A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years

Priya Moorjani, Sriram Sankararaman, Qiaomei Fu, Molly Przeworski, Nick Patterson, David Reich

Supplementary Material

Table of contents

Note S1: Methods and Materials

Datasets: Below we describe the details of the different datasets used in our analysis.

Neanderthal genome: We used the genotypes called from the high-coverage Altai Neanderthal sequence (1). We restricted our analysis to sites which passed the filters described in Prüfer et al. (1) and for which GQ >= 30. These filters discard sites that are identified as repeats by the Tandem Repeat Finder (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/simpleRepeat.txt.gz) or that have Phred-scaled MQ < 30, or that map to regions where the alignment is ambiguous or which fall within the upper or lower 2.5th percentile of the sample-specific coverage distribution (applied within the regions of unique mappability binned according to the GC-content of the reference genome). For the mappability filter, we used the map35_50% filter that requires that at least 50% of all 35-mers that overlap a position do not map to any other position in the genome allowing up to one mismatch.

1000 Genomes project: We used the sub-Saharan African populations Yoruba (YRI) and Luhya (LWK) for the ascertainment of SNPs that are informative of putative Neanderthal ancestry and the northern European samples (CEU) to estimate the date of Neanderthal admixture in extant samples for comparison (2). We used all SNPs from the 1000 Genomes Phase 1 Integrated call set (ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/20110521/) that passed the filters based on the Variant Quality Score Recalibration LOD value (which in turn was set to give 99.8% sensitivity on accessible HapMap3 variants (2)). This dataset has a total of 38.2 million SNPs across autosomes and chromosome X. We further restricted our analysis to sites that were biallelic across the Altai Neanderthal and the 1000 Genomes samples.

The ancestral allele was determined from the 6-primate EPO alignment (http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/supporting/ancestral_alignments/) and we further restricted our analysis to sites with confidently called ancestral alleles only. After filtering, we obtained 27.2 million SNPs.

Simons Genome Diversity Project (SGDP): We analyzed diverse west Eurasian populations (79 samples belonging to 39 groups) that were sequenced as part of SGDP (3), combined with genomes from the panel A individuals sequenced in an earlier study (1). Detailed information about the genotype calling and filtering of these samples is included in Mallick et al. 2016 (3). Briefly, the authors used Bwa-mem (4) alignments as an input for single-sample genotype calls built using a reference-bias-free modification of the Unified Genotyper from the Genome Analysis Toolkit (GATK) (5). Sites that were found to be both polymorphic in at least one sample compared with chimpanzee and that pass filters (at filter level 1) were compiled (62.6 million sites). At these discovered positions, genotype calls for samples were compiled at filter level 0 (the lower filter level is justified as the sites are known to be polymorphic in at least one sample). We and restricted our analysis to bi-allelic sites only and determined the ancestral allele as described above.

Ancient genomes: We applied our inference procedure to available ancient genomes with radiocarbon dates >10,000 years that were published at the time of the initial submission of this study: Clovis (6), Mal'ta1 (7), Kostenki14 (8), Ust'-Ishim (9) and Oase1 (10). For all ancient samples with less than 30x coverage (Clovis, Mal'ta1,

Kostenki14 and Oase1), we could not reliably call heterozygous genotypes and thus we used pseudo-homozygous majority genotypes, that is, we chose a single allele at each site for each sample by taking the most common (majority) allele observed in the reads mapped to that site. When the numbers of reference and non-reference alleles were the same, we chose a random allele at that site. For testing the robustness of the genotype calling approach, we used pseudo-homozygous calls based on choosing a random allele from the reads mapped to the site (pseudo-homozygous random). For samples with whole genome sequence data, we used the genotypes for all the SNPs that match the ascertainment scheme. For the Oase1 sample, where whole genome sequence data was not available, we used genotype calls for the 78,055 SNPs from the *Archaic Panel* (panel 4, Neanderthal subset; see (10) for details). To verify the robustness of inferences based on the *Archaic panel*, we used the same set of 78,055 SNPs to date the Neanderthal gene flow in Kostenki14 and Ust-Ishim genotyped using the same capture method (10).

Statistics for dating Neanderthal admixture: To date Neanderthal mixture, we first ascertained SNPs that were informative for Neanderthal ancestry. Unless otherwise stated, we used ascertainment 0 where the Altai Neanderthal carries at least one derived allele and 1000 Genomes Yoruba and Luhya samples are fixed for the ancestral allele. To reduce the effect of sequencing errors, we only considered variant sites where at least one derived allele has been observed among individuals included in the 1000 Genomes Project (except in Yoruba or Luhya). All ancient genomes included in the analysis had <35% missing data for ascertained SNPs.

To date Neanderthal admixture in the ancient genomes (Clovis, Mal'ta1, Kostenki14, Ust'-Ishim and Oase1), we used the single-sample statistic. For the extant samples (i.e., the 1000 Genomes CEU or SGDP west Eurasians), we used the population-sample statistic (11). Unless otherwise stated, we used genetic distances from the shared African American map (referred to as '*S map*' in (12)).

(a) Single-sample statistic

For all pairs of ascertained SNPs, $S(d) = (i,j)$, we computed $C(d)$, which measures the covariance between markers at genetic distance d Morgans apart. We ignored SNPs with missing data.

$$C(d) = \frac{1}{|S(d)|-1} \sum_{x=(i,j)\in S(d)} (g_{xi} - \overline{g_i})(g_{xj} - \overline{g_j})$$

where g_{xi}, g_{xj} are genotypes at i,j SNPs respectively and $\bar{g}_i$ is the mean of g_{xi} and $\bar{g}_j$ is the mean of g_{xj} for all SNPs at distance d. To estimate the date of mixture (λ), we used least squares to fit an exponential distribution with an affine term ($C(d) = Ae^{-d\lambda} + c$) for d in the range of 0.02 to 1cM in increments of 0.001cM (unless otherwise specified). The output based on a single genome is very noisy, and we have found in practice that the affine term is helpful in capturing some of the noise while not biasing the inferred date of admixture.

For Ust'-Ishim, we observed that the intercept at 1cM is substantially greater than 0, and hence we also ran the analysis to longer genetic distances, in the range of 0.02 to 10cM in increments of 0.001cM. To estimate the date of Neanderthal admixture, we tried two

models: single exponential ($C(d) = Ae^{-d\lambda} + c$) and double exponential ($C(d) = Ae^{-d\lambda_1} + Be^{-d\lambda_2} + c$) where, λ_1, λ_2 refer to the timing of the two waves of Neanderthal gene flow.

For Oase1, based on the recommendation in (10), we ran our single-sample statistic for the genetic distance range of 0.02 to 65cM. As the Neanderthal ancestry blocks are longer in this sample, we used a larger bin size of 0.1cM, which aids in visualization. Dates for different bin sizes between 0.001-1cM were consistent (Figure S4). As this sample has a history of multiple Neanderthal gene flow events, we tried two models: single exponential and double exponentials to estimate dates of Neanderthal gene flow.

(b) Population-sample statistic

For extant populations for which we had access to many individuals, we applied the population-sample statistic described in (11) to estimate the dates of Neanderthal admixture. For each pair for ascertained SNPs $S(d) = (i, j)$ at genetic distance d, we computed the statistic $D(i, j)$ and then considered the average across all pairs of SNPs to estimate $D(d)$.

$$D(d) = \frac{\sum_{(i,j)\in S(d)} D(i,j)}{|S(d)|}$$

Here, $D(i, j)$ denotes the standard signed measure of linkage disequilibrium, D, at the SNPs (i, j) (13). To estimate the time of mixture (λ), we used least squares to fit an exponential distribution ($D(d) = Ae^{-d\lambda}$) with d ranging from 0.02 to 1cM in increments of 0.001cM.

Likelihood ratio test to compare models of admixture: Following (14), we performed a likelihood ratio test to check if a null model of a single exponential ($y = Ae^{-nd} + c$), or the alternate model of sum of two exponentials ($y = Ae^{-dn_1} + Be^{-dn_2} + c$) is a better fit to the observed ancestry covariance patterns. Under the single pulse model, *n* reflects the time of the single pulse of admixture, and under the two-pulse mixture model, n_1 reflects the time of the first pulse of Neanderthal mixture and n_2 reflects the time of the second pulse of admixture. The difference between the log likelihood of the null and the alternate hypothesis (-2*$\log_e$(likelihood of null model) + 2*$\log_e$(likelihood of the alternate model)) is expected to be chi-square distributed with 2 degrees of freedom (14).

Genetic map correction and standard error estimation: To account for errors in the genetic map, we applied the genetic map correction described in (11). Briefly, this method models the true unobserved genetic distances (Z_i) using a gamma distribution that is a function of the observed genetic distances (g_i), and a scalar parameter α that measures per distance variance of the map (i.e. the fluctuation of the map).

$$Z_i = \Gamma(\alpha g_i, \alpha)$$

Thus, larger values of α imply a more accurate map. The method then compares the expected variance in number of crossovers (based on the number of meioses in the pedigree dataset) and the observed variance (based on the genetic map of interest) and

infers an approximate posterior distribution of α by Gibbs sampling. Simulations reported in (11) show that this method is effective in characterizing the map uncertainty.

Using this approach, we estimated α separately for the *S map*, African American (AA) map (12) and Oxford CEU LD map (15) as 3,414 ± 13, 2,802 ± 14 and 2,620 ± 9 per Morgan respectively, by comparing each map to the distribution of crossovers observed in the deCODE pedigree dataset containing 71,929 meiosis detected in Icelandic pedigrees (16). We did not use the pedigree based deCODE map (17) for our main analyses as we reserve it for computing our correction. As discussed in (18), our estimates of α for the *S map* and Oxford CEU LD map are higher than reported in (11), which used the Oxford CEU LD map and the pedigree dataset of 728 meiosis detected in European American Hutterite pedigrees (19) for correction. The finer resolution of the deCODE pedigree dataset is likely much more informative about the expected number of crossovers at shorter distances and hence we believe it is a more trustworthy dataset for the correction. A concern with the *S map*, however, is that it uses a subset of the deCODE pedigree data (15,000 meiosis) as a prior for the construction of the genetic map, and then uses a larger dataset (71,929 meiosis) for estimation of α. In principle, this could lead to an overestimation of α for the *S map*. To overcome this issue, similar to (18), we constrained the value of α by using the estimate of map uncertainty based on the African American map (that does not use the deCODE data as a prior but contains African-enriched hotspots) as the lower bound, and by using the estimate based on the *S map* (that may be inflated due to the use of overlapping data) as the upper bound. Specifically, we combined the posterior estimates of α based on the African American (2,802 ± 14 per M) and *S map* (3,414 ± 13 per M) and then assumed a flat prior over this range and sampled new values of α from the combined distribution. This has the effect of sampling from a distribution with the effective mean and uncertainty of 3,109 ± 308 per M.

