Cell Genomics, Volume *3*

Supplemental information

Extraordinary selection on the human X chromosome associated with archaic admixture

Laurits Skov, Moisès Coll Macià, Elise Anne Lucotte, Maria Izabel Alvez Cavassim, David Castellano, Mikkel Heide Schierup, and Kasper Munch

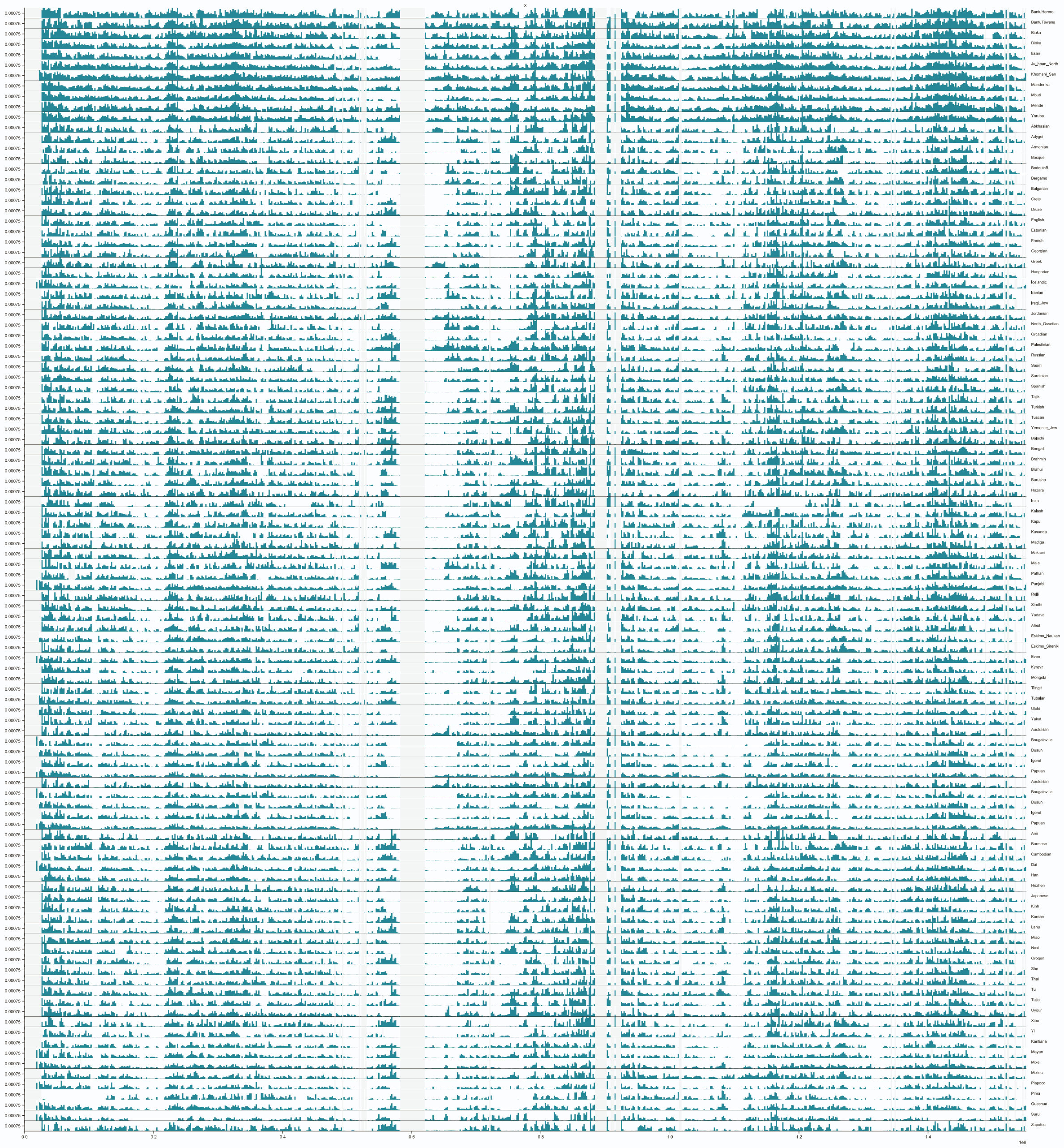

Figure S1: Mean pairwise differences in 200kb windows across chromosome X for each population (related to Figure 1). We discarded constituent 100kb windows with fewer than 50,000 called positions (50%) and considered these windows missing data. For a better visual comparison, the y-axis is truncated at 0.0015 to remove outliers. Light grey regions represent missing data.

