

# Supplementary information for Recovering signals of ghost archaic introgression in African populations

Arun Durvasula and Sriram Sankararaman

## Contents

|  |           |
|--|-----------|
| <b>S1 Current demographic models cannot explain the CSFS</b>   | <b>2</b>  |
| S1.1 Ancestral misidentification and genotyping error cannot explain the CSFS . . . . .                          | 3         |
| S1.2 Recurrent mutation cannot explain the CSFS . . . . .  | 3         |
| <b>S2 The CSFS cannot be explained by departures from panmixia in the ancestor of archaics and modern humans</b> | <b>4</b>  |
| <b>S3 Exploration of models of introgression into the ancestors of present-day Africans</b>                      | <b>4</b>  |
| S3.1 Model A.1 – Gene flow from modern human ancestor into the modern human ancestor before OOA . . . . .        | 5         |
| S3.2 Model A.2 – Gene flow from modern human ancestor into African populations after OOA . .                     | 6         |
| S3.3 Model B.1 – Gene flow from archaic branch into the modern human ancestor . . . . .                          | 6         |
| S3.4 Model B.2 – Gene flow from archaic branch into African populations . . . . .                                | 6         |
| S3.5 Model C.1 – Gene flow from a new unknown archaic branch into the modern human ancestor                      | 7         |
| S3.6 Model C.2 – Gene flow from a new unknown archaic branch into African populations . . . . .                  | 7         |
| <b>S4 Parameter exploration of model A</b>   | <b>8</b>  |
| S4.1 Approximate Bayesian computation with ancestral misidentification and genotyping error . .                  | 8         |
| S4.2 Genetic drift between archaics and modern humans constrains model A . . . . .                               | 9         |
| <b>S5 Estimating parameters for the best fit model of archaic introgression</b>                                  | <b>10</b> |
| S5.1 Method overview . . . . .   | 10        |
| S5.2 Parameter estimation . . . . .  | 10        |
| S5.3 Uncertainty in date estimates . . . . .   | 10        |
| <b>S6 Continuous migration versus a single pulse</b>   | <b>11</b> |
| <b>S7 Local ancestry inference</b>   | <b>11</b> |
| S7.1 Scaled divergence estimates . . . . .   | 12        |
| S7.2 Coalescent-based age estimates of introgressed segments . . . . .   | 12        |
| S7.3 Adaptive archaic introgression . . . . .  | 13        |
| <b>S8 Extended Discussion</b>  | <b>13</b> |
| S8.1 Results from the Luhya . . . . .  | 13        |
| S8.2 Pre- versus post- Out-of-Africa . . . . .   | 13        |
| <b>S9 ms command lines</b>   | <b>14</b> |
| <b>A The CSFS is uniform under structure in the archaic population</b>   | <b>16</b> |
| <b>B The CSFS is symmetric under Model A</b>   | <b>20</b> |

## S1 Current demographic models cannot explain the CSFS

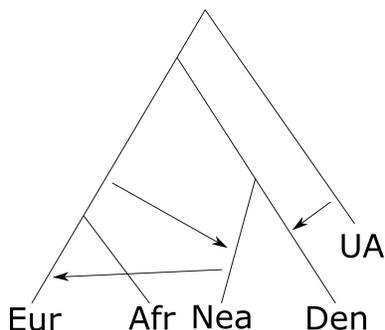


Figure S1: Demographic model from Prüfer et al. [2017] (15). See Section S1 for details.

We simulated genome sequence data according to the demographic history from Prüfer et al. [2017] (15) (we refer to this model as the ‘base model’, Figure S1), which contains several major admixture events among modern human populations, Neanderthals, and Denisovans and was fit to explain key population genetic statistics that relate these populations. We use a mutation rate (per basepair) of  $1.2 \times 10^{-8}$  and a recombination rate (per basepair) of  $1.3 \times 10^{-8}$  and simulate 3,000 1 Mb regions for a total of 3GB of simulated sequence.

The base model includes introgression from an unknown archaic hominin into the Denisovan population (6%, 250Ky B.P.) [Prüfer et al., 2014] (18), introgression from modern humans into Neanderthals (5%, 329Ky B.P.) [Kuhlwilm et al., 2016] (6) as well as an introgression event from Neanderthals into the ancestors of present-day non-Africans (3%, 50Ky B.P.). Population sizes are set to 10,000 individuals for all lineages except Neanderthal and Denisovan, which have population sizes of 1,000 individuals. The split times for pairs of populations come from Prüfer et al. [2017] (15) and are as follows:

- Unknown archaic - modern human ancestor: 910Ky B.P.
- Archaic-Modern human: 550Ky B.P.
- Neanderthal-Denisovan: 415Ky B.P.
- Africa-Out of Africa: 62.5Ky B.P.

In addition to the above introgression events, we also included introgression from the Neanderthal population into the African lineage 43Ky B.P. which was shown by Prüfer et al. [2017] (15) to lead to improved fits. We used introgression percentages of 1% and 0.015% (Figure S16ab). We see that 1% introgression fraction produces an increase in the counts of low frequency derived alleles in the CSFS and leads to a larger excess in Neanderthals relative to the Denisovan. This result is also seen in the 0.015% case, which is the best fitting parameter as observed by Prüfer et al. [2017] (15). However, we find that introgression from the ancestors of Neanderthals into the ancestors of Africans is insufficient to produce a U-shaped CSFS (Figure S16ab).

We also performed an additional simulation experiment where we modeled realistic variation in mutation and recombination rates along the genome (while fixing the demographic model to the base model). Mutation rates were calculated by estimating Watterson’s  $\theta$  [Watterson, 1975] (43) from the number of segregating sites within 1 Mb windows across 50 randomly sampled YRI genomes from the 1000 Genomes Project Phase 3 release and calculating the mutation rate:  $\mu = \theta/4N_eL$  where we set  $N_e = 10,000$  and  $L = 1$  Mb. Recombination rates were estimated from the combined, sex-averaged HapMap recombination map [International HapMap Consortium et al., 2007] (26) (Figure S12).

Next, we simulated data from the model inferred by Gravel et al. [2011] (17). We restricted our analysis to the African population in this model which we augmented with the archaic populations described above with demographic parameters fixed to those inferred in Prüfer et al. [2017] (15). The CSFS under this model is also uniform, suggesting the model is insufficient to explain the data (Figure S16c).

In addition, we also explored a model with migration between the Out-of-Africa population and the African population over the last 20Ky B.P. at a rate of 0.0001 migrants per generation as has been suggested by a number of demographic models [Gravel et al., 2011, Harris and Nielsen, 2013, Petr et al., 2019] (17, 44, 45). We see that such a model does produce a slight increase in the number of derived alleles at a high frequency but is insufficient to explain the signal observed in the data (Figure S16d).

Finally, we simulated the demographic model inferred in Hsieh et al. [2016] (19) relating the Yoruba and the Central African Pygmy (Baka and Biaka) populations. This model has gene flow between the three populations dating back to  $\simeq 40$ Ky B.P. Importantly, it is plausible that the gene flow over this period of time could produce the U-shape we see in the data. However, simulations indicate that the parameters inferred from Hsieh et al. [2016] (19) are insufficient to produce the U-shape of the spectrum we observe in the data ( $KS\ p < 2 \times 10^{-16}$ ) (Figure S13).

### S1.1 Ancestral misidentification and genotyping error cannot explain the CSFS

One possibility is that CSFS can be explained by a combination of demographic history and errors, in particular errors in identifying the ancestral allele (ancestral misidentification/mispolarization) and genotyping error in the archaic genome that we condition on (Neanderthal or Denisovan). To explore this possibility, we simulated data under the model from Prüfer et al 2017 and added ancestral misidentification ( $e_1$ ) and genotyping error ( $e_2$ ) over a grid of values  $e_1, e_2 \in \{0\%, 3\%, 15\%, 30\%\}$ . In Figure S11, we begin to observe a U-shaped CSFS with 3% ancestral misidentification and 0% genotyping error (Figure S11e). The best fit CSFS obtained when combining errors with the base model in Section S1 requires 15% ancestral misidentification and 3% genotyping error. In Paten *et al.* 2008, the authors estimate  $< 1\%$  error in the EPO alignment for the common ancestor of apes and monkeys and that the overall error rate for all ancestors is 3.2%, suggesting that the required ancestral misidentification rate is too large to explain the data [Paten et al., 2008] (16). In addition, Meyer *et al.* 2012 estimate the sequencing error in the high coverage Denisovan genome as 0.147% per read and a contamination rate of 0.224%, resulting in an overall error rate of 0.371% [Meyer et al., 2012] (4). In order to erroneously call a genotype at a coverage of 30x (the reported mean coverage for the Denisovan genome), 15 reads or more must support the alternate allele. Therefore, we estimate the genotyping error rate ( $e_2$ ) using the binomial distribution as:

$$e_2 = \sum_{i=15}^{30} \binom{30}{i} 0.00371^i (1 - 0.00371)^{30-i} = 5.1 \times 10^{-29}$$

The above computation of  $e_2$  relies on several simplifying assumptions such as independence of errors across reads. This is a reasonable assumption for sequencing errors but is less reasonable for contamination.

For the Vindija Neanderthal, Prüfer [2018] (46) estimates genotyping discordance rates by subsampling the Vindija genome and comparing the frequency of genotype discordance. They subsample read depths  $r \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12.5, 15, 17.5, 20, 25\}$ , where  $r$  is the average fold coverage, and obtain a genotype discordance rate for each subsample. We use this data to estimate the probability of genotyping error,  $P(g)$  as follows.

$$P(g) = \sum_{i \in r} P(g|i)P(i) = 4.3 \times 10^{-7}$$

In addition to this estimate, Prüfer et al. [2017] (15) report a nuclear contamination genotyping error rate of 0.6%. This represents the probability that the reported genotype comes from a modern human sample rather than the Neanderthal sample. This is an upper bound on the genotyping error rate (the probability that the called archaic genotype differs from the true archaic genotype). These numbers suggest the plausible error rates are lower than what would be required to produce the observed CSFS.

Finally, we note that the best fit model that includes ancestral misidentification and genotyping error is substantially worse than the best fitting model that we explore in Section S3 (model C;  $\Delta\mathcal{L}\mathcal{L} = -1, 416$ ).

### S1.2 Recurrent mutation cannot explain the CSFS

In principle, it is possible that recurrent mutations, or mutations that occur at the same site multiple times, could create a U-shaped CSFS. We explore that possibility here.

We simulated data using 1GB of sequence SLiM [Haller and Messer, 2017] (47) allowing for recurrent mutations under the base demographic model described in Section S1 using a mutation rate of  $\mu = 1.25 \times 10^{-8}$  and recombination rate  $r = 1 \times 10^{-8}$ . We plot the CSFS conditioned on a simulated Neanderthal genome in Figure S43. We find that the resulting spectrum is uniform, suggesting that recurrent mutation cannot explain the observed CSFS.

## S2 The CSFS cannot be explained by departures from panmixia in the ancestor of archaics and modern humans

One of the assumptions for the CSFS to be uniformly distributed across allele frequencies is the requirement of mutation-drift equilibrium at the archaic-modern human ancestor. In the absence of detailed demographic models of the ancestral population, we explored whether departures from equilibrium in the archaic-modern human ancestor can produce the signals we see in the data using the base model from Prüfer et al. [2017] (15) described in Section S1. We simulated a 5-fold expansion in the ancestor and see that the CSFS slightly increases in the low frequency bins relative to high frequency bins (Figure S17a). We simulated a 5-fold bottleneck in the ancestor and observed the opposite effect: a slight positive slope as the derived allele frequency increases (Figure S17b).

In addition, we turned our attention to models of population structure in the ancestor of archaic and modern humans. A model in which the ancestral population split 840 Ky B.P. into two sub-population followed by subsequent merging into one population 550Ky B.P. produces a uniform spectrum, as does a model with a split at 840Ky B.P. and continuous migration ( $5 \times 10^{-5}$  migrants per generation) until 550Ky B.P. (Figure S17cd).

Finally, we explored previous models of ancestral structure from the literature. In particular, we simulated data under models proposed by Yang et al. [2012] (14) and Sankararaman et al. [2012] (38) (Figure S18). Neither model could account for the shape of the CSFS although the model proposed in Yang et al. [2012] (14) does produce a slight U-shape. Overall, departures from mutation-drift equilibrium in the archaic-modern human ancestor are insufficient to produce the signals we see in the data.

## S3 Exploration of models of introgression into the ancestors of present-day Africans

We simulated data assuming a model of introgression into the ancestors of present-day Africans on top of the base model simulated in Section S1. We consider three sources for the gene flow: A) a population that diverged from the ancestors of modern humans after their split from the ancestors of archaic humans, B) a population that diverged from the ancestors of archaic (Neanderthal/Denisovan) humans after their split from the ancestors of modern humans, C) a population that diverged from the ancestors of archaic and modern humans prior to the archaic-modern human split. For each of these models, we considered two variants depending on the time of introgression: 1. introgression occurred in the modern human ancestor prior to the out-of-Africa (OOA) event, 2. introgression occurred in the African population after the OOA event.

We use ms [Hudson, 2002] (42) to simulate 3,000 unlinked 1 megabase (MB) regions (resulting in 3 gigabases (GB) of sequence). We provide representative ms command lines in Section S9. We used the following parameter values: ancestral effective population size  $N_e = 10,000$ , mutation rate  $\mu = 1.2 \times 10^{-8}$ /bp, region length  $L = 1,000,000$  base pairs, recombination rate  $r = 1.3 \times 10^{-8}$ . When scaling generations to years, we use a generation time of 29 years per generation as in Fenner [2005] (48). Unless otherwise stated, all population sizes are the same as in the base model (See Section S1).

We also perform simulations for simple versions of the models presented below where we only simulate the African, Neanderthal, and Denisovan branches (removing the Non-African and unknown archaic branches) without the following gene flow events: from modern humans into the archaic branch, from the unknown archaic into the Denisovan branch, from the Neanderthal into the non-African branch. We present these simulations in Figures S20, S22, S24, S26, S28, and S30. We see that these models are nearly identical to

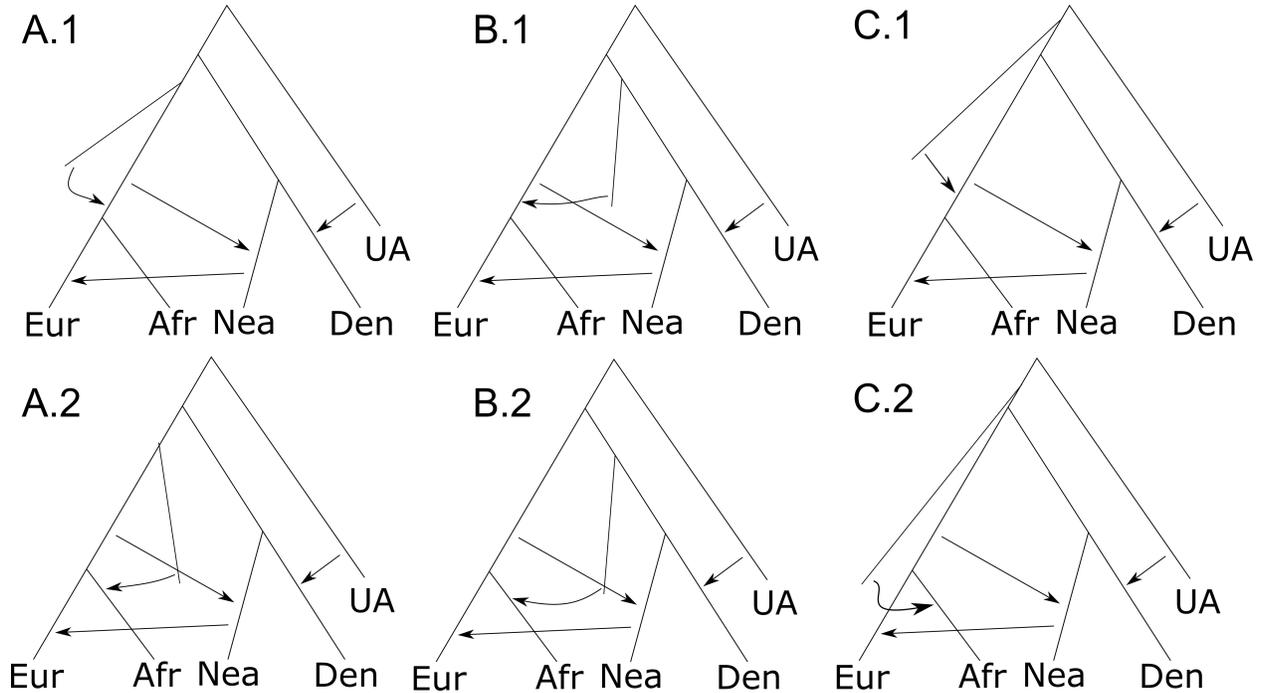


Figure S2: Demographic model topologies for introgression into the ancestors of present-day Africans

the full models. We also report the fits of the models in Table S5. We describe how these fits are calculated in the Methods of the main text (‘Model comparison’ and ‘Goodness of fit’).

In addition to the simulations with fixed parameter values described below, we also performed an exploration of the parameter space using approximate Bayesian computation (ABC). We used the proportional CSFS in YRI conditioned on the Vindija Neanderthal as a summary statistic and fit values for the split time of the introgressing lineage ( $t_s$ ), the effective population size of the introgressing lineage ( $N_e$ ), the admixture time ( $t_a$ ), and the admixture fraction  $\alpha$ . We drew these parameters from uniform distributions (details of the parameter values are described in the respective sections), simulated 1GB of sequence, and computed the resulting CSFS. We repeated this procedure 10,000 times for each model and obtained posterior distributions using the R package ‘abc’ [Csilléry et al., 2012] (41). We used the ‘neuralnet’ method for non-linear regression (number of networks= 50, number of hidden units= 7) and set the tolerance to 0.1. Then, we simulated 3GB of data from the best fit values and calculated a goodness-of-fit  $p$ -value (as described in the Methods section of the main text (‘Goodness of fit’)). We report times in terms of generations. Simulations of the best fitting models are shown in Figure S31.

### S3.1 Model A.1 – Gene flow from modern human ancestor into the modern human ancestor before OOA

In model A.1, the ancestor of modern humans splits into two populations which later admix. The split occurs at 174Ky B.P. while the admixture occurs at 63.8Ky B.P. (Figure S2). We see that as the admixture fraction increases, there is an increase in allele counts at the low and high frequency bins of the CSFS, which is potentially consistent with the U-shaped spectrum we observe in the data. However, this model is unable to capture the curvature of the spectrum even at large introgression fractions (Figure S19). Further, we show in Mathematical Appendix B that the CSFS under this model is symmetric.

To further explore if specific parameters of model A.1 could explain the CSFS, we performed ABC where

we drew parameters relevant to this demographic model from the following distributions:

$$\begin{aligned} ta &\sim \text{Uniform}(0.06, 0.19) \\ \alpha &\sim \text{Uniform}(0.0, 0.2) \\ ts &\sim \text{Uniform}(0.2, 0.47) \\ ne &\sim \text{Uniform}(0.1, 3) \end{aligned}$$

Based on the ABC, we found the posterior mean of the admixture time was 2,600 generations B.P. (95% HPD: 2,200-3,300), the admixture fraction was 0.05 (95% HPD: 0.03-0.08), the split time was 17,000 generations B.P. (95% HPD: 15,000-19,100), and the effective population size was 24,000 (95% HPD: 20,000-27,000). Simulations from the best fit values resulted in a KS  $p < 2 \times 10^{-16}$ .

### S3.2 Model A.2 – Gene flow from modern human ancestor into African populations after OOA

Model A.2 differs from model A.1 in the time of introgression being more recent, *i.e.*, 50 Ky B.P. (Figure S2). Similar to model A.1, we see an increase in allele counts at the low and high frequency bins. However, this model is still unable to capture the curvature of the spectrum (Figure S21).

For parameter exploration, we drew parameters from the following distributions:

$$\begin{aligned} ta &\sim \text{Uniform}(0.01, 0.053) \\ \alpha &\sim \text{Uniform}(0.0, 0.2) \\ ts &\sim \text{Uniform}(0.1, 0.47) \\ ne &\sim \text{Uniform}(0.1, 3) \end{aligned}$$

We found the admixture time was 800 generations B.P. (95% HPD: 500-1200), the admixture fraction was 0.07 (95% HPD: 0.03-0.10), the split time was 17,000 generations B.P. (95% HPD: 14,000-19,000), and the effective population size was 20,000 (95% HPD: 12,500-26,000). Simulations from the best fit values resulted in a KS  $p$ -value of  $3.3 \times 10^{-15}$ .

### S3.3 Model B.1 – Gene flow from archaic branch into the modern human ancestor

In model B.1, we simulate gene flow from the ancestors of the archaic branch containing Neanderthals and Denisovans (Figure S2). This lineage splits off from the ancestors of Neanderthals and Denisovans 464Ky B.P. and introgresses 200Ky B.P. into the modern human ancestor (Figure S23).

For parameter exploration, we drew parameters from the following distributions:

$$\begin{aligned} ta &\sim \text{Uniform}(0.06, 0.34) \\ \alpha &\sim \text{Uniform}(0.0, 0.2) \\ ts &\sim \text{Uniform}(0.35, 0.47) \\ ne &\sim \text{Uniform}(0.1, 3) \end{aligned}$$

We found the admixture time was 3,300 generations B.P. (95% HPD: 3,200-3,400), the admixture fraction was .08 (95% HPD: 0.07-0.08), the split time was 18,000 generations B.P. (95% HPD: 18,000-18,100), and the effective population size was 40,100 (95% HPD: 40,000-40,100). Simulations from the best fit values resulted in a KS  $p$ -value of  $9.8 \times 10^{-8}$ . The split time effectively reduces this model to a trifurcation.

### S3.4 Model B.2 – Gene flow from archaic branch into African populations

Model B.2 is identical to model B.1 except that the time of introgression into the African population branch is fairly recent, *i.e.*, 50 kY B.P.(Figure S2, S25).

