**Article (Methods)**

# An extended admixture pulse model reveals the limits to the dating of Human-Neandertal introgression

Iasi, Leonardo N. M. [1,2] and Peter , Benjamin M. [1,3]

[1]Department of Evloutionary Genetics,
Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[2] leonardo_iasi@eva.mpg.de
[3] benjamin_peter@eva.mpg.de

April 4, 2021

# Abstract

Neandertal DNA makes up 2-3% of the genomes of all non-African individuals on average. The length of Neandertal ancestry segments in modern humans has been used to estimate that the mean time of gene flow occurred during the expansion of modern humans into Eurasia, but the precise dates of this gene flow remain largely unknown. Here, we introduce an extended admixture pulse model that allows joint estimation of the timing and duration of gene flow. This model contains two parameters, one for the mean time of gene flow, and one for the duration of gene flow whilst retaining much of the mathematical simplicity of the simple pulse model. In simulations, we find that estimates of the mean time of admixture are largely robust to details in gene flow models. In contrast, the duration of the gene flow is much more difficult to recover, except under ideal circumstances where gene flow is recent or the exact recombination rate is known. We conclude that gene flow from Neandertals into modern humans could have happened over hundreds of generations. Ancient genomes from the time around the admixture event are thus likely required to resolve the question when, where, and for how long humans and Neandertals interacted.


# Introduction

The sequencing of Neandertal (Green et al., 2010; Prüfer et al., 2013, 2017; Mafessoni et al., 2020) and Denisovan genomes (Reich et al., 2010; Meyer et al., 2012) revealed that modern humans outside of Africa interacted, and received genes from these archaic hominins (Vernot and Akey, 2014; Fu et al., 2014; Sankararaman et al., 2014; Fu et al., 2015; Malaspinas et al., 2016; Sankararaman et al., 2016; Vernot et al., 2016). There are two major lines of evidence: First, Neandertals are genome-wide more similar to non-Africans than to Africans (Green et al., 2010). This shift can be explained by 2-4% of admixture from Neandertals into non-Africans (Green et al., 2010; Prüfer et al., 2013). Similarly, East Asians, Southeast Asians and Papuans are more similar to Denisovans than other human groups, which is likely due to gene flow from Denisovans (Meyer et al., 2012).

As a second line of evidence, all non-Africans carry genomic segments that are very similar to the sequenced archaic genomes. As these putative *admixture segments* are up to several hundred kilobases (kb) in length, it is unlikely that they were inherited from a common ancestor that predates the split of modern and archaic humans (Sankararaman et al., 2014; Vernot and Akey, 2014). Rather, they entered the modern human populations later through gene flow (Sankararaman et al., 2012, 2014; Vernot and Akey, 2014; Sankararaman et al., 2016; Vernot et al., 2016).

However, uncertainty remains about when, where, and over which period of time this gene flow happened. A better understanding of the location and timing of the gene flow would allow us to place constraints on the timing of movements of early modern humans. More certainty in the timing of gene flow also would affect assumptions about introgressed allele frequencies and their distribution in present-day human genomes and conclusions drawn about the phenotypic effects and selective pressures of introduced alleles.

Archaelogical evidence puts some boundaries on the times when Neandertals and modern humans might have interacted. The earliest modern human remains outside of Africa are dated to around 188 thousand years ago (kya) (Hershkovitz et al., 2018; Stringer and Galway-Witham, 2018) and the latest Neandertals are suggested to be between 37 kya and 39 kya old (Higham et al., 2014; Zilhão et al., 2017). Thus the time window where Neandertals and modern humans might have been in the same area stretches more than 140,000 years. However, there is less direct evidence of modern humans and Neandertals in the same geographical location at the same time. In Europe, for example, Neandertals and modern humans likely overlapped only for less than 10,000 years (Bard et al., 2020).

## Genetic dating of gene flow

The most common approach to learn about admixture dates from genetic data uses a *recombination clock* model: Conceptually, admixture segments are the result of the introduced chromosomes being broken down by recombination; the offspring of an archaic and a modern human parent will have one whole chromosome each of either ancestry. Thus, the offsprings' markers are in full ancestry linkage disequilibrium (ALD); all archaic variants are present on one DNA molecule, and all modern human one on the other one.

If this individual has offspring in a largely modern human population, in each generation meiotic recombination will reshuffle the chromosomes, progressively breaking down the ancestral chromosome down into shorter segments of archaic ancestry (Falush et al., 2003; Gravel, 2012; Liang and Nielsen, 2014), and ALD similarly decreases with each generation after gene flow (Chakraborty and Weiss, 1988; Stephens et al., 1994; Wall, 2000).

This inverse relationship between admixture time and either segment length or ALD is commonly used to infer the timing of gene flow (Pool and Nielsen, 2009; Moorjani et al., 2011; Pugach et al., 2011; Gravel, 2012; Sankararaman et al., 2012; Loh et al., 2013; Hellenthal et al., 2014; Liang and Nielsen, 2014; Sankararaman et al., 2016; Pugach et al., 2018; Jacobs et al., 2019). Most commonly, it is assumed that gene flow occurs over a very short duration, referred to as an *admixture pulse*, which is typically modelled as a single generation of gene flow (e.g Moorjani et al., 2011). This has the advantage that both the length distribution of admixture segment and the decay of ALD with distance will follow an exponential distribution, whose parameter is directly informative about the time of gene flow (Pool and Nielsen, 2009; Liang and Nielsen, 2014; Gravel, 2012).

In segment-based approaches, dating starts by identifying all admixture segments, which can be done using a variety of methods (Seguin-Orlando et al., 2014; Sankararaman et al., 2016; Vernot et al., 2016; Racimo et al., 2017; Skov et al., 2018). The length distribution of segments is then used as a summary for dating when gene flow happened.

Alternatively, ALD-based methods use linkage disequilibrium (LD) patterns, without explicitly inferring the genomic location of segments (Chimusa et al., 2018) (Figure 1 B). Instead, estimation

of admixture dates proceeds by fitting a decay curve of pairwise LD as a function of genetic distance, implicitly summing over all compatible segment lengths (Moorjani et al., 2011; Loh et al., 2013).

## Archaic gene flow estimates

Using this approach, Sankararaman et al. (2012) dated the Neandertal-human admixture pulse to be between 37–86 kya. Later, this date was refined to 41 – 54 kya ($C.I._{95\%}$) using an updated method, using a different marker ascertainment scheme combined with a refined genetic map for European populations (Moorjani et al., 2016). A date of 50 – 60 kya was obtained from the analysis of the genome of *Ust'-Ishim* a 45,000-year-old modern human from western Siberia.

The segments of putative Neanderthal origin in the *Ust'-Ishim* individual are substantially longer than those in present-day humans, which makes their detection easier, and adds further evidence that gene flow between Neandertals and modern humans has happened relatively recently before Ust'-Ishim lived (Fu et al., 2014).

## Limitations of the pulse model

The admixture pulse model assumes that gene flow occurs over a short time period; however it is currently unclear how short a time could still be consistent with the data. This makes admixture time estimates hard to interpret, as more complex admixture scenarios might be masked, and so gene flow could have happened tens of thousands of years before or after the estimated admixture time.

That admixture histories are often complicated has been shown in the context of Denisovan introgression into modern humans, where at least two distinct admixture events into East Asians and Papuans were proposed (Browning et al., 2018; Jacobs et al., 2019). While the length distributions of admixture segments are similar between the populations, there are significant differences in the geographic distribution of admixture fragments, and the similarity to the sequenced high-coverage genome (Browning et al., 2018; Massilani et al., 2020).

In contrast, all Neandertal admixture segments are most similar to the Vindija genome (Prüfer et al., 2017). However, differences in admixture proportions (Meyer et al., 2012; Wall et al., 2013; Kim and Lohmueller, 2015; Vernot and Akey, 2015; Villanea and Schraiber, 2019) and direct evidence from early modern humans with very recent Neandertal ancestry from Oase hint at more complex admixture histories (Fu et al., 2015).

One way to refine admixture time estimates is to include two or more distinct admixture pulses. The distribution of admixture segment lengths will then be a mixture of the fragments introduced from each event. This is especially useful if the events are very distinct in time, e.g. if one event is only a few generations back, and the other pulse occurred hundreds of generations ago (Fu et al., 2014, 2015). In this case, the admixture segments will be either very long if they are recent, or much shorter if they are older.

Zhou et al. (2017a) extended this model to continuous mixtures, using a polynomial function as a mixture density. However, they found that even for relatively short admixture events, the large number of parameters led to an underestimate of admixture duration (Zhou et al., 2017b).

## Extended Pulse Model

One drawback of these approaches is that they introduce a large number of parameters. Even a discrete mixture of two pulses requires at least three parameters (two pulse times and the relative magnitude of the two events) (Pickrell et al., 2014), and the more complex models require regularization schemes for fitting (Zhou et al., 2017b; Ralph and Coop, 2013).

