

RESEARCH ARTICLE

WILEY



Tests of fit of historically-informed models of African American Admixture

Jessica M Gross 

Department of Anthropology MSC01-1040,
Anthropology 1, University of New Mexico,
Albuquerque, New Mexico

Correspondence

Jessica M. Gross, Department of
Anthropology MSC01-1040, Anthropology
1, University of New Mexico, Albuquerque,
NM 87131.

Email: jgross@unm.edu

Abstract

Objectives: African American populations in the U.S. formed primarily by mating between Africans and Europeans over the last 500 years. To date, studies of admixture have focused on either a one-time admixture event or continuous input into the African American population from Europeans only. Our goal is to gain a better understanding of the admixture process by examining models that take into account (a) assortative mating by ancestry in the African American population, (b) continuous input from both Europeans and Africans, and (c) historically informed variation in the rate of African migration over time.

Materials and methods: We used a model-based clustering method to generate distributions of African ancestry in three samples comprised of 147 African Americans from two published sources. We used a log-likelihood method to examine the fit of four models to these distributions and used a log-likelihood ratio test to compare the relative fit of each model.

Results: The mean ancestry estimates for our datasets of 77% African/23% European to 83% African/17% European ancestry are consistent with previous studies. We find admixture models that incorporate continuous gene flow from Europeans fit significantly better than one-time event models, and that a model involving continuous gene flow from Africans and Europeans fits better than one with continuous gene flow from Europeans only for two samples. Importantly, models that involve continuous input from Africans necessitate a higher level of gene flow from Europeans than previously reported.

Discussion: We demonstrate that models that take into account information about the rate of African migration over the past 500 years fit observed patterns of African ancestry better than alternative models. Our approach will enrich our understanding of the admixture process in extant and past populations.

KEYWORDS

admixture, African American ancestry, gene flow

1 | INTRODUCTION

Anthropologists and geneticists have studied the admixture process in African Americans for decades. These studies typically estimate the proportion of African and European ancestry in African American individuals and populations. Some of these studies attempt to fit simple admixture models to extant patterns of population genetic diversity in African Americans to determine when admixture first began and the per-generation contribution of Europeans (Baharian et al., 2016; Bryc, Durand, Macpherson, Reich, & Mountain, 2015; Glass and Li, 1953;

Gravel, 2012; Kidd et al., 2012; Long, 1991). The results of these studies have the potential to assist in uncovering the causes of multifactorial disease and to identify and eliminate the social causes of racial disparity in health outcomes. They also have the potential to help us understand how populations have interacted with one another throughout human history, particularly in cases that involve substantial power asymmetries between the populations.

In terms of ancestry estimates at the population and individual levels, the chief findings of these studies are that (a) the mean African ancestry level in extant African American populations is always

substantially higher than the mean European level (Bryc et al., 2015; Glass and Li, 1953; Long, 1991; Parra et al., 2001), (b) the Native American ancestry level is relatively low throughout the United States (Bryc et al., 2015; Parra et al., 1998; Tishkoff et al., 2009), (c) despite high mean African ancestry proportions at the population level, African ancestry varies substantially among individuals, from a low of about 2% to a high of about 98% (Bryc et al., 2015), and (d) more mating has occurred between European males and African or African American females than between African American males and European females (Bryc et al., 2015; Parra et al., 1998, 2001).

In terms of the dynamics of the admixture process, studies have produced ambiguous results. Bryc et al. (2015), for example, fit a discrete two-stage admixture model to distributions of ancestry tract lengths and concluded that a single admixture event occurred between Africans and a combined European-Native American group six generations before the present. Baharian et al. (2016) found that a two-pulse admixture model, in which Europeans contributed to the African American population on two separate occasions, fit better than a one-time admixture event model, and that the first incident of admixture occurred around 1740. Glass and Li (1953) assumed a model of continuous gene flow into the African American population from Europeans, and estimated a per-generation rate of European contribution to the African American population of 3.58%. More recently, Parra et al. (2001) concluded that a model of continuous one-way gene flow from Europeans with a per-generation gene flow rate of between 2.0% and 3.1% is consistent with the distribution of European alleles in African American individuals. Even more recently, Jin et al. (2012) fit four admixture models, including a continuous two-way gene flow model, to the distribution of ancestry tracts for a sample of African Americans. The authors constrained their admixture onset time to be between 10 and 17 generations in the past in an effort to be consistent with African American history, and found that one-way gene flow from Europeans into the African American population with an onset of 14 generations in the past produced the best-fit distribution of ancestry tracts.

The results of these studies are broadly incompatible with one another, and, for the most part, they are inconsistent with our current understanding of African American history. Although there is a paucity of historical data about the timing and amount of admixture over the past 500 years, the historical record provides information that can be used to constrain the parameters of admixture models. We know, for example, that the ancestors of African Americans came from diverse locations in Africa, including areas where admixture with people from other regions occurred, such as regions bordering the Mediterranean. However, the vast majority of immigrants both during the slave trade and in recent years came from West Africa (Curtin, 1969; U.S. Census Bureau, 2010; Voyages Database, 2009). We also know that the first importation event to North America involved 20 slaves in Virginia in 1619 (Curtin, 1969). After this initial event, slave importation remained low until the beginning of the 18th century (see Figure 1), after which it continued unabated until 1860, even after importation became illegal in 1808 (Curtin, 1969; Smith, 1973). Reports made by slaves discuss forced mating between slaves and slave owners throughout African

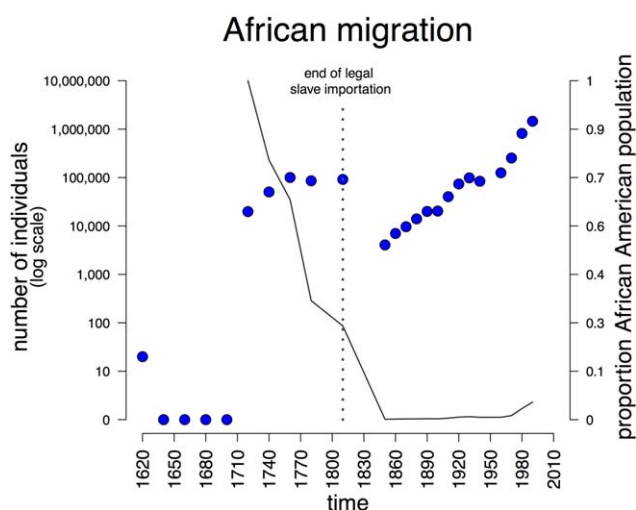


FIGURE 1 African migration. The blue points correspond to the vertical scale on the left, the number of African migrants. The solid line corresponds to the vertical axis on the right, the proportion of the African American population these migrants represent (the axis values are rounded to the nearest tenth). This line accounts for both the intrinsic rate of increase, as well as actual numbers of individuals. The dotted line is at 1808, and marks the end of legal slave importation; the points to the left of the dotted line represent people who were forcibly brought to North America to be slaves

American history (Federal Writers Project, 2001). Additionally, the passage of anti-miscegenation laws as early as 1664 suggest that African American-European mating occurred early in U.S. history (General Assembly of Maryland, 1664). More recently, the 2010 Census reports 314,400 immigrants from Northern African and 2,847,199 new immigrants from Sub-Saharan Africa (U.S. Census Bureau, 2010). Based on this information, it is likely that admixture between Europeans and African Americans began as early as 12 generations before the present and that it occurred continuously afterwards. It is also likely that the rate of gene European gene flow into the African American population varied dramatically over the past 500 years as a result of important historical events such as the U.S. Civil War, the passage of anti-miscegenation laws (pre- and post-civil war), the Great Migration, and the passage of civil rights legislation.

