



Shared Ancestors and the Birthday Problem

Lily Agranat-Tamir, Kennedy D. Agwamba, Jazlyn A. Mooney & Noah A. Rosenberg

To cite this article: Lily Agranat-Tamir, Kennedy D. Agwamba, Jazlyn A. Mooney & Noah A. Rosenberg (2026) Shared Ancestors and the Birthday Problem, The American Statistician, 80:3, 413-422, DOI: [10.1080/00031305.2025.2595972](https://doi.org/10.1080/00031305.2025.2595972)

To link to this article: <https://doi.org/10.1080/00031305.2025.2595972>



© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 10 Feb 2026.



Submit your article to this journal [↗](#)



Article views: 1291



View related articles [↗](#)



View Crossmark data [↗](#)

Shared Ancestors and the Birthday Problem

Lily Agranat-Tamir^a , Kennedy D. Agwamba^a , Jazlyn A. Mooney^b , and Noah A. Rosenberg^a 

^aDepartment of Biology, Stanford University, Stanford, CA; ^bDepartment of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA

ABSTRACT

The celebrated *birthday problem* asks for the probability that in a classroom of k people, there exists at least one pair with a shared birthday; a generalization seeks the probability that in two classrooms of $k^{(1)}$ and $k^{(2)}$ people, at least one shared birthday exists between classrooms. We examine a problem concerning genealogical ancestry as a variant of this ditype birthday problem. Suppose two people descend from the same demographic event. What is the probability that they possess at least one ancestor in common who experienced the event? We use the birthday problem to pose a question concerning genealogies in the African-American population, for which, due to the history of slavery, much genealogical information was not recorded or has been lost: what is the probability that two randomly chosen African-Americans descended from enslaved Africans forcibly transported to North America share at least one such transported ancestor in common? Using a demographic model and the ditype birthday problem, we find that the probability is approximately 19%–31% that two randomly chosen African-Americans born around 1960–1965 possess at least one shared African ancestor. The result contributes to a deeper understanding of African-American genealogies—and it illustrates an application of the birthday problem.

ARTICLE HISTORY

Received May 2025
Accepted November 2025

KEYWORDS

Ancestry; Birthday problem; Genealogy; Population genetics

1. Introduction

The *birthday problem* has long been recognized as a celebrated calculation of elementary probability (Ball and Coxeter 1939, pp. 45–46; Von Mises 1964; McKinney 1966; Diaconis and Mosteller 1989; Borja and Haigh 2007). The problem asks: in a classroom of $k \leq N$ students, where N denotes the number of days of the year, what is the probability of the event A that there exists at least one day shared as a birthday by two or more of the k students? It is solved by computing the probability of the complement of A , the probability that the k students each have distinct birthdays. The first student has N possible birthdays, the second has $N - 1$ possible birthdays, and so on, until the k th student has $N - (k - 1)$ possible birthdays. The total number of ways of assigning birthdays to k students is N^k . The problem has the familiar solution

$$\mathbb{P}[A] = 1 - \mathbb{P}[A^c] = 1 - \frac{\prod_{i=0}^{k-1} (N - i)}{N^k} = 1 - \frac{N!}{(N - k)! N^k}. \quad (1)$$

With its elementary solution and counterintuitively high answers, the birthday problem has produced many classrooms surprised by shared birthdays in small groups, inspired countless homework problems, and given rise to numerous variants and extensions (Good 1950; Holst 1986; Mathis 1991; Flajolet, Gardy, and Thimonier 1992). Of greatest interest here is the *multitype* birthday problem, and specifically, the *ditype* birthday problem,

in which the probability is computed that there exists at least one day shared as a birthday by a member of subgroup 1 of size $k^{(1)}$ and a member of subgroup 2 of size $k^{(2)}$. Denoting the event of at least one shared birthday between members of the two subgroups by B , we compute the probability of the complement of B . If it is known that the $k^{(1)}$ birthdays in subgroup 1 are all distinct and that the $k^{(2)}$ birthdays in subgroup 2 are all distinct, then

$$\begin{aligned} \mathbb{P}[B] &= 1 - \mathbb{P}[B^c] = 1 - \frac{\prod_{i=0}^{k^{(1)}+k^{(2)}-1} (N - i)}{[\prod_{i_1=0}^{k^{(1)}-1} (N - i_1)][\prod_{i_2=0}^{k^{(2)}-1} (N - i_2)]} \\ &= 1 - \frac{(N - k^{(1)})! (N - k^{(2)})!}{(N - k^{(1)} - k^{(2)})! N!}. \end{aligned} \quad (2)$$

If it is *not* known that the $k^{(1)}$ birthdays in subgroup 1 are distinct, then those $k^{(1)}$ birthdays can occur on j_1 distinct days for any value of j_1 satisfying $1 \leq j_1 \leq k^{(1)}$; the number of ways that the birthdays of $k^{(1)}$ people can occur on exactly j_1 distinct days is the number of ways to place $k^{(1)}$ labeled objects into j_1 unlabeled classes, the Stirling number of the second kind, $S_2(k^{(1)}, j_1)$ (Graham, Knuth, and Patashnik 2008, p. 258). Similarly, the $k^{(2)}$ birthdays in subgroup 2 occur on j_2 distinct days, $1 \leq j_2 \leq k^{(2)}$, in $S_2(k^{(2)}, j_2)$ ways. Once the j_1 and j_2 days have been selected, we calculate $\prod_{j=0}^{j_1+j_2-1} (N - j)$, the number of ways of selecting $j_1 + j_2$ distinct days for $j_1 + j_2$ birthdays,

in the same way as in (1). The total number of ways of selecting birthdays of $k^{(1)} + k^{(2)}$ people is $N^{k^{(1)}+k^{(2)}}$, so that (Wendl 2003)

$$\begin{aligned} \mathbb{P}[B] &= 1 - \mathbb{P}[B^C] \\ &= 1 - \frac{\sum_{j_1=1}^{k^{(1)}} \sum_{j_2=1}^{k^{(2)}} S_2(k^{(1)}, j_1) S_2(k^{(2)}, j_2) \prod_{j=0}^{j_1+j_2-1} (N-j)}{N^{k^{(1)}+k^{(2)}}}. \end{aligned} \quad (3)$$

2. Shared Genealogical Ancestors

We consider a birthday problem in the study of genealogical ancestry. An interest in mathematical aspects of genealogies—the structures of families, often of specific individuals—arises in areas as diverse as animal breeding (MacCluer et al. 1983; Farquharson, Hogg, and Grueber 2021), forensic science (Erich et al. 2018; Ertürk et al. 2022), human demography (Caswell 2019; Anderson-Trocmé et al. 2023), theoretical (Wakeley et al. 2012; Severson, Carmi, and Rosenberg 2021) and empirical (Pemberton 2008; Chen et al. 2019) population genetics, statistical genetics (Thompson 2000; The U.S.–Venezuela Collaborative Research Project & Wexler 2004), wildlife biology (vonHoldt et al. 2008; Galla et al. 2021), and the mathematical population biology of genealogies themselves (Derrida, Manrubia, and Zanette 2000; Lachance 2009). In the context of human history, features of genealogies contribute to understanding the events of recent generations and the relationship of modern descendants to those events. Our focus is on this phenomenon, in which major demographic events occurred in the recent past, and the genealogical connections of descendants to those events are of interest. If precise genealogical information has been lost, then a significant role can exist for a mathematical model to approximate features of unknown genealogies (Chang 1999; Rohde, Olson, and Chang 2004; Mooney et al. 2023).

What is the probability that two descendants of a group of individuals who experienced an event some time in the past share at least one ancestor in the group? Numerical computations concerning the relationship of descendants to a past demographic event help provide context for the impact of the event, accounting for the size of the population that experienced it, the time elapsed, and features of the modern population.