Here, we propose an *extended admixture pulse* model (Figure 1 A) to estimate the duration of an admixture event. It only adds one additional parameter, reflecting the duration of gene flow, while retaining much of the mathematical simplicity present in the simple pulse model. The extended pulse model assumes that the migration rate over time is Gamma distributed, so that the length distribution of admixture segment has a closed form (Figure 1 C & D) with two parameters, the mean admixture time and duration.

Conceptually, identifying an extended pulse requires us to establish that the length distribution of introgressed segments deviates from an exponential distribution. However, other sources of biases, such as the demography of the admixed population, the accuracy of the recombination map or details in the inference method parameters may also introduce similar signals. Thus, we have to carefully evaluate other potential sources of biases on whether they might lead to confounding signals. (Sankararaman et al., 2012; Fu et al., 2014; Moorjani et al., 2016).

Here, we first define the extended admixture pulse model and derive the resulting segment length and ALD distributions. We then introduce inference schemes for either data summary. Next, we use simulations under the extended pulse model to assess the effect of ongoing admixture on inference based on the simple pulse model, and investigate under which scenarios we can distinguish between the two models. We show that the simple and extended pulse models can be distinguished for very recent events. However, for the parameter values relevant for the case of archaic admixture, a simple pulses cannot be distinguished from continuous admixture over an extended time, or multiple admixture pulses that are close together in time. Using European genomes from the 1000 Genomes dataset (The 1000 Genomes Project Consortium, 2015) to estimate the timing of Neandertal introgression, we found that ALD inferred admixture times are consistent with a multitude of duration times, up to several tens of thousand of years.

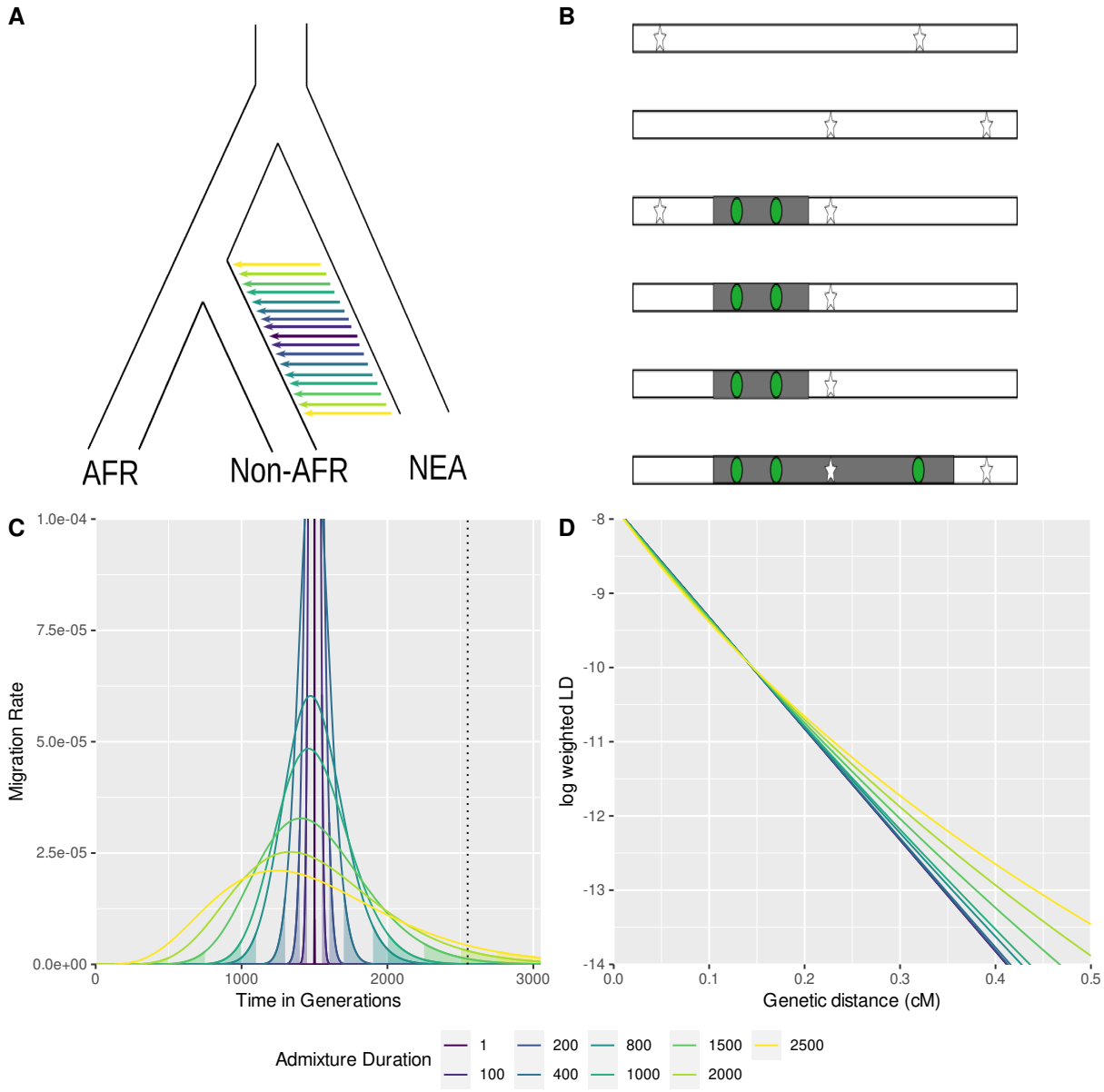


Figure 1: A) Neandertal introgression into non-Africans with a multitude of potential admixture durations. B) The time and duration of admixture results in different length distributions of introgressed chromosomal segments (grey) containing Neandertal variants (green circles) in high LD to each other compared to the background (human variants white stars). The ALD approach estimates linkage between the introgressed variants (green circles), whereas the haplotype approach tries to estimate the segment directly (grey area). C) Migration rate per generation modeled using the extended pulse model for different admixture durations (colored lines). The filled area under the curve indicates the boundaries of the discrete realization of the duration of gene flow $t_d$. The dotted line indicates the oldest possible time of gene flow (as defined in the simulations). D) The expected LD decay under the extended pulse model.

# New Approaches

In this section, we present the mathematical description of the admixture models we use in this paper, and introduce inference algorithms for estimating the admixture time and duration from both ALD and segment data.

## Admixture Models and Inference

We think of admixture as a series of "foreign" chromosomes introduced in a population. In the case of the simple pulse model, it is assumed that all admixture happens in the same generation, (*i.e.* all chromosomes are introduced to the population at the same time). To extend this model, we allow chromosomes to enter at potentially many different time points.

Formally, we assume we have $n$ admixture segments. We denote the length of the $i$-th segment as $L_i$. Further, the random variable $T_i$ represents the time when segment $i$ entered the population. We assume that the $L_i$ and $T_i$ are both realizations from more general distributions $L$ and $T$ that reflect the overall segment length and admixture time distributions, respectively.

Given $T_i$, we assume that $L_i$ is exponentially distributed with rate parameter given by the admixture time $t$ and the recombination rate $r$.

$$P(L_i = l | T_i = t) = tr \exp(-trl) \tag{1}$$

For simplicity, we measure the length of each segment $L_i$ in the recombination distance Morgan (M), so that $r = 1$ Morgan per generation, and is omitted from this section.

To model the unconditional distribution of admixture lengths, we need to integrate over all possible $T_i$, which are drawn from the admixture time distribution $T$:

$$P(L_i = l) = \int_0^\infty P(T_i = t) P(L_i = l | T_i = t) \, dt, \tag{2}$$

Thus, we can think of $L$ as an exponential mixture distribution with mixture density $T$ (Ralph and Coop, 2013; Zhou et al., 2017a).

## The Simple Pulse Model

In the simple pulse model, we assume that all fragments enter the population at the same time $t_m$. Therefore all $T_i$ have the same value $t_m$, and $T$ thus is a constant distribution. We can formalize this by using a Dirac delta function:

$$P(T_i) = \delta_{t_m}(T_i), \tag{3}$$

which integrates to one if the integration interval includes $t_m$, and zero otherwise.

Therefore, we obtain the exponential distribution of admixture fragments under this model (e.g. Pool and Nielsen, 2009):

$$P(L_i = l) = t_m e^{-t_m l} \tag{4}$$

The expected segment length under a simple pulse model is given by,

$$\mathbb{E}[L] = \frac{1}{t_m} \tag{5}$$

and the variance is

$$\mathrm{Var}[L] = \frac{1}{t_m^2}. \tag{6}$$

Note that this definition of the segment length distribution is independent from the migration rate $m$ at $t_m$, as we are only concerned about the length distribution of admixture segments, but not the total amount of Neandertal ancestry. However, as we do not consider the probability of admixture segments recombining with each other, we implicitly assume that $m$ is small (Gravel, 2012; Liang and Nielsen, 2014), which is reasonable for the Neandertal-modern human introgression, where $m \approx 0.02 - 0.03$ . Further we assume that there is no selection (Shchur et al., 2019) and the recombination rate is the same in different populations at all times (Gravel, 2012).