Given an estimate of (α) and date of Neanderthal gene flow (λ) (either estimated using *C(d)* or *D(d)),* we computed the corrected date of gene flow in generations (t_n in generations) using:

$$t_n = \alpha\left(exp\left(\frac{\lambda}{\alpha}\right) - 1\right)$$

Further, we assumed a uniform prior distribution on the number of years per generation of 25–33 years (20-23), and integrated over the uncertainty in generation interval to obtain corrected dates of gene flow in years (t_n in years).

Estimating the age of the ancient genomes: To estimate the age (t_c) of the ancient genome of interest (Clovis, Mal'ta1, Kostenki14, Ust'-Ishim and Oase1), we quantified the difference in dates of Neanderthal admixture in ancient genome (t_{na}) (estimated using the single-sample statistic $C(d)$) and extant CEU genomes (t_{ne}) (estimated using the population-sample statistic $D(d)$). We estimated standard errors based on the Bayesian framework described above. We report the mean and standard error of genetic ages in generations and years.

Web resources: The software implementing the dating approach will be made available for download from the following URLs:
single-sample statistic: https://github.com/priyamoorjani/Neanderthal_dating
population-sample statistic: https://github.com/sriramsr/dating

Note S2: Simulations

To assess the reliability of our method, we performed coalescent simulations for various demographic scenarios. For all simulations described below, we used the coalescent simulator *ms* (24) to generate data for 50 regions of 50 Mb each for Europeans, west Africans and Neanderthals with the following parameters. These parameters were chosen so that population differentiation (F_{ST}) between west Africans and Europeans and the D-statistic between west Africans, Europeans and Neanderthals (D(Y, E; N)) matched estimates from real data (F_{ST} = 0.15, D = 0.05 (11))

- Mutation rate = 1.5×10^{-8} per bp/generation
- Recombination rate = 1×10^{-8} per bp/generation
- Effective population size (N_e) of modern humans = 10,000.
- Effective population size (N_e) of Neanderthals = 2,500.
- Neanderthals were sampled 2,000 generations ago.
- Europeans split from Africans 3,000 generations before present.
- Extant Europeans split from ancient Europeans 1,900 generations before present.
- Modern humans split from Neanderthal 12,000 generation before present.
- Gene flow from Neanderthals into European ancestors occurred instantaneously and contributed 3% ancestry to Europeans.

To date the Neanderthal gene flow in Europeans, we ascertained Neanderthal informative SNPs based on ascertainment 0 where Neanderthals carry at least one derived allele and west Africans are fixed for the ancestral allele.

(a) Estimation of the date of Neanderthal admixture

We simulated data using *ms* for 44 haploid genomes: 20 Europeans, 20 west Africans, and four Neanderthals. We simulated data under a simple demographic model with gene flow from Neanderthals into Europeans that occurred instantaneously at varying times of mixture (100-2,500 generations ago). *ms* does not have an option to simulate ancient samples whose evolution stopped at a time *T* in the past. However, we can change the parameters of the program to exactly produce data such as are expected from an ancient sample that lived at time *T*. To achieve this, at time *T*, we split the population into two subpopulations (the idea is to use a single sample from each of these two populations—from *T* generations to present —to represent the two haplotypes of the ancient individual. We set the effective population size to a negligible value (order of 10^{-10}) for both the subpopulations just after their split. This has the effect of creating two inbred lines that freeze in the genetic variation that existed in the ancient haplotypes. There is no change in the frequencies of the alleles in these lineages afterward, as all alleles inherited from the ancestral population are fixed. (Recombination does not change patterns of genetic variation in these populations, since any recombination is between identical DNA). A complication in this procedure is that *ms* will continue to produce new mutations in the two sub-populations between time *T* and 0. However, any new mutations will not be shared between the two subpopulations. Thus, by restricting analysis to alleles that are shared in the two lineages, we have a snapshot of the variation that existed in the ancient individual, with all effects of new simulated mutation since time *T* canceled out. We use this approach to sample Neanderthals at *T*=2,000 generations ago. We combined two haploid chromosomes at random to generate one diploid chromosome.

The *ms* command line is as follows. Here, pop1 = west Africans, pop2 =

Europeans, pop3-6 = Neanderthals (to simulate branch shortening in *ms*, we generate four haploid chromosomes to obtain one diploid ancient genome) and t_n = time of Neanderthal gene flow (between 100 – 2,500 generations, appropriately scaled by $4N_e$ to be in the units needed for *ms*). Estimated F_{ST} and D-statistics for each simulation are shown in Table S2a.1.

ms 44 1 -r 20000 50000000 -t 30000 -I 6 20 20 1 1 1 1 -en 0 1 1 -en 0 2 1 -en 0 3 1e-10 -en 0 4 1e-10 -en 0 5 1e-10 -en 0 6 1e-10 -es t_n 2 0.97 -en 0.02500025 7 0.25 -en 0.02500025 2 1 -ej 0.05 4 3 -ej 0.05 6 5 -en 0.05000025 3 0.25 -en 0.05000025 5 0.25 -ej 0.0500025 5 3 -en 0.050005 3 0.25 -ej 0.075 2 1 -en 0.0750025 1 1 -ej 0.1 7 3 -en 0.1000025 3 0.25 -ej 0.3 3 1 -en 0.3000025 1 1

Table S2a.1: Summary statistics in simulated data

Expected date (in generations)	***F_{ST}(Y, E)***	***D(Y, E; N)***
500	0.127	0.0699
1,000	0.127	0.0659
1,500	0.129	0.0643
2,000	0.129	0.0670
2,500	0.131	0.0696

Note: Y= West Africans, E= Europeans and N = Neanderthals. We computed F_{ST}(Y, E) using smartpca (25) and we computed D-statistics used ADMIXTOOLS (26). Here, we don't need an outgroup as we have the ancestral allele in simulations.

We applied the single-sample statistic (Note S1) to all pairs of ascertained SNPs in five randomly chosen simulated diploid genomes. We fitted a single exponential distribution to estimate the date of the Neanderthal gene flow. To compute standard errors, we used a weighted block jackknife procedure (27), removing one region (50 Mb) in each run to estimate the variability in the estimated dates across the genome. Figure S2a.1 shows that for dates between 100-1,500 generations ago, we obtained accurate and precise dates of mixture. However, for dates $\geq$2000 generations, we observed decreased precision and a downward bias in the estimated dates.

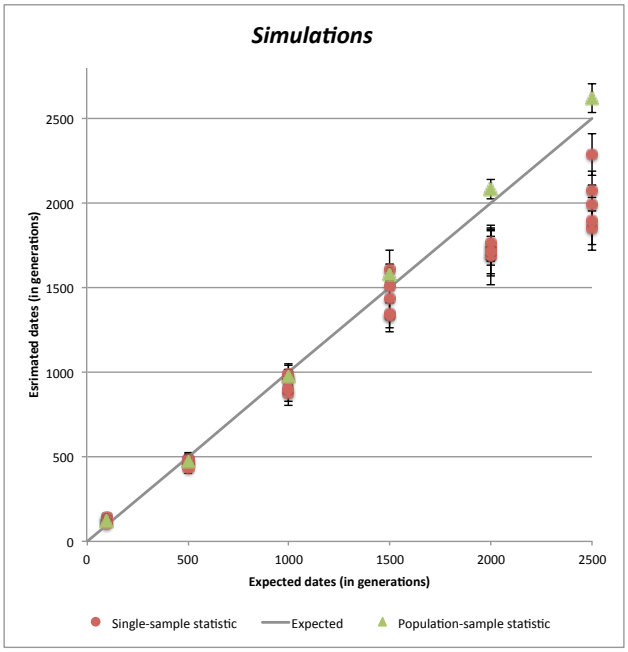


Figure S2a.1: Estimated dates of Neanderthal gene flow. For each time point (λ), we simulated diploid European genomes that derive 3% of their ancestry from Neanderthal gene flow that occurred λ generations ago. Shown is the true time against the estimated mean and standard error of the dates based on two statistics (single-sample and population-sample statistics) in generations for five diploid individuals. Standard errors were computed based on a weighted block jackknife (27), removing one region (50 Mb) in each run.

We applied the population-sample statistic (see Note S1) to date the Neanderthal admixture in five random simulated diploid ancient Europeans. For all time-depths, we obtained relatively unbiased dates of mixture (Figure S2a.1).

(b) Simulations of complex demographic scenarios

The simulations reported in Fu et al. (9) showed that under complex scenarios of severe bottlenecks and large population expansions, the single-sample statistic underestimates the older dates of Neanderthal admixture (>2,000 generations ago). In contrast, little or no bias is observed for recent dates of Neanderthal admixture (9). To test the performance of our statistics for inferring dates of Neanderthal admixture in this sample, we used the same data generated by Fu et al. (9) (Supplementary Information S18: Simulation 4). Briefly, the authors simulated data for 10 African, 20 European (Eu), 20 Asian (As), one Altai Neanderthal and one ancient non-African genome similar to Ust'-Ishim sampled 1,800 generations ago (ancient). Simulation parameters used in (9) were based on the demographic model described in (28) with slight modification to match the constraints from ancient DNA analyses (see (9) for details). Parameters for the simulations (that differ from Note S2 (a)) are as follows:

- Effective population size (N_e) of modern humans = 14,000.
- Both European and East Asian populations undergo expansions until the present (*t = 0*) and the present-day East Asians have N_e of 45,300 and present-day Europeans have N_e of 33,800.
- Europeans and East Asians split 2,000 generations ago with a reduction in N_e in East Asians to 550, and N_e in Europeans to 1,032.
- A population expansion occurred in the common ancestor of present-day humans at 6,000 generations ago with N_e increasing from 7,000 to 14,000.
- Gene flow from Neanderthals into the ancestors of present-day non-Africans occurred at 2,200 generations before present.
- Split time of non-Africans = 2,000 generations before present.
- We sample an ancient genome (similar to Ust'-Ishim) at 1,800 generations before present
- Altai splits from the introgressing Neanderthal 4,000 generations before present.
- We sample Altai genome at 2,400 generations before present.