Additionally, records of slave importation indicate that the contributions to African American populations from newly migrated Africans must have been large and persistent (Figure 1). Remarkably, to date, with the exception of those fit by Jin et al. (2012), models of the admixture process have ignored this African contribution. Jin et al. (2012), however, did not ground the per-generation source-group contributions in history. Rather, they used observed ancestry fractions, and divided the contribution from each source population evenly across the generations in their model.

This study builds on the work of Jin et al. (2012). Our novel contributions include the formal fitting of ancestry models to observed distributions of individual ancestry in three African American samples, the use of census and other historical records to vary the model-based per generation contribution of Africans to the African American population, and the incorporation of assortative mating by ancestry in the African

American population, which we will call “ancestry-related assortative mating” (AAM). AAM refers to a correlation in ancestry between mates. Such mating may have been common among African Americans, due, for example, to geographic structure in the distribution of newly imported slaves, or selective mating by phenotype among slave owners (Federal Writers Project, 2001). Although there is no historical information about AAM in African Americans, it has been reported for Hispanic populations in Mexico, the Bay Area of San Francisco, and Puerto Rico. In Mexican populations in Mexico City and San Francisco, Risch et al. (2009) found that the correlation in Native American ancestry between mates was 0.586 and 0.392, respectively. The same authors found that the correlation in African ancestry between mates was 0.328 in Puerto Rico. Interestingly, they were unable to identify the social mechanism for AAM. AAM is relevant to the study of African American admixture because it has the potential to affect estimates of the European contribution to the African American population, a parameter for which we have no direct historical information. By adding AAM to our analyses, we hope to refine our understanding of both the social causes of AAM as well as the European contribution to the African American population.

The goals of this study are to estimate African and European genetic ancestry in three African American samples and to compare the fit of discrete and one-way continuous models of admixture to historically informed admixture models that incorporate (a) continuous contributions from Africans and Europeans, (b) variable rates of per-generation contribution from Africans, and (c) assortative mating by ancestry in the African American population.

2 | METHODS

2.1 | Data

We use two datasets for our analyses. The first consists of 1,022,144 autosomal SNP genotypes from 112 Yoruban (YRI), 110 CEPH European (CEU), and 83 adults who self-identified primarily as African American from HapMap Phase 3 (ASW, African ancestry in SW USA) (The International HapMap Consortium 2003). All ASW stated that they had four African American grandparents from the U. S. Southwest.

We filtered the ASW sample to remove related individuals. For this step, we excluded children from parent–offspring trios and duos, resulting in a sample of 49 unrelated individuals. We filtered the SNPs in two ways. First, we retained SNPs that were common to the YRI, CEU, and ASW. This step reduced the number of SNPs to 992,601. Second, following the work of Pfaff, Barnholtz-Sloan, Wagner, and Long (2004), we limited our analyses to SNPs that were informative about the admixture process. Those authors showed that marker informativeness, captured by Fisher Information (FI), is a function not only of differences in allele frequencies between putative parental populations, but also of the allele frequencies themselves. Specifically, we used an FI cutoff of 2.5, which, while arbitrary, eliminated thousands of uninformative loci without inflating the error in individual ancestry estimates (see below). Data filtration was conducted in R.

TABLE 1 Sample sizes and number of loci in the three African American samples

Sample	<i>n</i>	Loci
ASW	49	7,392 SNPs
AME	98	645 STRs
ABT	44	645 STRs

The second sample consists of 645 autosomal short tandem repeat genotypes from 50 Yoruba, 29 French, and 98 self-identified African American individuals from four locations in the U.S. Midwest and East Coast: Baltimore, Chicago, North Carolina and Pittsburgh (Tishkoff et al., 2009); we refer to this sample as African Americans in the Midwest and East Coast (AME) We performed our analyses on both the full AME sample and separately on the Baltimore sample of 44 individuals (ABT) The Baltimore sample is the largest of the African American samples in the Tishkoff et al. (2009) dataset. We analyzed this sample separately to control for the possibility that the AME sample is structured with respect to ancestry simply because it is comprised of individuals from multiple geographic locations, each of which may have experienced a different admixture history.

The YRI and CEU samples served as parental source populations for analyses of the ASW, and the Yoruba and French samples served as parental source populations for analyses of the AME and ABT samples. We recognize that these samples are not the true parental source populations, which derived from diverse locations in Europe and Africa (Montinaro et al., 2015; Patin et al., 2017). For this reason, our individual-level ancestry estimates, and the mean population-level estimates, are unlikely to be accurate. This limitation is common to all admixture studies. In our study, this limitation could affect our estimates of the per generation contribution of the Europeans and Africans to the admixed population (see below). However, the absence of true parental sources is unlikely to affect the *shape* of the observed ancestry distributions because the error would be systematic, and therefore unlikely to change our conclusions about the relative fit of different models of the admixture process.

2.2 | Ancestry estimation and model construction

Our strategy for estimating African and European ancestry and for comparing different admixture models consisted of three steps. First, we estimated individual-level African and European ancestry in the ASW, AME, and ABT samples using the Bayesian model-based

TABLE 2 Model parameters each generation

Model	g_0	g_{1+}
1. One-time admixture	α, β	...
2. One-way gene flow	α, β	α, β
3. Two-way gene flow	α, β	α, β
4. Two-way gene flow + AAM	α, β, R	α, β, R

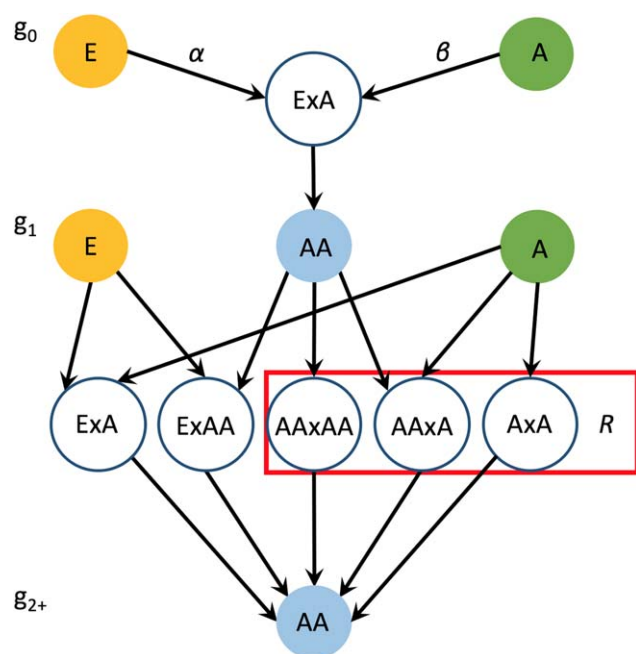


FIGURE 2 Generalized admixture model. A = African (green circles), E = European (yellow circles), AA = African American (blue circles). The African American population forms at g_1 from the fractional contributions of the European and African source populations, α_0 and β_0 , at time g_0 . R is the correlation in ancestry between mates, which affects the mating pairs in the red box

clustering algorithm implemented in the program STRUCTURE (Pritchard, Stephens, & Donnelly, 2000). For each sample, we ran STRUCTURE five times at $K = 2$ through $K = 6$ using a burnin phase of 25,000 steps and 15,000 MCMC repetitions. Otherwise, we used the default settings in STRUCTURE. Second, we fit four models to the distributions of individual African ancestry (see Table 2).