### The Extended Pulse Model

For the extended pulse model, we assume that the $T_i$ follow a Gamma distribution. It is convenient to parameterize the Gamma distribution by its mean $t_m$ and its shape parameter $k$. $\Gamma(k, \frac{t_m}{k})$, hence the density is

$$P(T_i = t) = \frac{1}{\Gamma(k)(\frac{t_m}{k})^k} t^{k-1} e^{-t\frac{k}{t_m}} \tag{7}$$

with, $t \geq 0$ and $k \geq 1$.

The expectation and variance for the admixture times are then

$$\begin{aligned} \mathbb{E}[T] &= t_m \\ Var[T] &= \frac{t_m^2}{k} = \left(\frac{t_d}{4}\right)^2. \end{aligned} \tag{8}$$

Here, we define the admixture duration $t_d = 4t_m k^{-\frac{1}{2}}$, which is a convenient measure for the duration of gene flow. If $k$ is low, then $t_d$ will be large and gene flow extends over many generations. In contrast, if $k$ is large, then $t_d \approx 0$ and we recover the simple pulse model (Figure 1 C & D).

The distribution of segment length is calculated by plugging equation 7 into equation 2 and integrating:

$$P(L = l|k, t_m) = t_m^{-k} \left( \frac{k}{l + \frac{k}{t_m}} \right)^{k+1}. \tag{9}$$

The distribution in equation 9 is known as a *Lomax* or *Pareto-II* distribution, which is a heavier-tailed relative of the Exponential distribution.

Under the extended pulse model, the expected segment length will be larger than under the simple pulse model (5):

$$\mathbb{E}[L] = \frac{k}{(k-1)} \frac{1}{t_m} \tag{10}$$

The fraction $\frac{k}{k-1}$ will be larger for low $k$, which fits previous results that the longest (*i.e.* most recently introgressed) admixture segments have a disproportionate impact on inference, and thus admixture time estimates from ongoing gene flow are biased towards more recent events (Moorjani et al., 2011, 2016).

The variance for $k > 2$ is

$$Var[L] = \frac{k^3}{(k-1)^2(k-2)} \frac{1}{t_m^2}. \tag{11}$$

Again, the term involving $k$'s is strictly larger than one, which matches our intuition that ongoing gene flow should result in a larger variance in admixture segments length. As $k$ approaches infinity, this term approaches one and we recover the moments of the exponential distribution, since $k$ is inversely proportional to the admixture duration.

## Admixture time estimates

For inference, either the admixture segment lengths or ALD can be used. In cases where the admixture segment length is known, equation 9 is the likelihood function and can be used for inference. For inference using ALD, we follow Moorjani et al. (2011) and use the decay of ALD with genetic distance as a statistic, and fit the tail function $P(L_i > d)$, which can be calculated by integrating equations 4 and 9.

The tail function under the simple pulse model is

$$P(L_i > d) = e^{-t_m\, d} \tag{12}$$

and under the extended pulse model we obtain

$$P(L_i > d) = \left( \frac{1}{1 + \frac{t_m}{k}\, d} \right)^k . \tag{13}$$

For fitting the data, following Moorjani et al. (2016), we add an intercept $A$ and a constant modelling background LD $c$, to the tail functions. The genetic distance $d$ is measured in centiMorgan (cM).

$$ALD \sim \; A\, e^{-t_m\, d} + c \tag{14}$$

$$ALD \sim \; A \left( \frac{1}{1 + \frac{t_m}{k}\, d} \right)^k + c \tag{15}$$

We fitted the distribution to the data with a non-linear least-square optimization algorithm using the `nls` function implemented in the R `stats v.3.6.2` package (R Core Team, 2019).

# Results

We test our ability to estimate the admixture time and duration from coalescent simulations using `msprime` (Kelleher et al., 2016). First, we compare the inference of simulations under simple and extended pulses of gene flow, while assuming only one generation of gene flow for both scenarios. We contrast this effect with the effect of other model assumptions and analysis parameters. Taking in consideration factors with the most substantial effects, we establish the conditions under which we can use the extended pulse model for gene flow duration inference. Finally, we apply our model to the 1000 Genomes Project data (The 1000 Genomes Project Consortium, 2015).

## Model comparisons

We first present the result of inference assuming a simple pulse model on simulations under the extended pulse model (Figure 2). Therefore we simulate 3 % Neandertal admixture into non-Africans using a simple demographic model (Supplement Figure 1A). To enrich for Neandertal informative sites we use the lower-enrichment ascertainment scheme (LES) which only considers sites that are fixed for the ancestral state in Africans and polymorphic or fixed derived in Neandertals. Pairs of SNPs with a minimal distance smaller 0.05 cM are excluded. In Figure 2A, we plot results for simulations under both the simple and extended pulse model, for $t_m$ between 250 and 2000

generations, and $t_d = \frac{t_m}{2}$ (i.e. if $t_m = 500$ generations, then $t_d = 250$ and most gene flow occurs between 375 and 625 generations ago). Even for these long periods of gene flow, the mean gene flow time is generally well-estimated, with a deviation of 3% to 6% for the simple pulse and 0% to 5% for the extended pulse.

To further investigate the effect of the duration of gene flow on mean admixture time estimates, we compare scenarios where $t_m$ is fixed at 1500 and $t_d$ varies between one (simple pulse model) and 1,500 generations (Figure 2B). Here, we find that the simple pulse model leads to a slight overestimate of $t_m$, and that longer pulses result in more recent admixture time estimates. However, even for very-long gene flow of $t_d = 1,500$ the relative deviation is less than 10%. Thus, we find that if we are interested in $t_m$, this parameter is generally well-estimated even under scenarios of longstanding gene flow.

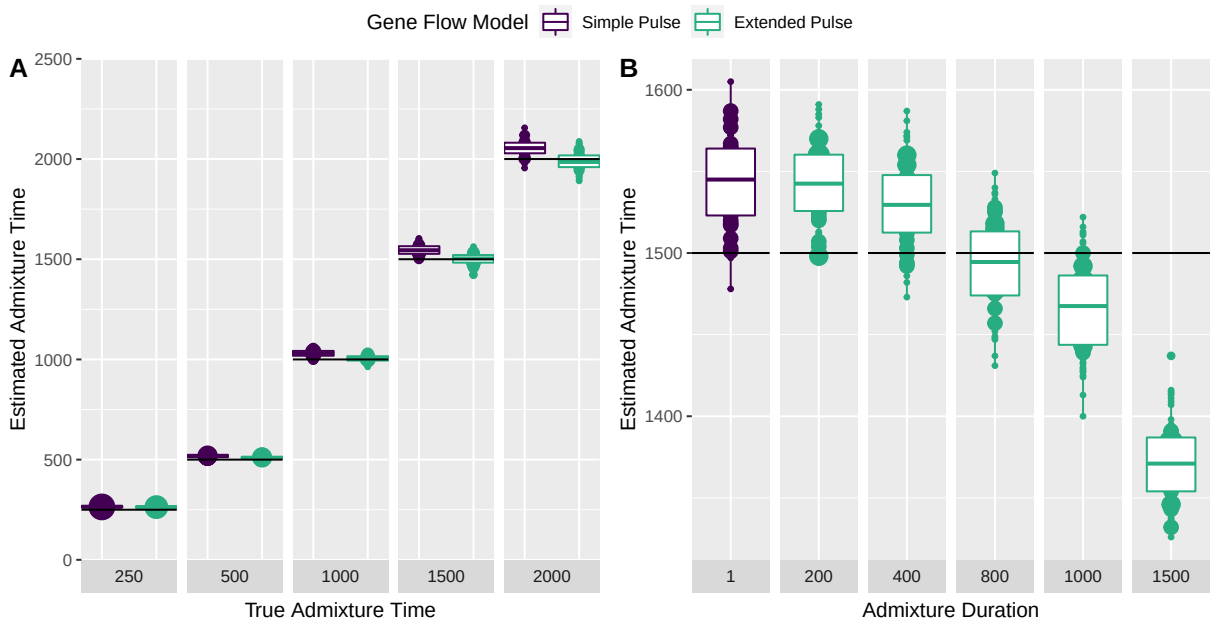


Figure 2: A) Comparison of mean admixture time estimates between simple and extended pulse gene flow for different admixture times in generations. The duration of continuous gene flow $t_d$ corresponds to 50% of the mean admixture time $t_m$, black line indicates true mean admixture time. B) Comparison of mean admixture time estimates for simulations with a mean time of admixture of 1500 generations ago, at varying durations of gene flow. Boxplot created from 100 simulation replicates.

## Comparing effect sizes for technical covariates

As the effect of ongoing gene flow is relatively minor, we next evaluate its relative importance for inference compared to other common assumptions made in the inference of admixture times.

We contrast the effect of extended gene flow on admixture time inference with the effects of a

complex demographic history (Supplement Figure 1B) and a variable recombination map (*i.e.* using an empirical map for simulations but assuming a constant rate for the analysis), as well as the impact of the ascertainment scheme used to amplify Neandertal introgressed variants in the analysis, and the minimum genetic distance between variants ,$d_0$, in the ALD estimation, using a simple and complex setting for each of these parameters.

