OXFORD GENETICS

GENETICS, 2023, 224(3), iyad079
https://doi.org/10.1093/genetics/iyad079
Investigation

On the number of genealogical ancestors tracing to the source groups of an admixed population

Jazlyn A. Mooney ,[1,2] Lily Agranat-Tamir,[1] Jonathan K. Pritchard,[1,3] Noah A. Rosenberg [1,*]

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

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

Abstract

Members of genetically admixed populations possess ancestry from multiple source groups, and studies of human genetic admixture frequently estimate ancestry components corresponding to fractions of individual genomes that trace to specific ancestral populations. However, the same numerical ancestry fraction can represent a wide array of admixture scenarios within an individual's genealogy. Using a mechanistic model of admixture, we consider admixture genealogically: how many *ancestors from the source populations* does the admixture represent? We consider African-Americans, for whom continent-level estimates produce a 75–85% value for African ancestry on average and 15–25% for European ancestry. Genetic studies together with key features of African-American demographic history suggest ranges for parameters of a simple three-epoch model. Considering parameter sets compatible with estimates of current ancestry levels, we infer that if all genealogical lines of a random African-American born during 1960–1965 are traced back until they reach members of source populations, the mean over parameter sets of the expected number of genealogical lines terminating with African individuals is 314 (interquartile range 240–376), and the mean of the expected number terminating in Europeans is 51 (interquartile range 32–69). Across discrete generations, the peak number of African genealogical ancestors occurs in birth cohorts from the early 1700s, and the probability exceeds 50% that at least one European ancestor was born more recently than 1835. Our genealogical perspective can contribute to further understanding the admixture processes that underlie admixed populations. For African-Americans, the results provide insight both on how many of the ancestors of a typical African-American might have been forcibly displaced in the Transatlantic Slave Trade and on how many separate European admixture events might exist in a typical African-American genealogy.

Keywords: admixture, ancestry, genealogy, population genetics

Introduction

Genetically admixed populations arise when two or more source groups combine to form a new population. After generations of mating among members of the incipient admixed population and new contributors from the source groups, typical individuals in the admixed group possess ancestry from multiple sources (Chakraborty 1986; Korunes and Goldberg 2021; Gopalan *et al.* 2022).

The genetic history of an admixed population can be represented by a temporal sequence of admixture contributions, starting with the founding of the new admixed group (Long 1991; Verdu and Rosenberg 2011; Gravel 2012). Among present-day members of the admixed population, genetic patterns such as the distribution of admixture levels estimated from individual genomes can then be used together with a model of the admixture process to uncover features such as the timing and magnitude of the genetic contributions that characterize the admixture (Verdu *et al.* 2014; Baharian *et al.* 2016; Zaitlen *et al.* 2017).

In studies that seek to infer population parameters from genetic patterns among individuals in the admixed population, each admixed individual is treated as a random outcome of the admixture process. The accumulation of data on many admixed individuals then provides information about the population history. In this perspective, for a given model of the admixture history, an individual possesses a random genealogy conditional on the parameters of the admixture process—a random pedigree. What information can be obtained about a random individual genealogy under the assumptions of an admixture model? In particular, for individual members of an admixed population, how many contributors from the source populations does their admixture represent?

Consider the example in Fig. 1, involving admixture of two source populations to form a third, admixed population. Tracing the genealogy of a member of the admixed population back in time on each genealogical line until the most recent member of a source population is reached, the example genealogy has six ancestors from source 1 (a grandfather, two great-grandmothers, a great-great-grandfather, and two great-great-grandmothers) and four from source 2 (a great-grandfather, two great-great-grandfathers, and a great-great-grandmother). Counts of the numbers of ancestors from source populations in a random individual genealogy depend both on the relative contributions of the source populations to the admixed group and on the timing of the admixture.

Received: January 24, 2023. **Accepted:** April 5, 2023

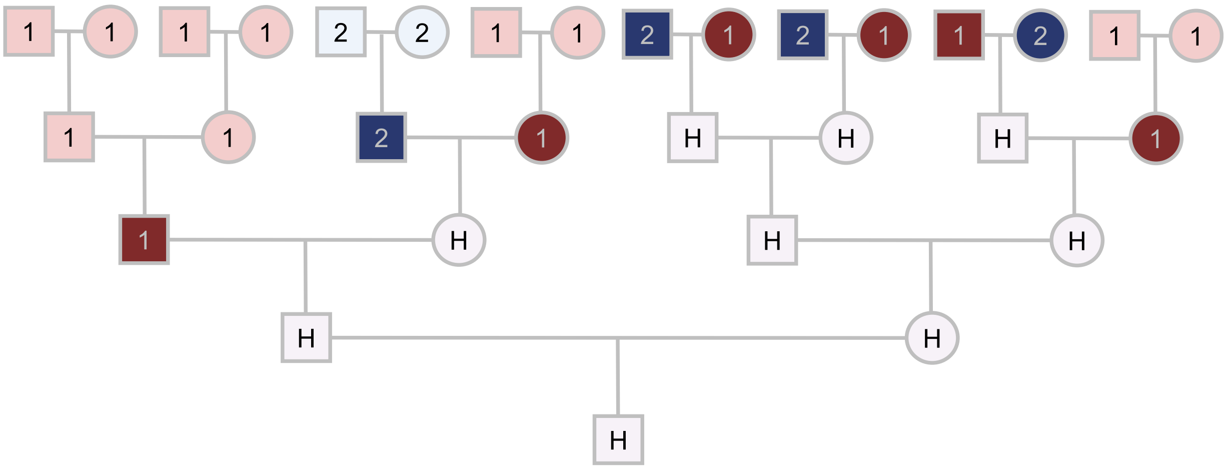

Fig. 1. Counting genealogical ancestors. The pedigree of the individual at the bottom of the diagram is traced back in time until ancestral populations are reached. Each individual in the pedigree is labeled by the population to which it belongs: source population 1 (light and dark red), source population 2 (light and dark blue), or admixed population *H* (purple). For the index individual, this pedigree shows six ancestors from source 1 and four from source 2. The count of genealogical ancestors from the source populations tabulates, along each ancestral line, the *first* individual reached who belongs to a source population: the six individuals from source 1 shown in dark red and the four individuals from source 2 shown in dark blue. The ancestry fractions for the individual are $\frac{11}{16}$ from source 1 and $\frac{5}{16}$ from source 2.

In human admixed populations, questions focused on random genealogies can provide information both about the population-level history of admixture and about the relationship of individuals to that history. Consider the case of the African-American admixed population in the United States. Living African-Americans descend primarily from an admixture of African and European source populations, much of the admixture having occurred during the period of enslavement of most African-Americans, 1619–1865. Owing to widespread patterns such as forcible fracturing of enslaved families by enslavers, lack of documentation of many of the enslaved even by first name in the written record, and a reticence of many formerly enslaved individuals to record genealogical information in the period after slavery, for many African-Americans, limited data are available about their individual ancestors prior to the middle or late 1800s (Gates 2009; Swarns 2012; Nelson 2016). An admixture model thus has potential to recover features of African-American genealogies that are otherwise difficult to obtain.

For an African-American chosen at random, considering genealogies in the last ~400 years, how many genealogical lines traced back from the present to a member of a source population reach an African individual? How many reach a European or European-American? The former quantity approximates the number of ancestors who traveled from Africa to the Western Hemisphere as forced enslaved migrants in the Transatlantic Slave Trade. The latter gives the number of occasions at which European admixture events occurred in a random African-American genealogy. Answers to such questions are informative not only for understanding the genealogies of individuals but also for contributing details of the admixture process that has given rise to the present-day population.

Model

Assumptions

We follow a mechanistic model in which admixture levels are explored in an admixed population over time (Verdu and Rosenberg 2011; Goldberg *et al.* 2014; Goldberg and Rosenberg 2015; Goldberg *et al.* 2020). Three populations are considered: source populations S_1 and S_2, and admixed population *H*. In each of a series of generations—indexed discretely with the index increasing forward in time—an individual in the admixed population *H* in generation *g* has a pair of parents probabilistically drawn from among individuals extant in generation $g-1$ in source populations S_1 and S_2 and admixed population *H* (Fig. 2).

Suppose that for an individual in generation *g*, the admixture contributions are $s_{1,g-1}$, $s_{2,g-1}$, and h_{g-1}, for populations S_1, S_2, and *H*, respectively. In other words, for an individual chosen at random in admixed population *H*, a parent chosen at random has probability $s_{1,g-1}$ of having originated from population S_1, $s_{2,g-1}$ for population S_2, and h_{g-1} for population *H*. We then have

$$s_{1,g-1} + h_{g-1} + s_{2,g-1} = 1. \tag{1}$$

The sampling probabilities $s_{1,g-1}$, $s_{2,g-1}$, and h_{g-1} can be interpreted as fractional contributions from source populations S_1, S_2, and *H* to autosomal genomes in population *H* in generation *g*. Generation $g=1$ represents the founding of the admixed population from members of the source population from generation $g=0$. The admixed population does not exist in generation $g=0$, so that $h_0=0$, and $s_{1,0}+s_{2,0}=1$.

Previous studies with these modeling assumptions have tracked properties of random variables that describe *ancestry proportions* in the source populations S_1 and S_2 at generation *g*. In particular, Verdu and Rosenberg (2011) studied recursions for the probability distribution and moments of a random variable $H_{1,g}$, representing the autosomal fraction of ancestry from source population 1 for an individual in the admixed population at generation *g*. We instead study the random variable $Z_{1,g}$, the *number of genealogical ancestors* from source population 1 for an individual in the admixed population at generation *g*, and $Z_{2,g}$, the number of genealogical ancestors from source population 2. In the sense in which we consider genealogical ancestors, once a source population is reached along a genealogical line in a specific ancestor, that ancestor is tabulated as a genealogical ancestor from the associated source population, and the line is not traced any farther back (Fig. 1).