Figure 2 shows a generalized version of the four models. Table 3 shows the ancestry of possible mating pairs in each generation of Models 3 and 4 after the first generation.

European \times European matings were excluded from the models. Thus, under this mating scheme, all individuals in the African American population at any given time must have at least one African ancestor.

Three parameters were associated with each model: the number of generations, g , since the onset of admixture, and the contributions from the European and African source populations α and β , respectively. A fourth parameter, R , the correlation in ancestry between mates, was associated with Model 4.

TABLE 3 Possible mating pairs after generation 1 in each model

Mating pairs	Models
African American \times African American	1–4
African American \times African	3 and 4
African \times African	3 and 4
African American \times European	2–4
European \times African	3 and 4

For all models, we assumed that admixture began $g = 12$ generations ago, in approximately 1700, when there was a surge in slave importation. For the one-time admixture event model (Model 1), the parameters α and β were set to zero after the initial admixture event at generation g_0 . Under this model, α_0 and β_0 are equal to the ancestry fractions in the current African American samples (ASW, AME, and ABT). For the one-way gene flow model, α was set to a constant rate per generation, β_0 was set to $1 - \alpha$, and β was set to zero for all subsequent generations. For the two-way gene flow models, we used the African slave import estimates from Curtin (1969) to approximate the per-generation African contribution, β , during the slave trade. These estimates are based on English slave trade data and estimates of slave trade importation (Curtin, 1969); thus the rates change each generation. We used information from the U.S. Census (Gibson and Lennon, 1999) on native-born African Americans vs. African immigrants to determine the β -values for each generation subsequent to the cessation of legal slave importation.

2.3 | Generating model-based individual African American ancestry distributions

For each model, we produce an expected distribution of individual African ancestry (IA) as follows. Our measure of African ancestry is an individual's number of African ancestors in a given generation. For a given model, the IA distribution for the first generations was formed by "mating" Europeans and Africans in the proportions α and β ($= 1 - \alpha$). In the first generation of existence, g_1 , individuals in this newly formed African American population could have either 50% or 100% African ancestry, that is, either one or two African ancestors. In subsequent generations, IA distributions were created by drawing mating pairs from the IA distribution in the previous generation in accord with the relevant model parameters (Tables 2 and 3). Thus, for all models, at any given generation, African American individuals had any of a discrete number of African ancestors, corresponding to African ancestry proportions between $1/(2^g)\%$ and 100%.

For Models 1–4 for each sample, the parameter α was set to values that produce the observed mean European ancestry in that sample after 12 generations. To achieve this outcome for Model 1, α was zero after generation 1. For Models 2–4, α was constant across generations.

2.4 | Ancestry-related assortative mating

For Model 4, we used a Monte Carlo simulation approach to incorporate AAM into our mechanistic admixture model. For each simulation, we began by creating a vector of 20,000 pairs of individual ancestry estimates, with each estimate being drawn at random (with replacement) from the previous generation's IA distribution. In this way, we effectively created a series of randomly mating couples. We then (a) calculated the correlation in ancestry between mates in the vector, (b) permuted one member of each of two mating pairs, (c) recalculated the correlation for the newly permuted vector, and (d) compared the new correlation to original correlation. At each step in the process, we retained permuted vectors that had a correlation in ancestry between

mates that was closer to our target correlation, and reverted back to the previous vector when the permutation produced a correlation that was further from our target. In this way, we were able to create new versions of Models 1–3 that incorporated correlations in ancestry between mates (R) ranging from 0 to 1 in increments of 0.01.

2.5 | African contribution to the African American population

In Models 3 and 4, which include continuous African gene flow into the African American population, for each generation during legal slave importation, the proportion of immigrant Africans was calculated as the number of imported slaves divided by the sum of the number of imported slaves including and prior to that generation. The post-1850 data were taken from the U.S. Census Bureau (Gibson and Lennon, 1999), which reports the number of native-born African Americans and the number of African immigrants in 10 year increments beginning in 1850. Since we used a generation time of 25 years, we calculate population sizes for each generation by combining population data for two 10-year census reports, then adding the midpoint value from a third report. β -values were then calculated as the number of African immigrants for each generation divided by the total African American population size according to the U.S. Census.

2.6 | Changes in the IA distributions over time

We iterated each of the four models until the expected distributions reached a steady-state to better understand the change in the expected IA distribution over time. For the one-time admixture event model (Model 1), we began with contributions $\alpha = 78\%$ European and $\beta = 1 - \alpha = 22\%$ African. For the one-way gene flow model (Model 2), we set α at 0.05 per generation. For the models involving two-way gene flow (Models 3 and 4), α was set at a constant-rate of 0.05, and β was set according to slave importation and U.S. Census data for the first 12 generations, then set to a constant rate of 0.05 for all remaining generations.

2.7 | Testing the fit of the models to the observed distribution of African American ancestry

We tested the fit of each model-based distribution to the IA distribution for each of the three samples (ASW, AME, and ABT). Based on historical information described in the introduction, we assume that admixture began 12 generations ago, and we compare the fit of each model at this 12-generation point. Our method for comparing the fit of the models was as follows.

Because individuals in admixed populations have a discrete number of ancestors from the parental source populations, for our statistical tests of fit, we first divide the individuals in each observed sample into 16 bins from $1/(2^3)\%$ to 100% African ancestry. Similarly, for each model, we binned the model-based probabilities into 16 bins from $1/(2^3)\%$ to 100% African ancestry.

We calculate the log-likelihood of multinomial cell probabilities for the IA distribution produced by each model using Equation 1,

TABLE 4 Mean and range of African American ancestry estimates for the ASW, AME, and ABT samples

Sample	Mean African ancestry	Individual African ancestry range
ASW	76.53%	58.43%–91.74%
AME	80.59%	45.25%–98.40%
ABT	83.06%	62.60%–98.40%

$$\ln(L) = \sum_{i=1}^m x_i \ln(p_i) \quad (1)$$

where m is the number of bins in the distribution ($m = 16$), x_i is the number of individuals in bin i in the observed distribution, and p_i is the probability of the number of African ancestors in bin i in the expected distribution.

After calculating the log-likelihood for each of the expected distributions, we assess their fit relative to one another using a likelihood ratio statistic, LLR (Equation 2) (Sokal and Rohlf, 2012).

$$LLR = -2(\ln(L_0) - \ln(L_1)) \quad (2)$$

For large samples, the distribution of LLR is approximated by the χ^2 distribution. In each case, we used the best-fit distribution of the previous model as the null hypothesis ($\ln(L_0)$).

