Differentiating mechanism from outcome for ancestry-assortative mating in admixed human populations

Dashiell J. Massey[1], Zachary A. Szpiech[2,3], Amy Goldberg[1]

[1] Department of Evolutionary Anthropology, Duke University, USA 27708

[2] Department of Biology, Pennsylvania State University, USA 16801

[3] Institute for Computational and Data Sciences, Pennsylvania State University, USA 16801

Abstract

Population genetic theory, and the empirical methods built upon it, often assume that individuals pair randomly for reproduction. However, natural populations frequently violate this assumption, which may potentially confound genome-wide association studies, selection scans, and demographic inference. Within several recently admixed human populations, empirical genetic studies have reported a correlation in global ancestry proportion between spouses, referred to as ancestry-assortative mating. Here, we use forward genomic simulations to link correlations in ancestry between mates to the underlying mechanistic mate-choice process. We consider the impacts of two types of mate-choice model, using either ancestry-based preferences or social groups as the basis for mate pairing. We find that multiple mate-choice models can produce the same correlations in ancestry proportion between spouses; however, we also highlight alternative analytic approaches and circumstances in which these models may be distinguished. With this work, we seek to highlight potential pitfalls when interpreting correlations in empirical data as evidence for a particular model of human mating practices, as well as to offer suggestions toward development of new best practices for analysis of human ancestry-assortative mating.

Introduction

Non-random mating has long been appreciated as an important source of genetic structure in natural populations (Fisher 1918; Wright 1921; Wright 1950; Nagylaki 1978). Positive assortative mating (hereafter, assortative mating), wherein genotypic or phenotypic trait values are positively correlated between mates, has been empirically observed across animal species (Jiang *et al.* 2013). Theoretical and empirical studies have demonstrated the consequences of assortative mating for speciation and hybridization (Kondrashov 1983; Otto *et al.* 2008; Tung *et al.* 2012; Schumer *et al.* 2017; Kopp *et al.* 2018; Powell *et al.* 2021; Muralidhar *et al.* 2022; Natola *et al.* 2022; Smadja *et al.* 2022; Robinson *et al.* 2023), and for the distributions of traits within populations (Wright 1921; Norris *et al.* 2019; Kim *et al.* 2021; Border *et al.* 2022; Muralidhar *et al.* 2022; Horwitz *et al.* 2023).

In non-human primates that live in complex and structured social environments, sociodemographic factors have been demonstrated to influence mate pair formation (Keddy-Hector 1992; Klinkova *et al.* 2005; Setchell and Jean Wickings 2006; Van Belle *et al.* 2009; Tung *et al.* 2012; Fogel *et al.* 2021). As such, when it comes to humans, assortative mating has been the purview not only of biologists, but also social scientists (*e.g.*, (Buss and Barnes 1986; Mare 1991; Kalmijn 1998; Luo and Klohnen 2005; Blossfeld 2009; Torche 2010; Schwartz 2013; Greenwood *et al.* 2014; Henz and Mills 2017;

SMIEJA AND STOLARSKI 2018; CHIAPPORI 2020; DE LA MARE AND LEE 2023)). Positive correlations have been reported between human spouses for a diverse array of phenotypes, including morphometric measurements, health outcomes, personality traits, lifestyle factors, age, socioeconomic status, educational attainment, religious affiliation, and language (NAGOSHI *et al.* 1990; ROBINSON *et al.* 2017; FIBLA *et al.* 2022; HORWITZ *et al.* 2023; YAMAMOTO *et al.* 2023). These associations are likely driven by multiple generative processes, including individual mate-choice preferences, phenotypic convergence over time facilitated by cohabitation, and social structures that restrict or promote particular pairings. These processes may act individually or in concert. For instance, sociological literature on assortative mating by educational attainment has revealed the influence of both a preference for mate similarity (KALMIJN 1998) and social barriers to marriage across socioeconomic class (TORCHE 2010), and that the strength of assortment by education over time is sensitive to the degree of temporal overlap between the end of schooling and average age at the time of marriage (MARE 1991), suggesting that social milieu is likely also an important factor in structuring mate pair outcomes.

Within the field of population genetics, relatively less attention has been paid to understanding the contributions of these generative processes to producing the observed correlations between mating pairs. Indeed, mechanism and outcome are often conflated or not clearly delineated when discussing assortative mating, which obscures these multiple mechanisms that may contribute to mate similarity. When studies do consider mechanism, they typically draw on foundational work in the sexual selection literature modeling the effects of female choice on male phenotype (LANDE 1981; KIRKPATRICK 1982; SEGER 1985), in which the mechanism driving assortative mating is assumed to be mate choice. To extend this framework beyond sexual dimorphism, the mate-choice mechanism is modeled as a preference for mates that "match" an individual's own phenotype (KOPP *et al.* 2018; GOLDBERG *et al.* 2020; KIM *et al.* 2021; MURALIDHAR *et al.* 2022). However, some studies have suggested temporal structure in the mating process (XIE *et al.* 2015; WOODMAN *et al.* 2023) or non-uniform ability to attract mates (BURLEY 1983) can generate phenotypic correlations between mates without an explicit preference for phenotypic similarity. In this paper, we will use the term "assortative mating" to refer specifically to an empirical observation of greater resemblance between mates than expected by chance, agnostic to mechanism.

Reports of assortative mating by ancestry in humans provide a particularly intriguing example for exploring mechanism because genetic ancestry is a complex quantitative "phenotype" that involves every locus in the genome and that can only be ascertained by sequencing. In recently admixed human populations, in which individuals derive ancestry from multiple source populations, a positive correlation in global ancestry proportion between spouses has been observed. This phenomenon, referred to as ancestry-assortative mating, has been reported in multiple Latino populations (RISCH *et al.* 2009; ZOU *et al.* 2015; SPEAR *et al.* 2020; MAS SANDOVAL *et al.* 2023), as well as in African-Americans (ZAITLEN *et al.* 2017; AVADHANAM AND WILLIAMS 2022), Cabo Verdeans (KORUNES *et al.* 2022), and ni-Vanuatu (ARAUNA *et al.* 2022). Ancestry-associated mating patterns, inferred from excess sharing of single-nucleotide variants, have also been reported in human populations with more distant histories of admixture, *e.g.*, non-Hispanic white spouse pairs in the United States (SEBRO *et al.* 2010; DOMINGUE *et al.* 2014; SEBRO *et al.* 2017). Empirical studies commonly use such correlations in ancestry between spouses to suggest that ancestry shapes mate choice, although this interaction would necessarily be mediated by a proxy phenotype of some type, not by

quantitative ancestry proportion. However, the relative contributions of individual preferences and broader social mechanisms to generating these correlations are unclear.

Accounting for assortative mating is critical for accurate statistical and population genetic analysis in admixed human populations: it has been shown that assortative mating confounds genome-wide association studies (HOWE *et al.* 2021; BORDER *et al.* 2022; VELLER AND COOP 2024) and heritability estimation (SEBRO AND RISCH 2012; TENESA *et al.* 2016; HUANG *et al.* 2023). In addition, ancestry-assortative mating specifically has been shown to produce estimates for the timing of admixture that are too recent (ZAITLEN *et al.* 2017; GOLDBERG *et al.* 2020; KORUNES *et al.* 2022). However, it remains challenging to establish a null expectation in the presence of assortative mating without understanding the underlying mechanism at play.