For parameter exploration, we drew parameters from the following distributions:

$$\begin{aligned}
 ta &\sim \text{Uniform}(0.01, 0.053) \\
 \alpha &\sim \text{Uniform}(0.0, 0.2) \\
 ts &\sim \text{Uniform}(0.35, 0.47) \\
 ne &\sim \text{Uniform}(0.1, 3)
 \end{aligned}$$

We found the admixture time was 700 generations B.P. (95% HPD: 680-710), the admixture fraction was 0.11 (95% HPD: 0.11-0.11), the split time was 18,800 (95% HPD: 18,700-18,900), and the effective population size was 25,300 (95% HPD: 25,100-25,600). Simulations from the best fit values resulted in a KS  $p$ -value of  $5.6 \times 10^{-6}$ . The split time effectively reduces this model to a trifurcation.

### S3.5 Model C.1 – Gene flow from a new unknown archaic branch into the modern human ancestor

In model C.1, we simulate gene flow from a branch population separates from the modern human and archaic common ancestor prior to their split from each other (696Ky B.P.) and then introgresses into the modern-human ancestor (63.8Ky B.P; Figure S2). We find that this model can produce the U-shaped spectrum we observe in the data, particularly at introgression fractions between 0.05-0.1 (Figure S27).

For parameter exploration, we drew parameters from the following distributions:

$$\begin{aligned}
 ta &\sim \text{Uniform}(0.06, 0.34) \\
 \alpha &\sim \text{Uniform}(0.0, 0.2) \\
 ts &\sim \text{Uniform}(0.48, 0.77) \\
 ne &\sim \text{Uniform}(0.1, 3)
 \end{aligned}$$

We found the admixture time was 2,700 generations B.P. (95% HPD: 2,200-3,700), the admixture fraction was 0.05 (95% HPD: 0.03-0.07), the split time was 26,000 generation B.P. (95% HPD: 24,700-27,000), and the effective population size was 26,000 (95% HPD: 22,000-29,000). Simulations from the best fit values resulted in a KS  $p$ -value of 0.09.

### S3.6 Model C.2 – Gene flow from a new unknown archaic branch into African populations

Model C.2 modifies model C.1 so that introgression into the African population is recent (50Ky B.P; Figure S2). We find that this model can also produce the U-shaped spectrum we observe in the data, again most notably at introgression fractions between 0.05-0.1 (Figure S29).

For parameter exploration, we drew parameters from the following distributions:

$$\begin{aligned}
 ta &\sim \text{Uniform}(0.01, 0.053) \\
 \alpha &\sim \text{Uniform}(0.0, 0.2) \\
 ts &\sim \text{Uniform}(0.48, 0.77) \\
 ne &\sim \text{Uniform}(0.1, 3)
 \end{aligned}$$

We found the admixture time was 1,360 generations B.P. (95% HPD: 1,300-1,400), the admixture fraction was 0.06 (95% HPD: 0.05-0.06), the split time was 24,300 generations B.P. (95% HPD: 24,000-24,600), and the effective population size was 31,000 (95% HPD: 30,500-31,300). Simulations from the best fit values resulted in a KS  $p$ -value of  $8.2 \times 10^{-6}$ .

## S4 Parameter exploration of model A

### S4.1 Approximate Bayesian computation with ancestral misidentification and genotyping error

We performed additional simulations to explore whether structure more recent than the archaic-modern human split can explain the CSFS.

First, since the CSFS is sensitive to the amount of drift on the introgressing lineage, we simulated model A.2 with several changes in the amount of drift. For all simulations, the introgressing lineage splits 6,000 generations ago, unless otherwise stated. We simulated the following changes to the introgressing lineage:

- $N_e = 100$  (Figure S32;  $\Delta\mathcal{L}\mathcal{L} = -4,504$ )
- $N_e = 1$  (Figure S33;  $\Delta\mathcal{L}\mathcal{L} = -4,695$ )
- $N_e = 1$  with subsequent migration between CEU-YRI (Figure S34;  $\Delta\mathcal{L}\mathcal{L} = -4,148$ )
- $N_e = 0.1$  from a population that splits 200Ky B.P from modern humans (Figure S35;  $\Delta\mathcal{L}\mathcal{L} = -3,517$ )
- a model where the introgressing population branches off the same time as the archaic population and has an  $N_e$  of 100 (a trifurcation) (Figure S36;  $\Delta\mathcal{L}\mathcal{L} = -2,713$ )

From these simulations, we find that increasing the amount of drift in the introgressing lineage while retaining a split time of 200Ky B.P. is insufficient to produce a U-shaped CSFS when the introgressing population branches off after the archaic-modern human split. On the other hand, we find that when the introgressing population is at least as old as the archaic population, we observe a U-shaped CSFS at admixture fractions between 1-20% (Figure S36).

Additionally, we wanted to explore whether any combination of parameters in model A, including genotyping error in the archaic genome or error in the ancestral genome could produce the observed CSFS. We performed two sets of analyses using approximate Bayesian computation. First, we simulated from model A.1, drawing parameters for the split time of the introgressing population ( $ts$  – time is in units of  $4*N_e$  generations where  $N_e = 10000$ ), admixture fraction ( $\alpha$ ), admixture time ( $ta$ ), effective population size of the introgressing population ( $N_e$ ), ancestral misidentification rate ( $e1$ ), and genotyping error in the archaic ( $e2$ ) from the following priors:

$$\begin{aligned} ta &\sim \text{Uniform}(0.06, 0.19) \\ \alpha &\sim \text{Uniform}(0.0, 0.2) \\ ts &\sim \text{Uniform}(0.1, 0.35) \\ ne &\sim \text{Uniform}(0.00001, 1) \\ e1 &\sim \text{Uniform}(0.00003, 0.3) \\ e2 &\sim \text{Uniform}(0.00003, 0.3) \end{aligned}$$

We constrained the admixture time  $ta$  to 2,400-7,600 generations ago and the admixture fraction  $\alpha$  to 0-20%, which corresponds to a split prior to the Out-of-Africa migration and a low to moderate admixture fraction. We constrained the split time of the introgressing lineage to be at least 4,000 generations ago with a maximum of 14,000 generations ago, corresponding to a period prior to the Neanderthal-modern human split. We constrain the population size of the introgressing lineage to 0.1-10,000 diploid individuals, keeping the size of the population equal to or lower than the ancestral modern human population size. Finally, we allow for a wide interval of 0.003%-30% for each of the errors, encompassing a large range of possibilities including those mentioned in Section S1.1.

We drew parameters 50,000 times, simulated 100 MB and used the R package ‘abc’ to compute posterior distributions [Csilléry et al., 2012] (41). We used the ‘neuralnet’ method to perform non-linear regression setting ‘numnet’ = 20 and ‘sizenet’ = 7 and a tolerance of 0.005. We find that the best fitting models have an ancestral misidentification rate of 10.32-14.88% (95% HPD) and an archaic genotyping error rate of 1.95-6.84% (95% HPD; Figure S38).



We applied the same method to model A.2, drawing from the same priors except for the following:

$$\begin{aligned} ta &\sim \text{Uniform}(0.01, 0.053) \\ ts &\sim \text{Uniform}(0.2, 0.35) \end{aligned}$$

We constrained the admixture time to before the Out-of-Africa migration and the split time of the introgressing lineage to at least 8,000 generations ago. For this model, we find that the ancestral misidentification rate is 7.59-13.41% and the archaic genotyping error is 0.63-6.12% (Figure S39).

## S4.2 Genetic drift between archaics and modern humans constrains model A

It is possible that mis-estimates in the split time of archaics and modern humans could allow model A to fit (that is, a model where the introgressing lineage splits after the archaic and modern human split). In particular, we wondered whether an older split time of archaics and modern humans could fit the data. We assessed this possibility by simulating CSFS from a demographic model that includes a trifurcation between modern humans, archaics, and the introgressing lineage with demographic parameters chosen to maximize drift in the modern human population since its split from the archaic population. Specifically, we chose an archaic-modern human split time of 765Ky B.P. (corresponding to the upper bound on the estimate of the archaic-modern human split time obtained by Prüfer et al. [2014] (18) using a mutation rate of  $0.5 \times 10^{-9}$ /bp/year). We converted this estimate from years to generations using a lower bound on the generation time of 25 years per generation. These choices maximize the amount of drift after the archaic-modern human split. We simulated admixture fractions of 1%, 5%, 10%, and 20% (Figure S37). None of these parameters result in a CSFS that yields an adequate fit ( $\text{KS } p < 1 \times 10^{-13}$ ), suggesting that demographic model A does not explain all the features of the CSFS.

Further, we also confirmed that the split time between archaic and modern humans is unlikely to be substantially larger than the values assumed in our simulations by comparing the drift on the modern human lineage produced by these models to estimates of drift in the modern human lineage since the split of modern and archaic humans. We estimated the drift on the modern human lineage since the split from the archaic humans using the property of the CSFS [Chen et al., 2007] (13):

$$\text{CSFS}_{\text{YRI},N} = \frac{\theta e^{-\tau}}{n+1}$$

Here  $\theta = 4N_e\mu L$  (where  $N_e$  is the effective population size of the ancestral population,  $\mu$  is the mutation rate per basepair, and  $L$  is the length of the sequence),  $\tau = \int_0^T \frac{dz}{2N_e(z)}$  (where  $T$  is the split time of the archaic and modern human populations in generations and  $N_e(z)$  denotes the effective population size of the modern human population  $z$  generations B.P.), and  $n$  is the number of chromosomes in the sample. We wish to solve for  $\tau$ , the amount of drift. These theoretical results rely on the assumptions that yield a CSFS that is uniform. Since the empirical CSFS ( $\text{CSFS}_{\text{YRI},N}$ ) is not uniform, we use the mean counts of the CSFS between the allele frequency bins 20%-80%. We take the CSFS computed on 108 YRI individuals from the 1000 Genomes data and project it down to 50 individuals to obtain a mean count of 8,800. Using a mutation rate of  $1.2 \times 10^{-8}$  mutations per basepair per generation, a sequence length of  $3 \times 10^9$  (approximately the length of the human genome), and an ancestral population size of 10,000, we obtain an estimate of  $\tau \approx 0.48$ , which corresponds to  $\approx 9600$  generations (assuming an effective population size in the history of present-day African populations of 10,000). Increasing the mutation rate to  $2 \times 10^{-8}$  (an upper bound), we obtain an estimate of  $\tau \approx 0.99$ , corresponding to  $\approx 20,000$  generations. Lowering the sequence length  $L$  to more conservative values of  $1.5 \times 10^9$  (approximately half the length of the human genome) results in a drift estimate of 0.3 ( $\approx 6,000$  generations). Thus, based on the data, we conservatively estimate the amount of drift as 0.3 – 0.99, or 6,000 – 20,000 generations.

On the other hand, our simulations assume values of drift of 1.53, or 30,600 generations. Therefore, conservative simulations in terms of the split time of archaics and modern humans produce much more drift than is supported by the data and still fail to fit the CSFS. We conclude that the introgressing lineage split from the ancestors of modern humans before the split of archaic and modern humans.

## S5 Estimating parameters for the best fit model of archaic introgression

### S5.1 Method overview

We used approximate Bayesian computation (ABC) [Csilléry et al., 2012] (41) to estimate parameters. We drew parameters from uniform distributions for the split time, admixture time, effective population size, and admixture fraction and simulated 100 MB of data for each draw (Section S5.2). We repeated this procedure 75,000 times for the model and computed the CSFS for each parameter draw, converting the counts into proportions. We used the R package ‘abc’ to estimate parameters for each model [Csilléry et al., 2012] (41). We used the ‘neuralnet’ algorithm, which uses a non-linear regression method for fitting parameters. We note that because we are using regression that is not bound (i.e. the value is unconstrained) to estimate the parameters, it is possible to get estimates or confidence intervals that are below 0, which are nonsensical. We truncate these values to 0 when they occur. We set the number of networks to 50 and set the size of the hidden nodes to 7. We used a tolerance of 0.02.

We verified that this method is valid in our parameter space of interest by simulating data under model A.2 with a split time of 0.15, admixture time of 0.043 and a 15% admixture fraction (time is in units of  $4*N_e$  generations where  $N_e = 10,000$ ). We performed ABC on this simulated data and inferred posterior distributions for each parameter. The mean posterior values are as follows (95% highest posterior density [HPD] in parenthesis): admixture time: 0.047 (0.0436-0.501), admixture fraction: 0.1163 (0.0613-0.1739), and split time: 0.152 (0.1137-0.1874). These values are close to the simulated values, suggesting that the ABC provides valid estimates for this problem.

### S5.2 Parameter estimation

We used ABC to fit 4 parameters from model C: the split time of the introgressing population ( $t_s$ ), the admixture fraction ( $\alpha$ ), the time of introgression ( $t_a$ ), and the effective population size of the introgressing population ( $n_e$ ). We sampled parameters from the following distributions:

$$\begin{aligned}t_a &\sim \text{Uniform}(0.0, 0.25862069) \\ \alpha &\sim \text{Uniform}(0.0, 1.0) \\ t_s &\sim \text{Uniform}(0.25862069, 1.72413793) \\ n_e &\sim \text{Uniform}(0.00001, 3)\end{aligned}$$

The times are given in units of  $4*N_e$  generations where  $N_e = 10,000$ . We applied this method across the 4 African populations in the 1000 Genomes project. Because each population had a different number of samples, we projected the CSFS down to 50 diploid samples using the hypergeometric distribution. In brief, we compute the hypergeometric probability mass function for one variant at a particular frequency class,  $i$ . Then, we multiply the projected SFS for variant  $i$  by the number of variants at that frequency. We perform this over all variant frequencies,  $i \in \{1..n - 1\}$ , where  $n$  is the original diploid sample size multiplied by 2.

We chose broad priors over the range of possible values for each of the parameters. Converting generations into years using a value of 29 years per generation,  $t_a$  ranges approximately from the present day to 300 Ky B.P and  $t_s$  ranges from 300 Ky B.P to 2 My B.P.  $n_e$  ranges from 1 diploid individual to 30,000 diploid individuals and  $\alpha$  includes all values from no contribution (0) to completely replacing the lineage (1).

### S5.3 Uncertainty in date estimates

We perform our inference measuring time in coalescent units. Converting these units to years is not trivial because of the uncertainty in generation time. To incorporate variation and uncertainty in generation time, we convolve our posterior distributions on generation times with a uniform distribution of generation times between 25 and 33. In particular, we sample from the posterior distribution of admixture time and split time and multiply by  $y$ , where  $y \sim \text{Uniform}(25, 33)$ . We repeat this procedure 10,000 times and report summaries of this posterior distribution.

## S6 Continuous migration versus a single pulse

We have modeled the introgression event as a single discrete admixture event in our simulations. It is plausible, however, that this event happened over an extended time scale with continuous migration. We wanted to determine whether a model of continuous migration could qualitatively affect our conclusions, in particular with respect to the split time of the introgressing lineage. In Henn et al. [2018] (21), the authors distinguish between a single origin range expansion with regional persistence model, where a single population expands and receives gene flow from other populations over the past  $\leq 200$ Ky B.P. and an African multiregional model where this split occurs further back in time, around 350 Ky B.P.

We simulated a lineage which split from the modern human lineage at various times ranging from 100 Ky B.P to 500Ky B.P with continuous migration with the modern human branch at rates  $m \in \{2.5 \times 10^{-5}, 5 \times 10^{-5}, 1 \times 10^{-4}, 2 \times 10^{-4}, 2 \times 10^{-3}, 2 \times 10^{-2}\}$  migrants per generation. We used a generation time of 29 years to convert these split times to generations. While this model can produce a U-shape with a split as recent as 175Ky B.P (Figure S14), we do not see a good fit to the data (KS  $p \leq 2.3 \times 10^{-5}$ ; Figure S15).

As a further check, we used the ABC framework from Section S5.2 to fit the continuous migration model to the data. We simulated a lineage that splits  $t_s$  generations ago and continuously exchanges migrants with the modern human branch until the present day at a rate of  $m$  migrants per generation. This branch has an effective population sizes of  $n_e$ . We drew parameters from the following uniform distributions

$$\begin{aligned}t_s &\sim \text{Uniform}(0, 18960) \\m &\sim \text{Uniform}(0, 8 \times 10^{-4}) \\n_e &\sim \text{Uniform}(0.00001, 3)\end{aligned}$$

we drew 25,000 parameter values and simulate 300 MB per parameter draw. We found that the best fitting parameters were a split time of 15,200 generations ago (95% HPD: 14,800-15,800), a migration rate of  $1.04 \times 10^{-4}$  individuals per generation (95% HPD:  $1.01 \times 10^{-4}, 1.06 \times 10^{-4}$ ), and an effective population size of 39,600 individuals (95% HPD: 35,000-44,000). We simulated data under these parameters and performed a goodness-of-fit test (see ‘Goodness of fit’ in the Methods section of the main text). We found that we could reject this model (KS  $p = 1.83 \times 10^{-6}$ ).

## S7 Local ancestry inference

We used ArchIE, a method we previously developed for archaic local ancestry inference that does not require access to an archaic reference genome [Durvasula and Sankararaman, 2019] (22). ArchIE is a machine learning method that uses data simulated under an assumed model and classifies 50KB windows as originating from an archaic population or not. ArchIE uses features that are informative of archaic introgression, including the number of private SNPs in a window, the S\*-statistic [Plagnol and Wall, 2006] (7), the minimum distance to an unadmixed reference population, and the distribution of pairwise differences between all segments in a sample.

To estimate the parameters of ArchIE, we simulated training data under a demography where an archaic population splits 12,000 generations before present and introgresses into a modern human population 2,000 generations B.P. with a 2% archaic admixture fraction. We use a fixed mutation rate of  $1.25 \times 10^{-8}$  and a recombination rate of  $1 \times 10^{-8}$ . We simulate two modern human populations, the target, which receives archaic gene flow, and the reference, which does not receive archaic gene flow. The reference population split from the target 2,500 generations B.P. All population sizes were set to 10,000 and we simulated 50 diploid individuals in both the reference and target populations. We simulated 50 KB segments. We have shown through extensive simulations [Durvasula and Sankararaman, 2019] (22) that ArchIE can accurately infer segments of deeply diverged ancestry even in settings where the true demographic model differs from the model used to estimate its parameters.

To further evaluate the accuracy of ArchIE in inferring diverged ancestry, we assessed its ability to recover introgressed lineages simulated our best fitting demographic model C (even though the parameters of ArchIE were estimated under a different demographic model described above). We simulated a lineage that split 625Ky B.P from modern humans and introgressed into the African branch 43Ky B.P with an admixture fraction of 11%. We included introgression from a Neanderthal branch (diverging 12,000 generations ago)

into the reference population (representing CEU) 2,000 generations ago at a 0.02 admixture fraction. We simulated a total of 2.5 MB of sequence in 50KB segments.

We report power and false discovery rates (FDR) for both of these scenarios at the 20% false discovery rate threshold obtained under the model used for generating training data ( $t_{arch} = P(\text{archaic}) = 0.862$ ). These quantities are defined as:

$$\text{power}(t) = \frac{TP(t)}{TP(t) + FN(t)}$$

$$FDR(t) = \frac{FP(t)}{FP(t) + TP(t)}$$

Here  $t$  is the probability cutoff (0.862),  $TP(t)$  is the number of true positives at threshold  $t$  (archaic segments that are assigned a probability of being archaic  $\geq t$ ),  $FN(t)$  is the number of false negatives at threshold  $t$  (archaic segments we do not call archaic),  $FP(t)$  is the number of false positives at threshold  $t$  (non-archaic segments we call archaic), and  $TN(t)$  is the number of true negatives at threshold  $t$  (non-archaic segments we call non-archaic). Under our best fitting model (using YRI data), we have 68% power to detect archaic segments and a false discovery rate of 7%. We note that we have low power to detect archaic segments that are shared between out-of-Africa and African populations (*i.e.* under model C.1) because the modern human reference population is less informative for these haplotypes (see Durvasula and Sankararaman [2019] (22) for more details about the impact of the modern human reference population).

We applied ArchIE to 50 individuals from the YRI population and 50 from the MSL population using a sample of 50 CEU individuals as the reference population. We applied the 1000 Genomes strict mask and restricted to segments where  $\geq 90\%$  of the segment is callable. At the threshold  $t_{arch}$ , ArchIE calls  $\simeq 6.6\%$  of the genome in YRI and  $\simeq 7.0\%$  in the MSL as putatively archaic on average.

Our local ancestry inference provides us with the opportunity to dive deeper into previous analyses of ancestral structure and archaic introgression. In Figure S42 we show the frequency of archaic segments in YRI and MSL across the genome. We observe that the maps are highly correlated (Spearman’s  $\rho = 0.71$   $p < 2 \times 10^{-16}$ ). Additionally, in Plagnol and Wall [2006] (7), the authors list 27 loci in the west African Yoruba that could be the result of ancient admixture. We looked at the frequency of archaic segments near these genes ( $\pm 1\text{MB}$  from the start and end sites). We computed the maximum frequency of archaic segments near these genes and found that the average was 15% in YRI. This is consistent with those genes harboring archaic ancestry and confirms the results of Plagnol and Wall [2006] (7).

## S7.1 Scaled divergence estimates

We sought to determine whether the segments labeled archaic by ArchIE are likely to trace their ancestry to Hunter Gatherer populations (Ju—hoan and Khomani San) Central African Pygmy populations (Mbuti and Biaka) or to the Neanderthal and Denisovan genomes. To do this, we comparing the divergence of these segments to genomes from each of these populations. If segments labeled archaic are closer to one of these populations than those labeled non-archaic, it suggests that population may be the source.

We denote the genome we are comparing the calls from ArchIE to as the ‘test’ genome. To estimate divergence of a segment, we count the number of mutations specific to the archaic segment (*i.e.*, not shared with the test genome) and subtract the number of mutations shared between the archaic segment and the test genome. We divide this number by the number of segregating sites in the window to control for local mutation rate variation.

We test the null hypothesis that the difference in mean divergence between archaic and non-archaic segments is 0. We compute  $p$ -values for this null using a block jackknife [Kunsch, 1989] (50). Traversing the genome, we delete 10 MB blocks and recalculate the difference in mean divergence.

We find that for all comparisons, the mean divergence is significantly higher for archaic segments than non-archaic segments.