Consider an event that occurred some time ago, in which an estimated N individuals participated. Suppose that a descendant has an estimated $k^{(1)}$ ancestors, not necessarily distinct, who participated in the event, and that another descendant has an estimated $k^{(2)}$ ancestors, not necessarily distinct, who also participated. The probability that the two descendants share an ancestor who participated in the event can be obtained by use of the ditype birthday problem. In the birthday problem analogy, each of the $k^{(1)}$ ancestors is analogous to a birthday of a person in subgroup 1, and each of the $k^{(2)}$ ancestors is analogous to a birthday of a person in subgroup 2. The N potential ancestors are analogous to the days of the year that could potentially be a person's birthday. The probability of a shared ancestor who participated in the event is obtained from (3), or (2) if the $k^{(1)}$ ancestors are all assumed to be distinct from one another, and the $k^{(2)}$ ancestors are also all assumed to be distinct from one another.

3. African Ancestors in the Transatlantic Slave Trade

Relying on calibrations from genetic data, we have been studying genealogical models of the relationship of descendant members of the African-American population of the United States to the Transatlantic Slave Trade (Mooney et al. 2023; Agranat-Tamir, Mooney, and Rosenberg 2024). Our models follow the basic structure of African-American demographic history (Berlin 2010; Eltis and Richardson 2010; Franklin and Higginbotham 2021), beginning with the traditional 1619 start date for the transport of enslaved African captives to the predecessor colonies of the United States. Over the duration of the Transatlantic Slave Trade, 400,000–500,000 enslaved Africans were transported to the United States and antecedent colonies (Eltis and Richardson 2010; Hacker 2020). Prior to the onset of substantial African immigration to the United States after 1965, the genetic ancestry of modern African-Americans traced largely to these enslaved Africans and to genetic admixture with Europeans and European-Americans, with offspring of a pair containing an African or African-American and a European or European-American generally regarded as African-American in most circumstances.

Tracing a genealogy of an African-American descendant backward in time from the present, each lineage eventually reaches a person from an ancestral source population for African-Americans—an African, or a European or European-American. Genetic studies have analyzed the genetic admixture of the population, estimating continent-level values at 75%–85% for African ancestry and 15%–25% for European ancestry (e.g., Bryc et al. 2015; Baharian et al. 2016; Micheletti et al. 2020). Hence, with other possible ancestries excluded, each lineage traced back from an African-American descendant reaches an African or a European or European-American who lived some time since 1619.

Using a mathematical model of African-American ancestry grounded in genetics and historical demography, we have estimated that for a randomly chosen African-American born during 1960–1965, if the genealogical lineages of that person are traced back to individuals originating in the African and European source populations, then the mean number of genealogical lines terminating with African individuals born prior to the end of slavery in 1865 is 313, and the mean number reaching Europeans or European-Americans is 51 (Table 1; Mooney et al. 2023; Agranat-Tamir, Mooney, and Rosenberg 2024). These numbers have historically meaningful interpretations. Because most Africans arriving in the United States and antecedent colonies prior to 1865 were forcibly transported, the number of African ancestors can be interpreted as an approximation of the number of a person's genealogical ancestors who were enslaved captives transported from Africa. The number of European or European-American ancestors can be interpreted as the number of separate admixture events in an individual genealogy—the number of mating pairs in a person's ancestry in which one parent was a European or European-American and the other was an African or African-American.

If an African-American born during 1960–1965 has 313 African genealogical lines tracing to forcibly transported enslaved captives, if all those lines represent distinct people, and if the total number of transported enslaved persons during the

Transatlantic Slave Trade was 450,000, then a single random person's genealogy contains a nontrivial fraction of all such captives, $313/450,000 \approx 6.96 \times 10^{-4}$. As a result, the probability that two randomly chosen genealogies share at least one African ancestor in common is also nontrivial.

We use the birthday problem to evaluate the probability that two randomly chosen African-Americans born during 1960–1965 possess at least one forcibly transported enslaved ancestor in common; in using the birthday problem to calculate the probability that two descendants share an ancestor who experienced a historical event, the “event” is the Transatlantic Slave Trade. In the birthday problem, each day of the year has the same probability to be the birthday of a person chosen at random. Hence, in our framework, each potential enslaved ancestor is initially treated as having the same probability to occur in a genealogy. To achieve randomness of ancestry among transported Africans, the problem considers two randomly chosen African-Americans whose African ancestors can be regarded as randomly drawn among potential ancestors.

The sequence of African-American migrations within the United States (Lemann 1991; Berlin 2010; Wilkerson 2010) contributes to our assumptions concerning the randomness of ancestors. During the slavery period, the legal sale of enslaved persons and the mobility of the enslavers who held them as property produced considerable migration of African-Americans within the United States. At the end of slavery in 1865, the African-American population was concentrated in southern states, after which the Great Migration of 1900–1970 brought large numbers of African-Americans north and west. A descendant of this migration born during 1960–1965 is likely to possess ancestors from multiple places in the southern states, ancestors who in turn likely descend from substantial earlier migration within those states. Hence, we approximate that we can regard such a person as drawing ancestors randomly from the $\sim 450,000$ transported Africans.

As one example, the first lady Michelle Obama was born in 1964 in the northern city of Chicago, with parents who had lived in Chicago, and grandparents and great-grandparents who had moved there during the Great Migration (Swarns 2012). A genealogical investigation of her ancestors found that the ancestors who arrived in Chicago and their own recent ancestors had lived in most of the states in which the African-American population had been concentrated: Alabama, Georgia, Kentucky, Louisiana, Mississippi, Missouri, North Carolina, South Carolina, Tennessee, and Virginia, as well as in southern Illinois.

Much information about individual African-American genealogies has been lost beyond a few generations in the past (Gates 2009; Swarns 2012; Nelson 2016), as family structures of enslaved persons were generally not recorded in legal documents, and oral histories often were not transmitted within families by the formerly enslaved in the immediate period after the end of slavery. Mathematical models hence have the potential to recover otherwise unavailable features of these genealogies. Use of the birthday problem to assess the probability that two African-Americans share at least one enslaved ancestor can therefore contribute to deeper understanding of relatedness and genealogical history in the African-American population.

4. Results

We now evaluate the probability that two modern African-American individuals share at least one ancestor who experienced the Transatlantic Slave Trade. We rely on estimates from the mathematical model of Mooney et al. (2023) of the number of African ancestors for a randomly sampled African-American. We use estimates of the number of transported enslaved captives from the demographic analysis of Hacker (2020).

4.1. An Initial Calculation

In the simplest calculation, we can compute (2) or (3) using our estimate that a typical African-American has a mean of 313 African genealogical ancestors born 1860–1865 or earlier. Our estimation in Mooney et al. (2023) did not assume that those ancestors were distinct, so that (3) provides a more appropriate calculation. We follow the demographic estimates of Table 1 of Hacker (2020), summing entries in columns 8 (“number imported”) and 9 (“additional number smuggled”) to estimate the total number of transported enslaved captives through 1865 at 463,098 (note that we sum the columns of Hacker (2020) ourselves, as the column-8 total reported by Hacker (2020) does not equal the sum of the numbers in the column).

Using $k^{(1)} = k^{(2)} = 313$ and $N = 463,098$, we obtain $\mathbb{P}[B] \approx 0.1907$ for the desired probability according to (3), $\mathbb{P}[B] \approx 0.1908$ according to (2) under the assumption that the $k^{(1)}$ ancestors were distinct from one another, and that the $k^{(2)}$ ancestors were also distinct from one another. In other words, with probability $\sim 19\%$, two African-Americans born in 1960–1965 possess at least one shared enslaved migrant ancestor.

