Supplementary Text and Figures
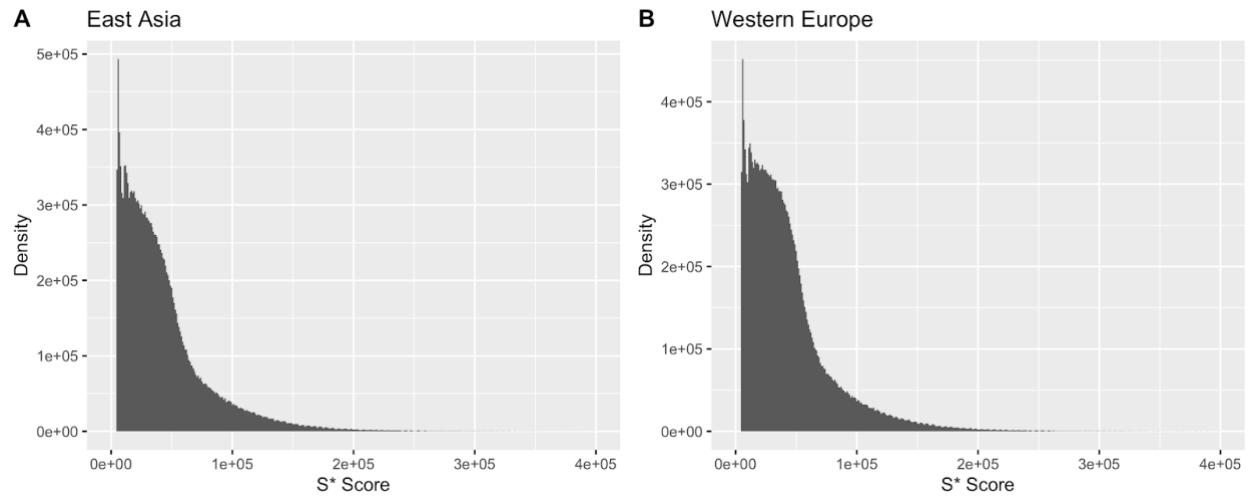

Figure S1. S* score Distributions. Density plots show the distributions of S*-scores of haplotypes detected in modern human genomes from East Asia (**A**) and Western Europe (**B**).
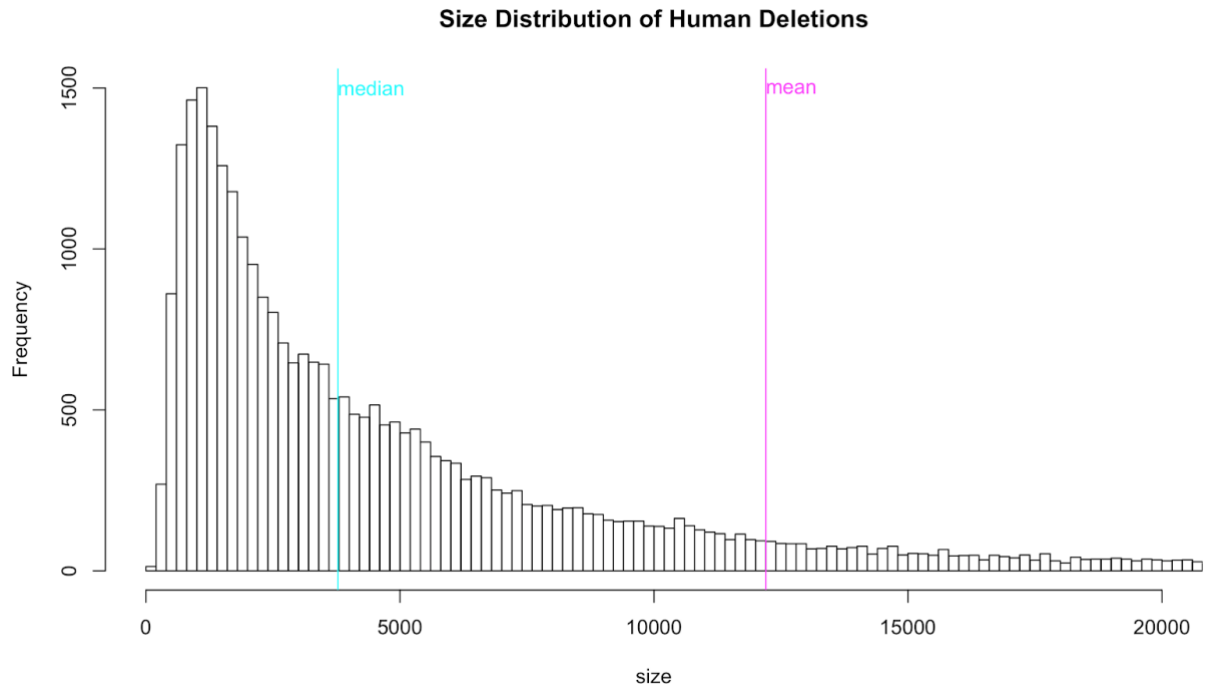