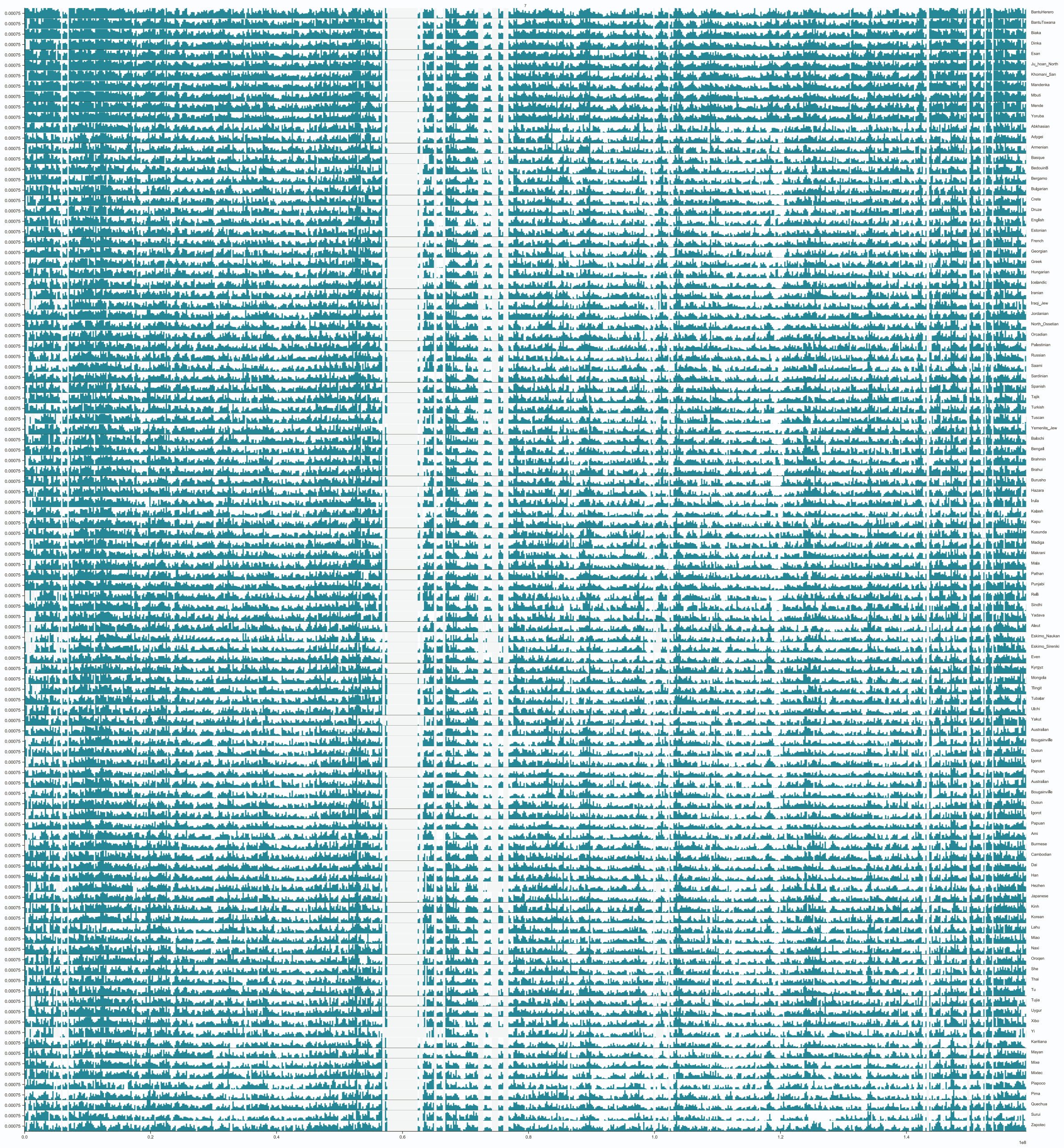

Figure S2: Mean pairwise differences in 100kb windows across chromosome 7 for each population (related to Figure 1). We discarded constituent 100kb windows with fewer than 50,000 called positions (50%) and considered these windows missing data. For a better visual comparison, the y-axis is truncated at 0.0015 to remove outliers. Grey regions represent missing data.

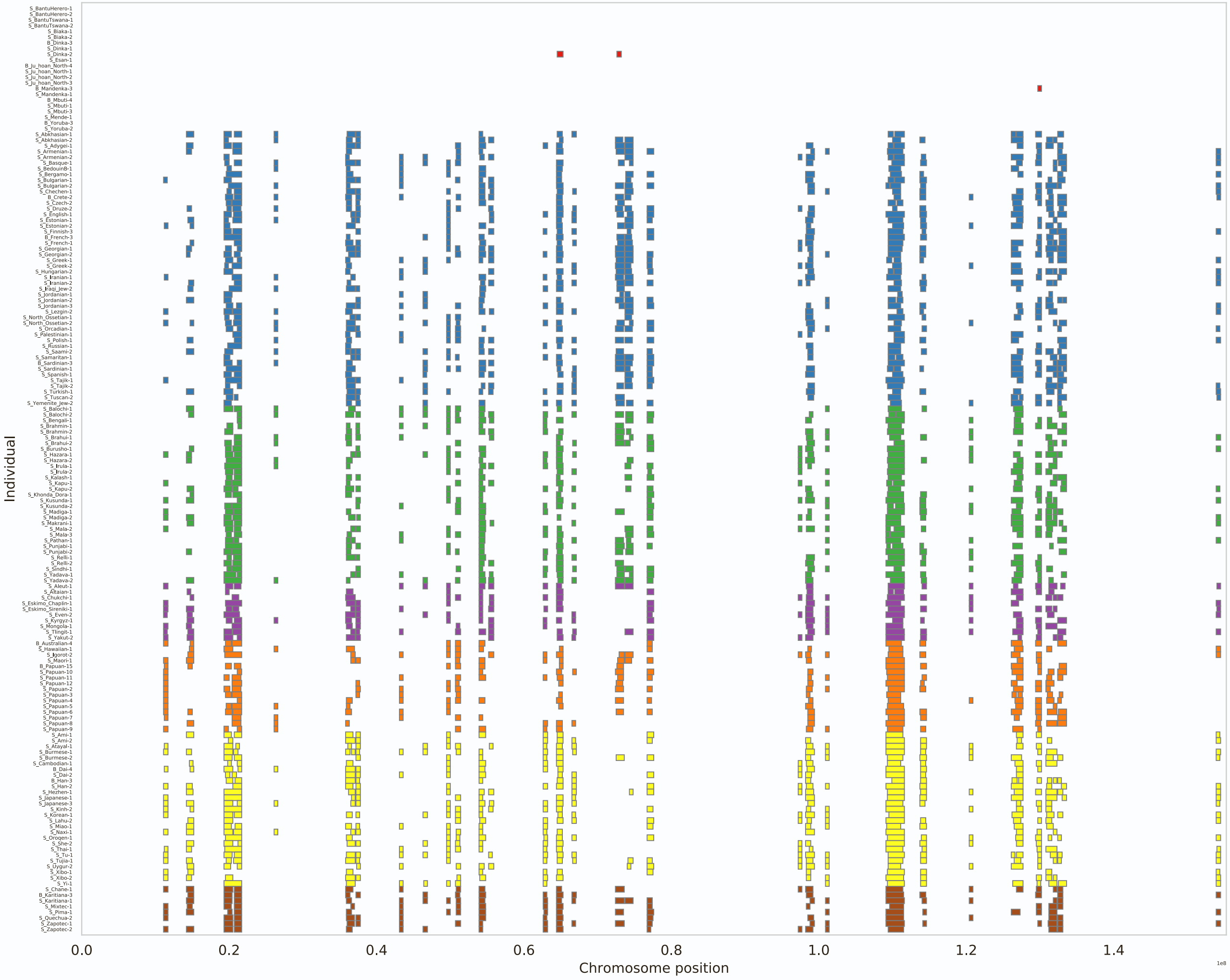

Figure S3: ECH inference using a clique size of 20% (related to Figure 2). ECHs identified on each male X chromosome using 20% as the minimum number of individuals included in the haplotype clades defining an ECH.