In Figure 3, we present the effect sizes of these four predictors on the admixture time inference. These effect sizes are estimated using a GLM on simulations under all possible parameter combinations (S. 2, Supplement Table 1).

As a baseline, for comparison, we define a standard model as one using the lower-enrichment ascertainment scheme (LES), $d_0 = 0.05$ cM, simple demography and constant recombination rate over the chromosome.

As shown in the previous section, under the standard model admixture times are well-estimated, with a mean standardized difference of -0.03 (-0.07 - 0.02 $C.I._{95\%}$) from the true admixture time.

We find that the inclusion of a variable recombination map leads to a large underestimate of admixture times -1.47 -1.53 – -1.42 $C.I._{95\%}$. In contrast, all other covariates had relatively moderate effect sizes. The extended pulse (-0.24, -0.29 – -0.19) is in the range of biases arising from the ascertainment scheme (-0.33, -0.38 – -0.28), a more complex demography (-0.25, -0.30 – -0.20) and only slightly more then caused by a different $d_0$ (-0.20, -0.16 – -0.26) (S. Figure 2).

Overall the bias introduced by the ascertainment, minimal distance cutoff, demography and admixture model are small: around +/-0.25 standard deviations or less. The major uncertainties in the admixture time estimate arise from assuming a constant recombination rate. The admixture time estimates for a simple or extended admixture pulse are similarly affected by the other modeling parameters.

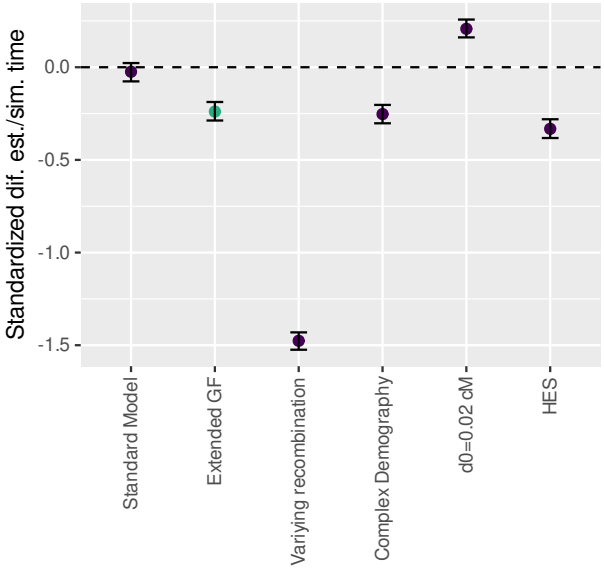


Figure 3: GLM effect size estimates and 95% C.I. for the parameters: gene flow model (simple/extended), recombination rate (constant/varying), demography (simple/complex), minimal genetic distance (0.02/0.05 cM) and ascertainment scheme (LES/HES), on the standardized difference between simulated and estimated admixture time. Estimates are calculated across all possible combinations of parameters and are given as the estimate of the standard model plus the respective parameter estimate. Dotted horizontal line indicates unbiased admixture estimates.

## Parameter estimation under the extended pulse model

In Figure 4, we show the accuracy of $t_d$ and $t_m$ estimates in two sets of simulations. In panels A and B we increase both $t_m$ and $t_d$ at a similar rate, such that gene flow ends 50 generations before sampling ($t_{end} = 50$, where $t_{end} = t_m - \frac{t_d}{2}$). In panels C and D $t_d$ is kept fixed at 800, but we increase $t_m$. We further vary recombination rate settings as i) inference and simulation under constant recombination rate, ii) simulation using the HapMap genetic map (HapMap Consortium, 2007), no correction, iii) simulation using HapMap, correction using African-American map (AAMap) (Hinch et al., 2011), iv) inference and simulation using HapMap.

Figure 4A and C show the mean time estimates received from the simple and extended pulse model fit. Figure 4B and D show the corresponding admixture duration estimate. For simulations under a constant recombination the mean time can be estimated confidently for different durations and different sampling times after the end of the admixture event in both cases.

Under constant recombination settings, we find that parameters under the extended pulse are well-estimated in all scenarios, although we observe a small overestimate of both $t_m$ and $t_d$, for very long admixture durations over 1000 generations in the very distant past (more than 600 generations ago). In contrast, inference under the simple pulse model results in a slight underestimate (Figure 2, 4A).

We also find that the simple pulse model is fairly robust to changes in recombination settings, both using the "correct" and an empirically similar recombination map. Only when we do not correct for recombination rate variation, we observe a substantial underestimate by almost a factor of two for more distant admixture (Figure 4A & C, left panels).

In contrast, estimates under the extended pulse model are substantially more spread out under variable recombination maps, but we do not observe the systematic bias towards more recent admixture times as in the simple-pulse model. In contrast, the uncorrected scenario results in extremely poor estimates with very high variation, and is clearly unsuitable for this purpose (Figure 4A & C, right panels)

We also find that estimates of $t_d$ differ substantially between recombination rate settings. Under constant rates, $t_d$ is well-estimated in all scenarios, and if the recombination rate is known, we get reasonably accurate estimates, at least if admixture is not older than 100 generations. However, under the scenarios where recombination rates differ between simulation and inference, results become more erratic; For the cases of $t_d = 200$ and $t_d = 400$, many simulations infer an admixture time of near zero; in contrast, very old admixture results in a large overestimate.

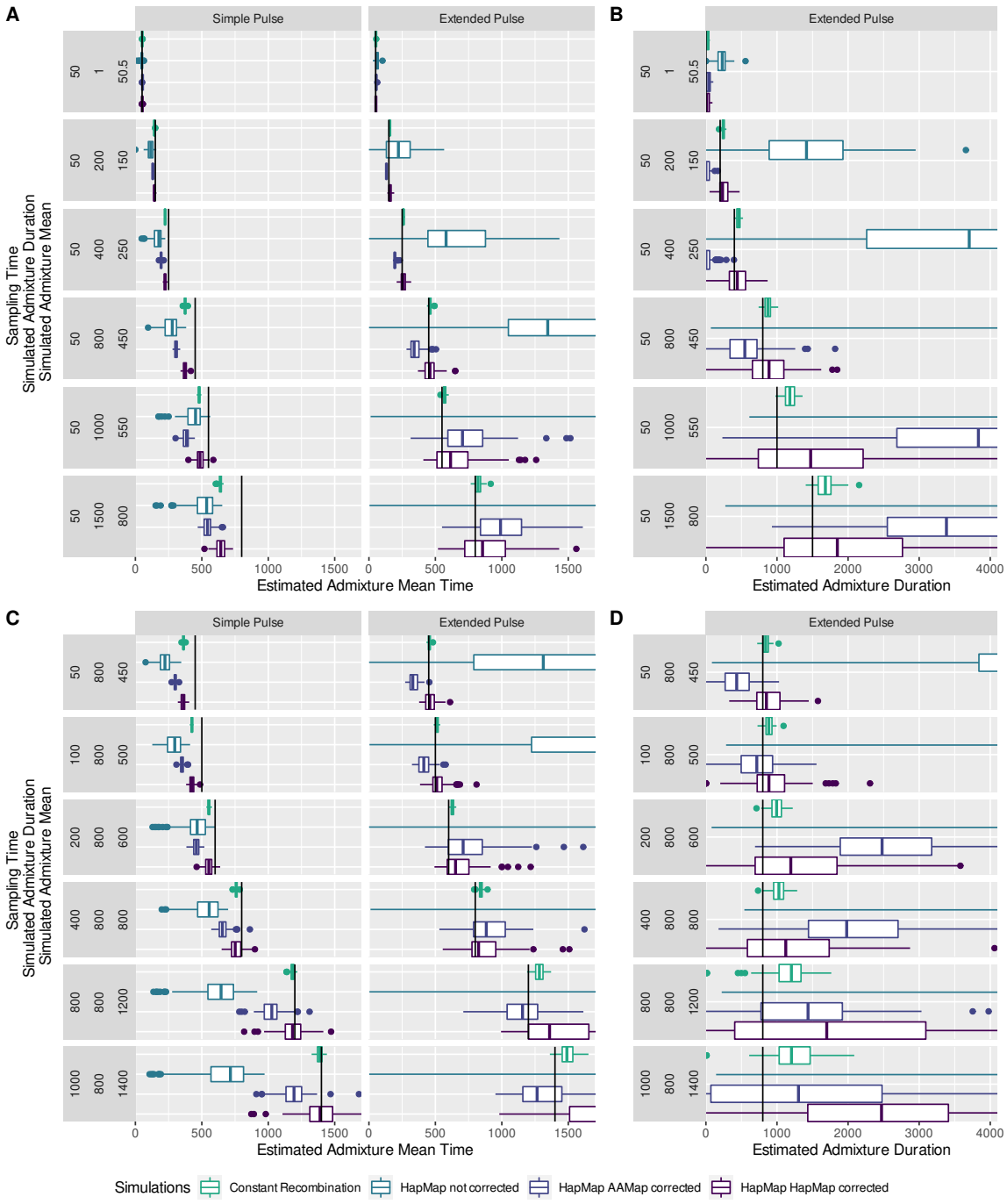


Figure 4: Comparison of parameter inference under the simple and extended pulse model A) Mean time estimates $t_m$ for different gene flow durations $t_d$ all sampled 50 generations after the gene flow ended. B) Duration estimate $t_d$ of the same scenario C) Mean time estimates for different sampling times following 800 generations of gene flow. D) Duration estimate of the same scenario. All scenarios are simulated either under a constant recombination rate or an empirical recombination map (HapMap). Genetic distances for simulations under an empirical map are assigned by: assuming a constant rate (HapMap not corrected), using a different map (HapMap AAMap corrected) and using the same map (HapMap HapMap corrected). All times are given in generations.