4.2. Male and Female Ancestors Distinguished

In the next calculation, we distinguish between male and female ancestors. Denoting the event of sharing at least one male ancestor by $B^{(m)}$ and the event of sharing at least one female ancestor by $B^{(f)}$, the event that at least one ancestor, male or female, is shared, is $B = B^{(m)} \cup B^{(f)}$. In the simplest approach, we suppose that the $k^{(1)} = k^{(2)}$ African ancestors of a descendant are equally divided between males and females. We also suppose that the possible ancestors follow the estimated overall slave trade sex ratio of $r = 2$ males to every female (Felipe-Gonzalez 2024), so that the $N = 463,098$ transported individuals include 308,732 males and 154,366 females. Finally, we suppose that events $B^{(m)}$ and $B^{(f)}$ are independent.

For males, with (3), we use $k^{(1)} = k^{(2)} = 157$, assigning the remainder of $313/2$ to be male rather than female to produce a conservative estimate of the probability of at least one shared ancestor, and $N_m = 308,732$. We obtain $\mathbb{P}[B^{(m)}] \approx 0.0767$. Similarly, for females, using $k^{(1)} = k^{(2)} = 156$ and $N_f = 154,366$, we have $\mathbb{P}[B^{(f)}] \approx 0.1459$. We then obtain

$$\mathbb{P}[B] = \mathbb{P}[B^{(m)}] + \mathbb{P}[B^{(f)}] - \mathbb{P}[B^{(m)}]\mathbb{P}[B^{(f)}] \approx 0.2114. \quad (4)$$

When we instead use (2), assuming that the male ancestors are distinct from one another and the female ancestors are distinct from one another, we obtain $\mathbb{P}[B^{(m)}] \approx 0.0768$, $\mathbb{P}[B^{(f)}] \approx 0.1460$, and $\mathbb{P}[B] \approx 0.2116$.

Note that if we do not suppose that $B^{(m)}$ and $B^{(f)}$ are independent, then we still have bounds $\mathbb{P}[B^{(f)}] \leq \mathbb{P}[B] \leq \mathbb{P}[B^{(m)}] + \mathbb{P}[B^{(f)}]$, so that $0.1459 \leq \mathbb{P}[B] \leq 0.2226$ with (3), or $0.1460 \leq \mathbb{P}[B] \leq 0.2228$ with (2).

4.3. Generation-Specific Ancestors

The most important distinction of the African-American shared ancestry problem from the classic birthday problem is that the Transatlantic Slave Trade lasted many generations, so that the 463,098 potential ancestors do not all have the same probability of being sampled as an ancestor in a genealogy. Our genealogical model of Mooney et al. (2023) provides an approach for addressing this aspect of the population history.

Table 1 summarizes estimates from our discrete-time model of Mooney et al. (2023), in which we estimated the mean number of African genealogical ancestors possessed by an African-American descendant in each of a discrete series of 25-year generations. Let B_ℓ denote the event that an African ancestor who was born in generation ℓ is shared by two descendants, where $0 \leq \ell < g$. The event that at least one ancestor is shared by two descendants in *some* generation is $B = \cup_{\ell=0}^{g-1} B_\ell$. Treating the event of a shared ancestor in generation ℓ as independent

of the event of a shared ancestor in generation $j \neq \ell$, by De Morgan's laws, we have

$$\mathbb{P}[B] = \mathbb{P}\left[\left(\bigcap_{\ell=0}^{g-1} B_\ell^c\right)^c\right] = 1 - \prod_{\ell=0}^{g-1} (1 - \mathbb{P}[B_\ell]). \quad (5)$$

Each $\mathbb{P}[B_\ell]$ can be calculated according to (3) or (2). Because most transported captives were adults at the time of transportation (Eltis and Engerman 1992, 1993), the first birth cohort that we assume not to have included transported enslaved migrants is generation $g = 10$. For generations ℓ from 0 to 9, we round the “African” column of Table 1 to obtain an integer for the number of African ancestors born in the generation—the classroom sizes $k_\ell^{(1)} = k_\ell^{(2)}$ in the birthday problem—reporting this value in Table 2.

We divide the $N = 463,098$ total African ancestors across the generations by use of tabulations of Hacker (2020). Again taking into account that most transported captives were adults (Eltis and Engerman 1992, 1993) who can be modeled as young adults at reproductive age (McClelland and Zeckhauser 1982, p. 48), transportation is assumed to have taken place one generation later than the birth generation. The total number of enslaved captives in a generation is obtained from counts reported in 10-year cohorts by Table 1 of Hacker (2020), summing the count

Table 1. Features of African, European/European-American, and African-American genealogical ancestors for an African-American born in 1960–1965.

Generation	Birth year	Mean number of genealogical ancestors		
		African	European/European-American	African-American
0	1610–1615	0.14	0	0
1	1635–1640	4.25	0.32	0.07
2	1660–1665	14.27	1.28	2.32
3	1685–1690	52.70	4.98	8.93
4	1710–1715	119.90	12.98	33.30
5	1735–1740	96.76	16.01	83.09
6	1760–1765	16.18	10.67	97.93
7	1785–1790	4.10	2.38	62.39
8	1810–1815	2.31	1.33	34.43
9	1835–1840	1.31	0.75	19.04
10	1860–1865	0.94	0.19	10.55
11	1885–1890	0.53	0.10	5.84
12	1910–1915	0.30	0.06	3.24
13	1935–1940	0.17	0.03	1.80
Total	1610–1940	313.86	51.08	362.93

NOTE: Estimates of the mean number of genealogical ancestors are taken from Tables 1 and 4 of Agranat-Tamir, Mooney, and Rosenberg (2024) based on the model of Mooney et al. (2023).

Table 2. Integer values used for birthday problem computations of the probability of a shared enslaved migrant ancestor for two African-Americans born during 1960–1965.

Generation	Birth year of ancestor	Mean number of African genealogical ancestors (“classroom size”)			Transportation year of ancestor	Number of potential ancestors (“days in the year”)		
		Total	Male	Female		Total	Male	Female
0	1590–1615	0	0	0	1615–1640	292	176	116
1	1615–1640	4	2	2	1640–1665	3155	1902	1253
2	1640–1665	14	7	7	1665–1690	10,048	6059	3989
3	1665–1690	53	27	26	1690–1715	33,450	21,114	12,336
4	1690–1715	120	60	60	1715–1740	116,817	75,931	40,886
5	1715–1740	97	49	48	1740–1765	100,952	65,619	35,333
6	1740–1765	16	8	8	1765–1790	96,012	62,408	33,604
7	1765–1790	4	2	2	1790–1815	96,155	63,751	32,404
8	1790–1815	2	1	1	1815–1840	5104	3649	1455
9	1815–1840	1	1	0	1840–1865	1113	796	317
Total	1590–1840	311	157	154	1615–1865	463,098	301,405	161,693

NOTE: The mean number of African genealogical ancestors is rounded from Table 1 and divided equally between males and females, rounding up for males and down for females. The number of potential ancestors is based on counts of transported individuals from Hacker (2020) and sex ratios from Eltis and Engerman (1993).

(sum of “imported” in column 8 and “smuggled” in column 9) for two 10-year cohorts covered in the 25-year period and half that of the third cohort spanned by the 25-year period. Remainders from taking half an odd number of individuals are assigned to the earlier generation. All transported captives from 1610 to 1620 are counted in the 1615–1640 transportation cohort, and all transported captives after 1860 are assigned to the 1840–1865 transportation cohort. We report the values of the numbers of potential ancestors in Table 2.