We used the *ms* command from (9) as follows:

Here, pop1= African, pop2=present-day Europeans, pop3=present-day Asians, pop4-7= ancient, pop8-11= Neanderthal, and g_E = 2.41 = population size in present-day Europeans (in units of N_e) and g_A = 3.24 = population size in present-day Asians (in units of N_e). $\alpha_A = 97.691, \alpha_E = 123.51$.

ms 68 1 -I 11 20 20 20 1 1 1 1 1 1 1 1 -en 0 1 1 -en 0 2 g_E -en 0 3 g_A -eg 0 2 α_A -eg 0 3 α_E -en 0 4 7.14285714285714e-11 -en 0 5 7.14285714285714e-11 -en 0 6 7.14285714285714e-11 -en 0 7 7.14285714285714e-11 -en 0 8 7.14285714285714e-11 -en 0 9 7.14285714285714e-11 -en 0 10 7.14285714285714e-11 -en 0 11 7.14285714285714e-11 -ej 0.0321428571428571 5 4 -ej 0.0321428571428571 7 6 -en 0.0321430357142857 4 0.714285714285714 -en 0.0321430357142857 6 0.714285714285714 -ej 0.0321446428571429 6 4 -en 0.0321464285714286 4 0.714285714285714 -ej 0.0357142857142857 2 3 -ej 0.0357142857142857 4 3 -en

0.0357160714285714 3 0.132857142857143 -en 0.0357160714285714 3 0.132857142857143 -es 0.0392857142857143 3 0.97 -en 0.0392875 12 0.178571428571429 -en 0.0392875 3 0.132857142857143 -ej 0.0428571428571429 9 8 -ej 0.0428571428571429 11 10 -en 0.0428573214285714 8 0.178571428571429 -en 0.0428573214285714 10 0.178571428571429 -ej 0.0428589285714286 10 8 -en 0.0428607142857143 8 0.178571428571429 -ej 0.0535714285714286 3 1 -en 0.0535732142857143 1 1 -ej 0.0714285714285714 12 8 -en 0.0714303571428571 8 0.178571428571429 -en 0.107142857142857 1 0.521428571428571 -ej 0.214285714285714 8 1 -en 0.2142875 1 0.714285714285714 -r 28000 50000000 -t 42000 -p 12 -seeds 46 47 48

Table S2b.1: Summary statistics in simulated data from Fu et al. (2014).

Statistic applied	Estimated
F_{ST}(Y, Eu)	0.236
F_{ST}(Y, As)	0.257
F_{ST}(Eu, As)	0.242
D(Y, Eu; N)	0.0745
D(Y, As; N)	0.0663

Note: Y= West Africans, Eu= Europeans, As=Asians and N = Neanderthals.

We note that the observed F_{ST} and D-statistics in this simulation are more extreme than observed in real data, likely because some of the parameters (such as bottleneck time or strength) are more extreme than in real data (Table S2b.1). To estimate the date of Neanderthal gene flow, we applied the single-sample and population-sample statistic to all pairs of ascertained SNPs (Note S1). We computed standard errors using a weighted Block Jackknife as described above. Table S2b.2 shows that while the dates of the Neanderthal admixture based on the single-sample statistic are downward biased for extant Europeans and Asians, the dates based on the population-sample statistic are accurate. We obtained reliable dates for the ancient genome (similar to Ust-Ishim).

Table S2b.2: Simulations of complex admixture scenario from Fu et al. 2015

Individual	Expected date (in generations)	Estimated date based on single-sample statistic	Estimated date based on population-sample statistic
Ancient	400	329 ± 50	n/a
Eu1	2,200	968 ± 306	2,207 ± 133*
Eu2	2,200	727 ± 198	
Eu3	2,200	820 ± 133	
Eu4	2,200	1,031 ± 236	
As1	2,200	797 ± 107	2,237 ± 182*
As2	2,200	910 ± 139	
As3	2,200	789 ± 133	
As4	2,200	746 ± 78	

Note: * indicates combined estimate for four samples. n/a = not applicable as the population-based statistic cannot be applied to a single sample. Here, ancient = simulated ancient genome, Eu = simulated European genomes and As = simulated Asian genomes. Standard errors are computed using a weighted block jackknife, removing one region (50 Mb) in each run.

To further test if a history of population expansion can cause a bias in our analysis, we performed two more simulations where we changed the growth rates to 10 times (population size in East Asians and Europeans during the expansion is 453,000 and 338,000 respectively) and 50 times (population size in East Asians and Europeans during the expansion is 2,265,000 and 1,690,000 respectively) the values used in Fu et al. (9). The other parameters of the simulation were same as before. Our results showed that our inferences were robust to recent expansions (Table S2b.3, Table S2b.4).

Table S2b.3: Simulations of complex admixture scenario similar to Fu et al. (9) **with ten times larger population size during the recent expansion**

Individual	Expected date (in generations)	Estimated date based on single-sample statistic	Estimated date based on population-sample statistic
ancient	400	375 ± 39	n/a
Eu1	2,200	1,000 ± 101	2,301 ± 123*
Eu2	2,200	813 ± 130	
Eu3	2,200	1,059 ± 159	
Eu4	2,200	1,240 ± 130	
As1	2,200	736 ± 86	2,195 ± 114*
As2	2,200	841 ± 105	
As3	2,200	795 ± 136	
As4	2,200	900 ± 143	

Note: * indicates combined estimate for four samples. n/a = not applicable as the population-based statistic cannot be applied to a single sample. Here, ancient = simulated ancient genome, Eu = simulated European genomes and As = simulated Asian genomes. We computed standard errors using a weighted block jackknife, removing one region (50 Mb) in each run. To match the data size to real data, we restricted our analysis to randomly sampled 250,000 ascertained SNPs.

Table S2b.4: Simulations of complex admixture scenario similar to Fu et al. (9) **with fifty times larger population size during the recent expansion**

Individual	Expected date (in generations)	Estimated date based on single-sample statistic	Estimated date based on population-sample statistic
ancient	400	367 ± 30	n/a
Eu1	2,200	1,104 ± 153	2,180 ± 94*
Eu2	2,200	1,054 ± 110	
Eu3	2,200	994 ± 113	
Eu4	2,200	951 ± 144	
As1	2,200	881 ± 121	2,272 ± 111*
As2	2,200	853 ± 111	
As3	2,200	761 ± 159	
As4	2,200	901 ± 140	

Note: * indicates combined estimate for four samples. n/a = not applicable as the population-based statistic cannot be applied to a single sample. Here, ancient = simulated ancient genome, Eu = simulated European genomes and As = simulated Asian genomes. We computed standard errors using a weighted block jackknife, removing one region (50 Mb) in each run. To match the data size to real data, we restricted our analysis to a randomly sampled set of 250,000 ascertained SNPs.

(c) Estimating the age of ancient genomes

To assess the reliability of our method for estimating the age of an ancient sample, we performed coalescent simulations where we sampled ancient Europeans at some time in the past (between 500-1,750 generations ago). We simulated the ancient genomes using the same set up as described for Neanderthals in Note S2 (a), such that the ancient genome of age *T* had a very small effective population size from time *T* to present and an additional subpopulation was simulated to remove mutations that might have occurred after *T*. Ancient and extant Europeans derived 3% of their ancestry from Neanderthals from a gene flow event that occurred 2,000 generations ago. The observed F_{ST} and D are as shown in Table S2c.1.

The *ms* command for this simulation was as follows. Here, pop1 = Africans, pop2 = extant Europeans, pop3-6 = ancient European, pop7-10 = Neanderthal and t_n = age of the ancient genome (between 500-1,750 generations ago, appropriately scaled by $4N_e$).

ms 48 1 -I 10 20 20 1 1 1 1 1 1 1 1 -en 0 1 1 -en 0 2 1 -en 0 3 1e-10 -en 0 4 1e-10 -en 0 5 1e-10 -en 0 6 1e-10 -en 0 7 1e-10 -en 0 8 1e-10 -en 0 9 1e-10 -en 0 10 1e-10 -ej t_n 4 3 -ej t_n 6 5 -en t_n 3 1 -en t_n 5 1 -ej t_n 5 3 -en t_n 3 1 -ej t_n 3 2 -en t_n 2 1 -ej 0.05 8 7 -ej 0.05 10 9 -es 0.05 2 0.97 -en 0.05000025 7 0.25 -en 0.05000025 9 0.25 -ej 0.0500025 9 7 -en 0.0500025 11 0.25 -en 0.0500025 2 1 -en 0.050005 7 0.25 -ej 0.075 2 1 -en 0.0750025 1 1 -ej 0.1 11 7 -en 0.1000025 7 0.25 -ej 0.3 7 1 -en 0.3000025 1 1 -r 20000 50000000 -t 30000 -p 12 -seeds 43 44 45

Table S2c.1: Summary statistics in simulated data.

Age of ancient sample (in generations)	***F_{ST}(Y, E)***	***D(Y, E; N)***
500	0.129	0.0681
1,000	0.129	0.0734
1,250	0.129	0.0719
1,500	0.129	0.0645
1,750	0.130	0.0668

To estimate the age of the ancient genome, we first estimated the dates of Neanderthal admixture in ancient genomes (using the single-sample statistic) and extant genomes (using the population-sample statistic). The difference in the dates of Neanderthal gene flow in the extant and ancient genomes translates into a direct estimate of the age of the ancient genome. Standard errors were computed using weighted Block Jackknife as described above. Table S2c.2 shows that our approach provides reliable age estimates for all time depths.

Table S2c.2: Estimated ages of simulated ancient genomes under simple admixture scenarios

Sim.	*Estimated date of gene flow in extant* (t_{na}) mean ± SE (Expected)	*Estimated date of gene flow in ancient* (t_{ne}) mean ± SE (Expected)	*Estimated age of ancient sample* (t_c) mean ± SE (Expected)
Sim 1	2060 ± 43 (2000)	1312 ± 123 (1500)	748 ± 130 (500)
Sim 2	2021 ± 37 (2000)	945 ± 71 (1000)	1075 ± 80 (1000)
Sim 3	2116 ± 44 (2000)	762 ± 46 (750)	1354 ± 64 (1250)
Sim 4	2092 ± 44 (2000)	561 ± 38 (500)	1531 ± 58 (1500)
Sim 5	2135 ± 46 (2000)	271 ± 25 (250)	1864 ± 52 (1750)

Note: Expected dates shown in brackets in red. Standard errors were computed using weighted block jackknife, removing one region (50 Mb) in each run.

To test the effect of complex admixture scenarios that includes bottlenecks and expansions on our inference; we repeated the simulations in Table S2b.2 (model from Fu et al. (9)) for ancient genomes that we sampled at varying time depths (between 700-1,800 generations ago). We inferred the age of the ancient genomes by comparing the dates of Neanderthal admixture in the ancient and extant European genomes. We obtained reliable age estimates (Table S2c.3).