## Application to Neandertal data

In the previous section, we have shown using simulations that it might be difficult to distinguish admixture scenarios of various durations particularly if these happened a long time ago (i.e. more than 500 generations ago). To evaluate whether this is also true for real data, we estimated the Neandertal admixture pulse from the 1000 genomes data, by fitting pulses with durations ranging from one generation up to 2,500 generations to the ALD decay curve (Figure 5, S. Table 2). Plotting these best-fit ALD curves (Figure 5A) on a y-axis in natural scale shows the extremely slight difference predicted under these drastically different gene flow scenarios. The difference between scenarios becomes more apparent if we log-transform the y-axis (Figure 5B), where we see that ongoing gene flow results in a heavier tail in the ALD distributions. However, these LD values are very close to zero, and are thus only very noisily estimated. Therefore, we find that all scenarios are compatible with the observed data, and that there is little power to differentiate different cases.

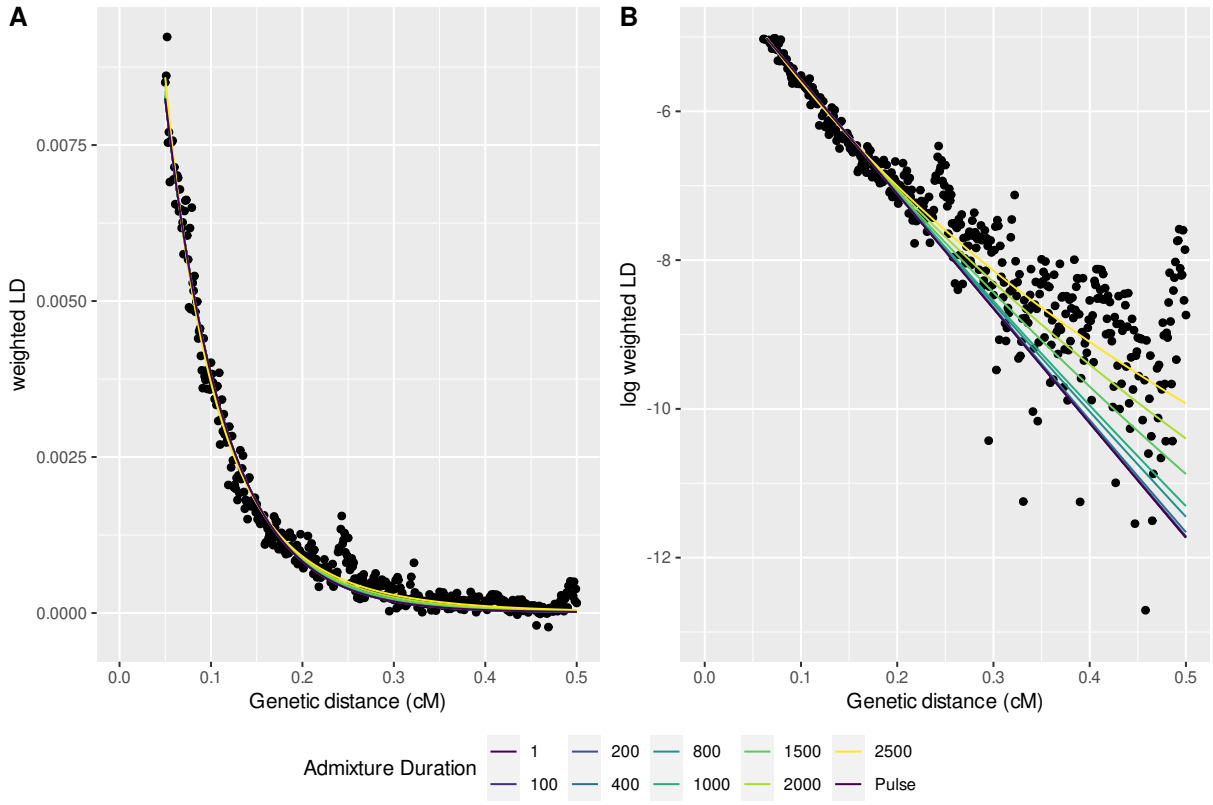


Figure 5: Different admixture duration models ranging from a one generation pulse to 2500 generations of gene flow for Neandertal interbreeding using all 1k Genome CEU individuals as the admixed population with all YRI and 3 high coverage Neandertals as reference populations. A) Weighted LD normal scaled B) Weighted LD log scaled.

# Discussion

Previous estimates to date Neandertal-human gene flow have focused almost entirely on the mean time of gene flow, for which reasonably tight credible intervals can be estimated (Sankararaman et al., 2012; Moorjani et al., 2016). Here, we show that this simple pulse model cannot be used to establish bounds on when gene flow happened, as scenarios involving hundreds of generations of gene flow are virtually indistinguishable from a short interval of gene flow. Thus, it is important to realize that previous estimates of the time of Neandertal gene flow are credible ranges for the mean time of gene flow. This means that while substantial amounts of gene flow happened around these mean times, they do not bound when gene flow happened. This is of great practical importance, as it might be tempting to link the genetic admixture date estimates with biogeographical events (Sankararaman et al., 2012; Lazaridis et al., 2016; Douka et al., 2019; Jacobs et al., 2019; Vyas and Mulligan, 2019). Indeed, our results show that substantial gene flow could also have happened tens of thousands of years before or after the mean time of gene flow.

The discovery of early modern human genomes dated to 40,000 - 45,000 ya with very recent Neandertal ancestors less than ten generations ago (Fu et al., 2014) illustrates that gene flow likely happened over at least several thousand years. In general, inference based on ancient genomes (Fu et al., 2014, 2015; Moorjani et al., 2016) promise to resolve some of these dating issues, as the time difference between gene flow and sampling time is much lower. However, using these genomes for dating leads to further hurdles, particularly pertaining to the spatial distribution of admixture events; whereas we perhaps assume that the spatial structure present in initial upper paleolithic modern humans might be largely homogenized in present-day people, the introgression signal observed in Oase could be partially private to this population, and thus these samples may have a different admixture time distribution than present-day people.

The uncertainty over the duration of Neandertal gene flow might also have some implications for selection on introgressed Neandertal haplotypes. Neandertal alleles have been suggested to be deleterious in modern human populations due to an increased mutation load (Harris and Nielsen, 2016; Juric et al., 2016). Some details of these models may be slightly affected if migration occurred over a longer time. For example, Harris and Nielsen (2016) suggested that an initial pulse of gene flow of up to 10% Neandertal ancestry might be necessary to explain current amounts of Neandertal ancestry, with very high variance in the first few generations after gene flow. More gradual introgression could mean that such high admixture proportions were never achieved, but rather a continuous migration-selection balance process persisted for the contact period, where deleterious Neandertal alleles continually entered the modern human populations, but were selected against immediately. However, in terms of the overall frequencies achieved, there is likely little difference. For example, Juric et al. (2016) showed using a two-locus model that the frequencies of Neandertal haplotypes alone cannot be used to distinguish different admixture histories.

In addition, we find that modelling and method assumptions have an impact on admixture time

estimates that are of a similar or larger magnitude than the effect of assuming a one-generation pulse. Especially recombination rate variation may pose a practical limitation to the accuracy of admixture date estimates, and has to be very carefully considered when making inferences about admixture times. This is because both an extended pulse, as well as a non-homogeneous recombination map, lead to an admixture segment distribution that deviates from the expected exponential distribution. Correcting for just one of these factors may thus potentially conflate these issues, and can lead to a substantial underestimate of mean admixture times (Sankararaman et al., 2012). Therefore, population-specific fine-scale recombination maps are needed for accurate admixture time estimates, at least for admixture that happened more than a thousand generations ago. Estimates of more recent admixture appear to be somewhat more robust, perhaps because coarser-scale recombination maps are better estimated and differ less between populations (Hinch et al., 2011).

To further refine admixture timing, time series data from more admixed early modern human and Neandertal genomes are needed. In particular, measures based on population differentiation (e.g Wall et al., 2013; Browning et al., 2018; Villanea and Schraiber, 2019) are very promising to understand the different events that contributed to archaic ancestry in present-day humans. While Neandertal ancestry in present-day people has been largely homogenized due to the substantial gene flow between populations, samples from both the Neandertal and early modern human populations immediately involved with the gene flow could likely refine when and where this gene flow happened.

