

Supplement for the timing of human adaptation from Neanderthal introgression

Sivan Yair^{*,†,1}, Kristin M. Lee[‡], and Graham Coop^{*,†}

^{*}Center for Population Biology and

[†]Department of Evolution and Ecology, University of California, Davis, California 95616

[‡]Previous address: Department of Biological Sciences, Columbia University, New York, New York 10027

S1 Supplementary Tables

Population	Migration Time (generations ago)	Admixture Proportion
EF	224	0.52
Steppe	155	0.36
WHG	226	0.12

Table S1: Timing and proportion of admixture between listed ancient populations and CEU (a European-ancestry population). Admixture proportions were determined from Haak *et al.* (2015). Simulations use adjusted admixture proportions so that CEU has the listed admixture proportions following the final admixture event. Our models assume a single time for all admixture among European-ancestry populations, which we define as the average of the migration times listed here.

Chrom.	Start to End Positions (hg19)	Genes	Plot Label
1	193772128 - 193999736		chr1:193885932
1	195854615 - 195922174		chr1:195888394
1	201038832 - 201102553	CACNA1S, ASCL5	CACNA1S,ASCL5
2	65395511 - 65533682	ACTR2	ACTR2
2	68347097 - 68507952	PPP3R1, PNO1	PPP3R1,PNO1
2	119536920 - 119693701	EN1	EN1
2	154435207 - 154551882		chr2:154493544

2	159794421 - 160204173	TANC1, WDSUB1, BAZ2B	TANC1...BAZ2B
2	160896628 - 161078020	PLA2R1, ITGB6	PLA2R1,ITGB6
4	5828060 - 5920067	EVC, CRMP1	EVC,CRMP1
4	189157618 - 189216507		chr4:189187062
5	507500 - 653905	SLC9A3	SLC9A3
5	167971758 - 168166025	SLIT3	SLIT3
6	1396838 - 1451472	FOXF2	FOXF2
6	66231951 - 66532520	EYS	EYS
7	129861223 - 129929228	CPA2	CPA2
7	136570934 - 136667889	CHRM2	CHRM2
8	13846885 - 13942904		chr8:13894894
8	13946965 - 14174597	SGCZ	SGCZ
9	16596898 - 16845056	BNC2	BNC2
11	11552692 - 11625808	GALNT18	GALNT18
11	70891472 - 70976597		chr11:70934034
12	54151324 - 54244563		chr12:54197944
12	84808525 - 84998584		chr12:84903554
12	113337379 - 113427536	OAS1, OAS3	OAS1,OAS3
14	100944952 - 101071031	WDR25, BEGAIN	WDR25,BEGAIN
15	43432954 - 44053859	TGM5, TGM7	TGM5,TGM7
15	74702840 - 75030779	SEMA7A, UBL7	SEMA7A,UBL7
15	85908189 - 86341507	AKAP13, KLHL25	AKAP13,KLHL25
16	51306915 - 51398178		chr16:51352546
16	77959647 - 78071863	VAT1L	VAT1L
18	60124891 - 60268585	ZCCHC2	ZCCHC2
18	60658804 - 60777888		chr18:60718346
19	33510019 - 33763643	RHPN2, GPATCH1, WDR88, LRP3, SLC7A10	RHPN2...SLC7A10
20	62156560 - 62210005	PPDPF, PTK6, HELZ2	PPDPF...HELZ2
22	49077496 - 49144645	FAM19A5	FAM19A5

Table S2: Putative regions of adaptive introgression that we analyzed. Genes correspond to those listed with signals of adaptive introgression in Racimo et al. (2017). We label intergenic regions by their chromosome and midpoint position.

Proposed selection coefficients (s)	Proposed waiting times until selection (t_b)
0.002, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.05	0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700

Table S3: Proposed parameters of s and t_b , each combination of which we calculated composite likelihoods. For a given run of the method, we did not evaluate a combination of s and t_b if the sweep would not complete (deterministically) before the present day, i.e. if the time until selection is recent enough and/or the sweep time t_s (which also depends on x_s) is long enough such that $t_b + t_s > t_i$.

Waiting time until selection (t_b)	Selected human populations	Combinations of s and t_s (references rows in ??)
0	All Eurasian	1, 2, 3, 4, 5
200	All Eurasian	7, 8, 9
400	All Eurasian All European	7, 8, 9 6, 7, 8, 9
600	All Eurasian All European All Eurasian except EurUP	7, 8, 9 6, 7, 8, 9 6, 7, 8, 9
800	All Eurasian All Eurasian except EurUP All Eurasian except East Asia, EurUP	7, 8, 9 6, 7, 8, 9 6, 7, 8, 9
1000	All Eurasian except EurUP All Eurasian except East Asia, EurUP	6, 7, 8, 9 6, 7, 8, 9
1200	All Eurasian except EurUP All Eurasian except East Asia, EurUP	6, 7, 8, 9 6, 7, 8, 9
1400	All Eurasian except EurUP All Eurasian except East Asia, EurUP	6, 7, 8, 9 6, 7, 8, 9

Table S4: Simulated combinations of waiting time until selection (t_b), group of selected populations, selection coefficient s , and sweep phase duration (t_s). Combinations of s and t_s are labeled by numbers in Table Table S5.

	Selection Coefficient (s)	Sweep Phase Duration (t_s)	Target x_s	# Runs
1	0.005	501	0.20	700
2	0.005	697	0.40	700
3	0.010	349	0.40	700
4	0.010	430	0.60	700
5	0.010	528	0.80	700
6	0.005	501	0.20	100
7	0.005	655	0.35	200
8	0.010	327	0.35	200
9	0.010	474	0.70	100

Table S5: Combinations of s and t_s used in selection simulations used to evaluate the method's performance. Target x_s corresponds to the average frequency the selected allele would reach at the end of the sweep phase if it started the sweep at frequency $g = 0.02$. The last column (# Runs) provides the number of replicates simulated for a provided waiting time until selection and group of populations experiencing selection.

S2 Method Details

S2.1 Method Validation

The method’s power to reject immediate selection increases with the true waiting time until selection (t_b) and the average frequency of the selected allele among putative selected populations (x_s) because they each lead to later estimates of t_b , which correspond to higher $\text{CLR}_{t_b > 0}$ (Figures S1-S5). Among the simulations in which our method rejects immediate selection, it reliably identifies late selection when $t_b > 800$. Notably, it can tell that selection is recent despite the selected allele having drifted to high frequency earlier in time. For example, the West Eurasian Upper Paleolithic (EurUP) were sampled 908 generations after admixture, such that whenever $t_b > 900$, allele frequencies in that population can only reach higher frequencies because of genetic drift. Despite these high frequency cases when the true $t_b > 900$, the method can tell that selection started after the EurUP were sampled.

When East Asians are the only Eurasian population that does not carry the allele at high frequency, the method is biased toward inferring earlier selection when $t_b < 800$ and does poorly when $t_b = 800$. When EurUP is the only population that does not carry the allele at high frequency, the method is biased toward inferring later selection when $t_b < 800$.