Figure S2. Size distribution of deletion variants detected for 2504 modern human genomes included in 1000 Genomes Project, Phase III. Median and mean of the distribution are shown with cyan and magenta vertical lines, respectively.
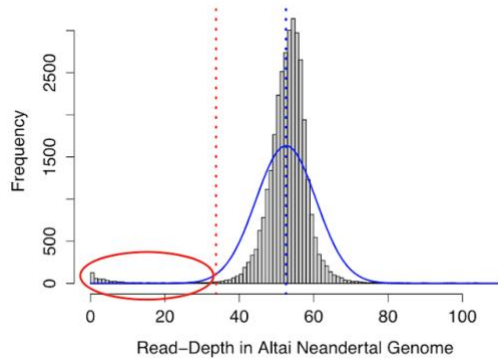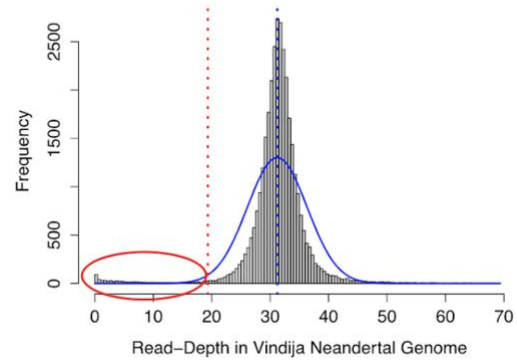
A