# Material & Methods

## Simulations

We test our approach with coalescence simulations using `msprime` (Kelleher et al., 2016). We focus on scenarios mimicking Neandertal admixture, and choose sample sizes to reflect those available from the 1000 Genomes data (The 1000 Genomes Project Consortium, 2015). For ALD simulations we simulate 176 diploid African individuals and 170 diploid non-Africans, corresponding to the number of Yoruba (YRI) and Central Europeans from Utah (CEU). Since three high coverage Neandertal genomes are available (Prüfer et al., 2013, 2017; Mafessoni et al., 2020) we simulate three diploid Neandertal genomes. For each individual we simulate 20 chromosomes with a length of 150 Mb each. The mutation rate is set for all simulations to $2 * 10^{-8}$ per base per generation. The recombination rate is set to $1 * 10^{-8}$ per base pair per generation unless specified otherwise. The demographic parameters are based on previous studies dating Neandertal admixture (Sankararaman et al., 2012; Fu et al., 2014; Moorjani et al., 2016). In the "simple" demographic model (S. Figure 1 A), the effective population size is assumed constant at $N_e = 10,000$ for all populations, the split time between modern humans and Neandertals is 10,000 generations ago and the split between Africans and non-Africans is 2,550 generations ago. The migration rate from Neandertals into non-Africans was set to zero before the split from Africans, to ensure that there is no Neandertal ancestry in Africans. Each simulation was repeated 100 times.

## Simulating admixture

We specify simulations under the extended pulse model using the mean admixture time $t_m$ and the duration $t_d$. We recover the simple pulse model by setting $t_d = 1$. To obtain the migration rates in each generation, we use a discretized version of the migration density (eq 7), which we then scale to the total amount of Neandertal ancestry in modern humans (here 0.03).

## Recombination maps

Uncertainties in the recombination map were previously shown to influence admixture time estimates (Sankararaman et al., 2012; Fu et al., 2014; Sankararaman et al., 2016). To investigate the effect of more realistic recombination rate variation, we perform simulations using empirical recombination maps. We use the African-American map (Hinch et al., 2011) for simulations to evaluate the relative importance of simulation and modeling parameters for admixture time estimates. The HapMap phase 3 map (HapMap Consortium, 2007) is used for evaluating inference using the extended pulse model. For simplicity, we use the same recombination map (150 Mb of chromosome 1, excluding the first 10 Mb) for all simulated chromosomes. When simulating under an empirical map, with the analysis assuming a constant rate (*i.e.* no correction), we use the mean recombination rate from the respective map to calculate the genetic distance from the physical distance for each SNP. The mean recombination rate is calculated from the 150 Mb map (1.017 $\frac{cM}{Mb}$ AAMap and 0.992 $\frac{cM}{Mb}$ HapMap).

## Complex demography

Demographic factors such as population size changes are known to influence LD patterns, and may thus influence admixture time estimates (Gravel, 2012; Liang and Nielsen, 2014). To test the impact of demographic history on admixture time estimates, we contrast our standard model with a more realistic and complex demographic history(S. Figure 1 B). The effective population sizes change over time, are based on estimates from Neandertal and present-day human genomes. In particular, we use the MSMC estimates for YRI and CEU to model Africans and Europeans, respectively (Schiffels and Durbin, 2014). The Neandertal population size history is modelled after PSMC estimates (Li and Durbin, 2011) using the high-coverage Vindija 33.19 genome (Prüfer et al., 2017).

The split times between Neandertals and modern humans is kept the same as in the simple demographic simulations (10,000 generations ago) (S. Figure 1 A). We add substructure within Africa starting from 3,200 till 2,550 generations ago with a per generation migration rate between the two subpopulations of 0.001. The population size of the common ancestor of Neandertals and humans is set to 18,296 (based on MSMC results). To mimic the age of the samples, Neandertals are sampled 750 generations before the Africans and non-Africans. Finally, we add recent gene flow between African and non-African populations with a total migration rate of 0.01 starting from 200 till 50 generations ago (Petr et al., 2019).

## Ascertainment scheme

Since ALD for ancient admixture events can be quite similar to the genomic background, SNPs need to be ascertained to enrich for Neandertal informative sites in the test population.This removes noise and amplifies the ALD signal (Sankararaman et al., 2012). We evaluate the impact of the ascertainment scheme by contrasting two distinct schemes (Sankararaman et al., 2012; Fu et al., 2014). The lower-enrichment ascertainment scheme (LES) only considers sites that are fixed for the ancestral state in Africans and polymorphic or fixed derived in Neandertals. The higher-enrichment ascertainment scheme (HES) is more restrictive in that it further excludes all sites that are not polymorphic in non-Africans.

## ALD calculation

The pairwise weighted LD between the ascertained SNPs a certain genetic distance $d$ apart is calculated using ALDER (Loh et al., 2013). A minimal genetic distance $d_0$ between SNPs is set either to 0.02 cM and 0.05 cM. This minimal distance cutoff removes extremely short-range LD, which might also be due to inheritance of segments from the ancestral population (incomplete lineage sorting ILS) and not gene flow.

## Modeling parameter effect sizes

We consider a set of five parameters:

- ascertainment scheme: $A_i$ = LES/HES
- minimal genetic distance: $M_i = 0.02cM/0.05cM$
- demography: $D_i$ = simple/complex
- recombination rate: $R_i$ = constant/variable
- gene flow model: $G_i$ = simple/extended

We perform simulations using all possible parameter combinations. We define a standard model having a constant recombination rate, simple demography and gene flow, LES ascertainment and $d_0 = 0.05$. The genetic distance is assigned from the physical position using the average recombination rate of the African-American genetic map (*i.e.*, assuming the recombination rate is constant over the simulated chromosome given by this value) for simulations under a variable recombination rate. For each of the possible sets of parameters, we simulate 100 replicates each and fit ALD decay curves. We excluded a very small number of simulations for which the curve could not be estimated (10 out of 3,200).

To estimate the effect size of the different parameters (eq. 16) we use a Bayesian Generalized Linear Model (GLM):

$$
\begin{aligned}
E_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha + \beta_a A_i + \beta_m M_i + \beta_d D_i + \beta_r R_i + \beta_g G_i \\
\alpha &\sim \text{Normal}(0, 2) \\
\beta_a, \beta_m, \beta_d, \beta_r, \beta_g &\sim \text{Normal}(0, 2) \\
\sigma &\sim \text{Exponential}(1)
\end{aligned} \tag{16}
$$

The response variable $E_i$ for this model are the residuals, *i.e.*, the difference between the estimated admixture time $t_{est}$ and simulated admixture time $t_{sim}$ for each simulation:

$$E_i = \frac{t_{est} - t_{sim}}{\sigma_{t_{est}}}$$

, where $\sigma$ is the standard deviation of $t_{est}$ , and the model variables $A, M, D, R$ and $G$ are binary predictors. We assume a Normal likelihood because it is the maximum entropy distribution in our case. We obtained the posterior probability using a Hamiltonian Monte Carlo MCMC algorithm, as implemented in STAN (Carpenter et al., 2017) using an R interface (Stan Development Team, 2018; McElreath, 2020). The Markov chains converged to the target distribution (Rhat = 1) and efficiently sampled from the posterior (S. Table 1).

### Estimating admixture time from real data

To evaluate when we can reject a given admixture durations (including a one generation pulse), we applied the extended pulse model on real data from the 1000 Genomes project (The 1000 Genomes Project Consortium, 2015).

We used the 1000 Genomes phase 3 data together with the Altai, Vindija and Chagyrskaya high coverage Neandertals, including the 107 unrelated individuals from the YRI as representatives of unadmixed Africans and all CEU as admixed Europeans. We only consider biallelic sites, and determine the ancestral allele using the Chimpanzee reference genome (panTro4). We used the CEU specific fine-scale recombination map (Spence and Song, 2019) to convert the physical distance between sites into genetic distance.

## Acknowledgments

We thank Svante Pääbo, Janet Kelso, Fabrizio Mafessoni, Stéphane Peyrégne, Laurits Skov, Divyaratan Popli, Alba Bossom Mesa and Arev Sümer for helpful comments and discussions. This work was supported by the Max Planck Society and the European Research Council (grant number: 694707) to Svante Pääbo.

## Data Availability

The simulation script, the analysis pipeline and the *Extended Admixture Pulse* model is available on GitHub: https://github.com/LeonardoIasi/Extended_Admixture_Pulse. No new data were generated for this study.

## Competing interests

The authors declare that they have no competing interests.

## References


E. Bard, T. J. Heaton, S. Talamo, B. Kromer, R. W. Reimer, and P. J. Reimer. Extended dilation of the radiocarbon time scale between 40,000 and 48,000 y BP and the overlap between Neanderthals and Homo sapiens. *Proceedings of the National Academy of Sciences*, 117(35): 21005–21007, Sept. 2020. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2012307117. URL https://www.pnas.org/content/117/35/21005.

S. R. Browning, B. L. Browning, Y. Zhou, S. Tucci, and J. M. Akey. Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell*, 173(1):53–61.e9, Mar. 2018. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2018.02.031. URL https://www.cell.com/cell/abstract/S0092-8674(18)30175-2.