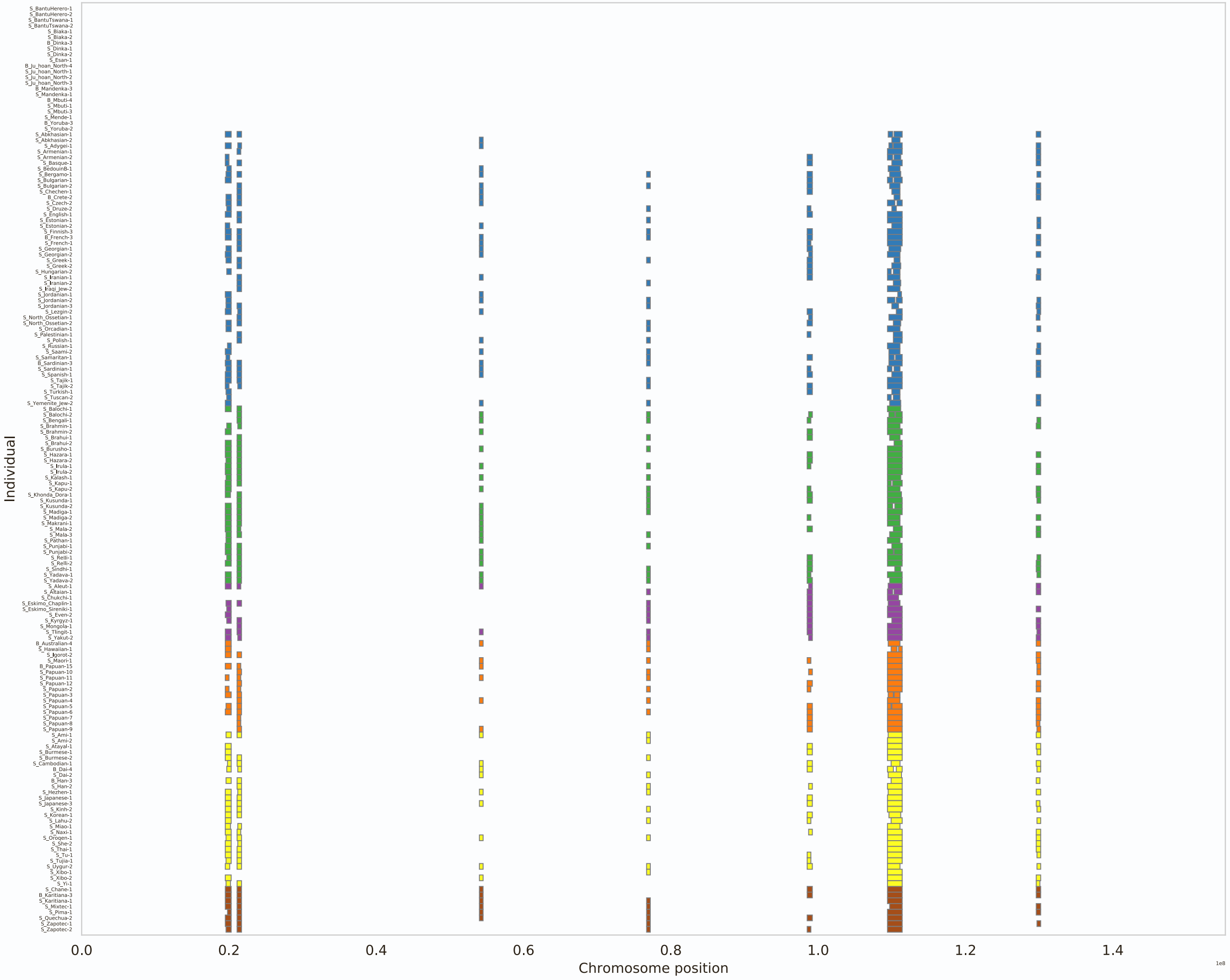

Figure S4: ECH inference using a clique size of 30% (related to Figure 2). ECHs identified on each male X chromosome using 30% as the minimum number of individuals included in the haplotype clades defining an ECH.

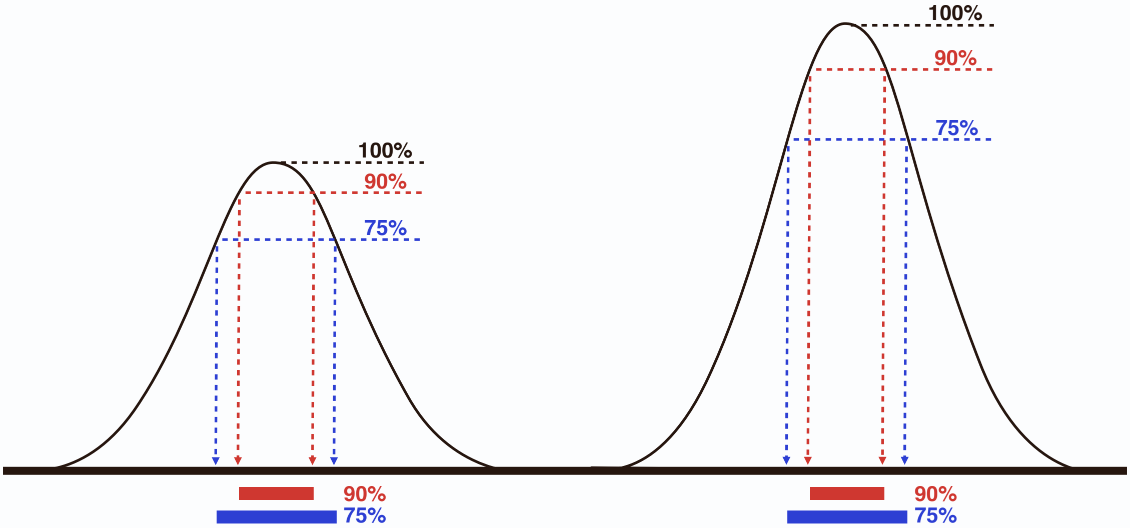

Figure S5: Core region of an ECH (related to Figure 4 and STAR methods). We operationally define a region around each peak where the proportion of individuals called as ECH is at least 90% of the peak value. Similarly, we define a wide region where the proportion of individuals called as ECH is above 75% of the peak value.