## S7.2 Coalescent-based age estimates of introgressed segments

We used the posterior decoding from MSMCv1.1 [Schiffels and Durbin, 2014] (23) to compute the age of putative archaic segments in the genome. When run on a pair of genomes, this method is called PSMC’.

Intuitively, this method is a hidden Markov model that uses the observed heterozygosity within an individual to determine the time to most recent common ancestor in blocks along the genome. We selected one individual from each population (HG03079 for MSL and NA18486 for YRI) and computed the posterior distribution of coalescent times in 1000bp bins. The times from MSMC are scaled by  $2N$ , where  $N$  is determined by the heterozygosity. We divide the observed  $\hat{\theta}$  (Watterson’s  $\theta$ , computed by MSMC. See methods section in Schiffels and Durbin [2014] (23)) by the mutation rate  $\mu = 1.25 \times 10^{-8}$ . We multiply the coalescent times outputted by MSMC by this factor in order to scale it to generations. Then, we multiplied the generations by 29 years per generation to obtain coalescence estimates in years [Fenner, 2005] (48). We then intersected these times with the segment calls from ArchIE, restricting the analysis to confidently called archaic segments (probability of being archaic  $P(\text{archaic}) \geq t_{\text{arch}}$ ) and confidently called non-archaic segments ( $P(\text{archaic}) \leq 1 - t_{\text{arch}}$ ). We restricted the analysis to windows where  $> 90\%$  of the bases are callable as defined by the 1000 Genomes strict mask.

We find that in YRI, the median coalescent time for archaic segments is 1.5 Mya and 0.89 Mya for non-archaic segments, a 1.69-fold increase in age. In MSL, we find the median coalescent time is 1.1 Mya and 0.93 Mya for archaic and non-archaic segments respectively.

### S7.3 Adaptive archaic introgression

We examined the maps of archaic introgression in YRI and MSL, computing the frequency of the archaic segments within 50 KB windows. We intersected our calls with gene annotations from Gencode v19 using bedtools [Quinlan and Hall, 2010] (50) and included a gene if any basepair in a single 50KB segment overlapped with the gene [Frankish et al., 2019] (51).

In order to evaluate whether the high frequency archaic segments could have risen to their present-day frequencies as a result of drift, we performed coalescent simulations using the parameters we estimated in S5.2. We simulated 100 replicates of 50KB with a 13% introgression pulse 1400 generations ago from a lineage that split from modern humans 23,000 generations ago. The introgressing lineage had an effective population size of 25,000. We chose the introgression pulse and introgression time to be on the upper bound of the 95% HPD estimates so as to maximize the drift in the population after introgression. Thus, we expect the distribution of archaic allele frequencies to be an upper bound on the neutral distribution. The distribution of allele frequencies for alleles originating in the introgressing lineage (Figure S41) gives us the null distribution of archaic allele frequencies expected under neutrality. Observing archaic alleles with frequencies above this threshold suggests that these alleles have risen to high frequencies due to selection.

## S8 Extended Discussion

### S8.1 Results from the Luhya

Examination of the CSFS in the Luhya from Kenya (LWK) also reveals a U-shaped spectrum, similar to the other African populations examined here (Figure S5). We performed the same ABC-based parameter estimation procedure and found that a lineage splitting 624,000 years ago (95% HPD: 290,100-1,137,000) with an  $N_e$  of 24,000 (95% HPD: 22,000-24,000) and introgressing 50,000 years ago (95% HPD: 11,000-113,000) is able to explain the shape of the spectrum. These parameters are similar to the other African populations. However, we find that an introgression fraction of 0.18 (95 % HPD: 0.09-0.31) is also required, which is much higher than the other populations. Inferences in the LWK could be complicated by non-African ancestry in the Luhya, resulting in increased estimates of the admixture fraction [Wall et al., 2013] (52). However, gene flow from the Neanderthals mediated through non-African ancestry into the LWK may not be enough to quantitatively explain this increase. Another possibility is other unmodeled gene flow events outside of the ones we describe here.

### S8.2 Pre- versus post- Out-of-Africa

We inspected the CSFS in the CEU conditioned on the Denisovan allele, since there is little Denisovan-related ancestry in Europeans [Mallick et al., 2016] (1). If the introgression event is shared in YRI and CEU, we expect to see a U-shaped CSFS in both populations (Figure S40abc). However, if the event is not

shared and is specific to African populations only, the signal is expected to only appear in the African CSFS (Figure S40def). We observe that the CEU CSFS, like the CSFS in the African populations, is U-shaped with an increase in the counts of high frequency alleles. (Figures S4 and S6). Further, we also observe a U-shape in the CSFS computed in the Han Chinese (CHB) population (Figure S6). These results suggest that a component of the archaic ancestry that we detect in African populations is shared with non-African populations.

We performed additional modeling considering the YRI and CEU spectra jointly. Since the right (high frequency derived) side of the spectrum is most informative for this question, we only considered the allele frequency bins above 50. We simulated data using the same priors in Section S5.2, but computed the spectrum for both YRI and CEU. We found that the best fitting parameters were an archaic split time of 27,000 generations ago (95% HPD: 26,000-28,000), admixture fraction of 0.09 (95% HPD: 0.04-0.17), admixture time of 3,000 generations ago (95% HPD: 2,800-3,400), and an effective population size of 19,700 individuals (95% HPD: 19,300-20,200). We find that the lower bound of the admixture time is further back than the simulated split between CEU and YRI (2155 generations ago), providing some evidence in favor of a pre-Out-of-Africa event. This model suggests that many populations outside of Africa should also contain haplotypes from this introgression event, though detection is difficult because many methods use unadmixed outgroups to detect introgressed haplotypes [Browning et al., 2018, Skov et al., 2018, Durvasula and Sankararaman, 2019] (5, 53, 22). It is also possible that some of these haplotypes were lost during the Out-of-Africa bottleneck.

## S9 ms command lines

Here, we provide the command lines used in the various models.

### Model A.1

```
ms 202 1 -t 480 -r 519.99948 1000000 -I 4 100 100 1 1 -en 0 3 0.1 -en 0 4 0.1
-es 0.04310345 1 0.97 -ej 0.05387931 5 3 -ej 0.05387931 2 1 -es 0.055 1 ${admix}
-ej 0.15 6 1 -es 0.2155172 4 0.94 -en 0.2155172 7 0.2 -es 0.2836207 3 0.95
-ej 0.2836207 8 1 -ej 0.3577586 4 3 -ej 0.4741379 3 1 -ej 0.7844828 7 1
```

### Model A.2

```
ms 202 1 -t 480 -r 519.99948 1000000 -I 4 100 100 1 1 -en 0 3 0.1 -en 0 4 0.1
-es 0.04310345 1 0.97 -es 0.04310345 2 ${admix} -ej 0.05387931 5 3 -ej 0.05387931 2 1
-ej 0.15 6 1 -es 0.2155172 4 0.94 -en 0.2155172 7 0.2 -es 0.2836207 3 0.95 -ej 0.2836207 8 1
-ej 0.3577586 4 3 -ej 0.4741379 3 1 -ej 0.7844828 7 1
```

### Model B.1

```
ms 202 1 -t 480 -r 519.99948 1000000 -I 4 100 100 1 1 -en 0 3 0.1 -en 0 4 0.1
-es 0.04310345 1 0.97 -ej 0.05387931 5 3 -ej 0.05387931 2 1 -es 0.1725 1 ${admix}
-es 0.2155172 4 0.94 -en 0.2155172 7 0.2 -es 0.2836207 3 0.95 -ej 0.2836207 8 1
-ej 0.3577586 4 3 -ej 0.40 6 3 -ej 0.4741379 3 1 -ej 0.7844828 7 1
```

### Model B.2

```
ms 202 1 -t 480 -r 519.99948 1000000 -I 4 100 100 1 1 -en 0 3 0.1 -en 0 4 0.1
-es 0.04310345 1 0.97 -es 0.04310345 2 ${admix} -ej 0.05387931 5 3 -ej 0.05387931 2 1
-es 0.2155172 4 0.94 -en 0.2155172 7 0.2 -es 0.2836207 3 0.95 -ej 0.2836207 8 1
-ej 0.3577586 4 3 -ej 0.4 6 3 -ej 0.4741379 3 1 -ej 0.7844828 7 1
```

### Model C.1

```
ms 202 1 -t 480 -r 519.99948 1000000 -I 4 100 100 1 1 -en 0 3 0.1 -en 0 4 0.1
-es 0.04310345 1 0.97 -ej 0.05387931 5 3 -ej 0.05387931 2 1
-es 0.055 1 ${admix} -es 0.2155172 4 0.94 -en 0.2155172 7 0.2
-es 0.2836207 3 0.95 -ej 0.2836207 8 1 -ej 0.3577586 4 3
-ej 0.4741379 3 1 -ej 0.6 6 1 -ej 0.7844828 7 1
```

## Model C.2

```
ms 202 1 -t 480 -r 519.99948 1000000 -I 4 100 100 1 1 -en 0 3 0.1 -en 0 4 0.1
-es 0.04310345 1 0.97 -es 0.04310345 2 ${admix} -ej 0.05387931 5 3 -ej 0.05387931 2 1
-es 0.2155172 4 0.94 -en 0.2155172 7 0.2 -es 0.2836207 3 0.95 -ej 0.2836207 8 1
-ej 0.3577586 4 3 -ej 0.4741379 3 1 -ej 0.6 6 1 -ej 0.7844828 7 1
```

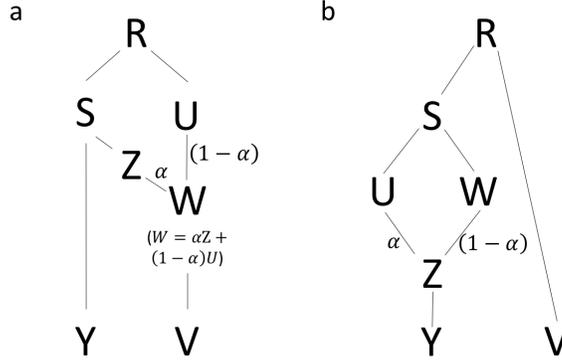


Figure S3: Demographic model topologies for mathematical results regarding a) the expectation of the CSFS when there is structure in the archaic population (Section A) and b) the symmetry of the CSFS under model A (Section B)

## A The CSFS is uniform under structure in the archaic population

We consider a simplified demographic model where the ancestors of the African and Neanderthal populations split with no subsequent gene flow. However, we consider the possibility of structure in the archaic population. This structure could have several forms that include the ancestral Neanderthal population being structured or it could involve gene flow from early modern humans into Neanderthals [Kuhlwilm et al., 2016] or as in the case of Denisovans, this could include gene flow from a highly-diverged archaic population [Prüfer et al., 2014].

To consider these possibilities, we consider the ancestral population ( $R$ ) that splits into two populations ( $S$  and  $U$ ). At a later time,  $S$  further splits into two populations: one of which evolves to become present-day Africans  $Y$  and the other  $Z$  admixes with  $U$  to give rise to population  $W$  (with admixture fraction  $\alpha : 1 - \alpha$ ) which then further evolves into the Neanderthal population  $V$ . See Figure S3a for a schematic of the topology. We use  $R$  to represent the random variable denoting the frequency of the derived allele in population  $R$  and  $r$  to denote the realized value of this random variable (and analogously for each of the other populations). We measure the time elapsed  $\tau$  from each of these populations to its parent population (where defined) in diffusion units, *i.e.*,  $\tau = \frac{t}{2N}$  where  $t$  is the elapsed time in generations and  $N$  is the diploid effective population size that is assumed constant. When the population size  $N(x)$  changes with time, we define  $\tau = \int_0^t \frac{dx}{2N(x)}$ . The drift leading to population  $S$  from its parent population  $R$  is denote  $\tau_S$  and so on. Within each of the populations, we assume random mating. Finally, we also assume that populations  $R$  and  $S$  are at mutation-drift equilibrium.

We assume no recurrent mutation and neutral evolution of alleles. Let  $X(\tau)$  denote the frequency of an allele at time  $\tau$  where is time scaled in diffusion units. The SNP is assumed polymorphic at time  $\tau = 0$ ,  $0 < X(0) < 1$ . The evolution of  $X(\tau)$  can be described by a Wright-Fisher diffusion that describes the effects of genetic drift on the allele frequency. Let  $P(y|x, \tau)$  denote the transition probability that the allele frequency is  $y$  at time  $\tau$  given that allele frequency at time 0 is  $x$  (we will overload  $P$  to refer to the probability density when  $0 < y < 1$  and the probability when  $y = 0$  or  $y = 1$ ).

We have:

$$\begin{aligned}
 P(y|x, \tau) &= P(y, \mathbf{1}(y = 0) | x, \tau) + P(y, \mathbf{1}(0 < y < 1) | x, \tau) + P(y, \mathbf{1}(y = 1) | x, \tau) \\
 &= \psi_0(x; \tau) + \phi(y, x; \tau) + \psi_1(x; \tau)
 \end{aligned} \tag{1}$$

The explicit forms of  $\psi_0$ ,  $\psi_1$ , and  $\phi$  have been solved by Kimura [1955].

We sample  $n$  chromosomes in Africans and 1 chromosome from the Neanderthal. The CSFS  $f_i$  is the number of SNPs at which  $i$  chromosomes ( $i \in \{1, \dots, n - 1\}$ ) in the African population carry the derived



allele and the chromosome sampled from Neanderthal also carries the derived allele.

We show that:

$$f_i = \text{constant} \quad (2)$$

We have:

$$f_i = \binom{n}{i} \int y^i (1-y)^{n-i} f(y) dy \quad (3)$$

Here  $f(y)$  denotes the population conditional frequency spectrum in Africans conditioned on observing a derived allele in Neanderthal. We note that the only SNPs that will contribute to  $f_i$  are those that are polymorphic at  $Y$  and contain the derived allele at  $V$ . This requires that these SNPs are also polymorphic at  $S$ . Further, it will be useful to partition the SNPs contributing to  $f(y)$  to those that were polymorphic at  $R$  (and also polymorphic at  $S$ ) ( $f_{old}$ ) and those mutations that arose in the ancestral population leading to  $S$  from  $R$  ( $f_{new}$ ).

$$f(y) = f_{old}(y) + f_{new}(y) \quad (4)$$

The assumption of mutation drift equilibrium at  $R$  and  $S$  implies:

$$f_0(s) = f_{0,new}(s) + \int \phi(s, r; \tau_S) f_0(r) dr \quad (5)$$

Here  $f_0$  is the frequency spectra at  $R$  and  $S$  respectively while  $f_{0,new}$  denotes the frequency spectrum due to mutations that arose after  $R$ .

Let  $P(r, s, u, z, w, y, v)$  denote the joint probability  $P(R = r, S = s, U = u, Z = z, W = w, Y = y, V = v)$ .

$$\begin{aligned}
f_{old}(y) &= \int P(r, s, u, z, w, y, v) v dr ds dudz dw dv \\
&= \int \left\{ P(r) P(s|r, \tau_S) P(u|r, \tau_U) P(z|s, \tau_Z) \right. \\
&\quad \left. P(v|w, \tau_V) P(y|s, \tau_Y) \mathbf{1}(w = \alpha z + (1 - \alpha)u) \right\} dr ds dudz dw dv \tag{6}
\end{aligned}$$

$$\begin{aligned}
&= \int \left\{ f_0(r) \phi(s, r; \tau_S) P(u|r, \tau_U) P(z|s, \tau_Z) \right. \\
&\quad \left. P(v|w, \tau_V) \phi(y, s; \tau_Y) \mathbf{1}(w = \alpha z + (1 - \alpha)u) \right\} dr ds dudz dw dv \tag{7}
\end{aligned}$$

$$\begin{aligned}
&= \int \left\{ f_0(r) \phi(s, r; \tau_S) P(u|r, \tau_U) P(z|s, \tau_Z) \right. \\
&\quad \left. \left( \int P(v|w, \tau_V) v dv \right) \phi(y, s; \tau_Y) \mathbf{1}(w = \alpha z + (1 - \alpha)u) \right\} dr ds dudz dw \\
&= \int \left\{ f_0(r) \phi(s, r; \tau_S) P(u|r, \tau_U) P(z|s, \tau_Z) \right. \\
&\quad \left. \phi(y, s; \tau_Y) \left( \int \mathbf{1}(w = \alpha z + (1 - \alpha)u) w dw \right) \right\} dr ds dudz \tag{8}
\end{aligned}$$

$$= \int \left\{ f_0(r) \phi(s, r; \tau_S) P(u|r, \tau_U) P(z|s, \tau_Z) \phi(y, s; \tau_Y) (\alpha z + (1 - \alpha)u) \right\} dr ds dudz \tag{9}$$

$$\begin{aligned}
&= \alpha \int \left\{ f_0(r) \phi(s, r; \tau_S) \left( \int P(u|r, \tau_U) du \right) \left( \int P(z|s, \tau_Z) z dz \right) \phi(y, s; \tau_Y) \right\} dr ds \\
&\quad + (1 - \alpha) \int \left\{ f_0(r) \phi(s, r; \tau_S) \left( \int P(u|r, \tau_U) u du \right) \left( \int P(z|s, \tau_Z) dz \right) \phi(y, s; \tau_Y) \right\} dr ds \tag{10}
\end{aligned}$$

$$= \alpha \int \left\{ f_0(r) \phi(s, r; \tau_S) s \phi(y, s; \tau_Y) \right\} dr ds + (1 - \alpha) \int \left\{ f_0(r) \phi(s, r; \tau_S) r \phi(y, s; \tau_Y) \right\} dr ds \tag{11}$$

$$= \alpha \int \left\{ \left( \int f_0(r) \phi(s, r; \tau_S) dr \right) s \phi(y, s; \tau_Y) \right\} ds + (1 - \alpha) \int \left\{ \left( \int f_0(r) \phi(s, r; \tau_S) r dr \right) \phi(y, s; \tau_Y) \right\} ds$$

$$= \alpha \int \left( \int f_0(r) \phi(s, r; \tau_S) dr \right) s \phi(y, s; \tau_Y) ds + (1 - \alpha) \theta \int \left( \int \phi(s, r; \tau_S) dr \right) \phi(y, s; \tau_Y) ds \tag{12}$$

$$= \alpha \int \left( \int f_0(r) \phi(s, r; \tau_S) dr \right) s \phi(y, s; \tau_Y) ds + (1 - \alpha) \theta \exp(-\tau_S) \int \phi(y, s; \tau_Y) ds \tag{13}$$

$$= \alpha \int \left( \int f_0(r) \phi(s, r; \tau_S) dr \right) s \phi(y, s; \tau_Y) ds + (1 - \alpha) \theta \exp(-\tau_S) \exp(-\tau_Y) \tag{14}$$

Here Equation 6 follows from the conditional independence statements implied by our model. Equation 7 follows because only the polymorphic SNPs at  $S$  and  $Y$  contribute to  $f(y)$ . Equations 8 and 11 follow from the fact that the Wright-Fisher diffusion is a Martingale so that  $\mathbb{E}[X(\tau)|X(0)] = X(0)$ . Equation 12 follows from the fact that the frequency spectrum of derived alleles at  $R$  under mutation-drift equilibrium is given by  $f_0(r) = \frac{\theta}{r}$  [Griffiths, 2003]. Equations 13 and 14 follow from explicit integration using the analytical solution for  $\phi(y, x; \tau)$  [Kimura, 1955, Chen et al., 2007].

$$\begin{aligned}
f_{new}(y) &= \int P(s, z, w, y, v) v ds dz dw dv \\
&= \int \left\{ P(s) P(z|s, \tau_Z) P(v|w, \tau_V) P(y|s, \tau_Y) \mathbf{1}(w = \alpha z) v \right\} ds dz dw dv \tag{15}
\end{aligned}$$

$$= \int \left\{ f_{0,new}(s) P(z|s, \tau_Z) P(v|w, \tau_V) \phi(y, s; \tau_Y) \mathbf{1}(w = \alpha z) v \right\} ds dz dw dv \tag{16}$$

$$\begin{aligned}
&= \int \left\{ f_{0,new}(s) P(z|s, \tau_Z) \left( \int P(v|w, \tau_V) v dv \right) \phi(y, s; \tau_Y) \mathbf{1}(w = \alpha z) \right\} ds dz dw \\
&= \int \left\{ f_{0,new}(s) P(z|s, \tau_Z) \phi(y, s; \tau_Y) \left( \int \mathbf{1}(w = \alpha z) w dw \right) \right\} ds dz \tag{17}
\end{aligned}$$

$$= \alpha \int \left\{ f_{0,new}(s) \left( \int P(z|s, \tau_Z) z dz \right) \phi(y, s; \tau_Y) \right\} ds \tag{18}$$

$$= \alpha \int f_{0,new}(s) s \phi(y, s; \tau_Y) ds \tag{19}$$

Here Equation 15 follows from the conditional independence statements implied by our model. Equation 16 follows because only the polymorphic SNPs at  $Y$  contribute to  $f(y)$ . Equations 17 and 19 follow from the Martingale property of Wright-Fisher diffusion.

From Equations 14 and 19:

$$\begin{aligned}
f(y) &= f_{old}(y) + f_{new}(y) \\
&= \alpha \int \left( \int f_0(r) \phi(s, r; \tau_S) dr + f_{0,new}(s) \right) s \phi(y, s; \tau_Y) ds + (1 - \alpha) \theta \exp(-(\tau_S + \tau_Y)) \\
&= \alpha \int f_0(s) s \phi(y, s; \tau_Y) ds + (1 - \alpha) \theta \exp(-(\tau_S + \tau_Y)) \tag{20}
\end{aligned}$$

$$= \alpha \theta \int \phi(y, s; \tau_Y) ds + (1 - \alpha) \theta \exp(-(\tau_S + \tau_Y)) \tag{21}$$

$$= \theta \left( \alpha \exp(-\tau_Y) + (1 - \alpha) \exp(-(\tau_S + \tau_Y)) \right) \tag{22}$$

Equation 20 follows from Equation 5 while Equation 21 follows from the mutation-drift equilibrium and Equation 22 follows from previous results [Chen et al., 2007].