Table S2c.3: Estimated ages of simulated ancient genomes under complex admixture scenarios

Sim.	*Estimated date of gene flow in extant* (t_{na}) mean ± SE (Expected)	*Estimated date of gene flow in ancient* (t_{ne}) mean ± SE (Expected)	*Estimated age of ancient sample* (t_c) mean ± SE (Expected)
Sim 1	2220 ± 61 (2200)	1285 ± 81 (1500)	935 ± 107 (700)
Sim 2	2098 ± 58 (2200)	1148 ± 89 (1250)	950 ± 106 (950)
Sim 3	2232 ± 77 (2200)	947 ± 67 (1000)	1285 ± 102 (1200)
Sim 4	2377 ± 70 (2200)	759 ± 60 (700)	1618 ± 93 (1500)
Sim 5*	2207 ± 137 (2200)	329 ± 50 (400)	1878 ± 142 (1800)

Note: * indicates data from Table S2b.2. To match the data size to real data, we restricted our analysis to randomly sampled 250,000 ascertained SNPs.

(d) Continuous admixture:

We performed coalescent simulations for a continuous admixture scenario where gene flow occurred gradually for a period of either 10 or 500 generations, starting at 2,000 generations ago. We chose all parameters to be similar to Table S2c.2. We estimated the dates of Neanderthal admixture under the assumption of a single gene flow event by fitting a single exponential distribution to the output of *C(d)* and *D(d)*.

The *ms* command for these simulations was as follows. Here, pop1 = Africans, pop2 = extant Europeans, pop 3-6 = ancient European, pop7-10 = Neanderthal, t_n = age of the ancient genome (between 500-1,750 generations ago, appropriately scaled by $4N_e$) and $t_{gf\text{-}start}$, $t_{gf\text{-}end}$ are set to the time of onset (2,000 generations) and cessation (2000 + x, where x = 10 or 500 generations) of the period of gene flow and m_i = 0.03/x is the per generation migration rate, which is set so that overall proportion of gene flow is 3%.

ms 48 1 -r 20000 50000000 -t 30000 -I 10 20 20 1 1 1 1 1 1 1 1 -en 0 1 1 -en 0 2 1 -en 0 3 1e-10 -en 0 4 1e-10 -en 0 5 1e-10 -en 0 6 1e-10 -en 0 7 1e-10 -en 0 8 1e-10 -en 0 9 1e-10 -en 0 10 1e-10 -ej t_n 4 3 -ej t_n 6 5 -en t_n 3 1 -en t_n 5 1 -ej t_n 5 3 -en t_n 3 1 -ej 0.0475 3 2 -en 0.0475025 2 1 -ej 0.0475 8 7 -ej 0.0475 10 9 -en 0.0475025 7 0.25 -en 0.0475025 9 0.25 -ej 0.0475025 9 7 -en 0.0475025 7 0.25 -em $t_{gf\text{-}start}$ 2 7 m_i -em $t_{gf\text{-}end}$ 2 7 0 -ej 0.075 2 1 -en 0.0750025 1 1 -ej 0.3 7 1 -en 0.3000025 1 1 -seeds 432122 598689 608300

Table S2d.1: Simulations for 10 generations of continuous gene flow, starting at 2,000 generations ago.

Sim.	*Estimated date of gene flow in extant* (t_{na}) mean ± SE (Expected)	*Estimated date of gene flow in ancient* (t_{ne}) mean ± SE (Expected)	*Estimated age of ancient sample* (t_c) mean ± SE (Expected)
Sim 1	2103 ± 58 (2000)	1394 ± 97 (1500)	709 ± 113 (500)
Sim 2	2061 ± 65 (2000)	859 ± 64 (1000)	1202 ± 91 (1000)
Sim 3	1998 ± 64 (2000)	731 ± 58 (750)	1267 ± 87 (1250)
Sim 4	2161 ± 77 (2000)	557 ± 37 (500)	1604 ± 86 (1500)
Sim 5	2166 ± 67 (2000)	277 ± 26 (250)	1889 ± 72 (1750)

Note: To match the data size to real data, we restricted our analysis to randomly sampled 250,000 ascertained SNPs.

Table S2d.2: Simulations for 500 generations of continuous gene flow, starting at 2,000 generations ago.

Sim.	*Estimated date of gene flow in extant* (t_{na}) mean ± SE (Expected)	*Estimated date of gene flow in ancient* (t_{ne}) mean ± SE (Expected)	*Estimated age of ancient sample* (t_c) mean ± SE (Expected)
Sim 1	2323 ± 80 (2000)	1569 ± 120 (1500)	754 ± 144 (500)
Sim 2	2273 ± 65 (2000)	1193 ± 98 (1000)	1080 ± 118 (1000)
Sim 3	2175 ± 91 (2000)	880 ± 48 (750)	1295 ± 103 (1250)
Sim 4	2263 ± 93 (2000)	701 ± 49 (500)	1562 ± 105 (1500)
Sim 5	2292 ± 85 (2000)	427 ± 37 (250)	1865 ± 92 (1750)

Note: To match the data size to real data, we restricted our analysis to randomly sampled 250,000 ascertained SNPs.

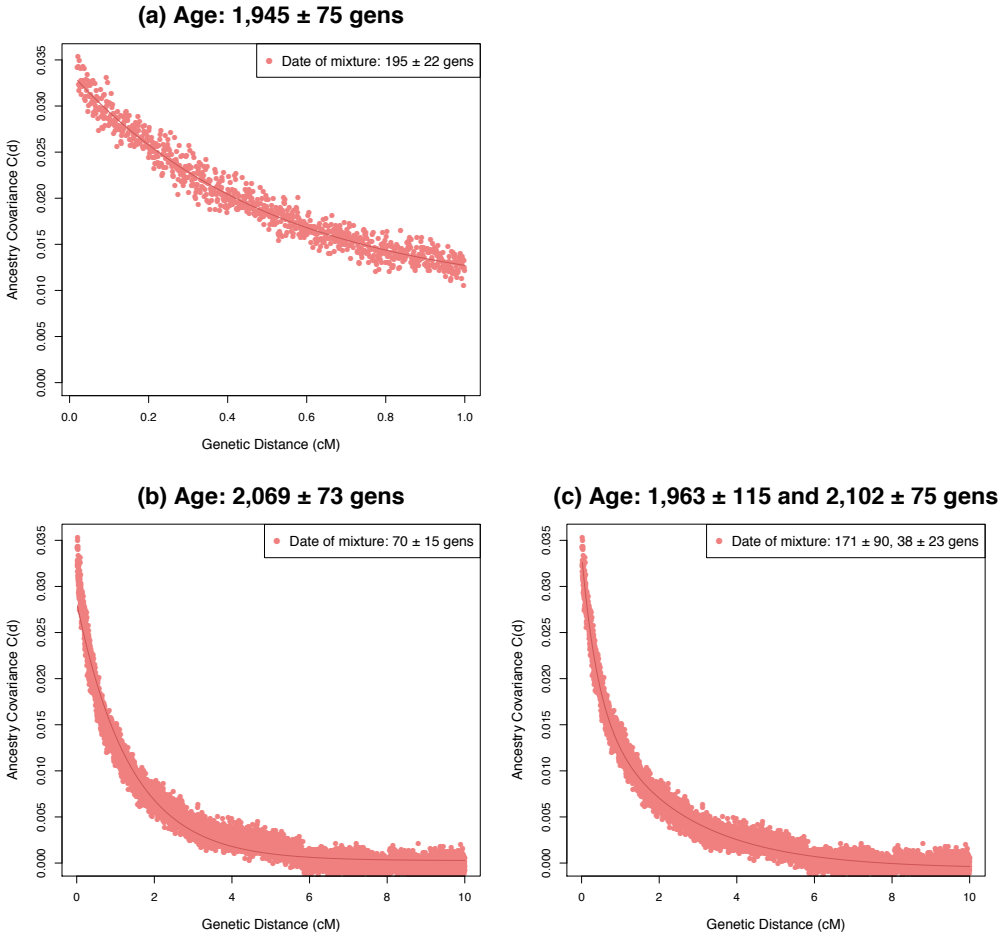
(e) Multiple pulse admixture:

We performed a simulation choosing parameters to be broadly similar to the history of Neanderthal gene flow in Ust'-Ishim. We generated data for a panel of extant genomes and one ancient sample (that died 1,750 generation ago). The ancient genome received Neanderthal ancestry from two distinct gene flow events, one that occurred 2,000 generations ago and was shared with extant Europeans and the other that occurred 1,800 generations ago. The latter was not shared with other Europeans. To date the Neanderthal admixture, we ran the single-sample statistic to 1cM as before. The ancient sample contained many Neanderthal segments that were longer than 1cM, as expected for a sample that has a history of recent admixture within the past 50 generations. Thus, to fully capture the signal of admixture, we ran the analysis up to 10cM and then fitted a single exponential and a sum of exponentials to the output of *C(d)* (Figure S2e.1).

The *ms* command for the simulation was as follows. Here, pop1 = Africans, pop2 = extant Europeans, pop3-6 = Ust, pop7-10 = Neanderthal.

ms 48 1 -r 20000 50000000 -t 30000 -I 10 20 20 1 1 1 1 1 1 1 1 -en 0 1 1 -en 0 2 1 -en 0 3 1e-10 -en 0 4 1e-10 -en 0 5 1e-10 -en 0 6 1e-10 -en 0 7 1e-10 -en 0 8 1e-10 -en 0 9 1e-10 -en 0 10 1e-10 -ej 0.04375 4 3 -ej 0.04375 6 5 -en 0.0437525 3 1 -en 0.0437525 5 1 -ej 0.0437525 5 3 -en 0.0437525 3 1 -em 0.045 3 7 400 -em 0.045025 3 7 0 -ej 0.0475 3 2 -en 0.0475025 2 1 -es 0.05 2 0.97 -en 0.0500025 11 0.25 -en 0.0500025 2 1 -ej 0.05 8 7 -ej 0.05 10 9 -en 0.0500025 7 0.25 -en 0.0500025 9 0.25 -ej 0.0500025 9 7 -en 0.0500025 7 0.25 -ej 0.075 2 1 -en 0.0750025 1 1 -ej 0.1 11 7 -en 0.1000025 7 0.25 -ej 0.3 7 1 -en 0.3000025 1 1 -seeds 510741 263809 345175

Figure S2e.1: Simulation results for a two-pulse model of Neanderthal admixture. Estimated dates of Neanderthal gene flow (mean ± SE) in an ancient European genome are shown in pink. We compared these to dates of Neanderthal admixture in extant samples of 2,140 ± 71 generations to infer the age of the ancient genomes (shown in the title). To match the data size to real data, we restricted our analysis to randomly sampled 250,000 ascertained SNPs. (a) results based on single exponential fit up to the genetic distance of 1 cM, (b) results based on single exponential fit up to the genetic distance of 10 cM, (c) results based on a sum of exponentials fit up to genetic distance of 10 cM.