Broad features of the numbers of genealogical ancestors $Z_{1,g}$ and $Z_{2,g}$ of an admixed individual from generation *g* can be

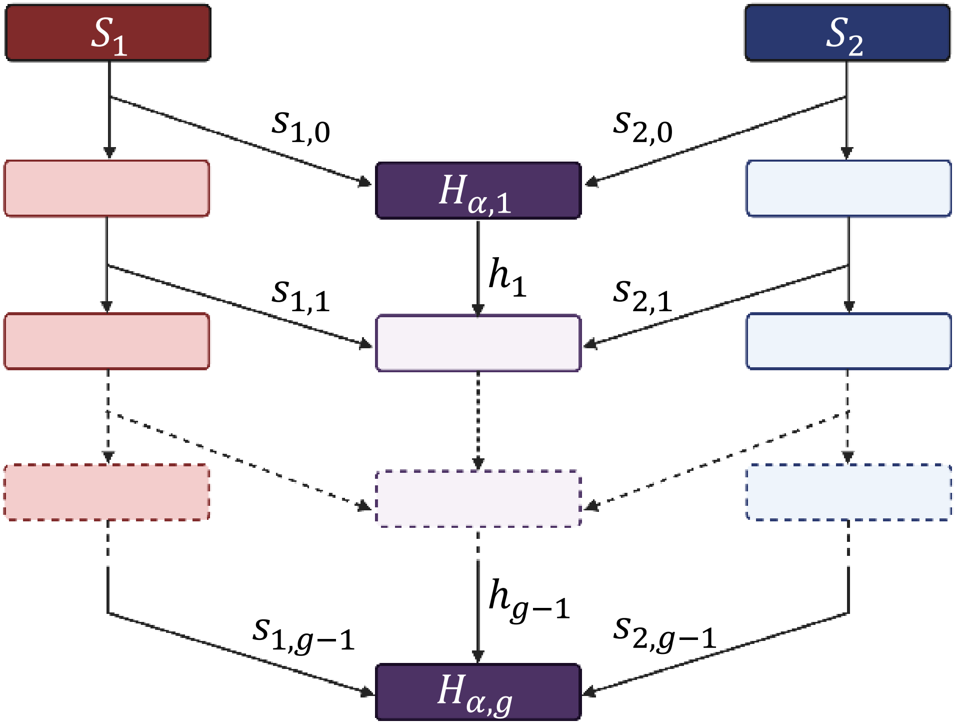

Fig. 2. Schematic of the admixture model. Source populations S_1 and S_2 contribute to an admixed population H. The members of H in generation g draw parents from the populations of generation $g-1$ from S_1 with probability $s_{1,g-1}$, from H with probability h_{g-1}, and from S_2 with probability $s_{2,g-1}$. Two parents are drawn independently. Random variable $H_{\alpha,g}$ denotes the random autosomal ancestry fraction from population α (1 for S_1, 2 for S_2) in an individual in population H in generation g.

understood in relation to admixture parameters $s_{1,0}, s_{1,1}, \ldots, s_{1,g-1}$ and $s_{2,0}, s_{2,1}, \ldots, s_{2,g-1}$. If per-generation genetic contributions from the source populations are large, then genealogical lines are likely to reach the sources in the most recent few generations. In the limiting case that $s_{1,g-1} + s_{2,g-1} = 1$ and all parents are from the source populations, a random individual has two parents from the source populations, and $Z_{1,g} + Z_{2,g} = 2$. If, however, the genetic contributions from the admixed population to itself predominate, then most genealogical lines reach the sources only many generations in the past. In the limit in which admixture occurred only in the initial generation, or $s_{1,0} + s_{2,0} = 1$ and $s_{1,i} + s_{2,i} = 0$ for each $i > 0$, then the source populations are reached only g generations in the past, when an individual has 2^g genealogical ancestors, and we have $Z_{1,g} + Z_{2,g} = 2^g$. Considering different admixture scenarios, the number of genealogical ancestors from source populations, $Z_{1,g} + Z_{2,g}$, is bounded between these extremes of 2 and 2^g.

Recursion for the number of genealogical ancestors

We review expressions that we will need for the mean and variance of autosomal admixture under the model (Verdu and Rosenberg 2011). The mean ancestry fraction from population 1 in generation g is (Verdu and Rosenberg 2011, equations (10) and (11)):

$$\mathbb{E}[H_{1,g}] = \begin{cases} s_{1,0}, & g = 1, \\ s_{1,g-1} + h_{g-1}\mathbb{E}[H_{1,g-1}], & g \geq 2. \end{cases} \tag{2}$$

The variance of the ancestry fraction from population 1 in generation g is (Verdu and Rosenberg 2011, equations (22) and (23))

$$\mathbb{V}[H_{1,g}] \\ = \begin{cases} \frac{s_{1,0}(1-s_{1,0})}{2}, & g = 1, \\ \frac{s_{1,g-1}(1-s_{1,g-1})}{2} - s_{1,g-1}h_{g-1}\mathbb{E}[H_{1,g-1}] + \frac{h_{g-1}(1-h_{g-1})}{2}\mathbb{E}[H_{1,g-1}]^2 \\ \qquad + \frac{h_{g-1}}{2}\mathbb{V}[H_{1,g-1}], & g \geq 2. \end{cases} \tag{3}$$

Note that the mean ancestry fraction from population 2 is one minus the mean ancestry fraction from population 1, and the variances of the two ancestry fractions are equal.

A recursion for the ancestry fraction $H_{1,g}$ (Verdu and Rosenberg 2011) can be modified to obtain a recursion for $Z_{1,g}$. Whereas the random autosomal ancestry fraction $H_{1,g}$ of an individual is the mean of the corresponding ancestry fractions of the parents of the individual, the random number of ancestors $Z_{1,g}$ is the *sum* of the numbers of ancestors of the parents (from population 1).

Let L be a random variable that gives the source populations of the parents of a random individual from the admixed population. Listing the mother first, L takes a value in the set $\mathcal{L} = \{S_1S_1, S_1H, S_1S_2, HS_1, HH, HS_2, S_2S_1, S_2H, S_2S_2\}$. Based on equations (1) and (2) of Verdu and Rosenberg (2011), for generation $g = 1$, we have

$$Z_{1,1} = \begin{cases} 2 & \text{if } L = S_1S_1, \text{ with } \mathbb{P}[L = S_1S_1] = s_{1,0}s_{1,0}, \\ 1 & \text{if } L = S_1S_2, \text{ with } \mathbb{P}[L = S_1S_2] = s_{1,0}s_{2,0}, \\ 1 & \text{if } L = S_2S_1, \text{ with } \mathbb{P}[L = S_2S_1] = s_{2,0}s_{1,0}, \\ 0 & \text{if } L = S_2S_2, \text{ with } \mathbb{P}[L = S_2S_2] = s_{2,0}s_{2,0}. \end{cases} \tag{4}$$

For subsequent generations, $g \geq 2$,

$$
\begin{aligned}
&Z_{1,g}\\
&= \begin{cases}
2 & \text{if } L = S_1S_1, \text{ with } \mathbb{P}[L = S_1S_1] = s_{1,g-1}s_{1,g-1},\\
1 + Z_{1,g-1} & \text{if } L = S_1H, \text{ with } \mathbb{P}[L = S_1H] = s_{1,g-1}h_{g-1},\\
1 & \text{if } L = S_1S_2, \text{ with } \mathbb{P}[L = S_1S_2] = s_{1,g-1}s_{2,g-1},\\
Z_{1,g-1} + 1 & \text{if } L = HS_1, \text{ with } \mathbb{P}[L = HS_1] = h_{g-1}s_{1,g-1},\\
Z_{1,g-1} + Z'_{1,g-1} & \text{if } L = HH, \text{ with } \mathbb{P}[L = HH] = h_{g-1}h_{g-1},\\
Z_{1,g-1} & \text{if } L = HS_2, \text{ with } \mathbb{P}[L = HS_2] = h_{g-1}s_{2,g-1},\\
1 & \text{if } L = S_2S_1, \text{ with } \mathbb{P}[L = S_2S_1] = s_{2,g-1}s_{1,g-1},\\
Z_{1,g-1} & \text{if } L = S_2H, \text{ with } \mathbb{P}[L = S_2H] = s_{2,g-1}h_{g-1},\\
0 & \text{if } L = S_2S_2, \text{ with } \mathbb{P}[L = S_2S_2] = s_{2,g-1}s_{2,g-1}.
\end{cases}
\end{aligned}
\tag{5}
$$

For $L = HH$, $Z_{1,g-1}$ and $Z'_{1,g-1}$ are independent and identically distributed copies of the same random variable.

Equations (4) and (5) enable us to compute the probability distribution of $Z_{1,g}$, the number of genealogical ancestors from population 1 for an individual in the admixed population in generation g. $Z_{1,g}$ and $Z_{2,g}$ range in $Q_g = \{0, 1, \ldots, 2^g\}$. For q in Q_g, we compute the probability $\mathbb{P}[Z_{1,g} = q]$ that a random individual from population H at generation g has q genealogical ancestors from population 1. Analogously to equations (3)–(5) of Verdu and Rosenberg (2011), we have for $g \geq 1$

$$
\mathbb{P}[Z_{1,1} = q] = \begin{cases}
s_{1,0}^2, & q = 2,\\
2s_{1,0}s_{2,0}, & q = 1,\\
s_{2,0}^2, & q = 0.
\end{cases}
\tag{6}
$$

For $g \geq 2$ and q in Q_g,

$$
\begin{aligned}
\mathbb{P}[Z_{1,g} = q] = h_{g-1}^2 \sum_{r=0}^{2^{g-1}} (\mathbb{P}[Z_{1,g-1} = r]\mathbb{P}[Z_{1,g-1} = q - r])\\
+ (2s_{1,g-1}h_{g-1})\mathbb{P}[Z_{1,g-1} = q - 1]\\
+ (2s_{2,g-1}h_{g-1})\mathbb{P}[Z_{1,g-1} = q] + I_g(q).
\end{aligned}
\tag{7}
$$

Function I_g is equal to

$$
I_g(q) = \begin{cases}
s_{1,g-1}^2, & q = 2,\\
2s_{1,g-1}s_{2,g-1}, & q = 1,\\
s_{2,g-1}^2, & q = 0,\\
0, & 3 \leq q \leq 2^q.
\end{cases}
\tag{8}
$$

Equation (7) sums over all possible parental pairings that lead to q ancestors from population 1 at generation g. Only three values of q are possible if neither parent is from the admixed population—$q = 0$, $q = 1$, and $q = 2$—producing the terms in equation (8).

Recursive mean and variance of the number of genealogical ancestors

Using the recursion for the probability distribution of the number of ancestors in equations (4) and (5), we follow Verdu and Rosenberg (2011) to obtain moments of $Z_{1,g}$. By the law of conditional expectation,

$$
\mathbb{E}[Z_{1,g}] = \mathbb{E}_L[\mathbb{E}[Z_{1,g}|L]] = \sum_{\ell \in \mathcal{L}} \mathbb{P}[L = \ell]\, \mathbb{E}[Z_{1,g}|L = \ell].
\tag{9}
$$

For each $\ell \in \mathcal{L}$, $\mathbb{E}[Z_{1,g}|L = \ell] = 2\mathbb{E}[H_{1,g}|L = \ell]$, so that the recursive computation of $\mathbb{E}[Z_{1,g}]$ follows that of $\mathbb{E}[H_{1,g}]$ in equations (6)–(11) of Verdu and Rosenberg (2011), multiplying by a factor of 2. We obtain $\mathbb{E}[Z_{1,g}] = 2\mathbb{E}[H_{1,g}]$, or

$$
\mathbb{E}[Z_{1,g}] = \begin{cases}
2s_{1,0}, & g = 1,\\
2s_{1,g-1} + 2h_{g-1}\mathbb{E}[Z_{1,g-1}], & g \geq 2.
\end{cases}
\tag{10}
$$

For the kth moment of $Z_{1,g}$, for each ℓ, $\mathbb{E}[Z_{1,g}^k \,|\, L = \ell] = 2^k\mathbb{E}[H_{1,g}^k \,|\, L = \ell]$. In particular, as $\mathbb{E}[Z_{1,g}^2 \,|\, L = \ell] = 4\mathbb{E}[H_{1,g}^2 \,|\, L = \ell]$, we obtain $\mathbb{E}[Z_{1,g}^2] = 4\mathbb{E}[H_{1,g}^2]$. Because $\mathbb{E}[Z_{1,g}]^2 = 4\mathbb{E}[H_{1,g}]^2$ and $\mathbb{E}[Z_{1,g}^2] = 4\mathbb{E}[H_{1,g}^2]$, we have $\mathbb{V}[Z_{1,g}] = 4\mathbb{V}[H_{1,g}]$. We apply equations (22) and (23) of Verdu and Rosenberg (2011) for $\mathbb{V}[H_{1,g}]$, obtaining

$$
\mathbb{V}[Z_{1,g}] = \begin{cases}
2s_{1,0}(1 - s_{1,0}), & g = 1,\\
2s_{1,g-1}(1 - s_{1,g-1}) - 4s_{1,g-1}h_{g-1}\mathbb{E}[Z_{1,g-1}] & \\
\quad +2h_{g-1}(1 - h_{g-1})\mathbb{E}[Z_{1,g-1}]^2 & \\
\quad +2h_{g-1}\mathbb{V}[Z_{1,g-1}], & g \geq 2.
\end{cases}
\tag{11}
$$

To obtain $\mathbb{P}[Z_{2,g} = q]$, $\mathbb{E}[Z_{2,g}]$, and $\mathbb{V}[Z_{2,g}]$, we substitute analogous quantities $s_{2,0}$ and $s_{2,g-1}$ in place of the quantities $s_{1,0}$ and $s_{1,g-1}$ used to produce $\mathbb{P}[Z_{1,g} = q]$, $\mathbb{E}[Z_{1,g}]$, and $\mathbb{V}[Z_{1,g}]$ in equations (4)–(11).