Substituting Equation 22 into Equation 3, we have:

$$f_i = \frac{1}{n+1} \theta \left( \alpha \exp(-\tau_Y) + (1 - \alpha) \exp(-(\tau_S + \tau_Y)) \right) \tag{23}$$

## B The CSFS is symmetric under Model A

We consider a simplified version of demographic model A where the ancestors of the African and Neanderthal populations split with no subsequent gene flow. However, in this model, the ancestral African population then splits into two sub-populations that later admix.

The ancestral population (labeled  $R$ ) splits into the two populations that are the ancestors of Africans and Neanderthals (labeled  $V$ ) respectively. The ancestral African population (labeled  $S$ ) splits into two sub-populations. These two sub-populations (labeled  $U$  and  $W$ ) admix in the ratio  $\alpha : 1 - \alpha$  to form population  $Z$  which then continues to evolve to give rise to the present-day African population (labeled  $Y$ ). We sample a single chromosome from the Neanderthal population (labeled  $V$ ). See Figure S3b for a schematic of the topology. We use  $R$  to represent the random variable denoting the frequency of the derived allele in population  $R$  and  $r$  to denote the realized value of this random variable (and analogously for each of the other populations). We measure the time elapsed  $\tau$  from each of these populations to its parent population (where defined) in diffusion units, *i.e.*,  $\tau = \frac{t}{2N}$  where  $t$  is the elapsed time in generations and  $N$  is the diploid effective population size that is assumed constant. When the population size  $N(x)$  changes with time, we define  $\tau = \int_0^t \frac{dx}{2N(x)}$ . The drift leading to population  $S$  from its parent population  $R$  is denote  $\tau_S$  and so on. Within each of the populations, we assume random mating. Finally, we also assume that population  $R$  is at mutation-drift equilibrium.

As in the previous section, we assume no recurrent mutation and neutral evolution of alleles. Let  $X(\tau)$  denote the frequency of an allele at time  $\tau$  where is time scaled in diffusion units. The evolution of  $X(\tau)$  can be described by a Wright-Fisher diffusion that describes the effects of genetic drift on the allele frequency and  $P(y|x, \tau)$  denotes the transition probability that the allele frequency is  $y$  at time  $\tau$  given that allele frequency at time 0 is  $x$ .

We sample  $n$  chromosomes in Africans and 1 chromosome from the Neanderthal. The CSFS  $f_i$  is the number of SNPs at which  $i$  chromosomes ( $i \in \{1, \dots, n-1\}$ ) in the African population carry the derived allele and the chromosome sampled from Neanderthal also carries the derived allele. We show that the CSFS is symmetric under this model:  $f_i = f_{n-i}, i \in \{1, \dots, n-1\}$ .

We then have:

$$f_i = \binom{n}{i} \int y^i (1-y)^{n-i} f(y) dy \quad (24)$$

$f(y)$  denotes the conditional frequency spectrum in Africans conditioned on observing a derived allele in Neanderthal. We note that the only SNPs that will contribute to  $f_i$  are those that are polymorphic at  $Y$  and contain the derived allele at  $V$ . This requires that these SNPs are also polymorphic at  $R$ ,  $S$ , and  $Z$ .

Let  $P(r, s, u, w, z, y, v)$  denote the joint probability  $P(R = r, S = s, U = u, W = w, Z = z, Y = y, V = v)$ .

$$\begin{aligned} f(y) &= \int P(r, s, u, w, z, y, v) v dr ds du dw dz dv \\ &= \int \left\{ P(r) P(v|r, \tau_V) P(s|r, \tau_S) P(u|s, \tau_U) P(w|s, \tau_W) P(y|z, \tau_Y) \mathbf{1}(z = \alpha u + (1-\alpha)w) \right\} v dr ds du dw dz dv \quad (25) \end{aligned}$$

$$\begin{aligned} &= \int \left\{ f_0(r) P(v|r, \tau_V) \phi(s, r; \tau_S) P(u|s, \tau_U) P(w|s, \tau_W) \phi(y, z; \tau_Y) \mathbf{1}(z = \alpha u + (1-\alpha)w) \right\} v dr ds du dw dz \\ &= \int \left\{ f_0(r) \left( \int P(v|r, \tau_V) v dv \right) \phi(s, r; \tau_S) P(u|s, \tau_U) P(w|s, \tau_W) \phi(y, z; \tau_Y) \mathbf{1}(z = \alpha u + (1-\alpha)w) \right\} dr ds du dw dz \end{aligned}$$

$$= \int \left\{ r f_0(r) \phi(s, r; \tau_S) P(u|s, \tau_U) P(w|s, \tau_W) \phi(y, z; \tau_Y) \mathbf{1}(z = \alpha u + (1-\alpha)w) \right\} dr ds du dw dz \quad (26)$$

$$= \theta \int \left\{ \phi(s, r, \tau_S) P(u|s, \tau_U) P(w|s, \tau_W) \phi(y, z; \tau_Y) \mathbf{1}(z = \alpha u + (1-\alpha)w) \right\} dr ds du dw dz \quad (27)$$

$$= \theta \int \left\{ \left( \int \phi(s, r, \tau_S) dr \right) P(u|s, \tau_U) P(w|s, \tau_W) \phi(y, z; \tau_Y) \mathbf{1}(z = \alpha u + (1-\alpha)w) \right\} ds du dw dz \quad (28)$$

$$= \theta \exp(-\tau_S) \int \left\{ P(u|s, \tau_U) P(w|s, \tau_W) \phi(y, z; \tau_Y) \mathbf{1}(z = \alpha u + (1-\alpha)w) \right\} ds du dw dz \quad (29)$$

Here Equation 25 follows from the conditional independence statements implied by our model. Equation 26 follows from the fact that the Wright-Fisher diffusion is a Martingale. Equation 27 follows from the fact that the frequency spectrum of derived alleles at  $R$  under mutation-drift equilibrium is given by  $P(r) = \frac{\theta}{r}$  [Griffiths, 2003]. Equation 28 follows from explicit integration using the analytical solution for  $P(s|r, \tau_S)$  [Kimura, 1955, Chen et al., 2007].

$$\begin{aligned}
f(y) &= \theta \exp(-\tau_S) \int \left\{ P(u|s, \tau_U) P(w|s, \tau_W) P(y|z, \tau_Y) \mathbf{1}(z = \alpha u + (1 - \alpha)w) \right\} ds du dw dz \\
&= \theta \exp(-\tau_S) \int \left\{ P(1 - u|1 - s, \tau_U) P(1 - w|1 - s, \tau_W) \right. \\
&\quad \left. P(1 - y|1 - z, \tau_Y) \mathbf{1}(1 - z = \alpha(1 - u) + (1 - \alpha)(1 - w)) \right\} ds du dw dz \tag{30}
\end{aligned}$$

$$= \theta \exp(-\tau_S) \int P(u|s, \tau_U) P(w|s, \tau_W) P(1 - y|z, \tau_Y) \mathbf{1}(z = \alpha u + (1 - \alpha)w) ds du dw dz \tag{31}$$

$$= f(1 - y) \tag{32}$$

Equation 30 uses the property of the transition probability that  $P(y|x, \tau) = P(1 - y|1 - x, \tau)$  while Equation 31 follows by a change of variables.

Combining Equations 32 and 24, we have  $f_i = f_{n-i}$  for all  $i \in \{1, \dots, n - 1\}$ .

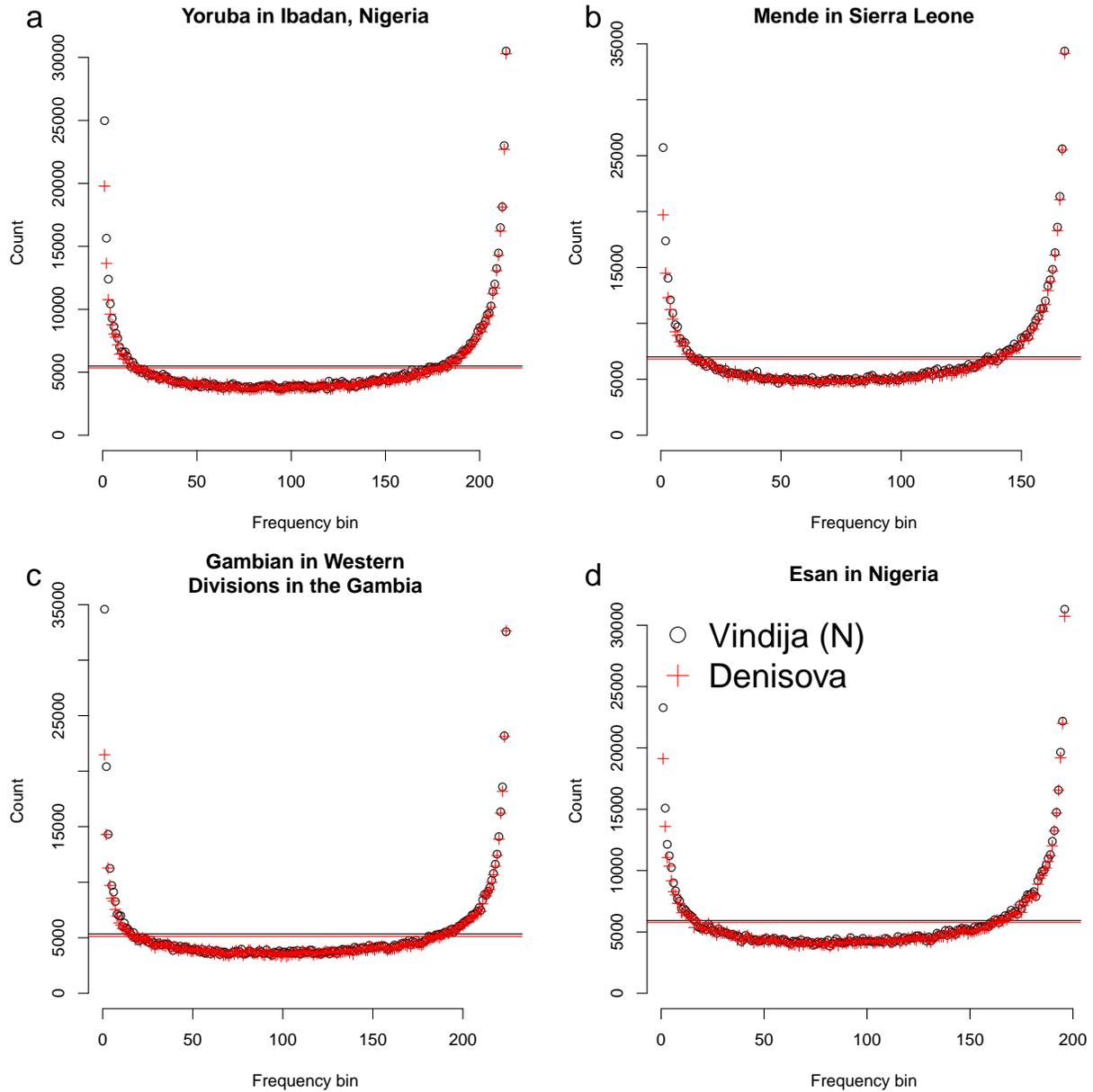


Figure S4: CSFS from 1000G Phase 3 data across all African populations included in the dataset. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4). The horizontal lines denote the mean count across all bins for both the Neanderthal (black) and Denisova (red) spectra.

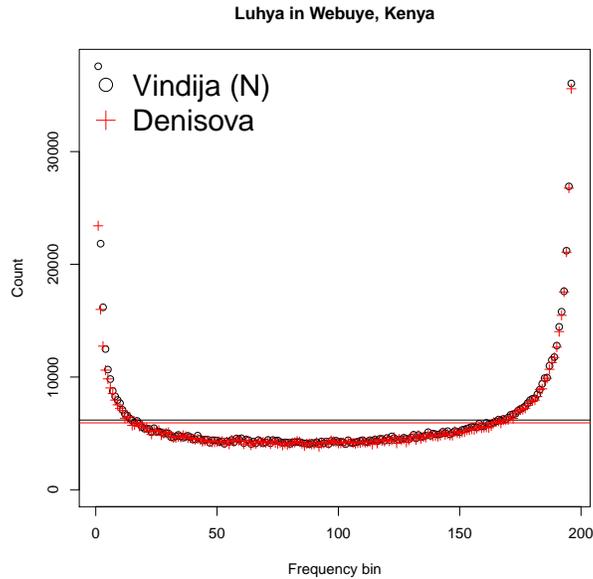


Figure S5: CSFS from 1000G Phase 3 data the Luhya population. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (15). The horizontal lines denote the mean count across all bins for both the Neanderthal (black) and Denisova (red) spectra.

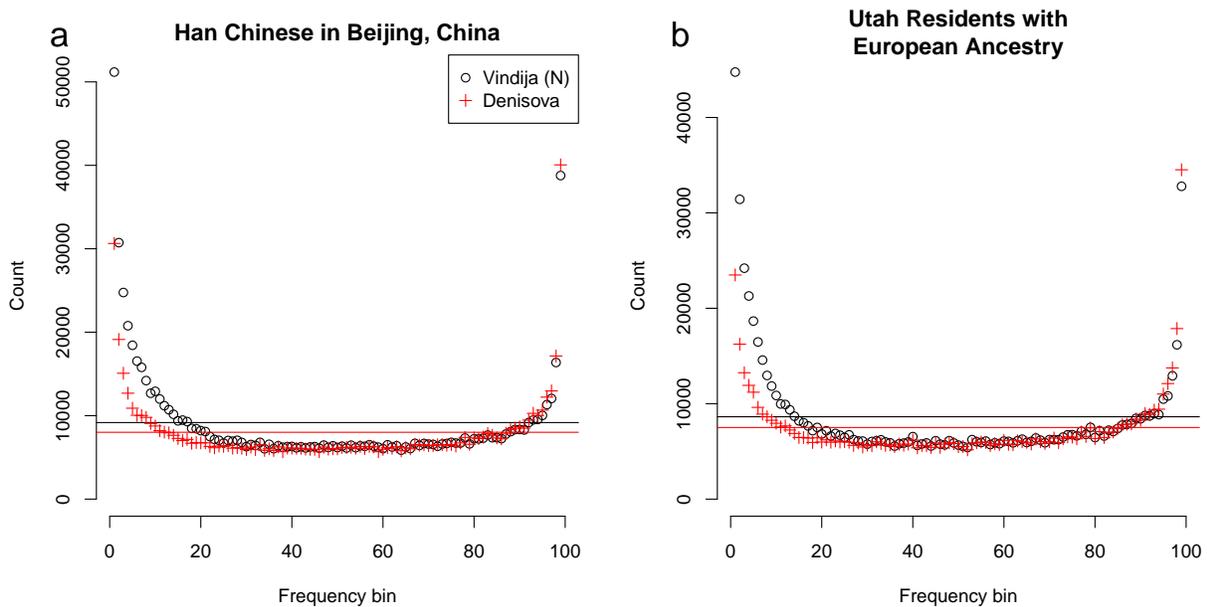


Figure S6: Conditional site frequency spectra (CSFS) from 1000 Genomes Phase 3 data. We show the number of sites in each frequency class in a sample of 50 individuals from each population, conditioned on observing a derived mutation in the Vindija Neanderthal (black points) and the Denisova (red plus). The lines denote the mean counts, which is the expectation under panmixia. (a) Han Chinese in Beijing (CHB) (b) European (CEU)

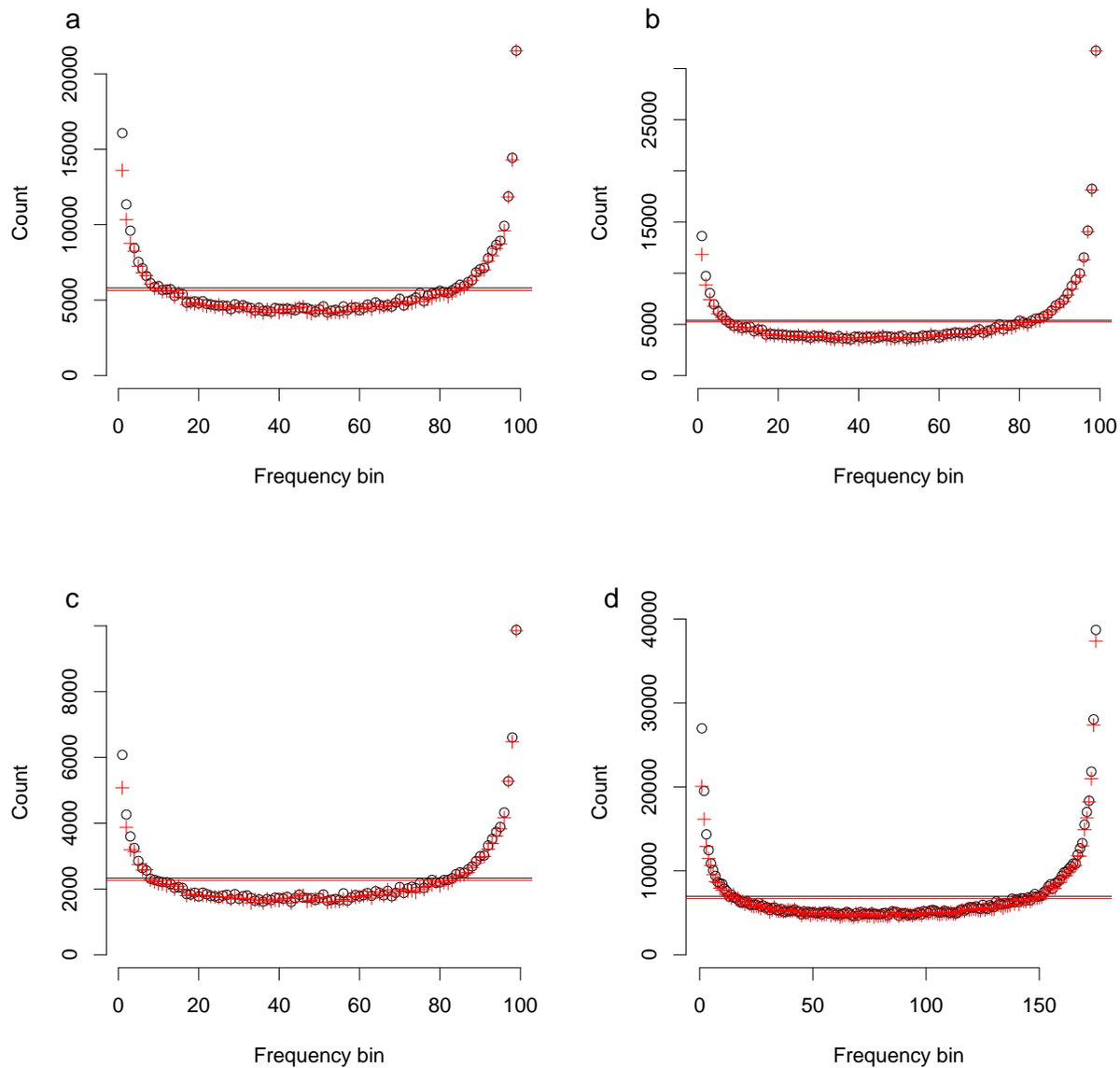


Figure S7: CSFS calculated for YRI with (a) CpG sites removed (b) transitions only (c) transversions only (d) Phase 1 1000 Genomes data (88 diploid individuals). Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4). The horizontal lines denote the mean count across all bins for both the Neanderthal (black) and Denisova (red) spectra.



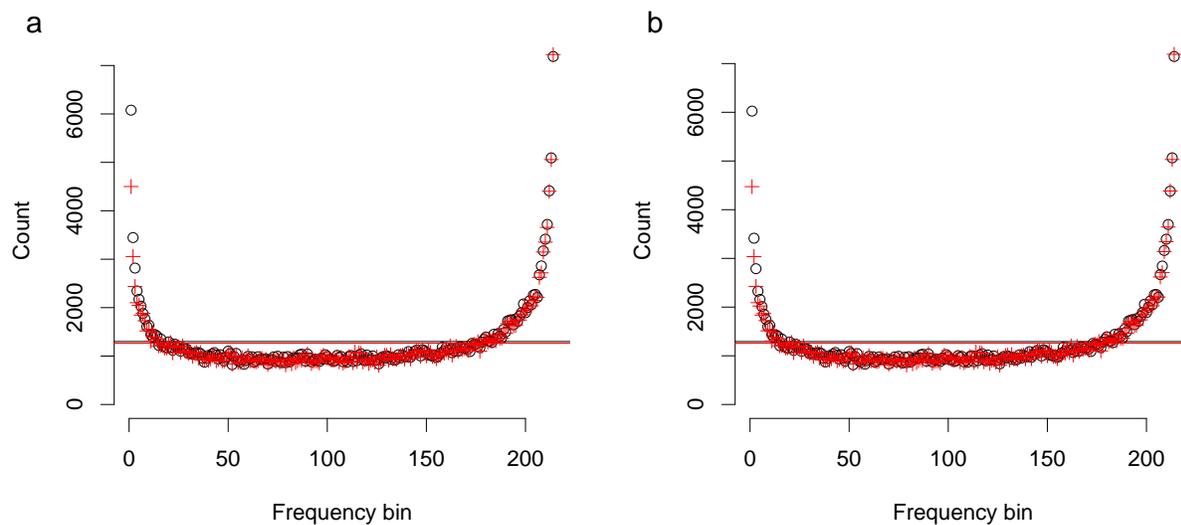


Figure S8: The CSFS in YRI restricted to sites where the archaic (Neanderthal or Denisovan) has a genotype quality (GQ) (a)  $\geq 30$  (b)  $\geq 50$ . Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4). The horizontal lines denote the mean count across all bins for both the Neanderthal (black) and Denisova (red) spectra.