Note S3: Robustness to data ascertainment

a. Robustness to how genotypes were determined: For most of our ancient samples, we did not have sufficient coverage to make reliable diploid genotype calls. Instead, we used pseudo-homozygous calls based on the majority allele observed in reads for each sample at each site (see Note S1). To verify the robustness of our inferences, we repeated our analysis with pseudo-homozygous calls where we sampled a random allele seen in the reads mapped to each site in each sample (pseudo-homozygous (random)). In addition, for our high coverage Ust'-Ishim genome, we compared inferences based on diploid and pseudo-homozygous calls (both majority and random sampling).

For all ancient genomes, we observed that the dates are consistent across different genotype calling approaches (Table S3a.1, Figure S2). These results highlight a strength of our method: that it works well even using pseudo-homozygous calls and for samples with low coverage (such as Mal'ta1 with an average coverage of 1.0x).

Table S3a.1: Effect of genotype calling approach

Age of sample	*Diploid*	*Pseudo-homozygous (major)*	*Pseudo-homozygous (random)*
Clovis	n/a	18,066 ± 5,112*	17,440 ± 5,146
Mal'ta1	n/a	24,935 ± 4,851*	24,868 ± 4,828
Kostenki14	n/a	41,189 ± 4,387*	40,358 ± 4,399
Ust'-Ishim (B)	44,560 ± 4,175*	44,572 ± 4,175	44,508 ± 4,175

Note: * indicates the approach used in main text. We selected SNPs based on ascertainment 0, the genetic distances were based on the *S map*, and made a genetic map correction based on the data from (17) (see Note S1). n/a – indicates that the coverage is not sufficient to make reliable diploid calls. The Ust'-Ishim dates are based on the model of a single exponential fitted to genetic distances up to 10cM.

For most of the analyses, we used the genotype calls based on low coverage data from 1000 Genomes CEU individuals to represent extant non-Africans. To evaluate whether our results are sensitive to errors introduced due to coverage or the use of northern Europeans as the comparison set, we repeated our analysis with high coverage data from west Eurasians that are part of the Simons Genome Diversity Project (SGDP) (29). This dataset includes between 1-5 samples from the following west Eurasian groups: Albanian, Czech Republic, Samaritan, Norwegian, Polish, Chechen, Abkhasian, Armenian, Bulgarian, English, Estonian, Saami, Basque, Georgian, Greek, Hungarian, Icelandic, Iranian, Iraqi Jew, Druze, Bedouin, Bergamo, Tuscan, Orcadian, Russian, Lezgin, North Ossetian, Adygei, Spanish, Tajik, Turkish, Yemenite Jew, Crete, Finnish. Palestinian, Jordanian, French and Sardinian. For sampling location and other details, please refer to (29). Applying the population sample statistic to SGDP west Eurasians, we estimate that Neanderthal gene flow occurred 38,168–51,270 years ago, similar to the results based on 1000 Genomes data. Age estimates based on these dates (Table S3a.2) were similar to those reported in Table S1 and main text.

Table S3a.2: Age estimated based on high coverage SGDP west Eurasians

Age of sample	*Using low coverage extant genomes: 1000 Genomes CEU**	*Using high coverage extant genomes: SGDP west Eurasians*
Clovis	18,066 ± 5112	15,299 ± 4712
Mal'ta	24,935 ± 4851	22,168 ± 4428
Kostenki	41,189 ± 4387	38,422 ± 3914
Ust'-Ishim (B)	44,560 ± 4175	41,718 ± 3957

Note: * indicates the approach used in main text. We selected SNPs based on ascertainment 0, used genetic distances from the *S map*, and performed a genetic map correction based on data from (17) (see Note S1).

b. *Robustness to the way SNPs were selected:* To study the effect of marker ascertainment, we analyzed data for additional SNP selection schemes shown to be informative for dating Neanderthal ancestry in (11). As previous studies have shown that Luhya have some recent West Eurasian ancestry (~2.4%) (30), which could affect the dating, we explored an ascertainment of selecting SNPs only using YRI and Altai Neanderthals (similar to ascertainment used for Oase1). The alternative ascertainment schemes considered were ones in which:

(Ascertainment 0) Altai Neanderthal carries at least one derived allele and all individuals in a panel of sub-Saharan Africans (1000 genomes YRI and LWK) carry the ancestral allele. This is the main ascertainment we use in the study.

(Ascertainment 1) Altai Neanderthal carries the derived allele, Africans (1000 genomes YRI and LWK) carry the ancestral allele, and Europeans (1000 genomes CEU) are polymorphic.

(Ascertainment 2) Altai Neanderthal carries the derived allele and Europeans (1000 genomes CEU) are polymorphic and have derived allele frequency of less than 20%. Our threshold is higher than the 10% used in (11). The 20% threshold provides more SNPs (thus improved precision) and is still informative about Neanderthal ancestry (11).

(Ascertainment 3) Altai Neanderthal carries the derived allele and Africans (1000 Genomes YRI) carry the ancestral allele.

The age estimates based on different ascertainments remain consistent for all four ancient genomes and are similar to the radiocarbon dates (Table S3b.1). However, ascertainments 1 and 2 are not symmetric with regard to ancient and extant genomes (we used extant Europeans for selecting SNPs while excluding ancient samples due to their low coverage). This asymmetry could in theory lead to a bias in case the two populations do not share the same admixture history. Thus, ascertainment 3 and ascertainment 0 seem preferable on first principles.

Table S3b.1: Effect of SNP ascertainment on age estimates

Sample	*ascertainment 0**	*ascertainment 1*	*ascertainment 2*	*ascertainment 3*
Number of ascertained SNPs included	221,482	164,995	486,804	280,441
Date of Neanderthal gene flow in CEU	47,236 ± 4339	43,693 ± 4026	48,891 ± 4564	42,253 ± 3818
Age of Clovis	18,066 ± 5112	19,600 ± 4615	18,272 ± 5306	12,958 ± 4677
Age of Mal'ta1	24,935 ± 4851	23,918 ± 4466	20,034 ± 5325	23,197 ± 4222
Age of Kostenki14	41,189 ± 4387	38,027 ± 4083	39,886 ± 4650	34,930 ± 3880
Age of Ust'-Ishim (B)	44,560 ± 4175	41,746 ± 3924	45,524 ± 4289	39,605 ± 3777

Note: * indicates the approach used in main text. For all four ascertainments, we used the genetic distances based on the *S map* and the genetic map correction based on data from (17). Ust'-Ishim dates are based on model of single exponential fitted to genetic distance of up to 10cM.

Note S4: Estimating the historical generation interval

To estimate the historical generation interval in humans since the Neanderthal introgression, we characterize the correlation between the dates of Neanderthal admixture (in generations) and radiocarbon dates (in years). A simplifying assumption in what follows is that the generation interval in males and females is the same and that the generation interval has remained constant since Neanderthal admixture.

One approach is to use ordinary least squares (OLS) to estimate the generation interval by regressing the radiocarbon dates on the estimated dates of Neanderthal admixture. However, OLS assumes that the independent variables are measured precisely, the measurement error is negligible and the residuals have constant variance. These assumptions are not met in our data. Instead, we implement a Bayesian approach using importance sampling to estimate these parameters (31).

Model:

We define:

$$y' = y + \theta_1$$

Here, y' = true radiocarbon date in years, y is the observed radiocarbon date and θ_1 is the error in radiocarbon dates.

We also define:

$$x' = x + \theta_2$$

Here, x' = true date of Neanderthal admixture in generations, x is the estimated date of admixture and θ_2 is the error in the estimated date.

We are interested in learning the parameters of the linear relationship between y' and x':

$$y' = \beta x' + c + \theta_3$$

Here, β reflects the generation interval (in years), c reflects the time of the shared date of Neanderthal admixture (in years) and θ_3 is the error in the model arising from the fact that generation interval is not exactly the same for all samples or may have changed over time, etc.

Thus,

$$y - \theta_1 = \beta(x + \theta_2) + c + \theta_3$$
$$y = \beta x + c + \theta_1 + \beta\theta_2 + \theta_3$$

Let $z = \beta x + c$

Thus, $y = z + n$
where, $n = \theta_1 + \beta\theta_2 + \theta_3$
$\mathrm{E}(n) = 0$ and $\sigma^2[n] = \mathrm{E}(n^2) - \big(\mathrm{E}(n)\big)^2 = \theta_1^2 + \beta^2\theta_2^2 + \theta_3^2$

The likelihood of z is as follows:

$$\mathcal{L}(z) = \prod_i \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(y_i - z_i)^2}{\sigma_i}\right)$$

As we do not know the exact distribution of our unknown parameters $(h(x))$, we use importance sampling (31) to estimate expected value of the parameters.

$$\int h(x)\, p(x) dx = \int h(x)\, \frac{p(x)}{g(x)}\, g(x) dx = \int h(x)\, w(x)\, g(x) dx$$

here, $g(x)$ is another density function whose sample space is similar to $p(x)$; $w(x)$ is the importance function. In our analysis, we use $g(x)$ as the prior probability distribution of $h(x)$ and use $\mathcal{L}(z)$ as the importance function.

Brief algorithm:

1. For each iteration (*k*):

- Sample a value of $\psi_k = (\beta_k, c_k, \theta_{3k})$

  We assume that:
  β is sampled from the prior probability for generation interval,
  c is the sampled from the distribution of the date of Neanderthal admixture
  θ_3 is normally distributed $(\mu = 0, \sigma = 3)$.

- Compute the importance function, $\mathcal{L}(z;\ \psi_k)$.

  We note that θ_1, θ_2 are the estimated errors in radiocarbon dates and dates of Neanderthal admixtures respectively that are directly estimated from the data.