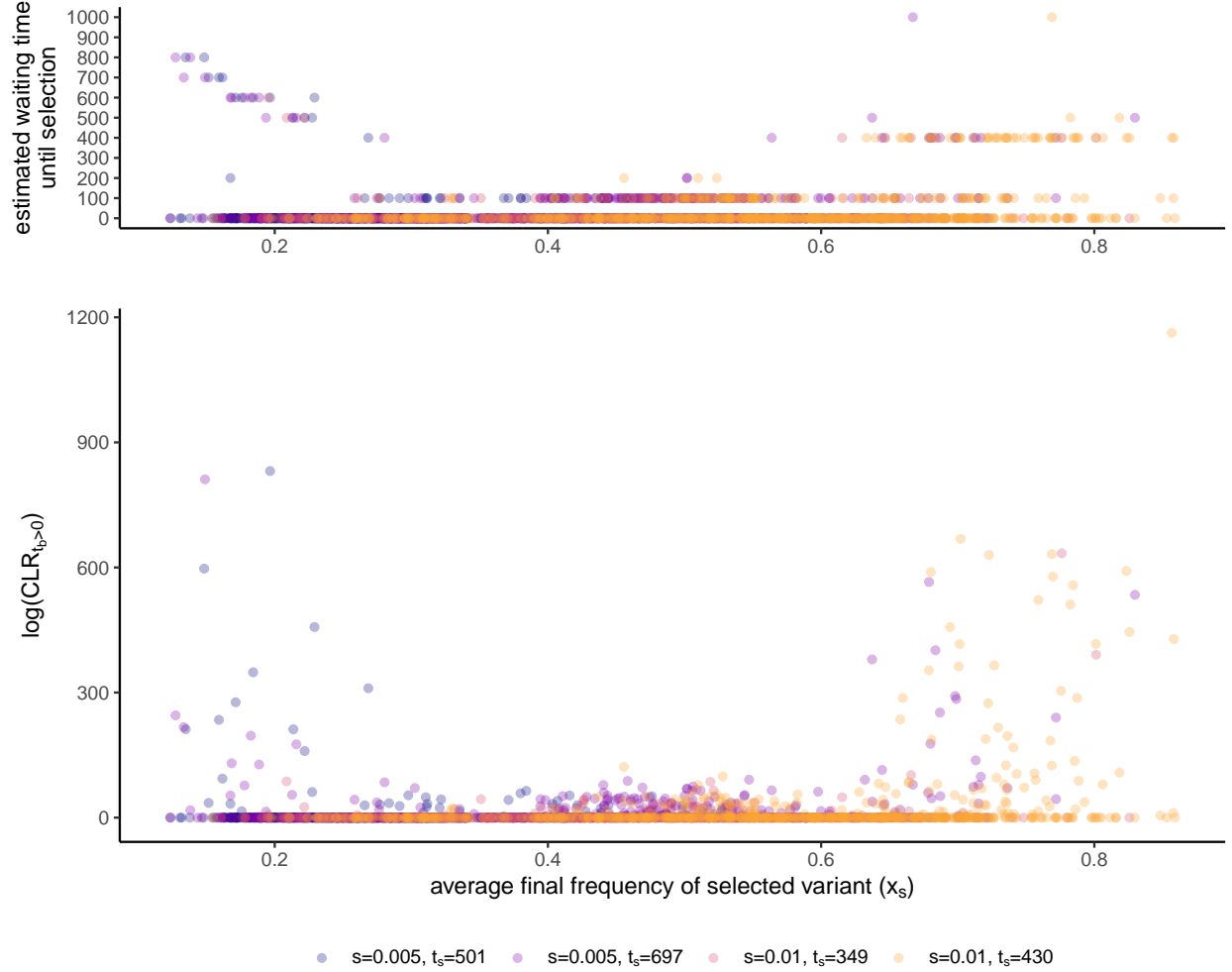


Figure S1: (Top) Relationship between the average final frequency of the selected variant (x_s) and the estimated waiting time until selection (\hat{t}_b) among immediate selection simulations. The method estimated immediate selection ($\hat{t}_b = 0$) in 85% of these simulations. (Bottom) Relationship between x_s and the log of the composite likelihood ratio between non-immediate and immediate selection, $\text{CLR}_{t_b > 0}$, among immediate selection simulations. $\log \text{CLR}_{t_b > 0}$ equals zero for the vast majority of low to intermediate x_s because at these lower frequencies the method is more biased toward inferring immediate selection. The relationship between x_s and $\log \text{CLR}_{t_b > 0}$ does not differ among combinations of s and t_s . See Table S5 for targeted x_s corresponding to each combination of s and t_s .

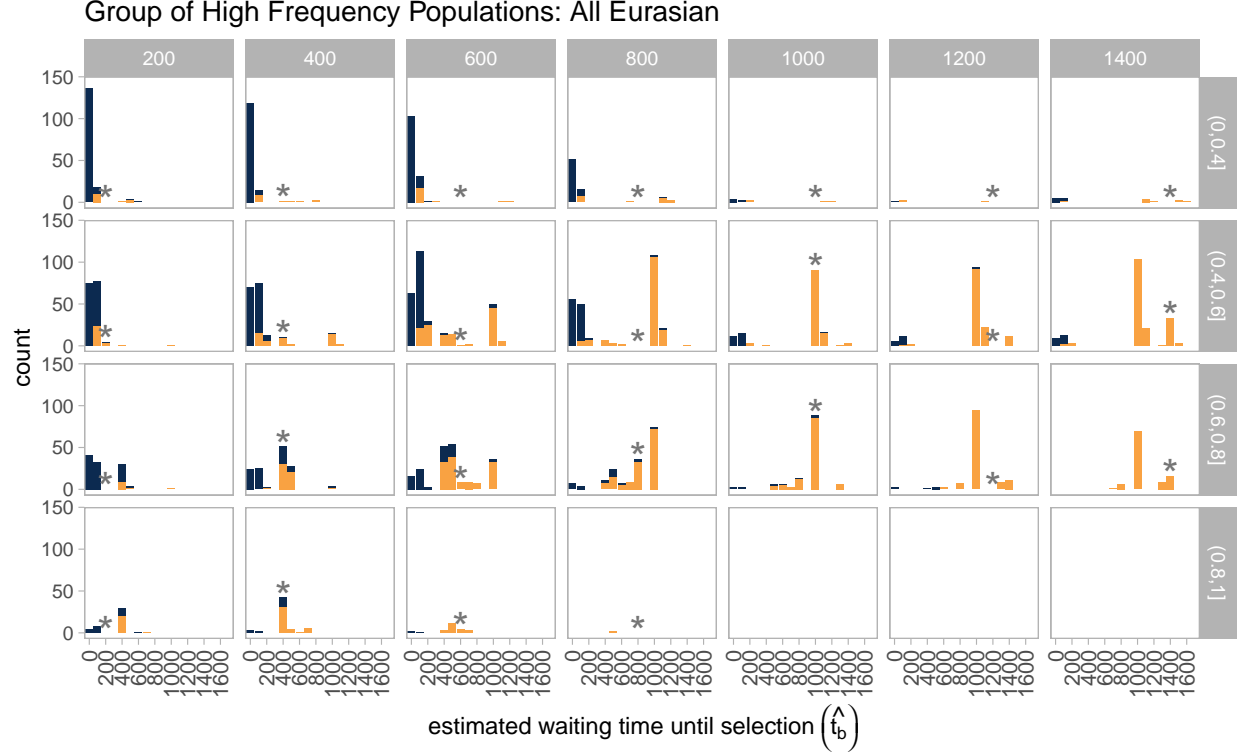


Figure S2: Distribution of estimated waiting times until selection among simulations in which the method considers all Eurasian populations to be selected. Blue bars represent cases in which the method did not reject immediate selection, and orange bars represent cases in which the method did reject immediate selection. Blue bars are stacked on top of orange bars (they do not overlap). Results are shown for each combination of true waiting times until selection (columns) and binned average final frequencies of the selected variant (x_s , rows). Asterisks mark the estimated waiting time until selection that matches the truth.

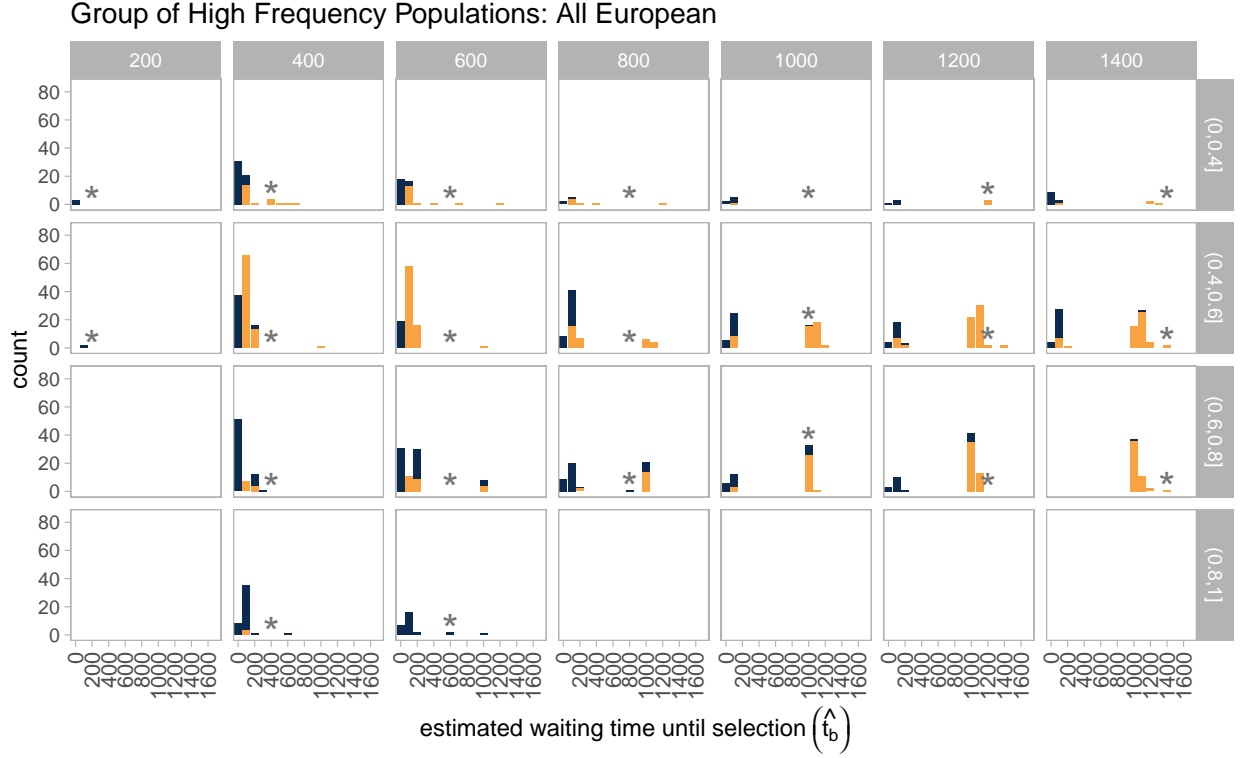


Figure S3: Distribution of estimated waiting times until selection among simulations in which the method considers all European populations (including the West Eurasian Upper Paleolithic) to be selected. Blue bars represent cases in which the method did not reject immediate selection, and orange bars represent cases in which the method did reject immediate selection. Blue bars are stacked on top of orange bars (they do not overlap). Results are shown for each combination of true waiting times until selection (columns) and binned average final frequencies of the selected variant (x_s , rows). Asterisks mark the estimated waiting time until selection that matches the truth.