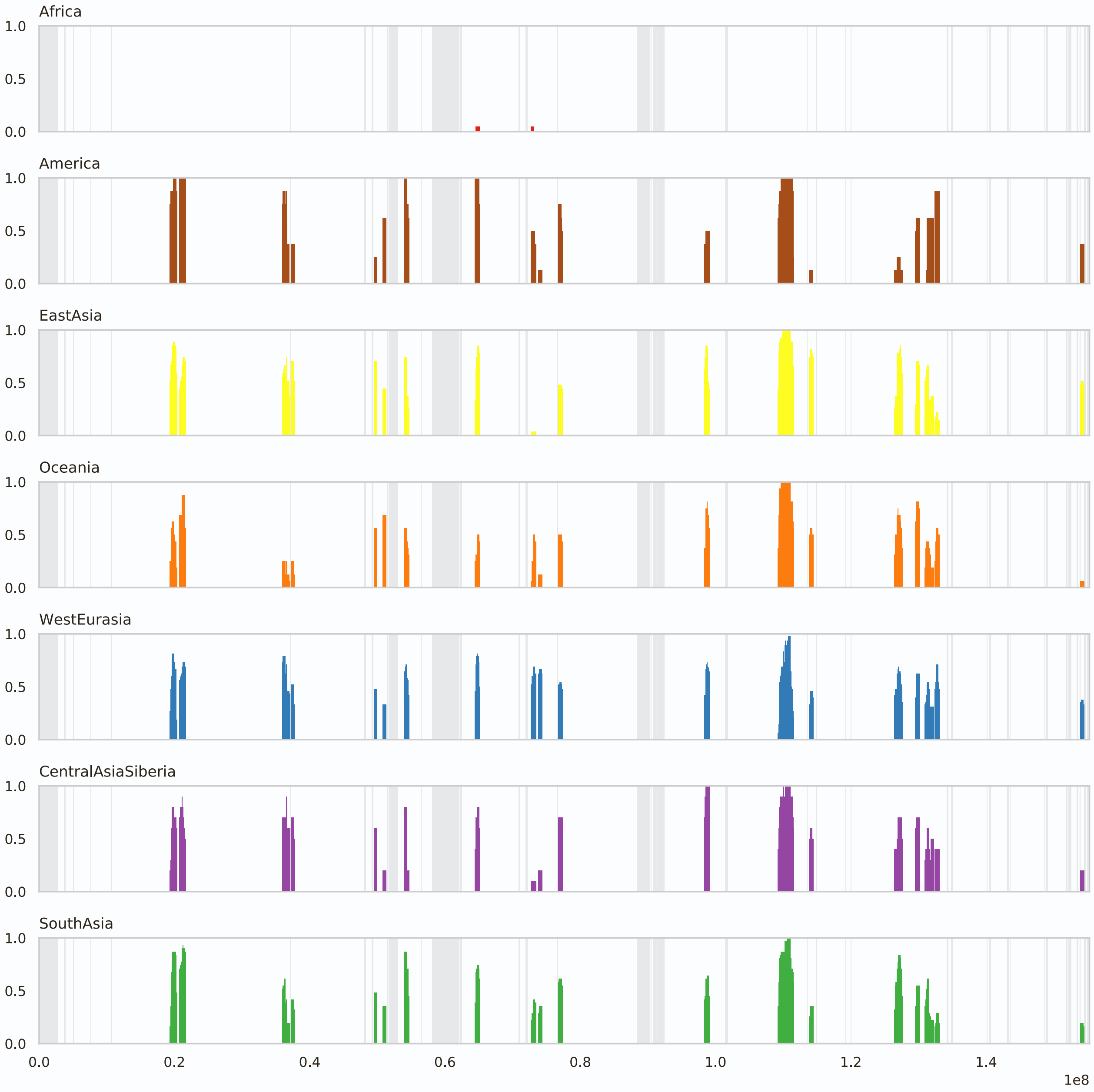

Figure S6: ECH frequencies by major geographical region (related to Figure 4). Proportion of individuals from each geographical region called as ECH in each 100kb window. Shaded regions represent regions with missing data.

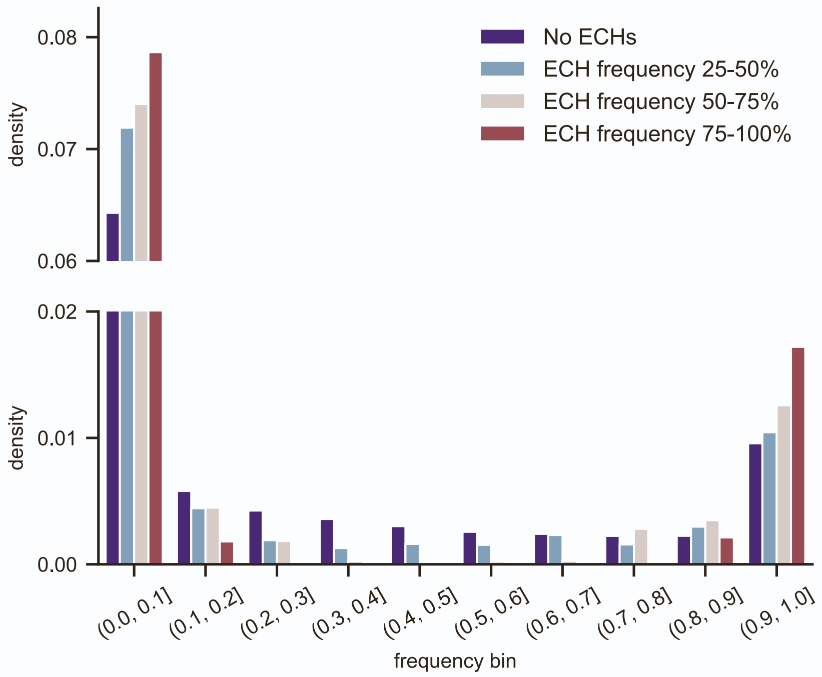

Figure S7 Site frequency spectrum by ECH frequency (related to Figure 2 and STAR methods): Unfolded site frequency spectrum in four bins of ECH frequency. The y-axis is cut to accommodate a large number of rare variants.