2. Compute the mean and variance of (τ) which is one of the three unknown parameters of interest (β, c, θ_3)

$$\mathrm{E}(\tau) = \frac{\sum_k w_k \tau_k}{\sum_k w_k}$$

$$\mathrm{E}(\tau^2) = \frac{\sum_k w_k {\tau_k}^2}{\sum_k w_k}$$

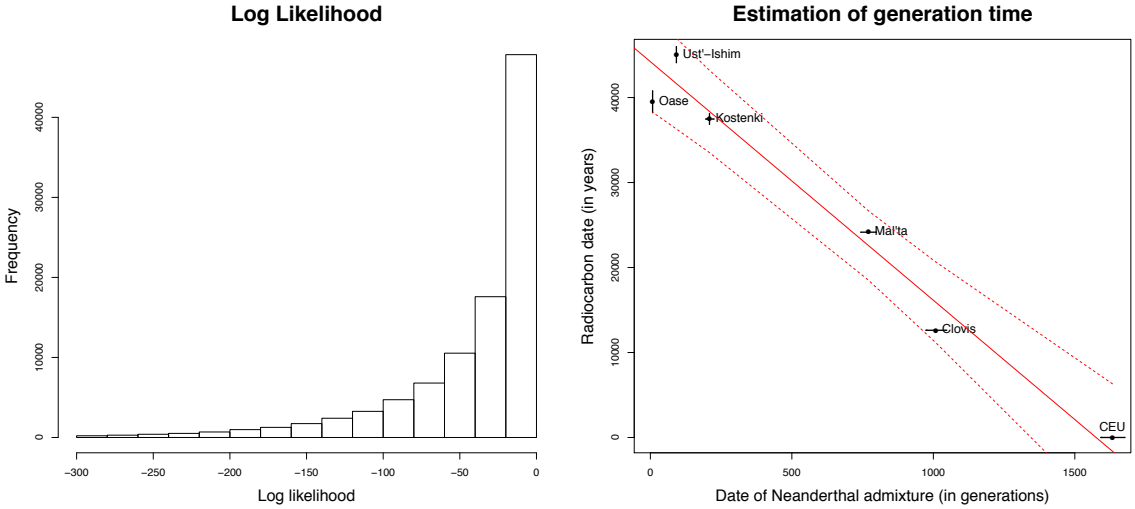
Thus, $\sigma^2(\tau) = \mathrm{E}(\tau^2) - (\mathrm{E}(\tau))^2$

Results:

We ran importance sampling for 100,000 iterations assuming β has a prior probability between 25 to 33 years (23) and c, which reflects the time of the shared Neanderthal admixture event, has a normal distribution $(\mu = 47{,}236, \sigma = 4339)$ (based on dates of Neanderthal admixture observed in CEU). Based on these parameters, we infer that the

value of $\beta = 28.1 \pm 0.7$ years and $c = 44{,}225 \pm 564$ years. (Figure S4.1)

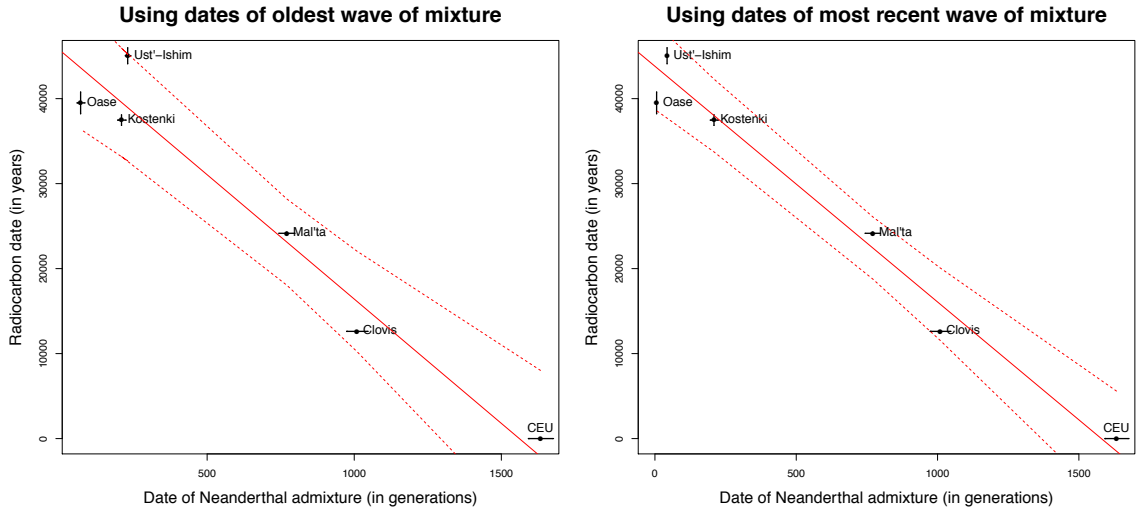
Figure S4.1. Estimate of generation interval. To estimate the generation interval and time of the shared Neanderthal admixture event, we used importance sampling. We assume β has a uniform prior probability between 25 to 33 years (23), c, which reflects the time of the shared Neanderthal admixture event has a normal distribution ($\mu = 47{,}236, \sigma = 4339$) (based on dates of Neanderthal admixture observed in CEU) and θ_3 is normally distributed ($\mu = 0, \sigma = 3$). Panel (a) shows the histogram of the log likelihood of the model for all iterations, (b) Linear relationship between dates of Neanderthal admixture and radiocarbon dates. We show mean ± SE for each sample.



To check the sensitivity of our inference to the priors assumed in the analysis, we repeated the analysis assuming β has a uniform prior probability between 1 to 100 years, c, is normally distributed with ($\mu = 70{,}000, \sigma = 20{,}000$) and θ_3 is normally distributed ($\mu = 0, \sigma = 3$). We inferred that the $\beta = 28.2 \pm 0.7$ years and $c = 44{,}299 \pm 587$ years, consistent with the previous estimate (Figure S4.1).

We have shown that Oase1 and Ust'-Ishim have a history of complex Neanderthal mixture, maybe involving more than one pulse of gene flow. To check how this history affects our estimate of the generation interval, we reran our analysis using the dates of the first or last pulse of gene flow in each of the samples. In both cases, we obtained consistent estimates for the generation interval as shown in Figure S4.2.

Figure S4.2. Robustness to the Oase1 and Ust'-Ishim histories of multiple pulses of Neanderthal mixture. We show the relationship between inferred dates of Neanderthal admixture and radiocarbon dates for the following scenarios:
(a) When we use the dates of the older Neanderthal gene flow in Oase1 and Ust'-Ishim, that the $\beta = 29.3 \pm 0.8$ years and $c = 45{,}678 \pm 615$ years,
(b) When we use the dates of the more recent Neanderthal gene flow in Oase1 and Ust'-Ishim, we estimate that the $\beta = 27.7 \pm 0.7$ years and $c = 43{,}813 \pm 547$ years.



We also used the linear model between the inferred dates of Neanderthal admixture and radiocarbon dates to estimate the age of ancient genomes.

To implement this idea, we estimate the parameters of the model (β, c) using *n-1* samples and then predicted the age of the n^{th} sample, given the date of Neanderthal admixture in the sample. We find relatively close corresponding between the true radiocarbon dates and the age estimated using this approach (Figure S5).

Figure S1: The model underlying our inference of the age of ancient genomes. We assume a simple demographic history relating Neanderthals, non-Africans and Africans. Neanderthal gene flow into non-African ancestors occurred t_n generations ago. This event was shared among all non-Africans and did not affect Africans. The ancient non-African genome was sampled at time t_c. To estimate the age of the ancient genome, we first estimate the dates of Neanderthal gene flow in ancient genomes $(\hat{t}_{na})$ and extant genomes $(\hat{t}_{ne})$. The difference in the inferred dates provides a direct estimate of the age of the ancient sample $(\hat{t}_c)$.

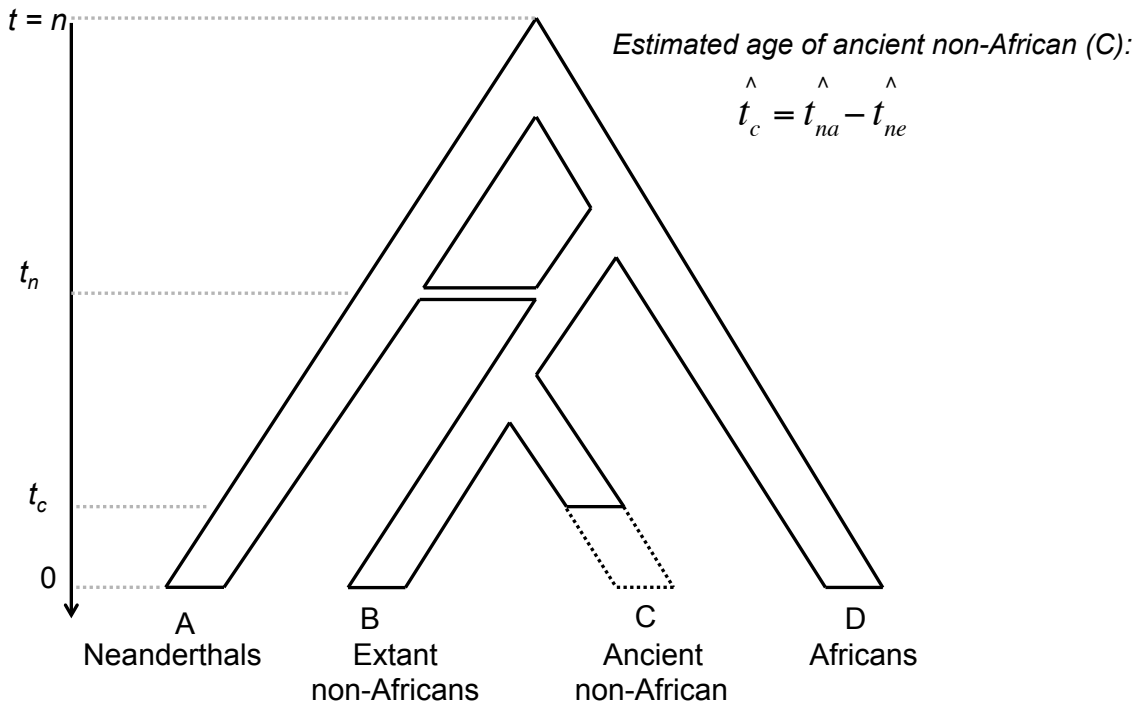


Figure S2: Dates of Neanderthal admixture in Ust'-Ishim based on three approaches of genotype determination. We determined genotypes for SNPs selected based on ascertainment 0 using (a) Pseudo-homozygous majority call, where we chose the major allele in case of heterozygous sites, (b) Pseudo-homozygous random call, where we chose a random allele at each site, and (c) Diploid call, where we considered heterozygous calls. See Note S1 for details. * indicates the results described in the main text.