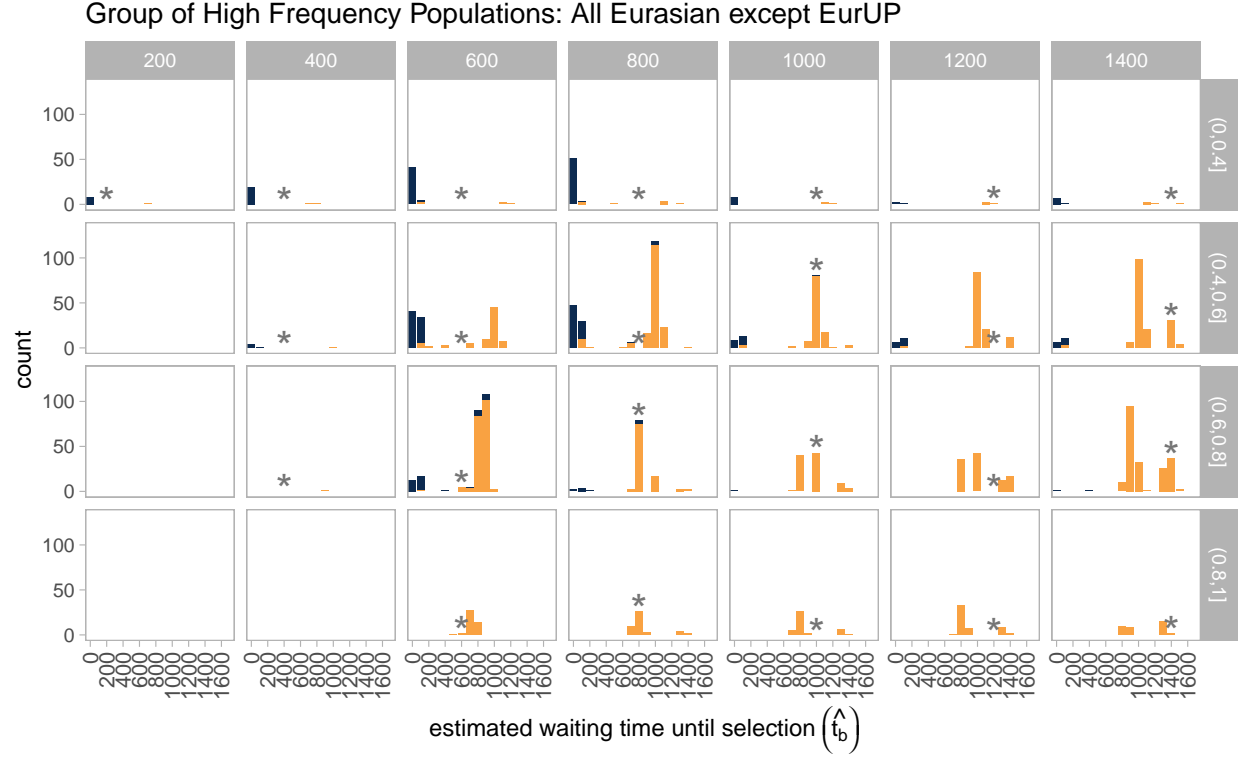


Figure S4: Distribution of estimated waiting times until selection among simulations in which the method considers all Eurasian populations except the West Eurasian Upper Paleolithic to be selected. Blue bars represent cases in which the method did not reject immediate selection, and orange bars represent cases in which the method did reject immediate selection. Blue bars are stacked on top of orange bars (they do not overlap). Results are shown for each combination of true waiting times until selection (columns) and binned average final frequencies of the selected variant (x_s , rows). Asterisks mark the estimated waiting time until selection that matches the truth.

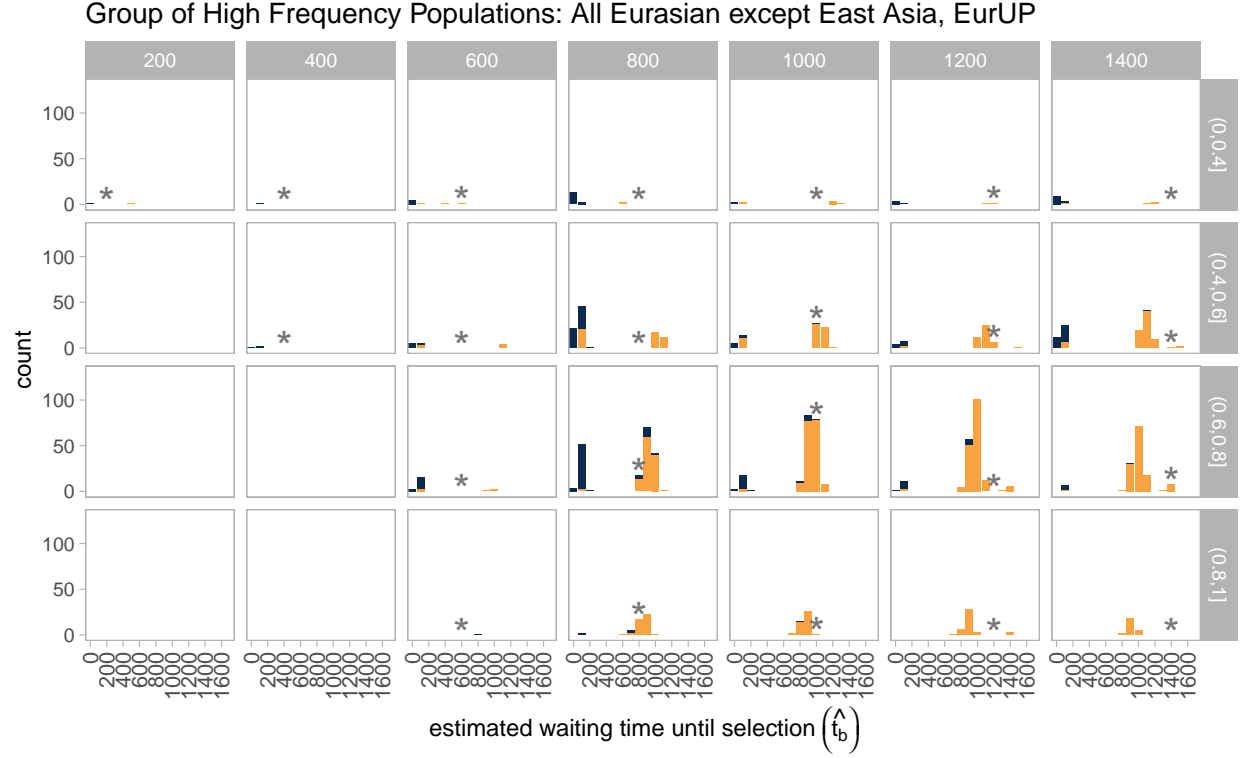


Figure S5: Distribution of estimated waiting times until selection among simulations in which the method considers all Eurasian populations except the West Eurasian Upper Paleolithic and East Asians (represented by CHB) to be selected. Blue bars represent cases in which the method did not reject immediate selection, and orange bars represent cases in which the method did reject immediate selection. Blue bars are stacked on top of orange bars (they do not overlap). Results are shown for each combination of true waiting times until selection (columns) and binned average final frequencies of the selected variant (x_s , rows). Asterisks mark the estimated waiting time until selection that matches the truth.