Nonrecursive mean number of genealogical ancestors

A nonrecursive solution for the mean number of genealogical ancestors from population 1, $\mathbb{E}[Z_{1,g}]$, can be obtained from equation (10). Iterating equation (10) from generation g back to generation 0, we have

$$
\mathbb{E}[Z_{1,g}] = \sum_{i=0}^{g-1} \left(2s_{1,i} \prod_{j=i+1}^{g-1} 2h_j \right), \quad g \geq 1.
\tag{12}
$$

The sum in equation (12) decomposes the expression for $\mathbb{E}[Z_{1,g}]$ into terms that represent ancestors from specific generations. The expected number of genealogical ancestors in generation g is a sum of values contributed by generations $0, 1, \ldots, g - 1$. In particular, the summand $2s_{1,i} \prod_{j=i+1}^{g-1} 2h_j$ represents the expected number of genealogical ancestors contributed by generation i, $0 \leq i \leq g - 1$, to a randomly chosen individual living in the admixed population in generation g. A similar nonrecursive expression can be obtained for $\mathbb{E}[Z_{2,g}]$, substituting $s_{2,i}$ in place of $s_{1,i}$.

Probability of at least one genealogical ancestor in a specified generation

The model also enables a calculation of the probability that an individual from the admixed population has *at least one* genealogical line terminating in a specified source population in a specified generation. Consider an individual in the admixed population H in generation g. For $i = 0, 1, \ldots, g - 1$, denote by X_i the number of the individual's genealogical ancestors from generation i who are also in H. Because generation i is separated by $g - i$ generations from generation g, X_i is a random variable ranging in $[0, 2^{g-i}]$. We define $X_g = 1$, as the individual in generation g is in the admixed population H. Each of the 2 parents of a random individual from generation $i + 1$ is a Bernoulli trial with probability h_i of being from H. Because a parent can be from the admixed population only if the offspring is from the admixed population, X_i is recursively distributed as $X_i \sim \text{Bin}(2X_{i+1}, h_i)$.

For each $i = 0, 1, \ldots, g - 1$, denote by U_i the random number of ancestral lines of a random member of H in generation g that reach S_1 precisely in generation i. For each $i = 0, 1, \ldots, g-1$, if $X_{i+1} = 0$, then $U_i = 0$; otherwise $U_i \sim \mathrm{Bin}(2X_{i+1}, s_{1,i})$.

For $i = 0, 1, \ldots, g-1$, we compute $1 - \mathbb{P}[U_i = 0]$, the probability that a random admixed individual has at least one ancestral line that reaches population S_1 in generation i. By the law of total probability,

$$\begin{aligned} \mathbb{P}[U_i = 0] &= \sum_{m=0}^{2^{g-(i+1)}} \mathbb{P}[U_i = 0 \mid X_{i+1} = m]\mathbb{P}[X_{i+1} = m] \\ &= \sum_{m=0}^{2^{g-(i+1)}} (1 - s_{1,i})^{2m} \mathbb{P}[X_{i+1} = m]. \end{aligned} \tag{13}$$

After recursively computing $\mathbb{P}[X_i = m]$ for each $i = g-1, g-2, \ldots, 0$ for all $m = 0, 1, \ldots, 2^{g-i}$, equation (13) can be evaluated as a function of the parameters $s_{1,i}$ and h_i for $i = 0, 1, \ldots, g-1$. We then obtain the desired probability $1 - \mathbb{P}[U_i = 0]$ for each i. A similar calculation can evaluate the probability that a random member of the admixed population has at least one ancestral line terminating in population S_2 in generation i; we simply substitute $s_{2,i}$ in place of $s_{1,i}$.

Application to African-American genealogies

Overview of the model for African-American admixture history

We use the admixture model to count genealogical ancestors for individuals chosen at random in the African-American population. Our approach involves fitting the model to data on African-American genetic ancestry. We thus estimate admixture parameters under the model, obtaining the expected numbers of African and European genealogical ancestors as byproducts of the estimation.

We constrain the model using known features of African-American demographic history (Berlin 2010; Eltis and Richardson 2010; Franklin and Higginbotham 2021). Starting from the founding of the African-American population, the admixture history of the population can be divided into three demographic epochs prior to 1965: 1619–1808, 1808–1865, and 1865–1965. In the first period, the population was formed from African and European sources, with both sources contributing to the emerging admixed population throughout the period. In the second period, with the end of legal importation of enslaved African captives into the United States, contributions from the African source were much reduced, with contributions from Europeans and European-Americans continuing. In the third period, the end of legal enslavement may have reduced contributions from the European and European-American source, with contributions from the African source remaining low. A three-epoch admixture model for births before 1965 accords with genetic evidence supporting such a division, with dates similar to those suggested by historical periods (Baharian *et al.* 2016).

We focus our attention on the birth cohort 1960–1965 as an endpoint for the model. This cohort is sensible first because much of the genetic data from which model parameters can be estimated traces largely to studies of adult diseases, representing individuals born approximately in this time period. Second, the period after 1965 would introduce a demographically distinct fourth epoch—with additional parameters to estimate—as African-American births after 1965 reflect increased contributions of the African source after an increase in African immigration, and increased contributions of the European source after relaxations of laws and norms limiting acceptance of unions between Africans or African-Americans and Europeans or European-Americans.

With a 25-year generation time, the third epoch contains four generational birth cohorts (1885–1890, 1910–1915, 1935–1940, 1960–1965), the second epoch has three, and the first has seven. Thus, the model has $g = 14$ generations, with generation 14 born during 1960–1965 (Fig. 3).

In our application of the model, we note subtleties of the meanings we use for "African" and "European" genealogical ancestors. First, the approach treats "European and European-American" genealogical ancestors as a single population category, not distinguishing individuals born in Europe from those born in North America. For simplicity, we abbreviate this population as "European."

Second, a person born in Africa who arrived in North America is regarded as "African"; in counting African ancestors, we count African migrants in the ancestry of an African-American. All births in the model take place in the admixed population in North America; a person born in this admixed population is regarded as an "African-American." It is possible for an African-American in the model to have all genealogical ancestors from Africa (or, in principle, from Europe, though this scenario is unlikely in the relevant portion of the parameter space). Irrespective of the person's genetic ancestry, however, such a person is regarded as an African-American.

Finally, owing to human origins in Africa, all humans ultimately have many genealogical ancestors there. Because our application is concerned only with the most recent ~400 years, we understand "African ancestors" to always refer to ancestors from this recent period.

Constraining the three-epoch model by demographic data

We treat the African source population as population 1 and the European source population as population 2. We set $s_{1,0} = 1$ and $s_{2,0} = 0$, founding the African-American population with Africans in the first generation $g = 1$. In the three-epoch model, after the founding with births in generation 1 to parents from generation 0, matings occur intragenerationally between members of generations 1–6 in epoch 1, 7–9 in epoch 2, and 10–13 in epoch 3.

We make use of demographic data to initialize the model for the duration of the first epoch (Hacker 2020). At the start of epoch 1, an individual born in the admixed African-American population in generation 1 has parents only from the African and European populations, and not from the African-American population—as the African-American population did not yet exist in the parental generation 0 (we further assume that all parents of the members of generation 1 are African). By the end of this epoch, an individual born in the admixed African-American population has a high probability of having one or both parents from the African-American population, as the size of the African-American population had grown to exceed the number of arriving Africans.

Let $c_{g-1} = s_{1,g-1}/(s_{1,g-1} + h_{g-1}) = s_{1,g-1}/(1 - s_{2,g-1})$, denoting, for individuals born in the African-American population in generation g, the fraction of their non-European parents who are African arrivals to North America rather than African-American residents. We assume that these parents are drawn in proportion to the population sizes of potential African and African-American parents available at the time of the birth of generation g. Hence, we write $c_{g-1} = \mathcal{S}_{1,g}/(\mathcal{S}_{1,g} + \mathcal{H}_{g-1})$, where $\mathcal{S}_{1,g}$ is the estimated number of African arrivals in generation g, entrants assumed to be of child-bearing age and hence potential parents of individuals born in generation g, and $\mathcal{H}$ is the number of births in the