3 | RESULTS

3.1 | Ancestry estimates

All but 7,392 of the SNPs in the full HapMap dataset had FI values below 2.5 (see Supporting Information Figure S1). This result is unsurprising given the relatively young age of our species and the fact that a small amount of the total variation in our species is unique to populations and regions (Rosenberg et al., 2002). Nonetheless, the remaining 7,392 loci were highly informative, resulting in average individual-level error estimates of $\pm 0.6\%$ (based on the 95% credible regions estimated in STRUCTURE).

Our STRUCTURE runs showed average individual-level Native American ancestry estimates of 0.08% at $K = 3$, 0.4% at $K = 5$, and 0.6% at $K = 6$ for the Tishkoff et al. (2009) dataset, which includes the AME and ABT samples (see Supporting Information Figure S2). The HapMap dataset does not include a Native American sample. However, previous studies have shown the average Native American ancestry to be $<1\%$ among African Americans across the United States, and $<2\%$ in African Americans in the U.S. Southwest (Bryc et al., 2015; Jin et al., 2012). Based on these results, we conduct our analyses under the assumption of dihybrid ancestry.

The mean and range of African American ancestry estimates for each sample are listed in Table 4. The mean estimates are consistent with those from previous studies of African American ancestry (Bryc et al., 2010, 2015; Falush, Stephens, & Pritchard, 2003; Glass and Li, 1953; Kidd et al., 2012; Oksenberg et al., 2004; Parra et al., 1998; Tishkoff et al., 2009).

The observed IA distribution for each sample are shown in Figure 3. Two predominant features of all three distributions are a relatively

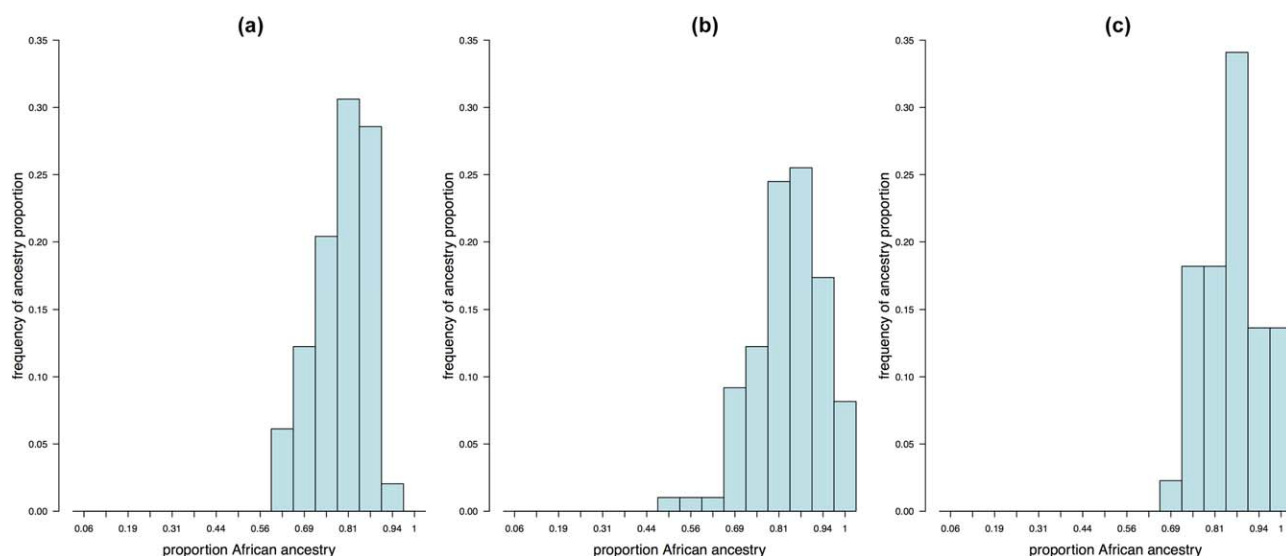


FIGURE 3 Individual ancestry distributions. (a) ASW, (b) AME, (c) ABT

high mean African ancestry level and a left skew. Differences between the distributions include an absence of individuals in the highest ancestry bin in the ASW sample, and a greater left skew towards low African ancestry in the AME sample compared with the other two samples.

3.2 | Model expectations over time

Figure 4 shows the change in the expected IA distribution for each model at 4, 12, 75, and 125 generations after the onset of admixture. We chose to begin at four generations because it produced the highest likelihood under Model 1, the one-time admixture event model, when fit to all three samples. Under this model, variation in African ancestry is lost each generation; by 20 generations, everyone in the population has a single African ancestry value that is equivalent to the original contributions from Africans and Europeans. Under Model 2, continuous one-way gene flow from Europeans and no contribution from newly migrated Africans, variation in African ancestry is reduced each generation until, by 125 generations, everyone has 0% African ancestry. Under Models 3 and 4, involving continuous gene flow from both Europeans and Africans, variation in African American ancestry is maintained among individuals over time, and the distribution eventually comes to a steady-state by 75 generations.

It is important to note that, even though the IA distributions for the models differ from one another over time, in some cases markedly, with the exception of Model 1, the expected IA distributions are similar at 12 generations. These results suggest that our statistical tests of fit will lead to a clear rejection of Model 1, but it may be considerably more difficult to distinguish between the remaining models.

3.3 | Model comparisons

Table 5 shows the log-likelihoods for each model (row 1), LLRs comparing the fit of the models (row 2), p -values for the LLR tests (row 3), α -values for Models 2–4 (row 4), and R -values for Model 4 (row 5). For all three samples, as predicted from the IA distributions in Figure 4,

Model 1 has by far the lowest log-likelihood (poorest fit), and Model 2 fit better than Model 1 (significantly higher log-likelihood at $p < .05$). For Models 2–4, the α -values (European contribution) that produced the highest log-likelihoods ranged from 0.037 (ABT, Model 2) to 0.071 (ASW, Model 4). The R -values for Model 4 ranged from 0.01 for the ASW sample to 0.15 for the ABT sample.

3.4 | Sample-specific results

Figure 5 shows the expected distributions for the best-fitting versions of Models 1–4 (at 12 generations) in pink and the observed IA distribution in pale blue for ASW. Figure 5a shows that Model 1 fits poorly because the observed ASW IA distribution retains substantial variation in African ancestry compared with that predicted under a one-time model. Model 2 (Figure 5b), involving continuous gene flow from Europeans, maintains substantially more variation in African ancestry after 12 generations, and therefore fits substantially better than Model 1. The differences in fit between the remaining models are more subtle. Model 3, involving continuous gene flow from both Europeans and Africans fits significantly worse than Model 2. Adding AAM (Model 4, Figure 5d), however, did not result in a difference in fit relative to Model 3.

One noteworthy feature of the expected IA distributions for Models 2–4 is the slight swell in frequency in the left tail of the distribution around 0.44. This swell is absent from the observed ASW distribution. We return to this issue in the discussion.

For the AME sample (Figure 6), Model 1 again fits poorly, and Model 2 fits significantly better than Model 1. In contrast to the ASW sample, Model 3 fits significantly better than Model 2, and Model 4, with AAM, fits significantly better than Model 3. The correlation in ancestry between mates, R , that produced the best fit for Model 4 was 0.08. Again, the expected distributions for Models 2–4 have a slight swell in frequency around 0.44 that is absent in the observed IA distribution.