To date, statistical methods to correct for ancestry-assortative mating in empirical data have tended to generate null models by simulating data to match the observed empirical correlation in ancestry proportion between mating pairs, often without a specific mechanistic model (ZAITLEN *et al.* 2017; SPEAR *et al.* 2020; HUANG *et al.* 2023; PFENNIG AND LACHANCE 2023). This approach makes two important assumptions: first, that the correlation coefficient for a given population has remained constant over time; and second, that matching the correlation coefficient is sufficient to recapitulate the dynamics of global and local ancestry. In contrast, theoretical population genetic studies have examined the behavior of these same parameters using mechanistic models of biased mate choice (GOLDBERG *et al.* 2020; KIM *et al.* 2021; MURALIDHAR *et al.* 2022). This approach assumes that ancestry-based preference is the primary mechanism driving non-random mating. These studies typically do not directly consider whether the model used results in an observed correlation in ancestry between mates, which leaves open the question of how applicable their conclusions are to the context of empirical studies of human ancestry-assortative mating.

Here, we used forward-in-time simulations to probe the relationship between mate-choice mechanism and observed correlation in ancestry, comparing variants of two classes of mate-choice model. The first model class considers only individual mate-choice preferences for similarity in ancestry proportion, treating genetic ancestry as a continuous quantitative trait. The second class considers only social groups, assigning each individual to one of two groups and limiting cross-group mating events. (Importantly, group membership is "inherited" randomly from one parent and group identities are initially associated with source populations; see **Models**.) We found that ancestry-similarity and social group models can produce similar correlations in ancestry between mates, suggesting that either or both mechanisms could be relevant in interpreting empirical data. Furthermore, the correlation coefficient was not constant over time in any of our simulations, and not all variants of the ancestry-similarity class or all parameter values of the social group model maintained sufficient population variance to sustain correlations in ancestry over tens of generations. Our results highlight important caveats about the assumptions behind both existing approaches and suggest additional analyses of empirical data that may help reveal the underlying mechanism of ancestry-assortative mating in human populations.

Models

Simulation framework

We performed forward-in-time simulations of admixture with equal contributions from two source populations using SLiM 4.0.1 (HALLER *et al.* 2019; HALLER AND MESSER 2023). We modeled sexually reproducing diploid individuals with 22 independently segregating chromosomes, representing the size distribution of human autosomes (total genome size $L = 2.88$ Gb), with a uniform recombination rate $r = 1 \times 10^{-8}$. We did not model distinct sexes; thus, mating could occur between any two individuals. Unless otherwise noted, there was a single pulse of admixture and no additional contributions from the source populations introduced at later generations. We recorded global and local ancestry for all individuals for each of 50 generations post-admixture, using a constant census population size of $N = 10{,}000$ individuals per generation. We focus primarily on results from the first 20 generations post-admixture, approximately corresponding to the timing of African-European admixture initiated by the trans-Atlantic slave trade (ZAITLEN *et al.* 2017; HAMID *et al.* 2021; KORUNES *et al.* 2022; MAS SANDOVAL *et al.* 2023; MOONEY *et al.* 2023), with a subset of results over longer time periods in the Supplement.

To investigate the impacts of different mate-choice models, we extended SLiM's built-in Wright-Fisher framework to include non-random mating using the `mateChoice()` callback. Specifically, we considered two broad classes of mate-choice models: (1) **ancestry-similarity**, wherein the probability of an individual of ancestry x_1 mating with an individual of ancestry x_i is defined in terms of $|x_1 - x_i|$; and (2) **social group**, wherein the probability of an individual belonging to social group s_1 mating with an individual belonging to social group s_i is defined by whether or not $s_1 = s_i$.

Under each model, we constructed mating pairs. First, mate 1 was uniformly sampled with replacement from the population. Second, each individual in the population was assigned a mating weight ψ_i, calculated by a mate-choice function (defined below) evaluated at $f(x_1, x_i)$ or $f(s_1, s_i)$, for the ancestry-similarity and social group models, respectively (**Figure 1**). Third, mate 2 was sampled with probability proportional to ψ_i, with incidental selfing explicitly prohibited. Each mating event generated a single individual in the subsequent generation.

Ancestry-similarity models

Individual-based mate preference models originate in the sexual selection literature, where they were developed for studying the co-evolutionary dynamics of male secondary sex characteristics and female mate choice (LANDE 1981; KIRKPATRICK 1982; SEGER 1985). Generally, under these models, one locus controls male phenotype while a second locus controls female preference for that phenotype. In the speciation literature, an alternative family of models are concerned with assortative mating as a mechanism of sympatric speciation and model mate preference based on phenotypic similarity for some ecologically relevant trait controlled by an arbitrary number of loci (DIECKMANN AND DOEBELI 1999; BURGER AND SCHNEIDER 2006; PENNINGS *et al.* 2008; RETTELBACH *et al.* 2013). These "matching rule" models (KOPP *et al.* 2018) are the basis for our ancestry-similarity model class, wherein individuals preferentially choose mates to maximize similarity in phenotype (*i.e.*, ancestry). Importantly for the models we consider here, these are fixed relative-preference models: a focal individual assigns each potential mate a preference value ψ_i that depends only on the potential mate's phenotype and is density independent (SEGER 1985).

Ancestry-similarity models define a mate-choice function $\psi_i = f(x_1, x_i)$ for calculating the sampling weight assigned to individual i with ancestry proportion x_i as a potential mate for a focal individual with ancestry proportion x_1. In the present study, we considered three variants of the ancestry-similarity model identified in a literature search, which differ in the precise definition of the mate-choice function f.

Models of ancestry-based assortative mating typically model $f(x_1, x_i)$ as an exponential function, such that ψ_i decays exponentially at rate c as $|x_1 - x_i|$ increases. We define,

$$\psi_i = e^{-c|x_1 - x_i|}.$$

We refer to this as the **stationary-preference model** ("like-with-like" model in (MURALIDHAR *et al.* 2022)): for a given difference in ancestry, the associated value of ψ_i does not change over time. It is important to note, however, that the range of ψ_i is constrained by x_1: a focal individual with ancestry $x_1 = 0.5$ is more similar to all potential mates than a focal individual with ancestry $x_1 = 0$, and is therefore less selective in choosing a mate (**Figure 1A**). Furthermore, without additional migration from the source populations, variance in ancestry within the population decreases over time (**Figure S2**), meaning that the pool of potential mates is becoming increasingly homogenous and the ratio of $\max(\psi_i)\,/\min\,(\psi_i)$ approaches 1 for all values of x_1 (**Figure S1A**). In other words, all individuals become less selective in choosing mates over successive generations.

