Counting the genetic ancestors from source populations in members of an admixed population

Lily Agranat-Tamir[1,*], Jazlyn A Mooney[2] and Noah A Rosenberg[1]

[1]Department of Biology, Stanford University, Stanford, CA 94305, USA.
[2]Department of Quantitative & Computational Biology, University of Southern California, Los Angeles, CA 90089, USA.

*Corresponding author: 327 Campus Drive, Stanford University, Stanford, CA 94305, USA. Email: lilyat@stanford.edu

Abstract

In a genetically admixed population, admixed individuals possess genealogical and genetic ancestry from multiple source groups. Under a mechanistic model of admixture, we study the number of distinct ancestors from the source populations that the admixture represents. Combining a mechanistic admixture model with a recombination model that describes the probability that a genealogical ancestor is a genetic ancestor, for a member of a genetically admixed population, we count *genetic* ancestors from the source populations—those genealogical ancestors from the source populations who contribute to the genome of the modern admixed individual. We compare patterns in the numbers of genealogical and genetic ancestors across the generations. To illustrate the enumeration of genetic ancestors from source populations in an admixed group, we apply the model to the African-American population, extending recent results on the numbers of African and European genealogical ancestors that contribute to the pedigree of an African-American chosen at random, so that we also evaluate the numbers of African and European *genetic* ancestors who contribute to random African-American genomes. The model suggests that the autosomal genome of a random African-American born in the interval 1960-1965 contains genetic contributions from a mean of 162 African (standard deviation 47, interquartile range 127 to 192) and 32 European ancestors (standard deviation 14, interquartile range 21 to 43). The enumeration of genetic ancestors can potentially be performed in other diploid species in which admixture and recombination models can be specified.

Keywords: admixture; African-American population; genealogies; genetic ancestors; pedigrees

Introduction

The genealogical pedigree of any individual person can be viewed as a structure that has been shaped by demographic events such as migrations and population admixtures. The pedigree contains the individual's recent ancestors, who have contributed in a genealogical sense to the individual, and with increasing probability as time proceeds toward the most recent generations, in a genetic sense as well.

The distinction between genealogical and genetic ancestry is inconsequential in recent generations: an individual necessarily contains genetic material from both parents, and almost certainly from all four grandparents and eight great-grandparents as well. However, genetic transmission involves chromosomal segments, the number of which is finite. Hence, going back in time, the number of genealogical ancestors increases rapidly, and proportionally fewer of them are genetic ancestors: individuals who contribute to the genetic material of the modern individual. In the memorable description of Donnelly (1983), *"This means that someone descended from the Scottish poet Robert Burns (born 1759) probably carries some of his genes, but that someone unilineally descended from the English playwright William Shakespeare (born 1564) is unlikely to have any genes in common with him."*

A number of studies have explored the peculiar consequences of the distinction between genealogical and genetic ancestors (Wiuf and Hein 1997; Baird *et al.* 2003; Matsen and Evans 2008; Gravel and Steel 2015; Buffalo *et al.* 2016; Kelleher

et al. 2016). For example, one simulation study (Rohde *et al.* 2004), based on earlier mathematical work (Chang 1999), argued that the most recent genealogical ancestor shared by all living humans might have lived as few as 5000 years ago, even though the most recent *genetic* ancestor lived much earlier. The rate at which recent genealogical ancestors dissipate from an individual's genetic ancestry has been studied by Coop (2013), who used approximations to the human recombination process in order to calculate the number of autosomal fragments a genealogical ancestor passes to a descendant. Through that quantity, Coop (2013) computed the probability that a genealogical ancestor k generations ago is also a genetic ancestor. This analysis finds that although the number of genealogical ancestors grows exponentially in the number of generations back from the present, the number of genetic ancestors grows only linearly.

Recent admixture introduces a new dimension to the challenge of understanding the distinction between genealogical and genetic ancestry. In a recently admixed population, genealogical ancestors ultimately trace to two or more source populations. Some of these genealogical ancestors are genetic ancestors and some are not, so that the fraction of the genetic ancestors that trace to a specific source group need not equal the corresponding fraction of the genealogical ancestors that trace to that source.

Building on a mechanistic admixture model (Verdu and Rosenberg 2011), we have devised a model for counting genealogical ancestors in an admixed individual's pedigree

(Mooney *et al.* 2023), evaluating the numbers of individuals that enter the pedigree from each specific source population. Our goal here is to extend this genealogical model of an admixed pedigree to count the *genetic* ancestors that enter the pedigree. That is, we seek to count genetic ancestors from a certain source population that contribute to an individual's genome, considering genetic ancestors in each generation in the pedigree.

To answer the new question posed by the study—*how many genetic ancestors from the source populations does the genetic admixture of a random member of an admixed population represent?*—we combine two mathematical approaches. The first is the extension of the admixture model studied by Mooney *et al.* (2023). The second is the method of Coop (2013) for approximating the probability that a genealogical ancestor is also a genetic ancestor. We develop a model that counts across the generations both genealogical and genetic ancestors from a certain source population of an admixed individual. We apply it to the African-American population, elaborating on the strictly genealogical approach of Mooney *et al.* (2023).

For this purpose, extending the work of Mooney *et al.* (2023), for a member of the admixed population, we study the random number of admixed *genealogical* ancestors in the pedigree in each generation by proceeding recursively back in time. From this random variable, we evaluate properties of the number of *genetic* ancestors from the admixed population and the number of genetic ancestors from the *source* populations, as well as the number of genealogical ancestors from the source populations as studied by Mooney *et al.* (2023).

The model

Admixture process

We build upon the model of Verdu and Rosenberg (2011) and Mooney *et al.* (2023), which considers the formation of a new admixed population. Two source populations that were present in generation 0 form the new admixed population in generation 1. After the initial admixture event, in each subsequent generation after generation 1, individuals from both source populations and the admixed population can be parents of an individual in the admixed population. Our interest is in an admixed individual in generation g after the initial admixture.

We call the source populations "source 1" and "source 2." For each $n = 1, 2, \ldots, g$, we denote by $s_{1,n-1}$ the probability that for an admixed individual in generation n (n generations after members of generation 0 admix to form generation 1), a specific parent is from source population 1. We denote by h_{n-1} the probability that the parent is from the admixed population, and by $s_{2,n-1}$ the corresponding probability for source 2. Therefore, for each $n = 1, 2, \ldots, g$, we have $s_{1,n-1} + h_{n-1} + s_{2,n-1} = 1$, recalling that $h_0 = 0$ (Figure 1). The two parents are independent and identically distributed, amounting to an assumption that they are exchangeable members of the previous generation. The population is large, so that the chance that a particular individual is sampled twice can be ignored.

Genealogical ancestors in a pedigree

Consider Figure 2A, describing the pedigree of an admixed individual. Tracing back from the admixed individual on each genealogical line, we eventually reach genealogical ancestors from the source populations. In each lineage that reaches ancestors who are only in source populations, we tabulate only the most recent one in our count of genealogical ancestors from source populations.

Generation

$s_{1,0} + s_{2,0} = 1$

$s_{1,n-1} + h_{n-1} + s_{2,n-1} = 1$
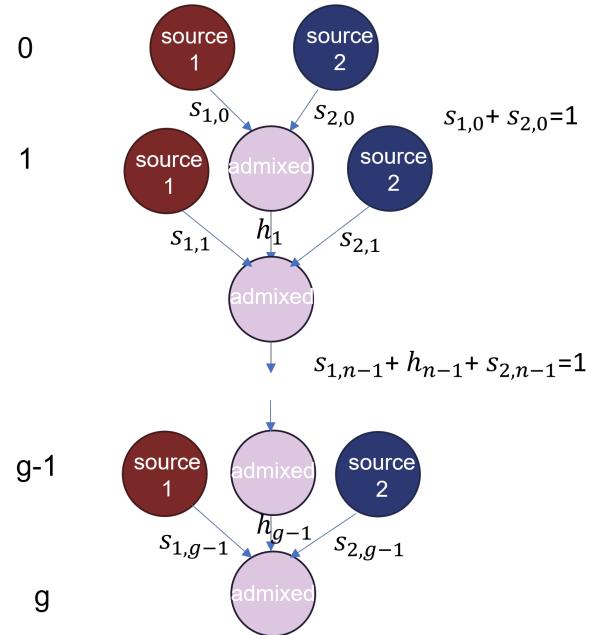
Figure 1 The general admixture model. Starting from generation 0, two source populations form an admixed population in generation 1, with admixture proportions $s_{1,0}$ and $s_{2,0}$. In the following generations, $n = 2, 3, \ldots, g$, the admixed population receives contributions from both the source populations and the admixed population, in proportions $s_{1,n-1}$, $s_{2,n-1}$, and h_{n-1}.

In the figure, some genealogical ancestors are genetic ancestors and some are not. In Mooney *et al.* (2023), we counted genealogical ancestors; the mathematical strategy followed previous studies (Verdu and Rosenberg 2011; Goldberg *et al.* 2014; Goldberg and Rosenberg 2015; Goldberg *et al.* 2020; Kim *et al.* 2021), in which source ancestry proportions were calculated recursively, beginning with the count of ancestors one generation after the initial admixture ($n = 1$), and moving forward in time.

To count genetic ancestors, the approach of Mooney *et al.* (2023) is not straightforward to apply, because the probability that a genealogical ancestor is a genetic ancestor depends on that ancestor's number of generations back from the present, even if the admixture process itself is constant in time. Further, a genetic ancestor of an individual in some generation $g - n$, with $0 < n \leq g$, is not necessarily a genetic ancestor of the individual of interest in generation g.

To address these problems, we develop a model in which we count genealogical and genetic ancestors by proceeding backward in time (Figure 2B,C). Tracing back from the admixed individual of interest in generation g, we examine, in each step, the parents of all the admixed individuals present in the pedigree. We tabulate those who are from a certain source population in our count of genealogical ancestors from that source population (Figure 2B). We tabulate as genetic ancestors those who, in addition to being genealogical ancestors from the source, are also genetic ancestors (Figure 2C). For this step, we use the calculations of Coop (2013) for generation-wise probabilities of genetic ancestry.
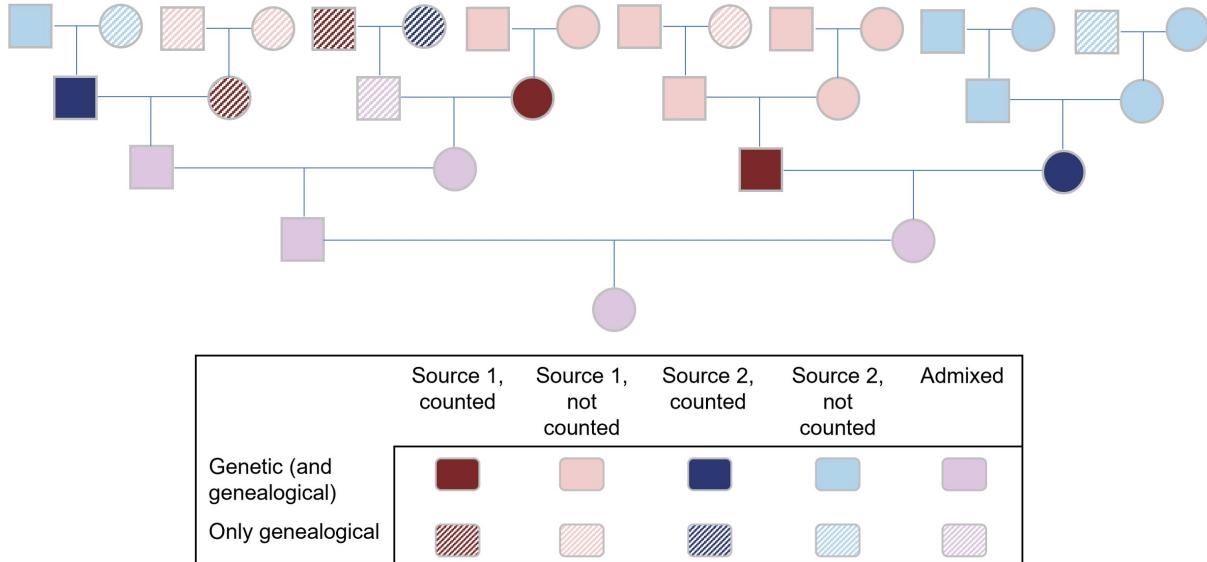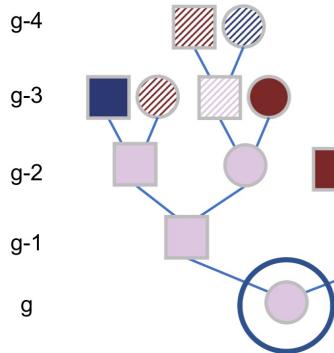
Figure 2 Counting genealogical and genetic ancestors from the source populations for an admixed individual. (A) Pedigree of an admixed individual. Ancestors can be from source populations (red, blue) or from the admixed population itself (pink). Ancestors from the source populations can be both genealogical *and* genetic ancestors (solid color), or genealogical ancestors only (striped). Along each genealogical line that reaches a source population, we count the most recent ancestor (dark color). (B) Counting genealogical ancestors from source populations. For the pedigree in (A), this panel goes back in time from an admixed individual in generation g (circled), on each line stopping when a source population is reached. The number of individuals from source 1 is 4 (red), and the number from source 2 is 3 (blue). (C) Counting *genetic* ancestors from source populations. As in (B), we traverse all admixed individuals in the pedigree, irrespective of genetic ancestry status. However, if a source-1 or 2 ancestor is not a genetic ancestor, then that individual is not tabulated. Note that for ease of interpretation, the figure contains a higher number of genealogical but non-genetic ancestors than is likely in real pedigrees.