The pattern of fit for the ABT sample, shown in Figure 7, is the same as that for the AME. In this case, the correlation in ancestry that

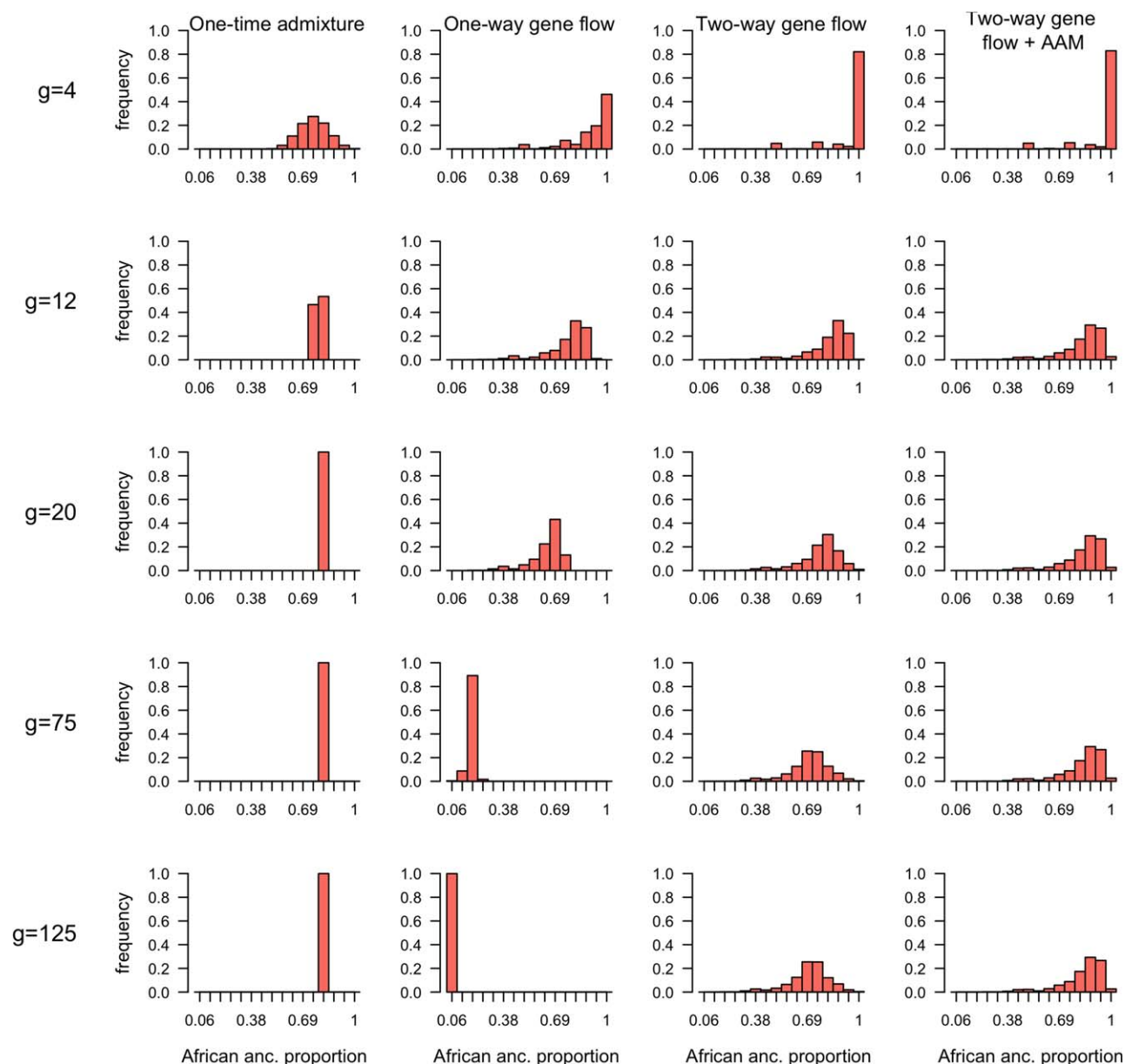


FIGURE 4 Expected model distributions over time. (a) One-time admixture event, (b) One-way gene flow, (c) Two-way gene flow, (d) Two-way gene flow with AAM. The choice of values for the generations were based on the best fit (highest likelihood) for Model 1 (4 generations), historical information about the onset of admixture (12 generations), the time to steady state at 80% African ancestry for Model 1 (20 generations), the time to steady state for Models 3 and 4 (75 generations), and the time to steady state at 0% African ancestry for Model 2 (125 generations)

produced the best fit for Model 4 was 0.15, almost twice as high as it was for the AME. Again, the observed IA distribution lacks the slight swell in frequency around 0.44.

4 | DISCUSSION

In this study, we fit four admixture models to the distribution of IA in three samples of American Americans. The models took into account historical information about immigration. In all models, the admixture process began 12 generations ago, approximating the onset of high

rates of slave importation into North American beginning around 1700 (Curtin, 1969). Models 2 - 4, in accord with slave reports, other historical accounts, and U.S. Census data (Federal Writers Project, 2001; Gordon-Reed, 1998), incorporated ongoing gene flow from Europeans. Models 3 and 4 incorporated ongoing gene flow from newly immigrated Africans in accord with historical data on slave importation (Curtin, 1969) and the U.S. Census Bureau (Gibson and Lennon, 1999). To our knowledge, to date, no study of genetic admixture has incorporated ongoing African input into the African American population, despite the fact that it may have been the most prevalent source of gene flow into the African American population for generations after

TABLE 5 Model-fitting results for the ASW, AME, and ABT samples at $g = 12$ generations

ASW	Model 1: one-time admixture	Model 2: one-way gene flow	Model 3: two-way gene flow	Model 4: two-way + AAM
<i>lnL</i>	−526.253	−81.425	−84.326	−83.838
<i>LLR</i>	...	889.656 ^a	−5.802 ^a	0.974
<i>p</i>	...	1.74E-195	0.016	0.324
α	...	0.051	0.068	0.071
<i>R</i>	0.010
AME	One-time Admixture	One-way gene flow	Two-way gene flow	Two-way + AAM
<i>lnL</i>	−1369.914	−335.781	−190.560	−188.435
<i>LLR</i>	...	2068.266 ^a	290.444 ^a	4.251 ^a
<i>p</i>	...	0.000	3.98E-65	0.039
α	...	0.042	0.057	0.059
<i>R</i>	0.080
ABT	One-time Admixture	One-way gene flow	Two-way gene flow	Two-way + AAM
<i>lnL</i>	−458.480	−184.888	−83.263	−80.832
<i>LLR</i>	...	547.184 ^a	203.251 ^a	4.861 ^a
<i>p</i>	...	5.16E-121	4.08E-46	0.027
α	...	0.037	0.050	0.057
<i>R</i>	0.150

lnL, log likelihood; *LLR*, log likelihood ratio; *p*, *p*-value; α , best-fit European contribution; *R*, best-fit correlation in ancestry between mates.

^a $p < .05$.