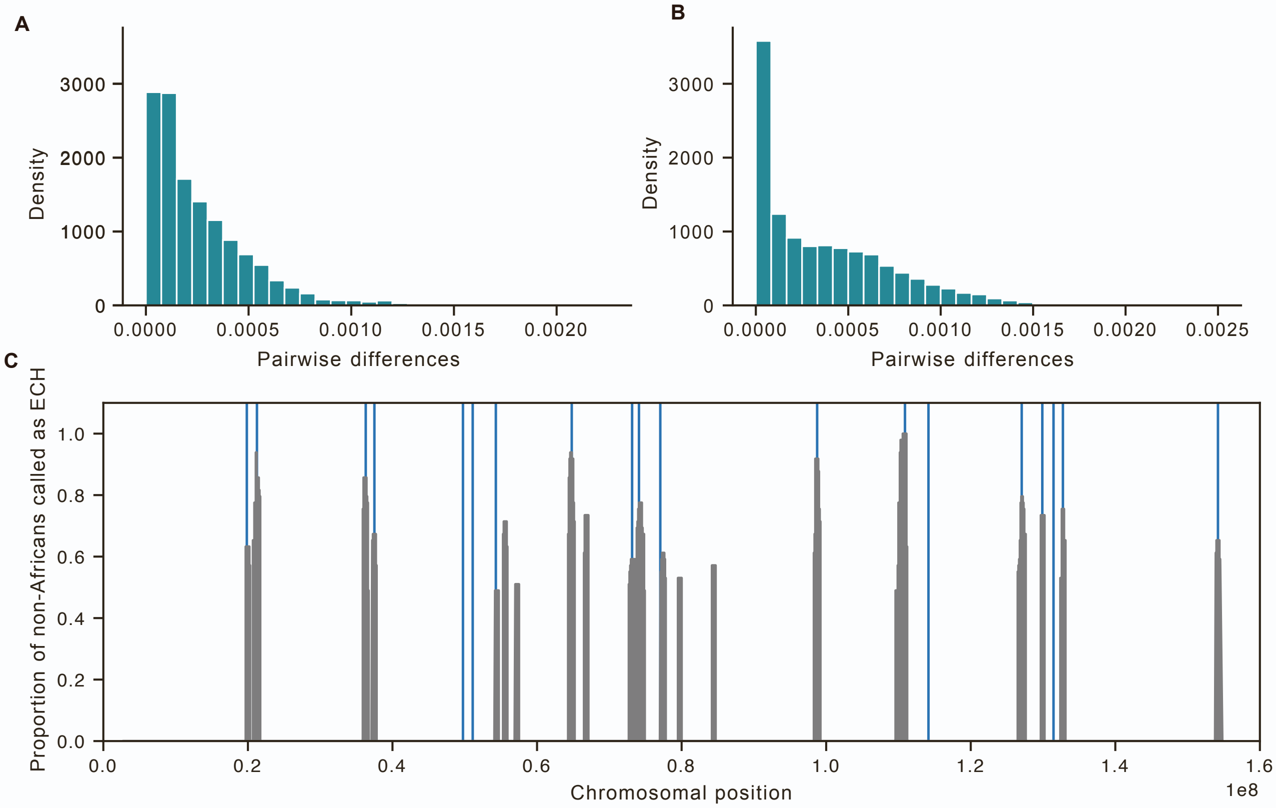

Figure S8: ECH inference in the CEU population (related to Figure 6 and STAR methods). (A) Distribution of mean pairwise differences in 200kb windows in neutral simulations of 49 CEU X chromosomes. (B) Distribution of mean pairwise differences in 200kb windows among the 49 male CEU samples analyzed. C) Frequency of ECHs along the X chromosome in the CEU population. Gray peaks show the proportion of 49 male haplotypes called as ECH in each 100kb window across the X chromosome. Vertical blue lines show the center of the corresponding peaks identified in the SGDP analysis.