Genetic ancestors and recombination

Coop (2013) used a model of recombination in humans to eval-
uate the probability that two individuals with an ancestor–
descendant relationship share at least one piece of DNA. In
other words, the model gives an approximate probability that
a descendant separated by k generations from a genealogical
ancestor possesses at least one genomic fragment from the an-
cestor. The model takes into account approximations to the
recombination process.

In the model of Coop (2013), the number of genomic frag-
ments that a genealogical ancestor passes to a descendant k
generations forward in time is treated as a random variable N_k.
This random variable is approximated as Poisson-distributed
owing to an assumption that recombination breakpoints are
Poisson-distributed. The probability p_k that a genealogical an-
cestor is a genetic ancestor to a k-generation descendant then
equals $1 - \mathbb{P}[N_k = 0] = 1 - e^{-\lambda_k}$, where λ_k is the Poisson mean
$\mathbb{E}[N_k]$.

Considering the autosomal genome, the mean number of
genomic pieces that a parent passes to its offspring, λ_1, is 22,
the number of autosomes. Each generation, on average ev-
ery 100 megabases (Mb) a crossover event occurs, adding one
piece. Because the haploid genome is about 3300 Mb long, each
generation after the first, 33 pieces are added on average. In
each generation back in time after the first, those pieces are dis-
tributed between two parents. Hence, in generation $k \geq 2$, the
total number of pieces for one of an individual's two genomic
copies, maternal or paternal, is $22 + 33(k - 1)$. Those pieces
trace to 2^{k-1} genealogical ancestors k generations back from
the present. Hence, the mean number of fragments contributed
by a specific ancestor k generations back from the present is
$\lambda_k = [22 + 33(k - 1)]/2^{k-1}$. The Poisson probability that *at least
one* fragment traces to such an ancestor then equals 1 minus the
probability that no fragments trace to the ancestor, or for $k \geq 2$,

$$p_k = 1 - e^{-\frac{22+33(k-1)}{2^{k-1}}}. \tag{1}$$

We also define $p_1 = 1$.

Figure 3 shows p_k across the generations, illustrating its de-
cline as k increases. With a 25-year generation time, the claim
(Donnelly 1983) that an individual living in 1983, say, born in
1960, probably possesses genetic material from a randomly cho-
sen genealogical ancestor born in 1759 corresponds to 8 genera-
tions and $p_8 = 0.8615$. The claim that the individual probably
does not possess genetic material from a randomly chosen ge-
nealogical ancestor born in 1564 corresponds to 16 generations
and $p_{16} = 0.0157$. Interestingly, the period in which this prob-
ability of sharing genetic material with an ancestor decreases
from a high to a low number corresponds to the period of in-
terest in the founding of the African-American population, on
which our example analysis focuses.

The human-specific Eq. 1 can be written in a more general
form suitable for other diploid organisms. Denote the number
of pairs of autosomes by q and the haploid genome length in
megabases by ℓ. Denote by m the distance in megabases over
which the mean number of crossover events is 1. As in the special
case for humans, $p_1 = 1$. The probability that a k-generation
($k > 1$) genealogical ancestor is also a genetic ancestor is

$$p_k = 1 - e^{-\frac{q+(\ell/m)(k-1)}{2^{k-1}}}. \tag{2}$$

The computation requires basic parameters of genomes and
recombination maps, quantities that are available for diverse
organisms (Milo and Phillips, 2015; Stapley *et al.*, 2017).
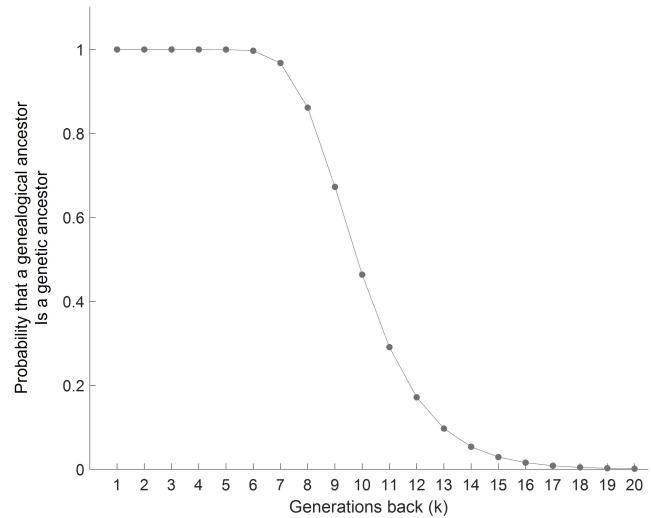
Figure 3 The probability p_k that a genealogical ancestor is an
(autosomal) genetic ancestor as a function of the number of
generations back in time from the present. This plot is based
on Eq. 1.

Results for the general model

To count genetic ancestors from source populations in a pedigree
of a random admixed individual, we first trace the pedigree
back, counting admixed individuals. We then use the count
of admixed genealogical ancestors to count genetic ancestors.
We also show how this approach can be used to recover the
distribution of the number of genealogical ancestors from source
populations in each generation, extending beyond calculations
from Mooney *et al.* (2023) that focused on the expectation.

Counting admixed individuals in a pedigree

Continuing to consider a model with g generations, we now
index generations by k, setting $k = 0$ in generation g, with k
increasing backward in time. Let X_k be the random number of
admixed individuals in the pedigree at step k. When $k = 0$, we
consider a random admixed individual of interest in generation
g, and $X_0 = 1$. For $1 \leq k \leq g$, we proceed backward in time.
At step k, or generation $g - k$, a randomly chosen parent of
an admixed individual in the previous step, or generation $g -
(k - 1)$, has probability h_{g-k} of being an admixed individual.
Consequently, because an individual has two parents, $X_k \sim$
Bin$(2X_{k-1}, h_{g-k})$.

The number of admixed individuals in the pedigree is a non-
homogeneous branching process going back in time. It follows
from Appendix A that for $0 \leq k \leq g$,

$$\mathbb{E}[X_k] = 2^k \prod_{i=1}^{k} h_{g-i}, \tag{3}$$

$$\mathrm{Var}[X_k] = \sum_{i=1}^{k} 2^{k-1+i}[1 - h_{g-(k+1)+i}]$$
$$\times \left[\left(\prod_{j=g-(k+1)+i}^{g-1} h_j \right) \left(\prod_{\ell=g-k}^{g-(k+2)+i} h_\ell^2 \right) \right]. \tag{4}$$

For the sum of the number of admixed genealogical ancestors
across all generations, computing the variance of the sum in

Appendix A, we have

$$\mathbb{E}\left[\sum_{k=1}^{g} X_k\right] = \sum_{k=1}^{g} \mathbb{E}[X_k], \tag{5}$$

$$\mathrm{Var}\left[\sum_{k=1}^{g} X_k\right] = \sum_{k=1}^{g} \mathrm{Var}[X_k] + \sum_{m=1}^{g-1}\sum_{n=m+1}^{g}\left(2^{n-m+1}\prod_{i=1}^{n-m} h_{g-(m+i)}\mathrm{Var}[X_m]\right). \tag{6}$$

Genealogical ancestors

In step k, $1 \le k \le g$, let U_k^i be the random number of source-1 genealogical ancestors of the generation-g admixed individual who are parents of individual i, one of the X_{k-1} admixed genealogical individuals in step $k-1$. Proceeding back in time, after step k, $\sum_{\ell=1}^{k}\sum_{i=1}^{X_{\ell-1}} U_\ell^i$ genealogical ancestors from source 1 have been counted (Figure 2B).

Random variable U_k^i takes values 0, 1, and 2, with probabilities as follows:

$$U_k^i = \begin{cases} 0, & h_{g-k}^2 + 2h_{g-k}s_{2,g-k} + s_{2,g-k}^2, \\ 1, & 2s_{1,g-k}h_{g-k} + 2s_{1,g-k}s_{2,g-k}, \\ 2, & s_{1,g-k}^2. \end{cases} \tag{7}$$

In fact, $U_k^i \sim \mathrm{Bin}(2, s_{1,g-k})$, as $1 - s_{1,g-k} = h_{g-k} + s_{2,g-k}$. The number of source-2 genealogical ancestors can be counted symmetrically by transposing subscripts 1 and 2 in Eq. 7.

The $\{U_k^i\}_{i=1}^{X_{k-1}}$ are independent and identically distributed. Therefore, using $U_k = \sum_{i=1}^{X_{k-1}} U_k^i$ to sum across all X_{k-1} admixed genealogical ancestors in step $k-1$, we have for each k, $1 \le k \le g$,

$$U_k \sim \mathrm{Bin}(2X_{k-1}, s_{1,g-k}). \tag{8}$$

Indeed, considering all parents of the admixed individuals in generation $g-k-1$, the distribution of the vector of counts of genealogical ancestors in source population 1, the admixed population, and source population 2 can be summarized by a multinomial distribution. If we denote by U_k' the number of source-2 genealogical ancestors reached in generation $g-k$, then

$$(U_k, X_k, U_k') \sim \mathrm{Mult}_3[2X_{k-1}, (s_{1,g-k}, h_{g-k}, s_{2,g-k})]. \tag{9}$$

By Eq. 3,

$$\mathbb{E}[U_k] = \mathbb{E}\left[\mathbb{E}[U_k|X_{k-1}]\right] = 2s_{1,g-k}\mathbb{E}[X_{k-1}]$$
$$= 2^k s_{1,g-k}\prod_{i=1}^{k-1} h_{g-i}. \tag{10}$$

This equation accords with the summand in Eq. 12 of Mooney *et al.* (2023), noting that generation i in that equation is equivalent to generation $g-k$ in Eq. 10. If we consider all 2^k genealogical ancestors of the generation-g admixed individual present in step k, $1 \le k \le g$, then the expected fraction of them who are source-1 individuals who are parents of step-$(k-1)$ admixed individuals is $\mathbb{E}[U_k]/2^k$.

We calculate the variance using the law of total variance together with Eqs. 3 and 4:

$$\mathrm{Var}[U_k] = \mathbb{E}\left[\mathrm{Var}[U_k|X_{k-1}]\right] + \mathrm{Var}\left[\mathbb{E}[U_k|X_{k-1}]\right]$$
$$= 2s_{1,g-k}(1-s_{1,g-k})\mathbb{E}[X_{k-1}] + (2s_{1,g-k})^2\mathrm{Var}[X_{k-1}]$$
$$= 2^k s_{1,g-k}(1-s_{1,g-k})\prod_{i=1}^{k-1} h_{g-i}$$
$$+ s_{1,g-k}^2\sum_{i=1}^{k-1} 2^{k+i}(1-h_{g-k+i})$$
$$\times\left[\left(\prod_{j=g-k+i}^{g-1} h_j\right)\left(\prod_{\ell=g-(k-1)}^{g-(k+1)+i} h_\ell^2\right)\right]. \tag{11}$$

We write $\tilde{s}_{1,g-k} = s_{1,g-k}/(1-h_{g-k})$ for convenience. Summing genealogical ancestors across generations in Eqs. 10 and 11 and computing the variance in Appendix B, we have

$$\mathbb{E}\left[\sum_{k=1}^{g} U_k\right] = \sum_{k=1}^{g} \mathbb{E}[U_k], \tag{12}$$

$$\mathrm{Var}\left[\sum_{k=1}^{g} U_k\right] = \sum_{k=1}^{g-1}[2\tilde{s}_{1,g-(k+1)} - \tilde{s}_{1,g-k}]^2\,\mathrm{Var}[X_k]$$
$$+\left[\sum_{m=1}^{g-2}\sum_{n=m+1}^{g-1} 2^{n-m+1}[2\tilde{s}_{1,g-(m+1)} - \tilde{s}_{1,g-m}]\right.$$
$$\times[2\tilde{s}_{1,g-(n+1)} - \tilde{s}_{1,g-n}]$$
$$\left.\times\prod_{i=1}^{n-m} h_{g-(m+i)}\,\mathrm{Var}[X_m]\right]$$
$$+\sum_{k=0}^{g-1} 2s_{1,g-(k+1)}[1 - \tilde{s}_{1,g-(k+1)}]\,\mathbb{E}[X_k]. \tag{13}$$

Genetic ancestors

Next, we count genetic ancestors. Let Y_k^i be the number of source-1 genetic ancestors of the generation-g admixed individual who are parents of individual i, one of the admixed genealogical ancestors in step $k-1$. Proceeding back in time, after step k, $\sum_{\ell=1}^{k}\sum_{i=1}^{X_{\ell-1}} Y_\ell^i$ genetic ancestors from source 1 have been counted. We have for $1 \le k \le g$ probabilities

$$Y_k^i = \begin{cases} 0, & h_{g-k}^2 + 2h_{g-k}s_{2,g-k} + s_{2,g-k}^2 \\ & + s_{1,g-k}^2(1-p_k)^2 + 2s_{1,g-k}h_{g-k}(1-p_k) \\ & + 2s_{1,g-k}s_{2,g-k}(1-p_k), \\ 1, & 2s_{1,g-k}s_{2,g-k}p_k + 2s_{1,g-k}h_{g-k}p_k \\ & + 2s_{1,g-k}^2 p_k(1-p_k), \\ 2, & s_{1,g-k}^2 p_k^2. \end{cases}$$

Here, p_k is the probability that a genealogical ancestor k generations ago is also a genetic ancestor (Eq. 1). The count of genetic ancestors from source 2 is obtained symmetrically.