With this approach, using (3) to calculate $\mathbb{P}[B_\ell]$, (5) gives $\mathbb{P}[B] \approx 0.2807$. Using (2) instead, (5) gives $\mathbb{P}[B] \approx 0.2810$. Table 3 reports the estimates of $\mathbb{P}[B_\ell]$, the probabilities for each generation of at least one shared ancestor born in that generation. The probabilities are highest in relatively early generations.

4.4. Generation-Specific Ancestors; Male and Female Ancestors Distinguished

Next, we proceed using both generation-specific and sex-specific ancestors. The genealogical ancestors in Table 2 are divided

Table 3. Probability of at least one shared enslaved migrant African ancestor for two African-Americans born during 1960–1965, with generation-specific computations.

Generation (ℓ)	Birth year of ancestor	$\mathbb{P}[B_\ell], (2)$	$\mathbb{P}[B_\ell], (3)$
0	1590–1615	0	0
1	1615–1640	0.0051	0.0051
2	1640–1665	0.0193	0.0193
3	1665–1690	0.0807	0.0805
4	1690–1715	0.1161	0.1160
5	1715–1740	0.0891	0.0890
6	1740–1765	0.0027	0.0027
7	1765–1790	0.0002	0.0002
8	1790–1815	0.0008	0.0008
9	1815–1840	0.0009	0.0009
Total	1590–1840	$\mathbb{P}[B]$ 0.2810	$\mathbb{P}[B]$ 0.2807

NOTE: Values of $k^{(1)} = k^{(2)}$ in birthday problem computations are taken from the generation-wise total mean number of African genealogical ancestors in Table 2, considering the generation-wise total number of potential ancestors in Table 2. These values are then used in (2) or (3). The total probability of at least one shared ancestor across all generations is computed by combining across rows to calculate (5).

evenly between males and females, assigning remainders of 0.5 to be male rather than female, again in order to conservatively estimate the probability of a shared ancestor.

To obtain the total generation-wise numbers of male and female potential ancestors, we begin with the generation-wise estimates of the number of potential ancestors N_ℓ in Table 2. We use sex ratios reported for different slave trade periods, as tabulated in Table 1 of Eltis and Engerman (1993). The number of males $N_\ell^{(m)}$ is obtained by multiplying the total number of individuals N_ℓ by constants $c_{16} = 0.603$ for transportation in 1700 or earlier, $c_{17} = 0.650$ for transportation during 1701–1809, and $c_{18} = 0.715$ for transportation during 1810–1865 (Eltis and Engerman 1993, Table 1), rounding fractional values of $N_\ell^{(m)}$ and $N_\ell^{(f)}$ up if they are greater than or equal to 0.5 and down otherwise, in the standard manner; we use $0.4c_{16} + 0.6c_{17}$ for the 1690–1715 transportation cohort and $0.8c_{17} + 0.2c_{18}$ for the 1790–1815 cohort.

Each $\mathbb{P}[B_\ell^{(m)}]$ and $\mathbb{P}[B_\ell^{(f)}]$ is then calculated from (2) or (3) and inserted into (4) to calculate $\mathbb{P}[B_\ell]$. Total values of $\mathbb{P}[B^{(m)}]$, $\mathbb{P}[B^{(f)}]$, and $\mathbb{P}[B]$ are then calculated by (5) from the generation-wise values.

Table 4 reports the estimates. We see that, as in the case of no distinction between male and female generation-specific ancestors (Table 3), the probability is nontrivial that at least one shared ancestor exists in each of generations 3, 4, and 5 and, to some extent, in generation 2. In other words, the probability of sharing an enslaved African ancestor is highest for ancestors born from 1640 to 1740.

The probabilities are nontrivial in generations 2–5 for sharing either a male or a female ancestor, but the sex-specific probabilities are also nontrivial. The probability of sharing a female ancestor exceeds that of sharing a male ancestor, owing to the fact that the number of imported enslaved males exceeds that for females, so that female ancestors are selected from a smaller pool of individuals. The unequal sex ratio leads to greater probabilities of sharing some ancestor when distinguishing between males and females than in the case that males and females are not distinguished (Table 3).

Table 4. Probability of at least one shared enslaved migrant African ancestor for two African-Americans born during 1960–1965, with generation-specific and sex-specific computations.

Generation (ℓ)	Birth year of ancestor	Using (2)			Using (3)		
		$\mathbb{P}[B_\ell^{(m)}]$	$\mathbb{P}[B_\ell^{(f)}]$	$\mathbb{P}[B_\ell]$	$\mathbb{P}[B_\ell^{(m)}]$	$\mathbb{P}[B_\ell^{(f)}]$	$\mathbb{P}[B_\ell]$
0	1590–1615	0	0	0	0	0	0
1	1615–1640	0.0021	0.0032	0.0053	0.0021	0.0032	0.0053
2	1640–1665	0.0081	0.0122	0.0202	0.0081	0.0122	0.0202
3	1665–1690	0.0340	0.0534	0.0856	0.0339	0.0533	0.0855
4	1690–1715	0.0463	0.0844	0.1268	0.0463	0.0843	0.1267
5	1715–1740	0.0360	0.0632	0.0969	0.0359	0.0631	0.0968
6	1740–1765	0.0010	0.0019	0.0029	0.0010	0.0019	0.0029
7	1765–1790	0.0001	0.0001	0.0002	0.0001	0.0001	0.0002
8	1790–1815	0.0003	0.0007	0.0010	0.0003	0.0007	0.0010
9	1815–1840	0.0013	0	0.0013	0.0013	0	0.0013
Total	1590–1840	$\mathbb{P}[B^{(m)}]$ 0.1232	$\mathbb{P}[B^{(f)}]$ 0.2028	$\mathbb{P}[B]$ 0.3010	$\mathbb{P}[B^{(m)}]$ 0.1231	$\mathbb{P}[B^{(f)}]$ 0.2025	$\mathbb{P}[B]$ 0.3006

NOTE: Values of $k^{(1)} = k^{(2)}$ in birthday problem computations are taken from the generation-wise male and female numbers of African genealogical ancestors in Table 2, considering the generation-wise male and female numbers of potential ancestors in Table 2. They are then used in (2) or (3) to obtain $\mathbb{P}[B_\ell^{(m)}]$ and $\mathbb{P}[B_\ell^{(f)}]$; $\mathbb{P}[B_\ell]$ is obtained generation-wise by (4). The total probability is computed by combining across rows to calculate (5).

4.5. Incorporating Uncertainty

The estimated mean numbers of African, European/European-American, and African-American genealogical ancestors in Table 1 represent mean values across 45,189 model parameter sets determined by Mooney et al. (2023) to be compatible with certain features of African-American genetic variation (mean African genetic ancestry 75%–85%, standard deviation 8%–15% across individuals). In this section, we return to those parameter sets (available in the supplementary data file for Agranat-Tamir, Mooney, and Rosenberg (2024)).

Each of the 45,189 parameter sets produces its own values of the mean number of African genealogical ancestors k (Figure 1(C)) and the numbers of genealogical ancestors k_ℓ in each generation ℓ . For each parameter set, we estimate the number of genealogical ancestors following the methods of Sections 4.3 and 4.4.

Figure 1 shows the distribution across the parameter sets of the probability of at least one shared ancestor, based on the expected value of the number of African genealogical ancestors calculated for each parameter set (Figure 1(C)). In Figure 1(A), using (2), pooling male and female ancestors together, the mean across the parameter sets of the probability is 0.2903, exceeding 0.2810, the probability calculated in Table 3 from the mean number of ancestors across the parameter sets, without taking the

individual parameter sets themselves into account. The median is 0.2609, and the distribution has a wide range, from 0.0438 to 0.8193, with interquartile range 0.1711–0.3848. When using (3), the results are similar: mean 0.2900, median 0.2606, range 0.0438–0.8184, and interquartile range 0.1709–0.3844.