Altai Neanderthal

B

Vindija Neanderthal

Figure S3. Genotyping of high-quality Neanderthal genomes for large deletion polymorphisms detected for 2504 modern human genomes included in 1000 Genomes Project, Phase III. Raw read count distributions of high-quality Altai (**A**) and Vindija (**B**) Neanderthal genomes are shown with gray vertical bars. Normal distributions with means (dotted blue vertical line) and standard deviations of the raw read count distributions that were fitted on the distributions are shown with blue lines. The values marking the 0.01 quantiles of the normal distributions are shown with red vertical dotted lines. Regions for which there are lower numbers of reads for the two Neanderthal genomes than the 0.01 quantile values of the respective normal distributions are shown within red circles and categorized as deletions for the two Neanderthals.
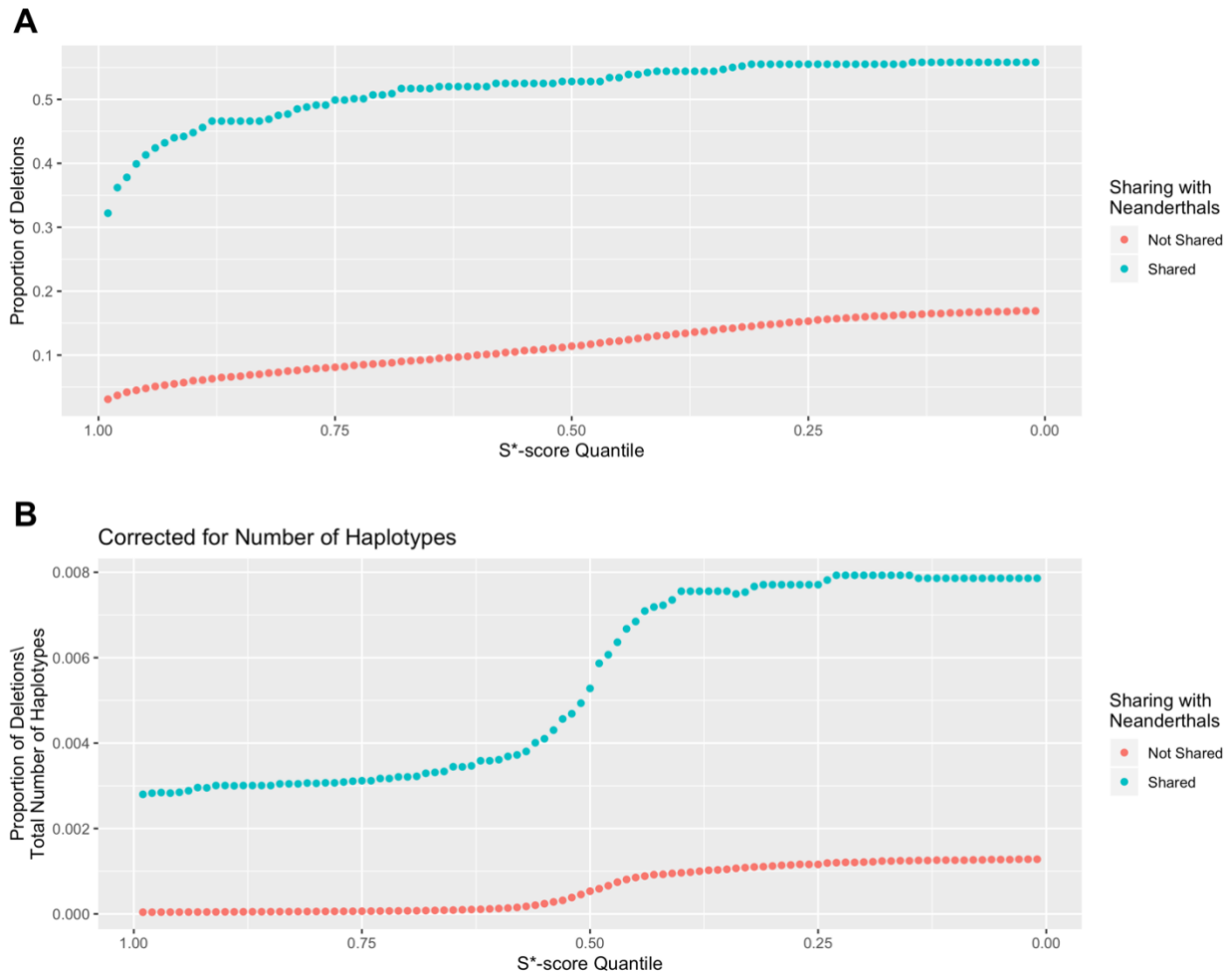
A.

B.

Corrected for Number of Haplotypes

Figure S4. S* Enrichment for Introgressed Deletions. Proportion of human polymorphic deletions found within the *S**-haplotypes detected for the same human genome carrying the deletion are shown on the y-axis. The x-axis shows the *S**-score quantiles. **A.** Percentage of human-polymorphic deletions found within *S**-haplotypes that are shared (red circles) and not-shared with Neanderthals (green circles). **B.** Percentage of human-polymorphic deletions corrected for the total number of S*-haplotypes carrying deletions shared and not-shared with Neanderthals.

Can one pulse of introgression from a lineage closer to Vindija Neanderthal explain the substantial number of introgressed haplotypes closer to the Altai Neanderthal genome?