We can also see that $Y_k^i \sim \mathrm{Bin}(2, s_{1,g-k}p_k)$, as

$$\mathbb{P}[Y_k^i = 1] = 2s_{1,g-k}p_k(1 - s_{1,g-k}p_k),$$
$$\mathbb{P}[Y_k^i = 2] = (s_{1,g-k}p_k)^2.$$

We write $Y_k = \sum_{i=1}^{X_{k-1}} Y_k^i$ for the number of genetic ancestors tabulated in step k. By analogy with the tabulation of genealogical ancestors, we conclude by Eqs. 3 and 4 that for $1 \leq k \leq g$,

$$Y_k \quad \sim \quad \text{Bin}(2X_{k-1}, s_{1,g-k}p_k), \tag{14}$$

$$\mathbb{E}[Y_k] \quad = \quad 2^k s_{1,g-k} p_k \prod_{i=1}^{k-1} h_{g-i}, \tag{15}$$

$$\text{Var}[Y_k] \quad = \quad 2^k s_{1,g-k} p_k (1 - s_{1,g-k} p_k) \prod_{i=1}^{k-1} h_{g-i}$$
$$+ (s_{1,g-k} p_k)^2 \sum_{i=1}^{k-1} 2^{k+i}(1 - h_{g-k+i})$$
$$\times \left[\left(\prod_{j=g-k+i}^{g-1} h_j \right) \left(\prod_{\ell=g-(k-1)}^{g-(k+1)+i} h_\ell^2 \right) \right]. \tag{16}$$

For the sum of the number of genetic ancestors across all generations, computing the variance in Appendix B, we have

$$\mathbb{E}\left[\sum_{k=1}^{g} Y_k \right] \quad = \quad \sum_{k=1}^{g} \mathbb{E}[Y_k], \tag{17}$$

$$\text{Var}\left[\sum_{k=1}^{g} Y_k \right] \quad = \quad \sum_{k=1}^{g-1} [2\tilde{s}_{1,g-(k+1)}p_{k+1} - \tilde{s}_{1,g-k}p_k]^2$$
$$\times \text{Var}[X_m] + \left[\sum_{m=1}^{g-2} \sum_{n=m+1}^{g-1} 2^{n-m+1} \right.$$
$$\times [2\tilde{s}_{1,g-(m+1)}p_{m+1} - \tilde{s}_{1,g-m}p_m]$$
$$\times [2\tilde{s}_{1,g-(n+1)}p_{n+1} - \tilde{s}_{1,g-n}p_n]$$
$$\times \left. \prod_{i=1}^{n-m} h_{g-(m+i)} \text{Var}[X_m] \right]$$
$$+ \sum_{k=0}^{g-1} 2s_{1,g-(k+1)}p_{k+1}$$
$$\times [1 - \tilde{s}_{1,g-(k+1)}p_{k+1}] \mathbb{E}[X_k]. \tag{18}$$

Among all 2^k genealogical ancestors of the generation-g admixed individual who are present in step k, $1 \leq k \leq g$, the expected fraction of them who are source-1 individuals who are parents of step-$(k-1)$ admixed individuals and are genetic ancestors is $\mathbb{E}[Y_k]/2^k$.

In the same way that we count genetic ancestors among the genealogical ancestors from the source populations, we can count the number of admixed genealogical ancestors who are also genetic ancestors. Denoting the random number of admixed genetic ancestors in step k by X_k^*, this random variable is binomially distributed for $1 \leq k \leq g$, so that

$$X_k^* \quad \sim \quad \text{Bin}(2X_{k-1}, h_{g-k}p_k), \tag{19}$$

$$\mathbb{E}[X_k^*] \quad = \quad 2^k h_{g-k} p_k \prod_{i=1}^{k-1} h_{g-i}, \tag{20}$$

$$\text{Var}[X_k^*] \quad = \quad 2^k h_{g-k} p_k (1 - h_{g-k} p_k) \prod_{i=1}^{k-1} h_{g-i}$$
$$+ (h_{g-k}p_k)^2 \sum_{i=1}^{k-1} 2^{k+i}(1 - h_{g-k+i})$$
$$\times \left[\left(\prod_{j=g-k+i}^{g-1} h_j \right) \left(\prod_{\ell=g-(k-1)}^{g-(k+1)+i} h_\ell^2 \right) \right]. \tag{21}$$

The expected fraction of the 2^k genealogical ancestors of the generation-g admixed individual who are themselves admixed individuals and who are also genetic ancestors is $\mathbb{E}[X_k^*]/2^k$.

Considering all parents of the admixed individuals in generation $g - k - 1$, the distribution of the vector of counts of genetic ancestors in source population 1, the admixed population, and source population 2 follows a multinomial distribution. If we denote by Y_k' the number of source-2 genealogical ancestors reached in generation $g - k$, then

$$(Y_k, X_k^*, Y_k') \sim \text{Mult}_3[2X_{k-1}, (s_{1,g-k}p_k, h_{g-k}p_k, s_{2,g-k}p_k)]. \tag{22}$$

For the sum of the number of genetic ancestors across generations, we have

$$\mathbb{E}\left[\sum_{k=1}^{g} X_k^* \right] \quad = \quad \sum_{k=1}^{g} \mathbb{E}[X_k^*], \tag{23}$$

$$\text{Var}\left[\sum_{k=1}^{g} X_k^* \right] \quad = \quad \sum_{k=1}^{g} \text{Var}[X_k^*] + \sum_{m=1}^{g-1} \sum_{n=m+1}^{g} 2^{n-m+1}$$
$$\times \prod_{i=1}^{n-m} h_{g-(m+i)} p_{m+i} \text{Var}[X_m^*]. \tag{24}$$

A single admixture event

We now consider two specific cases of the admixture model, where after the initial generation of admixture, the contributions from the two sources and from the admixed population are constant across generations. First, we study the case in which the constants are 0. We examine the situation in which no subsequent admixture occurs after the admixed population is founded: in other words, $s_{1,0}, s_{2,0} > 0$ and for all n, $1 \leq n \leq g - 1$, $s_{1,n} = s_{2,n} = 0$ and $h_n = 1$.

For each $k = 1, 2, \ldots, g - 1$, the random number of admixed individuals in the pedigree of a randomly chosen admixed individual follows $X_k \sim \text{Bin}(2X_{k-1}, 1)$. Recalling that $X_0 = 1$ for the single admixed individual in generation g, we have $X_k = 2^k$ for all $k = 0, 1, 2, \ldots, g - 1$: all 2^k ancestors of an individual k generations back from the present are admixed.

To consider genealogical ancestors from the source populations, we separate between two cases, $1 \leq k \leq g - 1$ and $k = g$. For $1 \leq k \leq g - 1$, $U_k \sim \text{Bin}(2^k, 0)$ and no individuals from sources 1 and 2 are reached. Consequently, $U_k = 0$ for all k with $1 \leq k \leq g - 1$.

Next, we proceed one generation back from the case of $k = g - 1$. If $k = g$, then by Eq. 8, $U_g \sim \text{Bin}(2 \cdot 2^{g-1}, s_{1,0})$. Therefore, $\mathbb{E}[U_g] = 2^g s_{1,0}$ and $\text{Var}[U_g] = 2^g s_{1,0}(1 - s_{1,0})$.

For genetic ancestors, we again separate $1 \leq k \leq g - 1$ from $k = g$. For $1 \leq k \leq g - 1$, $Y_k \sim \text{Bin}(2^k, 0)$, and the count of genetic ancestors is $Y_k = 0$ for all k with $1 \leq k \leq g - 1$, as is seen with genealogical ancestors. For $k = g$, by Eq. 14, $Y_g \sim \text{Bin}(2^g, s_{1,0}p_g)$. Therefore, $\mathbb{E}[Y_g] = 2^g s_{1,0}p_g$ and $\text{Var}[Y_g] = 2^g s_{1,0}p_g(1 - s_{1,0}p_g)$. The numbers of genetic ancestors from the source populations, like the corresponding numbers of genealogical ancestors, are determined by parameters of the initial admixture, as tabulated by $n = 0$ looking forward in time, or by $k = g$ looking backward.

Constant positive admixture

We now examine the situation in which $s_{1,0}, s_{2,0} > 0$, after which the contributions from the sources are constant and positive. We denote $s_{1,n} = s_1$ and $s_{2,n} = s_2$ for all n, $1 \leq n \leq g - 1$, with

$s_1, s_2 > 0$. Then $h_n = 1 - s_{1,n} - s_{2,n}$ is also constant for all n, $1 \le n \le g - 1$; we denote this constant by $h_n = h$.

Mathematical results

The number of admixed genealogical ancestors X_k follows a homogeneous branching process. For $k = 0$, $\mathbb{E}[X_k] = 1$. By Eq. 3, for $k = 1, 2, \ldots, g - 1$,

$$\mathbb{E}[X_k] = (2h)^k. \tag{25}$$

For $k = g$, $\mathbb{E}[X_k] = 0$.

For the variance of the number of admixed genealogical ancestors, by Eq. 4, $\text{Var}[X_0] = 0$ and for $1 \le k \le g - 1$,

$$
\begin{aligned}
\text{Var}[X_k] &= \sum_{i=1}^{k} 2^{k-1+i}(1-h) \\
&\quad \times \left[\left(\prod_{j=g-(k+1)+i}^{g-1} h \right) \left(\prod_{\ell=g-k}^{g-(k+2)+i} h^2 \right) \right] \\
&= \sum_{i=1}^{k} (1-h) 2^{k-1+i} h^{k-1+i} \\
&= \begin{cases} \frac{1-h}{1-2h}(2h)^k [1 - (2h)^k], & h \ne \frac{1}{2}, \\ \frac{k}{2}, & h = \frac{1}{2}. \end{cases}
\end{aligned} \tag{26}
$$

For $k = g$, $\text{Var}[X_k] = 0$.

To count genealogical and genetic ancestors, we again separate $1 \le k \le g - 1$ from $k = g$. When $k = g$, by Eq. 8, $U_g \sim \text{Bin}(2X_{g-1}, s_{1,0})$. Hence, by Eqs. 10 and 25, for genealogical ancestors, we have

$$\mathbb{E}[U_g] = 2^g h^{g-1} s_{1,0}. \tag{27}$$

For the variance, starting from Eq. 11 and applying Eqs. 25 and 26, we have

$$\text{Var}[U_g] = 2s_{1,0}(1 - s_{1,0})\mathbb{E}[X_{g-1}] + (2s_{1,0})^2 \text{Var}[X_{g-1}]$$

$$
= \begin{cases} 2s_{1,0}(1-s_{1,0})2^{g-1}h^{g-1} \\ + (2s_{1,0})^2(\frac{1-h}{1-2h})(2h)^{g-1}[1-(2h)^{g-1}], & h \ne \frac{1}{2}, \\ 2s_{1,0}(1-s_{1,0})2^{g-1}(\frac{1}{2})^{g-1} + (2s_{1,0})^2(\frac{g-1}{2}), & h = \frac{1}{2}. \end{cases}
$$

$$
= \begin{cases} 2s_{1,0}(2h)^{g-1} \\ \times \{1 - s_{1,0} + 2s_{1,0}(\frac{1-h}{1-2h})[1-(2h)^{g-1}]\}, & h \ne \frac{1}{2}, \\ 2s_{1,0}[1 + s_{1,0}(g-2)], & h = \frac{1}{2}. \end{cases} \tag{28}
$$

For $1 \le k \le g - 1$, by Eq. 8, $U_k \sim \text{Bin}(2X_{k-1}, s_1)$. By Eqs. 10 and 25,

$$\mathbb{E}[U_k] = 2^k h^{k-1} s_1. \tag{29}$$

We then obtain, by Eqs. 11, 25, and 26,

$$
\text{Var}[U_k] = \begin{cases} 2s_1(2h)^{k-1} \\ \times \{1 - s_1 + 2s_1(\frac{1-h}{1-2h})[1-(2h)^{k-1}]\}, & h \ne \frac{1}{2}, \\ 2s_1[1 + s_1(k-2)], & h = \frac{1}{2}. \end{cases} \tag{30}
$$

For genetic ancestors, when $k = g$, similarly to the calculations for genealogical ancestors, we use Eq. 14 to obtain $Y_g \sim \text{Bin}(2X_{g-1}, s_{1,0}p_g)$. By Eqs. 15 and 25,

$$\mathbb{E}[Y_g] = 2^g h^{g-1} s_{1,0} p_g. \tag{31}$$

Following the reasoning underlying Eq. 16 with Eqs. 25 and 26,

$$\text{Var}[Y_g] = 2s_{1,0}p_g(1 - s_{1,0}p_g)\mathbb{E}[X_{g-1}] + (2s_{1,0}p_g)^2 \text{Var}[X_{g-1}]$$

$$
= \begin{cases} 2s_{1,0}p_g(2h)^{g-1} \\ \times \{1 - s_{1,0}p_g + 2s_{1,0}p_g(\frac{1-h}{1-2h})[1-(2h)^{g-1}]\}, & h \ne \frac{1}{2}, \\ 2s_{1,0}p_g[1 + s_{1,0}p_g(g-2)], & h = \frac{1}{2}. \end{cases} \tag{32}
$$

For $1 \le k \le g - 1$, by Eq. 14, $Y_k \sim \text{Bin}(2X_{k-1}, s_1 p_k)$. Hence, by Eqs. 15 and 25,

$$\mathbb{E}[Y_k] = 2^k h^{k-1} s_1 p_k. \tag{33}$$

By Eqs. 16, 25, and 26,

$$
\text{Var}[Y_k] = \begin{cases} 2s_1 p_k(2h)^{k-1} \\ \times \{1 - s_1 p_k + 2s_1 p_k(\frac{1-h}{1-2h})[1-(2h)^{k-1}]\}, & h \ne \frac{1}{2}, \\ 2s_1 p_k[1 + s_1 p_k(k-2)], & h = \frac{1}{2}. \end{cases} \tag{34}
$$

Analysis of temporal trends

In the case of constant positive admixture, we analyze the way in which genealogical and genetic ancestors accumulate across the generations of the admixture process. Comparing generation k, $2 \le k \le g - 1$, to the generation $k - 1$ of its offspring, Eq. 29 gives

$$\frac{\mathbb{E}[U_k]}{\mathbb{E}[U_{k-1}]} = 2h.$$

If $h < \frac{1}{2}$, then $2h < 1$ and $\mathbb{E}[U_k]$ decreases with increasing k and hence decreasing $n = g - k$ (Figure 4A). The number of admixed ancestors is small, so that the source populations are likely to be reached in a small number of generations back from the present; hence, the numbers of genealogical ancestors from the source populations are also small. The contribution from the admixed population is low enough and the contributions from the source populations are high enough that the number of genealogical ancestors from the source populations is greatest in the most recent generations.