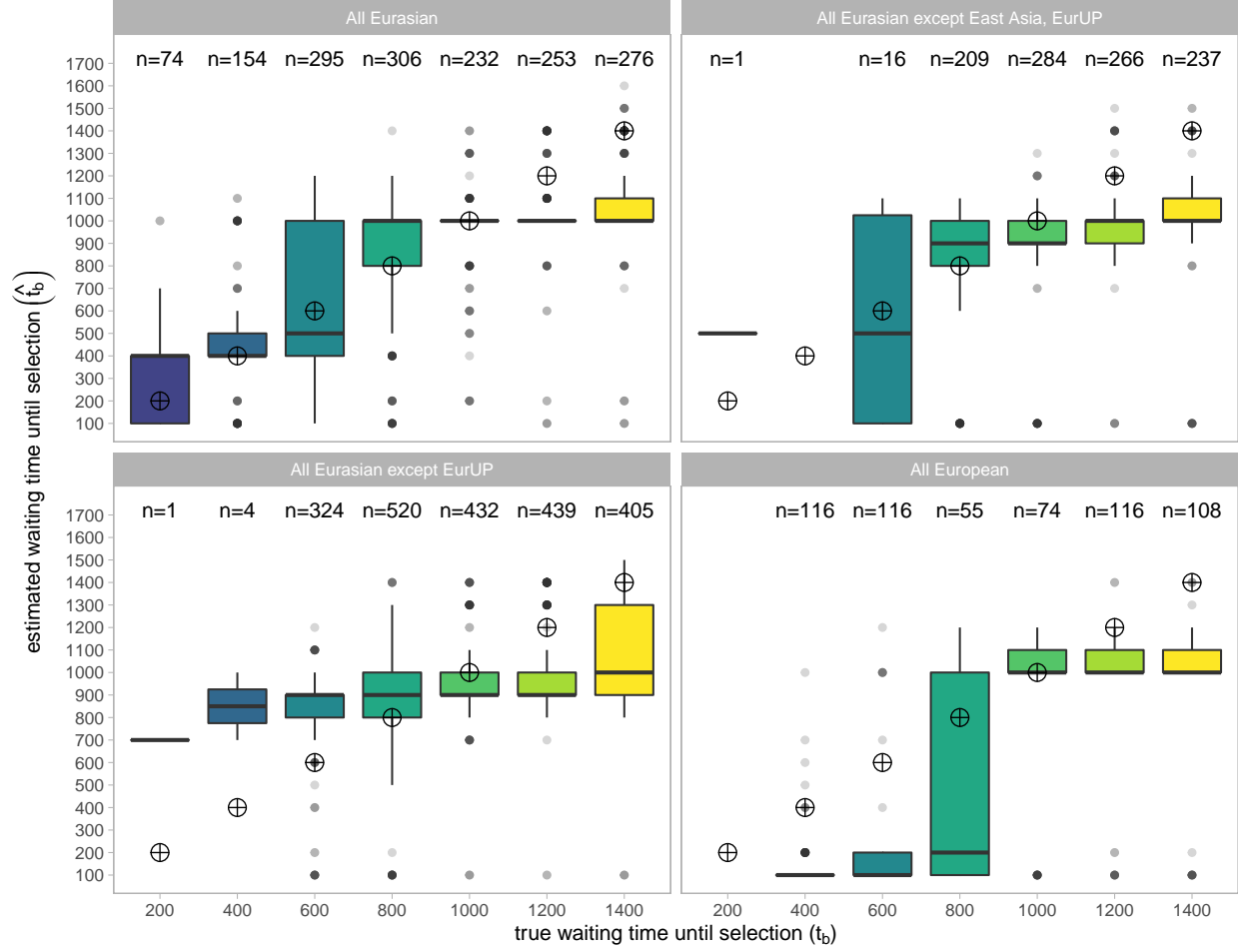


Figure S6: Estimated waiting time until selection (\hat{t}_b) among simulations in which the method rejected immediate selection. Target points mark the true waiting time until selection that we aim to estimate. Earlier true waiting times have fewer observations because the method has less power to reject immediate selection in this parameter space. Each panel corresponds to a different group of populations that the method considers selected.

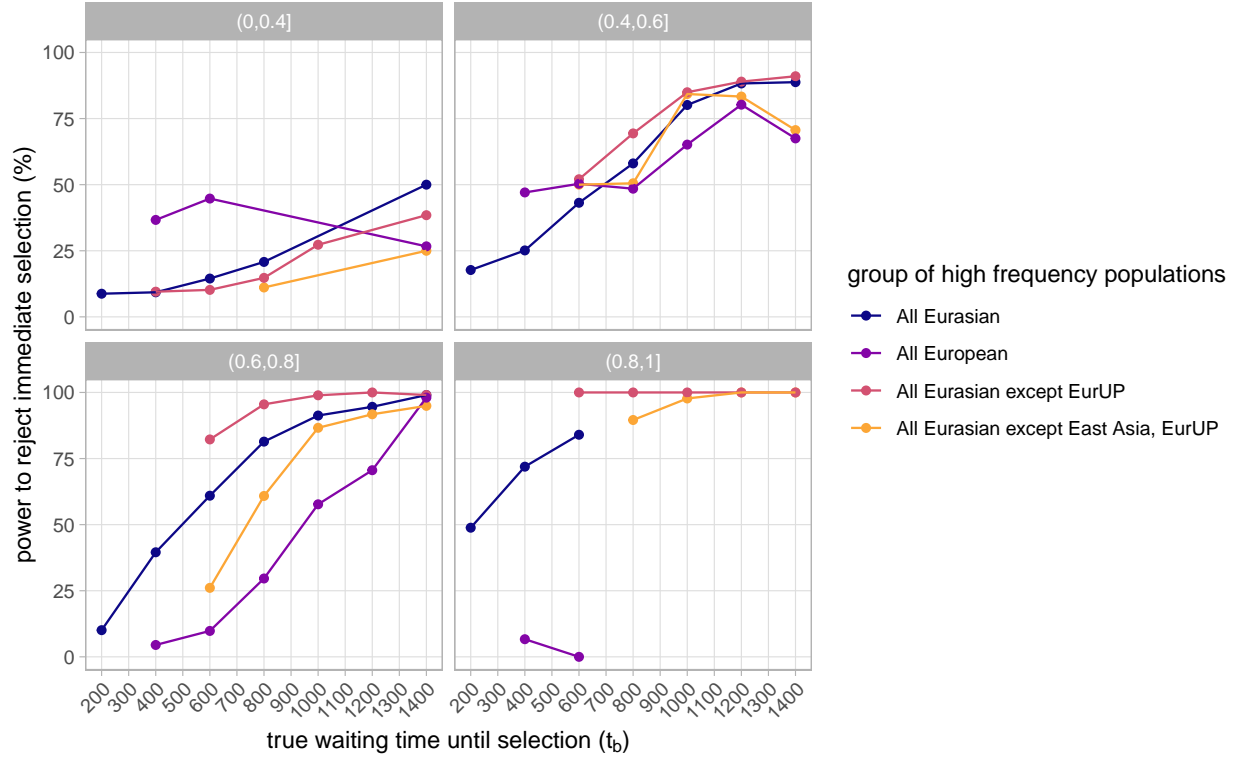


Figure S7: Power to reject immediate selection as the true waiting time until selection (t_b) increases. Lines correspond to simulations with different groups of populations considered selected. Points with fewer than 10 observations were removed. Panels correspond to bins of x_s , the selected allele's average frequency among populations considered selected.