One approach to account for the decreased variance in ancestry proportion over time, and thus preserve mate-selectiveness over time, is to scale the exponential decay rate c by the generation-specific population variance in ancestry, σ_g^x:

$$\psi_i = e^{\frac{-c}{\sigma_g^x}|x_1 - x_i|}.$$

We refer to this as the **increasing-preference model** (EQUATION A2 in (KIM *et al.* 2021)). Assuming an infinite-sites model, $\sigma_g^x = 2^{-g/2}$, where g is the number of generations post-admixture (VERDU AND ROSENBERG 2011), although it has been demonstrated that this variance is greater than $2^{-g/2}$ when accounting for recombination on finite-length chromosomes, as we are here (LIANG AND NIELSEN 2014). Under this model, as under the stationary-preference model, a focal individual with ancestry $x_1 = 0.5$ is less selective than a focal individual with ancestry $x_1 = 0$ (**Figure 1B**). However, because σ_g^x decreases over time, there is an increase in mate-selectiveness over successive generations under this model for a given x_1 (**Figure S1B; Figure S3**).

We also considered a third ancestry-similarity model, employing a Gaussian – rather than exponential – mate-choice function. This approach is commonly used in models of assortative mating based on a quantitative trait (*e.g.,* (DIECKMANN AND DOEBELI 1999; PENNINGS *et al.* 2008; FUNK *et al.* 2021)), although it has not to our knowledge been used to model assortative mating based on ancestry proportion. Under this model, ψ_i decreases following a normal distribution as $(x_1 - x_i)^2$ increases, with the strength of mate-choice preference set by a fixed parameter, σ:

$$\psi_i = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_1 - x_i)^2}{2\sigma^2}}.$$

We refer to this as the **broad-preference model** because $f(x_1, x_i)$ decays more gradually with increasing differences in ancestry proportion, relative to the stationary-preference and increasing-preference models (**Figure 1C**). Thus, a focal individual prefers potential mates with similar ancestry proportion to themself over potential mates with dissimilar ancestry proportion, but smaller differences within the pool of "similar" mates are not weighed strongly. Compared to the other two models, individuals are less selective in choosing mates under the broad-preference model; however, as under the stationary-preference model, they also become less selective over time (**Figure S1C; Figure S4**).

Social group model

An alternative class of models considers preferential mating based on a categorical, rather than quantitative, trait. This type of model is often used in the context of interspecific hybrid zones, wherein non-random mating can maintain species boundaries and/or promote hybrid speciation (MELO *et al.* 2009; SCHUMER *et al.* 2017; POWELL *et al.* 2021; NATOLA *et al.* 2022; SMADJA *et al.* 2022). Mechanistically, this can be construed as a model of biased mating by species or source population (GOLDBERG *et al.* 2020). However, application of this type of model in humans is not straightforward: the source populations in question are not biologically distinct species, but rather labels describing socially defined boundaries. Furthermore, it is unclear how we ought to model the mating behavior of admixed individuals: neither random mating by admixed individuals nor a simple preference of admixed individuals to mate with one another seem likely to produce the observed correlation in global ancestry proportion between mates. Yet, the strongest evidence for ancestry-assortative mating in humans is observed in contexts where both spouses have admixed ancestry (RISCH *et al.* 2009; ZOU *et al.* 2015; ZAITLEN *et al.* 2017; SPEAR *et al.* 2020; AVADHANAM AND WILLIAMS 2022; KORUNES *et al.* 2022; MAS SANDOVAL *et al.* 2023).

Social identity is an important organizer of pair formation in humans, as evidenced by widespread cultural practices of endogamy based on race, ethnicity, religion, socioeconomic status, and caste. Although distinct from genetic ancestry, social categorizations often interact with ancestry. For instance, skin pigmentation, a prominent contributor to how an individual is racialized, is correlated with West African-related ancestry proportion in multiple populations with recent admixture history (PARRA *et al.* 2003; SHRIVER *et al.* 2003; BONILLA *et al.* 2004; BELEZA *et al.* 2012). Race has also served as the basis for numerous sociolegal barriers to mating across contexts, including anti-miscegenation laws (BROWNING 1951). This type of interaction opens the possibility that social barriers to mating might be "recorded" in patterns of ancestry-assortative mating detectable by genomic analyses.

We considered a simple **social group model** with two groups, in which a mate-choice function $\psi_i = f(s_1, s_i)$ determines the sampling weight assigned to individual i belonging to social group s_i as a potential mate for a focal individual belonging to social group s_1 as a function of mating bias parameter α (**Figure 1D; Figure S5**):

$$\psi_i = \begin{cases} 1 - \dfrac{1}{\alpha + 1}, & s_1 = s_i \\ \dfrac{1}{\alpha + 1}, & s_1 \neq s_i. \end{cases}$$

To generate an association between social group membership and ancestry proportion, we assigned individuals in each source population to the same social group prior to admixture. However, from the onset of population contact, individuals "inherited" their social group membership from the first parent in the mating pair. Thus, individuals whose parents belonged to opposite social groups were equally likely to be assigned to either social group.

Several features distinguish the social group model from the ancestry-similarity models described above, or from one in which individuals use an observable quantitative trait (*e.g.*, skin pigmentation) as a proxy for an unobserved trait (genetic ancestry). First, social group membership is correlated with – but not determined by – genotype. This matters because it has previously been demonstrated that recombination is expected to decouple the relationship between ancestry proportion and a proxy quantitative trait over time, leading to only a transient correlation between the two phenotypes (Kim *et al.* 2021). Because social group membership is not genetic, it is not impacted by recombination. In addition, individuals are uniformly likely to cross the social group barrier in choosing a mate and do not express relative preferences among potential mates within a social group. Thus, in contrast to an ancestry-similarity model, individuals with intermediate ancestry are not less selective than individuals with more source-like ancestry in choosing a mate.

However, these features of the social group model also mean that the association between ancestry proportion and social group membership is expected to decay over successive generations, and we explored this relationship quantitatively over time.