If, on the other hand, $h > \frac{1}{2}$, then $2h > 1$ and $\mathbb{E}[U_k]$ increases with increasing k and decreasing $n = g - k$ (Figure 4B). The number of admixed genealogical ancestors is larger than with $h < \frac{1}{2}$, so that the number of genealogical ancestors from the source populations is also larger. With a high contribution from the admixed population to itself, the number of genealogical ancestors from the source populations is greatest farther back in time. A transition occurs at $h = \frac{1}{2}$, where $2h = 1$ and $\mathbb{E}[U_k]$ is constant in time, equaling $2s_1$ by Eq. 29 (Figure 4C).

For genetic ancestors, for $2 \le k \le g - 1$, Eq. 33 gives

$$\frac{\mathbb{E}[Y_k]}{\mathbb{E}[Y_{k-1}]} = 2h\frac{p_k}{p_{k-1}}.$$

Although the admixture process is constant in time after the founding of the admixed population, the dependence of p_k on
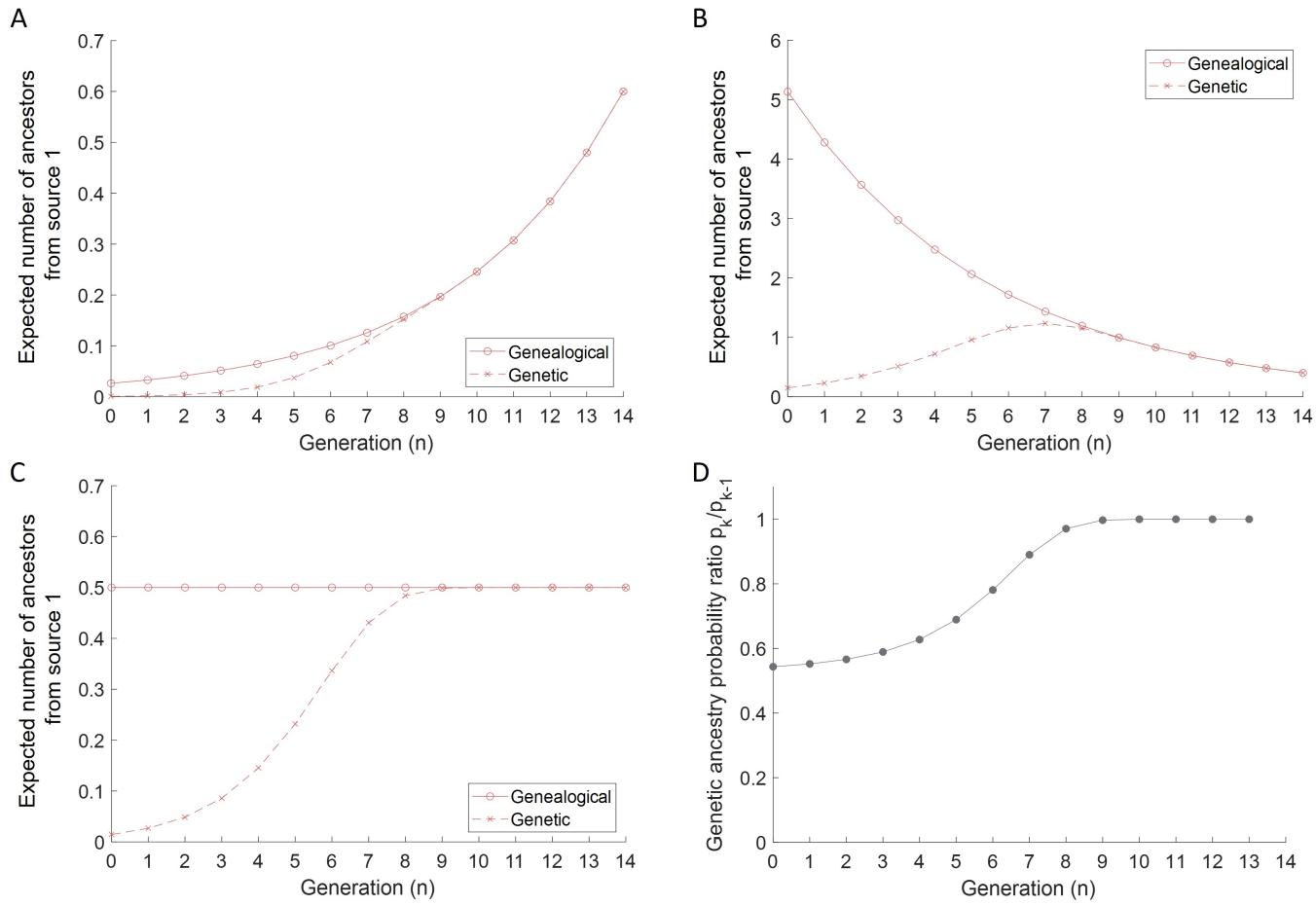
Figure 4 Genealogical and genetic ancestors in a model of constant admixture with $g = 15$, evaluated forward in time from generation $n = 0$ to generation $n = g - 1 = 14$. The forward-time generation n corresponds to the backward-time generation $k = g - n$. (A-C) Expected number of source-1 genealogical ancestors (Eqs. 27, 29) and genetic ancestors (Eqs. 31, 33). The three panels use $s_{1,0} = s_{2,0} = 0.5$ and $s_1 = s_2$ with different values of h. (A) $h = 0.4$. (B) $h = 0.6$. (C) $h = 0.5$. (D) The ratio of the conditional probabilities of genetic ancestry given genealogical ancestry for generations k and $k - 1$, p_k/p_{k-1} (Eq. 1), where $k = 0$ in generation $g = 15$ and $n = g - k$. Note that this plot stops at $n = 13$ and $k = 2$ with the value of p_2/p_1.

k (Eq. 1) affects the time at which genetic ancestors from the sources accumulate.

We now examine p_k/p_{k-1}. Denote the event "a k-generation genealogical ancestor is a k-generation genetic ancestor" by A_k, $k \geq 1$. Irrespective of the form chosen for $\mathbb{P}[A_k]$, we argue that

$$\frac{1}{2} \leq \frac{\mathbb{P}[A_k]}{\mathbb{P}[A_{k-1}]} = \frac{p_k}{p_{k-1}} \leq 1. \tag{35}$$

For the right-hand side of Eq. 35, a necessary condition for a k-generation genealogical ancestor of a descendant to be a k-generation genetic ancestor is that it is a $(k - 1)$-generation genetic ancestor of the parent of the descendant. In other words, $A_k \subseteq A_{k-1}$ and $\mathbb{P}[A_k] \leq \mathbb{P}[A_{k-1}]$.

For the left-hand side of Eq. 35, because $A_k \subseteq A_{k-1}$,

$$\frac{\mathbb{P}[A_k]}{\mathbb{P}[A_{k-1}]} = \frac{\mathbb{P}[A_k \cap A_{k-1}]}{A_{k-1}} = \mathbb{P}[A_k|A_{k-1}].$$

$A_k|A_{k-1}$ is the event that conditional on a k-generation ancestor transmitting at least one genomic segment to the parent of a descendant, the k-generation ancestor transmits at least one segment to the descendant itself. The probability that a parent transmits a certain segment to an offspring is $\frac{1}{2}$, and therefore $\frac{1}{2} \leq \mathbb{P}[A_k|A_{k-1}]$.

For the functional form of $\mathbb{P}[A_k]$ used by Coop (2013), Eq. 1, a proof that $\frac{1}{2} < p_k < 1$ for all $k \geq 2$ appears in Appendix C. An example of p_k/p_{k-1} appears in Figure 4D, illustrating a decrease in p_k/p_{k-1} with increasing k and decreasing $n = g - k$.

Application to African-Americans

Model and methods

We apply our model to count genetic ancestors for a random individual in the African-American population in the United States. In Mooney et al. (2023), relying on demographic data on the history of the population, we considered a model with $g = 14$ generations, ending in 1960-1965. Using information on current patterns of genetic admixture, we inferred admixture parameters $(s_{1,n}, h_n, s_{2,n})$, with source 1 representing Africans and source 2 representing Europeans. The model divided the demographic history of the population into three epochs: 1619-1808, during which the population was founded, with importation of enslaved African captives and admixture with Europeans; 1808-1865, during which enslavement and admixture continued but

Table 1 Generation-specific expectations of the numbers of African-American genealogical and genetic ancestors across accepted parameter sets.

Generation (*n*)	Birth year	Genealogical			Genetic		
		Mean of expectation	Standard deviation of expectation	Mean of standard deviation	Mean of expectation	Standard deviation of expectation	Mean of standard deviation
0	1610-1615	-	-	-	-	-	-
1	1635-1640	0.07	0.03	0.27	0.01	0.00	0.08
2	1660-1665	2.32	0.98	1.83	0.40	0.17	0.66
3	1685-1690	8.93	3.38	4.41	2.60	0.98	1.88
4	1710-1715	33.30	11.15	12.71	15.44	5.17	6.60
5	1735-1740	83.09	24.37	28.93	55.91	16.39	20.05
6	1760-1765	97.93	25.08	33.17	84.36	21.61	28.99
7	1785-1790	62.39	14.22	21.05	60.39	13.76	20.57
8	1810-1815	34.43	6.97	11.51	34.33	6.95	11.52
9	1835-1840	19.04	3.52	6.28	19.04	3.52	6.28
10	1860-1865	10.55	1.87	3.40	10.55	1.87	3.40
11	1885-1890	5.84	0.77	1.83	5.84	0.77	1.83
12	1910-1915	3.24	0.28	0.94	3.24	0.28	0.94
13	1935-1940	1.80	0.08	0.42	1.80	0.08	0.42
Total	-	362.93	90.16	119.94	293.89	69.66	99.54

Suppose θ_i denotes an accepted parameter set and $\boldsymbol{\theta} = \{\theta_i\}_{i=1}^{|\boldsymbol{\theta}|}$ denotes the collection of all accepted parameter sets. For each generation $n = g - k$ with $g = 14$ ($k = 1, 2, \ldots, g$), the mean of the expectation of the genealogical ancestors is $\mathrm{Mean}_{\boldsymbol{\theta}}\{\mathbb{E}[X_k(\theta_i)]\}$ (Eq. 3; Eq. 20 for genetic ancestors); the standard deviation of the expectation is $\sigma_{\boldsymbol{\theta}}\{\mathbb{E}[X_k(\theta_i)]\}$; the mean of the standard deviation is $\mathrm{Mean}_{\boldsymbol{\theta}}\{\sqrt{\mathrm{Var}[X_k(\theta_i)]}\}$ (Eq. 4; Eq. 21 for the genetic ancestors). For the total, the mean of the expectation of the genealogical ancestors is $\mathrm{Mean}_{\boldsymbol{\theta}}\{\mathbb{E}[\sum_{k=1}^{g} X_k(\theta_i)]\}$ (Eq. 5; Eq. 23 for the genetic ancestors); the standard deviation of the expectation is $\sigma_{\boldsymbol{\theta}}\{\mathbb{E}[\sum_{k=1}^{g} X_k(\theta_i)]\}$; the mean of the standard deviation is $\mathrm{Mean}_{\boldsymbol{\theta}}\{\sqrt{\mathrm{Var}[\sum_{k=1}^{g} X_k(\theta_i)]}\}$ (Eq. 6; Eq. 24 for genetic ancestors). The table shows the generation-wise values plotted in Figure 5B for the mean and standard deviation of the expectation.
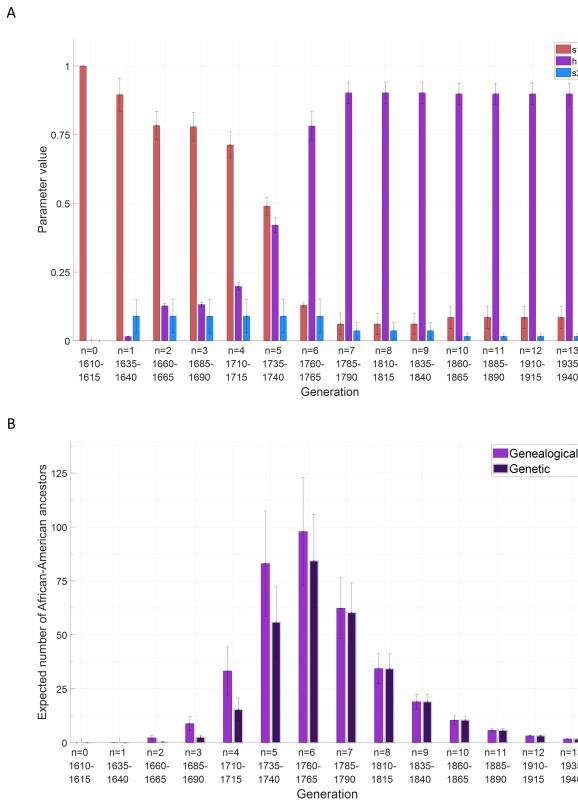
Figure 5 Generation-specific genealogical and genetic ancestry features for African-Americans. (A) Generation-wise mean admixture contributions s_1 (African), h (African-American) and s_2 (European) across accepted parameter sets. Error bars show standard deviations. (B) Means across accepted parameter sets of the expected numbers of African-American genealogical and genetic ancestors possessed by a random individual, as calculated by Eqs. 3 and 20. Error bars show the standard deviations of these expected numbers across accepted parameter sets. The values plotted in (A) are obtained by summarizing the distributions underlying Figure 4 of Mooney et al. (2023). The values in (B) are given in Table 1.

importation of enslaved persons was illegal; and 1865-1965, after the end of legal enslavement. The 1965 endpoint for the model was chosen to accord with the approximate timing of the birth of individuals in whom genetic ancestry has been measured, and to precede subsequent major demographic changes.

