[3] Wong et al. Deep Whole-Genome Sequencing of 100 Southeast Asian Malays. Am J Hum Genet. 2013.

[4] Wong et al. Insights into the Genetic Structure and Diversity of 38 South Asian Indians from Deep Whole-Genome Sequencing. PLoS Genet. 2014.

[5] Zhou et al. Whole-Genome Sequencing Uncovers the Genetic Basis of Chronic Mountain Sickness in Andean Highlanders. Am J Hum Genet. 2013.

[6] Jeong et al. Admixture facilitates genetic adaptations to high altitude in Tibet. Nat Commun. 2014.

Table S3 | List of ancient genomes from other studies

Sample name	Geographic region	Age	Sequence coverage	Reference	Culture
Neanderthal	Altai mountains in Siberia	50,000	50x	Prüfer et al. 2014	.
Denisova	Altai mountains in Siberia	50,000	30x	Meyer et al. 2012	.
Ust'-Ishim	Omsk Oblast, Russian Federation	45,000	42x	Fu et al. 2014	.
MA-1	Mal'ta in south-central Siberia	24,000	1x	Raghavan et al. 2014	Ancient north Eurasian
AG-2	Afontova Gora-2 in south-central Siberia	17,000	0.1x	Raghavan et al. 2014	Ancient north Eurasian
Anzick-1	Western Montana, America	12,500	14.4x	Rasmussen et al. 2014	.
Loschbour	Heffingen, Luxembourg	8,000	22x	Lazaridis et al. 2014	Western European hunter-gatherers
Stuttgart	Stuttgart-Mühlhausen, Germany	7,000	19x	Lazaridis et al. 2014	Early European farmers
La Braña	León, Spain	7,000	3.4x	Olalde et al. 2014	Western European hunter-gatherers
NE1	Polgár-Ferenci-hát, Hungary	5,000	22x	Gamba et al. 2014	Middle Neolithic Alföld Linear Pottery
Ajv52	Gotland, Sweden	5,000	< 0.1x	Skoglund et al. 2012	Scandinavian Neolithic hunter-gatherers
Ire8	Gotland, Sweden	5,000	< 0.1x	Skoglund et al. 2012	Scandinavian Neolithic hunter-gatherers
Saqqaq	Qeqertasussuk, north-western Greenland	4,000	20x	Rasmussen et al. 2010	.
Australian aboriginal	Southern western Australia	100	6.4x	Rasmussen et al. 2011	.

Datasets analyzed using previously calculated SNPs

Yamnaya	Samara oblast, Russia	5,300-4,700	N/A	Haak et al. 2015	Yamnaya
Samara HG	Samara oblast, Russia	8,000-6,600	N/A	Haak et al. 2015	Eastern European hunter gatherers
Karelia HG	Karelia, Western Russia	8,000-6,600	N/A	Haak et al. 2015	Eastern European hunter gatherers
Motala HG (Motala12)	Östergötland, Sweden	8,000-6,600	N/A	Lazaridis et al. 2014	Scandinavian Neolithic hunter gatherers
Hungary EN	Various locations in Hungary	8,000-7,500	N/A	Haak et al. 2015	Early Neolithic
Hungary HG (KO1)	Tiszaszőlös-Domaháza, Hungary	8,000-6,600	N/A	Gamba et al. 2014	Western European hunter-gatherers
Sweden MHG	Gotland, Sweden	8,000-6,600	N/A	Skoglund et al. 2014	Scandinavian Neolithic hunter gatherers
Sweden NHG	Gotland, Sweden	5,300-4,700	N/A	Skoglund et al. 2014	Scandinavian Neolithic hunter gatherers

[1] Prüfer et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. 2014.

[2] Meyer et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science*. 2012.

[3] Raghavan et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*. 2014.

[4] Rasmussen et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature*. 2014.

[5] Lazaridis et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. 2014.

[6] Olalde et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature*. 2014.

[7] Skoglund et al. Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science*. 2012.

[8] Rasmussen et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*. 2010.

[9] Rasmussen et al. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science*. 2011.

[10] Fu et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*. 2014.

[11] Gamba et al. Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications*. 2014.

[12] Haak et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. 2014.

[13] Skoglund et al. Genomic diversity and admixture differs for Stone-Age Scandinavian Foragers and Farmers. *Nature*. 2014.