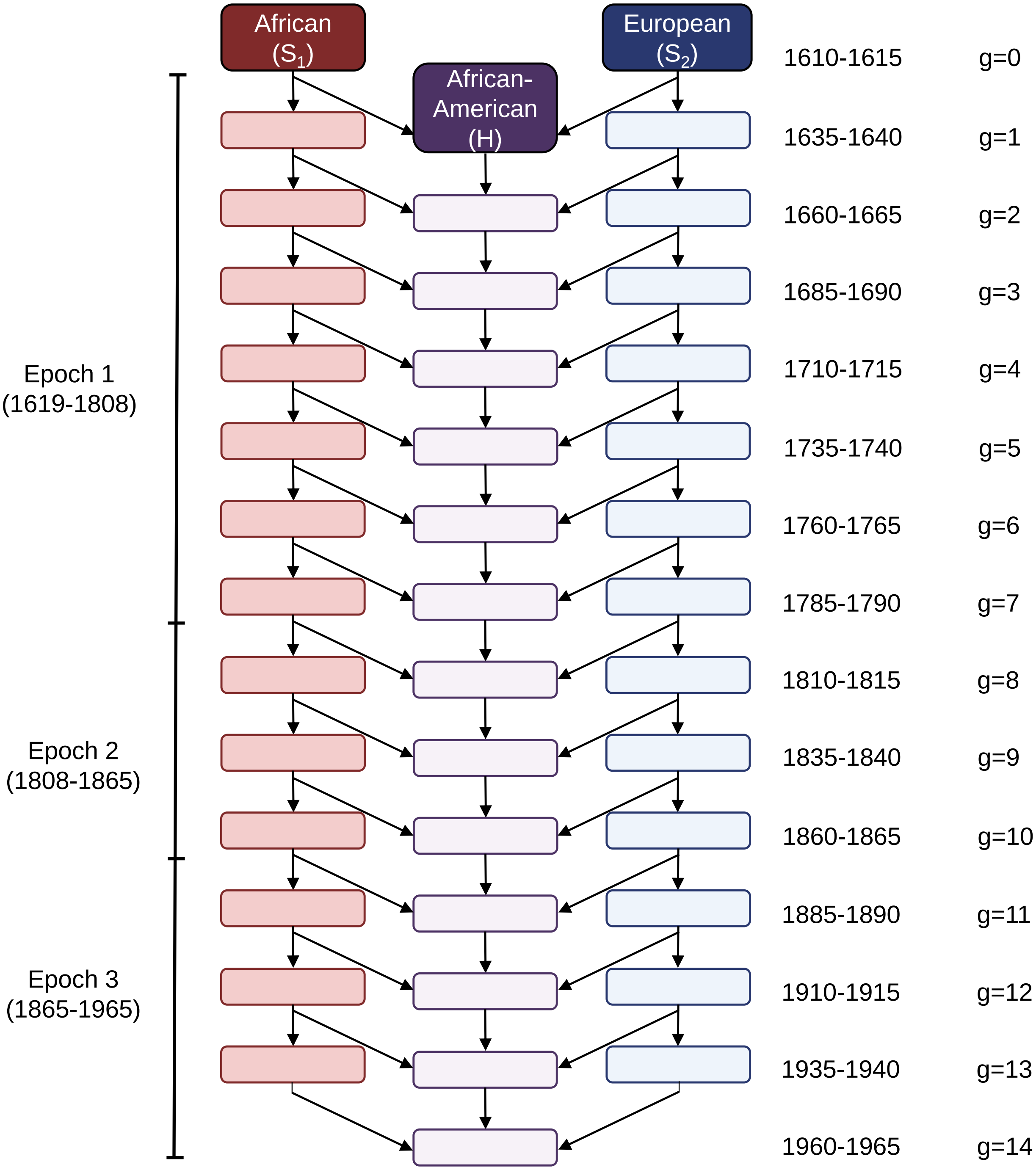

Fig. 3. The admixture model for African-Americans. The model is a special case of Fig. 2. S_1 denotes Africans, S_2 denotes Europeans and European-Americans, and H denotes African-Americans. We consider the births in a 5-year interval to be a discrete generation g, with $g = 0$ corresponding to 1610–1615 and $g = 14$ to 1960–1965, and we assume a 25-year generation time. The model has three epochs, with epochs 1, 2, and 3 corresponding to generations 1–7, 8–10, and 11–14, respectively.

African-American population in generation $g - 1$, members of the previous generation who are also potential parents of individuals born in generation g.

To choose values for c_{g-1}, we use estimated numbers of migrants and births from demographic analysis of the enslaved population (Hacker 2020, columns 7 and 8 of Table 1). Each of our generations is a 5-year interval; we use data reported for the 10-year interval of which that 5-year interval is a subinterval. Thus, for example, c_2, representing the fraction of non-European parents of African-Americans born in generation 3 (1685–1690) who are African, is the ratio of the estimated number of African migrants in 1680–1690 to the sum of this quantity and the estimated number of African-American births 1660–1670 (representing generation 2, 1660–1665). Note that the demographic study (Hacker 2020) focuses on enslaved Africans and African-Americans; we assume that its parameters apply to the entire population of Africans and African-Americans.

With this approach, in epoch 1, for each generation 1–7, we seek to estimate the model parameters $(s_{1,g-1}, h_{g-1}, s_{2,g-1})$ subject to the constraints that for each g from 1 to 7, $s_{1,g-1} = c_{g-1}(1 - s_{2,g-1})$ and

Table 1. Parametrizing a historically informed model.

Generation g	Birth year	Epoch	c_{g-1}
1	1635–1640	1	1
2	1660–1665	1	0.9835
3	1685–1690	1	0.8602
4	1710–1715	1	0.8551
5	1735–1740	1	0.7826
6	1760–1765	1	0.5380
7	1785–1790	1	0.1418
8	1810–1815	2	—
9	1835–1840	2	—
10	1860–1865	2	—
11	1885–1890	3	—
12	1910–1915	3	—
13	1935–1940	3	—
14	1960–1965	3	—

For the non-European contributions in generations 1–7, the model enforces specified ratios of the African to the African-American contributions. For all generations g in epoch 1 (g from 1 to 7), the quantity $c_{g-1} = s_{1,g-1}/(1 - s_{2,g-1})$ denotes, for individuals born in the African-American population in generation g, the fraction of their non-European parents who are African arrivals to North America rather than African-American residents. In our model, we inserted numerical values for this quantity estimated based on demographic data.

$h_{g-1} = (1 - c_{g-1})(1 - s_{2,g-1})$, with each c_{g-1} fixed according to the entries of Table 1 and with $s_{2,g-1}$ equal to the same value for each g from 1 to 7 (to be precise, note that for $g = 1$, no estimation is needed, as $s_{1,0}$ is fixed at 1). In the more recent epochs 2 and 3, we estimate all model parameters $(s_{1,g-1}, h_{g-1}, s_{2,g-1})$ associated with births in generation g, without such constraints. Across the generations within epochs 2 and 3, we assume parameter values are constant, and we index parameters by the first of the contributing generations: 7 and 10. Thus, model parameters for these epochs are $(s_{1,7}, h_7, s_{2,7})$ and $(s_{1,10}, h_{10}, s_{2,10})$, with only two of each parameter trio being free to vary, and the third equaling one minus the sum of the other two (equation (1)). Because model parameters are constant within epochs 2 and 3, we treat model parameters as equal across generations in the recursions that give rise to generations 8–10 and in those that give rise to generations 11–14.

Fitting the model

To fit the model, we search the parameter space, for each choice of model parameters computing the mean and variance of autosomal admixture in generation $g = 14$. We compute $\mathbb{E}[H_{1,14}]$ and $\mathbb{V}[H_{1,14}]$ by recursively applying equations (2) and (3); we proceed similarly for $\mathbb{E}[H_{2,14}]$ and $\mathbb{V}[H_{2,14}]$.

Estimates of African and European ancestry in studies of African-American admixture in different locations and in different conditions of health and disease have been generally concordant, with values of ~80% for the mean African ancestry and ~10% for the standard deviation. For example, in 14 data sets on African-American admixture tabulated by Cheng *et al.* (2009), mean estimated autosomal ancestry from a European ancestral group in African-Americans has range 15–25%, with standard deviation 8–15%. Comparable values have been observed in subsequent studies (Bryc *et al.* 2015; Baharian *et al.* 2016; Micheletti *et al.* 2020).

Because we treat the African-American population as a two-source group, we assume the African and European ancestry components sum to 1. As $\mathbb{V}[X] = \mathbb{V}[1 - X]$ for a random variable X, we assume the two ancestry components have the same variance. Hence, to find parameter sets that give rise to admixture estimates that match those seen by Cheng *et al.* (2009), we search the parameter space for parameter sets that satisfy (i) the mean African ancestry, $\mathbb{E}[H_{1,14}]$, lies in [0.75, 0.85] and (ii) the standard deviation of the African ancestry, $\sqrt{\mathbb{V}[H_{1,14}]}$, lies in [0.08, 0.15].

We choose model parameters on a grid, and we then retain those sets of parameter values that satisfy the required conditions. For each parameter set that is retained, we calculate the mean, variance, and distribution of $Z_{1,14}$ and $Z_{2,14}$ by equations (10), (11), and (7), respectively. We also compute the contributions of specific generations to the mean number of genealogical ancestors, following equation (12). We characterize the properties of the parameter sets that we retain; for each parameter, we summarize the distribution of its accepted values.

The analysis has one free parameter for epoch 1 (the European contribution, say, $s_{2,1}$); for epochs 2 and 3, it has three parameters each ($s_{1,7}$, h_7, $s_{2,7}$ and $s_{1,10}$, h_{10}, $s_{2,10}$), with two of three free to vary in each trio, as the trio necessarily sums to 1. We consider all possible points on a grid with increment 0.01 for each parameter, enforcing an upper bound on the European contributions in all epochs due to the understanding that the African and African-American contributions predominate, an upper bound on the African contribution in epochs 2 and 3 due to comparatively low African immigration in these periods, and a lower bound on the African-American contribution in epochs 2 and 3 as a result of its equaling one minus the European and African contributions (Supplementary Table 1).

Estimated model parameters

Distributions of the estimated model parameter sets that produce a mean and variance of African ancestry within permissible ranges appear in Fig. 4, and they are summarized in Table 2; Supplementary Fig. 1 visualizes these parameter sets on ternary plots in which the constraints that parameters place on one another can be seen. In epoch 1, the generationwise European ancestry contribution lies near the low end of the assumed range (Fig. 4c, Supplementary Fig. 1a), with a median of 0.08 (Table 2). For this epoch, the African and African-American ancestry contributions are determined from demographic information and the European contribution (see Table 1); the estimated African contribution decreases from one generation to the next from the beginning to the end of the epoch (Fig. 4a), and the African-American component increases (Fig. 4b).

In epoch 2, the European contribution has median 0.03 (Table 2), and the distribution of this contribution is concentrated at smaller values than in epoch 1 (Fig. 4f, Supplementary Fig. 1b). The African ancestry contribution is also small (Fig. 4d), with median 0.06; most of the ancestry lies in the African-American component (Fig. 4e).

Finally, in epoch 3, the European contribution decreases further to a median of 0.02 (Table 2), with all the weight placed in the first two bins in Fig. 4i. The African and African-American contributions are similar to those seen in epoch 2 (Fig. 4g and h, Supplementary Fig. 1c), with a slight increase in the median African component (Table 2).