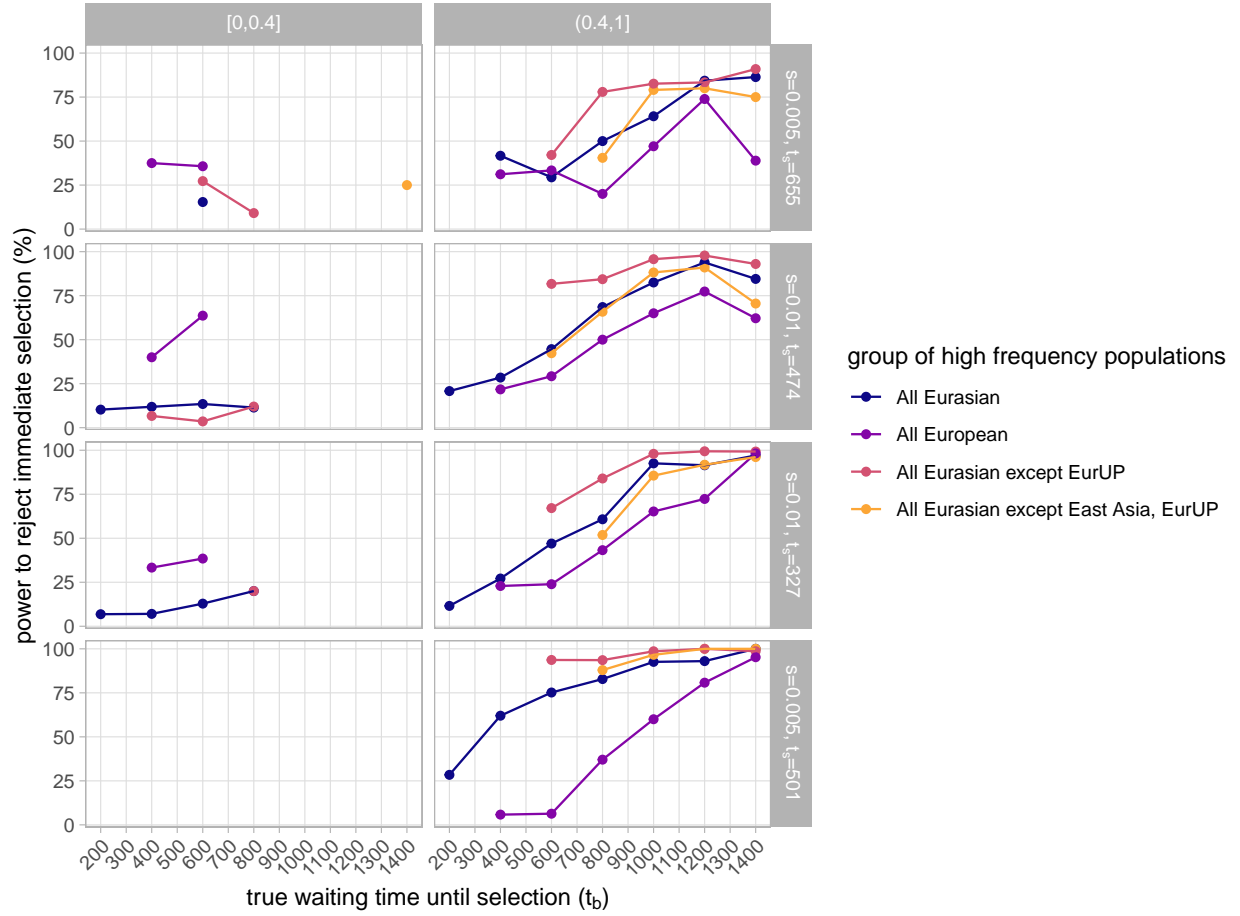


Figure S8: Power to reject immediate selection as the true waiting time until selection (t_b) increases. Lines correspond to simulations with different groups of populations considered selected. Points with fewer than 10 observations were removed. Columns correspond to bins of x_s , the selected allele's average frequency among populations considered selected, and rows correspond to combinations of s and t_s used in simulations.

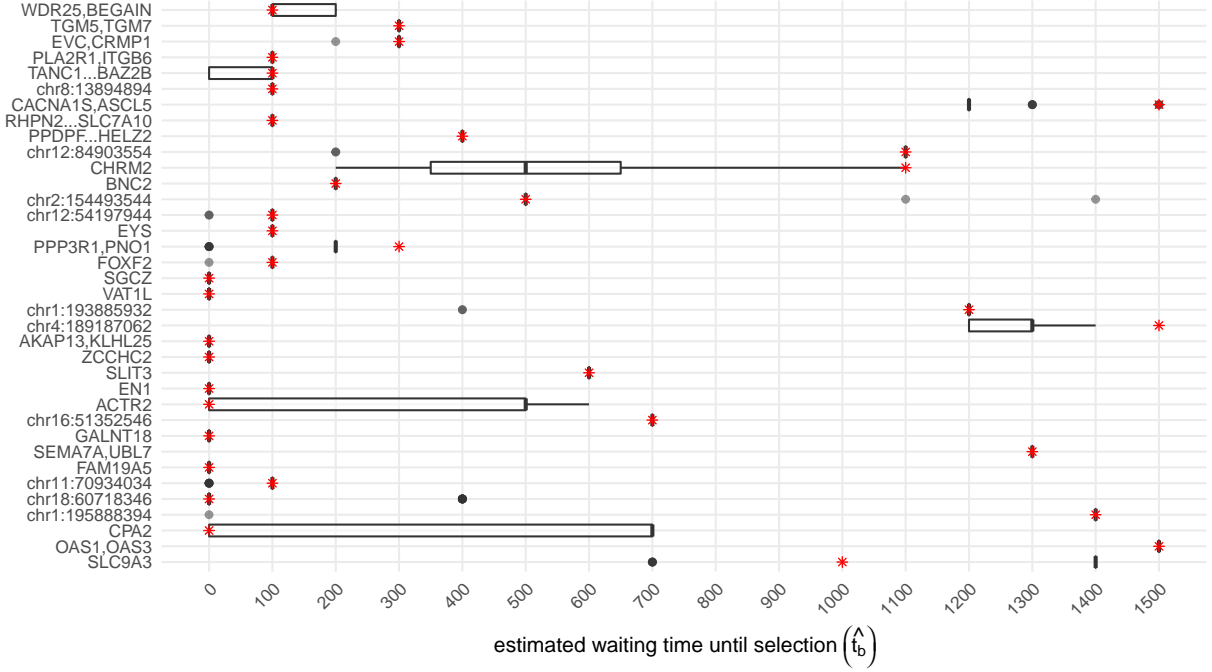


Figure S9: Each region's estimated waiting time until selection (\hat{t}_b) when different sets of ancient samples were randomly assigned to each population's genotype partitions according to their posterior genotype probabilities from imputation. Boxplots show the distribution under 40 resampled datasets. The red asterisks mark each region's \hat{t}_b under the dataset in which ancient partitions are assigned based on the maximum likelihood genotypes. Regions are presented in the same order as in Figure 2.

S2.2 Results under different datasets and demographic assumptions

S2.2.1 Inclusion of the Altai Neanderthal Sample

Population structure among Neanderthals and among Neanderthal-introgressed haplotypes could lead to erroneous results if we do not include the true introgressing populations in our analysis. To investigate our method's sensitivity to our choice of Neanderthal population, we compared our results between using only the Vindija sample and using both the Vindija and Altai samples to calculate allele frequencies for the Neanderthal population. We made a single comparison for each region at the 'best' partition site whose maximum composite likelihood under the selection model was greatest relative to the null model. In the vast majority of regions, our estimated waiting times until selection (\hat{t}_b) did not change, and if they did they only moved within 200 generations (Figure S10). Our estimates changed greatly in two regions, CHRM2 and chr11:70934034, both of which we have identified elsewhere as difficult to infer their timing of selection onset (see the above and below sections). Our results are likely not sensitive to specification of the Neanderthal population because of their low effective population sizes leading to high probabilities of coalescing before the common ancestor of Neanderthals and modern humans. When the Neanderthal population is composed of Vindija alone, we estimate that the probability of coalescing $f_{nn} = 0.95$. For a population composed of Altai and Vindija ($n=4$), we estimate $f_{nn} = 0.91$. Therefore, the ancestral lineages of an allele sampled from the Altai Neanderthal and an allele sampled from the Vindija

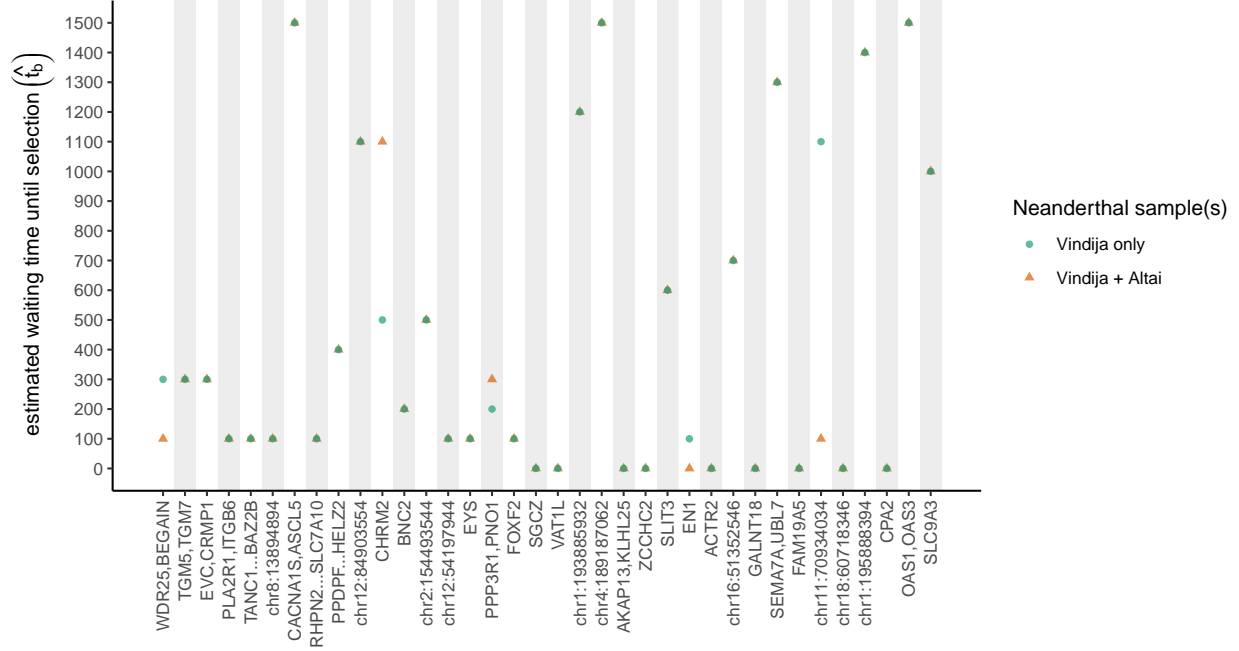


Figure S10: Each region's estimated waiting time until selection \hat{t}_b under different sets of samples comprising the Neanderthal population. Regions are presented in the same order as in Figure 2.