The model considered 25-year generations, initializing the population solely with Africans ($s_{1,0} = 1, s_{2,0} = 0$). The first epoch had seven generations (1635-1640, 1660-1665, 1685-1690, 1710-1715, 1735-1740, 1760-1765, 1785-1790; $n = 1$ to 7), the second epoch had three (1810-1815, 1835-1840, 1860-1865; $n = 8$ to 10) and the third had four (1885-1890, 1910-1915, 1935-1940, 1960-1965; $n = 11$ to 14). In the first epoch, $s_{2,n}$ was kept constant, and the values of $s_{1,n}$ and h_n were specified by estimating the value of $s_{1,n}/(s_{1,n} + h_n)$ using demographic data (Hacker 2020) about newly transported enslaved individuals from Africa and births in the African-American populations. In both the second and third epochs, $s_{1,n}$, h_n, and $s_{2,n}$ were maintained as constants for all generations in the epoch.

Mooney et al. (2023) identified sets of parameter values that recovered features of genetic ancestry measured in African-

Americans: an expected African genetic ancestry in $[0.75, 0.85]$ with standard deviation in $[0.08, 0.15]$. A summary of generation-wise mean parameter values across all accepted parameter sets appears in Figure 5A. The figure reports mean values of s_1, h, and s_2, summarizing distributions that appear in Figure 4 of Mooney et al. (2023). It shows the high African contribution to the African-American population in the earliest generations (s_1), with an increasing contribution of the African-American population to itself (h), and with European contributions occurring across the generations (s_2). For each set of accepted parameters, Mooney et al. (2023) calculated the generation-wise expected numbers of African and European genealogical ancestors associated with the set.

Here, using these parameter sets, we calculate the generation-wise expected numbers of African-American genealogical ancestors and the expected numbers of African, European, and African-American *genetic* ancestors (Eq. 15), in a pedigree of a person drawn randomly from the African-American population born between 1960-1965. We also show the distribution across parameter sets, in each generation, of the expected numbers of genealogical ancestors from each population.

Genealogical ancestors

For each accepted parameter set, using Eq. 3, we evaluated the generation-wise expected number of African-American ancestors that appear in a random genealogy, represented by $\mathbb{E}[X_k]$. The mean across accepted parameter sets is shown in Figure 5B and Table 1. Forward in time, the mean number of African-American genealogical ancestors is initially small, increasing to a peak in generation 6 (1760-1765) with a value of 98. It decreases toward the end of the admixture process.

At each generation n, genealogical ancestry is split across five groups: Africans reached in generation n, African-Americans present in generation n, Europeans reached in generation n, Africans who are ancestors to Africans reached subsequent to generation n, and Europeans who are ancestors to Europeans reached subsequent to generation n. The first and third of these categories were studied by Mooney et al. (2023). The fourth and fifth are individuals who are genealogical ancestors of individuals who contributed directly to the African-American population, but who are not themselves parents of African-Americans; the expected number of Africans who are ancestors to African genealogical ancestors reached only subsequent to generation n is obtained from Eq. 10 by $\sum_{i=n+1}^{13} 2^{i-n}\mathbb{E}[U_{14-i}]$. A similar computation can be performed for Europeans.

Figure 6A plots the fractions among all genealogical ancestors assigned to the five categories, and the values plotted appear in Table 2. In the earliest generations, all genealogical ancestors are Africans and Europeans who do not directly contribute to the African-American population. As the admixture continues, African and European genealogical ancestors who directly contribute are reached, and eventually, African-Americans represent most of the genealogical ancestors. In generation 0 (1610-1615), ~79% of genealogical ancestors are African and ~21% are European, reflecting the fractions of an African-American genome that trace to African genetic ancestry and to European genetic ancestry.

Genetic ancestors

Considering the accepted parameter sets from Mooney et al. (2023), we used Eq. 15 to calculate generation-wise expected numbers of African and European *genetic* ancestors. These val-
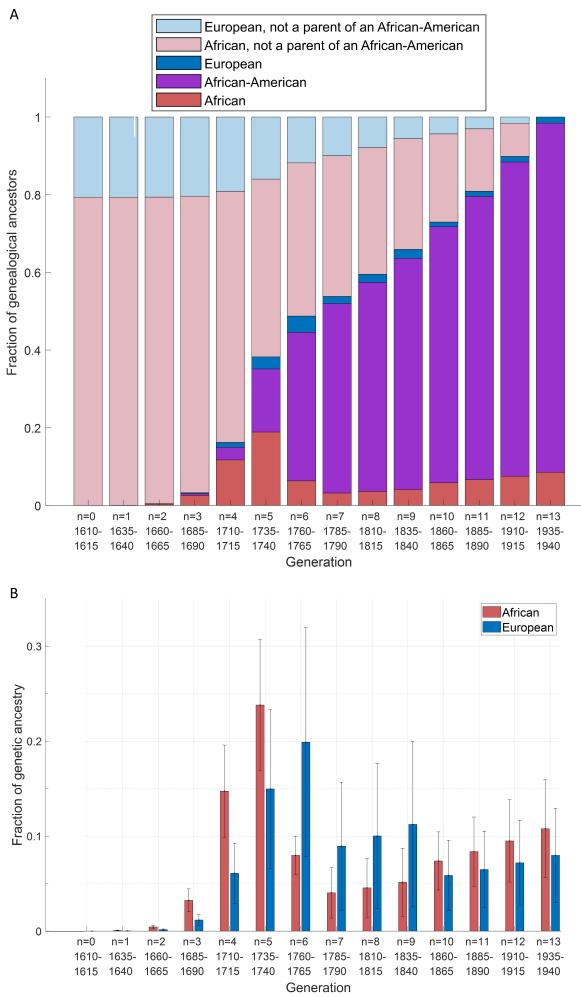
Figure 6 Generation-specific genealogical and genetic ancestry fractions for African-Americans. (A) Generation-wise genealogical ancestry for a random African-American individual, partitioned across five categories and averaged across accepted parameter sets. The fraction of genealogical ancestors who are Africans in generation n who contribute directly to the African-American population is obtained from Eq. 10 as $\mathbb{E}[U_{14-n}]/2^{14-n}$; the fraction of genealogical ancestors who are African but who only contribute to the African-American population through their subsequent African descendants is $(\sum_{i=n+1}^{13} 2^{i-n}\mathbb{E}[U_{14-i}])/2^{14-n}$. Similar calculations are performed for Europeans. The fraction of genealogical ancestors who are African-American is $\mathbb{E}[X_{14-n}]/2^{14-n}$, calculated using Eq. 3. The values plotted appear in Table 2. (B) Generation-wise expected African genetic ancestry contributed to a descendant as a fraction of the total expected African genetic ancestry in the descendant, and expected European genetic ancestry contributed to the descendant as a fraction of the total expected European genetic ancestry in the descendant. The values are obtained from Eq. 36, with $n = 14 - k$. Error bars represent standard deviations of the values from Eq. 36 across accepted parameter sets.

Table 2 Generation-specific expectations of the fractions of genealogical ancestry assigned to five categories, across accepted parameter sets.

Generation (n)	Birth year	Fraction of genealogical ancestors				
		African	African-American	European	African, not counted	European, not counted
0	1610-1615	0.0000	-	-	0.7934	0.2066
1	1635-1640	0.0005	0.0000	0.0000	0.7929	0.2066
2	1660-1665	0.0035	0.0006	0.0003	0.7894	0.2062
3	1685-1690	0.0257	0.0044	0.0024	0.7637	0.2038
4	1710-1715	0.1171	0.0325	0.0127	0.6466	0.1911
5	1735-1740	0.1890	0.1623	0.0313	0.4575	0.1599
6	1760-1765	0.0632	0.3825	0.0417	0.3944	0.1182
7	1785-1790	0.0320	0.4874	0.0186	0.3624	0.0996
8	1810-1815	0.0362	0.5380	0.0208	0.3262	0.0788
9	1835-1840	0.0409	0.5950	0.0234	0.2853	0.0554
10	1860-1865	0.0584	0.6593	0.0118	0.2269	0.0436
11	1885-1890	0.0663	0.7296	0.0130	0.1606	0.0305
12	1910-1915	0.0752	0.8088	0.0145	0.0854	0.0161
13	1935-1940	0.0854	0.8985	0.0161	-	-

The table shows the values plotted in Figure 6A.

ues enable evaluation of expected fractions of the total African and European ancestry that have contributed to a descendant genome by each generation of genetic ancestors. For example, the fraction of the genome that traces to a specific African genetic ancestor from k generations before the descendant is, on average, $1/W_k$, where W_k is the number of genetic ancestors in that generation. W_k has expectation $2^k p_k$, the product of the number of genealogical ancestors k generations ago and the probability that a genealogical ancestor is a genetic ancestor. Therefore, the expected contribution to the African genetic ancestry fraction from all African genetic ancestors k generations before the present can be approximated by $\mathbb{E}[Y_k]/(2^k p_k)$, the ratio of the expected number of African genetic ancestors k generations prior to the descendant and the expected total number of genetic ancestors in that generation. By Eqs. 10 and 15, $\mathbb{E}[Y_k]/(2^k p_k) = \mathbb{E}[U_k]/2^k$.

Figure 6B shows the expected African and European genetic ancestry contributed by the genetic ancestors from each generation as fractions of the total African and European genetic ancestry, or

$$\frac{\mathbb{E}[Y_k]/(2^k p_k)}{\sum_{\ell=1}^{14} \mathbb{E}[Y_\ell]/(2^\ell p_\ell)}. \tag{36}$$

The figure converts between the backward-time perspective indexed by k and the forward-time $n = g - k$. Because a genetic ancestor from the more recent generations (large n) contributes more genetic ancestry on average than a genetic ancestor in previous generations (small n), we observe non-negligible contributions from these later generations. However, ~40% of the African genetic ancestry traces to generations 4 and 5, and ~35% of the European genetic ancestry traces to generations 5 and 6, with an additional ~30% of European genetic ancestry tracing to generations 7, 8, and 9.

The generation-wise mean values across parameter sets of the expected numbers of genetic ancestors appear in Figure 7, alongside expected numbers of genealogical ancestors for comparison. Replotting values from Figure 7 of Mooney *et al.* (2023), the numbers of genealogical ancestors are greater for Africans than for Europeans, and the expected total numbers of genealogical ancestors, summing across generations, are 314 Africans and 51 Europeans (Tables 3 and 4). Looking forward in time from the founding of the population, the numbers of genealogical ancestors increase to peak values and then decrease. The numbers

Table 3 Summary statistics for the expected numbers of African, European, and African-American genealogical and genetic ancestors for a random individual from the African-American population across the accepted parameter sets.