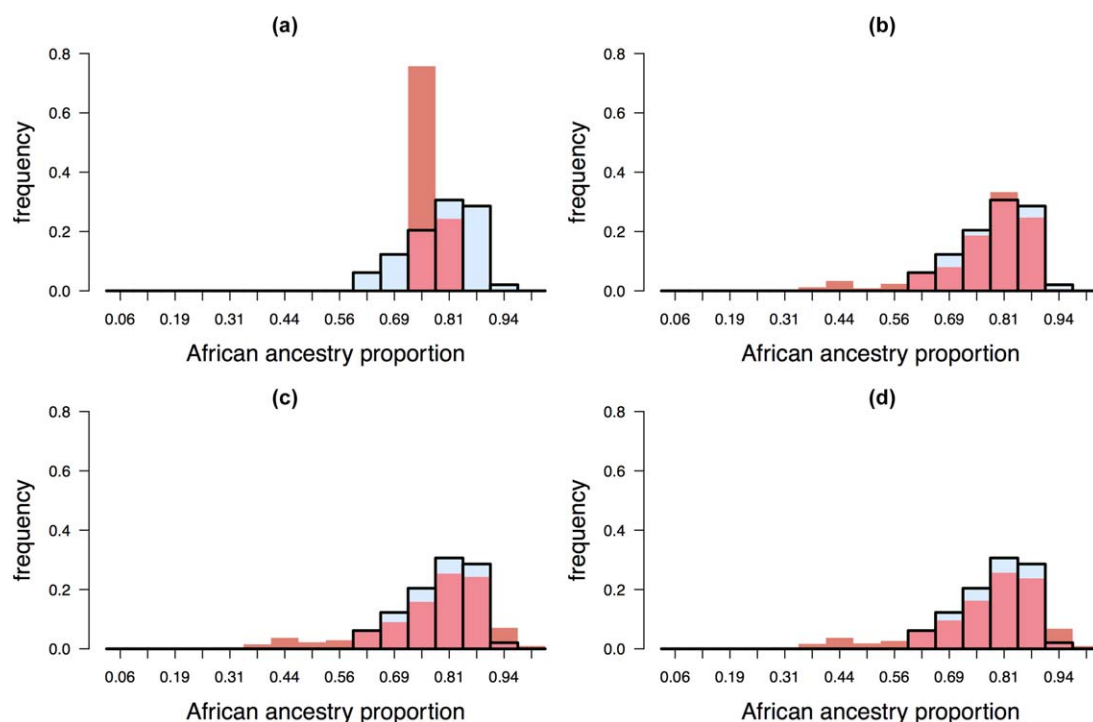


FIGURE 5 ASW. IA distributions for 12 generation models. (a). One-time admixture, (b) One-way gene flow, (c) Two-way gene flow, (d) Two-way gene flow with AAM. Model distributions are shown in pink, and the observed distribution is shown as transparent blue bars with thick black outlines

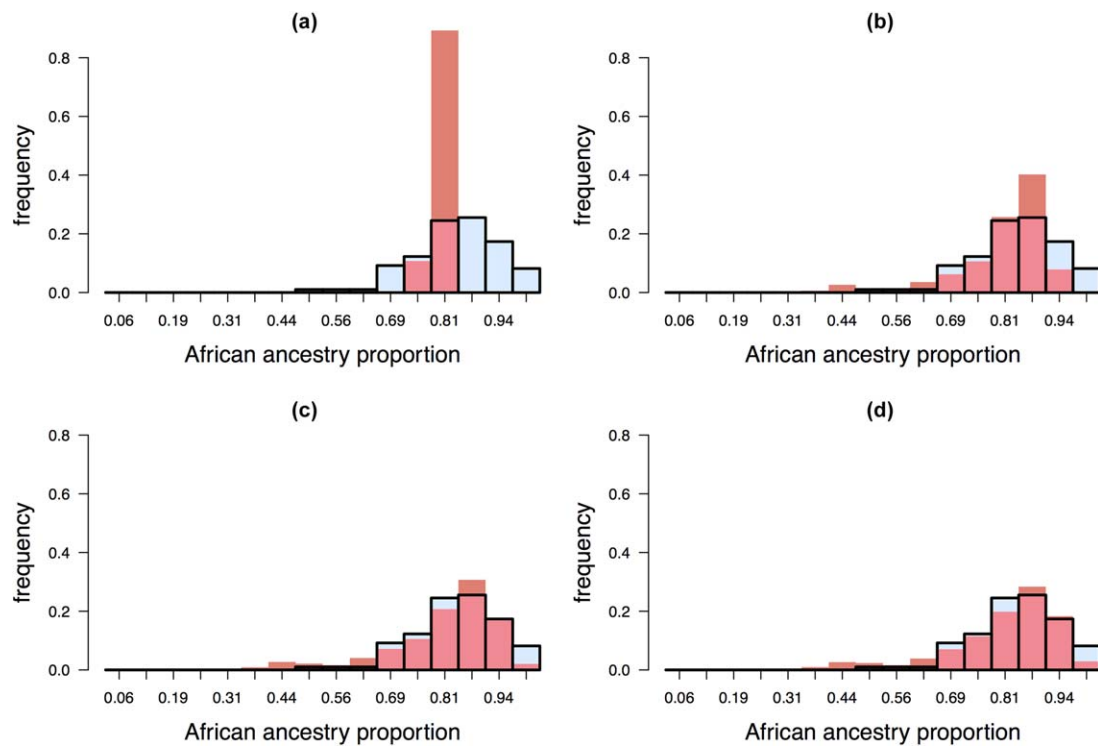


FIGURE 6 AME. IA distributions for 12 generation models. (a) One-time admixture, (b) One-way gene flow, (c) Two-way gene flow, (d) Two-way gene flow with AAM. Model distributions are shown in pink, and the observed distribution is shown as transparent blue bars with thick black outlines

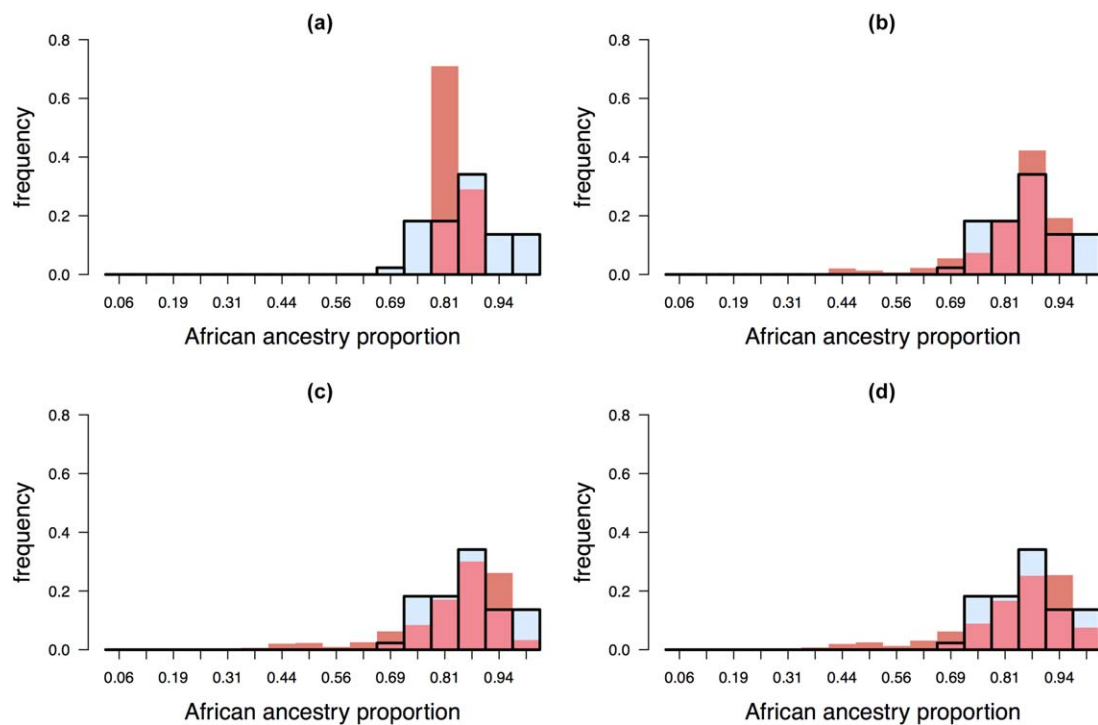


FIGURE 7 ABT. IA distributions for 12 generation models. (a) One-time admixture, (b) One-way gene flow, (c) Two-way gene flow, (d) Two-way gene flow with AAM. The model distributions are shown in pink, and the observed distribution is shown as transparent blue bars with thick black outlines