Estimated numbers of genealogical ancestors

Each accepted parameter set generates values for the expected numbers of African and European genealogical ancestors, and the distributions of these quantities appear in Fig. 5 and Table 3. The expected number of African ancestors has a mean of 314 and a median of 299, with an interquartile range from 240 to 376 and a minimum of 124 and maximum of 680 (Table 3). The expected number of European ancestors is smaller and more concentrated, with mean 51, median 51, and

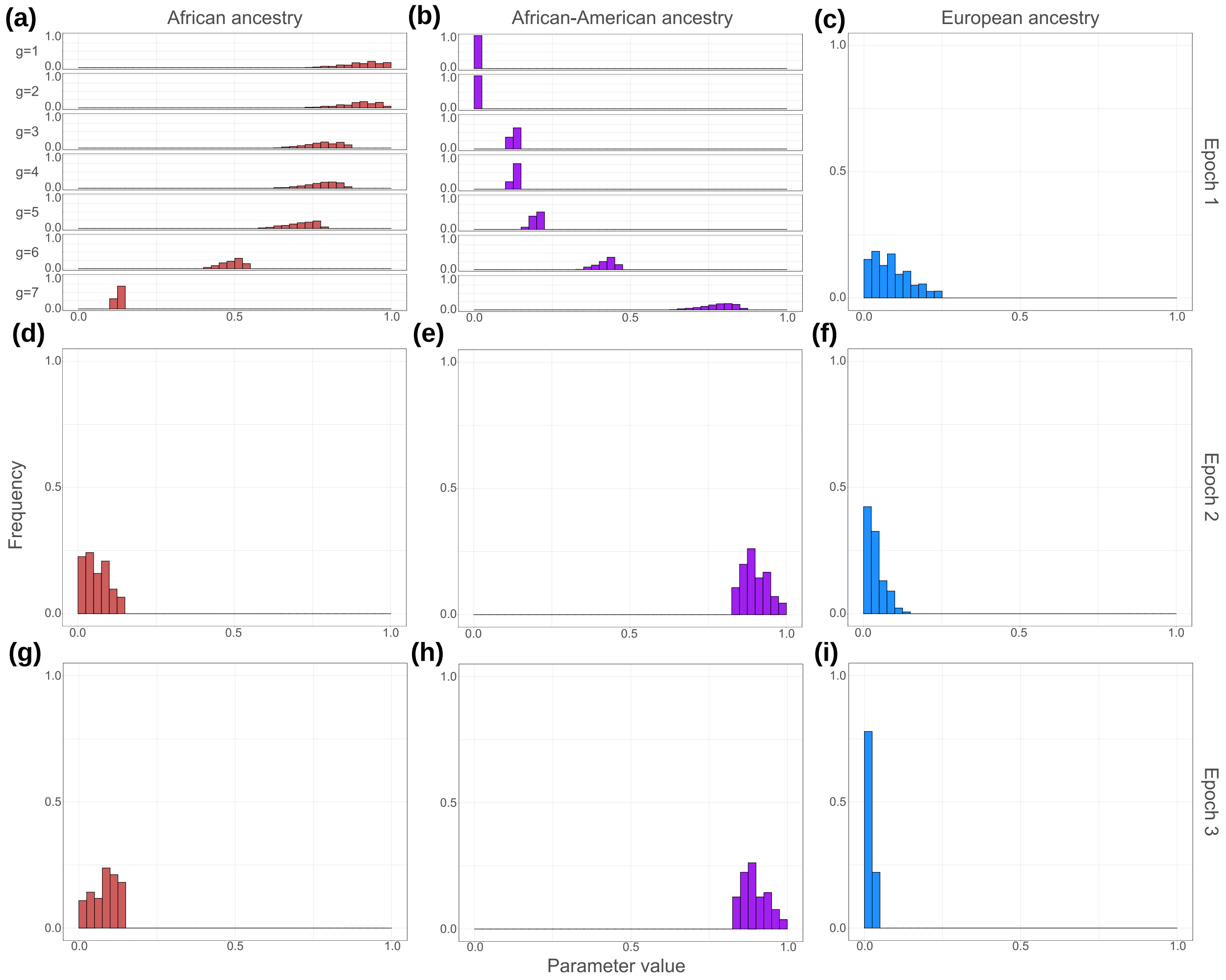

Fig. 4. Distributions of generationwise ancestry contributions estimated for African-Americans. Generationwise ancestry contributions are estimated for Africans, African-Americans, and Europeans and European-Americans. For each population in epochs 2 and 3, and for Europeans in epoch 1, the contribution from that population is assumed to be equal across generations within the epoch; for Africans and African-Americans in epoch 1, the contribution changes across generations according to Table 1. The histograms are constructed from among accepted parameter sets that satisfied specified criteria. In epoch 1, the plots labeled with generation g are the estimates of the parameters that contributed to births of individuals in generation g, representing $s_{1,g-1}$ and h_{g-1}. Parameter values are binned in intervals [0, 0.025], (0.025, 0.05], …, (0.975, 1], half-open in all cases except the closed first bin. (a) African ancestry, epoch 1. (b) African-American ancestry, epoch 1. (c) European ancestry, epoch 1. (d) African ancestry, epoch 2. (e) African-American ancestry, epoch 2. (f) European ancestry, epoch 2. (g) African ancestry, epoch 3. (h) African-American ancestry, epoch 3. (i) European ancestry, epoch 3.

Table 2. Estimated model parameters for a 3-epoch model of African-American demographic history.

Epoch	Population	Mean	Standard deviation	Minimum	First quartile	Median	Third quartile	Maximum
Epoch 1	European ($s_{2,1}$)	0.089	0.061	0	0.04	0.08	0.13	0.25
Epoch 2	African ($s_{1,7}$)	0.061	0.040	0	0.03	0.06	0.09	0.15
	African-American (h_7)	0.902	0.039	0.85	0.87	0.90	0.93	1.00
	European ($s_{2,7}$)	0.037	0.030	0	0.01	0.03	0.05	0.15
Epoch 3	African ($s_{1,10}$)	0.085	0.041	0	0.05	0.09	0.12	0.15
	African-American (h_{10})	0.899	0.039	0.85	0.87	0.89	0.93	0.99
	European ($s_{2,10}$)	0.016	0.010	0	0.01	0.02	0.02	0.03

The table summarizes the parameter sets that produce permissible values for the expectation and variance of $H_{1,14}$, the African ancestry fraction in generation 14. Note that in epoch 1, the African and African-American parameter values are generation-specific, set according to the values in Table 1 rather than estimated. The table is based on 45,189 accepted parameter sets, ~9% of the 480,896 sets examined.

interquartile range from 32 to 69; the minimum is 4 and the maximum is 125.

Considering the expected numbers of African and European ancestors jointly, we observe that across accepted parameter sets, they are negatively correlated ($r = -0.455$, Fig. 6a). For both Africans and Europeans, the standard deviation of the number of ancestors increases with the associated expectation ($r = 0.434$ for Africans, Fig. 6b; $r = 0.900$ for Europeans, Fig. 6c).

Separating the African and European ancestors by their generational timing (Fig. 7 and Supplementary Table 2), we see that the greatest numbers trace to epoch 1, particularly generations 3–5 for Africans (1685–1740) and 4–6 for Europeans (1710–1765). Nonzero values for both quantities continue, decreasing to small values in the most recent generations.

Probability of at least one genealogical ancestor

Applying the estimated means for the admixture parameters, we used equation (13) to evaluate the probability for each generation that an African-American individual has at least one African genealogical ancestor in that generation, and the corresponding probability that an African-American individual has at least one European genealogical ancestor.

Figure 8 plots this probability. For African ancestors, the probability is small for generation 0, increasing to large values for generations 2–6, and then decreasing. For each of generations 2–6, the probability exceeds 0.975 that a random African-American has at least one African ancestor in that generation (Supplementary Table 3). In other words, the probability is near 1 that in each of generations 3–7, the offspring of generations 2–6, at least one individual in a random genealogy has an African parent.

In Fig. 8, in each generation, the probability of at least one European ancestor has a similar pattern, with its largest values in generations 4–6. It remains above 0.5 in each of generations 7–9, and it is substantially lower in generations 10–13.

Discussion

Under models of admixture, we have evaluated the numbers of genealogical lines that trace to particular source populations. The results provide a new perspective on admixture models, focusing on properties of individual genealogies. We have applied this perspective to the case of African-Americans, finding that under a model calibrated by demographic and genetic data, a random African-American genealogy traced back in time from birth in 1960–1965 reaches a mean of 314 African individuals and 51 European or European-American individuals.

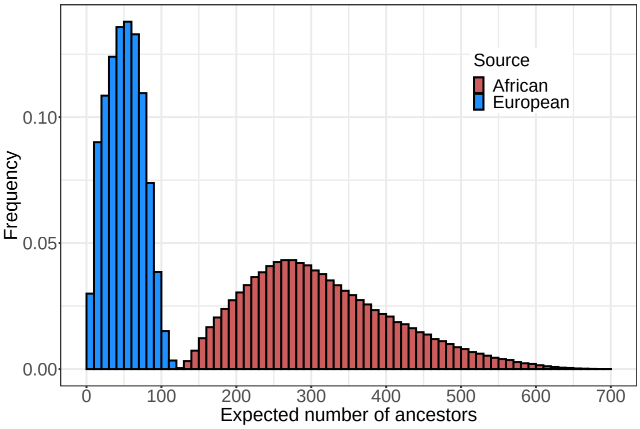

Fig. 5. Distribution of the expectation of the numbers of African and European ancestors across accepted parameter sets. For each accepted set of parameter values, the expected number of African ancestors and the expected number of European ancestors are computed from equation (10). Summaries of the figure appear in Table 3.

Admixture models

Our approach builds on mechanistic admixture models that have characterized the distribution of admixture levels over time as a function of model parameters. The quantities that we examine —properties of the distributions of the number of ancestors from the source populations—are obtained as functions of model parameters in a manner similar to the computation of the distributions of admixture levels. Estimated individual-level genomic ancestry fractions are used to calibrate the models, from which aspects of the numbers of ancestors are calculated in terms of model parameters.

In standard coalescent approaches, the genealogy of a single locus is traced among many individuals back to a common ancestor—disregarding diploid pedigrees. Recent genealogical analyses have sought to also include pedigrees and to examine stochastic processes involving gene lineages on those pedigrees (Wollenberg and Avise 1998; Wakeley et al. 2012; Campbell 2015; Wakeley et al. 2016; Wilton et al. 2017; King et al. 2018; Severson et al. 2019; Cotter et al. 2021; Severson et al. 2021; Cotter et al. 2022). Such studies often analyze properties of genealogical rather than genetic ancestry, using theoretical and simulation-based approaches (Chang 1999; Rohde et al. 2004; Matsen and Evans 2008; Lachance 2009; Gravel and Steel 2015; Kelleher et al. 2016; Edge and Coop 2020). Not all genealogical ancestors are genetic ancestors, and an analysis of the distinction requires detailed consideration of features of genetic transmission from parent to offspring. Our investigation of genealogical lines in admixed populations continues a series of studies that investigates admixed biparental genealogies in the most recent generations (Verdu and Rosenberg 2011; Gravel 2012; Goldberg et al. 2014; Liang and Nielsen 2014; Goldberg and Rosenberg 2015; Goldberg et al. 2020; Kim et al. 2021), and potentially enables extensions for studying genetic ancestors.

African-American demographic history

The results provide insight into African-American history. First, the model suggests that patterns seen in African-American genetic ancestry correspond to a mean of 0.089 for the generationwise European ancestry component in epoch 1, 0.037 in epoch 2, and 0.016 in epoch 3 (Table 2). These values have comparable magnitude to values in other studies that have estimated similar quantities, but without a 3-epoch perspective (Glass and Li 1953; Gross 2018). The European ancestry parameter decreases from the initial period through the last generations of enslavement, decreasing again after the end of slavery.

We estimate that a random African-American born during 1960–1965 has a mean of 314 African ancestors and 51 European

Table 3. Summary statistics for the expected numbers of African and European ancestors for a random individual from the African-American population ($\mathbb{E}[Z_{1,14}]$ and $\mathbb{E}[Z_{2,14}]$).