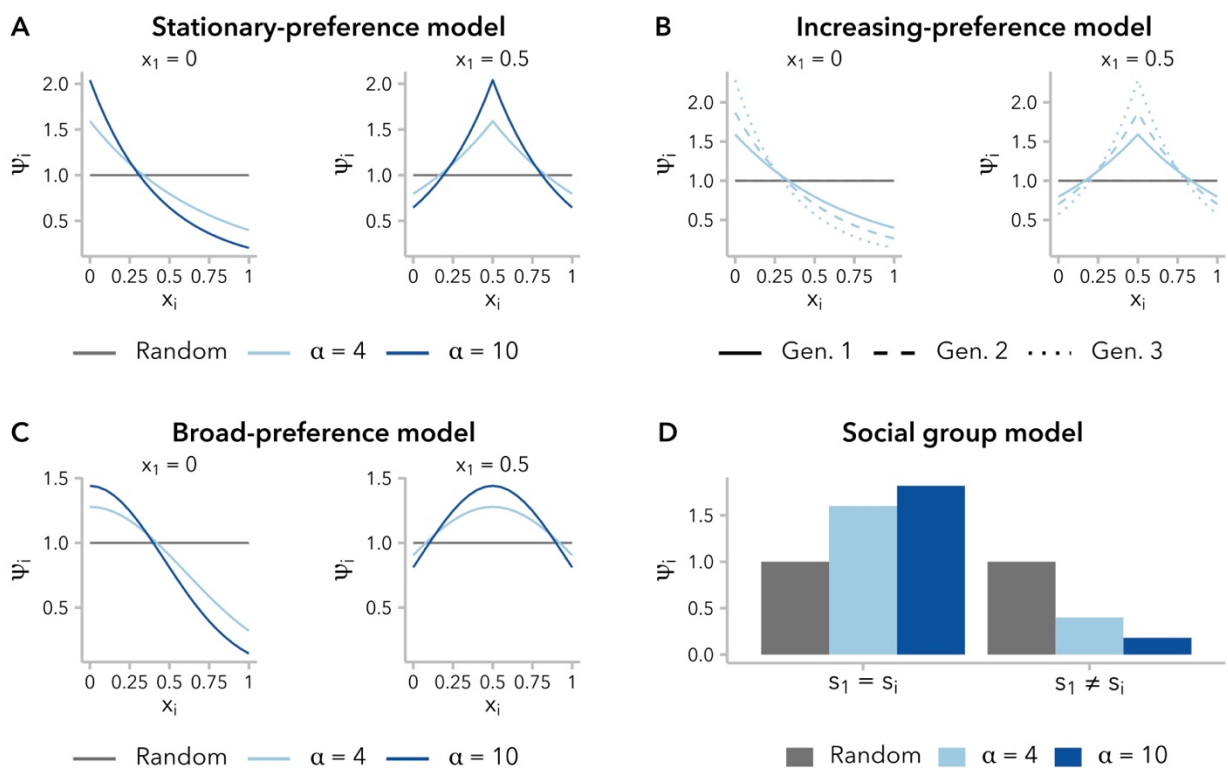

Figure 1. Four functions for defining the sampling weight, ψ_i, assigned to individual i as a potential mate for a focal individual, based on global ancestry proportion x_i (**A-C**) or social group s_i (**D**). For each mating event, mate 1 was uniformly sampled from the population, and mate 2 was chosen by weighted sampling. **(A)** The stationary-

preference variant of the ancestry-similarity model defines ψ_i as a function of the difference in ancestry proportion between individuals, $|x_1 - x_i|$. **(B)** Under the increasing-preference variant of the mate-preference model, the value of ψ_i is scaled to the variance in ancestry proportion across all potential mates, resulting in increasing choosiness over time. ψ_i values are shown for the first three generations post-admixture for $\alpha = 4$. **(C)** Under the broad-preference variant of the ancestry-similarity model, ψ_i follows the probability density function of a normal distribution with mean x_1. **(D)** Under the social group model, ψ_i can take one of two discrete values, determined by whether a potential mate belongs to the same social group as the focal individual ($s_1 = s_i$) or to the other social group ($s_1 \neq s_i$).

Defining a common parameter for model comparison

The mate-choice functions we considered differ in how they parameterize the strength of mating bias; to aid in comparison across models, we define an alternate metric of the strength of mating bias, α, which measures the frequency of mating events between individuals from opposite source populations. Specifically, we define α to be inversely related to the proportion of admixed offspring in the initial generation post-contact, A. We have,

$$\alpha = \frac{1}{\mathrm{A}} - 1.$$

That is, at time of population contact, individuals are α times more likely choose a mate from within their own source population than from the opposite source population. α must be ≥ 0, with increasing values of α representing an increasing bias toward endogamy. $\alpha = 0$ corresponds to complete disassortative mating (all mating pairs are cross-source population), $\alpha \in (0,1)$ corresponds to negative assortative mating, $\alpha = 1$ corresponds to random mating, and $\alpha > 1$ corresponds to positive assortative mating. Given our focus on mechanisms that might explain empirically observed positive correlations, we focus here on $\alpha \geq 1$. While models differ in their dynamics over time, comparing simulations with the same α ensures that the correlation in ancestry between mates at generation 1, $r_{t=1}$, is the same. For instance, for the scenario in which both sources contribute equally to a single-pulse admixture event:

$$r_{t=1} = \frac{0.25 - (\mathrm{A}/2)}{0.25} = 1 - 2\mathrm{A} = 1 - 2(\frac{1}{\alpha + 1}).$$

Results

All four models produce a positive correlation in global ancestry between mates, at least transiently

Ancestry-assortative mating is often inferred from a positive correlation between spouses at single point in time, and statistical methods to account for non-random mating typically generate null models using simulations where potential mate pairs are permuted until the empirically observed correlation is achieved (ZAITLEN *et al.* 2017; HUANG *et al.* 2023; PFENNIG AND LACHANCE 2023). This approach assumes that the correlation observed in the present-day samples has remained constant since the start of admixture. **Figure 2** plots the Pearson correlation coefficient between mates over time under three variants of the ancestry-similarity model or the social group model, for multiple strengths of assortative mating (values of α) and following a single pulse of admixture. In all scenarios of non-random mating that we considered, we observed positive correlations in ancestry proportion between mates (**Figure 2; Figure S6**). However, contrary to the assumption that the correlation in ancestry observed in a contemporary sample should be modeled as a fixed parameter, the observed correlation coefficients in our simulations decayed over time. In simulations performed under the increasing-preference model, correlation between mates reached a stable plateau within the first 10 generations post-admixture across values of α (**Figure S7A**). In contrast, simulations performed under the other two variants of the ancestry-similarity model (stationary-preference or broad-preference) approached correlations of zero within 20 generations post-admixture, unless α was sufficiently large to disrupt the admixture process (**Figure S8; Figure S9**). Under the social group model, positive correlations between mates could be observed 20 generations post-admixture for $\alpha > 2$, although these correlations decayed to zero on longer time scales (**Figure S7B**). Thus, not all variants of the ancestry-similarity model generate a correlation in ancestry between mates observable 20 generations post-admixture while also allowing for a large population of individuals within intermediate ancestry. Conversely, results for the social group model demonstrate that even relatively weak social barriers can produce correlations in ancestry between mates on the timescale of tens of generations, even as the association between social group membership and ancestry proportion diminishes over time **(Figure S10).**