the onset of slavery (see Figure 1) (Curtin, 1969; Gibson and Lennon, 1999). To our knowledge, this is also the first study to incorporate AAM into models of the admixture process in a population. Although there is no historical information about assortative mating by ancestry in African Americans, it has been documented in Hispanic populations (Risch et al., 2009).

We found that, in all cases, the one-time admixture event model fit the observed IA distributions poorly relative to other models, which involve ongoing gene flow from one or both source populations. The better fit is, in part, a result of the fact that ongoing gene flow maintains variation in ancestry in admixed populations; this variation erodes rapidly under a model involving a single admixture event (Verdu and Rosenberg, 2011). A one-time admixture event model is also inconsistent with slave narratives that describe rape by white slave owners and abortion attempts for the potential offspring of European-African/African American unions (Federal Writers Project, 2001), and it is inconsistent with data from the U.S. Census Bureau. "Mulatto", for example, was a category on the U.S. Census from 1850 to 1890 and again from 1910 to 1920. More recently, the U.S. Census has allowed individuals to choose multiple races. According to the 2010 U.S. Census, 9.0 million U.S. residents reported multiple-race ancestry (Jones and Bullock, 2012). Of these, about 2.3 million individuals chose combinations involving white and black race. These data may underreport mixed race ancestry in the United States; an independent analysis of census data by the Pew Research Center (Parker, Morin, Horowitz, Lopez, & Rohal, 2015) found that 6.9% of American residents had multiple race origins, as opposed to the 2.1% identified on 2013 American Community Survey. Furthermore, in 2014, according to the U.S. Census Bureau, 7% of African American men were married to European American women, and 4% of African American women were married to European American men (Current Population Survey, 2014). This marriage rate is consistent with the per-generation European gene flow rate (α) in our best-fit one-, two-, and two-way gene flow with AAM models.

Although these data and our results are inconsistent with a one-time admixture event model, they are at odds with findings from recent high-profile genetic studies (Baharian et al., 2016; Bryc et al., 2015; Gravel, 2012). Recently, Bryc et al. (2015), for example, fit a two-event model of admixture to a large, nation-wide sample of African American genetic data. The best fit-version of this model involved a one-time admixture event between Native Americans and Europeans 12 generations ago followed by a one-time admixture event between this group and Africans six generations ago. This result makes sense in terms of the amount of variation in ancestry in African Americans. One-time admixture events that occurred earlier would result in less variation, and one-time events that occurred later (more recently) would result in a wider range of variation (Verdu and Rosenberg, 2011). However, this result does not make sense in terms of the *shape* of the IA distribution. One time models produce symmetric IA distributions (as do continuous models with equal contributions from both parental sources). Neither the distribution from Bryc et al. (2015) nor the observed distributions for the ASW, AME, and ABT samples are symmetric; they all have strong skews toward low African ancestry. Our modeling results (Figure 4), as well as those from Verdu and

Rosenberg (2011), demonstrate that skewed distributions are the result of asymmetric contributions from parental sources under continuous gene flow models (including zero contribution from one of the parental sources). Furthermore, this one-time admixture event model, and, for that matter, any one-time admixture event model, is inconsistent with census and other historical records documenting mating between Americans of African and European descent. Based on these results, we reject the hypothesis of one-time admixture in African Americans.

The history of continuous slave importation from about 1700 to 1860, as well as the continued post-slavery migration of Africans to what is now the United States, led us to predict that the two-way gene flow model would fit the three observed IA distributions better than one-way gene flow model. We were therefore surprised by the finding for the ASW sample that the two-way gene flow model fit significantly worse than the one-way model. It is possible that African Americans in the U.S. Southwest were relatively isolated from African Americans along the east coast beginning in the 19th century; this result may be consistent with such a history. However, the lack of fit could also reflect the sampling scheme used to collect the ASW data. This scheme excluded individuals with any non-African American parent or grandparent (The International HapMap Consortium, 2003), that is, it excluded individuals with African and/or European ancestors in the previous two generations. This sampling scheme could explain the low average African ancestry in the ASW sample relative to the AME and ABT (Table 4), and it could explain the fact that the ASW sample was the only one of the three to lack individuals with >98% African ancestry. These results highlight the importance in studies of admixture models of collecting representative samples of admixed populations.

In contrast, as expected from the historical and census data, Model 4, with two-way gene flow and AAM, fit best for the AME and ABT samples. An important finding of this study is that input from Africans into the African American population necessitates a concomitant increase in the per-generation contribution from Europeans compared with a one-way gene flow model. AAM also necessitates a higher contribution from Europeans compared with models lacking AAM. For example, for the ABT sample, the per-generation European contributions under the one-, two-, and two-way with AAM models respectively were 3.0%, 5.0%, and 5.7%. The values were even higher for the ASW and AME samples. These results imply that the per-generation contributions to African American populations from Europeans may have been in excess of 5% throughout U.S. history.

No information is provided about the sampling scheme for the AME sample (of which the ABT is a subset). However, the pattern of the lack of fit, an observed excess of African ancestry in the highest bins and a deficit in the lowest, may indicate an absence of recent European contribution, which is inconsistent with census data but consistent with a sampling scheme that excluded individuals with recent European ancestors.

Along the same line, none of the observed distributions had a slight swell in frequency around 0.44. This slight swell was seen in all of the models that included continuous gene flow from Europeans. These swells are actually distributions that are produced by matings between African Americans and Europeans each generation. The same

phenomenon occurs with continuous contribution from Africans to the African American population; however, these distributions are not as apparent because they are contained within main distribution. The absence of this feature in the observed IA distributions reflects a lack of European contribution in recent generations because there was no European contribution, because individuals with recent European ancestry do not self-identify as African American, or because the sampling scheme excluded recent European ancestors.

Independent of the sampling scheme, there are other possible reasons for residual lack of fit of the models to the observed distributions. These reasons include the failure of our relatively simple models to capture the true complexity of African American history. Our models, for example, do not take into account the potential effects of population substructure due to processes other than AAM, for example, the AME sample is comprised of individuals from four locations that may have experienced limited gene flow, or whose ancestors came from different places. Such substructure has led to heterogeneity in the distribution of African ancestry among African Americans in different regions of the country (Bryc et al., 2015). Other possible reasons for the lack of fit include reduced power associated with low sample sizes and limitations of the ancestry-estimation methods, for example, a lack of correct source populations.