Quantity	Mean	Standard deviation	Minimum	1st quartile	Median	3rd quartile	Maximum
African genealogical ancestors	314	99	124	240	299	376	680
African genetic ancestors	162	47	72	127	155	192	332
European genealogical ancestors	51	24	4	32	51	69	125
European genetic ancestors	32	14	4	21	32	42	77
African-American genealogical ancestors	363	90	202	294	345	418	709
African-American genetic ancestors	294	70	172	240	280	336	566

The estimates consider random individuals in the 1960-1965 birth cohort, assumed to be generation $g = 14$ in a 3-epoch model. The standard deviations are standard deviations of the means across accepted parameter sets; means and standard deviations are rounded from Tables 4 and 1. The values for African and European genealogical ancestors appear in Table 3 in Mooney *et al.* (2023).
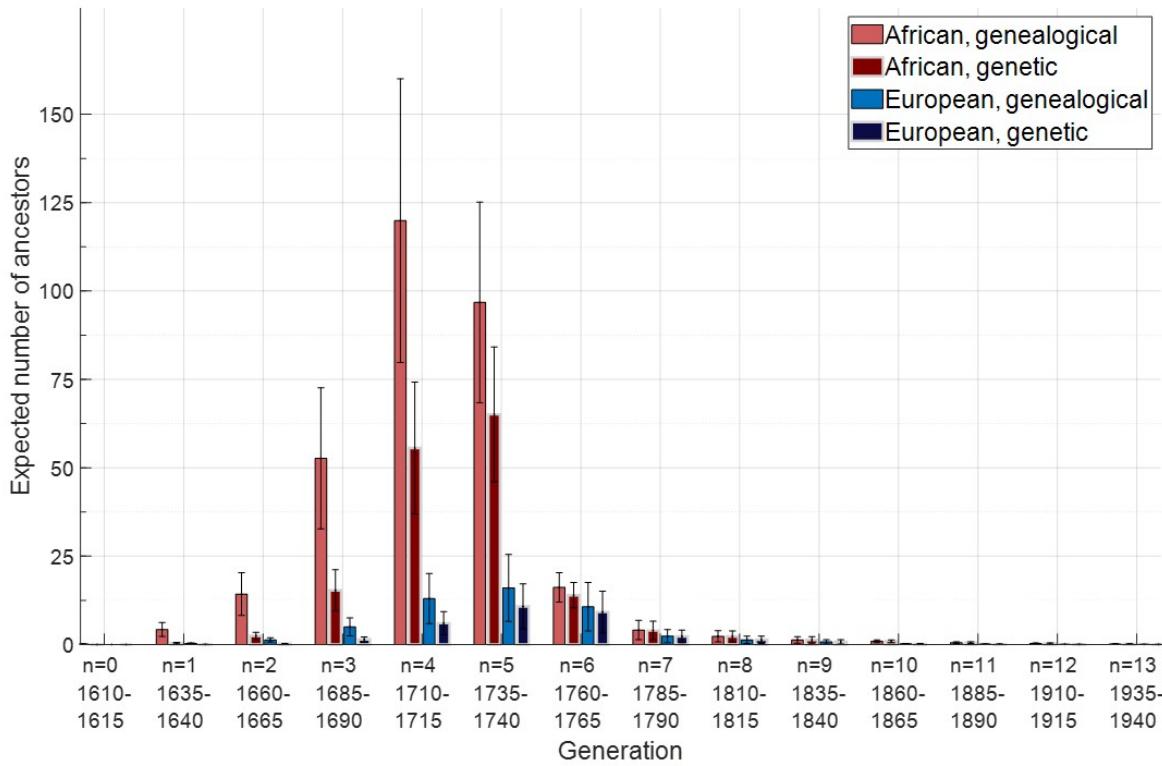
Figure 7 Generation-specific expectations of the numbers of African and European genealogical and genetic ancestors. The expected number of African genealogical ancestors is calculated according to Eq. 10 (standard deviation, Eq. 11). The expected number of African genetic ancestors is calculated according to Eq. 15 (standard deviation, Eq. 16). Similar calculations are performed for Europeans. The plot shows means of the expectation and standard deviation across expected parameter sets. The values plotted appear in Table 4.

of genetic ancestors also reach peaks and decrease toward the present. The expected total numbers of genetic ancestors are 162 Africans and 32 Europeans.

By a similar computation, Figure 5B provides the generation-wise expected numbers of African-American genetic ancestors, comparing them to corresponding numbers of genealogical ancestors. The expected total number of African-American genealogical ancestors, summing from generations 0 to 13, is 363, and the expected total for genetic ancestors is 294 (Tables 1 and 3).

In Figure 7, the peak expected number of African genealogical ancestors appears in generation 4 (1735-1740). However, the corresponding peak for genetic ancestors occurs in generation 5. The difference occurs because the peak for African genealogical ancestors occurs far enough back in time that the probability of genetic ancestry for those genealogical ancestors is well below 1 ($p_{10} = p_{14-4} \approx 0.4637$ by Eq. 1); the number of genetic ancestors among the smaller number of generation-5 genealogical ancestors is greater than among the larger number of generation-4 genealogical ancestors.

For Europeans, the peak of genealogical ancestors occurs later than for Africans, in generation 5 (1760-1765). In that later generation, the fraction of genealogical ancestors who are also genetic ancestors is greater than in generation 4 ($p_9 = p_{14-5} \approx 0.6728$ by Eq. 1). Because the peak in genealogical ancestors occurs later for Europeans, the fraction of all European genealogical ancestors who are genetic ancestors ($\frac{32}{51} \approx 0.63$) exceeds the corresponding fraction for Africans ($\frac{162}{314} \approx 0.52$).

This observation can be illustrated in a computation shown in Figure 8, which compares the ratio of African and European genetic ancestors to the ratio of African and European genealogical ancestors across accepted parameter sets. The African:European ratio of genetic ancestors is consistently lower than the African:European ratio of genealogical ancestors. The comparative recency of the European genealogical ancestors—and the resulting increased probability of genetic ancestry for those genealogical ancestors—produces a greater value for the fraction of all genetic ancestors who are European compared to the fraction of all genealogical ancestors who are European.

In Figure 5B, the peak number of African-American genealogical ancestors appears still later than the peaks for African and European genealogical ancestors, in generation 6 (1785-1790). In that generation, the fraction of genealogical ancestors who are also genetic ancestors is $p_8 = p_{14-6} \approx 0.8615$ (by Eq. 1). Hence, the fraction of African-American genealogical ancestors who are also genetic ancestors ($\frac{294}{363} \approx 0.81$) exceeds corresponding fractions for Africans and Europeans.

Discussion

We have developed an approach to counting genetic ancestors of an admixed individual, estimating the number of genetic ancestors who contributed directly to the admixed population and the number of genetic ancestors belonging to the admixed population itself. The approach proceeds by recursively treating the number of such ancestors in a given generation as a random variable that is binomially distributed based on a corresponding random variable for the subsequent generation. We used an admixture model together with a model of African-American demographic history to estimate that a random African-American born between 1960 and 1965 has an estimated mean of 162 for the number of African genetic ancestors (standard deviation
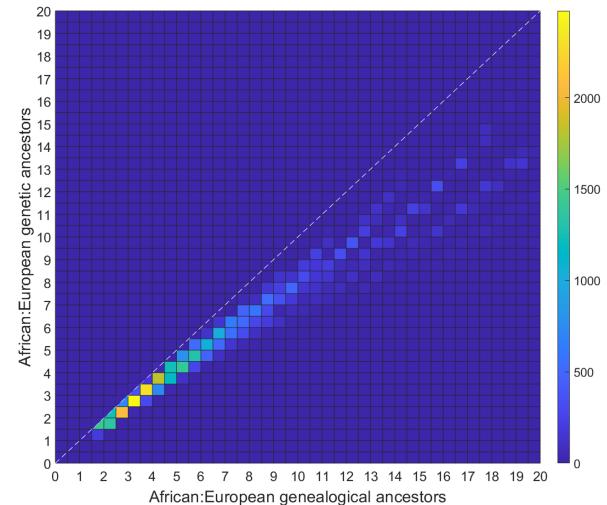
Figure 8 Ratios of the number of African ancestors to the number of European ancestors. The x-axis shows the ratio for genealogical ancestors, and the y-axis shows the ratio for genetic ancestors. For each of 45,189 accepted parameter sets, we calculated $((\sum_{n=0}^{13} \mathbb{E}[U_{14-n}])/(\sum_{n=0}^{13} \mathbb{E}[U'_{14-n}])$, $(\sum_{n=0}^{13} \mathbb{E}[Y_{14-n}])/(\sum_{n=0}^{13} \mathbb{E}[Y'_{14-n}]))$, visualizing the ordered pair of ratios in a density plot. The 89% of the pairs (40,201) that have both ratios below 20 are presented in the plot, with the color of a $\frac{1}{2} \times \frac{1}{2}$ square representing the number of pairs located in that square. The mean ratios across all accepted parameter sets are $(9.99, 7.07)$, and the standard deviations are $(11.87, 6.32)$, with covariance 73.56. For the 89% of points shown, the mean ratios are $(6.74, 5.36)$, with standard deviations $(4.18, 2.96)$ and covariance 12.15. The $y = x$ line is shown for comparison. Among the accepted parameter sets, the ratio we observed for genetic ancestors was always smaller than the ratio for genealogical ancestors; hence, for squares along the diagonal, only the lower triangle is colored. Note that although a smaller value for the ratio of genetic ancestors compared to the ratio of genealogical ancestors was always observed, such a relationship need not hold in principle.

47) and 32 for the number of European genetic ancestors (standard deviation 14) who contributed to the African-American population directly from the source populations, and 294 total African-American genetic ancestors (standard deviation 70).

Genetic and genealogical ancestors

In population-genetic studies of genetically admixed populations, genetic ancestry that traces to the source populations has generally been analyzed by evaluation of estimated admixture fractions in members of an admixed population. The statistical models used for this estimation consider admixture in terms of the fractions of genomes contributed rather than via contributions of specific ancestors. With the increasing use of these genomic contributions to report information to individuals about their own genealogies, the meaning of concepts of genetic ancestry and admixture—and their estimates—have been increasingly queried (Weiss and Long 2009; Lawson *et al.* 2018; Mathieson and Scally 2020). Our use of mechanistic admixture models enables new perspectives on the interpretation of genetic admixture and

Table 4 Generation-specific expectations of the numbers of African and European genealogical and genetic ancestors across accepted parameter sets.

Generation (n)	Birth year	Number of African ancestors					
		Genealogical			Genetic		
		Mean of expectation	Standard deviation of expectation	Mean of standard deviation	Mean of expectation	Standard deviation of expectation	Mean of standard deviation
0	1610-1615	0.14	0.07	0.54	0.01	0.00	0.09
1	1635-1640	4.25	1.99	3.41	0.41	0.19	0.71
2	1660-1665	14.27	6.04	7.28	2.45	1.03	1.93
3	1685-1690	52.70	19.95	20.47	15.33	5.80	6.91
4	1710-1715	119.90	40.15	42.22	55.60	18.62	20.50
5	1735-1740	96.76	28.37	33.45	65.10	19.09	23.12
6	1760-1765	16.18	4.14	6.60	13.94	3.57	5.88
7	1785-1790	4.10	2.71	2.50	3.97	2.62	2.45
8	1810-1815	2.31	1.57	1.70	2.31	1.57	1.70
9	1835-1840	1.31	0.91	1.20	1.31	0.91	1.19
10	1860-1865	0.94	0.39	0.99	0.94	0.39	0.99
11	1885-1890	0.53	0.23	0.72	0.53	0.23	0.72
12	1910-1915	0.30	0.14	0.53	0.30	0.14	0.53
13	1935-1940	0.17	0.08	0.39	0.17	0.08	0.39
Total	-	313.86	98.58	102.62	162.37	46.72	52.66

Generation (n)	Birth year	Number of European ancestors					
		Genealogical			Genetic		
		Mean of expectation	Standard deviation of expectation	Mean of standard deviation	Mean of expectation	Standard deviation of expectation	Mean of standard deviation
0	1610-1615	-	-	-	-	-	-
1	1635-1640	0.32	0.14	0.61	0.03	0.01	0.18
2	1660-1665	1.28	0.61	1.31	0.22	0.10	0.48
3	1685-1690	4.98	2.52	3.10	1.45	0.73	1.36
4	1710-1715	12.98	7.07	6.50	6.02	3.28	3.53
5	1735-1740	16.01	9.44	7.80	10.77	6.35	5.60
6	1760-1765	10.67	6.85	5.58	9.19	5.90	4.96
7	1785-1790	2.38	1.84	1.84	2.30	1.78	1.81
8	1810-1815	1.33	1.04	1.27	1.33	1.04	1.27
9	1835-1840	0.75	0.60	0.90	0.75	0.60	0.90
10	1860-1865	0.19	0.12	0.44	0.19	0.12	0.44
11	1885-1890	0.10	0.06	0.32	0.10	0.06	0.32
12	1910-1915	0.06	0.03	0.24	0.06	0.03	0.24
13	1935-1940	0.03	0.02	0.18	0.03	0.02	0.18
Total	-	51.08	24.32	18.68	32.44	14.31	12.18

Values are calculated as in Table 1. The mean of the expectation for African genealogical ancestors is obtained by averaging values of Eq. 10 across accepted parameter sets (Eq. 15 for genetic ancestors); the standard deviation of the expectation takes the standard deviation of those values. The mean of the standard deviation for African genealogical ancestors is obtained as the mean of Eq. 11 across accepted parameter sets (Eq. 16 for genetic ancestors). For the total, the mean of the expectation of the sum of the African genealogical ancestors is calculated by averaging values of Eq. 12 across accepted parameter sets (Eq. 17 for genetic ancestors); the standard deviation of the expectation takes the standard deviation of those values. The mean of the standard deviation for the total African genealogical ancestors is obtained as the mean of Eq. 13 across accepted parameter sets (Eq. 18 for genetic ancestors). Corresponding quantities for European ancestors are calculated by replacing each $s_{1,g-k}$ with $s_{2,g-k}$. The values of the total means for the expectation and standard deviation of African and European genealogical ancestors are those that appear in Table 3 of Mooney *et al.* (2023). The table shows the generation-wise values plotted in Figure 7 for the means and standard deviations of the expectation across the accepted parameter sets.

ancestry estimates, seeking to describe the timing at which the ancestors entered pedigrees of individuals and to count genetic ancestors across the length of the admixture process.

The number of genetic ancestors is bounded above by the number of genealogical ancestors, as each genetic ancestor must also be a genealogical ancestor. Both for genealogical and for genetic ancestors, the number of ancestors in a given generation is binomially distributed based on the number of genealogical ancestors in the subsequent generation (Eqs. 3, 10, 15, 20). The difference between the distributions of genealogical and genetic ancestors is in the binomial probability of success. For genealogical ancestors, the distribution depends only on parameters of the admixture process (Eqs. 3, 10), whereas for genetic ancestors, it depends also on a genetic ancestry probability for a genealogical ancestor separated from a descendant by a specified number of generations (Eqs. 15, 20). Depending on the features of the admixture process, the number of genetic ancestors from a source population can be close to the number of genealogical ancestors, or far smaller (Figure 4).