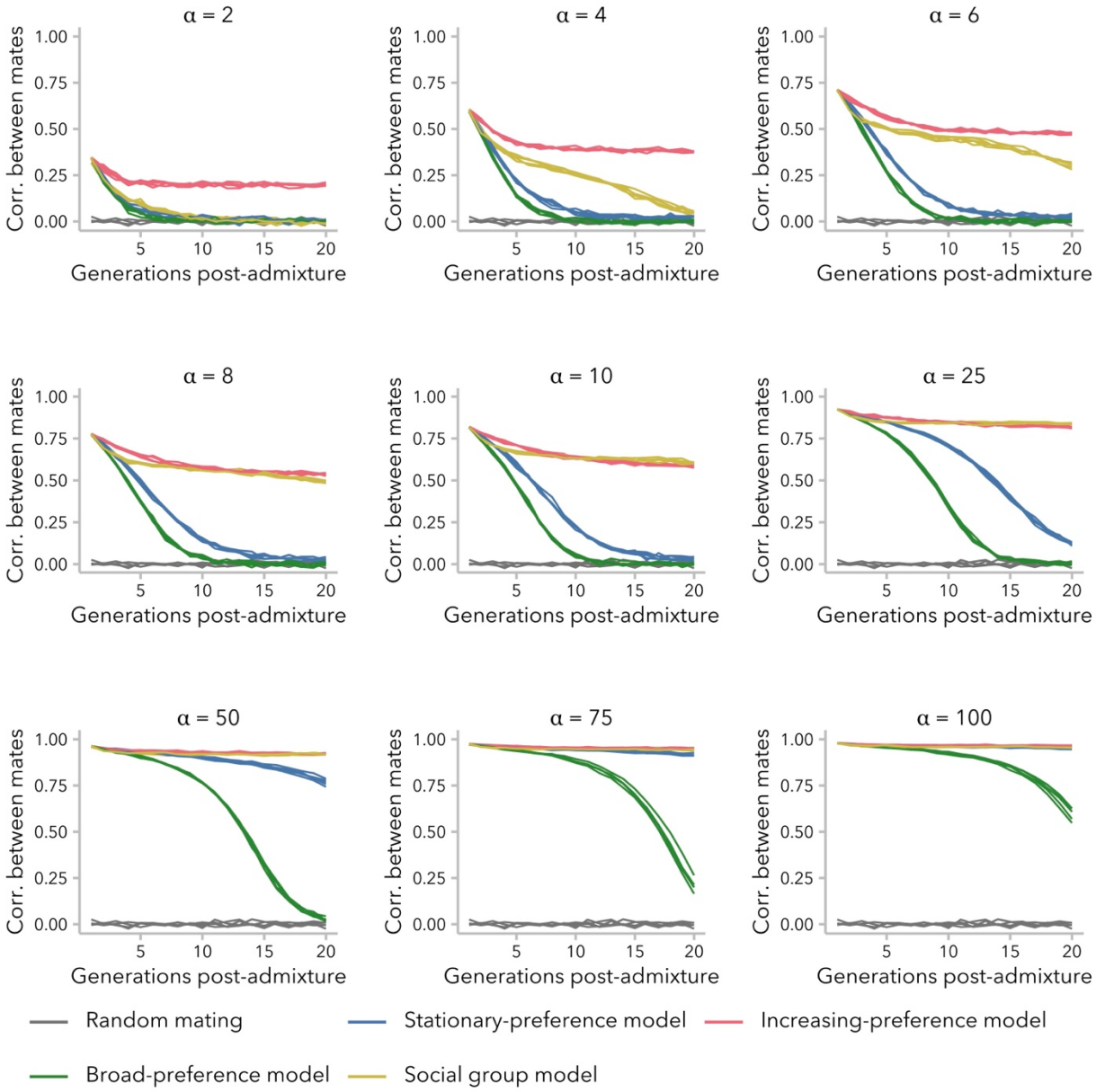

Figure 2. Correlation in ancestry between mates was not constant over time and decayed to near-zero within 20 generations post-admixture in some simulations. Simulations under the stationary-preference (blue) and broad-preference (green) models did not maintain a positive correlation while also allowing for widespread admixture, whereas those under the increasing-preference (pink) and social group (yellow) models did for some α. By definition, simulations performed using the same value of α have the same correlation in generation 1, regardless of model. The same replicate random-mating simulations ($\alpha = 1$; gray) are reproduced in each subplot. See **Figure S6**.

To understand mechanisms that could potentially explain the correlations between spouses observed in empirical data, in the following sections we will focus on the two models that produce measurable positive correlations between mates 20 generations post-admixture: the increasing-preference variant of the ancestry-similarity model, and the social group model. Specifically, we

compare pairs of simulations under these models that produced the same correlation coefficient at generation 20.

Similar correlation coefficients may be produced by different underlying distributions of ancestry proportion

Though commonly used for empirical genetic studies, the correlation between mating pairs is not particularly informative about the underlying distribution of global ancestry proportion, across individuals or mating pairs. We next extended our perspective beyond correlation coefficients to broader patterns of ancestry proportion under the increasing-preference and social group models.

Over time, the distribution of ancestry proportion x shifts from two discrete populations to a continuous distribution of intermediate values. Random mating produces a unimodal distribution of ancestry proportion, here seen from generation 2; under ancestry-biased mating, individuals with more extreme values of x are likely to mate with one another, meaning that we expect the shift from bimodal to unimodal distribution to be delayed. However, the timing of this shift is delayed more under the social group model than under the increasing-preference model for the same α (**Figure S11**). Furthermore, this difference in the distribution of ancestry proportion under the two models is observed even when comparing pairs of simulations that have similar correlation coefficients.

Specifically, we analyzed the distributions of global ancestry within the admixed population at generation 20, approximating a common time since admixture for multiple human populations with African and European ancestry (ZAITLEN *et al.* 2017; HAMID *et al.* 2021; KORUNES *et al.* 2022; MAS SANDOVAL *et al.* 2023; MOONEY *et al.* 2023). At $g = 20$, we found a unimodal distribution of ancestry proportion for simulations under the increasing-preference model for all values of α, whereas the social group model produced a bimodal distribution for simulations with $\alpha \geq 7$. For simulations under the social group model with $\alpha < 7$, ancestry proportion was unimodal but with higher variance than simulations under the increasing-preference model with the same correlation between mates **(Figure S12)**.

The difference in the distribution of ancestry proportion at generation 20 between the two models could also be observed in plotting the correlation between mating pairs (**Figure 3A-B**): under the increasing-preference model, there was a single cluster of mating pairs with the highest density around $x = 0.5$, corresponding to the original admixture contributions. In contrast, under the social group model, we were able to distinguish two clusters of mating pairs, symmetric around the population mean. These clusters were apparent when the correlation in ancestry between mates was ≥ 0.4 **(Figure S13)**. Thus, variants of both the ancestry-similarity and the social group model can produce similar correlations, but these models may be distinguishable by examining underlying structure in mating pairs, especially when the correlation in ancestry between mates is sufficiently strong.

Given that evidence for ancestry-assortative mating is often assessed in much smaller empirical datasets, we also examined whether these differences in mating structure could be observed in a subsample of our simulation population ($n = 100$ individuals). Correlation in ancestry between mates is typically visualized as a dot-plot (*e.g.*, (ZOU *et al.* 2015; KORUNES *et al.* 2022)); we found that the two models looked similar when taking this approach (**Figure S14**). However, even with 100

individuals, density plots were suggestive of discrete clusters of mate pairs in simulations under the social group model with high α, and thus high correlation in ancestry between mates (**Figure 3C-D; Figure S15**).