We detected a substantial number of haplotypes closer to Altai Neanderthal than to Vindija Neanderthal. There are two possible scenarios to explain this. First, due to the random nature of the coalescent process and under the assumption of one pulse of introgression from a lineage closer to Vindija Neanderthal, it is plausible that these haplotypes find a common ancestor with the Altai Neanderthal genome before they find a common ancestor with the Vindija Neanderthal genome in the ancestral population of the two Neanderthals. Second, it is plausible that these haplotypes can be traced back to an independent introgression event from a Neanderthal population closer to Altai Neanderthal than it is to Vindija Neanderthal.
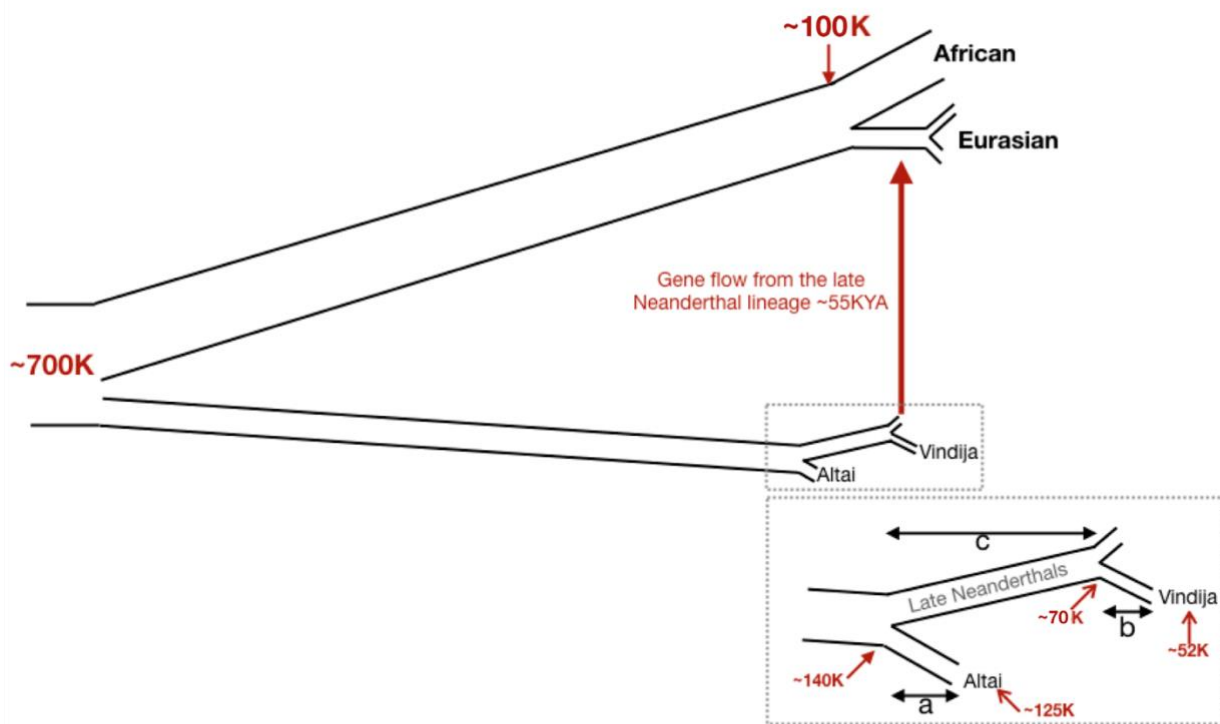
To distinguish between these scenarios, we first asked if a mainstream model of a single pulse of introgression into the ancestral Eurasian population from a branch closer to Vindija Neanderthal can explain our observations. Under this model, haplotypes that match more closely to Altai Neanderthal descend from a population that was ancestral to Vindija and Altai Neanderthals, rather than descending from a more recent introgression event from the Altai Neanderthal lineage. It then follows under this model that the number of haplotypes matching the Altai Neanderthal genome more closely than the Vindija Neanderthal genome should be a function of the total branch length of the common ancestor of the admixing Neanderthal population and late Neanderthals (c in **fig. S5A**). This expectation depends on the assumption that the effective population size in both Neanderthal lineages is similar, the mutation rate did not change for different branches of Neanderthals, and more importantly that the lengths b and a are negligible as compared to the branch length c.