The evaluation of genetic ancestors extends the mechanistic admixture of model of Mooney *et al.* (2023). From a mathematical perspective, the focus on genealogical ancestors by Mooney *et al.* (2023) proceeded by adding a well-placed factor of 2 to the work of Verdu and Rosenberg (2011), converting a genomic fraction in a single-generation recursion into a genealogical ancestor count. The mathematical extension here is substantial, incorporating into the admixture model not only the factor of 2 but also the time-varying probability that a genealogical ancestor is a genetic ancestor.

Viewed from the perspective of the recombination-based genetic ancestry model of Coop (2013), our approach extends the analysis of genetic ancestors by separating them across source populations. If we were to follow Coop (2013) and consider all populations together as one, then Eq. 3 would reduce to $\mathbb{E}[X_k] = 2^k$, and our count of the random number of genetic ancestors in generation k would reduce Eq. 20 to $X_k^* \sim \text{Bin}(2^k, p_k)$. In other words, with no ancestry proportion considered—or alternatively, with all genealogical ancestors treated as members of the admixed population—the number of genealogical ancestors in generation k is 2^k, and the probability that a genealogical ancestor is tabulated as a genetic ancestor depends only on the genetic ancestry probability p_k. The expectation of this random variable gives the Coop (2013) calculation of the expected number of genetic ancestors in generation k, $\mathbb{E}[X_k^*] = 2^k p_k$ (Eq. 1, Figure 3).

African-American demographic history

With the Mooney *et al.* (2023) 14-generation model of African-American demographic history, we examined the expected numbers of genetic ancestors from Africa, Europe, and the African-American population itself, for random African-Americans born 1960-1965. We found for the mean numbers of genetic ancestors 162 Africans and 32 Europeans (Figure 7, Table 3), smaller than the corresponding numbers of genealogical ancestors, 314 Africans and 51 Europeans (Mooney *et al.* 2023). Tabulating ancestors within the African-American population itself, the expected numbers of genealogical and genetic ancestors are 363 and 294, respectively (Figure 5B, Table 1).

The peak number of genealogical ancestors occurs in generation 4 for Africans (1710-1715), generation 5 for Europeans (1735-1740), and generation 6 for African-Americans (1760-1765, Tables 4 and 1). Tracing genealogical ancestors back in time, noting

that the total number of genealogical ancestors doubles in each generation, we find that the proportion of African-Americans among genealogical ancestors is greatest in generation 13, decreasing back in time (Figure 6A, Table 2). The highest proportion occurs for Africans in generation 5 and for Europeans in generation 6. Eventually, African and European genealogical ancestors are reached who are parents solely of Africans or of Europeans; the proportions of these Africans and Europeans increase back in time until all genealogical ancestors are in these categories, in an approximate ratio of 79% Africans to 21% Europeans (Table 2). These quantities, which estimate fractions of all genealogical ancestors tracing to Africans and Europeans, lie in the range of permissible mean empirical genomic ancestry coefficients (Mooney *et al.* 2023).

For genetic ancestors, the contribution to African genetic ancestry is greatest for generations 4 and 5; the European genetic ancestry is highest in generations 5 and 6 (Figure 6B). The peak number of genetic ancestors occurs in generation 5 for Europeans and generation 6 for African-Americans, matching corresponding peaks for genealogical ancestors (Tables 1 and 4). However, the peak for African genetic ancestors occurs in generation 5, one generation later than for African genealogical ancestors (Table 4). Many African genealogical ancestors are far enough back in time that many of them are not genetic ancestors—so that the peak for genetic ancestors occurs later for genealogical ancestors. The fact that African genealogical ancestors occur on average farther in the past than European genealogical ancestors means that the 314:51 ratio of the mean numbers of African and European genealogical ancestors is smaller than the 162:32 ratio of the mean numbers of African and European genetic ancestors (Figure 8), as a larger fraction of the African genealogical ancestors have been lost as genetic ancestors. In effect, the fact that the European genealogical ancestors are later on average than the African genealogical ancestors has the result that the probability that a European genealogical ancestor is also a genetic ancestor exceeds the corresponding probability for Africans.

An interesting difference occurs between the peak of the African ancestor counts and the subsequent peak of the Transatlantic Slave Trade. The fraction of Africans transported by 1760 is about half of the total (Hacker 2020, Table 1); however, the comparable fraction of African genealogical ancestors, individuals born in generation 5 (born 1735-1740, reproductive age at 1760) or earlier, is 92% (Table 4). Hence, although the many transported Africans born in generations 6 and 7 certainly contributed in great numbers to the African-American population, a typical pedigree likely contains multiple lines that trace to the earlier enslaved migrants of generations 5 and earlier. In other words, by the time of the birth of generations 6 (1760-1765) and 7 (1785-1790), the African-American population was large enough that among all genealogical lines of a person born 1960-1965, many trace to genealogical ancestors who were already resident in the African-American population at the time of those generations. Indeed, for generation 6 onward and even for generation 5, African-Americans are a nontrivial fraction of the genealogical ancestors of a modern person (Figure 6), from ∼38% in generation 6 up to ∼90% in generation 13 (Table 2). The other major component in generation 6 onward is African genealogical ancestors who did not contribute directly to the African-American population. These Africans are the genealogical ancestors of Africans newly contributing to the African-American population. The substantial fraction for this category results from the accumulation of many African genealogical ancestors who con-

tributed to pedigrees in generations later than generation 5.

Limitations and extensions

As our approach follows the assumptions of Mooney *et al.* (2023), it is subject to many of the same limitations. For example, we do not consider a Native American component of admixture in African-Americans. Our treatment of a "random African American" born in the 1960-1965 window does not take into consideration regional variation across the African-American population in admixture processes or other demographic phenomena. We also disregard the possibility that the same genealogical ancestor might occur in multiple positions in a pedigree, so that our count of the number of ancestors might double-count some individuals; the time over which this assumption is sensible is the period in wnich the number of genealogical ancestors in a pedigree is small in relation to the pool of potential ancestors. Our discretization of the generations oversimplifies the demographic history, as does our three-epoch model, though this model does accord with the perspective of one of the most comprehensive empirical analyses of African-American genetic admixture (Baharian *et al.* 2016). Another limitation is that our model in principle allows an unlikely scenario in which the two parents of an African-American are two Europeans. We also do not consider distinct ancestry parameters for males and females. Each of these limitations is shared between the assessment of genealogical ancestors by Mooney *et al.* (2023) and our analysis of genetic ancestors here. As is discussed by Mooney *et al.* (2023, p. 13), each is possible to address by extensions and modifications of the model, potentially leading to further understanding of both genealogical and genetic ancestors.

Additional limitations not shared in the work of Mooney *et al.* (2023), which focused solely on genealogical ancestors, arise from the use of the Coop (2013) model to evaluate the probability that a genealogical ancestor is a genetic ancestor. This approach does not account for recombination phenomena such as recombination-rate variation across the genome, gene conversion, the particular sizes of chromosomes, crossover interference that perturbs the Poisson distribution assumed for the number of new genomic segments each generation, differing male and female recombination rates, or the X chromosome. With its simple treatment of the recombination process, the Coop (2013) model ignores many complexities that affect the probability that some segment from a genealogical ancestor might be retained in a descendant. Although extensions to accommodate such phenomena could be developed, in a single simple equation (Eq. 1), the Coop (2013) recombination model does capture the basic phenomenon—as explained by Donnelly (1983)—that as the time between ancestor and descendant increases, the probability that the descendant retains a segment from the ancestor decreases (Figure 3), and a steep drop in probability occurs when the separation increases from 7-8 generations (Robert Burns and descendants born 1960-1965) to 15-16 generations (descendants of William Shakespeare).

Our empirical focus has been on an example from human populations, but the model can be applied more generally to diploid species in which mechanistic admixture models and recombination models can be specified. To take one example, Armstrong *et al.* (2023) have studied genetic variation in captive tigers, a population formed through admixture of wild source populations from several different parts of Asia. Armstrong *et al.* (2023) have estimated genomic proportions that trace to the various source populations. With a generalization to permit more than two sources, our model can assist in understanding the properties of the genetic ancestors that have given rise to typical individual captive tigers.

Conclusions

Further study of a mechanistic admixture model has deepened the analysis of the number of genealogical ancestors who contribute from a source population to an admixed pedigree, and it has also introduced an approach to evaluating the number of contributing genetic ancestors. For African-Americans, the distinction between genealogical and genetic ancestors suggests that although the number of African genealogical ancestors in a pedigree greatly exceeds the number of European genealogical ancestors, because the African genealogical ancestors are on average earlier in time than the European genealogical ancestors, the number of African *genetic* ancestors does not exceed the number of European genetic ancestors by as great a margin. More generally, the calculations contribute to understanding the relationship between an admixed population's demographic history, its ancestral individuals who have given rise to the modern population, and the genomes of its current members.

Data availability

The 45,189 sets of accepted parameter values $(s_{1,0}, h_0, s_{2,0})$, $(s_{1,1}, h_1, s_{2,1}), \ldots, (s_{1,13}, h_{13}, s_{2,13})$ from Mooney *et al.* (2023), on which the analysis of the African-American population is based, are available in File S1.

Conflicts of interest

The authors declare no conflicts of interest. .

Literature cited

Armstrong EE, Mooney JA, Solari KA, Kim BY, Barsh GS, Grant V, Greenbaum G, Kaelin CB, Panchenko K, Pickrell JK *et al*. 2023. Unraveling the genomic diversity and evolutionary history of captive tigers in the United States. bioRxiv. p. https://doi.org/10.1101/2023.06.19.545608.

Baharian S, Barakatt M, Gignoux CR, Shringarpure S, Errington J, Blot WJ, Bustamante CD, Kenny EE, Williams SM, Aldrich MC *et al*. 2016. The Great Migration and African-American genomic diversity. PLoS Genet. 12:e1006059.

Baird SJE, Barton NH, Etheridge AM. 2003. The distribution of the surviving blocks of an ancestral genome. Theor. Pop. Biol. 64:451–471.

Buffalo V, Mount SM, Coop G. 2016. A genealogical look at shared ancestry on the X chromosome. Genetics. 204:57–75.

Chang JT. 1999. Recent common ancestors of all present-day individuals. Adv. Appl. Probab. 31:1002–1026.

Coop G. 2013. How many genetic ancestors do I have? https://gcbias.org/2013/11/11/how-does-your-number-of-genetic-ancestors-grow-back-over-time/.

Donnelly KP. 1983. The probability that related individuals share some section of genome identical by descent. Theor. Pop. Biol. 23:34–63.

Goldberg A, Rastogi A, Rosenberg NA. 2020. Assortative mating by population of origin in a mechanistic model of admixture. Theor. Pop. Biol. 134:129–146.

Goldberg A, Rosenberg NA. 2015. Beyond 2/3 and 1/3: the complex signatures of sex-biased admixture on the X chromosome. Genetics. 201:263–279.

Goldberg A, Verdu P, Rosenberg NA. 2014. Autosomal admixture levels are informative about sex bias in admixed populations. Genetics. 198:1209–1229.

Gravel S, Steel M. 2015. The existence and abundance of ghost ancestors in biparental populations. Theor. Pop. Biol. 101:47–53.

Hacker JD. 2020. From '20. and odd' to 10 million: the growth of the slave population of the United States. Slavery Abol. 41:840–855.

Kelleher J, Etheridge AM, Véber A, Barton NH. 2016. Spread of pedigree versus genetic ancestry in spatially distributed populations. Theor. Pop. Biol. 108:1–12.

Kim J, Edge MD, Goldberg A, Rosenberg NA. 2021. Skin deep: the decoupling of genetic admixture levels from phenotypes that differed between source populations. Am. J. Phys. Anthropol. 175:406–421.

Lawson DJ, van Dorp L, Falush D. 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nature Commun. 9:3258.

Mathieson I, Scally A. 2020. What is ancestry? PLoS Genet. 16:e1008624.

Matsen FA, Evans SN. 2008. To what extent does genealogical ancestry imply genetic ancestry? Theor. Pop. Biol. 74:182–190.

Milo R, Phillips R. 2015. *Cell biology by the numbers*. New York: Garland Science.

Mooney JA, Agranat-Tamir L, Pritchard JK, Rosenberg NA. 2023. On the number of genealogical ancestors tracing to the source groups of an admixed population. Genetics. 224:iyad079.

Rohde DLT, Olson S, Chang JT. 2004. Modelling the recent common ancestry of all living humans. Nature. 431:562–566.

Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philos. Trans. R. Soc. B. 372:20160455.

Verdu P, Rosenberg NA. 2011. A general mechanistic model for admixture histories of hybrid populations. Genetics. 189:1413–1426.

Weiss KM, Long JC. 2009. Non-Darwinian estimation: My ancestors, my genes' ancestors. Genome Res. 19:703–710.

Wiuf C, Hein J. 1997. On the number of ancestors to a DNA sequence. Genetics. 147:1459–1468.