When distinguishing male and female ancestors in Figure 1(B), using (2), as in the analyses that do not consider individual parameter sets, the mean probability across parameter sets is higher than when not distinguishing male and female ancestors, equaling 0.3091. It also slightly exceeds the value obtained without considering individual parameter sets, 0.3010, as calculated in Table 4. The median is 0.2796, the range is 0.0476 to 0.8436, and the interquartile range is 0.1842 to 0.4094. Using (3), the results are similar: mean 0.3088, median 0.2793, range 0.0476 to 0.8427, and interquartile range 0.1840 to 0.4089.

Figure 2(A) shows the distribution across the parameter sets of the probability of at least one shared ancestor for each generation 1–9. Because calculations according to (2) and (3) are similar in Figure 1, we proceed using only one of these equations, namely (3). In Figure 2(A), showing the distributions of the probabilities when we do not distinguish between male and female ancestors, the generation-wise means are 0.0070, 0.0234, 0.0873, 0.1243, 0.0942, 0.0029, 0.0003, 0.0015, 0.0024 for generations 1 to 9, respectively. As in Figure 1, the means

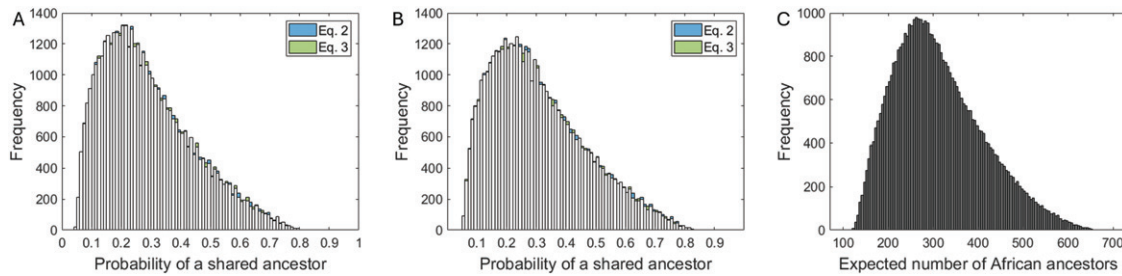


Figure 1. Distribution across demographic parameter sets of the probability of at least one shared ancestor for a pair of African-Americans born in the 1960–1965 cohort. (A) Distribution computed using (5) as in Table 3, without taking into account a distinction between males and females. (B) Distribution computed using (5) and (4) as in Table 4, taking into account a distinction between males and females. (C) Distribution of the expected number of African ancestors across the 45,189 demographic parameter sets used in the computations of panels (A) and (B) (modified from Mooney et al. 2023).

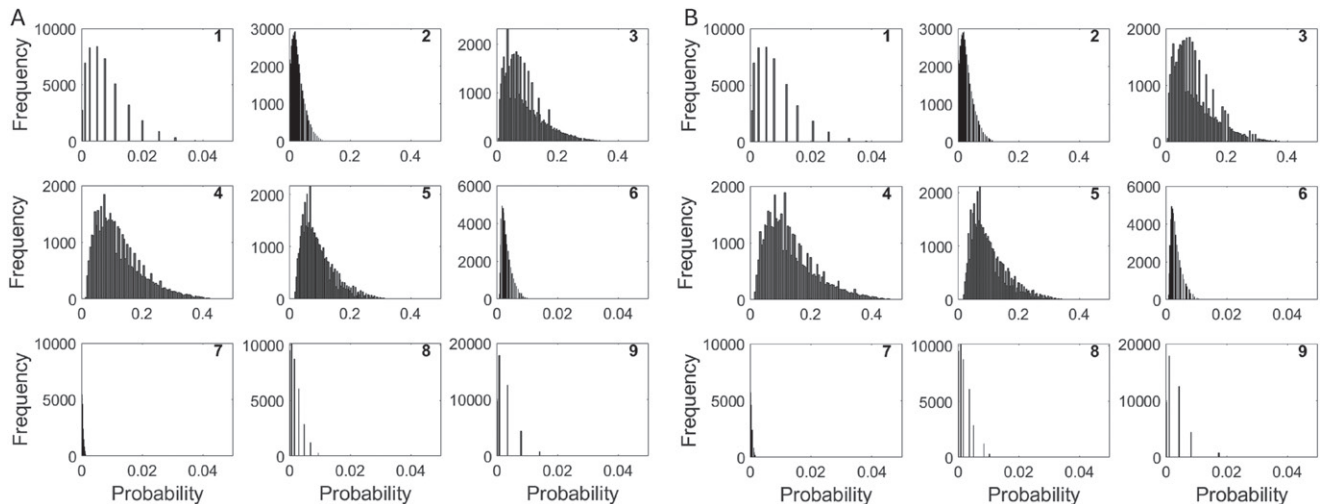


Figure 2. Distribution across demographic parameter sets of the probability of at least one shared ancestor for a pair of African-Americans born in the 1960–1965 cohort, computed separately in generations 1–9. (A) Distribution computed using (5) as in Table 3, without taking into account a distinction between males and females. (B) Distribution computed using (5) and (4) as in Table 4, taking into account a distinction between males and females. Both panels rely on (3). In each panel, the number for the generation depicted appears in the upper right corner.

exceed those obtained in Table 3 without taking into account the parameter sets. All distributions are right-skewed, with wide ranges particularly for generations with large values for the mean.

Distinguishing male and female ancestors, Figure 2(B) shows the distribution across parameter sets of the generation-wise probability of a shared ancestor. The patterns are similar to Figure 2(A), with slightly greater values: 0.0071, 0.0242, 0.0940, 0.1351, 0.1026, 0.0031, 0.0003, 0.0018, and 0.0028 for generations 1–9.

5. Discussion

We have studied the probability that two descendants share at least one ancestor from a larger collection of ancestors, treating the problem as an instance of the birthday problem. Using the birthday problem together with increasingly detailed assumptions about the demography of the African-American population, we have evaluated the probability that two randomly chosen African-Americans born in the years 1960–1965 possess at least one shared ancestor among Africans transported to North America. We estimate this probability at 19%–31%, with more detailed models producing values near the higher end of this range (Table 5).

5.1. Interpretation for African-American Genealogies

The analysis finds a relatively high probability that a pair of African-Americans share at least one African ancestor among individuals transported to North America during the slavery era—reminiscent of the relatively high probability of a shared birthday among students in a classroom. All the models place this quantity above 10%, and those that most carefully consider the population history estimate it near 30%.

We can comment on two features of likely shared ancestors. First, when we consider males and females separately, the probability of at least one shared female ancestor substantially exceeds the probability of at least one shared male ancestor (Table 5). This result traces to the fact that the number of transported males exceeded the number of transported females (Eltis and Engerman 1992, 1993), so that the number of “days of the

year” is smaller for females than for males, leading to a higher probability of a “shared birthday.”

Second, the probability of at least one shared ancestor is greatest in generations 2–5, covering birth cohorts from 1640 to 1740, with the highest probability occurring in generation 4, the 1690–1715 birth cohort (Tables 3 and 4, Figure 2). The probability of at least one shared ancestor is low both in the earliest generations 0–1, representing births prior to 1640, and in the later generations 6–9, corresponding to births from 1740 to 1840. These patterns can be understood in relation to the estimated mean number of African genealogical ancestors in a genealogy (Mooney et al. 2023) and the size of the transported population (Table 2).