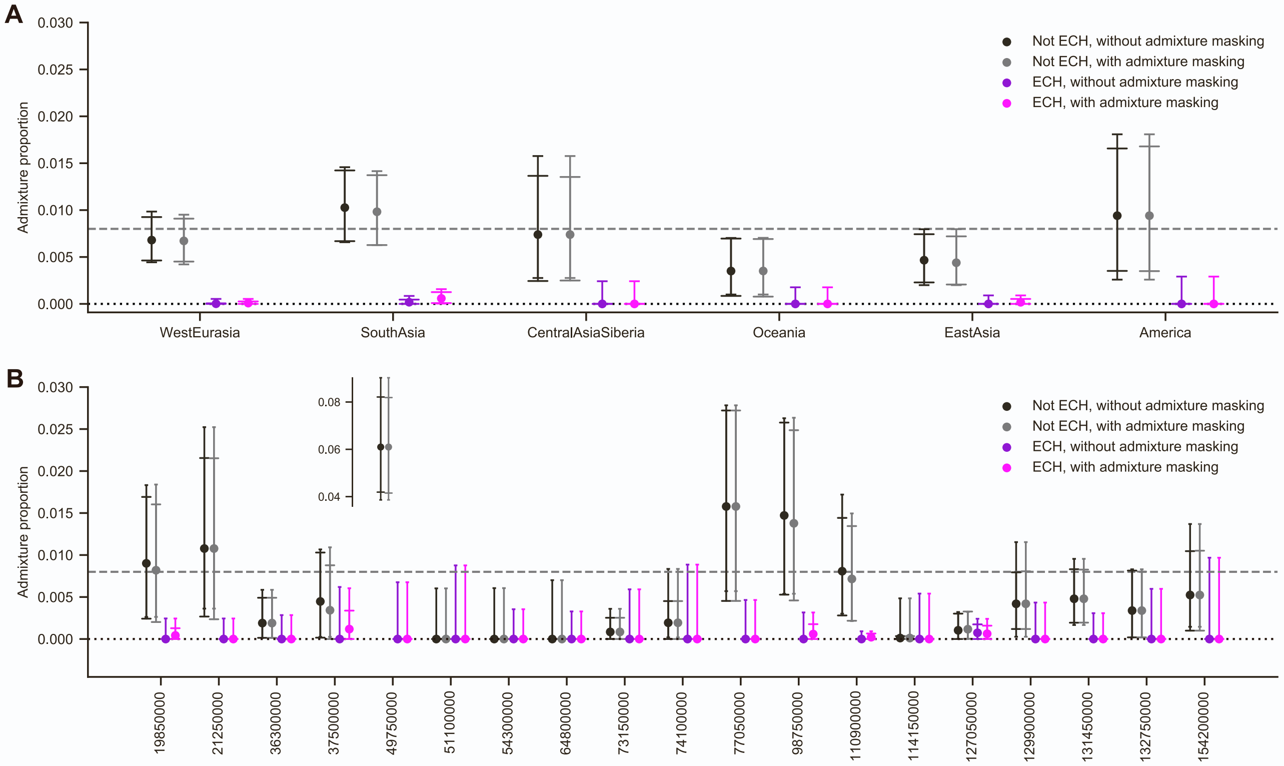

Figure S9: Admixture proportions in chromosomal regions of partial sweeps (related to Figure 7). Mean admixture proportions of ECHs (purple) and in remaining haplotypes (black) are computed separately for each 100kb window. Pink and gray represent inference of ECHs after masking archaic admixture (STAR methods). Error bars with wide caps designate the standard 95% confidence intervals obtained from 10,000 bootstrapping iterations. Error bars with narrow caps designate Jeffrey's binomial confidence interval, which better represents confidence when frequencies are zero or very low (in computing this interval, we assume that the sample size equals the number of 100kb sequence windows, STAR methods). The dotted line represents zero admixture. The dashed line shows the mean admixture proportion on the X chromosome. (A) Mean admixture proportions of haplotypes from each geographical region. (B) Admixture proportions at each individual ECH. X-axis labels represent chromosomal positions where each ECH has the highest frequency (peaks in Figure 4). For legibility, one outlier is shown separately.

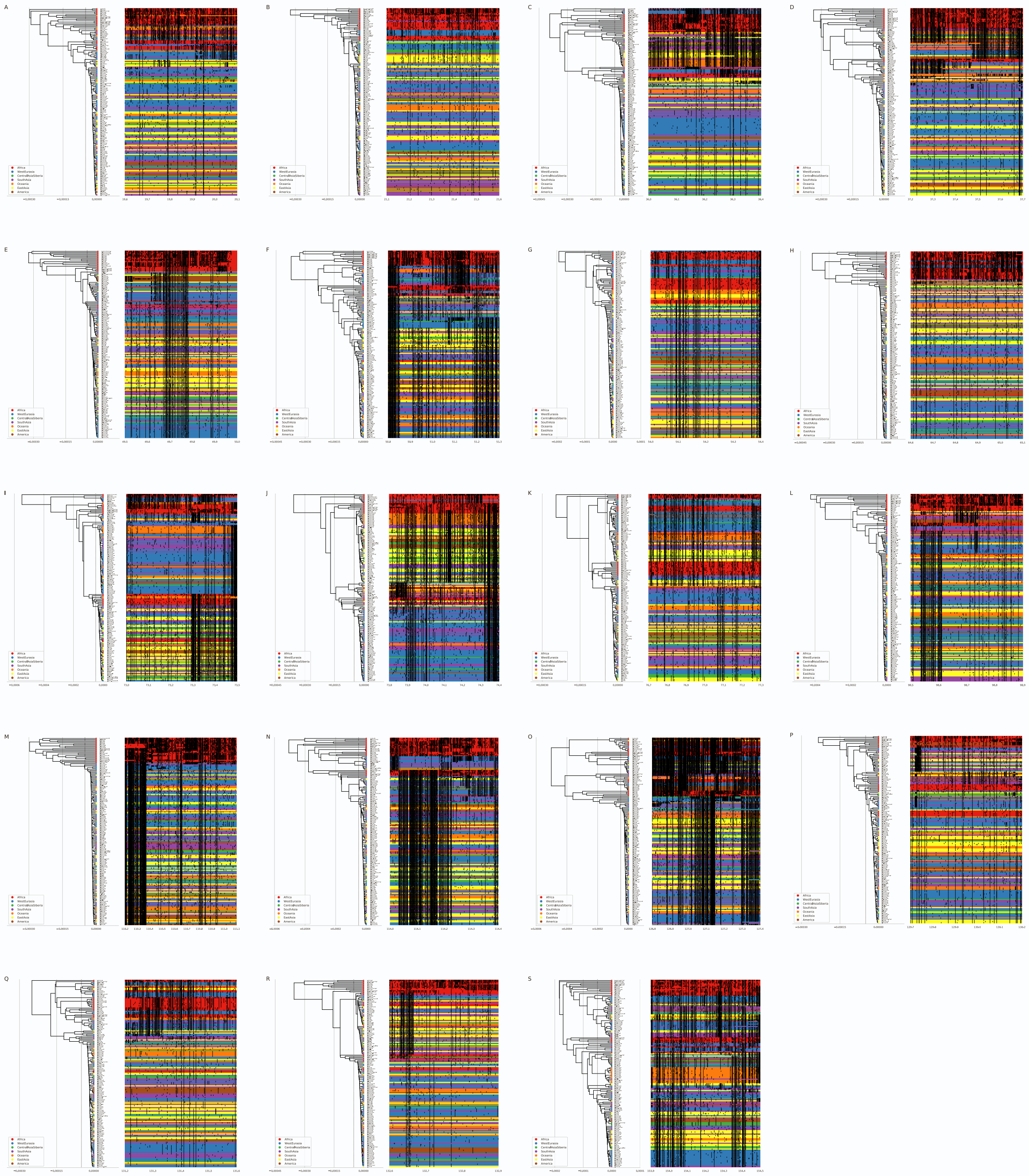

Figure S10: Haplotype visualizations of all ECHs (related to Figure 3 and Table S2). The callout label of each sub-figure refers to the Data S1 columns of table S1. The left side of each sub-figure is a UPGMA tree with the x-axis representing Jukes-Cantor corrected sequence distances between male haplotypes. The individual haplotypes are shown as horizontal lines on the right with the x-axis in hg19 megabase coordinates. Haplotypes are color-coded according to geographical region. The ancient Ust'-Ishim individual is marked with grey. Vertical black bars on each haplotype represents non-reference SNPs