B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1):1–32, Jan. 2017. ISSN 1548-7660. doi: 10.18637/jss.v076.i01. URL https://www.jstatsoft.org/index.php/jss/article/view/v076i01.

R. Chakraborty and K. M. Weiss. 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, Dec. 1988. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.85.23.9119. URL https://www.pnas.org/content/85/23/9119.

E. R. Chimusa, J. Defo, P. K. Thami, D. Awany, D. D. Mulisa, I. Allali, H. Ghazal, A. Moussa, and G. K. Mazandu. Dating admixture events is unsolved problem in multi-way admixed populations. *Briefings in Bioinformatics*, Nov. 2018. doi: 10.1093/bib/bby112. URL https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bby112/5168590.

K. Douka, V. Slon, Z. Jacobs, C. B. Ramsey, M. V. Shunkov, A. P. Derevianko, F. Mafessoni, M. B. Kozlikin, B. Li, R. Grün, et al. Age estimates for hominin fossils and the onset of the Upper Palaeolithic at Denisova Cave. *Nature*, 565(7741):640–644, Jan. 2019. ISSN 1476-4687. doi: 10.1038/s41586-018-0870-z. URL https://www.nature.com/articles/s41586-018-0870-z.

D. Falush, M. Stephens, and J. K. Pritchard. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics*, 164(4):1567–1587, Aug. 2003. ISSN 0016-6731, 1943-2631. URL https://www.genetics.org/content/164/4/1567.

Q. Fu, H. Li, P. Moorjani, F. Jay, S. M. Slepchenko, A. A. Bondarev, P. L. F. Johnson, A. Aximu-Petri, K. Prüfer, C. de Filippo, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*, 514:445, Oct. 2014. URL https://doi.org/10.1038/nature13810.

Q. Fu, M. Hajdinjak, O. T. Moldovan, S. Constantin, S. Mallick, P. Skoglund, N. Patterson, N. Rohland, I. Lazaridis, B. Nickel, et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature*, 524(7564):216–219, Aug. 2015. ISSN 1476-4687. doi: 10.1038/nature14558. URL https://www.nature.com/articles/nature14558.

S. Gravel. Population Genetics Models of Local Ancestry. *Genetics*, 191(2):607–619, June 2012. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.112.139808. URL https://www.genetics.org/content/191/2/607.

R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H.-Y. Fritz, et al. A Draft Sequence of the Neandertal Genome. *Science*, 328(5979): 710, May 2010. doi: 10.1126/science.1188021. URL http://science.sciencemag.org/content/328/5979/710.abstract.

HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164):851–861, Oct. 2007. ISSN 0028-0836. doi: 10.1038/nature06258. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2689609/.

K. Harris and R. Nielsen. The Genetic Cost of Neanderthal Introgression. *Genetics*, 203(2): 881–891, June 2016. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.116.186890. URL https://www.genetics.org/content/203/2/881.

G. Hellenthal, G. B. J. Busby, G. Band, J. F. Wilson, C. Capelli, D. Falush, and S. Myers. A Genetic Atlas of Human Admixture History. *Science*, 343(6172):747–751, Feb. 2014. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1243518. URL https://science.sciencemag.org/content/343/6172/747.

I. Hershkovitz, G. W. Weber, R. Quam, M. Duval, R. Grün, L. Kinsley, A. Ayalon, M. Bar-Matthews, H. Valladas, N. Mercier, et al. The earliest modern humans outside Africa. *Science*, 359 (6374):456–459, Jan. 2018. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aap8369. URL https://science.sciencemag.org/content/359/6374/456.

T. Higham, K. Douka, R. Wood, C. B. Ramsey, F. Brock, L. Basell, M. Camps, A. Arrizabalaga, J. Baena, C. Barroso-Ruíz, et al. The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature*, 512(7514):306–309, Aug. 2014. ISSN 1476-4687. doi: 10.1038/nature13621. URL https://www.nature.com/articles/nature13621.

A. G. Hinch, A. Tandon, N. Patterson, Y. Song, N. Rohland, C. D. Palmer, G. K. Chen, K. Wang, S. G. Buxbaum, E. L. Akylbekova, et al. The landscape of recombination in African Americans. *Nature*, 476:170, July 2011. URL https://doi.org/10.1038/nature10336.

G. S. Jacobs, G. Hudjashov, L. Saag, P. Kusuma, C. C. Darusallam, D. J. Lawson, M. Mondal, L. Pagani, F.-X. Ricaut, M. Stoneking, et al. Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell*, 177(4):1010–1021.e32, May 2019. ISSN 0092-8674, 1097-4172. doi: 10.1016/j.cell.2019.02.035. URL https://www.cell.com/cell/abstract/S0092-8674(19)30218-1.

I. Juric, S. Aeschbacher, and G. Coop. The Strength of Selection against Neanderthal Introgression. *PLOS Genetics*, 12(11):e1006340, Nov. 2016. ISSN 1553-7404. doi: 10.1371/journal.pgen.1006340. URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006340.

J. Kelleher, A. M. Etheridge, and G. McVean. Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLOS Computational Biology*, 12(5):e1004842, May 2016. doi: 10.1371/journal.pcbi.1004842. URL https://doi.org/10.1371/journal.pcbi.1004842.

B. Kim and K. Lohmueller. Selection and Reduced Population Size Cannot Explain Higher Amounts of Neandertal Ancestry in East Asian than in European Human Populations. *American Journal of Human Genetics*, 96(3):454–461, Mar. 2015. ISSN 0002-9297. doi: 10.1016/j.ajhg.2014.12.029. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4375557/.

I. Lazaridis, D. Nadel, G. Rollefson, D. C. Merrett, N. Rohland, S. Mallick, D. Fernandes, M. Novak, B. Gamarra, K. Sirak, et al. Genomic insights into the origin of farming in the ancient Near East. *Nature*, 536(7617):419–424, Aug. 2016. ISSN 1476-4687. doi: 10.1038/nature19310. URL https://www.nature.com/articles/nature19310.

H. Li and R. Durbin. Inference of human population history from individual whole-genome sequences. *Nature*, 475:493, July 2011. URL https://doi.org/10.1038/nature10231.

M. Liang and R. Nielsen. The Lengths of Admixture Tracts. *Genetics*, 197(3):953–967, July 2014. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.114.162362. URL https://www.genetics.org/content/197/3/953.

P.-R. Loh, M. Lipson, N. Patterson, P. Moorjani, J. K. Pickrell, D. Reich, and B. Berger. Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. *Genetics*, 193(4):1233, Apr. 2013. doi: 10.1534/genetics.112.147330. URL http://www.genetics.org/content/193/4/1233.abstract.

F. Mafessoni, S. Grote, C. d. Filippo, V. Slon, K. A. Kolobova, B. Viola, S. V. Markin, M. Chintalapati, S. Peyrégne, L. Skov, et al. A high-coverage Neandertal genome from Chagyrskaya Cave. *Proceedings of the National Academy of Sciences*, June 2020. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2004944117. URL https://www.pnas.org/content/early/2020/06/15/2004944117.

A.-S. Malaspinas, M. C. Westaway, C. Muller, V. C. Sousa, O. Lao, I. Alves, A. Bergström, G. Athanasiadis, J. Y. Cheng, J. E. Crawford, et al. A genomic history of Aboriginal Australia. *Nature*, 538(7624):207–214, Oct. 2016. ISSN 1476-4687. doi: 10.1038/nature18299. URL https://www.nature.com/articles/nature18299.

D. Massilani, L. Skov, M. Hajdinjak, B. Gunchinsuren, D. Tseveendorj, S. Yi, J. Lee, S. Nagel, B. Nickel, T. Devièse, et al. Denisovan ancestry and population history of early East Asians. *Science*, 370(6516):579–583, Oct. 2020. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abc1166. URL https://science.sciencemag.org/content/370/6516/579.

R. McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*. CRC Press, Mar. 2020. ISBN 978-0-429-63914-2. Google-Books-ID: 6H_WDwAAQBAJ.

M. Meyer, M. Kircher, M.-T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, F. Jay, K. Prüfer, C. de Filippo, et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science (New York, N.Y.)*, 338(6104):222–226, Oct. 2012. ISSN 1095-9203. doi: 10.1126/science.1224344.

P. Moorjani, N. Patterson, J. N. Hirschhorn, A. Keinan, L. Hao, G. Atzmon, E. Burns, H. Ostrer, A. L. Price, and D. Reich. The History of African Gene Flow into Southern Europeans, Levantines, and Jews. *PLOS Genetics*, 7(4):e1001373, Apr. 2011. doi: 10.1371/journal.pgen.1001373. URL https://doi.org/10.1371/journal.pgen.1001373.

P. Moorjani, S. Sankararaman, Q. Fu, M. Przeworski, N. Patterson, and D. Reich. A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proceedings of the National Academy of Sciences*, 113(20):5652, May 2016. doi: 10.1073/pnas.1514696113. URL http://www.pnas.org/content/113/20/5652.abstract.