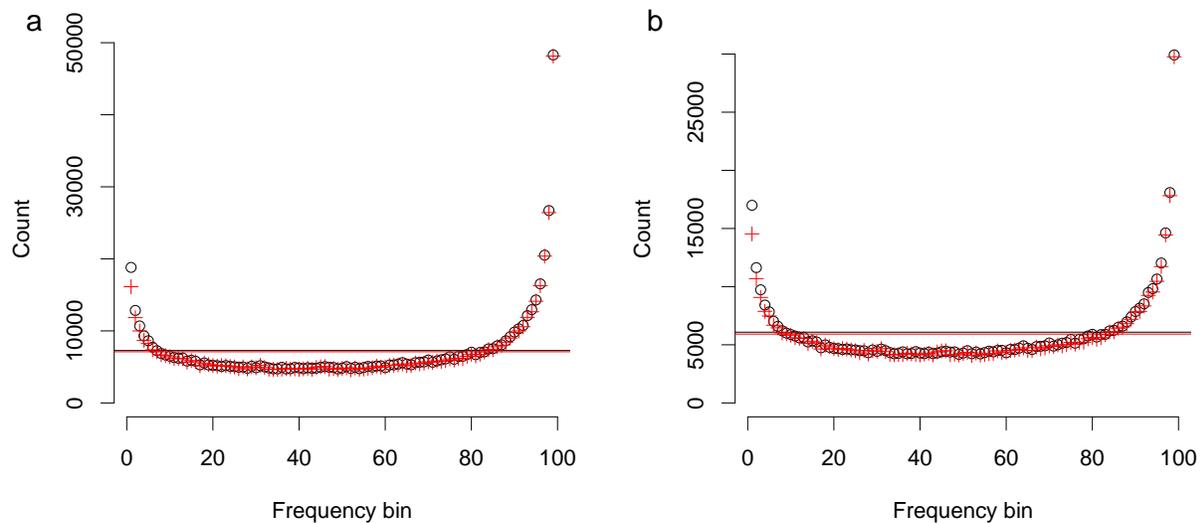


Figure S9: The CSFS in YRI when using alternate sources for the ancestral allele. (a) The CSFS in YRI using the chimpanzee genome to polarize the ancestral allele. (b) The CSFS in YRI restricting to sites where the chimpanzee and orangutan genomes have the same allele, which is used as the ancestral allele. These results are nearly identical to results using the EPO alignment to polarize the ancestral allele although the total number of sites is reduced. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4). The horizontal lines denote the mean count across all bins for both the Neanderthal (black) and Denisova (red) spectra.

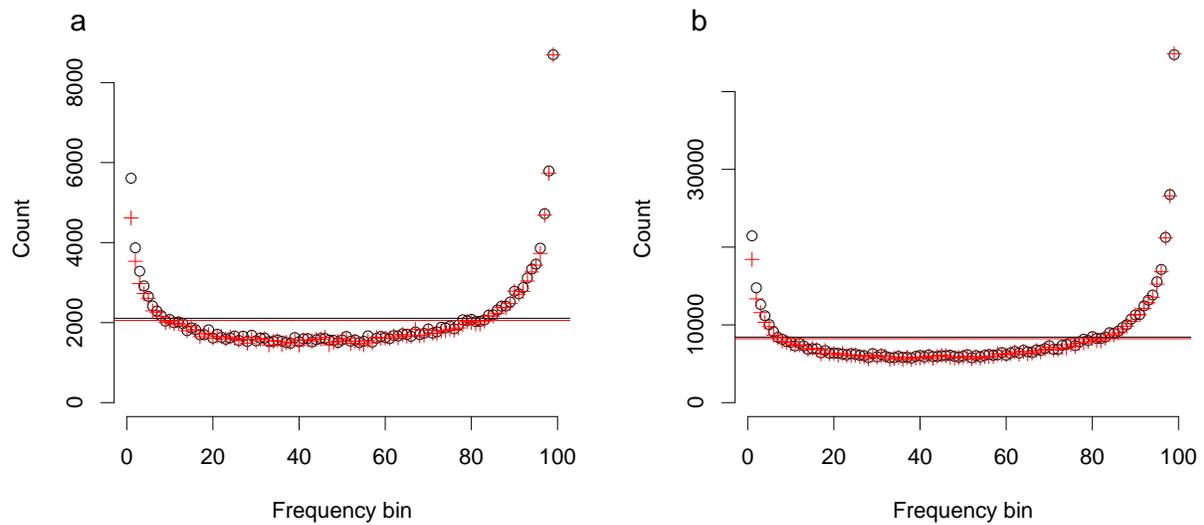


Figure S10: The CSFS in YRI when controlling for biased gene conversion and background selection. (a) The CSFS in YRI excluding mutations that may be affected by biased gene conversion. We excluded weak to strong and strong to weak mutations and recomputed the CSFS. (b) The CSFS in YRI in neutral regions. We subset to neutral sites, defined as a background selection statistic  $\geq 800$  from [McVicker et al., 2009] (40). The horizontal lines denote the mean count across all bins for both the Neanderthal (black) and Denisova (red) spectra.

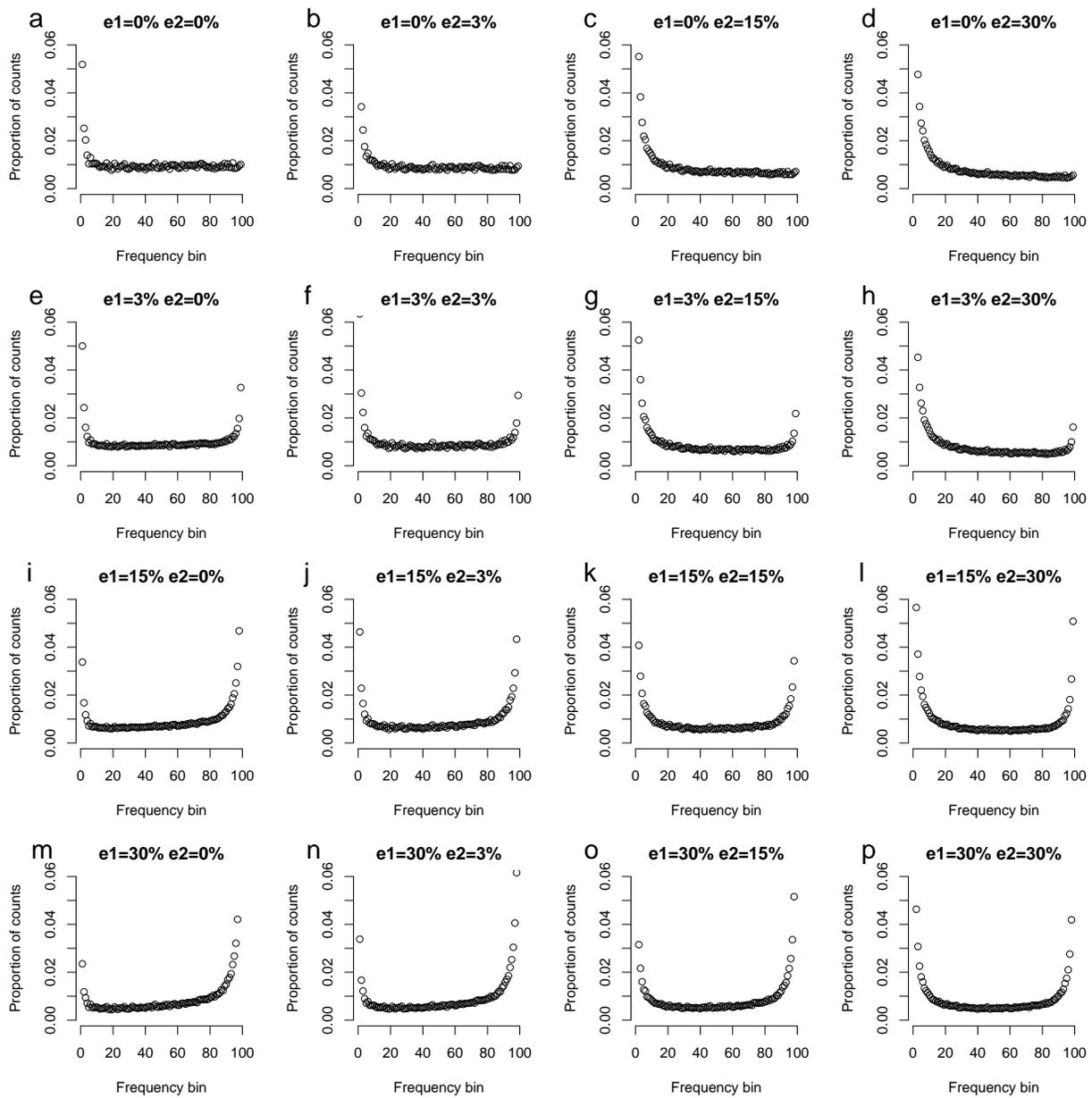


Figure S11: Simulations of the baseline model (Section S1) with both ancestral misidentification ( $e_1$ ) and genotyping error in the archaic ( $e_2$ ). Model likelihoods are presented in table S2

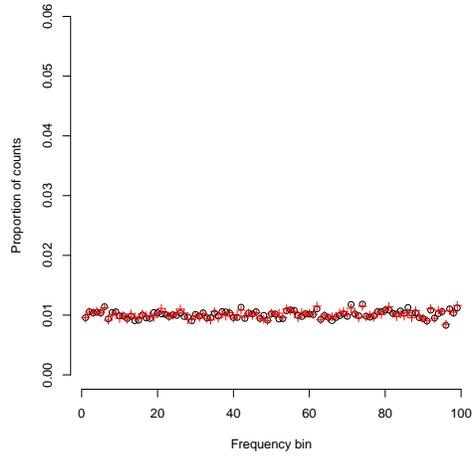


Figure S12: Mutation rate and recombination rate variation: We simulated data from the base model (Section S1) in 1 MB chunks and varied the mutation rate and recombination rate according to mutation rates estimated using Watterson's  $\theta$  (See Section S1) and recombination rates from International HapMap Consortium et al. [2007] (26). Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4).

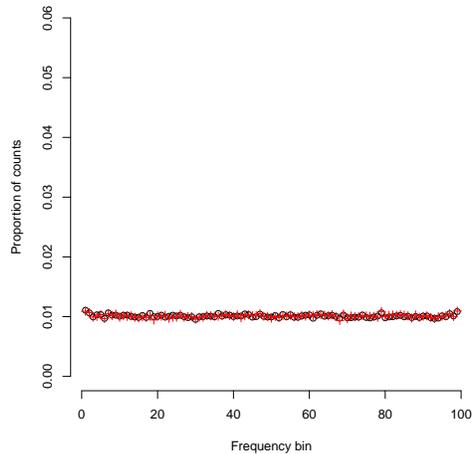


Figure S13: Simulations of the demographic model inferred from Hsieh et al. [2016] (19) relating the Yoruba, Baka, and Biaka populations. This model has gene flow between the three populations dating back to  $\simeq 40$ Ky B.P. This gene flow is insufficient to produce the U-shape of the spectrum we observe in the data. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4).

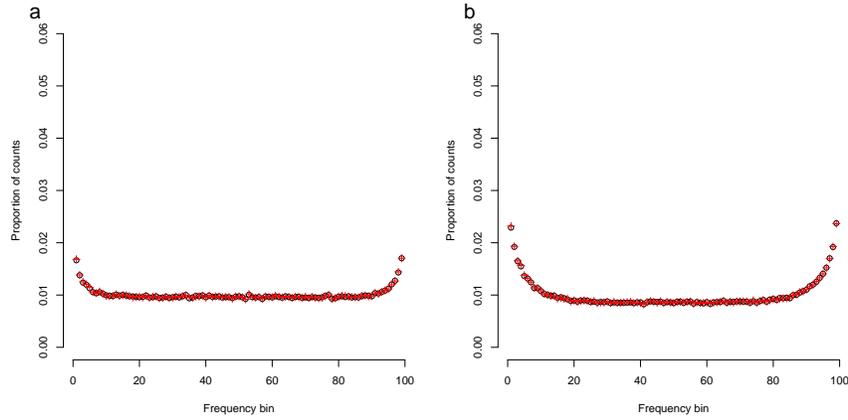


Figure S14: A model with continuous migration between lineages in Africa over the last (a) 6,000 generations ( $\approx 175$ Ky B.P.). This model is similar to the Single Origin Range expansion with Regional Persistence model from Henn et al. [2018] (21). (b) 12,000 generations ( $\approx 350$ Ky B.P.) This model is similar to the African Multiregionalism model in Henn et al. [2018] (21). Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4).

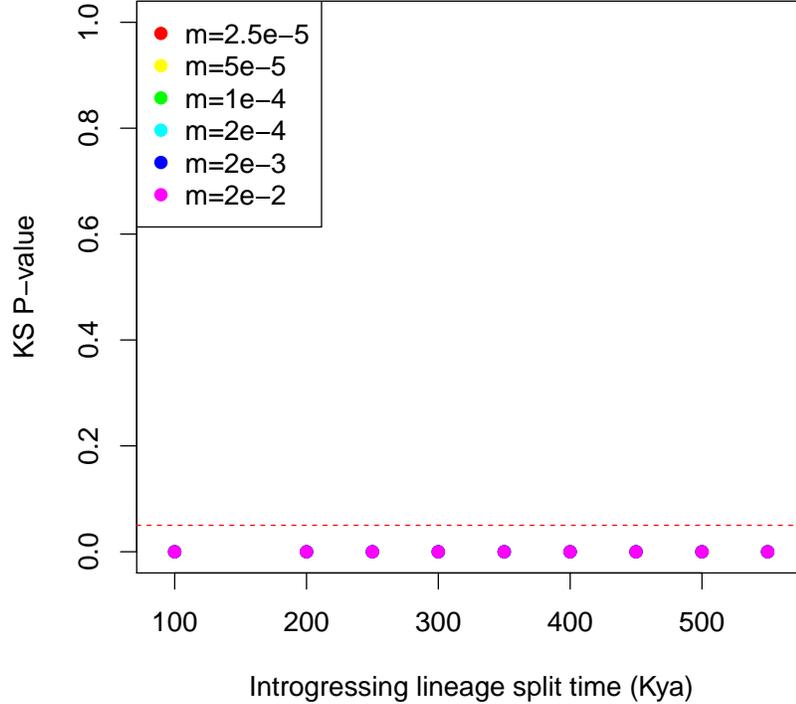


Figure S15: Models with continuous migration ( $m$  in units of migrants per generation) since the introgressing lineages lineage splits. We compute a  $p$ -value assessing the model fit to the data (See ‘Goodness of fit’ section in the main text). The red line denotes a 0.05 cutoff. A non-significant  $p$ -value indicates that we cannot reject the null hypothesis that the residuals are normally distributed, indicating a good fit of the model to the data. We see that a model with continuous migration does not fit well.

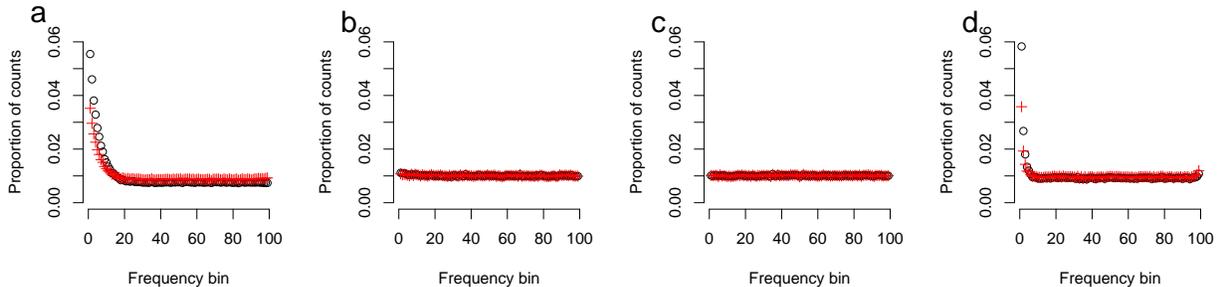


Figure S16: Current demographic models from the literature cannot explain the shape of the CSFS observed in Figure S4. (a) The model from Prüfer et al. [2017] (15), with 1% admixture from the Neanderthal branch into Yoruba 43Ky B.P. (b) the same model, but with a 0.015% admixture pulse, which is the best fitting model in Prüfer et al. [2017] (15) (S9b). (c) The model from Gravel et al. [2011] (17) with archaic lineages and admixture events included (d) The model from Petr et al. [2019] (45), which introduces Neanderthal DNA into Yoruba via migration between Europeans and Yorubans over the last 20Ky B.P. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4).

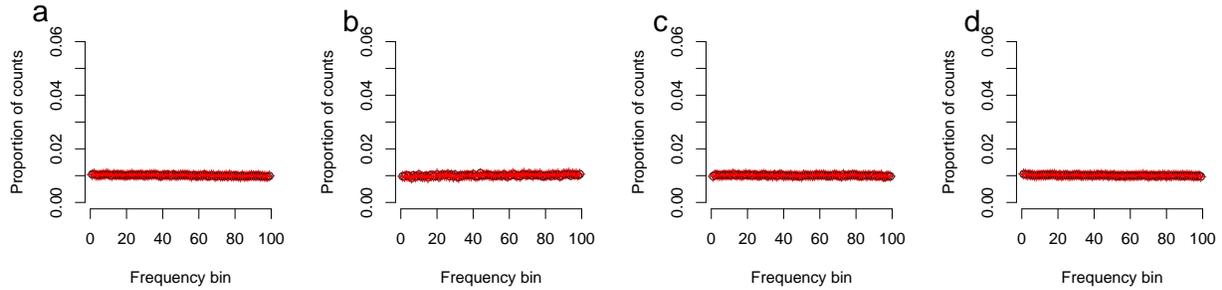


Figure S17: Models involving structure in the ancestor of modern humans and archaics cannot explain the observed CSFS. (a)  $5N_a$  population expansion in the ancestor of modern humans and archaics. (b) a  $0.2N_a$  bottleneck in the ancestor. (c) Population subdivision (840Ky B.P.) and subsequent merging in the ancestor at 550Ky B.P. (d) Population subdivision (840Ky B.P.) and continuous migration in the ancestor until 550Ky B.P. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012].

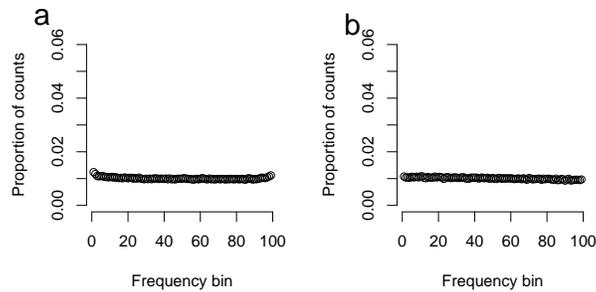


Figure S18: Models involving ancestral structure from the literature cannot explain the observed CSFS. (a) the ‘Ancient structure model’ (Figure 1B) from Yang et al. [2012], with continuous gene flow between Neanderthals and modern humans. (b) the ‘Ancient structure model’ from Sankararaman et al. [2012] with continuous gene flow between Neanderthals and modern humans.

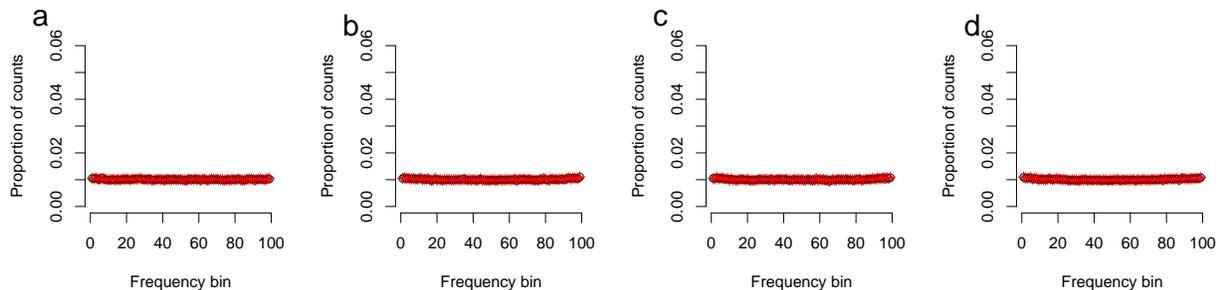


Figure S19: Model A.1: Gene flow from the modern human ancestor branch back into the modern human ancestor prior to the out of Africa event. See section S3.1 for a detailed description. Panels show different amounts of gene flow into the modern human ancestor (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4).



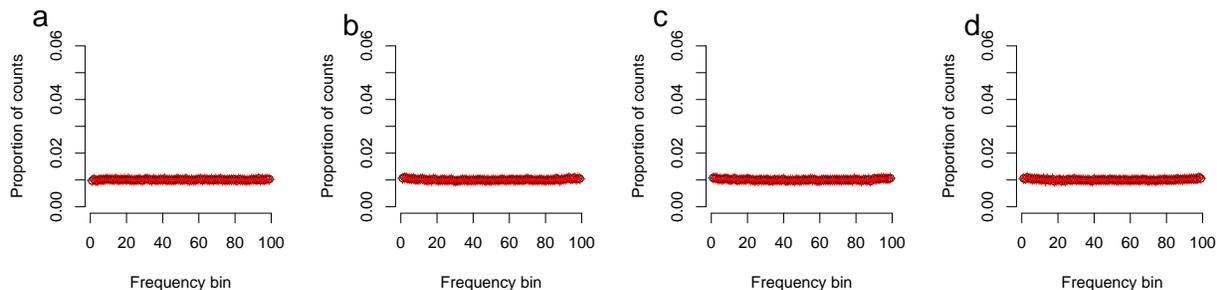


Figure S20: Model sA.1: We simulate the model in Figure S19, but only include the African, Neanderthal, and Denisovan branches (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4).