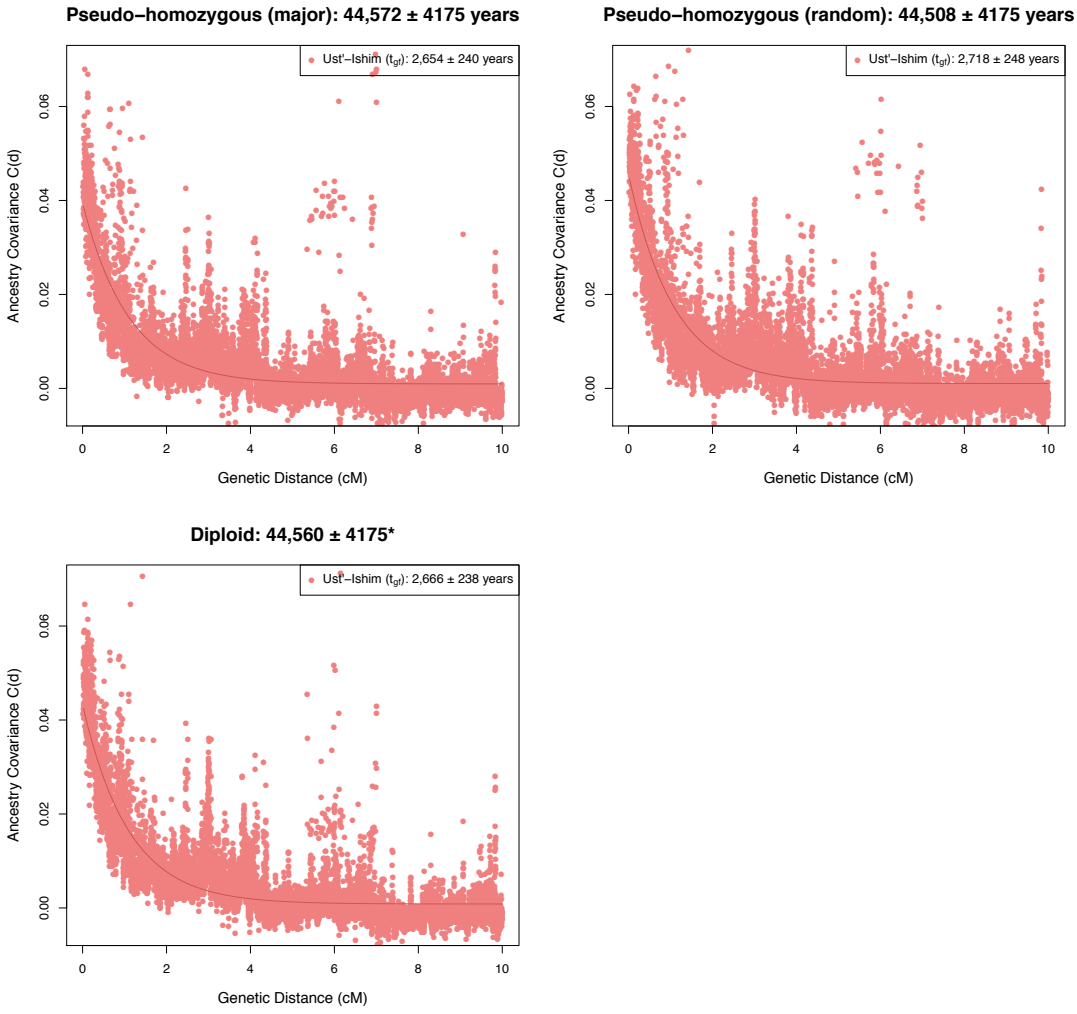


Figure S3: Estimated age of Ust'-Ishim and Oase1 genomes using a model of two Neanderthal admixture events. Estimated dates of Neanderthal gene flow in extant Europeans (1000 Genomes Europeans (CEU)) shown in blue and ancient Eurasians (either Ust'-Ishim or Oase1) shown in pink. Estimated ages of the ancient genome (mean ± SE) shown in the title. For Ust'-Ishim, we show results based on double exponential fit up to the genetic distance of 10cM. For Oase, we show results based on double exponential fit up to the genetic distance of 65cM and bin size of 0.1cM. For Oase, we do not show CEU as the analysis was based on a different bin size and maximum distance.

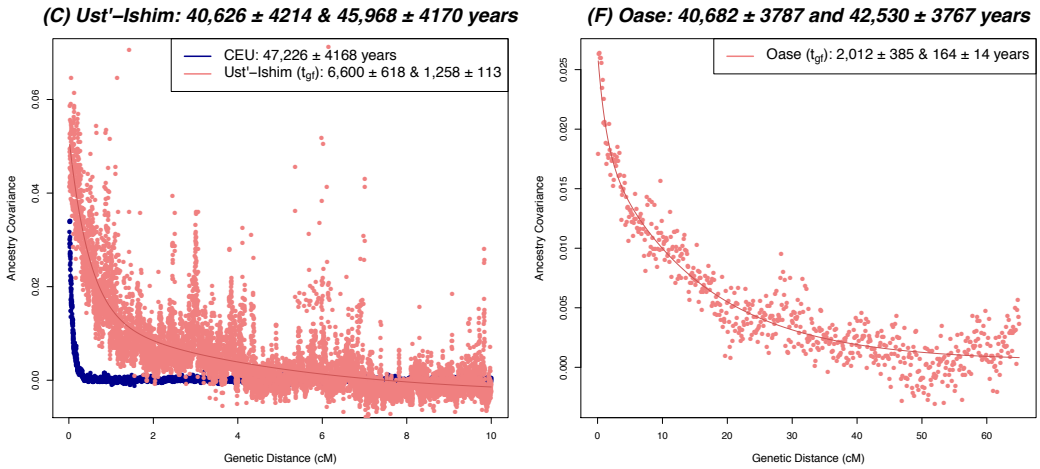


Figure S4: Age estimates for Oase1 for different bin sizes. We estimate the date of Neanderthal admixture in Oase1 using different bin sizes between 0.001 cM (as used for other genomes) - 1cM. In each case, we compared the dates of Neanderthal gene flow in Oase1 with CEU where the Neanderthal gene flow occurred 42,694 ± 3,767 yr BP (for this ascertainment based on a binsize of 0.001cM). The title of each sub-figure shows the bin size used for Oase1 and the estimated age of this specimen. * indicates the estimates used in main text.

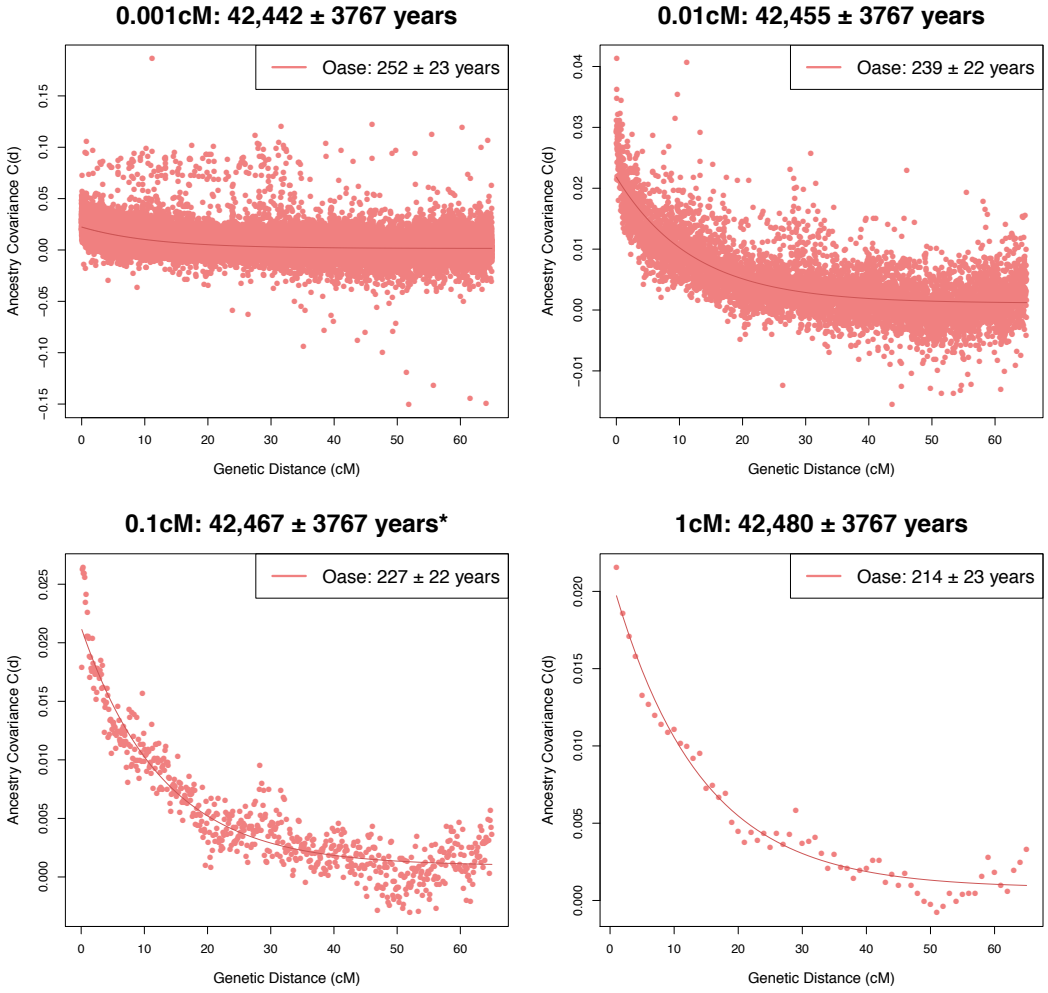


Figure S5: Predicted age based on calibration curve. We estimate the parameters of the model (β, c) using *n-1* samples (shown in black) and then predict the radiocarbon dates of the nth sample (shown in red), given the mean date of Neanderthal admixture in the sample. Standard error (in red) reflects the uncertainty of the model.