M. Petr, S. Pääbo, J. Kelso, and B. Vernot. Limits of long-term selection against Neandertal introgression. *Proceedings of the National Academy of Sciences*, 116(5):1639–1644, Jan. 2019. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1814338116. URL https://www.pnas.org/content/116/5/1639.

J. K. Pickrell, N. Patterson, P.-R. Loh, M. Lipson, B. Berger, M. Stoneking, B. Pakendorf, and D. Reich. Ancient west Eurasian ancestry in southern and eastern Africa. *Proceedings of the National Academy of Sciences*, 111(7):2632–2637, Feb. 2014. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1313787111. URL https://www.pnas.org/content/111/7/2632.

J. E. Pool and R. Nielsen. Inference of Historical Changes in Migration Rate From the Lengths of Migrant Tracts. *Genetics*, 181(2):711–719, Feb. 2009. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.108.098095. URL https://www.genetics.org/content/181/2/711.

K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 505:43, Dec. 2013. URL https://doi.org/10.1038/nature12886.

K. Prüfer, C. de Filippo, S. Grote, F. Mafessoni, P. Korlević, M. Hajdinjak, B. Vernot, L. Skov, P. Hsieh, S. Peyrégne, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*, 358(6363):655, Nov. 2017. doi: 10.1126/science.aao1887. URL http://science.sciencemag.org/content/358/6363/655.abstract.

I. Pugach, R. Matveyev, A. Wollstein, M. Kayser, and M. Stoneking. Dating the age of admixture via wavelet transform analysis of genome-wide data. *Genome Biology*, 12(2):R19, Feb. 2011. ISSN 1474-760X. doi: 10.1186/gb-2011-12-2-r19. URL https://doi.org/10.1186/gb-2011-12-2-r19.

I. Pugach, A. T. Duggan, D. A. Merriwether, F. R. Friedlaender, J. S. Friedlaender, and M. Stoneking. The Gateway from Near into Remote Oceania: New Insights from Genome-Wide Data. *Molecular Biology and Evolution*, 35(4):871–886, Apr. 2018. ISSN 0737-4038. doi: 10.1093/molbev/msx333. URL https://academic.oup.com/mbe/article/35/4/871/4782510.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/.

F. Racimo, D. Marnetto, and E. Huerta-Sánchez. Signatures of Archaic Adaptive Introgression in Present-Day Human Populations. *Molecular Biology and Evolution*, 34(2):296–317, Feb. 2017. ISSN 0737-4038. doi: 10.1093/molbev/msw216. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5400396/.

P. Ralph and G. Coop. The Geography of Recent Genetic Ancestry across Europe. *PLOS Biology*, 11(5):e1001555, May 2013. ISSN 1545-7885. doi: 10.1371/journal.pbio.1001555. URL https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001555.

D. Reich, R. E. Green, M. Kircher, J. Krause, N. Patterson, E. Y. Durand, B. Viola, A. W. Briggs, U. Stenzel, P. L. F. Johnson, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, 468:1053, Dec. 2010. URL https://doi.org/10.1038/nature09710.

S. Sankararaman, N. Patterson, H. Li, S. Pääbo, and D. Reich. The Date of Interbreeding between Neandertals and Modern Humans. *PLOS Genetics*, 8(10):e1002947, Oct. 2012. doi: 10.1371/journal.pgen.1002947. URL https://doi.org/10.1371/journal.pgen.1002947.

S. Sankararaman, S. Mallick, M. Dannemann, K. Prüfer, J. Kelso, S. Pääbo, N. Patterson, and D. Reich. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, 507 (7492):354–357, Mar. 2014. ISSN 1476-4687. doi: 10.1038/nature12961. URL https://www.nature.com/articles/nature12961.

S. Sankararaman, S. Mallick, N. Patterson, and D. Reich. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. *Current Biology*, 26(9):1241–1247, May 2016. ISSN 0960-9822. doi: 10.1016/j.cub.2016.03.037. URL http://www.sciencedirect.com/science/article/pii/S0960982216302470.

S. Schiffels and R. Durbin. Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, 46:919, June 2014. URL https://doi.org/10.1038/ng.3015.

A. Seguin-Orlando, T. S. Korneliussen, M. Sikora, A.-S. Malaspinas, A. Manica, I. Moltke, A. Albrechtsen, A. Ko, A. Margaryan, V. Moiseyev, et al. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science (New York, N.Y.)*, 346(6213):1113–1118, Nov. 2014. ISSN 1095-9203. doi: 10.1126/science.aaa0114.

V. Shchur, J. Svedberg, P. Medina, R. Corbett-Detig, and R. Nielsen. On the distribution of tract lengths during adaptive introgression. *bioRxiv*, page 724815, Aug. 2019. doi: 10.1101/724815. URL https://www.biorxiv.org/content/10.1101/724815v1.

L. Skov, R. Hui, V. Shchur, A. Hobolth, A. Scally, M. H. Schierup, and R. Durbin. Detecting archaic introgression using an unadmixed outgroup. *PLOS Genetics*, 14(9):e1007641, Sept. 2018. ISSN 1553-7404. doi: 10.1371/journal.pgen.1007641. URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007641.

J. P. Spence and Y. S. Song. Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Science Advances*, 5(10):eaaw9206, Oct. 2019. ISSN 2375-2548. doi: 10.1126/sciadv.aaw9206. URL https://advances.sciencemag.org/content/5/10/eaaw9206.

Stan Development Team. RStan: the R interface to Stan, 2018.

J. C. Stephens, D. Briscoe, and S. J. O'Brien. Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. *American Journal of Human Genetics*, 55(4):809–824, Oct. 1994. ISSN 0002-9297.

C. Stringer and J. Galway-Witham. When did modern humans leave Africa? *Science*, 359 (6374):389–390, Jan. 2018. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aas8954. URL https://science.sciencemag.org/content/359/6374/389.

The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*, 526:68, Sept. 2015. URL https://doi.org/10.1038/nature15393.

B. Vernot and J. Akey. Complex History of Admixture between Modern Humans and Neandertals. *American Journal of Human Genetics*, 96(3):448–453, Mar. 2015. ISSN 0002-9297. doi: 10.1016/j.ajhg.2015.01.006. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4375686/.

B. Vernot and J. M. Akey. Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. *Science*, 343(6174):1017–1021, Feb. 2014. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1245938. URL https://science.sciencemag.org/content/343/6174/1017.

B. Vernot, S. Tucci, J. Kelso, J. G. Schraiber, A. B. Wolf, R. M. Gittelman, M. Dannemann, S. Grote, R. C. McCoy, H. Norton, et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science*, 352(6282):235–239, Apr. 2016. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aad9416. URL https://science.sciencemag.org/content/352/6282/235.

F. A. Villanea and J. G. Schraiber. Multiple episodes of interbreeding between Neanderthal and modern humans. *Nature Ecology & Evolution*, 3(1):39–44, Jan. 2019. ISSN 2397-334X. doi: 10.1038/s41559-018-0735-8. URL https://www.nature.com/articles/s41559-018-0735-8.

D. N. Vyas and C. J. Mulligan. Analyses of Neanderthal introgression suggest that Levantine and southern Arabian populations have a shared population history. *American Journal of Physical Anthropology*, 169(2):227–239, 2019. ISSN 1096-8644. doi: 10.1002/ajpa.23818. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ajpa.23818.

J. D. Wall. Detecting ancient admixture in humans using sequence polymorphism data. *Genetics*, 154(3):1271–1279, Mar. 2000. ISSN 0016-6731. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1460992/.

J. D. Wall, M. A. Yang, F. Jay, S. K. Kim, E. Y. Durand, L. S. Stevison, C. Gignoux, A. Woerner, M. F. Hammer, and M. Slatkin. Higher Levels of Neanderthal Ancestry in East Asians than in Europeans. *Genetics*, 194(1):199–209, May 2013. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.112.148213. URL https://www.genetics.org/content/194/1/199.

Y. Zhou, H. Qiu, and S. Xu. Modeling Continuous Admixture Using Admixture-Induced Linkage Disequilibrium. *Scientific Reports*, 7(1):1–10, Feb. 2017a. ISSN 2045-2322. doi: 10.1038/srep43054. URL https://www.nature.com/articles/srep43054.

Y. Zhou, K. Yuan, Y. Yu, X. Ni, P. Xie, E. P. Xing, and S. Xu. Inference of multiple-wave population admixture by modeling decay of linkage disequilibrium with polynomial functions. *Heredity*, 118(5):503–510, May 2017b. ISSN 1365-2540. doi: 10.1038/hdy.2017.5. URL https://www.nature.com/articles/hdy20175.

J. Zilhão, D. Anesin, T. Aubry, E. Badal, D. Cabanes, M. Kehl, N. Klasen, A. Lucena, I. Martín-Lerma, S. Martínez, et al. Precise dating of the Middle-to-Upper Paleolithic transition in Murcia (Spain) supports late Neandertal persistence in Iberia. *Heliyon*, 3(11), Nov. 2017. ISSN 2405-8440. doi: 10.1016/j.heliyon.2017.e00435. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5696381/.