Additionally, we did not include Native American contributions in our models. This choice was justified in part by the fact that the Native American contribution is low. Our STRUCTURE analyses showed the mean Native American ancestry to be below 1% for values of K between three and six, and according to the large-scale analyses of thousands of African Americans by Bryc et al. (2015), the mean Native American ancestry among African Americans is 0.8%. Although the average ancestry proportion was higher in the west and southwest, it was still <2%. Furthermore, Jin et al. (2012) found only negligible amounts of ancestry from groups other than Europeans and Africans in their sample of 1,890 African Americans. Importantly, the failure to include Native American ancestry would not affect the fit of one- and two-way gene flow models unless contributing African and European sources themselves contained substantial Native American ancestry prior to mating with individuals in the African American population. Even in this case, for which there is no historical evidence, the shapes of the model-based distributions and the pattern of lack of fit would not be affected.

5 | CONCLUSIONS

We conclude that (a) admixture models that are informed by our understanding of African American history fit better than simplistic models involving one-time admixture events, (b) historically-informed models suggest that the European contribution to African American populations has been higher than previously reported, (c) future studies of the admixture process should collect representative samples of admixed populations, and (d) future studies of the admixture process may benefit from exploring AAM.

ACKNOWLEDGMENTS

The author thanks Heather JH Edgar for her comments on drafts of both the manuscript and the response to reviewers, and acknowledges Jeffrey Long.

ORCID

Jessica M Gross  <http://orcid.org/0000-0002-5221-8218>

REFERENCES

- Baharian, S., Barakatt, M., Gignoux, C. R., Shringarpure, S., Errington, J., Blot, W. J., ... Gravel, S. (2016). The great migration and African-American genomic diversity. *PLOS Genetics*, 12, e1006059.
- Bryc, K., Auton, A., Nelson, M. R., Oksenberg, J. R., Hauser, S. L., Williams, S., ... Bustamante, C. D. (2010). Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 786–791.
- Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., & 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.
- Current Population Survey. (2014). 2014 Annual Social and Economic Supplement Technical Documentation. Washington, DC: US Census Bureau. <ftp://ftp2.census.gov/programs-surveys/cps/techdocs/cpsmar14.pdf>
- Curtin, P. D. (1969). *The Atlantic Slave Trade: A Census*. Madison, WI: The University of Wisconsin Press.
- Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164, 1567–1587.
- Federal Writers Project. (2001). *Born in Slavery: Slave Narratives from the Federal Writers' Project, 1936–1938*. Washington, DC: Library of Congress.
- General Assembly of Maryland. (1664). *Proceedings and Acts: An act concerning Negroes and other slaves*. Available from: <http://web.utk.edu/~mfitzge1/docs/374/MDlaw.pdf>
- The International HapMap Consortium (2003). The International HapMap Project. *Nature*, 426, 789–796.
- Gibson, C. J., & Lennon, E. (1999). *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 1990*. Available from: <https://www.census.gov/population/www/documentation/twps0029/twps0029.html>
- Glass, B., & Li, C. C. (1953). The dynamics of racial intermixture—an analysis based on the American Negro. *Am J Hum Genet*, 5, 1–20.
- Gordon-Reed, A. (1998). *Thomas Jefferson and Sally Hemings: An American Controversy*. Charlottesville, VA: University of Virginia Press.
- Gravel, S. (2012). Population genetics models of local ancestry. *Genetics* 191:607–619.
- Jin, W., Xu, S., Wang, H., Yu, Y., Shen, Y., Wu, B., & Jin, L. (2012). Genome-wide detection of natural selection in African Americans pre- and post-admixture. *Genome Research*, 22:519–527.
- Jones, N. A., & Bullock, J. (2012). *The Two or More Races Population: 2010*. Washington, DC: US Census Bureau. Available from: <https://www.census.gov/prod/cen2010/briefs/c2010br-13.pdf>
- Kidd, J. M., Gravel, S., Byrnes, J., Moreno-Estrada, A., Musharoff, S., Bryc, K., ... Bustamante, C. D. (2012). Population genetic inference from personal genome data: impact of ancestry and admixture on

- human genomic variation. *American Journal of Human Genetics*, 91, 660–671.
- Long, J. C. (1991). The genetic structure of admixed populations. *Genetics*, 127, 417–428.
- Montinaro, F., Busby, G. B. J., Pascali, V. L., Myers, S., Hellenthal, G., & Capelli, C. (2015). Unravelling the hidden ancestry of American admixed populations. *Nature Communications*, 6, 6596.
- Oksenberg, J. R., Barcellos, L. F., Cree, B. A. C., Baranzini, S. E., Bugawan, T. L., Khan, O., ... Hauser, S. L. (2004). Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. *American Journal of Human Genetics*, 74, 160–167.
- Parker, K., Morin, R., Horowitz, J. M., Lopez, M. H., & Rohal, M. (2015). Multiracial in America: Proud, diverse and growing in numbers. Washington, DC: Pew Research Center
- Parra, E. J., Kittles, R. A., Argyropoulos, G., Pfaff, C. L., Hiester, K., Bonilla, C., ... Shriver, M. D. (2001). Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *American Journal of Physical Anthropology*, 114, 18–29.
- Parra, E. J., Marcini, A., Akey, J., Martinson, J., Batzer, M. A., Cooper, R., ... Shriver, M. D. (1998). Estimating African American admixture proportions by use of population-specific alleles. *American Journal of Human Genetics*, 63, 1839–1851.
- Patin, E., Lopez, M., Grollemund, R., Verdu, P., Harmant, C., Quach, H., ... Quintana-Murci, L. (2017). Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science*, 356, 543–546.
- Pfaff, C. L., Barnholtz-Sloan, J., Wagner, J. K., & Long, J. C. (2004). Information on ancestry from genetic markers. *Genetics of Epidemiology*, 26, 305–315.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Risch, N., Choudhry, S., Via, M., Basu, A., Sebro, R., Eng, C., ... Gonzalez Burchard, E. (2009). Ancestry-related assortative mating in Latino populations. *Genome Biology*, 10, R132.
- Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic structure of human populations. *Science*, 298, 2381–2385.
- Smith, J. F. (1973). *Slavery and Plantation Growth in Antebellum Florida, 1821 - 1860*. Gainesville: University of Florida Press.
- Sokal, R. R., & Rohlf, F. J. (2012). *Biometry: The Principles and Practice of Statistics in Biological Research*. 4th ed. New York: W.H. Freeman and Co.
- Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., ... Williams, S. M. (2009). The Genetic structure and history of Africans and African Americans. *Science*, 324, 1035–1044.
- US Census Bureau. (2010). American FactFinder. Available from: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_10_1YR_B04003&prodType=table
- Verdu, P., & Rosenberg, N. A. (2011). A general mechanistic model for admixture histories of hybrid populations. *Genetics*, 189, 1413–1426
- Voyages Database. (2009). Voyages: The Trans-Atlantic Slave Trade Database. Available from: <http://www.slavevoyages.org/assessment/estimates>

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Gross JM. Tests of fit of historically-informed models of African American Admixture. *Am J Phys Anthropol*. 2017;00:000–000. <https://doi.org/10.1002/ajpa.23343>