Neanderthal still have an extremely high probability of coalescing. Since in our models we mainly track whether sampled lineages descended from Neanderthals, such that those lineages coalesce while segregating in a Neanderthal population, the Neanderthal source population has a negligible effect on our predictions.

S2.2.2 Sensitivity to Admixture Graphs

In our application, we ran our method on all partition sites while assuming the admixture graph specified in Figure 1, which we call option A. To evaluate the sensitivity of our estimates to these assumptions, we re-ran the method using three other graphs and the dataset of the highest maximum composite likelihood partition site relative to the null model under option A. Each of these additional graphs contain a single modification from the original. In option B, we changed the time of admixture with Neanderthals from 2068 generations in the past (60kya with a generation time of 29 years) to 1988 generations (55kya with a generation time of 29 years). In option C, we did not model present day Europeans as a mixture of the Bronze Age Steppe, Early Neolithic Farmers, and Western Hunter Gatherers. While there is strong evidence that present day Europeans are indeed a mixture of these three populations, we chose an extreme option to generally demonstrate the insensitivity of our method to specifications of admixture among human populations. In option D, we modified the divergence time among present day Europeans, the Bronze Age Steppe, Early Neolithic Farmers, and Western Hunter Gatherers to be 333 generations (10k years with a generation time of 29 years) later (from 40kya to 30kya). Each region's timing estimates when assuming each admixture graph are shown in Figure S11.

Overall, estimated waiting times until selection for the majority of regions did not change with

each admixture graph, or changed very little. In particular, our modification of the divergence time among European populations (option D) did not change the estimated waiting time until selection from option A with the exception of 3 regions (PPPDF..HELZ2; PPP3R1,PNO1; and chr11:7093403). Generally, estimates more often changed when the admixture time with Neanderthals shifted to be more recent. Among those regions in which there was a change in the estimate from option A when using option C, the estimated waiting time until selection was one or two hundred generations shorter, therefore corresponding to a very similar fixed time that selection began. Thus, changing the admixture time resulted in consistency either in the waiting time until selection or how recently selection began. However, three regions showed large inconsistencies in their estimates as we changed the admixture time, transitioning from very early to very late selection: chr12:54197944, FAM19A5, and chr16:51352546. In the former two regions, we did not reject immediate selection under option A. In chr12:54197944, we would potentially reject immediate selection if we simulated immediate selection under the new admixture graph, and therefore caution that the history of selection in this region is unclear. As for FAM19A5, the composite likelihood ratio between non-immediate and immediate selection remains low such that we would likely still reject immediate selection. In chr16:51352546, we previously rejected immediate selection and likely still would, however we caution that our estimated waiting time until selection is not informative in this region. When we remove any migration among modern human populations, the regions with a large change (> 300 generations) are chr12:84903554, chr1:193885932, ACTR2, SEMA7A, UBL7, and CPA2. Based on the composite likelihood ratio between non-immediate and immediate selection, we would likely continue to reject immediate selection in chr12:84903554, chr1:193885932, and SEMA7A, UBL7. In each of those regions the removal of migration is the only change that alters results, and in each case moves our estimates from recent to early selection. These earlier estimates may be unrealistic given that we have strong evidence for migration among European ancestry populations. It is unclear whether we would continue not to reject immediate selection in ACTR2 and CPA2, though we note these two regions also had wide variation in their estimates when we re-partitioned ancient populations (Figure S9), and so we refrain from making any claims about the timing of selection in these regions.

S3 Model Derivation

S3.1 Between selected populations that diverged during neutral phase I

Here, we illustrate another example of how we predict probabilities of coalescing between different categories of populations. We focus in depth on the probability of coalescing between pairs of selected populations who diverged from each other before the sweep started, i.e. during neutral phase I. Afterward, we describe how we change these predictions among other categorical pairs of populations. As a reminder, we consider three phases of the selected allele frequency trajectory: during neutral phase I the selected allele segregates at frequency g for t_b generations, during the sweep phase the selected allele rises from frequency g to frequency x_s over t_s generations, and during neutral phase II the selected allele segregates at frequency x_s until the present day. Thus neutral phase II lasts a duration of $t_I - t_b - t_s$ generations, where t_I is the time between introgression and the present.

If a pair of populations (i, j) diverged from each other d_{ij} generations in the past, and if this divergence occurred during neutral phase I, then they share a common ancestor for $t_I - d_{ij}$

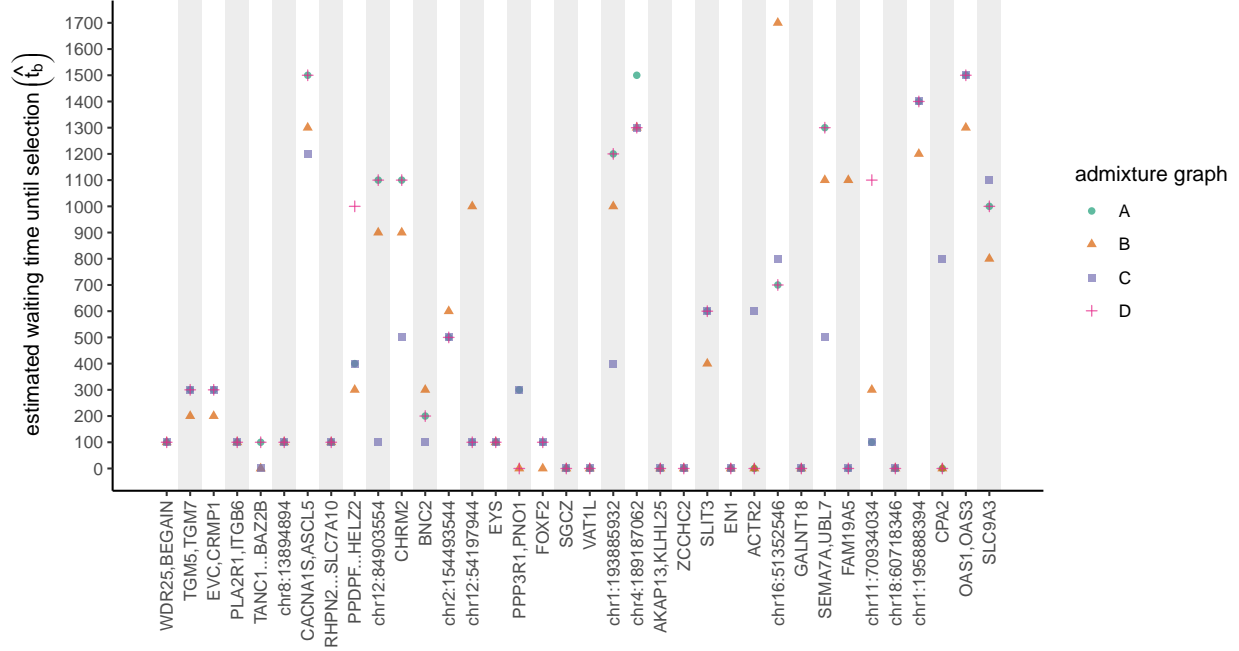


Figure S11: Each region’s estimated waiting time until selection \hat{t}_b under different admixture graphs. Each admixture graph specified by a letter is introduced in section S2.2.2. Regions are presented in the same order as in Figure 2.

generations before the time of introgression. Since the duration of neutral phase I is t_b , we also know that these populations were isolated for the last $t_b - t_I + d_{ij}$ generations of neutral phase I. To get the probability that the ancestral lineages of an allele sampled in population i and an allele sampled in population j coalesce, we consider the ancestry background at the selected site that each ancestral lineage is associated with at each phase transition. In the main text, we described the probabilities that a lineage is linked to the selected or non-selected background at the transition between neutral phase II and the sweep completion (Equation 5), in addition to the probability that a lineage remains strictly associated with this ancestry background for the duration of the sweep phase (Equation 6). With probability $\exp(-(t_b - t_I + d_{ij}))$ a lineage does not recombine out of the background that it was associated with at the transition between the sweep start and neutral phase I, until the time at which the two populations share a common ancestor in neutral phase I.