Quantity	Mean	Standard deviation	Minimum	First quartile	Median	Third quartile	Maximum
African ancestors	314	99	124	240	299	376	680
European ancestors	51	24	4	32	51	69	125

The estimates consider random individuals in the 1960–1965 birth cohort, assumed to be generation $g = 14$ in a 3-epoch model. The quantities in the table summarize results plotted in Fig. 5. Note that the standard deviations shown here are standard deviations of the means $\mathbb{E}[Z_{1,14}]$ and $\mathbb{E}[Z_{2,14}]$ across accepted parameter sets, not standard deviations of $Z_{1,14}$ and $Z_{2,14}$.

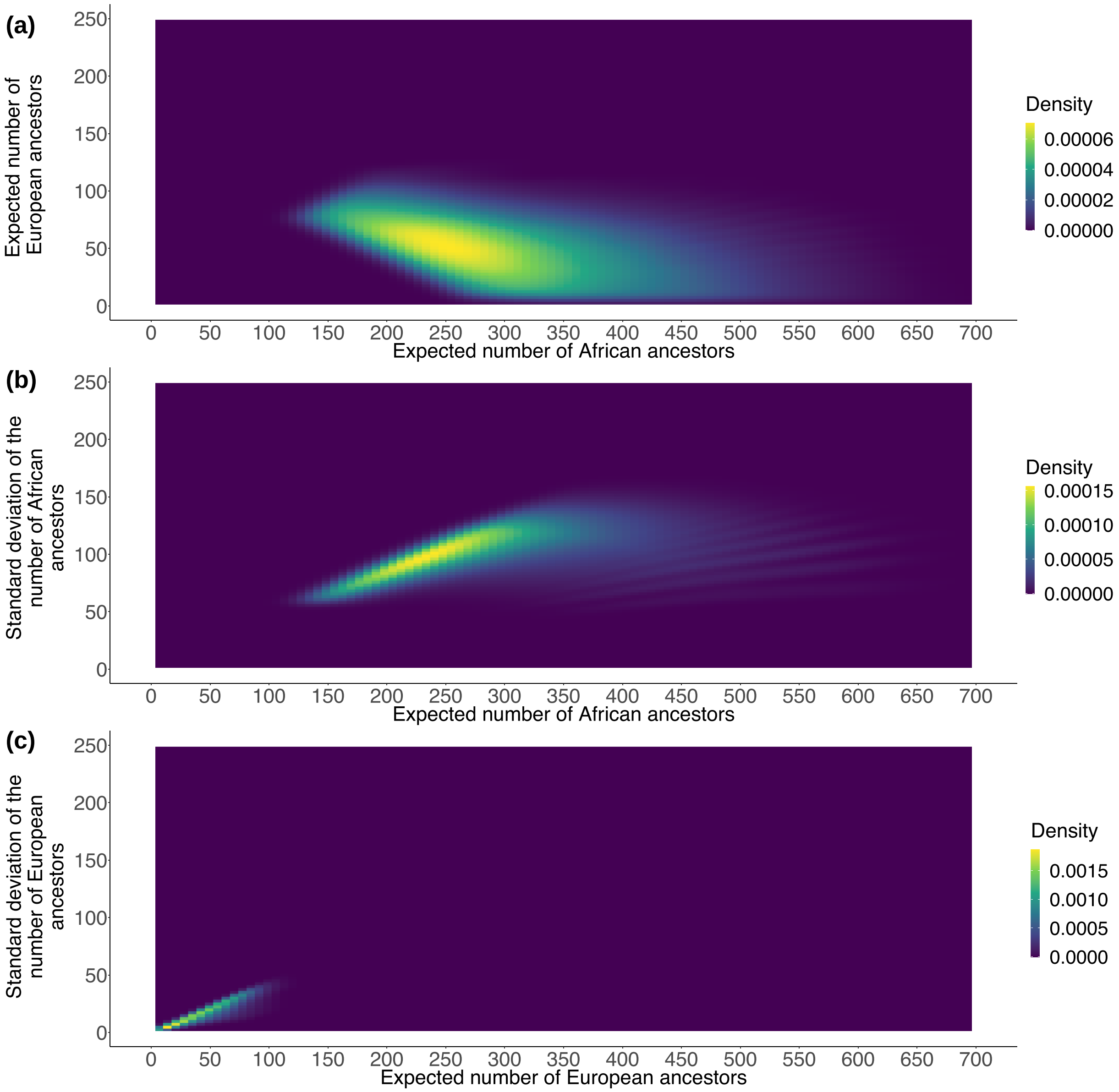

Fig. 6. Joint distributions of the expectations and standard deviations of the numbers of African and European ancestors across accepted parameter sets. For each accepted set of parameter values, the expected number of African ancestors and the expected number of European ancestors are computed from equation (10); the associated standard deviations are computed from equation (11). a) Expected number of European ancestors and expected number of African ancestors. b) Standard deviation of the number of African ancestors and expected number of African ancestors. c) Standard deviation of the number of European ancestors and expected number of European ancestors.

and European-American ancestors (Fig. 5 and Table 3). The model finds that most genealogical lines trace back through African-American ancestors for several generations; at that point, the number of African-American ancestors is large, and some have African parents, European parents, or both. The mean of $314 + 51 = 365$ total African and European ancestors lies between $2^8 = 256$ and $2^9 = 512$, the total numbers of genealogical lines in a pedigree 8 and 9 generations ago; with 365 ancestors from the source populations, some must precede generation 6, which has only 256 total genealogical lines. Indeed, most ancestors from the source populations, both African and European, appear in generations 3–6, 1685–1765 (Fig. 7), with near 100 African ancestors each in generations 4 and 5 (Supplementary Table 2). These results accord with the substantial decrease between generation 6 and generation 7 in the African contribution to the next generation (Table 1)—by which it is sensible that many of the African ancestors trace to generation 6 or earlier.

The proportion of the sum of the mean numbers of African and European ancestors due to African ancestors, $314/(314 + 51) \approx 86\%$, is near the 75–85% range for the African genetic ancestry fraction. That it slightly exceeds this range accords with the observation that European ancestors are slightly more recent than African ancestors in Fig. 7; a European ancestor chosen at random would then have contributed slightly more to a genome than a random African ancestor—with the smaller number of European ancestors, $51/(314 + 51) \approx 14\%$, reflecting 15–25% of the genome.

As a genealogy proceeds back in time, for those genealogical lines that are not from the source populations, the number of lines doubles each generation, potentially driving the temporal maximum for the number of genealogical ancestors early in the

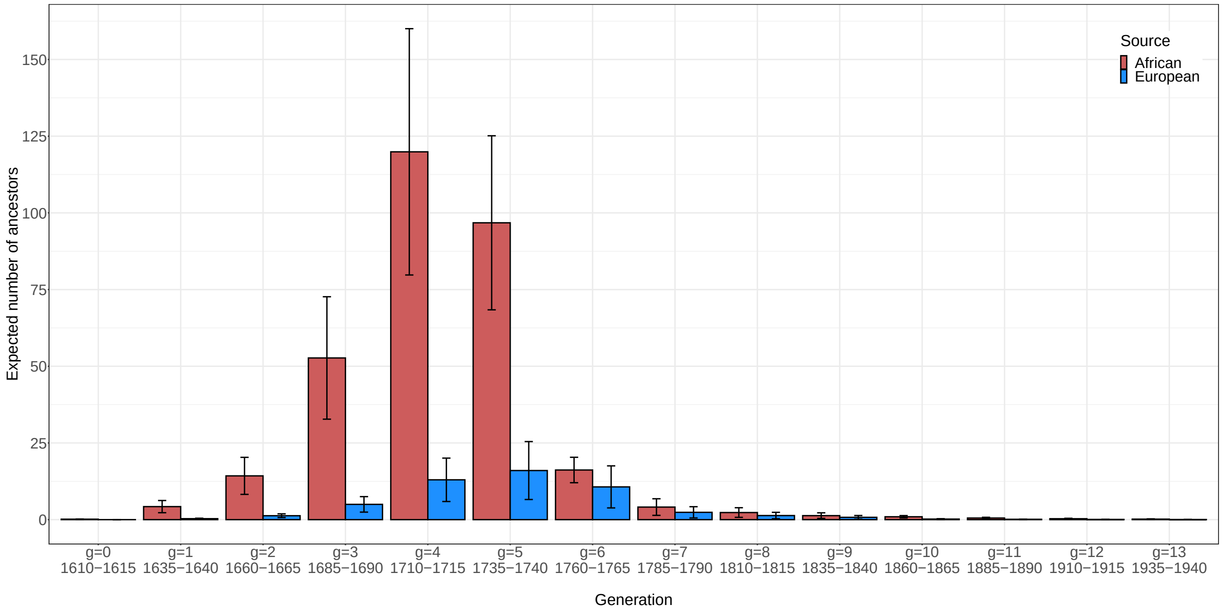

Fig. 7. Generation-specific expectations of the numbers of African and European ancestors across accepted parameter sets. For each accepted set of parameter values, the generation-specific expected number of African ancestors and the generation-specific expected number of European ancestors are computed from equation (12). The height of a bar represents the mean across accepted parameter sets of the generation-specific expected number of ancestors, and the error bars represent standard deviations.