A typical genealogy has few transported African ancestors in the earliest generations 0 and 1 (born 1590–1640), so that the probability that two individuals share an ancestor in those generations is small. In generations 4 and 5 (1690–1740), the number of transported Africans in a genealogy is large, so that the probability of a shared ancestor is high. Generation 3 has fewer transported African ancestors in a genealogy, but the transported population size is also smaller than in generations 4 and 5, so that the probability of at least one shared ancestor is comparable. In generations 6 and 7 (1740–1790), the transported population size is high, but the “classroom size” is smaller, so that the probability of a shared ancestor drops precipitously. It increases slightly in generations 8 and 9, which have lower values for the transported population size.

The results can be understood with a specific example. A discovery of shared ancestors is described in the book *The 272*, about 272 enslaved people whose sale from Maryland to Louisiana in 1838 by a branch of the Catholic church contributed to the financial support of Georgetown University (Swarns 2023). Living descendants Melissa Kemp, a biologist and PhD graduate of our Stanford department, and Jeremy Alexander, an employee of Georgetown University, discover that they have shared enslaved ancestors Harry Mahoney (born ~1771) and Anna Mahoney (born ~1775)—and an earlier shared ancestor Ann Joice, born free in 1660, likely to an unknown African mother and European father, and eventually enslaved in Maryland. One child of Harry and Anna Mahoney, Anna Mahoney Jones, born around 1811, is an ancestor of Jeremy Alexander and was among “the 272.” Her sister, Louisa Mahoney Mason, born in 1812, is an ancestor of Melissa Kemp; she escaped the sale and remained in Maryland.

The implication of our computations is that perhaps 19–31% of pairs of African-Americans from the 1960 to 1965 age cohort—even if they are unaware of any genealogical relationship—are similar to (somewhat younger) Jeremy Alexander and Melissa Kemp in having a shared ancestor such as Ann Joice. Like in this example, it is perhaps a female ancestor, and a person born between 1640 and 1740. For most pairs of living African-Americans, the existence of the shared ancestor can only be probabilistically postulated, rather than identified as a named person. Indeed, even in this comparatively well-documented case of descendants of Ann Joice—who was born in Barbados and transported to England and then to Maryland as an indentured servant—the shared African-born ancestor is the *mother* of Ann Joice, a person whose name has been lost.

Table 5. Summary of estimates.

Method	Probability of at least one shared ancestor		
	Male	Female	Any
Section 4.1, (2)	–	–	0.1908
Section 4.1, (3)	–	–	0.1907
Section 4.2, (2), (4)	0.0768	0.1460	0.2116
Section 4.2, (3), (4)	0.0767	0.1459	0.2114
Section 4.3, (2), (5)	–	–	0.2810
Section 4.3, (3), (5)	–	–	0.2807
Section 4.4, (2), (4), (5)	0.1232	0.2028	0.3010
Section 4.4, (3)–(5)	0.1231	0.2025	0.3006
Section 4.5, (2), (4), (5)	0.1335	0.2139	0.3091
Section 4.5, (3)–(5)	0.1334	0.2136	0.3088

NOTE: Each row estimates the probability that two African-Americans born in 1960–1965 share at least one enslaved African migrant ancestor, according to several different birthday problem computations.

5.2. Differences Among Different Methods

We have considered several versions of the birthday problem. Equation (2) assumes that the k ancestors of a person, representing the “classroom” in the birthday problem, are distinct individuals; equation (3) does not require these individuals to be distinct. Owing to this difference, numerical calculations according to (2) are predicted to give greater estimates of the probability of at least one shared ancestor than (3). Although this prediction is satisfied in our analyses (Tables 3–5), the difference between the two computations is generally negligible. Scenarios in which the African genealogical ancestors in a genealogy include many repeated individuals are unusual, and they have little effect on the overall probability.

When we consider a temporal model that accounts for the fact that later African entrants to the African-American population arrived in a population that already included descendants of earlier such entrants, the probabilities of at least one shared ancestor produce larger values (Table 5). This result is explained by a distribution of African ancestors across the generations that differs from the corresponding distribution of the potential ancestors. Generations 3–5 contribute 87% of African ancestors (Table 2), but only 54% of potential ancestors. The probabilities of shared ancestors in these generations exceed those that would be obtained if the distributions were identical. The generation-specific computation uses a more detailed model and may represent a more accurate estimate.

Similarly, distinguishing between males and females among the potential African ancestors leads to a greater probability of at least one shared ancestor, owing to the substantially smaller pool of African females than African males among potential ancestors (Table 5). The effect of distinguishing males and females in increasing the probability of at least one shared ancestor is smaller than that of separating ancestors by generation; the two effects combine to increase the probability from $\sim 19\%$ to $\sim 31\%$.

5.3. Limitations of the African-American History Model

In using the birthday problem to evaluate the probability that a pair of African-American descendants possess at least one shared transported enslaved ancestor, we have made several assumptions that may affect the value of the estimated probability and its meaning. First, the calculation via the birthday problem assumes that each potential ancestor in a specified category—by cohort, by sex, or both—is equiprobable as an actual ancestor. This assumption disregards variation over geographic space in the population of potential ancestors, which could produce systematic variation in the probability that a particular potential ancestor appears in the genealogy of a randomly chosen descendant. However, the sequences of migrations that have occurred among descendants of transported captives—south and west during slavery, and then north and further west during the Great Migration, for example—have reduced this structuring. Our model using the birthday problem gives an approximation that can be deepened by further considering the detailed history of this structure. For example, with careful genealogical analysis, census records of the formerly enslaved could provide probability distributions of the geographic locations of enslaved ancestors given the location of a modern

descendant, and of the locations of descendants given that of an ancestor; the latter computation may also be feasible for DNA from burials of enslaved individuals (Harney et al. 2023).

Second, although we have considered randomness in choosing actual ancestors from among potential ancestors, we have omitted other sources of randomness. For example, our calculations of the probability of repeated ancestors in the genealogy considered only the expected number of African ancestors given model parameters, rather than full distributions of genealogies for random pairs of descendants, conditional on the demographic model. Further, in our generation-specific analyses with (5), we have treated the probability of a shared ancestor in one specific generation to be independent of the corresponding probability for other generations. With additional uncertainty considered, potentially involving simulations of entire pedigrees rather than birthday-problem calculations, estimates of the probability of at least one shared transported ancestor might extend outside the range of 19%–31%.

Third, we have treated the probability of at least one shared transported African ancestor as interchangeable with the probability of at least one transported *enslaved* African ancestor. The estimation of the count of potential ancestors based on Hacker (2020) (the “number of days in the year” in the birthday problem) focuses on transported enslaved ancestors, so that the potential ancestors in our computations are enslaved ancestors. As is illustrated in the story of Ann Joice, some transported African individuals were not enslaved at the time of transportation. Adding non-enslaved African arrivals to counts of potential ancestors would enlarge the pool of potential ancestors, decreasing the birthday-problem probability of at least one shared ancestor. However, the size of the group of Africans arriving as non-enslaved was small. For example, in the earliest United States censuses, enslaved individuals greatly outnumbered the relatively few free African-Americans; further, free individuals generally were not new free arrivals, and instead were freed descendants of African ancestors who were originally transported as enslaved (and who are already tabulated in our pool of potential ancestors).