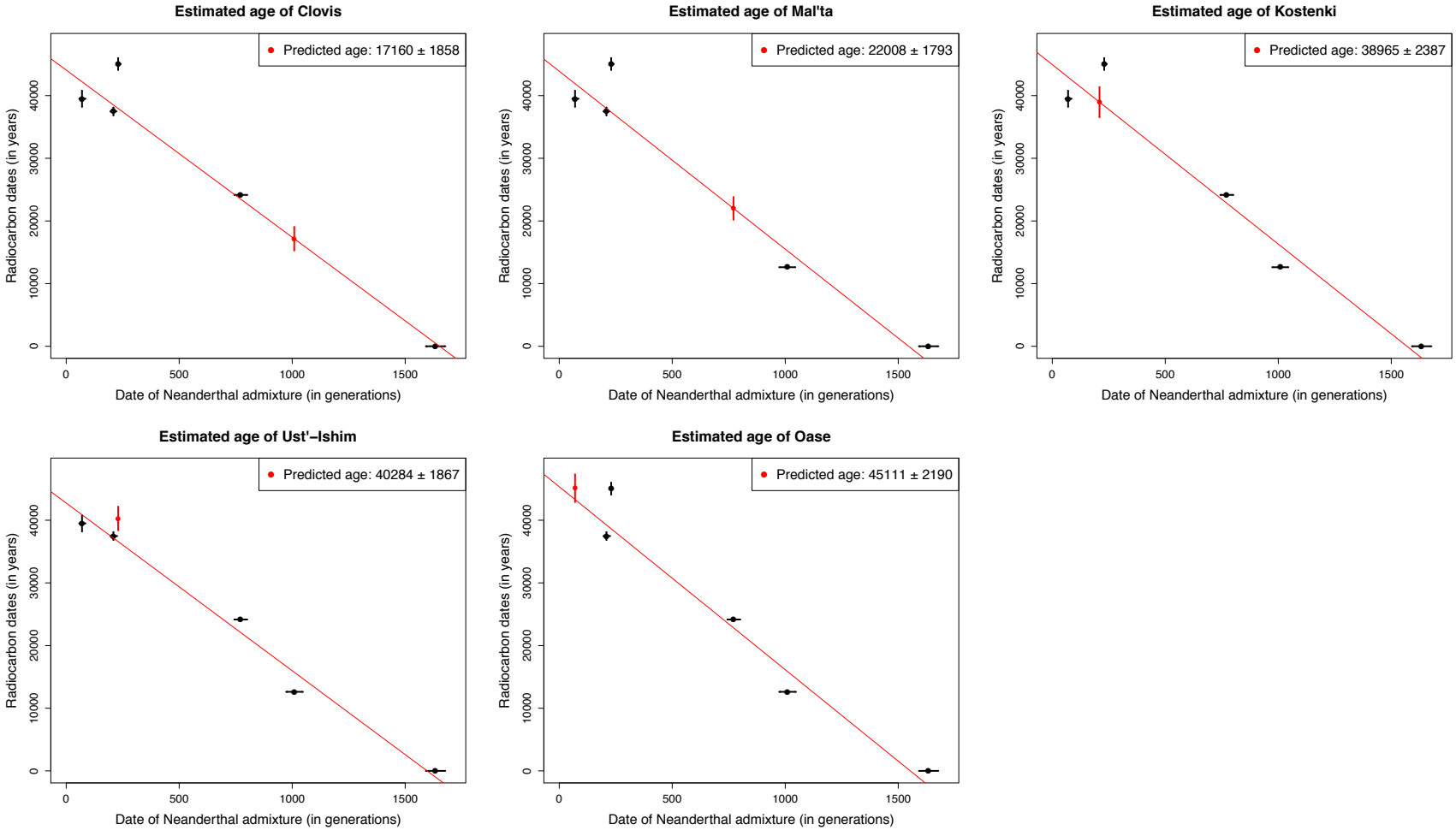


Table S1: Estimated ages of ancient genomes

Sample	*Calibrated radiocarbon dates (calBP)*	*genetic age in generations (mean* ± SE*)*	*genetic age in years (mean* ± SE*)*
Clovis	12,509 – 12,722	624 ± 54	18,066 ± 5112
Mal'ta1	23,891 – 24,423	862 ± 49	24,935 ± 4851
Kostenki14	36,262 – 38,684	1,424 ± 43	41,189 ± 4387
Ust'-Ishim[a]	43,210 – 46,880	1,373 ± 46	39,715 ± 4422
Ust'-Ishim[b]	43,210 – 46,880	1,541 ± 37	44,560 ± 4175
Ust'-Ishim[c]	43,210 – 46,880	1,403 ± 38	40,626 ± 4214
Ust'-Ishim[d]	43,210 – 46,880	1,590 ± 37	45,968 ± 4170
Oase1[e]	37,000 – 42,000	1,468 ± 31	42,467 ± 3767
Oase1[f]	37,000 – 42,000	1,406 ± 33	40,682 ± 3787
Oase1[g]	37,000 – 42,000	1,470 ± 31	42,530 ± 3767

Note: We selected SNPs based on ascertainment 0, used genetic distances based on the *S map*, and a genetic correction based on deCODE data and ae generation interval of 25-33 years (see Note S1).
[a-d] Ust'-Ishim – we tried different models for Ust'-Ishim dating (see results and Figure 2). (A) Dates based on single exponential fit up to genetic distance of 1cM, (B) Dates based on single exponential fit up to genetic distance of 10cM, (C, D) Dates based on double exponential fit ($y = Ae^{-d\lambda_1} + Be^{-d\lambda_2} + c$) to up to genetic distance of 10cM.
[e-g] Oase1 – we tried different models for Oase1 dating (see results and Figure 2). (E) Dates based on single exponential fit up to genetic distance of 65cM, (F, G) Dates based on double exponential fit to up to genetic distance of 65cM.

Table S2: Effect of genetic map on age estimates

Sample	*S map**	*AA map*	*Oxford CEU map*
Date of Neanderthal gene flow in CEU	47,236 ± 4339	42,317 ± 3712	43,376 ± 3820
Clovis	18,066 ± 5112	17,111 ± 3901	19,474 ± 3999
Mal'ta1	24,935 ± 4851	22,594 ± 3808	26,616 ± 3925
Kostenki14	41,189 ± 4387	36,655 ± 3747	38,466 ± 3853
Ust'-Ishim (B)	44,560 ± 4175	39,590 ± 3667	40,204 ± 3807

Note: * indicates the approach used in main text. We used ascertainment 0 described in Note S1. For each map, we separately estimated the genetic map correction based on data from (17). Ust'-Ishim dates are based on the model of single exponential fitted to genetic distance of up to 10cM.

Table S3: Effect of natural selection on age estimates

Sample	*Standard Approach**	*Remove putative targets of selection*
Number of ascertained SNPs included	221,482	213,423
Clovis	18,066 ± 5112	17,731 ± 5091
Mal'ta1	24,935 ± 4851	24,716 ± 4815
Kostenki14	41,189 ± 4387	40,836 ± 4373
Ust'-Ishim (B)	44,560 ± 4175	44,468 ± 4261

Note: * indicates the approach used in main text. SNPs were selected based on ascertainment 0, the genetic distances were based on the *S map*, and the genetic correction was based on data from (17) (see Note S1). Ust'-Ishim dates are based on model of single exponential fitted to genetic distance of up to 10cM. To account for putative targets of selection, we remove ascertained SNPs belonging to exons (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz) or conserved elements in primates (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/phastConsElements46wayPrimates.txt.gz).

Table S4: Effect of B-statistic on dates of Neanderthal gene flow

Date of Neanderthal gene flow	*Standard Approach**	*Low bins of B-stats (b=0-4)*	*High bins of B-stats (b=5-9)*
CEU	47,236 ± 4339	50,608 ± 4512	41,882 ± 3774
Clovis	29,170 ± 2703	29,684 ± 3030	23,314 ± 2336
Mal'ta1	22,301 ± 2169	32,162 ± 3663	22,592 ± 2344
Kostenki14	6,047 ± 649	1,958 ± 672	12,246 ± 1208[+]
Ust'-Ishim (B)	2,666 ± 238	2,727 ± 261	2,027 ± 178

Note: * indicates the approach used in main text. We selected SNPs based on ascertainment 0, used genetic distances based on the *S map*, and made a genetic correction based on data from (17) (see Note S1). Ust'-Ishim dates are based on model of single exponential fitted to genetic distance of up to 10cM.
[+] indicates that this estimate is not very reliable based on manual inspection of the fit. This is likely due to the fact that the sample has very low coverage and we are only using a subset of SNPs in this analysis. We note that for the low bins of B-stats, there were 103,731 SNPs and for the high bins of B-stat, there were 117,702 SNPs.

References:

1. Prüfer K*, et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43-49.
2. The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature* 467(7319):1061-1073.
3. Swapan Mallick HL, Mark Lipson, Iain Mathieson, Melissa Gymrek,*, et al.* (2016) The landscape of human genome diversity. *Nature* (in press).
4. Li H (2014) Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* 30(20):2843-2851.
5. DePristo MA*, et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* 43(5):491-498.
6. Rasmussen M*, et al.* (2014) The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506(7487):225-229.
7. Raghavan M*, et al.* (2013) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*.
8. Seguin-Orlando A*, et al.* (2014) Genomic structure in Europeans dating back at least 36,200 years. *Science* 346(6213):1113-1118.
9. Fu Q*, et al.* (2014) Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514(7523):445-449.
10. Fu Q*, et al.* (2015) An early modern human from Romania with a recent Neanderthal ancestor. *Nature*.
11. Sankararaman S, Patterson N, Li H, Paabo S, & Reich D (2012) The date of interbreeding between Neandertals and modern humans. *PLoS Genet* 8(10):e1002947.
12. Hinch AG*, et al.* (2011) The landscape of recombination in African Americans. *Nature* 476(7359):170-175.
13. Chakraborty R & Weiss KM (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proceedings of the National Academy of Sciences* 85(23):9119-9123.
14. Moorjani P*, et al.* (2013) Genetic evidence for recent population mixture in India. *American journal of human genetics* 93(3):422-438.
15. Myers S, Bottolo L, Freeman C, McVean G, & Donnelly P (2005) A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310(5746):321-324.
16. Kong A*, et al.* (2014) Common and low-frequency variants associated with genome-wide recombination rate. *Nature genetics* 46(1):11-16.
17. Kong A*, et al.* (2010) Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* 467(7319):1099-1103.
18. Lipson M*, et al.* (2015) Calibrating the Human Mutation Rate via Ancestral Recombination Density in Diploid Genomes. *bioRxiv*:015560.
19. Coop G, Wen X, Ober C, Pritchard JK, & Przeworski M (2008) High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. *Science* 319(5868):1395-1398.

20. Fenner J (2005) Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *American Journal of Physical Anthropology* 128(2):415.
21. Helgason A, Hrafnkelsson B, Gulcher JR, Ward R, & Stef√° nsson Kr (2003) A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *The American Journal of Human Genetics* 72(6):1370-1388.
22. Amster G, Sella G (2015) Life history effects on the molecular clock of autosomes and sex chromosomes. Proc Natl Acad Sci USA 113(6):1588–1593.
23. Sun JX*, et al.* (2012) A direct characterization of human mutation based on microsatellites. *Nature genetics* 44(10):1161-1165.
24. Hudson RR (2002) Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* 18(2):337-338.
25. Patterson N, Price AL, & Reich D (2006) Population structure and eigenanalysis. *PLoS genetics* 2(12):e190.
26. Patterson N*, et al.* (2012) Ancient Admixture in Human History. *Genetics* 192(3):1065-1093.
27. Busing FM, Meijer E, & Van Der Leeden R (1999) Delete-m jackknife for unequal m. *Statistics and Computing* 9(1):3-8.
28. Gravel S*, et al.* (2011) Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences* 108(29):11983-11988.
29. Sudmant PH*, et al.* (2015) Global diversity, population stratification, and selection of human copy-number variation. *Science* 349(6253):aab3761.
30. Pickrell JK*, et al.* (2014) Ancient west Eurasian ancestry in southern and eastern Africa. *Proceedings of the National Academy of Sciences* 111(7):2632-2637.
31. Bishop C (2007) Pattern Recognition and Machine Learning (Information Science and Statistics), 1st edn. 2006. corr. 2nd printing edn. (Springer, New York).