The first and third assumptions are robust: the two Neanderthal lineages have similar effective population sizes (Prufer et al. 2014; Prufer et al. 2017) and that the branch length c is much greater than branch lengths a and b (**fig. S5A**). As to the second assumption, the mutation rate is not expected to vary for closely related populations. Specifically, under these assumptions, the branch length c is expected to be inversely proportional to the number of haplotypes more closely matching the Altai Neanderthal genome than the Vindija Neanderthal genome. That is, as the branch length c gets larger the probability of coalescence of introgressed Neanderthal variants with the Vindija Neanderthal genome gets larger. In other words, looking backwards in time, the probability that introgressed Neanderthal haplotypes find a common ancestor with the Altai Neanderthal genome before they find a common ancestor with the Vindija Neanderthal genome is an inverse exponential function of branch length c (**fig. S5A**). As it is a function of branch lengths and hence, the waiting times for the coalescence of different haplotypes at a common ancestor, and assuming that the mutation process occurs independently of the coalescence process and at a constant rate, we expect that the average nucleotide differences between the introgressed haplotypes and the two Neanderthal genomes should follow an exponential distribution. Similarly, we expect that the difference between these two π values, where the introgressed haplotypes are closer to the Altai Neanderthal genome (*i.e.,* $\pi_{I\text{-}V} - \pi_{I\text{-}A} > 0$) should follow an exponential distribution.
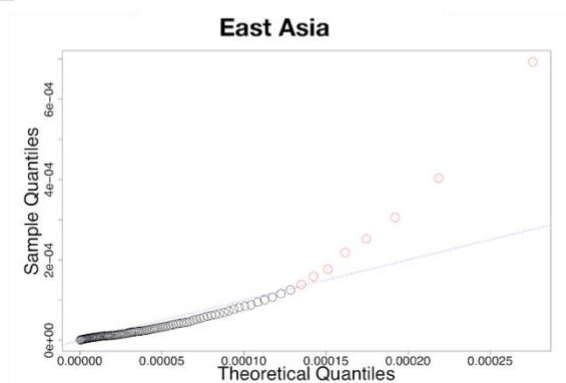
To test whether there is agreement between empirical data and the theoretical expectation of exponential distribution, we used average pairwise nucleotide differences between S*-significant putatively introgressed haplotypes and the two high quality Neanderthal genomes ($\pi_{I\text{-}V}$ for π between S*-haplotypes and the Vindija Neanderthal genome and $\pi_{I\text{-}A}$ for π between S*-haplotypes and the Altai Neanderthal genome). These are two variables shown in **fig 2A**. We fitted an exponential distribution on the difference between these two π values where $\pi_{I\text{-}V} - \pi_{I\text{-}A} > 0$ with a mean of the empirical distribution. We observed that more than expected number of haplotypes fall into >99% confidence intervals of exponential distributions for both Western Europe and East Asia: 127 and 136 haplotypes fall into >0.99 tail of the cumulative density distribution of the exponential distributions as opposed to theoretical expectations of 34 and 35 haplotypes for East Asians and Western Europeans, respectively. Quantile-quantile plots neatly show this deviation from the expectation for the higher quantiles representing the haplotypes substantially closer to the Altai Neanderthal genome than to the Vindija Neanderthal genome (*i.e.*, haplotypes for which the value of $\pi_{I\text{-}V} - \pi_{I\text{-}A}$ is the greatest) for both East Asia and Europe (**fig. S5B and C**). It is important to note here that this particular analysis depends on several assumptions with regards to Neanderthal demography, the distribution of nucleotide differences, etc. and thus should be verified when additional Neanderthal sequences are available.

None of the haplotypes in this category closely match the Denisovan genome. In fact, the average pairwise distances to the Denisovan genome is significantly greater for this category of haplotypes compared to all S*-significant haplotypes (Wilcoxon rank-sum test, W = 1876600, p-value = 1.05 X 10[-10]). This rules out the possibility of confounding haplotypes originating from the Neanderthal introgression with those finding the most recent common ancestor in the Denisovan lineage. The disparity between the expected and observed haplotype distances to the Altai and Vindija Neanderthal genomes cannot be directly explained by a single introgression event, and instead, invoke other potential scenarios including an additional introgression event from a branch related to Altai Neanderthal.

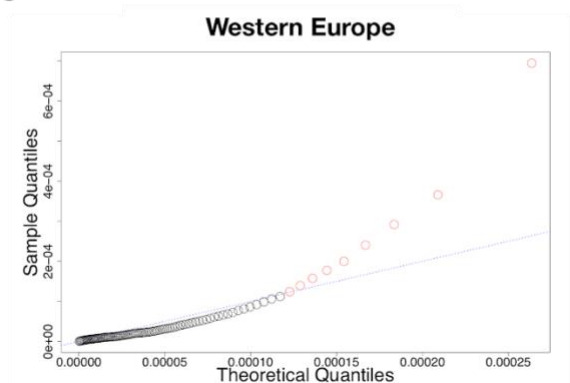**Figure S5. More Than Expected Number of Introgressed Haplotypes Closer to the Altai Neanderthal Genome Than to the Vindija Neanderthal Genome. A.** A base demographic model including one pulse of introgression from a Neanderthal lineage closer to Vindija Neanderthal into the common ancestor of Eurasians. Inset figure shows the population split times (human-Neanderthal split time: ~700 kya, Prüfer *et al.* 2014; African-Eurasian human populations