In general, model limitations trace to the profound loss of genealogical information that resulted from the practices of the slavery era. We can compare the African-American setting to that of populations known for extensive genealogical record-keeping—French Canadians (Anderson-Trocmé et al. 2023), Icelanders (Helgason et al. 2003), and Italians (Cavalli-Sforza, Moroni, and Zei 2004), for example—in which genealogical modeling efforts have potential to directly measure features of family structure, such as the distribution of the number of offspring in families and the ages of parents at the birth of their offspring. By contrast, African-American genealogical studies (Gates 2009; Swarns 2012; Nelson 2016) discuss many factors that have contributed to the loss of such genealogical data: failure to record enslaved persons by name in official records; regular sundering of families by enslavers, separating enslaved individuals from their forebears and hence from the genealogical information known to older family members; illiteracy among the enslaved, often due to significant personal risk required to learn to read, limiting their potential to record genealogical data themselves; and a reticence of the newly freed to discuss the period of enslavement, and by consequence, to

transmit genealogical information. Our reliance on assumptions such as a general uniformity of age at transportation and age at the birth of offspring trace to the lack of a documentary record, and it is meaningful that our multiple approaches to the birthday problem calculation produce results of comparable magnitude. A benefit of our calculations, building on previous studies that have incorporated slave trade data and modern genetic data to construct quantitative historical models (Mooney et al. 2023; Agranat-Tamir, Mooney, and Rosenberg 2024), is that they can probabilistically recover aspects of lost genealogies in aggregate.

Finally, we note that the ancestors considered in our birthday problem calculations are transported ancestors, and we do not examine the probability of a shared ancestor in the generations that precede a transported person in a pedigree. In addition, we have focused on the birthday problem for shared *genealogical* rather than *genetic* ancestors. Tracing back from the present, owing to features of the mechanism of genetic descent via finite segments of DNA, not every genealogical ancestor of a person is a genetic ancestor. By incorporating a model of the declining probability back in time that a genealogical ancestor is a genetic ancestor (Agranat-Tamir, Mooney, and Rosenberg 2024), a birthday-problem analysis similar to our genealogical analysis could be posed about genetic ancestors. With such a model, the birthday-problem probability of a shared genetic ancestor would be bounded above by the corresponding probability of a shared genealogical ancestor.

5.4. Conclusions

In this study, we have used the birthday problem of probability and statistics to understand the properties of genealogies, with ancestors viewed as students in a classroom and potential ancestors viewed as the days of the year. The analogy has contributed to illuminating features of the demographic history of the United States, by analysis of a demographic model of African-American genealogies.

A frequent formulation of the birthday problem asks for the minimum classroom size for the probability to exceed $\frac{1}{2}$ that two people in the class have a shared birthday. With (2) and the African-American population, the probability of at least one shared ancestor in the slavery period exceeds $\frac{1}{2}$ if the number of African ancestors k in a genealogy satisfies $k \geq 567$. As $k > 300$ for the mean number of African genealogical ancestors for descendants born in the 1960–1965 cohort, a value exceeding 567 is achieved one generation forward, for African-Americans born to two parents from the 1960–1965 cohort—who can be expected to have close to twice as many African ancestors as each parent. Hence, replacing $N = 365$ and $k = 23$ with $N = 463,098$ and $k = 567$, in pairs of African-American descendants born around 1990 to African-American parents born in the early 1960s, a “shared birthday” exists in their genealogies with probability above $\frac{1}{2}$.

In an incident that took place on the 1620 voyage of the *Mayflower* from England to what is now Massachusetts, a passenger John Howland fell overboard and was fortuitously rescued (Philbrick 2006, p. 32). This event is well-remembered for its genealogical implications, as Howland went on to have 10 children, 88 grandchildren, and possibly millions of eventual

descendants in total, so that 400 years later, a nontrivial fraction of pairs of individuals in the United States may have him as a shared ancestor in their pedigrees. The high probability of at least one shared transported ancestor for pairs of African-Americans suggests that there might exist analogous and similarly unlikely ancestors who would appear in thousands or millions of African-American pedigrees, were such pedigrees possible to reconstruct.

Disclosure Statement

The authors report that there are no competing interests to declare.

Funding

We acknowledge support from National Science Foundation grant BCS-2116322 and from a National Science Foundation Postdoctoral Research Fellowship in Biology.

ORCID

Lily Agranat-Tamir  <http://orcid.org/0000-0003-2211-0369>
Kennedy D. Agwamba  <http://orcid.org/0000-0002-4510-9427>
Jazlyn A. Mooney  <http://orcid.org/0000-0002-2369-0855>
Noah A. Rosenberg  <http://orcid.org/0000-0002-1829-8664>

References

- Agranat-Tamir, L., Mooney, J. A., and Rosenberg, N. A. (2024), “Counting the Genetic Ancestors from Source Populations in Members of an Admixed Population,” *Genetics*, 226, iyae011. [414,416,418,421]
- Anderson-Trocmé, L., Nelson, D., Zabad, S., Diaz-Papkovich, A., Kryukov, I., Baya, N., Touvier, M., Jeffery, B., Dina, C., Vézina, H., Kelleher, J., and Gravel, S. (2023), “On the Genes, Genealogies, and Geographies of Quebec,” *Science*, 380, 849–855. [414,420]
- Baharian, S., Barakatt, M., Gignoux, C. R., Shringarpure, S., Errington, J., Blot, W. J., Bustamante, C. D., Kenny, E. E., Williams, S. M., Aldrich, M. C., and Gravel, S. (2016), “The Great Migration and African-American Genomic Diversity,” *PLoS Genetics*, 12, e1006059. [414]
- Ball, W. W. R., and Coxeter, H. S. M. (1939), *Mathematical Recreations and Essays* (11th ed.), London: Macmillan. [413]
- Berlin, I. (2010), *The Making of African America: the Four Great Migrations*, New York: Viking. [414,415]
- Borja, M. C., and Haigh, J. (2007), “The Birthday Problem,” *Significance*, 4, 124–127. [413]
- Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., and Mountain, J. L. (2015), “The Genetic Ancestry of African Americans, Latinos, and European Americans across the United States,” *American Journal of Human Genetics*, 96, 37–53. [414]
- Caswell, H. (2019), “The Formal Demography of Kinship: A Matrix Formulation,” *Demographic Research*, 41, 679–712. [414]
- Cavalli-Sforza, L. L., Moroni, A., and Zei, G. (2004). *Consanguinity, Inbreeding, and Genetic Drift in Italy*, Princeton, NJ: Princeton University Press. [420]
- Chang, J. T. (1999), “Recent Common Ancestors of All Present-Day Individuals,” *Advances in Applied Probability*, 31, 1002–1026. [414]
- Chen, N., Juric, I., Cosgrove, E. J., Bowman, R., Fitzpatrick, J. W., Schoech, S. J., Clark, A. G., and Coop, G. (2019), “Allele Frequency Dynamics in a Pedigreed Natural Population,” *Proceedings of the National Academy of Sciences of the USA*, 116, 2158–2164. [414]
- Derrida, B., Manrubia, S. C., and Zanette, D. H. (2000), “On the Genealogy of a Population of Biparental Individuals,” *Journal of Theoretical Biology*, 203, 303–315. [414]
- Diaconis, P., and Mosteller, F. (1989), “Methods for Studying Coincidences,” *Journal of the American Statistical Association*, 84, 853–861. [413]