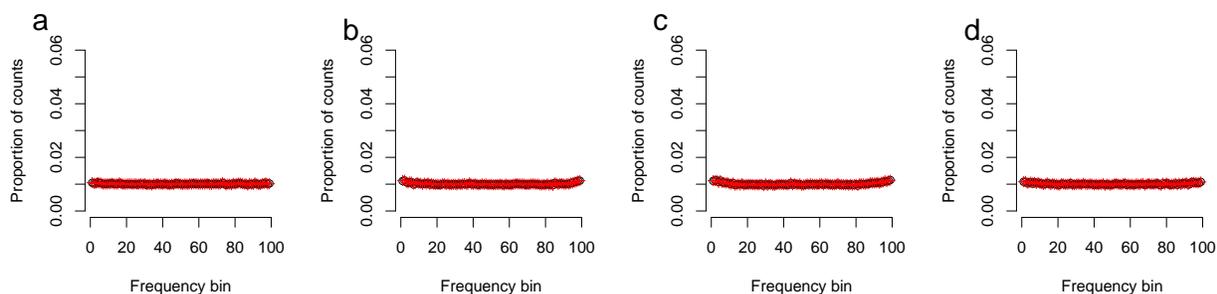


Figure S21: Model A.2: Gene flow from the modern human ancestor branch into the African branch after the out of Africa event. See section S3.2 for a detailed description. Panels show different amounts of gene flow into the African branch (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

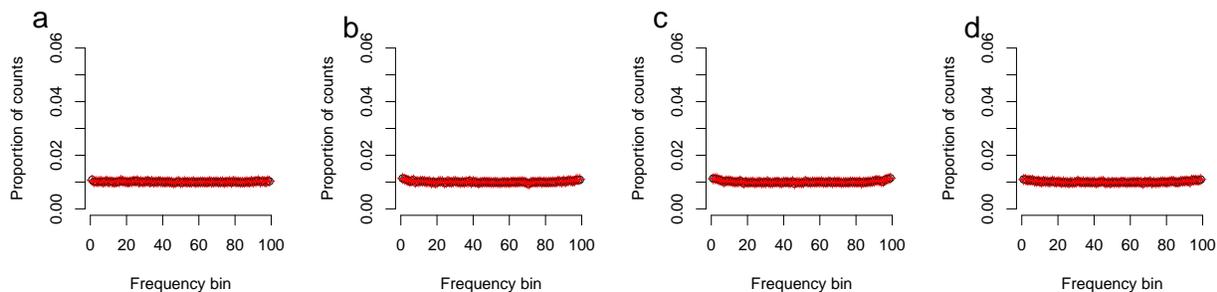


Figure S22: Model sA.2: We simulate the model in Figure S21, but only include the African, Neanderthal, and Denisovan branches (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

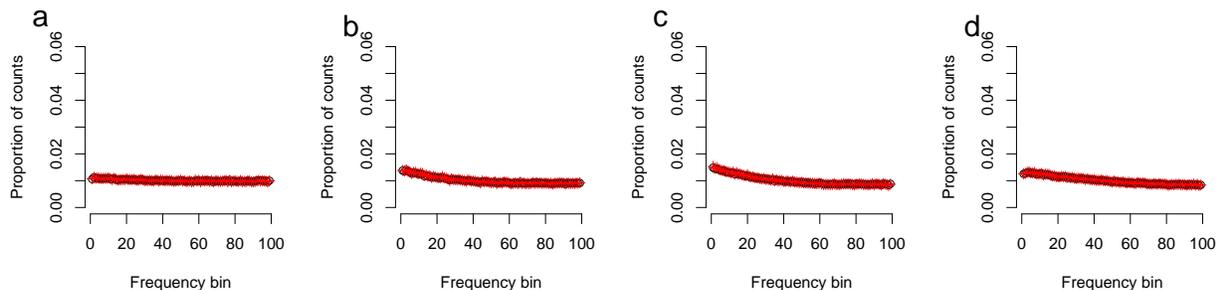


Figure S23: Model B.1: Gene flow from the archaic branch into the modern human ancestor prior to the out of Africa event. See section S3.3 for a detailed description. Panels show different amounts of gene flow into the modern human ancestor (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

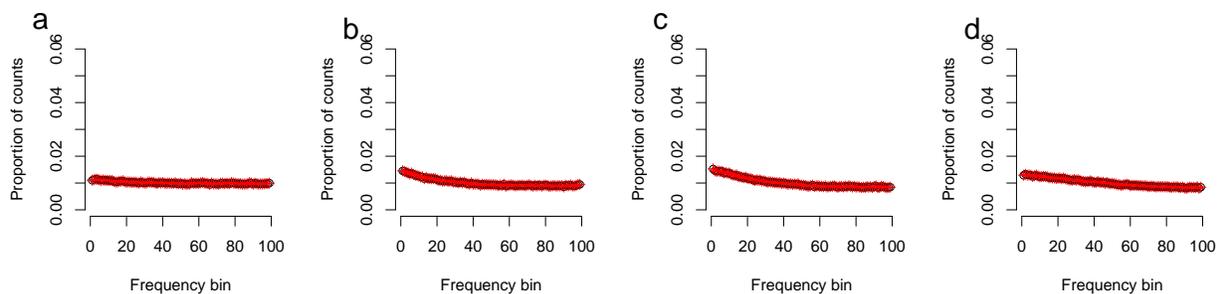


Figure S24: Model sB.1: We simulate the model in Figure S23, but only include the African, Neanderthal, and Denisovan branches (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

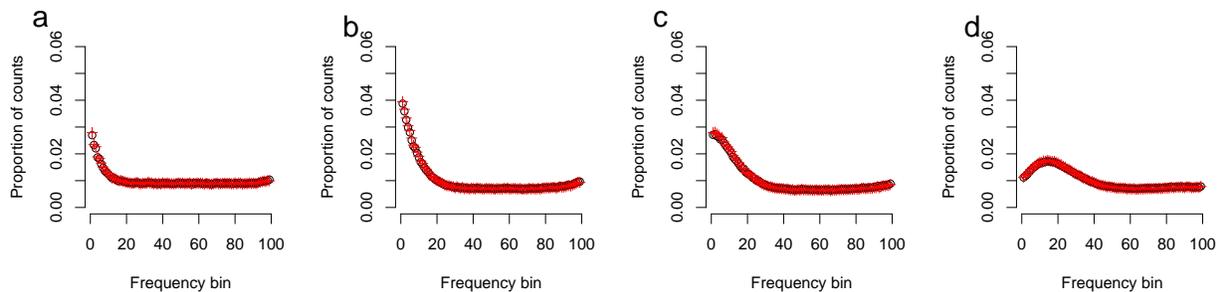


Figure S25: Model B.2: Gene flow from the archaic branch into the African branch after the out of Africa event. See section S3.4 for a detailed description. Panels show different amounts of gene flow into the African branch (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

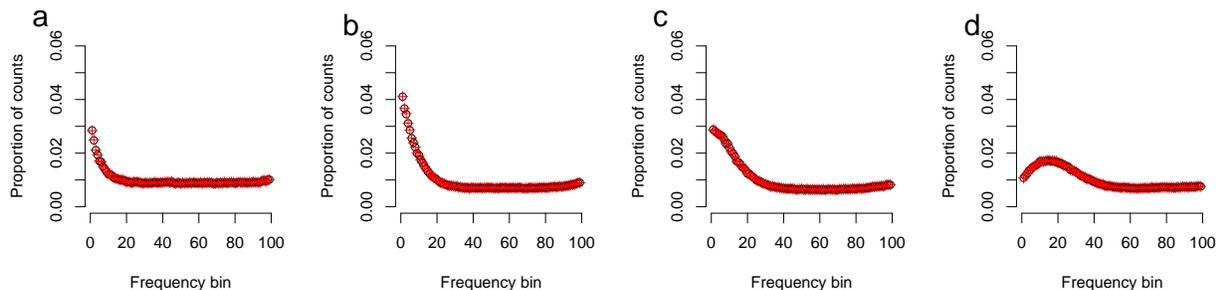


Figure S26: Model sB.2: We simulate the model in Figure S25, but only include the African, Neanderthal, and Denisovan branches (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

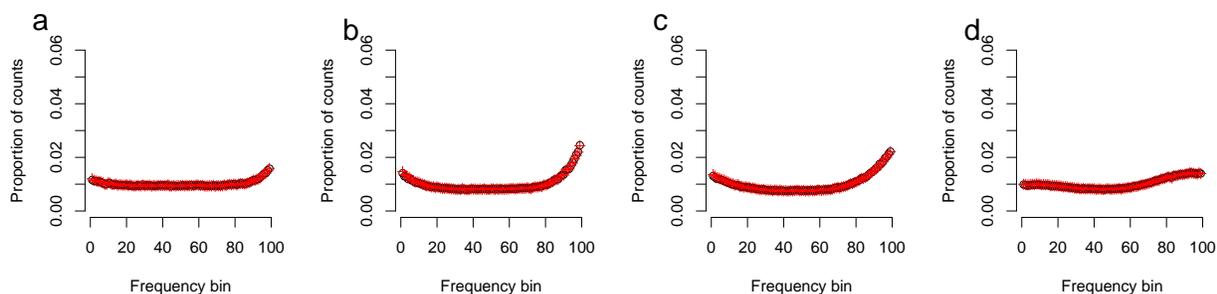


Figure S27: Model C.1: Gene flow from an unknown archaic branch into the modern human ancestor prior to the out of Africa event. See section S3.5 for a detailed description. Panels show different amounts of gene flow into the modern human ancestor (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

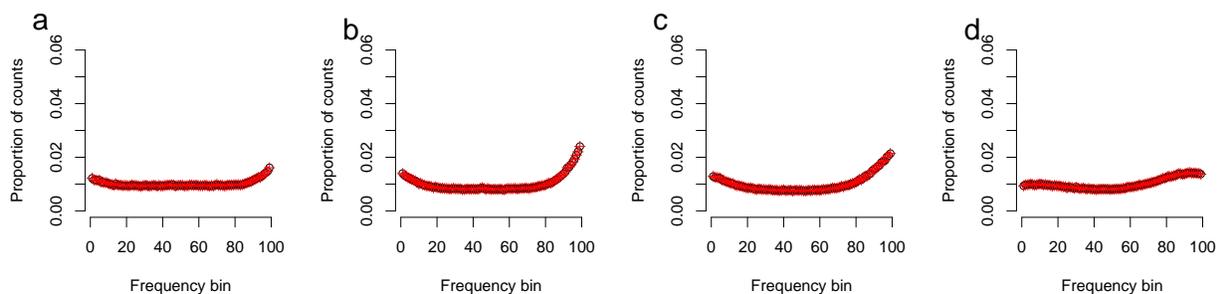


Figure S28: Model sC.1: We simulate the model in Figure S27, but only include the African, Neanderthal, and Denisovan branches (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

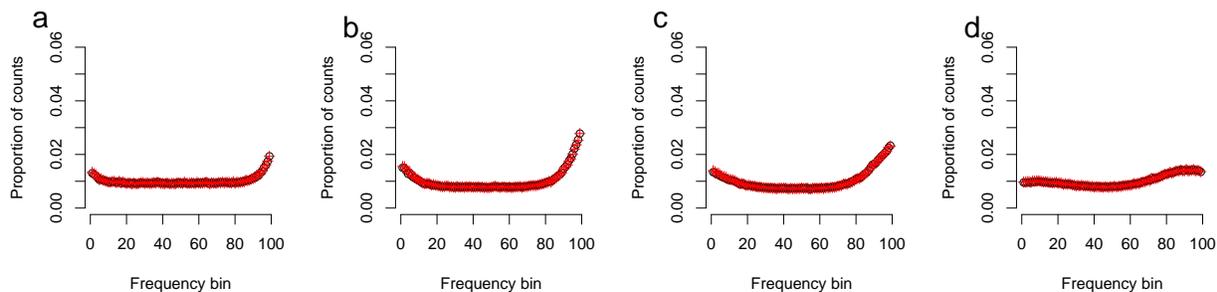


Figure S29: Model C.2: Gene flow from an unknown archaic branch into the African branch after the out of Africa event. See section S3.6 for a detailed description. Panels show different amounts of gene flow into the African branch (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

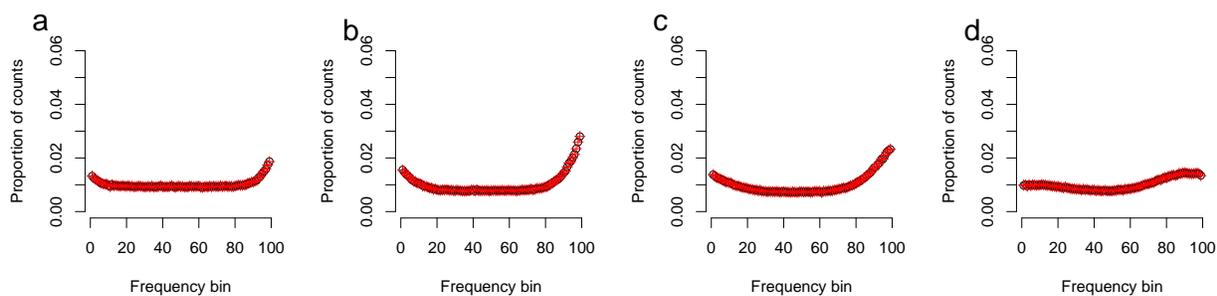


Figure S30: Model sC.2: We simulate the model in Figure S29, but only include the African, Neanderthal, and Denisovan branches (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

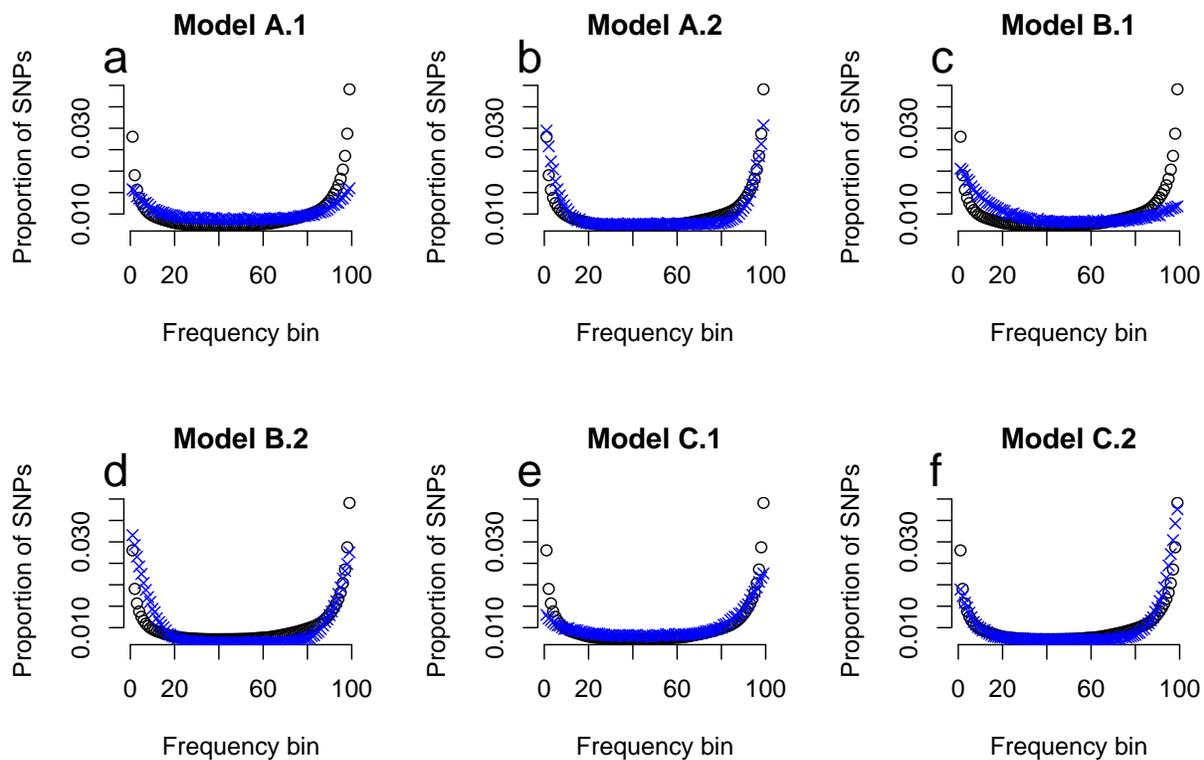


Figure S31: Simulations of the best fitting parameters for models A, B, C (Section S3). Blue crosses denote the simulated spectra and black circles denote the observed  $CSFS_{YRI,N}$ .

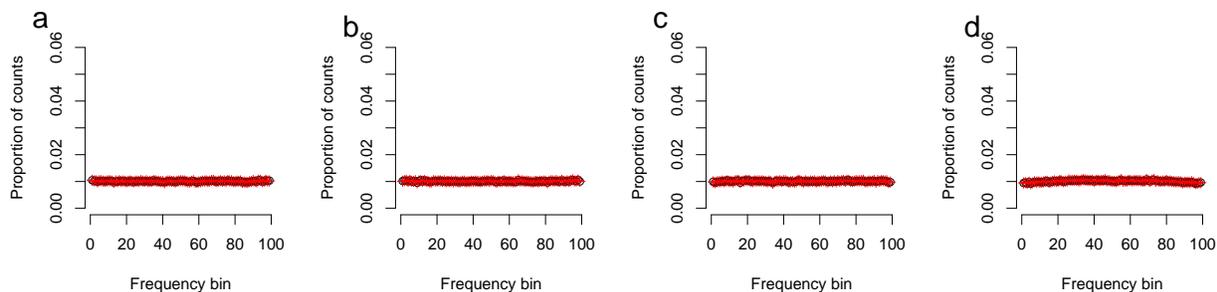


Figure S32: Model A.2 with a population size of  $0.01N_a$  in the introgressing population. The panels represent different levels of gene flow (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

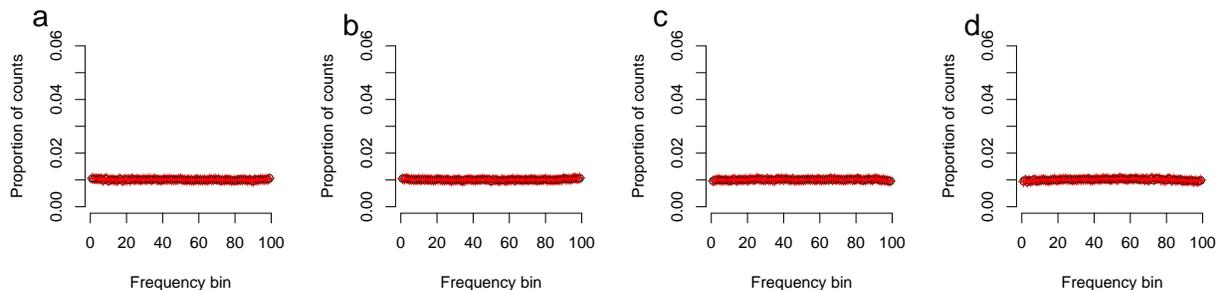


Figure S33: Model A.2 with a population size of  $1 \times 10^{-4} N_a$  in the introgressing population. The panels represent different levels of gene flow (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

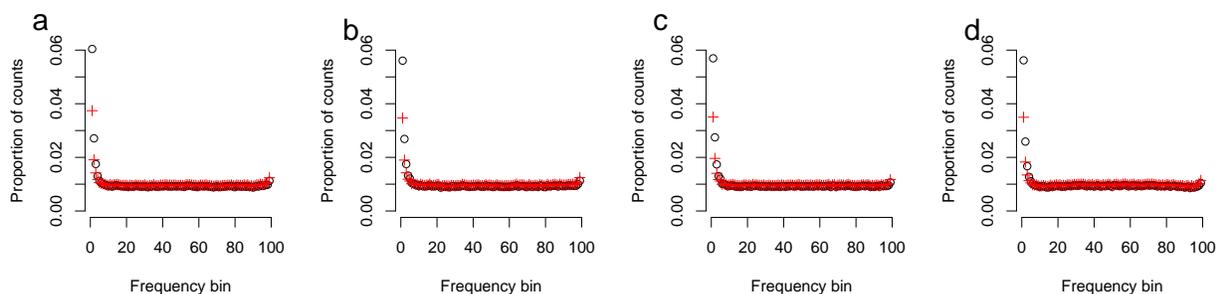


Figure S34: Model A.2 with a population size of  $1 \times 10^{-4} N_a$  in the introgressing population and migration between CEU and YRI over the last 20Ky B.P. The panels represent different levels of gene flow (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

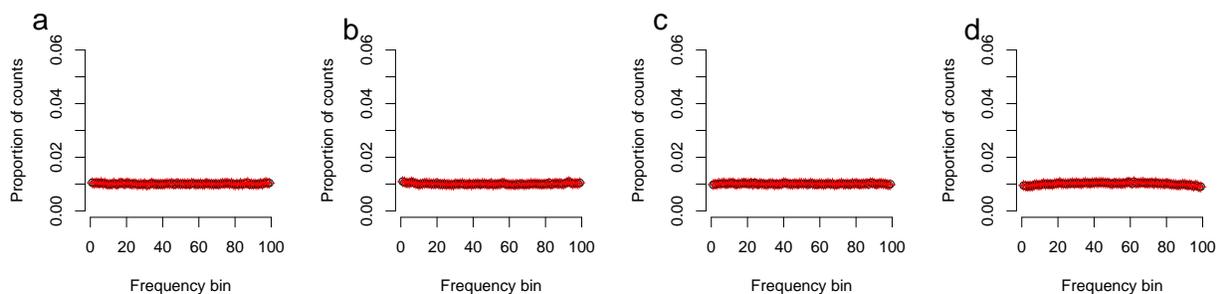


Figure S35: Model A.2 with a population size of  $1 \times 10^{-5} N_a$  in the introgressing population, which branches off 200Ky B.P. The panels represent different levels of gene flow (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

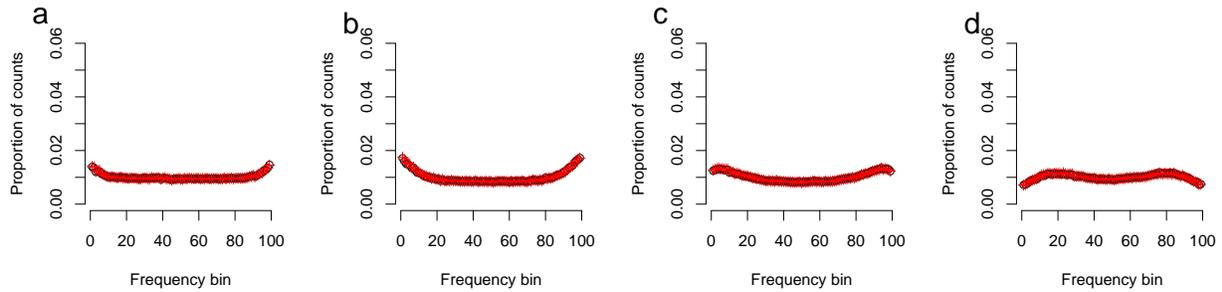


Figure S36: Model A.2 where the introgressing population splits at the same time as the archaic population (550Ky B.P.) with a population size of  $0.01N_a$ . The panels represent different levels of gene flow (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