We begin by describing the probability that the pair of lineages coalesce conditional on them both being linked to the selected allele at the transition between neutral phase II and the sweep finish. We say a lineage is “strictly associated” with the sweep if it never recombines out of that background during the sweep phase and does not recombine at all during the preceding generations of neutral phase I when the lineages are in separate populations, such that

$$\Pr(\text{strictly associated with sweep}) = \Pr(\text{always linked to selected allele during sweep}) \exp(-r(t_b - t_I + d_{ij})). \quad (1)$$

If both lineages remain strictly associated with the sweep, then we consider the possibility that they coalesce because they are associated with a background that is now relatively rare during neutral

phase I. This could be due to their descent from the same introgressed allele, or the same non-introgressed allele that became associated with the selected background before the sweep began. As these lineages can only coalesce if they remain on the same background, we treat coalescence and recombination as competing Poisson processes, with recombination occurring at rate $2r$ and coalescence at rate $\frac{1}{2N_e g}$, where N_e is the mean effective population size of populations (i, j) and g is the neutral admixture proportion, or similarly the frequency of the selected background during neutral phase I. The probability that neither of these events occur before the time of introgression is

$$\Pr(\text{no event}) = \exp\left(-(t_I - d_{ij})\left(2r + \frac{1}{2N_e g}\right)\right). \quad (2)$$

In that case, both lineages are still associated with the selected background at the time of introgression and thus must have descended from Neanderthals, so they coalesce with probability f_{nn} (the probability a pair of alleles sampled from Neanderthals coalesce). If at least one coalescence or recombination event does occur, we care about the outcome of the first event. The relative probability of coalescence is $\frac{1}{1+4N_e gr}$ and of recombination is $\frac{4N_e gr}{1+4N_e gr}$. If the event is recombination of one of the lineages, then we condition on whether the other lineage recombines at some point before introgression. With approximate probability $\exp(-r(t_I - d_{ij}))$ it does not recombine, and therefore remains associated with the selected background such that it descended from Neanderthals. Thus the lineages can only coalesce if the lineage that did recombine also descended from Neanderthals, with probability g . If both lineages recombined during this shared portion of neutral phase I, then they coalesce with neutral probability f_{ij} , which already accounts for whether they are each introgressed or not. We use a similar logic for the remainder of cases that involve one or no lineages remaining strictly associated with the selected background before the pair are in the same population, such that

$\Pr(\text{coalesce} \mid \text{both linked to selected allele at sweep finish}) =$

$$\begin{aligned} & \Pr(\text{strictly associated with sweep})^2 \left[\Pr(\text{no event}) f_{nn} + (1 - \Pr(\text{no event})) \left(\frac{1}{1 + 4N_e gr} + \right. \right. \\ & \quad \left. \left. \frac{4N_e gr}{1 + 4N_e gr} (\exp(-r(t_I - d_{ij})) g f_{nn} + (1 - \exp(-r(t_I - d_{ij}))) f_{ij}) \right) \right] + \\ & 2 \Pr(\text{strictly associated with sweep}) (1 - \Pr(\text{strictly associated with sweep})) \left[\exp(-r(t_I - d_{ij})) g f_{nn} \right. \\ & \quad \left. + (1 - \exp(-r(t_I - d_{ij}))) f_{ij} \right] + \\ & (1 - \Pr(\text{strictly associated with sweep}))^2 f_{ij}. \end{aligned} \quad (3)$$

Next we describe the probability that the pair of lineages coalesce conditional on one being associated with the selected allele and the other being associated with the non-selected allele at the transition between neutral phase II and the sweep finish. Since it is relatively rare for the lineage linked to the non-selected allele to switch to the background of the selected allele during the sweep phase, we do not consider the possibility of increased coalescence during neutral phase I. Therefore, we are simply interested in the ancestry background of each lineage at the time of introgression. If neither of them become disassociated with their backgrounds during the sweep

phase and do not recombine before the time of introgression, then we know that one allele descended from Neanderthals while the other did not, such that they cannot coalesce. If the lineage originally associated with the non-selected background remains associated with this background during the sweep and does not recombine before introgression, whereas the other lineage does disassociate at some point, they can only coalesce if the disassociating allele did not descend from Neanderthals, with probability $1 - g$. At this point, we need to slightly modify their probability of coalescing from the neutral estimate because we are dealing with a case in which we exclude descent from Neanderthals as a possibility. The probability that two lineages sampled from Neanderthal admixed populations coalesce, conditional on them both not descending from Neanderthals, is

$$\Pr(\text{coalesce} \mid \text{both non-Neanderthal}) = \frac{f_{ij} - g^2 f_{nn}}{(1 - g)^2}, \quad (4)$$

where g is the neutral admixture proportion. We have previously described our predictions when one of the lineages descends from Neanderthals due to their consistent association with the selected background while the other lineage may not maintain a consistent association, in addition to when neither lineage maintains a consistent association. Therefore we obtain the following with conditional probabilities ordered in the same manner we discussed them,

$$\begin{aligned} \Pr(\text{coalesce} \mid \text{linked to different allele types at sweep finish}) = & \\ & \Pr(\text{always linked to non-selected allele during sweep})e^{-rt_b} \\ & \times (1 - \Pr(\text{always linked to non-selected allele during sweep})e^{-rt_b})(1 - g) \\ & \times \Pr(\text{coalesce} \mid \text{both non-Neanderthal}) + \\ & \Pr(\text{always linked to selected allele during sweep})e^{-rt_b} \\ & \times (1 - \Pr(\text{always linked to non-selected allele during sweep})e^{-rt_b})gf_{nn} + \\ & (1 - \Pr(\text{always linked to non-selected allele during sweep})e^{-rt_b}) \\ & \times (1 - \Pr(\text{always linked to selected allele during sweep})e^{-rt_b})f_{ij} \end{aligned} \quad (5)$$

Finally, we describe the probability that the pair of lineages coalesce conditional on both being associated with the non-selected allele at the transition between neutral phase II and the sweep finish. We again condition on their background associations at the time of introgression. If both lineages never disassociate from this background during the sweep and never recombine during neutral phase I, then neither descended from Neanderthals and they coalesce with a slightly higher probability than neutrality as we previously described. If this happens with just one of the lineages, then the other lineage must not have descended from Neanderthals for them to coalesce, again with a slightly higher probability than neutrality. In the remainder of cases, the lineages coalesce neutrally. These cases are summarized in the following,

$$\begin{aligned}
& \Pr(\text{coalesce} \mid \text{both linked to non-selected allele at sweep finish}) = \\
& \quad (\Pr(\text{always linked to non-selected allele during sweep})e^{-rt_b})^2 \Pr(\text{coalesce} \mid \text{both non-Neanderthal}) + \\
& \quad \Pr(\text{always linked to non-selected allele during sweep})e^{-rt_b} \\
& \quad \times (1 - \Pr(\text{always linked to non-selected allele during sweep})e^{-rt_b})(1 - g) \\
& \quad \times \Pr(\text{coalesce} \mid \text{both non-Neanderthal}) + \\
& \quad (1 - \Pr(\text{always linked to non-selected allele during sweep})e^{-rt_b})^2 f_{ij}.
\end{aligned} \tag{6}$$

Together, the full probability that a pair of lineages coalesce before the root is

$$\begin{aligned}
f_{p_i p_j}^{(S)} = & \Pr(\text{linked to selected allele at sweep finish})^2 \Pr(\text{coalesce} \mid \text{both linked to selected allele at sweep finish}) + \\
& 2 \Pr(\text{linked to selected allele at sweep finish})(1 - \Pr(\text{linked to selected allele at sweep finish})) \\
& \times \Pr(\text{coalesce} \mid \text{linked to different allele types at sweep finish}) + \\
& (1 - \Pr(\text{linked to selected allele at sweep finish}))^2 \\
& \times \Pr(\text{coalesce} \mid \text{both linked to non-selected allele at sweep finish}),
\end{aligned} \tag{7}$$

where subscript p_i refers to any partition of population i and subscript p_j refers to any partition of population j .

S3.2 Modifications under other scenarios among pairs of selected and non-selected populations

If the pair of selected populations diverged during the sweep phase, we modify our above predictions so that the lineages are in the same population for all of neutral phase I by substituting any term $t_I - d_{ij}$ with t_b (these terms represent the time during neutral phase I in which the pair of lineages are segregating in the same population). If the pair of selected populations are either the same population or diverged from each other during neutral phase II (when the selected allele segregates at frequency x_s), and if the pair of lineages are linked to the same allele type at the most recent time they share a common ancestor, we consider the additional possibility that they coalesce during neutral phase II because of their associations with a subpopulation of haplotypes. We account for this using the same Poisson processes described above, except during neutral phase II the frequency of the selected background is x_s and the frequency of the non-selected background is $1 - x_s$. Recall that in cases of earlier selection, it is possible that the common ancestor of a set of selected populations also has descendant populations that do not carry the allele at high frequency. For those populations, our models assume the selected allele has the same frequency trajectory as other selected populations until the time it diverges from them, after which we switch the frequency of the selected allele in this population to be its sampled frequency. This influences the probability that this population's ancestral lineages were linked to the selected allele at the time its sweep phase begins.