Peak region where frequency is maximal			“90%-region” where frequency is at least 90% of its value in the peak region (see Figure S5)					“75%-region” where frequency is at least 75% of its value in the peak region (see Figure S5)					Data S1 Subplot label	Ust’-Ishim male shares ECH (✓/÷)
Start	End	Frequency	Start	End	Length	Lowest frequency in region	Frequency of ECHs spanning the region	Start	End	Length	Lowest frequency in region	Frequency of ECHs spanning the region		
198	199	0,83	196	201	5	0,77	0,71	195	203	8	0,65	0,54	A	÷
212	213	0,81	211	216	5	0,79	0,74	207	217	10	0,62	0,52	B	✓
362	364	0,66	360	364	4	0,64	0,64	359	367	8	0,51	0,31	C	✓
373	377	0,51	372	377	5	0,5	0,5	372	377	5	0,5	0,5	D	÷
495	500	0,53	495	500	5	0,53	0,53	495	500	5	0,53	0,53	E	÷
508	514	0,34	508	513	5	0,41	0,41	508	513	5	0,41	0,41	F	✓
542	544	0,76	540	544	4	0,74	0,74	539	544	5	0,61	0,61	G	÷
647	649	0,78	646	651	5	0,71	0,68	645	652	7	0,59	0,49	H	✓
730	733	0,43	730	735	5	0,39	0,39	728	735	7	0,34	0,3	I	✓
739	743	0,34	738	744	6	0,31	0,3	738	744	6	0,31	0,3	J	✓
769	772	0,56	767	773	6	0,55	0,54	767	774	7	0,5	0,49	K	✓
987	988	0,75	985	989	4	0,71	0,69	985	991	6	0,57	0,54	L	✓
1107	1111	0,99	1102	1111	9	0,95	0,91	1094	1112	18	0,75	0,53	M	✓
1140	1143	0,51	1140	1144	4	0,49	0,49	1138	1145	7	0,41	0,36	N	✓
1270	1271	0,71	1268	1274	6	0,65	0,57	1268	1275	7	0,6	0,49	O	✓
1297	1301	0,65	1297	1302	5	0,64	0,64	1297	1302	5	0,64	0,64	P	✓
1313	1316	0,58	1312	1316	4	0,56	0,56	1311	1316	5	0,46	0,46	Q	÷
1326	1329	0,49	1326	1329	3	0,49	0,49	1325	1331	6	0,37	0,27	R	÷
1540	1544	0,31	1539	1545	6	0,29	0,27	1539	1545	6	0,29	0,27	S	÷

Table S1: Coordinates and frequencies of ECHs (related to Figure 4). All coordinates are listed in 100kb units of the hg19 reference.

		Fraction of region Introgressed (%)		
Geographical region	Fragment type	X chromosome (with autosomal parameters for decoding)	X chromosome (with haploid parameters for decoding)	Autosome
America	total	0,144	0,49	2,1
	Denisova/Altai/Vindija	0,122	0,237	1,676
	Altai	0,011	0,024	0,233
	Vindija	0,082	0,109	0,961
	Denisova	0,029	0,027	0,135
	Ambiguous	0	0,078	0,348
	Sankararaman et al. 2016 [9]	0,26	0,26	1,42
CentralAsiaSiberia	total	0,178	0,417	2,264
	Denisova/Altai/Vindija	0,125	0,144	1,84
	Altai	0,021	0,056	0,229
	Vindija	0,032	0,027	1,083
	Denisova	0,021	0,006	0,15
	Ambiguous	0,051	0,055	0,378
	Sankararaman et al. 2016 [9]	0,23	0,23	1,45
EastAsia	total	0,288	0,709	2,292
	Denisova/Altai/Vindija	0,211	0,36	1,873
	Altai	0,014	0,058	0,246
	Vindija	0,151	0,211	1,097
	Denisova	0,009	0,009	0,168
	Ambiguous	0,036	0,083	0,363
	Sankararaman et al. 2016 [9]	0,32	0,32	1,45
Oceania	total	0,515	1,189	3,745
	Denisova/Altai/Vindija	0,414	0,745	3,342
	Altai	0,004	0,056	0,289
	Vindija	0,26	0,282	1,132
	Denisova	0,11	0,229	1,501
	Ambiguous	0,04	0,178	0,42
	Sankararaman et al. 2016 [9]	0,6	0,6	2,39
SouthAsia	total	0,273	0,71	2,081
	Denisova/Altai/Vindija	0,222	0,405	1,658
	Altai	0,017	0,048	0,226
	Vindija	0,152	0,228	0,881
	Denisova	0,014	0,026	0,179
	Ambiguous	0,039	0,103	0,371
	Sankararaman et al. 2016 [9]	0,41	0,41	1,25
WestEurasia	total	0,222	0,527	1,822
	Denisova/Altai/Vindija	0,139	0,202	1,406
	Altai	0,014	0,04	0,214
	Vindija	0,069	0,097	0,772
	Denisova	0,007	0,009	0,088
	Ambiguous	0,048	0,056	0,332
	Sankararaman et al. 2016 [9]	0,18	0,18	1,08

Table S2: Summary of archaic admixture inference (related to Figure 7)