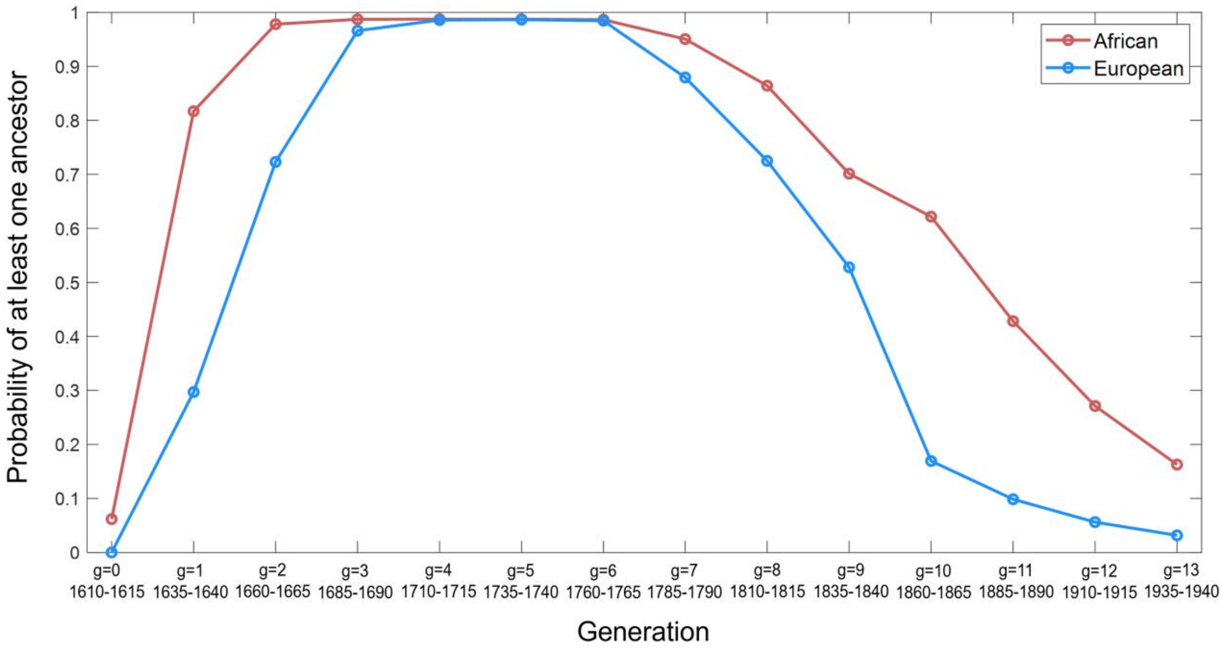

Fig. 8. The probability of having at least one ancestor from a source population in a specified generation. Considering the means among accepted parameter sets, $(s_{1,0}, s_{1,1}, \ldots, s_{1,13}) = (1, 0.896, 0.783, 0.779, 0.713, 0.490, 0.129, 0.061, 0.061, 0.061, 0.085, 0.085, 0.085, 0.085)$, $(h_0, h_1, \ldots, h_{13}) = (0, 0.015, 0.127, 0.132, 0.198, 0.421, 0.781, 0.902, 0.902, 0.902, 0.899, 0.899, 0.899, 0.899)$, and $(s_{2,0}, s_{2,1}, \ldots, s_{2,13}) = (0, 0.089, 0.089, 0.089, 0.089, 0.089, 0.089, 0.037, 0.037, 0.037, 0.016, 0.016, 0.016, 0.016)$ (Tables 1 and 2), the generation-specific probabilities of at least one African ancestor and at least one European ancestor are computed from equation (13).

history of the African-American population. In the early generations, the number of African parents is high relative to African-American parents, so that large numbers of African ancestors accumulate in a pedigree in those generations; in generations after generation 6, the number of African-American parents relative to African parents is high enough that fewer Africans appear. Interestingly, the peak importation of enslaved individuals did not occur until later in the 1700s than the African-ancestor peak (Eltis and Richardson 2010, p. 200); by the time of the importation peak, the fraction of parents of a generation's offspring who were African-American rather than African was already relatively high (Hacker 2020).

The ancestor counts can be approached by a focus on the earliest African ancestor: for a random African-American, what is the distribution of the generation in which the earliest African ancestor lived? In Fig. 8 and Supplementary Table 3, for each of generations 2–6, the probability exceeds 97% that a random African-American contains at least one African ancestor in that generation. In other words, the probability exceeds 97% that in each of generations 3–7, the offspring generations of generations 2–6, at least one individual in a random genealogy is an African-American with an African parent. Considering the earliest of these generations, under the model, a typical African-American born in 1960–1965 likely has at least one ancestor from generation 4 (1710–1715) who was an African-American with an African parent, and it is also likely that such an individual has at least one African-American ancestor from generation 3 as well (1685–1690).

For European ancestors, we find that under the model, the probability is high (>96%) that a random African-American individual has at least one European ancestor in each of generations 3–6, the parents of generations 4–7 (Fig. 8). Although fewer European ancestors are present in generations 7–9 than 3–6, the probability of a European ancestor exceeds 50% in each of generations 7–9. In other words, for example, the probability is above 50% that a random African-American individual has a European ancestor born in generation 9 (1835–1840).

Among the parameter estimates, 0.085 for African ancestry in epoch 3 is potentially misaligned with historical information; this value is large given low levels of African immigration during the period (Reimers 2005; Gates 2009; Berlin 2010). The estimate may reflect any of a number of phenomena. First, individuals from the Caribbean potentially have high African ancestry fractions (Mathias *et al.* 2016; Adhikari *et al.* 2017; Micheletti *et al.* 2020); some of the apparent African immigration detected in epoch 3 might, instead, be misattributed immigration from the Caribbean, a source of more migrants than Africa during the period, though still a small number relative to the resident African-American population (Henke 2001; Reimers 2005; Berlin 2010). Second, the African-American and African ancestry components might be less identifiable than the European component: because the admixed African-American population has greater genetic similarity to the African than to the European population, parameter sets that exchange African-American for African contributions or vice versa might produce similar distributions of ancestry fractions, decreasing identifiability for the African and African-American components. Indeed, these components are negatively correlated across accepted parameter sets (Supplementary Table 4), and their levels of uncertainty in epoch 3 exceed that of the European component (Fig. 4, g to i). An overestimation of the African ancestry component in epoch 3—when the true African ancestry traces to earlier epochs—means that the model may be placing larger fractions of individual pedigrees in the African source population in recent generations than is warranted, though not enough to increase the standard deviation of African ancestry across individuals outside the 8–15% range. To produce the desired mean African ancestry level, one African ancestor in epoch 3 contributes the same amount of African ancestry as multiple African ancestors from earlier epochs. Hence, if the African component in epoch 3 overestimates the true value, then the model may be undercounting the true number of African ancestors—so that a count of 314 may in fact *underestimate* the true value.

Interpretation in relation to a single African-American genealogy

As limitations of African-American genealogical research impede the use of documentary evidence to count genealogical lineages that reach individual African and European ancestors in genealogies of specific individuals (Gates 2009; Swarns 2012; Nelson 2016), our claim that a random African-American born during 1960–1965 has a mean of 314 African and 51 European ancestors provides information that extends beyond what can typically be documented in individual genealogies. To illustrate the meaning of the results, we examine them in the context of a single specific genealogy.

Consider a genealogical study (Swarns 2012) of a prominent African-American: Michelle Obama, born in 1964, corresponding to generation 14 of our model. As her family history has many features typical of African-American genealogies (Swarns 2012), we treat it as an instance of a "random" genealogy. The genealogy has 2 African-American parents, 4 African-American grandparents, 8 African-American great-grandparents, and 10 named African-American great-great-grandparents; the 6 unnamed great-great-grandparents can be inferred to be African Americans as well (2 are described, and the information available about their offspring is suggestive for the other 4 (Swarns 2012, pp. 31, 73, 150)). In the great-great-great-grandparental generation (generation 9 in our model), one European is identified, Charles Shields (born 1839), the father of African-American great-great-grandparent Dolphus Shields born circa 1859, with enslaved African-American mother Melvinia Shields (born c. 1844),

In one of the most extensively investigated African-American genealogies, in tracing back 5 generations—to generation 9 in our model—1 specific named European is reached. From photographs, oral histories, and written records, it can be inferred that at least six other lineages spanning all four grandparental lines likely terminate in a European in that generation or one that precedes it (the Fraser Robinson Sr., James Preston Johnson, Melvinia Shields, Peter Jumper Sr., Dolly Jumper, and Eliza Wade lineages (Swarns 2012, pp. 31, 147, 185, 211, 299). No African ancestors are identifiable by name.

Michelle Obama's ancestors of the last 2–3 generations (generations 11–12) were part of a migration of millions of African-Americans from the American South to northern cities (Lemann 1991; Berlin 2010; Wilkerson 2010). Her ancestors 3–4 generations ago (generations 10–11) were African-Americans living throughout the American South. The large number of southern locations from which they arrived in her home city of Chicago suggests that they can be viewed as an approximately random sample from the region. Her ancestors in the fourth generation back from the present (generation 10) primarily included enslaved individuals and some free African-Americans prior to 1865. The fifth generation (generation 9) includes the likely most recent European appearance in a genealogy that consisted in that generation primarily of enslaved African-Americans. Note that generation 9 is precisely the most recent generation identified in Fig. 8 during which the probability of a European ancestor exceeds 50%.

The small numbers of African and European ancestors who can be named in an African-American genealogy that is, in many ways, typical—1 European, 0 Africans—can be compared with our much larger estimates of the numbers of ancestral lines that, in a typical genealogy, reach the source populations. As the numbers of African and European ancestors in the two most recent generations (12 and 13) are small in the model (Fig. 7), our estimates of 314 African and 51 European ancestors approximately correspond to a claim that for a random African-American born during 1960–1965 with 4 African-American grandparents, each grandparent has a mean of perhaps $\frac{314}{4} = 78.5$ African and $\frac{51}{4} = 12.75$ European ancestors.

In an additional interpretation, the African ancestors largely belong to the groups of individuals who survived forced voyages of enslaved migrants from Africa to the North American mainland, voyages with a collective fatality rate estimated at ~12–29% (Eltis and Richardson 2010, p. 167). Under the model, if it is assumed that almost all the African ancestors before 1808 were enslaved migrants and that no ancestor is an ancestor by multiple paths through a pedigree, then a random African-American born in 1960–1965 is descended from, on average, ~300 separate survivors of these journeys. For the European ancestors, although genetic studies have found that African-Americans have ~20% European ancestry on average, the equivalent of more than one European great-grandparent (12.5% ancestry), African-Americans whose recent ancestors are all African-Americans might have no European ancestors specifically known to them: for Michelle Obama, the most recent European ancestor was discovered by a genealogist (Swarns 2012). Our estimate of a mean of 51 European ancestors amounts to a claim that for a typical African-American genealogy of a person born during 1960–1965, the generations since the founding of the population contain a mean of 51 separate mating events between a European or European-American and an African or African-American.

Limitations

Our analyses of African-Americans make use of empirical estimates of admixture levels together with information on the demographics of enslavement (Hacker 2020). However, we note that the model relies on many assumptions, and it does not consider a variety of known phenomena of African-American demographic history.

First, we assume a fixed generation time of 25 years, with discrete generations and mating that is only intragenerational. We fit the mechanistic model only using genome-wide genetic ancestry levels; more informative length distributions of genomic segments from different source populations could potentially be employed along with extensions of the model predictions. The mating assumptions are simple, and to obtain recursions, we allow as a rare case a historically implausible scenario in which an African-American possesses two European parents. This scenario is unlikely in the model, occurring in a specific birth in generation g with probability equal to the square of the European admixture parameter $s_{2,g-1}$, or 1 in 100 births at 10% for this parameter, 1 in 400 births at 5%, and 1 in 2500 births at 2%; because the numbers of admixed ancestors in pedigrees on the relevant time scale have similar magnitude to these values, few instances of this scenario are expected in any pedigree. A greater limitation is that we treat males and females equivalently, not considering sex-biased admixture; a model with sex bias (Goldberg *et al.* 2014) could potentially be explored, though its larger number of parameters would complicate the estimation.

We have analyzed the African-American population as the outcome of admixture only between African and European sources, and we have not considered Native-American or other sources. Genomic studies generally find that the Native-American contribution is small (Bryc *et al.* 2015; Baharian *et al.* 2016), 3% or less, and that the distribution across African-Americans of the Native-American ancestry component is more difficult to accurately estimate than the African and European contributions. With a model that includes the Native-American contributions as a third source, the distribution of the number of Native-American ancestors could potentially be estimated. Attribution of a small portion of genetic ancestry to Native-American sources would decrease genetic ancestry slightly for both Africans and Europeans, so that some of the African and European ancestors in the model would be replaced by Native-American ancestors.