In non-selected Neanderthal admixed populations, we track whether a lineage recombines out of its initial ancestry association, specified by its partition, over the entire time to introgression. If it loses this association, the only way it can have a non-neutral probability of coalescing with another lineage from a similarly non-selected population would be if that other lineage never recombined out

of its initial background. As for a lineage from a selected population, they would have a non-neutral probability of coalescing if the lineage from the selected population maintained an association with a selected or non-selected background by the time of introgression. In the non-selected population, the probability that a lineage could have descended from Neanderthals, conditional on it recombining out of its background at sampling, is g throughout the whole time to introgression. We show how our predictions change with t_b and x_s in Figure S12 and Figure S13.

S3.3 Modifications if a population was sampled in the past

If a population was sampled in the past, we modify the probabilities that its sampled lineages are linked to each background according to the number of generations allotted to a given phase. For example, consider a population i that was sampled τ_i generations ago, and carries the selected allele at high frequency such that in our application we consider it selected. If its sampling time falls within neutral phase II, then to calculate the probability that a lineage sampled from partition B of population i is linked to the selected background at the transition between neutral phase II and the sweep completion we simply reduce the amount of time for recombination in neutral phase II by τ_i . If the provided waiting time until selection t_b and selection coefficient s make it so that population i was instead sampled during the sweep phase, then at the transition between neutral phase II and the sweep completion, the sampled lineages are linked to the background denoted by their partition with probability 1. Then, we calculate that lineage's probabilities of remaining associated with either background during the sweep phase as if the selected allele reached the lower frequency expected under the provided s and the allotted time for the sweep, $t_I - t_b - \tau_i$. If population i was sampled during neutral phase I, then we do not consider it a selected population, and it has $t_I - \tau_i$ generations for possible recombination by the introgression time.

S3.4 Incorporating migration among Neanderthal admixed human populations

In our application, most admixture graphs we use make CEU (the European ancestry representative population) a mixture of ancient populations WHG, EF, and Steppe. In our models, we modify predictions for partition B of CEU, assuming that ancient populations contributed only haplotypes with selected alleles to CEU. We thus consider donor populations to be only those contributing ancient populations with a partition B in the dataset. We set a single migration time t_m to be the average migration time from all human donor populations (see Table S1). With probability $\exp(-rt_m)$ a lineage sampled from partition B of CEU does not recombine before the migration time. In this case, the probability this lineage coalesces with a lineage sampled from any other partition p_j is an average of the probability that each donor population's partition A coalesces with the lineage from p_j , weighted by the relative admixture proportions of each of these donor populations. If the lineage does recombine by the time of migration, the probability it coalesces with a lineage from p_j is that of partition b of CEU. To predict coalescent probabilities for a pair of alleles sampled within CEU's partition B , we average over results from the possible migration histories of both lineages. We acknowledge that our approach to treat migration does not reflect a realistic coalescent history. Other options did not allow our predictions to converge to CEU's neutral probabilities of coalescing with increasing genetic distance from the selected site.

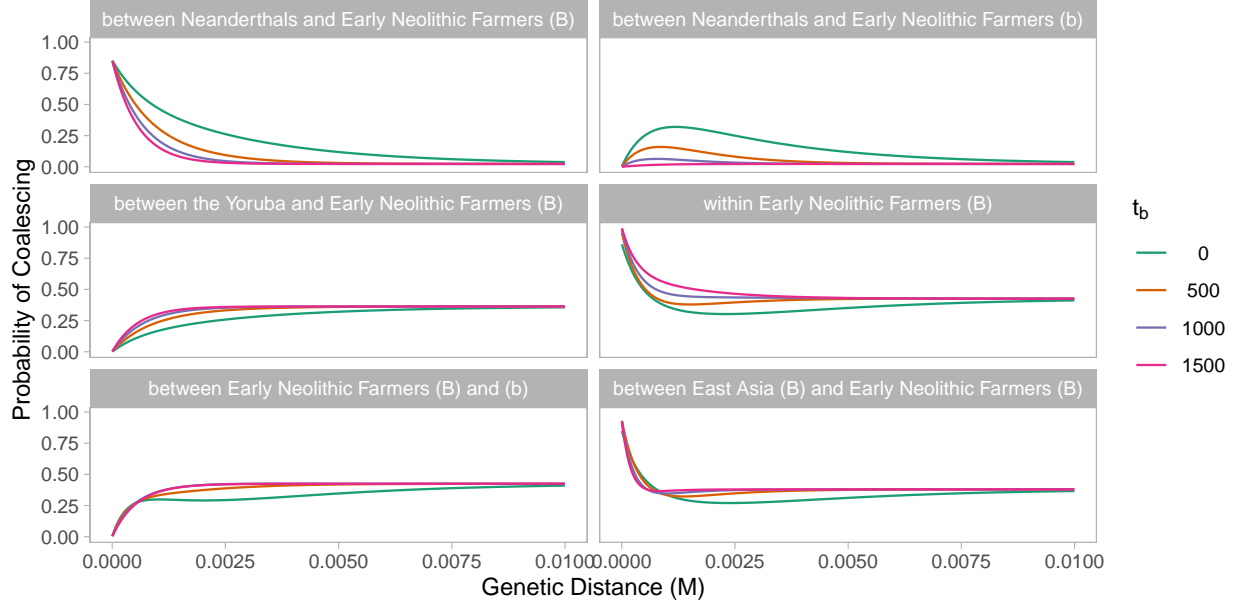


Figure S12: Model predictions under different values of t_b (the waiting time until selection) when $s = 0.01$ and $x_s = 0.7$.

S3.5 Predictions for genotype partition Bb

After making modifications for migration, we make predictions for all Bb partitions. The probability a lineage sampled from this partition in population i coalesces with any other partition from any other population p_j is the mean of this prediction for population i 's partition B and partition b . This reflects that half of the time we are sampling a neutral allele that is linked to the selected allele, whereas in the remainder of cases the neutral allele is linked to the non-selected allele. The probability of coalescing between partition Bb of population i and partition Bb of population j is the mean of the probability of coalescing of all four combinations of partitions A and a in populations i and j , in which there is a 50% chance either partition is sampled in either population. For probabilities of coalescing between lineages that were both sampled from partition Bb of population i , we take a weighted average of the probabilities of coalescing within partition B , within partition b , and between partition B and partition b , all partitions corresponding to population i . The first two cases occur with probability 0.25 and the last case occurs with probability 0.5.

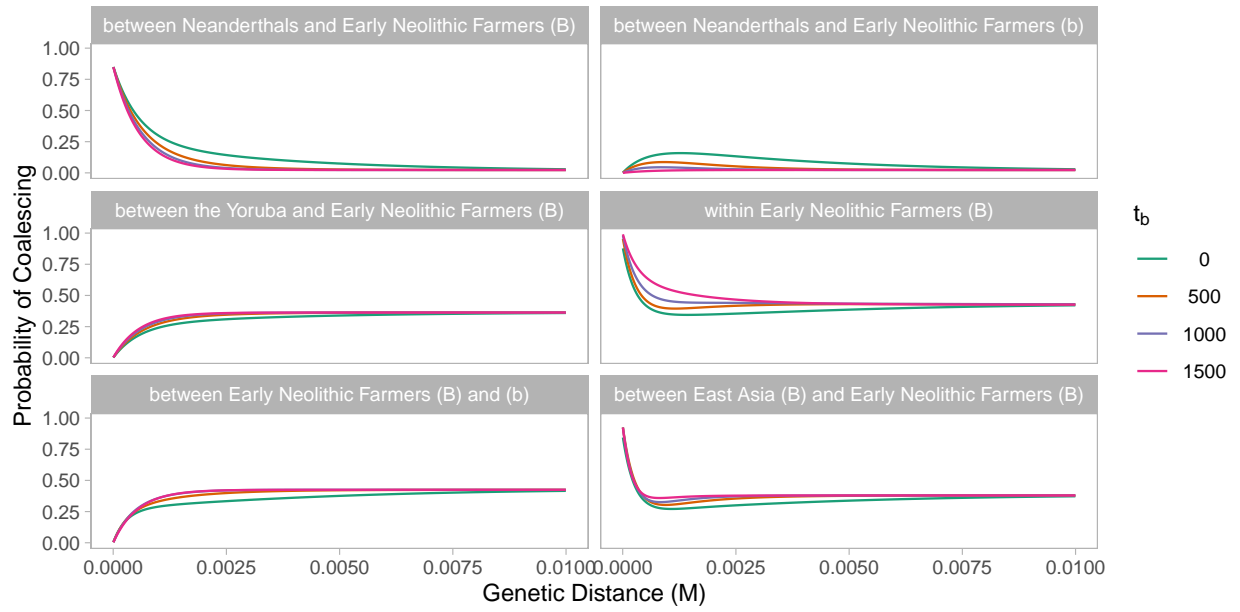


Figure S13: Model predictions under different values of t_b (the waiting time until selection) when $s = 0.01$ and $x_s = 0.3$. Note that predictions of t_b become less distinguishable when x_s decreases (see Figure S12 for comparison to $x_s = 0.7$).