Sample ID	Population	Region		Sample ID	Population	Region
S_BantuHerero-1	BantuHerero	Africa		S_Papuan-12	Papuan	Oceania
S_BantuHerero-2	BantuHerero	Africa		S_Papuan-2	Papuan	Oceania
S_BantuTswana-1	BantuTswana	Africa		S_Papuan-3	Papuan	Oceania
S_BantuTswana-2	BantuTswana	Africa		S_Papuan-4	Papuan	Oceania
S_Biaka-1	Biaka	Africa		S_Papuan-5	Papuan	Oceania
S_Biaka-2	Biaka	Africa		S_Papuan-6	Papuan	Oceania
B_Dinka-3	Dinka	Africa		S_Papuan-7	Papuan	Oceania
S_Dinka-1	Dinka	Africa		S_Papuan-8	Papuan	Oceania
S_Dinka-2	Dinka	Africa		S_Papuan-9	Papuan	Oceania
S_Esan-1	Esan	Africa		S_Balochi-1	Balochi	SouthAsia
B_Ju_hoan_North-4	Ju_hoan_North	Africa		S_Balochi-2	Balochi	SouthAsia
S_Ju_hoan_North-1	Ju_hoan_North	Africa		S_Bengali-1	Bengali	SouthAsia
S_Ju_hoan_North-2	Ju_hoan_North	Africa		S_Brahmin-1	Brahmin	SouthAsia
S_Ju_hoan_North-3	Ju_hoan_North	Africa		S_Brahmin-2	Brahmin	SouthAsia
B_Mandenka-3	Mandenka	Africa		S_Brahui-1	Brahui	SouthAsia
S_Mandenka-1	Mandenka	Africa		S_Brahui-2	Brahui	SouthAsia
B_Mbuti-4	Mbuti	Africa		S_Burusho-1	Burusho	SouthAsia
S_Mbuti-1	Mbuti	Africa		S_Hazara-1	Hazara	SouthAsia
S_Mbuti-3	Mbuti	Africa		S_Hazara-2	Hazara	SouthAsia
S_Mende-1	Mende	Africa		S_Irula-1	Irula	SouthAsia
B_Yoruba-3	Yoruba	Africa		S_Irula-2	Irula	SouthAsia
S_Yoruba-2	Yoruba	Africa		S_Kalash-1	Kalash	SouthAsia
S_Chane-1	Chane	America		S_Kapu-1	Kapu	SouthAsia
B_Karitiana-3	Karitiana	America		S_Kapu-2	Kapu	SouthAsia
S_Karitiana-1	Karitiana	America		S_Khonda_Dora-1	Khonda_Dora	SouthAsia
S_Mixtec-1	Mixtec	America		S_Kusunda-1	Kusunda	SouthAsia
S_Pima-1	Pima	America		S_Kusunda-2	Kusunda	SouthAsia
S_Quechua-2	Quechua	America		S_Madiga-1	Madiga	SouthAsia
S_Zapotec-1	Zapotec	America		S_Madiga-2	Madiga	SouthAsia
S_Zapotec-2	Zapotec	America		S_Makrani-1	Makrani	SouthAsia
S_Aleut-1	Aleut	CentralAsiaSiberia		S_Mala-2	Mala	SouthAsia
S_Altaian-1	Altaian	CentralAsiaSiberia		S_Mala-3	Mala	SouthAsia
S_Chukchi-1	Chukchi	CentralAsiaSiberia		S_Pathan-1	Pathan	SouthAsia
S_Eskimo_Chaplin-1	Eskimo_Chaplin	CentralAsiaSiberia		S_Punjabi-1	Punjabi	SouthAsia
S_Eskimo_Sireniki-1	Eskimo_Sireniki	CentralAsiaSiberia		S_Punjabi-2	Punjabi	SouthAsia
S_Even-2	Even	CentralAsiaSiberia		S_Relli-1	Relli	SouthAsia
S_Kyrgyz-1	Kyrgyz	CentralAsiaSiberia		S_Relli-2	Relli	SouthAsia
S_Mongola-1	Mongola	CentralAsiaSiberia		S_Sindhi-1	Sindhi	SouthAsia
S_Tlingit-1	Tlingit	CentralAsiaSiberia		S_Yadava-1	Yadava	SouthAsia
S_Yakut-2	Yakut	CentralAsiaSiberia		S_Yadava-2	Yadava	SouthAsia
S_Ami-1	Ami	EastAsia		S_Abkhasian-1	Abkhasian	WestEurasia
S_Ami-2	Ami	EastAsia		S_Abkhasian-2	Abkhasian	WestEurasia
S_Atayal-1	Atayal	EastAsia		S_Adygei-1	Adygei	WestEurasia
S_Burmese-1	Burmese	EastAsia		S_Armenian-1	Armenian	WestEurasia
S_Burmese-2	Burmese	EastAsia		S_Armenian-2	Armenian	WestEurasia
S_Cambodian-1	Cambodian	EastAsia		S_Basque-1	Basque	WestEurasia
B_Dai-4	Dai	EastAsia		S_BedouinB-1	BedouinB	WestEurasia
S_Dai-2	Dai	EastAsia		S_Bergamo-1	Bergamo	WestEurasia
B_Han-3	Han	EastAsia		S_Bulgarian-1	Bulgarian	WestEurasia
S_Han-2	Han	EastAsia		S_Bulgarian-2	Bulgarian	WestEurasia
S_Hezhen-1	Hezhen	EastAsia		S_Chechen-1	Chechen	WestEurasia
S_Japanese-1	Japanese	EastAsia		B_Crete-2	Crete	WestEurasia
S_Japanese-3	Japanese	EastAsia		S_Czech-2	Czech	WestEurasia
S_Kinh-2	Kinh	EastAsia		S_Druze-2	Druze	WestEurasia
S_Korean-1	Korean	EastAsia		S_English-1	English	WestEurasia
S_Lahu-2	Lahu	EastAsia		S_Estonian-1	Estonian	WestEurasia
S_Miao-1	Miao	EastAsia		S_Estonian-2	Estonian	WestEurasia
S_Naxi-1	Naxi	EastAsia		S_Finnish-3	Finnish	WestEurasia
S_Oroqen-1	Oroqen	EastAsia		B_French-3	French	WestEurasia
S_She-2	She	EastAsia		S_French-1	French	WestEurasia
S_Thai-1	Thai	EastAsia		S_Georgian-1	Georgian	WestEurasia
S_Tu-1	Tu	EastAsia		S_Georgian-2	Georgian	WestEurasia
S_Tujia-1	Tujia	EastAsia		S_Greek-1	Greek	WestEurasia
S_Uygur-2	Uygur	EastAsia		S_Greek-2	Greek	WestEurasia
S_Xibo-1	Xibo	EastAsia		S_Hungarian-2	Hungarian	WestEurasia
S_Xibo-2	Xibo	EastAsia		S_Iranian-1	Iranian	WestEurasia
S_Yi-1	Yi	EastAsia		S_Iranian-2	Iranian	WestEurasia
B_Australian-4	Australian	Oceania		S_Iraqi_Jew-2	Iraqi_Jew	WestEurasia
S_Hawaiian-1	Hawaiian	Oceania		S_Jordanian-1	Jordanian	WestEurasia
S_Igorot-2	Igorot	Oceania		S_Jordanian-2	Jordanian	WestEurasia
S_Maori-1	Maori	Oceania		S_Jordanian-3	Jordanian	WestEurasia
B_Papuan-15	Papuan	Oceania		S_Lezgin-2	Lezgin	WestEurasia
S_Papuan-10	Papuan	Oceania		S_North_Ossetian-1	North_Ossetian	WestEurasia
S_Papuan-11	Papuan	Oceania		S_North_Ossetian-2	North_Ossetian	WestEurasia

Table S3: Male samples included in the analysis (related to Figure 2)