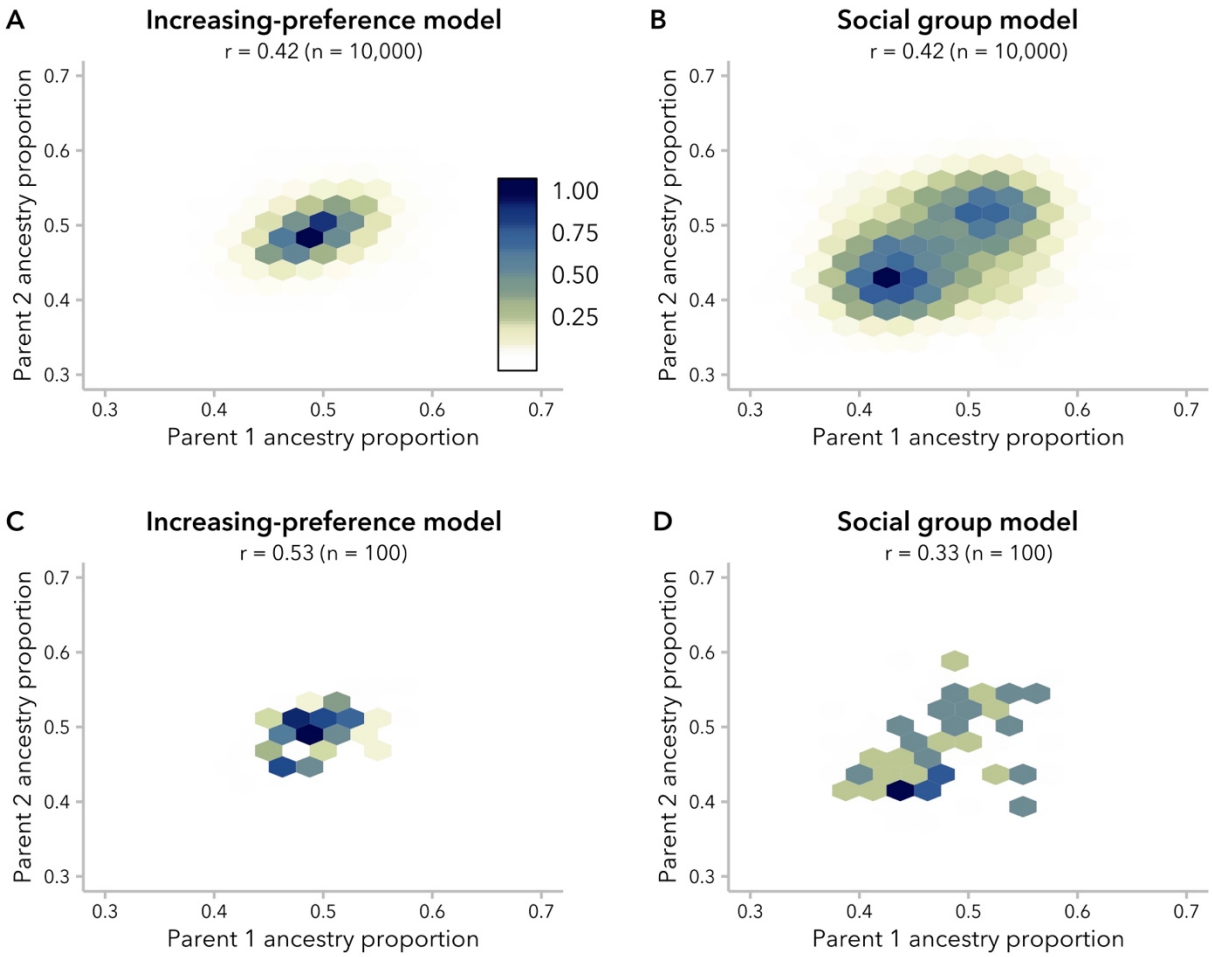

Figure 3. Simulations under the increasing-preference ($\alpha = 5$) and social group ($\alpha = 7$) models produced similar correlation coefficients with different underlying mating structure. "Hexbin" plots represent correlation in ancestry proportion between the two parents of individuals in generation 20, with each hexagon corresponding to a bin of 0.025 ancestry proportion units. The density of mating pairs in each ancestry bin is represented by the color, with darker colors representing higher density. (Density values are scaled to a maximum of 1). **(A, B)** Under the increasing-preference model, mating pairs cluster around a single bin of maximum density, whereas under the social group model, there are two centers of density. All 10,000 individuals are shown. **(C, D)** Density plots are suggestive of differences in mating structure between models, even with a sample size of 100 individuals.

Variance in ancestry proportion determines whether or not correlation in ancestry between mates is maintained over time

Although all of the models we consider define functions for non-random mate choice, we observed in **Figure 2** that for the stationary-preference and broad-preference models, the correlation in ancestry proportion between mates either decayed to zero within the first 20 generations post-

admixture ($\alpha \leq 10$) or permitted very limited admixture ($\alpha \geq 25$). To understand why this correlation approaches zero under these models, but not under the increasing-preference and social group models, we compared the population variance in ancestry proportion over time under each model. Intuitively, variance is necessary for observing non-random mating: if all potential mates are identical, mate choice will appear arbitrary. Thus, we would expect our simulations to become indistinguishable from random mating, even though we are modeling biased mating, if variance in ancestry proportion becomes sufficiently small.

Consistent with prior analytic models, non-random mating resulted in higher variance in ancestry proportion relative to random mating at a given generation, regardless of mating model (WRIGHT 1921; VERDU AND ROSENBERG 2011; LIANG AND NIELSEN 2014; ZAITLEN *et al.* 2017; GOLDBERG *et al.* 2020), and variance decayed more slowly when α was larger (**Figure 4**). However, variance in ancestry proportion for simulations under the stationary-preference and broad-preference models approached to zero within ~15 generations post-admixture under our set of parameters. This timing corresponds to the decay to zero in correlation between mates observed in **Figure 2** under these two models (**Figure S16**). In contrast, for the increasing-preference and social group models, which sustained a correlation between mates over 20 generations, variance in ancestry proportion was greater than for random mating at each time point (all α for increasing preference; $\alpha > 2$ for social group).