We also have not considered variation in African and European admixture across the United States. To calibrate the model, we chose a range of admixture estimates for African and European admixture, based on studies in many locations. Parameter estimates for our model of African-American admixture history represent a composite of many subpopulations; in some regions, the numbers of African and European genealogical ancestors might differ from these composite values.

Finally, in counting genealogical ancestors, we have assumed that ancestral individuals do not appear in a genealogy on multiple paths. In a genealogy, multiple genealogical lineages might reach the same ancestor; we have assumed that such ancestor-sharing events are rare in individual genealogies of the last ~400 years. The number of enslaved African migrants brought to the United States has been estimated near ~400,000 prior to 1825 (Eltis and Richardson 2010, p. 200). With 314 African ancestors for a random individual, it is possible that two or more genealogical lines reach the same individual among the ~400,000. Duplication of lines is most likely in the early history of the admixed population, in which the population had the smallest size, and in which many of the ancestors are assigned (Fig. 7). However, as 314 is small in relation to 400,000, any possible overestimation of the number of ancestors due to these duplications is likely to be relatively small.

Conclusions

This study introduces new quantities into the genetic study of admixed populations, namely the numbers of genealogical ancestors in an individual genealogy who were members of the source populations. We have shown how to calculate these quantities from a mechanistic model of ancestry whose parameters can be estimated from admixture levels in an admixed population. The approach yields new information for understanding the history of admixed populations, and in the case of African-Americans, it sheds light on an admixture process many of whose genealogical and demographic aspects are difficult to access by other means.

Data availability

The study uses data that can be found in Table 1 of Hacker (2020). Supplemental material is available at *GENETICS* online.

Acknowledgments

We thank Shai Carmi, Janina Jeff, Tina Lasisi, Magnus Nordborg, John Novembre, John Thornton, and Naomi Zack for discussions and comments.

Funding

Support was provided by National Science Foundation, grant BCS-2116322, by a National Science Foundation Postdoctoral Research Fellowship in Biology, by a Council for Higher Education of Israel Scholarship for Outstanding Postdoctoral Fellows in Data Science, and by the Stanford Center for Computational, Evolutionary, and Human Genomics.

Conflicts of interest

The authors declare no conflict of interest.

Literature cited

Adhikari K, Chacón-Duque JC, Mendoza-Revilla J, Fuentes-Guajardo M, Ruiz-Linares A. The genetic diversity of the Americas. Annu Rev Genomics Hum Genet. 2017;18:277–296. doi:10.1146/annurev-genom-083115-022331

Baharian S, Barakatt M, Gignoux CR, Shringarpure S, Errington J, Blot WJ, Bustamante CD, Kenny EE, Williams SM, Aldrich MC, *et al.* The great migration and African-American genomic diversity. PLoS Genet. 2016;12:e1006059. doi:10.1371/journal.pgen.1006059

Berlin I. The Making of African America: the Four Great Migrations. New York: Viking; 2010.

Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. Am J Hum Genet. 2015;96:37–53. doi:10.1016/j.ajhg.2014.11.010

Campbell RB. The effect of inbreeding constraints and offspring distribution on time to the most recent common ancestor. J Theor Biol. 2015;382:74–80. doi:10.1016/j.jtbi.2015.06.037

Chakraborty R. Gene admixture in human populations: models and predictions. Yrbk Phys Anthropol. 1986;29:1–43. doi:10.1002/ajpa.1330290502

Chang JT. Recent common ancestors of all present-day individuals. Adv Appl Probab. 1999;31:1002–1026. doi:10.1239/aap/1029955256

Cheng C-Y, Kao WHL, Patterson N, Tandon A, Haiman CA, Harris TB, Xing C, John EM, Ambrosone CB, Brancati FL, *et al.* Admixture mapping of 15,280 African Americans identifies obesity susceptibility loci on chromosomes 5 and X. PLoS Genet. 2009;5:e1000490. doi:10.1371/journal.pgen.1000490

Cotter DJ, Severson AL, Carmi S, Rosenberg NA. Limiting distribution of X-chromosomal coalescence times under first-cousin consanguineous mating. Theor Popul Biol. 2022;147:1–15. doi:10.1016/j.tpb.2022.07.002

Cotter DJ, Severson AL, Rosenberg NA. The effect of consanguinity on coalescence times on the X chromosome. Theor Popul Biol. 2021;140:32–43. doi:10.1016/j.tpb.2021.03.004

Edge MD, Coop G. Donnelly (1983) and the limits of genetic genealogy. Theor Popul Biol. 2020;133:23–24. doi:10.1016/j.tpb.2019.08.002

Eltis D, Richardson D. Atlas of the Transatlantic Slave Trade. New Haven: Yale University Press; 2010.

Franklin JH, Higginbotham EB. From Slavery to Freedom: A History of African Americans. 10th ed. New York: McGraw-Hill; 2021.

Gates HL. In Search of Our Roots: How 19 Extraordinary African Americans Reclaimed their Past. New York: Crown Publishers; 2009.

Glass B, Li CC. The dynamics of racial intermixture—an analysis based on the American Negro. Am J Hum Genet. 1953;5:1–20.

Goldberg A, Rastogi A, Rosenberg NA. Assortative mating by population of origin in a mechanistic model of admixture. Theor Popul Biol. 2020;134:129–146. doi:10.1016/j.tpb.2020.02.004

Goldberg A, Rosenberg NA. Beyond 2/3 and 1/3: the complex signatures of sex-biased admixture on the X chromosome. Genetics. 2015;201:263–279. doi:10.1534/genetics.115.178509

Goldberg A, Verdu P, Rosenberg NA. Autosomal admixture levels are informative about sex bias in admixed populations. Genetics. 2014;198:1209–1229. doi:10.1534/genetics.114.166793

Gopalan S, Smith SP, Korunes K, Hamid I, Ramachandran S, Goldberg A. Human genetic admixture through the lens of population genomics. Phil Trans R Soc Lond B. 2022;377:20200410. doi:10.1098/rstb.2020.0410

Gravel S. Population genetics models of local ancestry. Genetics. 2012;191:607–619. doi:10.1534/genetics.112.139808

Gravel S, Steel M. The existence and abundance of ghost ancestors in biparental populations. Theor Popul Biol. 2015;101:47–53. doi:10.1016/j.tpb.2015.02.002

Gross JM. Tests of fit of historically-informed models of African American admixture. Am J Phys Anthropol. 2018;165:211–222. doi:10.1002/ajpa.23343

Hacker JD. From '20. and odd' to 10 million: the growth of the slave population of the United States. Slavery Abol. 2020;41:840–855. doi:10.1080/0144039X.2020.1755502

Henke H. The West Indian Americans. Westport (CT): Greenwood Press; 2001.

Kelleher J, Etheridge AM, Véber A, Barton NH. Spread of pedigree versus genetic ancestry in spatially distributed populations. Theor Popul Biol. 2016;108:1–12. doi:10.1016/j.tpb.2015.10.008

Kim J, Edge MD, Goldberg A, Rosenberg NA. Skin deep: the decoupling of genetic admixture levels from phenotypes that differed between source populations. Am J Phys Anthropol. 2021;175:406–421. doi:10.1002/ajpa.24261

King L, Wakeley J, Carmi S. A non-zero variance of Tajima's estimator for two sequences even for infinitely many unlinked loci. Theor Popul Biol. 2018;122:22–29. doi:10.1016/j.tpb.2017.03.002

Korunes KL, Goldberg A. Human genetic admixture. PLoS Genet. 2021;17:e1009374. doi:10.1371/journal.pgen.1009374

Lachance J. Inbreeding, pedigree size, and the most recent common ancestor of humanity. J Theor Biol. 2009;261:238–247. doi:10.1016/j.jtbi.2009.08.006

Lemann N. The Promised Land: The Great Black Migration and How It Changed America. New York: Random House; 1991.

Liang M, Nielsen R. The lengths of admixture tracts. Genetics. 2014;197:953–967. doi:10.1534/genetics.114.162362

Long JC. The genetic structure of admixed populations. Genetics. 1991;127:417–428. doi:10.1093/genetics/127.2.417

Mathias RA, *et al.* A continuum of admixture in the Western Hemisphere revealed by the African Disapora genome. Nat Commun. 2016;7:12522. doi:10.1038/ncomms12522

Matsen FA, Evans SN. To what extent does genealogical ancestry imply genetic ancestry? Theor Popul Biol. 2008;74:182–190. doi:10.1016/j.tpb.2008.06.003

Micheletti SJ, Bryc K, Esselmann SGA, Freyman WA, Moreno ME, Poznik GD, Shastri AJ, Beleza S, Mountain JL. 23andMe Research Team. Genetic consequences of the Transatlantic Slave Trade in the Americas. Am J Hum Genet. 2020;107:265–277. doi:10.1016/j.ajhg.2020.06.012

Nelson A. The Social Life of DNA: Race, Reparations, and Reconciliation after the Genome. Boston: Beacon Press; 2016.

Reimers DM. Other Immigrants: the Global Origins of the American People. New York: New York University Press; 2005.

Rohde DLT, Olson S, Chang JT. Modelling the recent common ancestry of all living humans. Nature. 2004;431:562–566. doi:10.1038/nature02842

Severson AL, Carmi S, Rosenberg NA. The effect of consanguinity on between-individual identity-by-descent sharing. Genetics. 2019;212:305–316. doi:10.1534/genetics.119.302136

Severson AL, Carmi S, Rosenberg NA. Variance and limiting distribution of coalescence times in a diploid model of a consanguineous population. Theor Popul Biol. 2021;139:50–65. doi:10.1016/j.tpb.2021.02.002

Swarns RL. American Tapestry: the Story of the Black, White, and Multiracial Ancestors of Michelle Obama. New York: Amistad; 2012.

Verdu P, *et al.* Patterns of admixture and population structure in native populations of northwest North America. PLoS Genet. 2014;10:e1004530. doi:10.1371/journal.pgen.1004530

Verdu P, Rosenberg NA. A general mechanistic model for admixture histories of hybrid populations. Genetics. 2011;189:1413–1426. doi:10.1534/genetics.111.132787

Wakeley J, King L, Low BS, Ramachandran S. Gene genealogies within a fixed pedigree, and the robustness of Kingman's coalescent. Genetics. 2012;190:1433–1445. doi:10.1534/genetics.111.135574

Wakeley J, King L, Wilton PR. Effects of the population pedigree on genetic signatures of historical demographic events. Proc Natl Acad Sci USA. 2016;113:7994–8001. doi:10.1073/pnas.1601080113

Wilkerson I. The Warmth of Other Suns. New York: Random House; 2010.

Wilton PR, Baduel P, Landon MM, Wakeley J. Population structure and coalescence in pedigrees: comparisons to the structured coalescent and a framework for inference. Theor Popul Biol. 2017; 115:1–12. doi:10.1016/j.tpb.2017.01.004

Wollenberg K, Avise JC. Sampling properties of genealogical pathways underlying population pedigrees. Evolution. 1998;52: 957–966. doi:10.2307/2411228

Zaitlen N, *et al.* The effects of migration and assortative mating on admixture linkage disequilibrium. Genetics. 2017;205:375–383. doi:10.1534/genetics.116.192138

Editor: J. Novembre