split time: ~100 kya, Schlebusch *et al.* 2017; Altai Neanderthal-Vindija Neanderthal split time: ~140 kya, Prüfer *et al.* 2017; population split times within late Neanderthals: ~70 kya, Hajdinjak *et al.* 2018), sampling times of Neanderthals (Altai Neanderthal sampling time: ~125 kya, Vindija Neanderthal sampling time: ~52 kya, Prüfer *et al.* 2017), time of the gene-flow from Neanderthals into Eurasian human populations (~55 kya, Fu *et al.* 2014) and branch lengths of different Neanderthal lineages. Admixture time as well as population split times are shown in red. **B.** and **C.** Exponential quantile-quantile (Q-Q) plots for π_{I-V}- π_{I-A}: the difference between the average pairwise nucleotide differences between the introgressed haplotypes and the Vindija Neanderthal genome (π_{I-V}) and the same statistic for the Altai Neanderthal genome (π_{I-A}). Q-Q plots for introgressed haplotypes detected for the East Asian and western Asian genomes are shown in B and C, respectively. Only haplotypes that are closer to the Altai Neanderthal genome than to the Vindija Neanderthal genome are included in data. Points above the *X=Y* line are shown in red. EA: East Asia, WE: Western Europe.

To further investigate the underlying scenarios, we then considered the potential effect of introgression from an early branch of anatomically modern humans into Neanderthal lineages (with an estimated time of ~200-300 kya, Hubisz *et al.* 2019). Specifically, an earlier study found evidence for introgression from humans specifically to the Altai Neanderthal lineage (Kuhlwilm *et al.* 2016). This may explain at least some of the *S**-significant haplotypes that are closer to Altai Neanderthal than to Vindija Neanderthal (Altai-matching haplotypes). Concordant with this hypothesis, we found that Altai-matching haplotypes are found in higher allele frequencies as compared to haplotypes closer to Vindija Neanderthal (Vindija-matching haplotypes) in sub-Saharan African populations (**fig. S6)**. Thus, at least some of the haplotypes that are shared between extant human genomes and the Altai Neanderthal genome can be traced back to the putative introgression event from an earlier, enigmatic population of modern human ancestors to Altai Neanderthals.