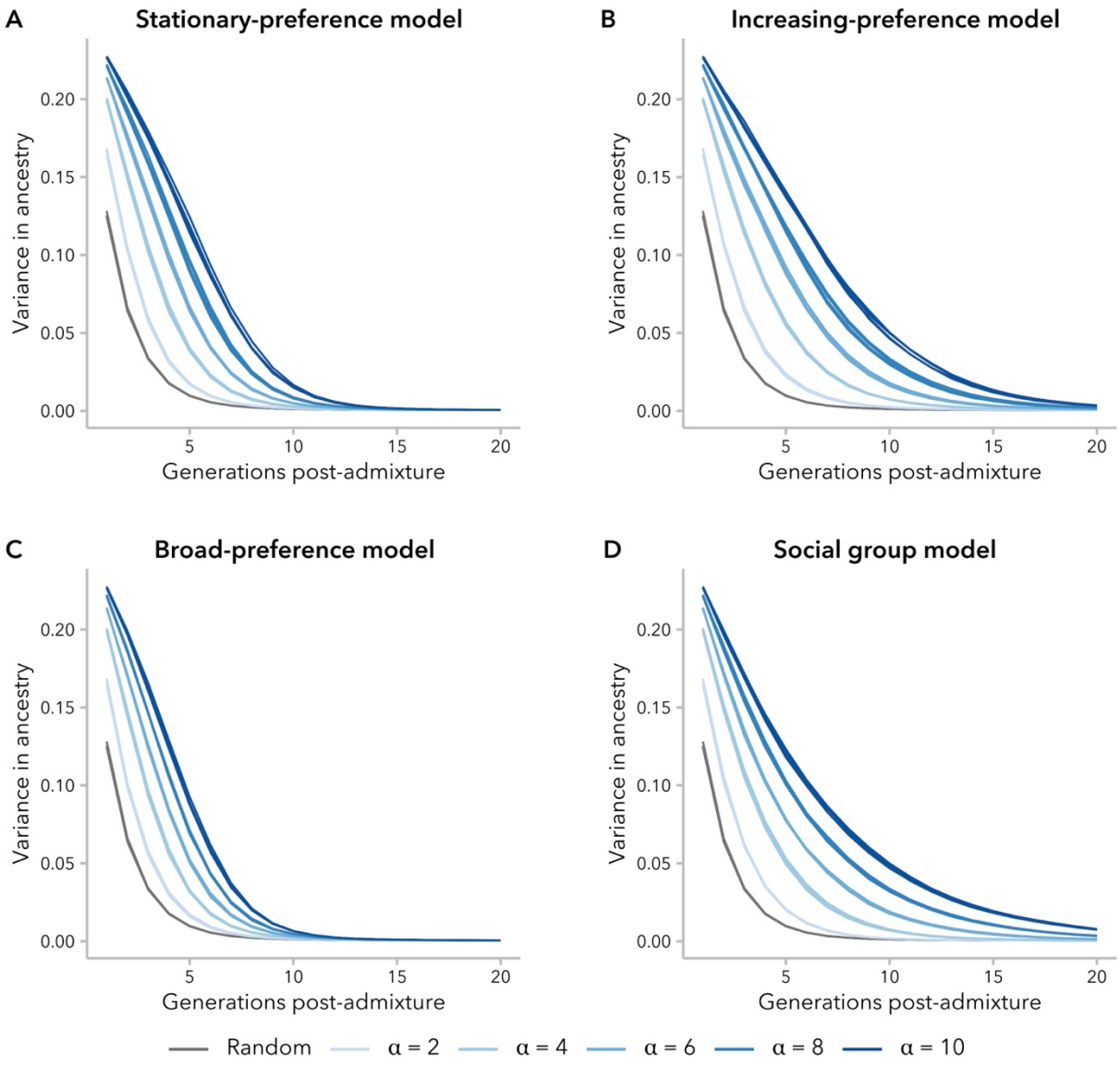

Figure 4. Variance in ancestry proportion decayed to zero over time for a single-pulse admixture scenario, but the decay is slower for larger values of α. At twenty generations post-admixture, only simulations under the increasing-preference and social group functions **(B, D)** had a greater variance than the random-mating control simulations.

To quantitatively assess whether the decay in correlation between mates meant that mating was becoming effectively random over time, we performed permutation tests comparing the correlation coefficient for the observed mating pairs to randomized mating pairs (1,000 permutations). We found that this was not necessarily the case: for $\alpha = 10$, correlation between mates under the stationary-preference model was significantly different from random 20 generations post-admixture even in a simulation for which $r < 0.05$. However, under the broad-preference model, the correlation between mates for $\alpha = 10$ was indistinguishable from random by around 10 generations post-admixture (**Figure S17**).

Given the limited information from a single summary statistic, the correlation coefficient of global ancestry for mating pairs, we next explored an alternative metric of biased mating: the absolute difference in ancestry proportion between mate pairs, $\Delta_x = |x_1 - x_i|$. A small mean value of Δ_x indicates a tendency for individuals to have similar ancestry proportion to their mates. Importantly, however, the value of Δ_x cannot be interpreted outside the context of the population-specific variance in ancestry; $\overline{\Delta}_x$ is smaller under random mating than for a simulation with $\alpha = 10$ because the mean difference between any pair of individuals is smaller (**Figure S18**). Permutation tests for Δ_x were consistent with those for the correlation coefficient: the stationary-preference model was significantly different from random mating at generation 20 for $\alpha > 5$ ($p = 0.001$), whereas the broad-preference model became "random-like" within the first 5-10 generations post-admixture (**Figure S19**).

To compare between models, we devised a statistic that we named expressed mating bias, B, to quantify how much smaller $\overline{\Delta}_x$ is than expected given the simulation- and generation-specific variance of the population from mate pairs are drawn. Specifically, we define B as 1 less the ratio of $\overline{\Delta}_{observed}$ to the mean $\overline{\Delta}_{permuted}$ across permutations:

$$B = 1 - \overline{\Delta}_{observed} \Big/ \overline{\overline{\Delta}}_{permuted}.$$

We define B in this manner so that $B = 0$ when mating is random ($\overline{\Delta}_{observed} = \overline{\overline{\Delta}}_{permuted}$) and $B = 1$ when mating is so biased that no admixture is permitted ($\overline{\Delta}_{observed} = 0$). Within this range, increasing B corresponds to more biased mating.

We observed a similar relationship between the observed correlation between mates and B across models and time points, confirming that correlation coefficient largely reflects the degree of similarity between mates relative to the possible alternative pairs (**Figure 5A; Figure S20**). We observe a different trend when considering $\overline{\Delta}_{observed}$: individuals are less similar to their mates ($\overline{\Delta}_{observed}$ is greater) for the same correlation under the social group model compared to the increasing-preference model (**Figure S21**). This underscores the importance of accounting for bias in mate choice *relative to variance*.

As expected, higher variance in ancestry proportion corresponded to higher B across models. However, for the same value of B, simulations under the increasing-preference model had lower variance than those under the social group model (**Figure 5B**). This emphasizes a key distinction between the two models: the social group model sustains high correlation in ancestry between mates over tens of generations by maintaining higher variance in the population, whereas the increasing-preference model instead compensates for very low variance by explicitly scaling the mate-choice parameter c by the variance, increasing the strength of preference.