- Eltis, D., and Engerman, S. L. (1992), "Was the Slave Trade Dominated by Men?" *Journal of Interdisciplinary History*, 23, 237–257. [416,419]
- (1993), "Fluctuations in Sex and Age Ratios in the Transatlantic Slave Trade, 1663–1864," *The Economic History Review*, 46, 308–323. [416,417,419]
- Eltis, D., and Richardson, D. (2010), *Atlas of the Transatlantic Slave Trade*, New Haven, CT: Yale University Press. [414]
- Erlach, Y., Shor, T., Pe'er, I., and Carmi, S. (2018), "Identity Inference of Genomic Data Using Long-Range Familial Searches," *Science*, 362, 690–694. [414]
- Ertürk, M. S., Fitzpatrick, C., Press, M., and Wein, L. M. (2022), "Analysis of the Genealogy Process in Forensic Genetic Genealogy," *Journal of Forensic Sciences*, 67, 2218–2229. [414]
- Farquharson, K. A., Hogg, C. J., and Grueber, C. E. (2021), "Offspring Survival Changes Over Generations of Captive Breeding," *Nature Communication*, 12, 3045. [414]
- Felipe-Gonzalez, J. (2024), "Demography of the Transatlantic Slave Trade," *Oxford Research Encyclopedia of African History*. DOI:10.1093/acrefore/9780190277734.013.894 [415]
- Flajolet, P., Gardy, D., and Thimonier, L. (1992), "Birthday Paradox, Coupon Collectors, Caching Algorithms and Self-Organizing Search," *Discrete Applied Mathematics*, 39, 207–229. [413]
- Franklin, J. H., and Higginbotham, E. B. (2021), *From Slavery to Freedom: a History of African Americans* (10th ed.), New York: McGraw-Hill. [414]
- Galla, S. J., Brown, L., Couch-Lewis (Ngāi Tahu: Te Hapūo Ngāti Wheke, Ngāti Waewae), Y., Cubrinovska, I., Eason, D., Gooley, R. M., Hamilton, J. A., Heath, J. A., Hauser, S. S., Latch, E. K., Matocq, M. D., Richardson, A., Wold, J. R., Hogg, C. J., Santure, A. W., and Steeves, T. E. (2021), "The Relevance of Pedigrees in the Conservation Genomics Era," *Molecular Ecology*, 31, 41–54. [414]
- Gates, H. L. (2009), *In Search of Our Roots: How 19 Extraordinary African Americans Reclaimed Their Past*, New York: Crown Publishers. [415,420]
- Good, I. J. (1950), *Probability and the Weighing of Evidence*, London: Charles Griffin and Company Limited. [413]
- Graham, R. L., Knuth, D. E., and Patashnik, O. (2008), *Concrete Mathematics* (2nd ed.), Boston: Addison-Wesley. [413]
- Hacker, J. D. (2020), "From '20. and Odd' to 10 Million: The Growth of the Slave Population of the United States," *Slavery and Abolition*, 41, 840–855. [414,415,416,420]
- Harney, E., Micheletti, S., Bruwelheide, K. S., Freyman, W. A., Bryc, K., Akbari, A., Jewett, E., Comer, E., Gates Jr., H. L., Heywood, L., Thornton, J., Curry, R., Esselmann, S. A., Barca, K. G., Sedig, J., Sirak, K., Olalde, I., Adamski, N., Bernardos, R., Broomandkhoshbacht, N., Ferry, M., Qiu, L., Stewardson, K., Workman, J. N., Zalzal, F., Mallick, S., Micco, A., Mah, M., Zhang, Z., 23andMe Research Team, Rohland, N., Mountain, J. L., Owsley, D. W., and Reich, D. (2023), "The Genetic Legacy of African Americans from Catotian Furnace," *Science*, 381, eade4995. [420]
- Helgason, A., Hrafnkelsson, B., Gulcher, J. R., and Ward, R. (2003), "A Populationwide Coalescent Analysis of Icelandic Matrilineal and Patrilineal Genealogies: Evidence for a Faster Evolutionary Rate of mtDNA Lineages than Y Chromosomes," *American Journal of Human Genetics*, 72, 1370–1388. [420]
- Holst, L. (1986), "On Birthday, Collectors, Occupancy and Other Classical Urn Problems," *International Statistical Review*, 54, 15–27. [413]
- Lachance, J. (2009), "Inbreeding, Pedigree Size, and the Most Recent Common Ancestor of Humanity," *Journal of Theoretical Biology*, 261, 238–247. [414]
- Lemann, N. (1991), *The Promised Land: the Great Black Migration and How It Changed America*, New York: Random House. [415]
- MacCluer, J. W., Boyce, A. J., Dyke, B., Weitkamp, L. R., Pfenning, D. W., and Parsons, C. J. (1983), "Inbreeding and Pedigree Structure in Standardbred Horses," *Journal of Heredity*, 74, 394–399. [414]
- Mathis, F. H. (1991), "A Generalized Birthday Problem," *SIAM Review*, 33, 265–270. [413]
- McClelland, P. D. and Zeckhauser, R. J. (1982), *Demographic Dimensions of the New Republic: American Interregional Migration, Vital Statistics, and Manumissions, 1800–1860*, Cambridge: Cambridge University Press. [416]
- McKinney, E. H. (1966), "Generalized Birthday Problem," *The American Mathematical Monthly*, 73, 385–387. [413]
- Micheletti, S. J., Bryc, K., Esselmann, S. G. A., Freyman, W. A., Moreno, M. E., Poznik, G. D., Shastri, A. J., 23andMe Research Team, Beleza, S., and Mountain, J. L. (2020), "Genetic Consequences of the Transatlantic Slave Trade in the Americas," *American Journal of Human Genetics*, 107, 265–277. [414]
- Mooney, J. A., Agranat-Tamir, L., Pritchard, J. K., and Rosenberg, N. A. (2023), "On the Number of Genealogical Ancestors Tracing to the Source Groups of an Admixed Population," *Genetics*, 224, iyad079. [414,415,416,418,419,421]
- Nelson, A. (2016), *The Social Life of DNA: Race, Reparations, and Reconciliation after the Genome*, Boston: Beacon Press. [415,420]
- Pemberton, J. M. (2008), "Wild Pedigrees: The Way Forward," *Proceedings of the Royal Society B*, 275, 613–621. [414]
- Philbrick, N. (2006), *Mayflower*, New York: Viking. [421]
- Rohde, D. L. T., Olson, S., and Chang, J. T. (2004), "Modelling the Recent Common Ancestry of All Living Humans," *Nature*, 431, 562–566. [414]
- Severson, A. L., Carmi, S., and Rosenberg, N. A. (2021), "Variance and Limiting Distribution of Coalescence Times in a Diploid Model of a Consanguineous Population," *Theoretical Population Biology*, 139, 50–65. [414]
- Swarns, R. L. (2012), *American Tapestry: The Story of the Black, White, and Multiracial Ancestors of Michelle Obama*, New York: Amistad. [415,420]
- Swarns, R. L. (2023), *The 272: the Families Who Were Enslaved and Sold to Build the American Catholic Church*, New York: Random House. [419]
- The U.S.–Venezuela Collaborative Research Project and Wexler, N. S. (2004), "Venezuelan Kindreds Reveal that Genetic and Environmental Factors Modulate Huntington's Disease Age of Onset," *Proceedings of the National Academy of Sciences of the USA*, 101, 3498–3503. [414]
- Thompson, E. A. (2000), *Statistical Inference from Genetic Data on Pedigrees*, Beachwood, OH: Institute of Mathematical Statistics. [414]
- Von Mises, R. (1964), "Ueber Aufteilungs- und Besetzungswahrscheinlichkeiten," in *Selected Papers of Richard von Mises* (Vol. 2), eds. P. Frank, S. Goldstein, M. Kac, W. Prager, G. Szegő, and G. Birkhoff, pp. 313–334, Providence, RI: American Mathematical Society. [413]
- vonHoldt, B. M., Stahler, D. R., Smith, D. W., Earl, D. A., Pollinger, J. P., and Wayne, R. K. (2008), "The Genealogy and Genetic Viability of Reintroduced Yellowstone Grey Wolves," *Molecular Ecology*, 17, 252–274. [414]
- Wakeley, J., King, L., Low, B. S., and Ramachandran, S. (2012), "Gene Genealogies Within a Fixed Pedigree, and the Robustness of Kingman's Coalescent," *Genetics*, 190, 1433–1445. [414]
- Wendl, M. C. (2003), "Collision Probability between Sets of Random Variables," *Statistics & Probability Letters*, 64, 249–254. [414]
- Wilkerson, I. (2010), *The Warmth of Other Suns*, New York: Random House. [415]