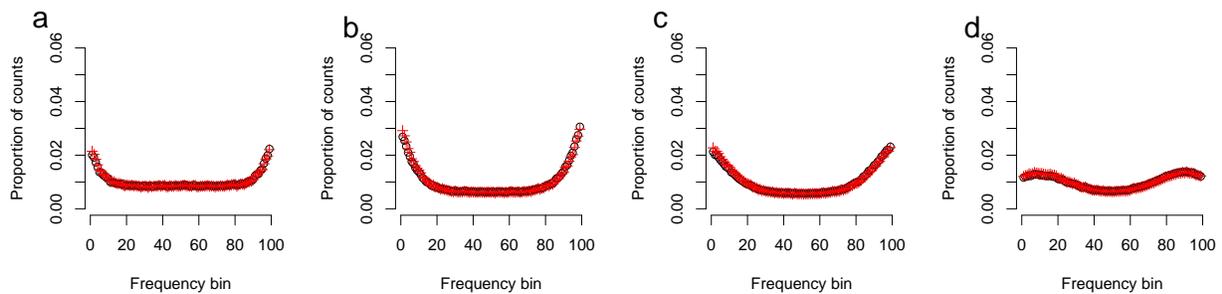


Figure S37: Model A.2 where the introgressing population splits at the same time as the archaic population, 765Ky B.P. This simulation shows that uncertainty in the split time between archaics and modern humans cannot explain the CSFS. The panels represent different levels of gene flow (a) 1% (b) 5% (c) 10% (d) 20%. Black circles denote the CSFS conditioned on the Vindija Neanderthal [Prüfer et al., 2017] (15) and red pluses denote the CSFS conditioned on the high coverage Denisovan [Meyer et al., 2012] (4)

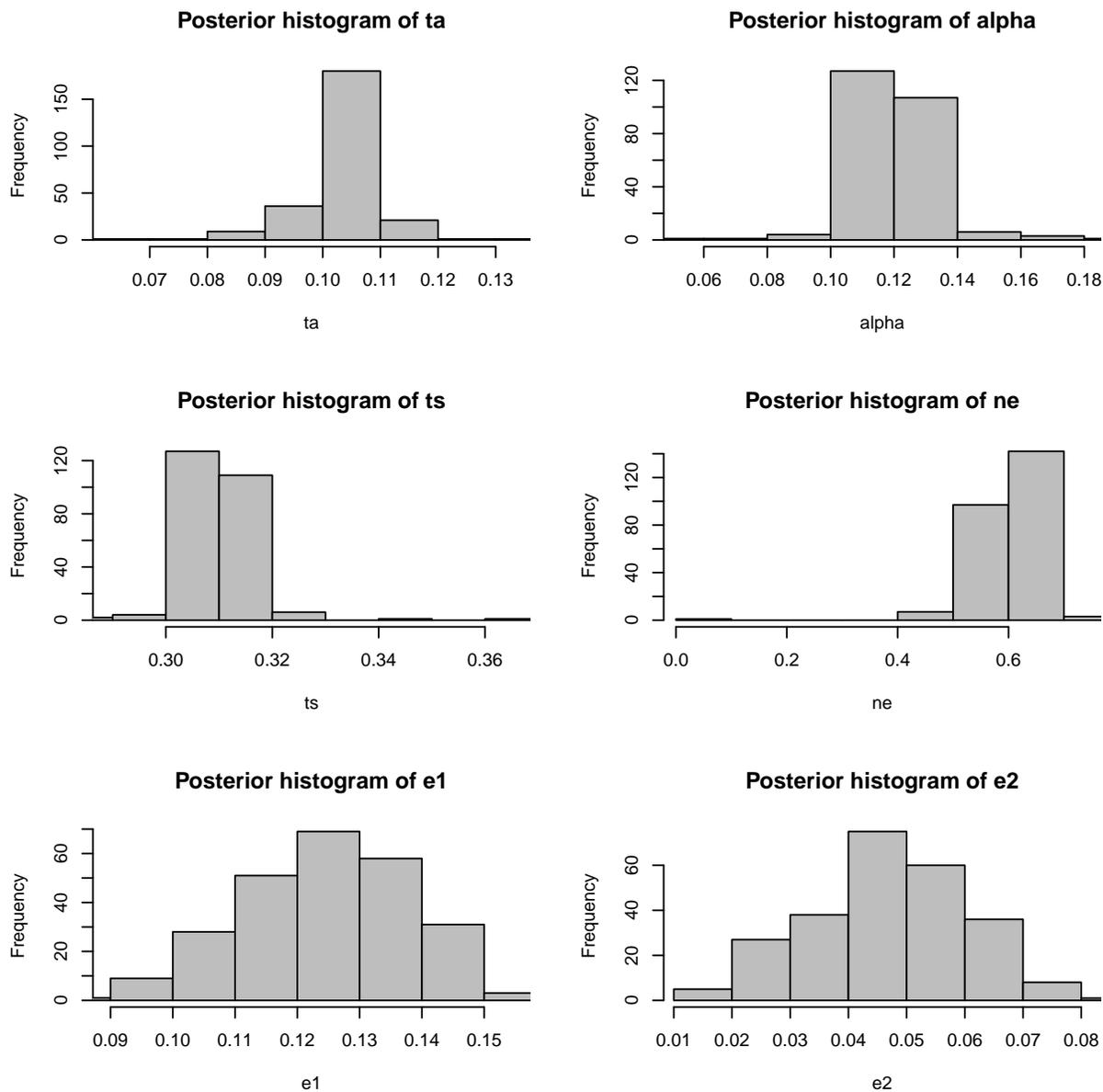


Figure S38: Parameter estimates using ABC for Model A.1 including ancestral misidentification (e1) and genotyping error in the archaic (e2). These estimates are made on a per-base pair rate. 'ta' is the time of admixture, scaled by  $4N_a$ , where  $N_a$  is the ancestral population size (10,000). 'alpha' is the admixture fraction. 'ts' is the split time scaled by  $4N_a$  and 'ne' is the effective population size of the introgressing lineage.



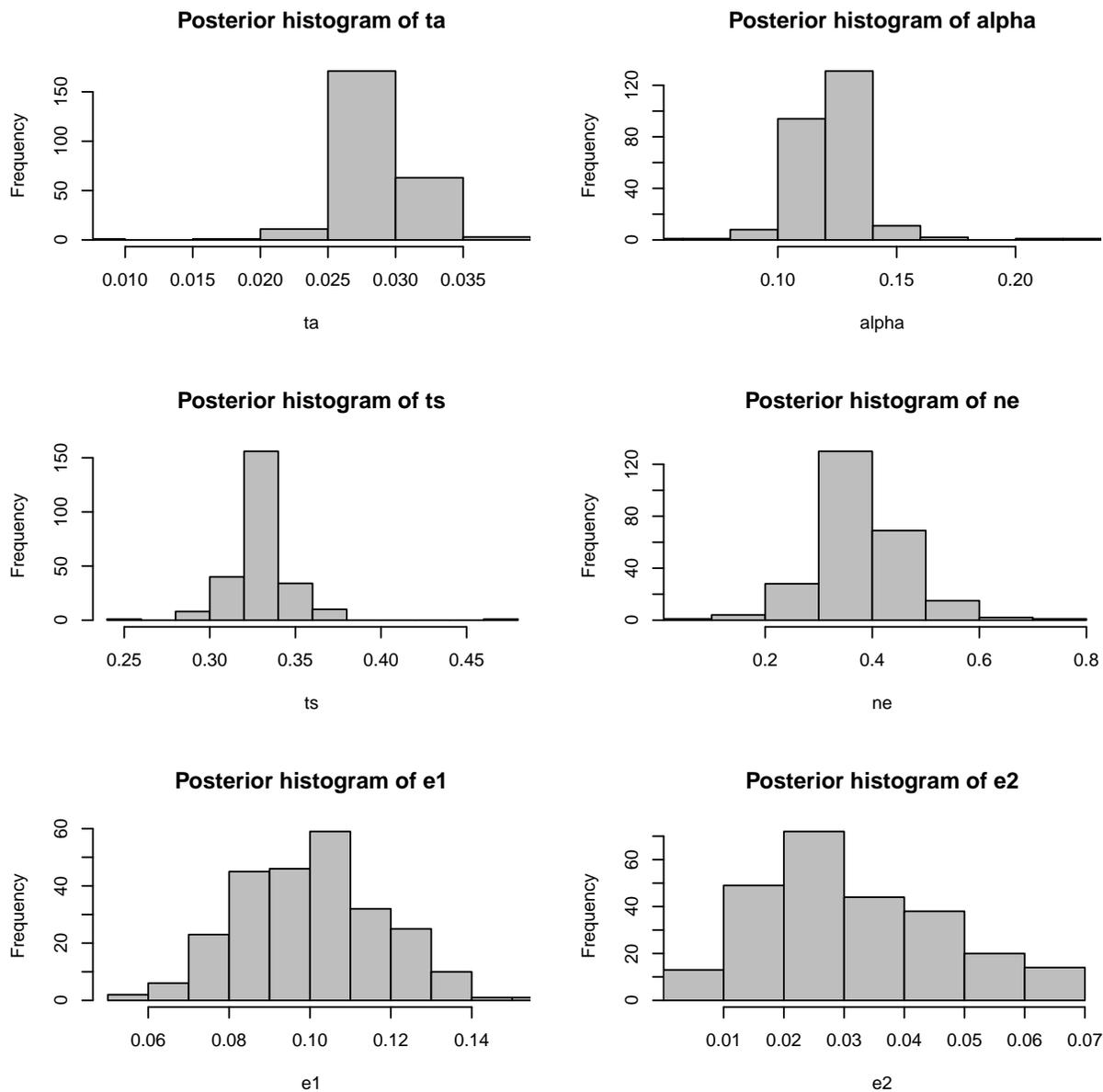


Figure S39: Parameter estimates using ABC for Model A.2 including ancestral misidentification (e1) and genotyping error in the archaic (e2). These estimates are made on a per-base pair rate. ‘ta’ is the time of admixture, scaled by  $4N_a$ , where  $N_a$  is the ancestral population size (10,000). ‘alpha’ is the admixture fraction. ‘ts’ is the split time scaled by  $4N_a$  and ‘ne’ is the effective population size of the introgressing lineage.

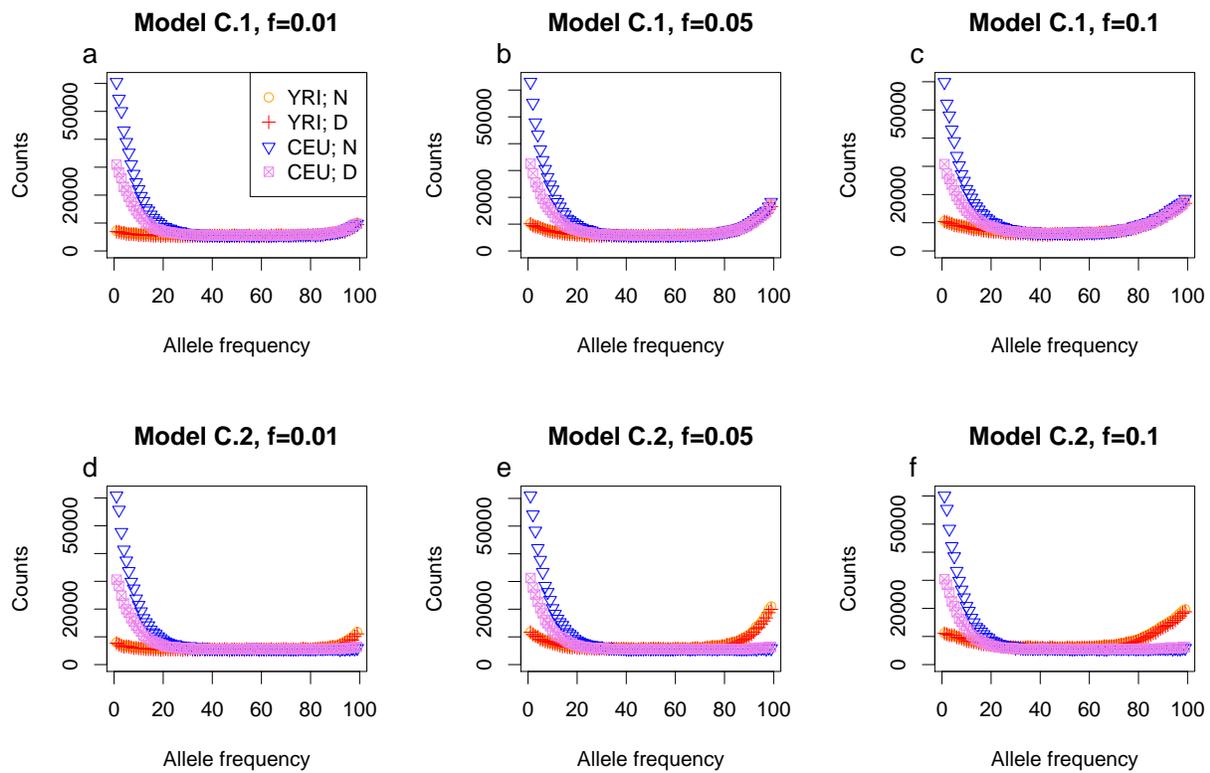


Figure S40: Marginalized joint CSFS of YRI and CEU from simulations. We contrast introgression into the modern human branch before the Out-of-Africa split (C.1 – top row) against introgression into the African branch specifically (C.2 – bottom row). We see that model C.1 produces an increase in high frequency derived alleles for both CEU and YRI, while C.2 only produces that signal in YRI.

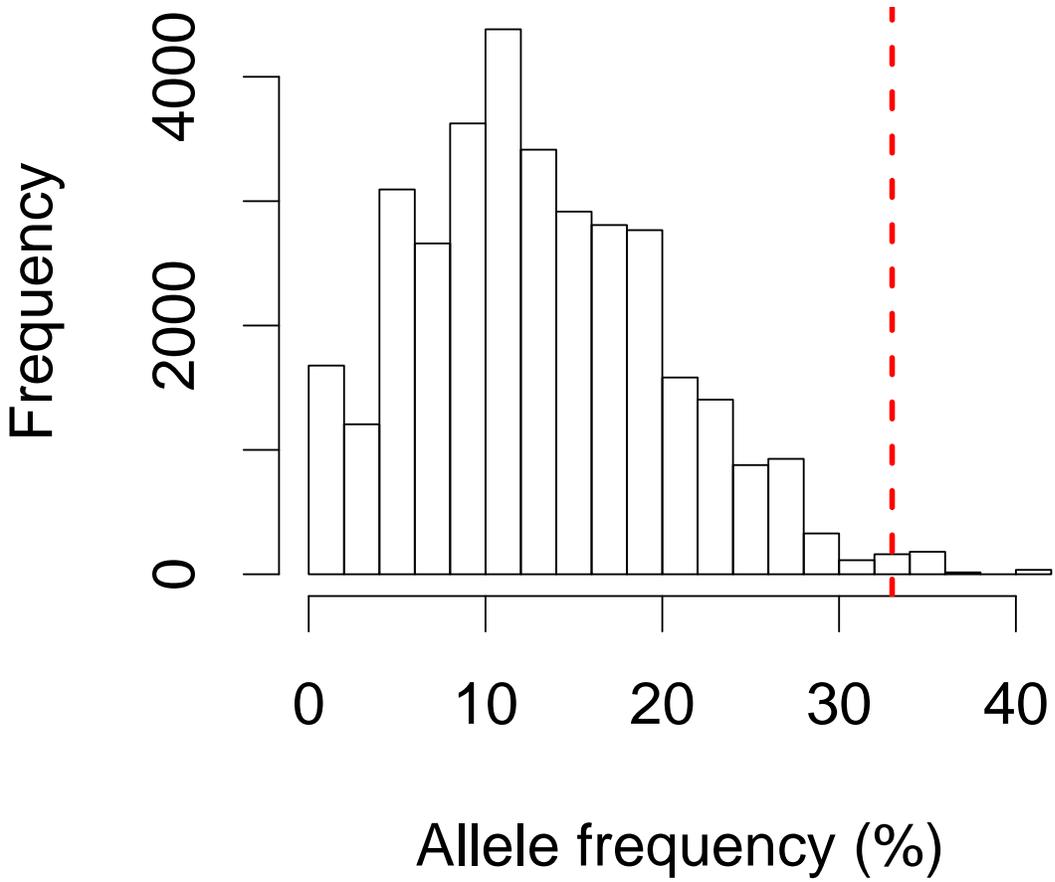


Figure S41: Distribution of allele frequencies for neutral archaic SNPs from model C with 13% introgression and an introgression time of 42Ky B.P. The red line denotes the 99% quantile, which corresponds to an allele frequency of 33%.

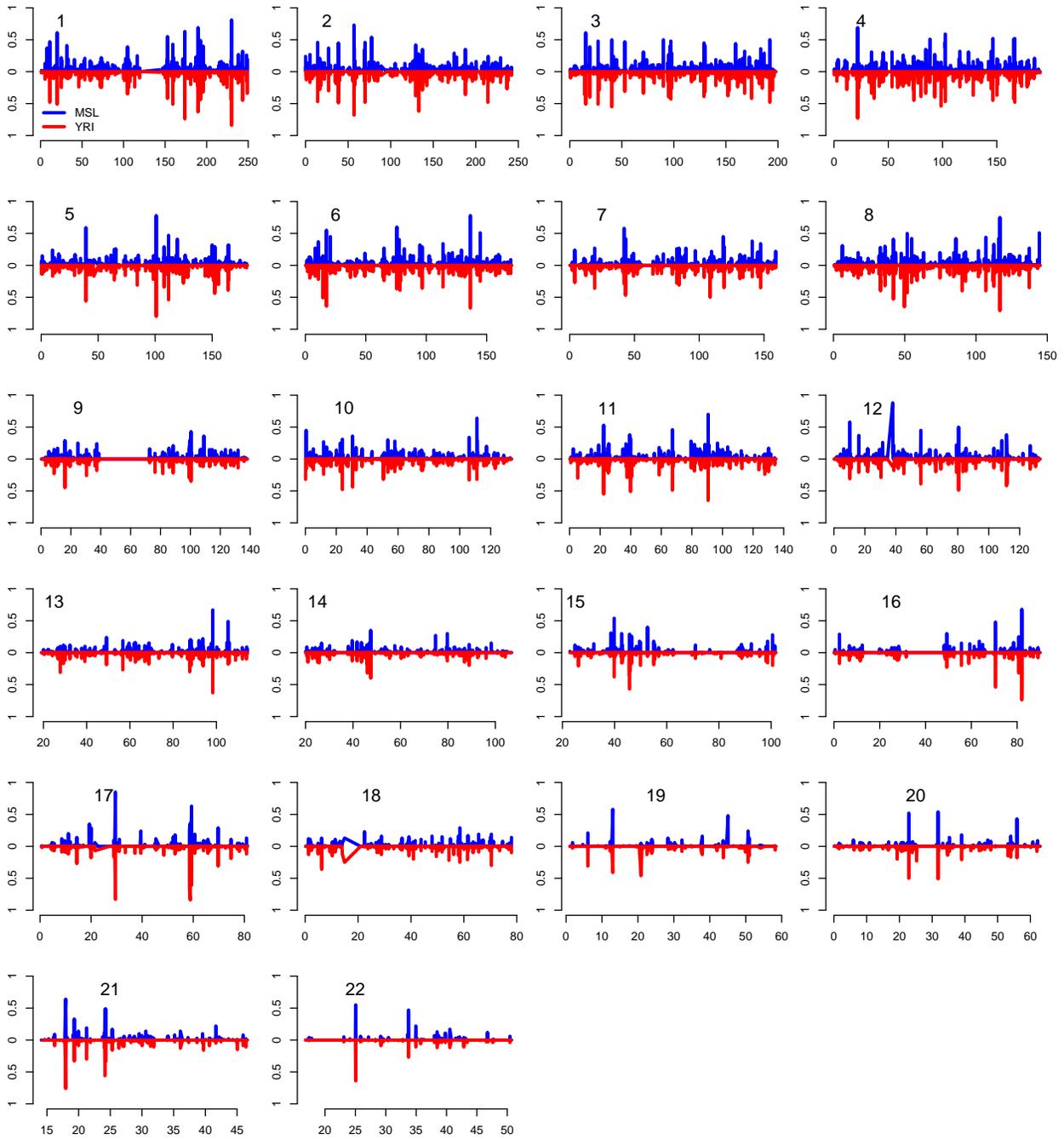


Figure S42: Archaic segment frequency map for MSL and YRI. Spearman's  $\rho$  between the frequencies of the two maps is 0.71 ( $p < 2 \times 10^{-16}$ ).