Appendix A. Proofs of Eqs. 3, 4 and 6

We prove Eq. 3, describing $\mathbb{E}[X_k]$, by induction. For $k = 0$,

$$X_k = 1 = 2^0 \prod_{i=1}^{0} h_{g-i}.$$

We assume that for $k - 1$,

$$\mathbb{E}[X_{k-1}] = 2^{k-1} \prod_{i=1}^{k-1} h_{g-i}.$$

Using the inductive hypothesis and the fact that $X_k \sim \text{Bin}(2X_{k-1}, h_{g-k})$ for $1 \leq k \leq g$, we obtain

$$\mathbb{E}[X_k] = \mathbb{E}\left[\mathbb{E}[X_k \mid X_{k-1}]\right] = 2h_{g-k}\mathbb{E}[X_{k-1}]$$

$$= 2h_{g-k}2^{k-1}\prod_{i=1}^{k-1} h_{g-i} = 2^k \prod_{i=1}^{k} h_{g-i}.$$

Next, we prove Eq. 4, again by induction. For $k = 0$, $X_0 = 1$ has variance 0; Eq. 4 holds trivially, as it is an empty sum. For $k = 1$, $X_1 \sim \text{Bin}(2, h_{g-1})$, and therefore,

$$\begin{aligned}
\text{Var}[X_1] &= 2h_{g-1}(1 - h_{g-1}) \\
&= \sum_{i=1}^{1} 2^{1-1+i}[1 - h_{g-(1+1)+i}] \\
&\quad \times \left[\left(\prod_{j=g-(1+1)+i}^{g-1} h_j\right)\left(\prod_{\ell=g-1}^{g-(1+2)+i} h_\ell^2\right)\right].
\end{aligned}$$

We assume that for $k - 1$,

$$\begin{aligned}
\text{Var}[X_{k-1}] &= \sum_{i=1}^{k-1} 2^{k-2+i}(1 - h_{g-k+i}) \\
&\quad \times \left[\left(\prod_{j=g-k+i}^{g-1} h_j\right)\left(\prod_{\ell=g-(k-1)}^{g-(k+1)+i} h_\ell^2\right)\right].
\end{aligned}$$

We use the law of total variance with Eq. 3 and the inductive

hypothesis. We have

$$\begin{aligned}
\text{Var}[X_k] &= \mathbb{E}\left[\text{Var}[X_k \mid X_{k-1}]\right] + \text{Var}\left[\mathbb{E}[X_k \mid X_{k-1}]\right] \\
&= \mathbb{E}[2h_{g-k}(1 - h_{g-k})X_{k-1}] + \text{Var}[2h_{g-k}X_{k-1}] \\
&= 2h_{g-k}(1 - h_{g-k})2^{k-1}\left(\prod_{i=1}^{k-1} h_{g-i}\right) \\
&\quad + (2h_{g-k})^2 \sum_{i=1}^{k-1} 2^{k-2+i}(1 - h_{g-k+i}) \\
&\quad \times \left[\left(\prod_{j=g-k+i}^{g-1} h_j\right)\left(\prod_{\ell=g-(k-1)}^{g-(k+1)+i} h_\ell^2\right)\right] \\
&= 2^k(1 - h_{g-k})\left(\prod_{j=g-k}^{g-1} h_j\right) \\
&\quad + \sum_{i=2}^{k} 2^{2+k-2+i-1}(1 - h_{g-k+i-1}) \\
&\quad \times \left[\left(\prod_{j=g-k+i-1}^{g-1} h_j\right)\left(\prod_{\ell=g-k}^{g-(k+1)+i-1} h_\ell^2\right)\right] \\
&= \sum_{i=1}^{k} 2^{k-1+i}[1 - h_{g-(k+1)+i}] \\
&\quad \times \left[\left(\prod_{j=g-(k+1)+i}^{g-1} h_j\right)\left(\prod_{\ell=g-k}^{g-(k+2)+i} h_\ell^2\right)\right].
\end{aligned}$$

Finally, we prove Eq. 6. First, we prove that if $0 \leq m < n \leq g$, then

$$\text{Cov}[X_n, X_m] = 2^{n-m} \prod_{i=1}^{n-m} h_{g-(m+i)} \text{Var}[X_m].$$

Fixing m with $0 \leq m \leq g - 1$, we proceed by induction on n. For $n = m + 1$, we have

$$\begin{aligned}
\text{Cov}[X_{m+1}, X_m] &= \mathbb{E}[X_{m+1}X_m] - \mathbb{E}[X_{m+1}]\mathbb{E}[X_m] \\
&= \mathbb{E}\left[\mathbb{E}[X_{m+1}X_m | X_m]\right] \\
&\quad - 2h_{g-(m+1)}\mathbb{E}[X_m]\mathbb{E}[X_m] \\
&= \mathbb{E}[2h_{g-(m+1)}X_m^2] - 2h_{g-(m+1)}\mathbb{E}[X_m]\mathbb{E}[X_m] \\
&= 2^{m+1-m}h_{g-(m+1)}\text{Var}[X_m].
\end{aligned}$$

We now assume that for (n, m) with $0 \leq m < n \leq g$ and $n \geq m + 2$,

$$\text{Cov}[X_{n-1}X_m] = 2^{n-1-m} \prod_{i=1}^{n-1-m} h_{g-(m+i)} \text{Var}[X_m].$$

Then

$$\begin{aligned}
\text{Cov}[X_n, X_m] &= \mathbb{E}[X_nX_m] - \mathbb{E}[X_n]\mathbb{E}[X_m] \\
&= \mathbb{E}\left[\mathbb{E}[X_nX_m | X_m, X_{n-1}]\right] \\
&\quad - 2h_{g-n}\mathbb{E}[X_{n-1}]\mathbb{E}[X_m] \\
&= 2h_{g-n}\left(\mathbb{E}[X_{n-1}X_m] - \mathbb{E}[X_{n-1}]\mathbb{E}[X_m]\right) \\
&= 2h_{g-n}2^{n-1-m} \prod_{i=1}^{n-1-m} h_{g-(m+i)} \text{Var}[X_m] \\
&= 2^{n-m} \prod_{i=1}^{n-m} h_{g-(m+i)} \text{Var}[X_m].
\end{aligned}$$

Having obtained the covariance $\text{Cov}[X_n, X_m]$, we conclude

$$
\begin{aligned}
\text{Var}\left[\sum_{k=1}^{g} X_k\right] &= \sum_{k=1}^{g} \text{Var}[X_k] \\
&\quad + 2\sum_{m=1}^{g-1}\sum_{n=m+1}^{g} \text{Cov}[X_n, X_m] \\
&= \sum_{k=1}^{g} \text{Var}[X_k] \\
&\quad + \sum_{m=1}^{g-1}\sum_{n=m+1}^{g} 2^{n-m+1}\prod_{i=1}^{n-m} h_{g-(m+i)}\text{Var}[X_m].
\end{aligned}
$$

Appendix B. Proofs of Eqs. 13 and 18

We prove Eq. 13, starting with the law of total variance.

$$
\begin{aligned}
\text{Var}\left[\sum_{k=1}^{g} U_k\right] &\overset{(i)}{=} \text{Var}\left[\mathbb{E}\left[\sum_{k=1}^{g} U_k | X_0, X_1, \ldots, X_k\right]\right] \\
&\quad + \mathbb{E}\left[\text{Var}\left[\sum_{k=1}^{g} U_k | X_0, X_1, \ldots, X_k\right]\right] \\
&\overset{(ii)}{=} \text{Var}\left[\sum_{k=1}^{g} \tilde{s}_{1,g-k}(2X_{k-1} - X_k)\right] \\
&\quad + \mathbb{E}\left[\sum_{k=1}^{g} \tilde{s}_{1,g-k}(1 - \tilde{s}_{1,g-k})(2X_{k-1} - X_k)\right] \\
&\overset{(iii)}{=} \text{Var}\left[\sum_{k=1}^{g-1} [2\tilde{s}_{1,g-(k+1)} - \tilde{s}_{1,g-k}]X_k\right] \\
&\quad + \sum_{k=1}^{g} \tilde{s}_{1,g-k}(1 - \tilde{s}_{1,g-k})\,\mathbb{E}\left[\mathbb{E}[2X_{k-1} - X_k | X_{k-1}]\right] \\
&\overset{(iv)}{=} \sum_{k=1}^{g-1} [2\tilde{s}_{1,g-(k+1)} - \tilde{s}_{1,g-k}]^2\,\text{Var}[X_k] \\
&\quad + \sum_{m=1}^{g-2}\sum_{n=m+1}^{g-1} 2^{n-m+1}[2\tilde{s}_{1,g-(m+1)} - \tilde{s}_{1,g-m}] \\
&\quad\quad \times [2\tilde{s}_{1,g-(n+1)} - \tilde{s}_{1,g-n}]\prod_{i=1}^{n-m} h_{g-(m+i)}\,\text{Var}[X_m] \\
&\quad + \sum_{k=1}^{g} \tilde{s}_{1,g-k}(1 - \tilde{s}_{1,g-k})\,(1 - h_{g-k})\,(2\mathbb{E}[X_{k-1}]) \\
&\overset{(v)}{=} \sum_{k=1}^{g-1} [2\tilde{s}_{1,g-(k+1)} - \tilde{s}_{1,g-k}]^2\,\text{Var}[X_k] \\
&\quad + \sum_{m=1}^{g-2}\sum_{n=m+1}^{g-1} 2^{n-m+1}[2\tilde{s}_{1,g-(m+1)} - \tilde{s}_{1,g-m}] \\
&\quad\quad \times [2\tilde{s}_{1,g-(n+1)} - \tilde{s}_{1,g-n}]\prod_{i=1}^{n-m} h_{g-(m+i)}\,\text{Var}[X_m] \\
&\quad + \sum_{k=0}^{g-1} 2s_{1,g-(k+1)}[1 - \tilde{s}_{1,g-(k+1)}]\,\mathbb{E}[X_k].
\end{aligned}
$$

For line (ii), given $X_0, \ldots, X_{k-1}, X_k, U_k$ depends only on X_{k-1} and X_k. Among the genealogical ancestors in step k of the descendant from step 0, $2X_{k-1}$ are parents of admixed individuals from step $k-1$, and X_k are admixed individuals in step k; $2X_{k-1} - X_k$ reach a source population in step k, with binomial probabilities $s_{1,g-k}/(s_{1,g-k} + s_{2,g-k}) = s_{1,g-k}/(1 - $

$h_{g-k}) = \tilde{s}_{1,g-k}$ for source 1 and $s_{2,g-k}/(s_{1,g-k} + s_{2,g-k}) = s_{2,g-k}/(1 - h_{g-k}) = \tilde{s}_{2,g-k}$ for source 2, respectively. In other words, $U_k | X_{k-1}, X_k \sim \text{Bin}(2X_{k-1} - X_k, \tilde{s}_{1,g-k})$.

For line (iii), in the sum $\sum_{k=1}^{g} \tilde{s}_{1,g-k}(2X_{k-1} - X_k)$, for $k = 1, 2, \ldots, g-1$, $X_0 = 1$ and $X_g = 0$ are constants and have zero variance. We also use the law of total expectation. Line (iv) follows from Eq. 6 and from the binomial distribution of $X_k | X_{k-1}$, so that $\mathbb{E}[2X_{k-1} - X_k | X_{k-1}] = 2\mathbb{E}[X_{k-1}] - 2h_{g-k}\mathbb{E}[X_{k-1}] = (1 - h_{g-k})(2\mathbb{E}[X_{k-1}])$. Finally, for (v), we simplify $\tilde{s}_{1,g-k}(1 - h_{g-k}) = s_{1,g-k}$.

Similarly, we also use the law of total variance to prove Eq. 18:

$$
\begin{aligned}
\text{Var}\left[\sum_{k=1}^{g} Y_k\right] &= \text{Var}\left[\mathbb{E}\left[\sum_{k=1}^{g} Y_k | X_0, X_1, \ldots, X_k\right]\right] \\
&\quad + \mathbb{E}\left[\text{Var}\left[\sum_{k=1}^{g} Y_k | X_0, X_1, \ldots, X_k\right]\right].
\end{aligned}
$$

The proof is entirely analogous, except that $\tilde{s}_{1,g-k}p_k$ appears in place of $\tilde{s}_{1,g-k}$.

Appendix C. Proof of Eq. 35

We prove inequalities concerning p_k/p_{k-1}: (i) $p_k/p_{k-1} < 1$ for $k \geq 2$; (ii) $p_k/p_{k-1} > \frac{1}{2}$ for $k \geq 2$.

(i) By Eq. 1, $p_k/p_{k-1} = [1 - e^{-a(k)}]/[1 - e^{-b(k)}]$ for $k \geq 3$, where $a(k) = (33k - 11)/2^{k-1}$ and $b(k) = (33k - 44)/2^{k-2}$. For $k \geq 3$, $0 < a(k) < b(k)$, and hence, $1 - e^{-a(k)} < 1 - e^{-b(k)}$ and $p_k/p_{k-1} < 1$. For $k = 2$, $p_k/p_{k-1} < 1$ as $p_k < 1$ by Eq. 1 and $p_{k-1} = 1$.

(ii) For $k = 2$, $p_k/p_{k-1} = p_2 = 1 - e^{-55/2} > \frac{1}{2}$. For $k \geq 3$, we rearrange Eq. 1 to find that the inequality $p_k/p_{k-1} > \frac{1}{2}$ is equivalent to

$$
e^{\frac{66}{2^{k-1}}} e^{\frac{33k-77}{2^{k-1}}} + e^{-\frac{33k-77}{2^{k-1}}} > 2. \tag{37}
$$

The inequality $e^x + e^{-x} \geq 2$ holds for all x, as it is equivalent to $\cosh x \geq 1$. Hence, for $c > 1$, $ce^x + e^{-x} > e^x + e^{-x} \geq 2$. We see that Eq. 37 then follows, with

$$
\left(e^{\frac{66}{2^{k-1}}}, \frac{33k - 77}{2^{k-1}}\right)
$$

in place of (c, x). As Eq. 37 holds, we conclude $p_k/p_{k-1} > \frac{1}{2}$.