Table S4 | Quality control and concordance of SNP calls

Samples	Number of SNPs (GQ>30)		Overlapping SNPs	Genotype concordance	Number of unique variants		% of this study unique SNPs in original study GQ <= 30 SNPs	Ti/Tv					
	This study	Original study			This study	Original study		This (all SNPs)	Original (all)	This (dbSNP)	Original (dbSNP)	This (unique)	Original (unique)
French2	3,568,386	4,307,164	3,484,376	100%	82,581	821,359	69.45	2.04	1.87	2.07	1.99	1.15	0.98
Sardinian2	3,536,527	4,297,034	3,449,074	100%	85,951	846,458	80.86	2.04	1.86	2.07	1.99	1.2	0.99
Han3	3,674,350	4,348,837	3,465,461	100%	96,037	881,904	81.36	2.02	1.81	2.05	1.97	1.12	0.86
San2	4,664,486	5,566,763	4,571,703	100%	91,053	993,330	86.99	2.05	1.88	2.07	2	1.84	1.39
Mandenka2	4,312,669	5,137,427	4,220,231	100%	90,856	915,614	81.50	2.05	1.88	2.07	2.01	1.46	1.05
Dinka2	4,253,154	5,029,340	4,150,525	100%	101,068	877,254	71.74	2.04	1.88	2.07	2.01	1.48	1.07
Yoruba2	4,364,781	5,210,848	4,269,200	100%	93,917	939,984	66.59	2.04	1.87	2.07	2	1.31	1.03
Andean_AA52	3,536,316	2,961,941	2,921,655	100%	613,888	39,513	3.67	1.86	2.12	2.05	2.12	0.46	2.08
Andean_BS93	3,580,208	3,001,366	2,964,128	100%	615,308	36,466	3.13	1.86	2.11	2.04	2.11	0.48	2.07
Andean_DA72	3,581,161	3,001,784	2,965,944	100%	614,387	35,010	3.10	1.85	2.11	2.04	2.11	0.44	2.07
Malaysian_SSM018	3,574,845	3,338,291	3,172,476	100%	402,321	165,767	8.01	2.01	2.15	2.06	2.15	1.05	2.04
Malaysian_SSM093	3,533,037	3,317,283	3,150,499	100%	382,495	166,741	8.42	2.03	2.15	2.06	2.15	1.28	2.05
Indian_SSI009	3,658,802	3,256,573	3,185,948	100%	472,808	70,579	5.17	2.04	2.13	2.06	2.13	1.51	2.1
Indian_SSI029	3,630,134	3,156,530	3,096,590	100%	533,495	59,891	4.62	2.03	2.13	2.06	2.13	1.43	2.12
Indian_SSI032	3,622,253	3,143,683	3,092,284	100%	529,935	51,365	3.45	2.04	2.13	2.06	2.13	1.58	2.15
Indian_SSI032_real	3,666,319	3,165,225	3,116,769	100%	549,514	48,420	2.85	2.02	2.13	2.06	2.13	1.4	2.11
Indian_SSI036	3,643,094	3,248,539	3,177,951	100%	465,101	70,546	5.35	2.04	2.13	2.06	2.13	1.55	2.11
Sherpa_1	3,760,291	3,594,092	3,368,025	100%	391,191	224,992	9.09	1.82	1.98	2.04	2.06	0.53	1.97
Sherpa_2	3,769,261	3,550,161	3,344,293	100%	423,904	204,804	9.04	1.78	2.01	2.05	2.06	0.43	1

The table shows comparison of SNP calls from this study (GATK) and original study SNPs using 19 high-coverage genomes ('Samples' column). The number of SNPs represents high-quality variants using GQ>30 criteria. We compared genotype concordance on the set of common SNPs between this study and original study variant calls. Ti/Tv ratios were calculated for each set of shown SNPs to investigate quality of variant calls.

Supp Table 5 | Ti/Tv ratios of genomes from Siberian and Eastern European populations

Ethnic group	Samples	Ti/Tv ratios		
		All SNPs	dbSNPs	Non-dbSNP
Komi Izhma	2	2.03	2.05	1.22
		2.01	2.05	1.08
Komi Ob'yachevo	2	2.02	2.05	1.16
		2.02	2.05	1.09
Veps	2	2.02	2.05	1.14
		2.03	2.05	1.17
Karelians	1	1.89	2.06	0.50
Russians (Mezen)	2	2.03	2.05	1.24
		2.03	2.05	1.29
Russians (Ustyuzhna)	2	1.69	2.02	0.39
		1.84	2.04	0.44
Russians (Andreapol)	2	1.70	2.03	0.38
		1.85	2.04	0.47
Belarusians	1	1.90	2.06	0.53
Mansi	3	1.97	2.03	1.22
		1.94	2.03	1.11
		1.96	2.03	1.20
Khanty	4	2.02	2.04	1.25
		2.00	2.04	1.10
		1.78	2.03	0.47
		1.79	2.03	0.47
Nenets	1	2.01	2.04	1.22
Evens	1	1.96	2.04	0.86
Evenki	1	1.96	2.04	0.81
Yakuts	1	1.94	2.06	0.69
Altayans	1	1.96	2.05	0.82
Buryats	1	1.97	2.05	0.92
Kalmyks	1	1.97	2.05	0.86