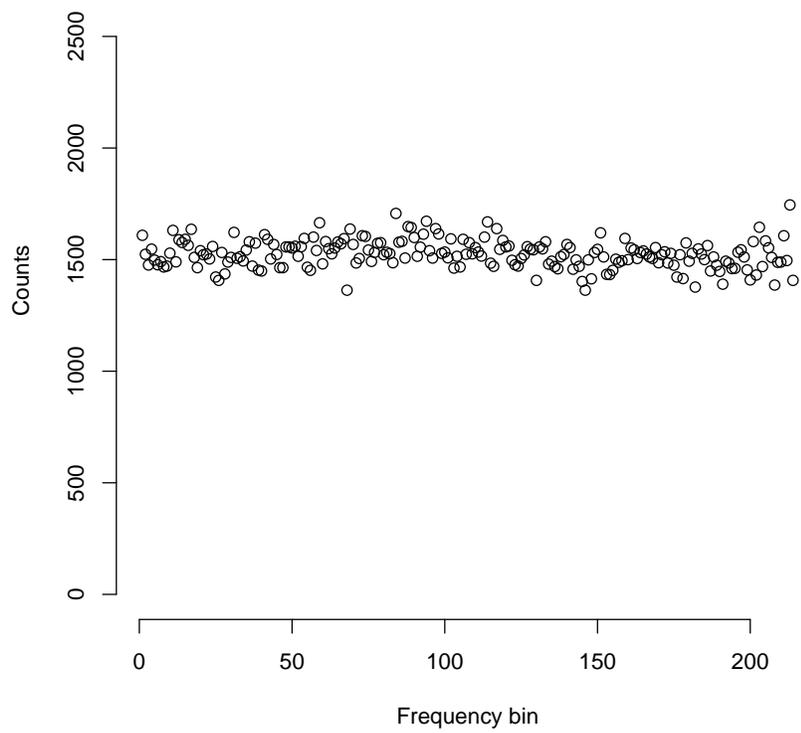


Figure S43: Conditional site frequency spectrum (CSFS) from the baseline model allowing for recurrent mutations.

| <b>Model</b>                                | <b>Description</b>  |
|---|---|
| A.1   | Gene flow into the modern human ancestor from a lineage that splits from the modern human ancestor 6K generations ago   |
| A.2   | Gene flow into the African branch from a lineage that splits from the modern human ancestor 6K generations ago  |
| A.2 contraction                             | Model A.2 with an effective population size of 100 in the introgressing branch  |
| A.2 contraction 2                           | Model A.2 with an effective population size of 1 in the introgressing branch  |
| A.2 contraction 2<br>CEU-YRI migra-<br>tion | Model A.2 with an effective population size of 100 in the introgressing branch and migration between YRI and CEU over the last 20Kya  |
| A.2 contraction 3                           | Model A.2 with an effective population size of 0.1 in the introgressing branch  |
| A.2 trifurcation                            | Model A.2 with the introgressing lineage splitting off the same time as the archaic branch ( 550Kya) and an effective population size of 100  |
| B.1   | Gene flow into the modern human ancestor from a lineage that splits from the archaic branch (ancestor of Denisovan and Neanderthal) 16K generations ago   |
| B.2   | Gene flow into the African branch from a lineage that splits from the archaic branch (ancestor of Denisovan and Neanderthal) 16K generations ago  |
| Baseline with errors                        | A model including the major demographic events (see Figure 1) with ancestral misidentification error and archaic sequencing error   |
| C.1   | Gene flow into the modern human ancestor from a lineage that splits from the modern human ancestor 24K generations ago  |
| C.2   | Gene flow into the African branch from a lineage that splits from the modern human ancestor 24K generations ago   |
| Continuous Migra-<br>tion                   | Continuous bidirectional migration between the African branch and a lineage that splits 174Kya over the past 174Kya   |
| Continuous Migra-<br>tion 2                 | Continuous bidirectional migration between the African branch and a lineage that splits 350Kya over the past 350Kya   |
| sA.1  | A simplified version of model A.1 simulating on the African Neanderthal and Denisovan branches  |
| sA.2  | A simplified version of model A.2 simulating on the African Neanderthal and Denisovan branches  |
| sB.1  | A simplified version of model B.1 simulating on the African Neanderthal and Denisovan branches  |
| sB.2  | A simplified version of model B.2 simulating on the African Neanderthal and Denisovan branches  |
| sC.1  | A simplified version of model C.1 simulating on the African Neanderthal and Denisovan branches  |
| sC.2  | A simplified version of model C.2 simulating on the African Neanderthal and Denisovan branches  |
| SIM004                                      | The model from Pruefer et al 2017; including gene flow from the super archaic into the Denisovan; modern humans into Neanderthals; Neanderthal into Europeans; and a small contribution from the Neanderthal directly to the African branch |
| SIM005                                      | The model from Pruefer et al 2017; with migration between CEU and YRI over the last 20;000 years  |
| SIM006                                      | The model from Pruefer et al 2017; with a 5x population expansion in the modern human-archaic ancestor  |
| SIM007                                      | The model from Pruefer et al 2017; with a 5x population bottleneck in the modern human-archaic ancestor   |
| SIM008                                      | The model from Pruefer et al 2017; with a population that branches off from and rejoins the modern human-archaic ancestor   |
| SIM009                                      | The model from Pruefer et al 2017; with a population that branches off from the modern human-archaic ancestor and includes bidirectional migration between the lineages   |
| SIM010                                      | The model from Pruefer et al 2017; with a population that branches off from and rejoins the modern human-archaic ancestor with an effective population size of 100  |
| SIM012                                      | The ancestral structure model from Yang et al 2012  |
| SIM013                                      | The ancestral structure model from Sankararaman et al 2012  |

Table S1: Description of the models examined in this work.

| Model                      | Ancestral misidentification rate (e1) | Genotype error in archaic (e2) | $\Delta\mathcal{L}\mathcal{L}$ (N) |
|----------------------------|---------------------------------------|--------------------------------|------------------------------------|
| Baseline model with errors | 0.03                                  | 0.03                           | -6058                              |
| Baseline model with errors | 0.03                                  | 0.15                           | -16084                             |
| Baseline model with errors | 0.03                                  | 0.3                            | -35357                             |
| Baseline model with errors | 0.03                                  | 0                              | -4230                              |
| Baseline model with errors | 0.15                                  | 0.03                           | -1462                              |
| Baseline model with errors | 0.15                                  | 0.15                           | -2409                              |
| Baseline model with errors | 0.15                                  | 0.3                            | -13074                             |
| Baseline model with errors | 0.15                                  | 0                              | -2849                              |
| Baseline model with errors | 0.3                                   | 0.03                           | -9430                              |
| Baseline model with errors | 0.3                                   | 0.15                           | -1416                              |
| Baseline model with errors | 0.3                                   | 0.3                            | -3880                              |
| Baseline model with errors | 0.3                                   | 0                              | -13998                             |
| Baseline model with errors | 0                                     | 0.03                           | -10312                             |
| Baseline model with errors | 0                                     | 0.15                           | -23531                             |
| Baseline model with errors | 0                                     | 0.3                            | -45726                             |
| Baseline model with errors | 0                                     | 0                              | -9106                              |

Table S2: We simulated data from the Prüfer *et al.* 2017 model and added in ancestral misidentification error and genotyping error in the archaic.  $\Delta\mathcal{L}\mathcal{L}$  values come from comparing to model C. **N** refers to the fit of the simulated model compared to  $CSFS_{YRI,N}$ , the site frequency spectrum in Yoruba conditional on observing a derived allele in the Vindija Neanderthal. **D** refers to  $CSFS_{YRI,D}$ , the same spectrum conditional on observing a derived allele in the Denisovan.

| Model                  | Admixture frac | $\Delta\mathcal{L}\mathcal{L}$ (N) | $\Delta\mathcal{L}\mathcal{L}$ (D) |
|------------------------|----------------|------------------------------------|------------------------------------|
| SIM004                 | 0.00001        | -7976                              | -7740                              |
| SIM004                 | 0.00015        | -8263                              | -8048                              |
| SIM004                 | 0.001          | -8268                              | -7931                              |
| SIM004                 | 0.01           | -13838                             | -8492                              |
| SIM004                 | 0.02           | -28564                             | -12432                             |
| SIM005                 | 0              | -8122                              | -7769                              |
| SIM006                 | 0              | -9428                              | -8804                              |
| SIM007                 | 0              | -8431                              | -7449                              |
| SIM008                 | 0              | -8317                              | -7745                              |
| SIM009                 | 0              | -8329                              | -7894                              |
| SIM010                 | 0              | -8373                              | -7871                              |
| SIM012                 | 0              | -14916                             | -7983                              |
| SIM013                 | 0              | -31553                             | -9619                              |
| SIM014                 | 0              | -8958                              | -8609                              |
| SIM015                 | 0              | -8109                              | -7702                              |
| Continuous Migration   | 0              | -7799                              | -7405                              |
| Continuous Migration 2 | 0              | -4077                              | -3683                              |

Table S3: Model fits for null models including structure, and departures from panmixia in the MH ancestor.

| Model     | Admixture frac | $\Delta\mathcal{L}\mathcal{L}$ (N) | $\Delta\mathcal{L}\mathcal{L}$ (D) |
|-----------|----------------|------------------------------------|------------------------------------|
| Model A.1 | 0              | -7817                              | -7527                              |
| Model A.1 | 0.01           | -8092                              | -7633                              |
| Model A.1 | 0.05           | -7448                              | -6939                              |
| Model A.1 | 0.1            | -7677                              | -7261                              |
| Model A.1 | 0.2            | -6806                              | -6240                              |
| Model A.1 | 0.3            | -7460                              | -7219                              |
| Model A.1 | 0.4            | -8096                              | -7670                              |
| Model A.1 | 0.5            | -7907                              | -7384                              |
| Model A.2 | 0              | -8313                              | -7809                              |
| Model A.2 | 0.01           | -7876                              | -7383                              |
| Model A.2 | 0.05           | -7697                              | -7341                              |
| Model A.2 | 0.1            | -7307                              | -6826                              |
| Model A.2 | 0.2            | -7668                              | -7253                              |
| Model A.2 | 0.3            | -7282                              | -6975                              |
| Model A.2 | 0.4            | -8055                              | -7496                              |
| Model A.2 | 0.5            | -8248                              | -7935                              |
| Model B.1 | 0              | -8139                              | -7765                              |
| Model B.1 | 0.01           | -9058                              | -8638                              |
| Model B.1 | 0.05           | -13985                             | -13752                             |
| Model B.1 | 0.1            | -18786                             | -18190                             |
| Model B.1 | 0.2            | -19658                             | -19065                             |
| Model B.1 | 0.3            | -15490                             | -14903                             |
| Model B.1 | 0.4            | -10720                             | -10497                             |
| Model B.1 | 0.5            | -10585                             | -11429                             |
| Model B.2 | 0              | -8311                              | -7979                              |
| Model B.2 | 0.01           | -8811                              | -8434                              |
| Model B.2 | 0.05           | -27579                             | -28156                             |
| Model B.2 | 0.1            | -51449                             | -53124                             |
| Model B.2 | 0.2            | -58645                             | -60131                             |
| Model B.2 | 0.3            | -46865                             | -46987                             |
| Model B.2 | 0.4            | -49247                             | -49217                             |
| Model B.2 | 0.5            | -64301                             | -68542                             |
| Model C.1 | 0              | -8029                              | -7675                              |
| Model C.1 | 0.01           | -4997                              | -4774                              |
| Model C.1 | 0.05           | -363                               | -162                               |
| Model C.1 | 0.1            | 0                                  | 0                                  |
| Model C.1 | 0.2            | -151                               | -127                               |
| Model C.1 | 0.3            | -4844                              | -4631                              |
| Model C.1 | 0.4            | -16355                             | -15762                             |
| Model C.1 | 0.5            | -24115                             | -23062                             |
| Model C.2 | 0              | -8345                              | -7880                              |
| Model C.2 | 0.01           | -5396                              | -4954                              |
| Model C.2 | 0.05           | -424                               | -283                               |
| Model C.2 | 0.1            | -976                               | -1033                              |
| Model C.2 | 0.2            | -434                               | -531                               |
| Model C.2 | 0.3            | -5266                              | -5133                              |
| Model C.2 | 0.4            | -18908                             | -18293                             |
| Model C.2 | 0.5            | -28835                             | -27534                             |

Table S4: Model fits for alternate models including admixture from other lineages.



| Model      | Admixture frac | $\Delta\mathcal{L}\mathcal{L}$ (N) | $\Delta\mathcal{L}\mathcal{L}$ (D) |
|------------|----------------|------------------------------------|------------------------------------|
| Model sA.1 | 0              | -8524                              | -8127                              |
| Model sA.1 | 0.01           | -8353                              | -7955                              |
| Model sA.1 | 0.05           | -7539                              | -7176                              |
| Model sA.1 | 0.1            | -8023                              | -7594                              |
| Model sA.1 | 0.2            | -7635                              | -7181                              |
| Model sA.1 | 0.3            | -7495                              | -7109                              |
| Model sA.1 | 0.4            | -8245                              | -7649                              |
| Model sA.1 | 0.5            | -8742                              | -8438                              |
| Model sA.2 | 0              | -7990                              | -7515                              |
| Model sA.2 | 0.01           | -8272                              | -7767                              |
| Model sA.2 | 0.05           | -7761                              | -7275                              |
| Model sA.2 | 0.1            | -7571                              | -7101                              |
| Model sA.2 | 0.2            | -7289                              | -7004                              |
| Model sA.2 | 0.3            | -7573                              | -7275                              |
| Model sA.2 | 0.4            | -7923                              | -7617                              |
| Model sA.2 | 0.5            | -7964                              | -7551                              |
| Model sB.1 | 0              | -7970                              | -7545                              |
| Model sB.1 | 0.01           | -8951                              | -8453                              |
| Model sB.1 | 0.05           | -14284                             | -13641                             |
| Model sB.1 | 0.1            | -20217                             | -19009                             |
| Model sB.1 | 0.2            | -21393                             | -20120                             |
| Model sB.1 | 0.3            | -15401                             | -14732                             |
| Model sB.1 | 0.4            | -11116                             | -10807                             |
| Model sB.1 | 0.5            | -12002                             | -12105                             |
| Model sB.2 | 0              | -7883                              | -7564                              |
| Model sB.2 | 0.01           | -8761                              | -8120                              |
| Model sB.2 | 0.05           | -29636                             | -28036                             |
| Model sB.2 | 0.1            | -56985                             | -54354                             |
| Model sB.2 | 0.2            | -64927                             | -62281                             |
| Model sB.2 | 0.3            | -49752                             | -48286                             |
| Model sB.2 | 0.4            | -50901                             | -50110                             |
| Model sB.2 | 0.5            | -70994                             | -70502                             |
| Model sC.1 | 0              | -8322                              | -7885                              |
| Model sC.1 | 0.01           | -5365                              | -5037                              |
| Model sC.1 | 0.05           | -441                               | -175                               |
| Model sC.1 | 0.1            | 14                                 | 145                                |
| Model sC.1 | 0.2            | -97                                | -166                               |
| Model sC.1 | 0.3            | -4655                              | -4667                              |
| Model sC.1 | 0.4            | -16010                             | -15794                             |
| Model sC.1 | 0.5            | -24263                             | -23490                             |
| Model sC.2 | 0              | -8162                              | -7701                              |
| Model sC.2 | 0.01           | -5374                              | -4962                              |
| Model sC.2 | 0.05           | -591                               | -284                               |
| Model sC.2 | 0.1            | -978                               | -776                               |
| Model sC.2 | 0.2            | -705                               | -814                               |
| Model sC.2 | 0.3            | -5121                              | -5187                              |
| Model sC.2 | 0.4            | -18625                             | -18191                             |
| Model sC.2 | 0.5            | -28545                             | -27798                             |

Table S5: Model fits for alternate models using a simplified demography. See Section S3 for more information.

| <b>Model</b>                             | <b>Admixture frac</b> | $\Delta\mathcal{L}\mathcal{L}$ (N) | $\Delta\mathcal{L}\mathcal{L}$ (D) |
|--|-----------------------|------------------------------------|------------------------------------|
| Model A.2 contraction2                   | 0.01                  | -8824                              | -8438                              |
| Model A.2 contraction2                   | 0.05                  | -7871                              | -7454                              |
| Model A.2 contraction2                   | 0.1                   | -8336                              | -7777                              |
| Model A.2 contraction2                   | 0.2                   | -9163                              | -8855                              |
| Model A.2 contraction2                   | 0.3                   | -8804                              | -8492                              |
| Model A.2 contraction2                   | 0.4                   | -4621                              | -4332                              |
| Model A.2 contraction2                   | 0.5                   | -4695                              | -4230                              |
| Model A.2 contraction2                   | 0                     | -8636                              | -8224                              |
| Model A.2 contraction2 CEU-YRI migration | 0.01                  | -8243                              | -7905                              |
| Model A.2 contraction2 CEU-YRI migration | 0.05                  | -7421                              | -7111                              |
| Model A.2 contraction2 CEU-YRI migration | 0.1                   | -7892                              | -7638                              |
| Model A.2 contraction2 CEU-YRI migration | 0.2                   | -9144                              | -8775                              |
| Model A.2 contraction2 CEU-YRI migration | 0.3                   | -9360                              | -8909                              |
| Model A.2 contraction2 CEU-YRI migration | 0.4                   | -4730                              | -4391                              |
| Model A.2 contraction2 CEU-YRI migration | 0.5                   | -4148                              | -3404                              |
| Model A.2 contraction2 CEU-YRI migration | 0                     | -7894                              | -7450                              |
| Model A.2 contraction3                   | 0.01                  | -8025                              | -7495                              |
| Model A.2 contraction3                   | 0.05                  | -7584                              | -7214                              |
| Model A.2 contraction3                   | 0.1                   | -7782                              | -7317                              |
| Model A.2 contraction3                   | 0.2                   | -10225                             | -9831                              |
| Model A.2 contraction3                   | 0.3                   | -10617                             | -10405                             |
| Model A.2 contraction3                   | 0.4                   | -6466                              | -6045                              |
| Model A.2 contraction3                   | 0.5                   | -3517                              | -2903                              |
| Model A.2 contraction3                   | 0                     | -7650                              | -7302                              |
| Model A.2 contraction                    | 0.01                  | -8469                              | -8257                              |
| Model A.2 contraction                    | 0.05                  | -7919                              | -7335                              |
| Model A.2 contraction                    | 0.1                   | -7966                              | -7443                              |
| Model A.2 contraction                    | 0.2                   | -9879                              | -9560                              |
| Model A.2 contraction                    | 0.3                   | -9169                              | -8727                              |
| Model A.2 contraction                    | 0.4                   | -4719                              | -4335                              |
| Model A.2 contraction                    | 0.5                   | -4504                              | -3991                              |
| Model A.2 contraction                    | 0                     | -8525                              | -7995                              |
| Model A.2 trifurcation                   | 0.01                  | -6775                              | -6360                              |
| Model A.2 trifurcation                   | 0.05                  | -3226                              | -2983                              |
| Model A.2 trifurcation                   | 0.1                   | -2713                              | -2597                              |
| Model A.2 trifurcation                   | 0.2                   | -5290                              | -5442                              |
| Model A.2 trifurcation                   | 0.3                   | -22625                             | -22443                             |
| Model A.2 trifurcation                   | 0.4                   | -42168                             | -41835                             |
| Model A.2 trifurcation                   | 0.5                   | -34267                             | -33836                             |
| Model A.2 trifurcation                   | 0                     | -8324                              | -7849                              |

Table S6: Model fits for variations of model A.

| Parameter                       | Value              | 95% HPD                                | Pop |
|---------------------------------|--------------------|--|-----|
| Branch split time ( $t_s$ )     | $62.5 \times 10^4$ | $36.5 \times 10^4 - 102 \times 10^4$   | ESN |
| Admixture fraction ( $\alpha$ ) | 0.09               | 0.02-0.19                              | ESN |
| Admixture time ( $t_a$ )        | $3.8 \times 10^4$  | $0 - 10.8 \times 10^4$                 | ESN |
| Eff. population size ( $N_e$ )  | $2.3 \times 10^4$  | $2.2 \times 10^4 - 2.5 \times 10^4$    | ESN |
| Branch split time ( $t_s$ )     | $66 \times 10^4$   | $41.5 \times 10^4 - 99.9 \times 10^4$  | GWD |
| Admixture fraction ( $\alpha$ ) | 0.12               | 0.06-0.19                              | GWD |
| Admixture time ( $t_a$ )        | $5.9 \times 10^4$  | $3.2 \times 10^4 - 10.4 \times 10^4$   | GWD |
| Eff. population size ( $N_e$ )  | $2.4 \times 10^4$  | $2.3 \times 10^4 - 2.5 \times 10^4$    | GWD |
| Branch split time ( $t_s$ )     | $62.4 \times 10^4$ | $29.1 \times 10^4 - 113.7 \times 10^4$ | LWK |
| Admixture fraction ( $\alpha$ ) | 0.18               | 0.09-0.31                              | LWK |
| Admixture time ( $t_a$ )        | $5 \times 10^4$    | $1.1 \times 10^4 - 13.3 \times 10^4$   | LWK |
| Eff. population size ( $N_e$ )  | $2.4 \times 10^4$  | $2.2 \times 10^4 - 2.6 \times 10^4$    | LWK |
| Branch split time ( $t_s$ )     | $61.1 \times 10^4$ | $39 \times 10^4 - 88.4 \times 10^4$    | MSL |
| Admixture fraction ( $\alpha$ ) | 0.09               | 0.05-0.16                              | MSL |
| Admixture time ( $t_a$ )        | $3.7 \times 10^4$  | $0.82 \times 10^4 - 9.8 \times 10^4$   | MSL |
| Eff. population size ( $N_e$ )  | $2.6 \times 10^4$  | $2.4 \times 10^4 - 2.8 \times 10^4$    | MSL |
| Branch split time ( $t_s$ )     | $62.5 \times 10^4$ | $36 \times 10^4 - 97.5 \times 10^4$    | YRI |
| Admixture fraction ( $\alpha$ ) | 0.11               | 0.045-0.19                             | YRI |
| Admixture time ( $t_a$ )        | $4.3 \times 10^4$  | $0.6 \times 10^4 - 12.4 \times 10^4$   | YRI |
| Eff. population size ( $N_e$ )  | $2.5 \times 10^4$  | $2.3 \times 10^4 - 2.7 \times 10^4$    | YRI |

Table S7: Best fitting parameter values for all populations using approximate Bayesian computation. Times are expressed in years. See Section S5.3 for details on the conversion from generations to years.

| Model | $p$ -value            |
|-------|-----------------------|
| A.1   | $6.7 \times 10^{-16}$ |
| A.2   | $3.2 \times 10^{-9}$  |
| B.1   | $1.4 \times 10^{-11}$ |
| B.2   | $6.7 \times 10^{-16}$ |
| C.1   | 0.003                 |
| C.2   | $9.5 \times 10^{-6}$  |

Table S8: P-values of a test of goodness of fit for the best fitting parameters for each class of demographic models. We computed the residuals comparing the empirical CSFS to the CSFS simulated under the best-fitting parameters for each model. We assessed goodness-of-fit of the model by testing if the residuals are approximately normally distributed using the approach outlined in the ‘Goodness of fit’ section of the main text. We computed goodness-of-fit restricting to derived allele counts ranging between 6 and 95. While all the compared are formally rejected (likely due to additional unmodeled complexities in the true underlying demographic model), we find that model C has substantially higher p-values than the other demographic models consistent with the observations in the main text.