We then asked whether the increased allele frequency in African populations is also observed among haplotypes that are highly specifically matching to the Altai Neanderthal genome. These are the haplotypes for which π_{I-V} - π_{I-A} fall above the 99th quantile value of the exponential distribution fit with the mean of the empirical data. We found that this small number of haplotypes (n = 263) actually do not have different allele frequency in Yoruba population, and raises the possibility that they were introgressed from the Altai branch after Eurasian populations separate from sub-Saharan African populations. Nevertheless, the structure in the ancestral population of Altai and Vindija Neanderthals may also explain this observation. Unfortunately, the early sampling time of the Altai Neanderthal very close to the branching-time of the Altai and Vindija lineages diminishes the power to accurately distinguish between specific introgression from the Altai Neanderthal and structure in the Neanderthal populations. Indeed, we attempted to resolve this issue using an ABC framework but failed to find a well-supported solution (see **Methods**). Regardless, our haplotype level analysis is consistent with the recently accumulating evidence (Villanea and Schraiber 2019) that there may be multiple introgression events throughout Eurasia.
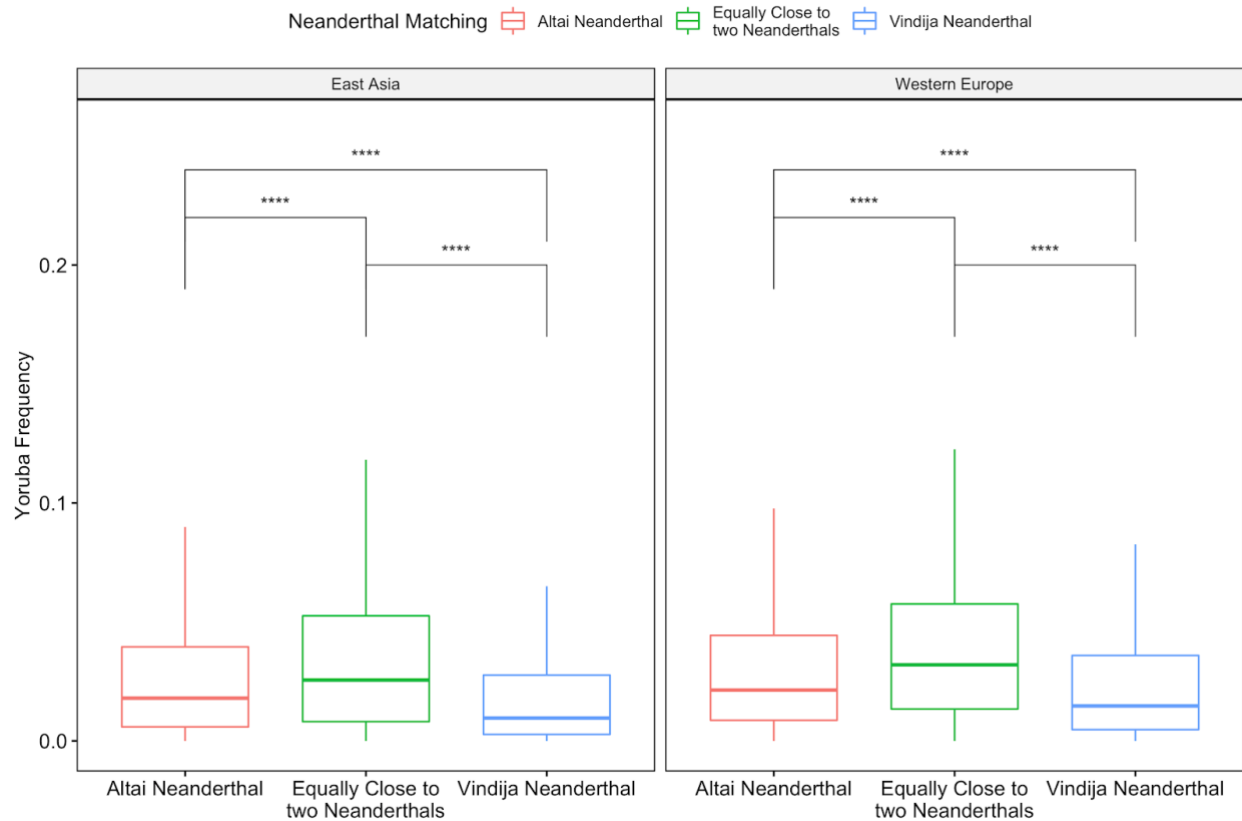
Figure S6. Yoruba Frequency of S*-significant haplotypes. Yoruba frequency of the S*-significant haplotypes that are closer to Altai Neanderthal and Vindija Neanderthal genomes are shown with the red and blue boxes, respectively. Yoruba frequency of the haplotypes that are equally close to the two Neanderthal genomes are shown with the green boxes. Significant differences between groups are shown with asterisks on the top of the boxes. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

REFERENCES

Fu Q., H. Li, P. Moorjani, F. Jay, S. M. Slepchenko, *et al.*, 2014 Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 514: 445–449.

Hubisz M. J., A. L. Williams, and A. Siepel, 2019 Mapping gene flow between ancient hominins through demography-aware inference of the ancestral recombination graph. bioRxiv 687368.

Hajdinjak M., Q. Fu, A. Hübner, M. Petr, F. Mafessoni, *et al.*, 2018 Reconstructing the genetic history of late Neanderthals. Nature 555: 652–656.

Kuhlwilm M., I. Gronau, M. J. Hubisz, C. de Filippo, J. Prado-Martinez, *et al.*, 2016 Ancient gene flow from early modern humans into Eastern Neanderthals. Nature 530: 429–433.

Prüfer K., F. Racimo, N. Patterson, F. Jay, S. Sankararaman, *et al.*, 2014 The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505: 43–49.

Prüfer K., C. de Filippo, S. Grote, F. Mafessoni, P. Korlević, *et al.*, 2017 A high-coverage Neandertal genome from Vindija Cave in Croatia. Science 358: 655–658.

Schlebusch C. M., H. Malmström, T. Günther, P. Sjödin, A. Coutinho, *et al.*, 2017 Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. Science 358: 652–655.

Villanea F. A., and J. G. Schraiber, 2019 Multiple episodes of interbreeding between Neanderthal and modern humans. Nat Ecol Evol 3: 39–44.