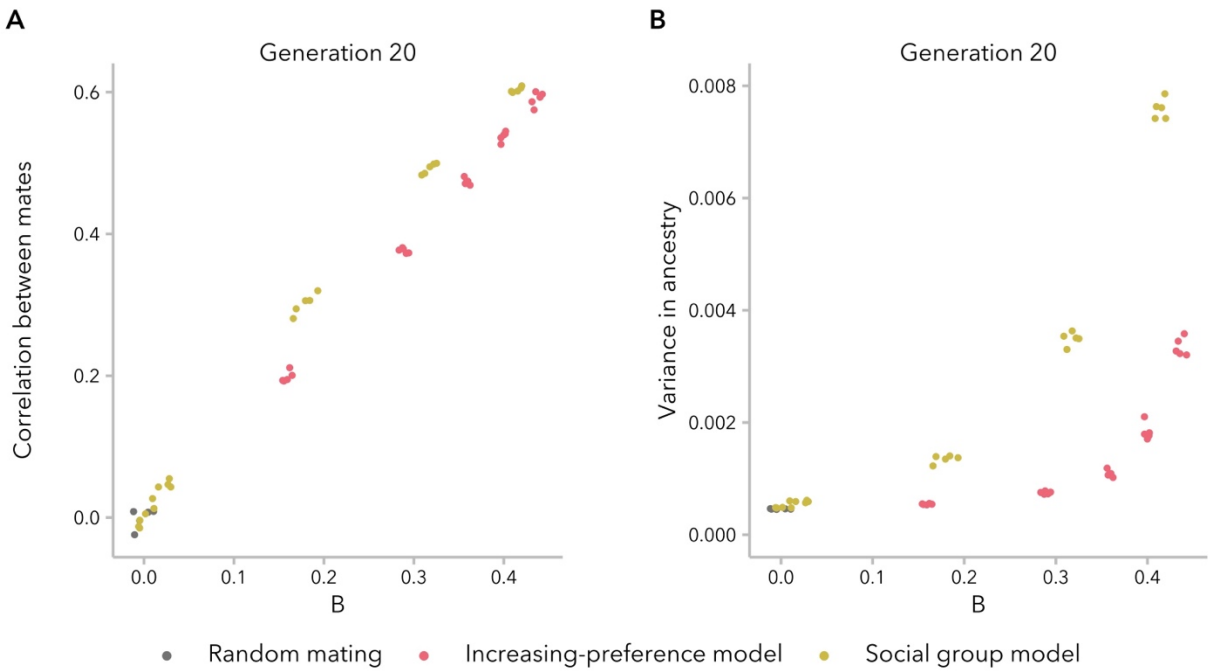

Figure 5. Variance in ancestry proportion within the population constrains the possibility of observing a correlation in ancestry between mates. **(A)** Correlation coefficients are correlated with an alternative metric of assortative mating, expressed mating bias B, which accounts for variance. Each dot represents one simulation. **(B)** Variance in ancestry increases with increasing B, but not uniformly across models. A given value of B corresponds to lower variance in ancestry under the increasing-preference model, emphasizing that biased mating is maintained in these simulations in the face of low variance.

Ancestry-assortative mating produced in an excess of long local-ancestry tracts relative to the exponential distribution under both model classes

Thus far, we have focused on the relationship between mating model and the distribution of global ancestry proportion. We next considered the impact of model choice on local-ancestry tract length. Local-ancestry tracts are used in population-genetic methods to infer demographic parameters, including the time since the onset of admixture (Moorjani *et al.* 2011; Gravel 2012; Loh *et al.* 2013; Hellenthal *et al.* 2014), and prior work has shown that non-random mating results in longer local-ancestry tracts, leading to underestimation of this parameter (Zaitlen *et al.* 2017; Korunes *et al.* 2022). Thus, some methods to correct for the effects of assortative mating attempt to recapitulate local-ancestry tract dynamics by matching the empirical correlation between spouses (Zaitlen *et al.* 2017).

We first compared the median local-ancestry tract length 20 generations post-admixture in simulations under the increasing-preference and social group models, matched for correlation between mates. Differences in median tract length were modest when comparing across α values within a single model, between these two mate-choice models, and between both models and random mating (**Figure 6A**). However, we did observe across all four mate-choice models that larger values of α were associated with longer median tract lengths at generation 20 (**Figure S22**). Furthermore, under the increasing-preference and social group models — the two models under which there was still a positive correlation between mates at generation 20 — greater correlations

between mates were also associated with longer median tract lengths. Additionally, for a given correlation between mates, median local tract length tended to be longest under the social group model (**Figure 6B**).

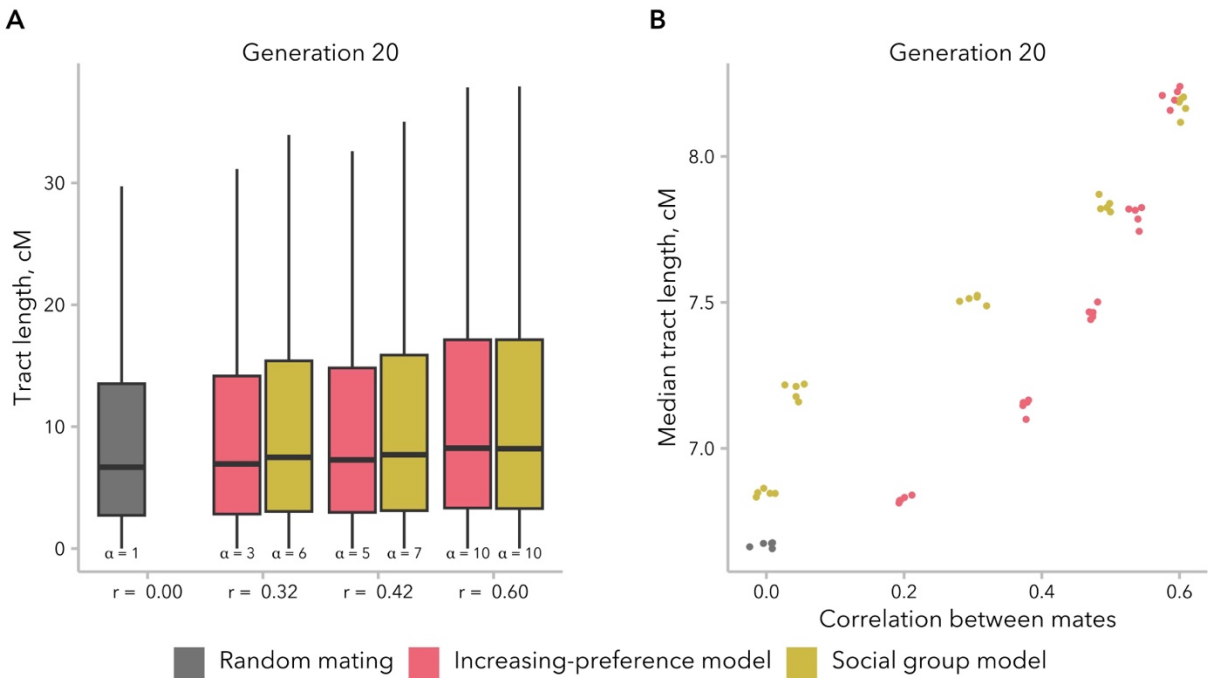

Figure 6. Local-ancestry tract length distributions were similar for simulations under the increasing-preference and social group models, when comparing simulations with similar correlation in ancestry between mates. **(A)** The distribution of local-ancestry tract lengths at generation 20 is shown for three representative pairs of simulations, compared to a random mating control. The y-axis is truncated at the top whisker (75th percentile plus 1.5 times the inter-quartile range). **(B)** Median local-ancestry tract length was similar across models and values of α. However, the median tract length tended to be longer for simulations under the social group model relative to those under the increasing-preference model when comparing simulations with the same correlation in ancestry between mates. Each dot represents one simulation (5 replicates per α).

To directly test the effects of the increasing-preference and social group models on downstream inference of admixture timing, we fit an exponential model to the local-ancestry tract length distribution to infer the time since a single admixture event (EQUATION 1 from (GRAVEL 2012)). We observed that the empirical distribution had a heavier right tail than the exponential fit, particularly as the correlation between mates increased (**Figure 7A, B**). Although slight, this mismatch in distribution shape may prove useful as an indication in empirical data that the effects of non-random mating should be taken into consideration, although it may also be confused for evidence of continuous migration.

As expected, we underestimated the time since admixture when using local-ancestry tract length distributions, and the discrepancy between the truth and our estimate increased as the correlation in ancestry between mates increased (**Figure 7C**). Intriguingly, we found that this discrepancy was established in early generations and grew relatively slowly over time, suggesting that there might be a plateauing of the effect on longer timescales (**Figure S23**). For instance, time since admixture

was underestimated by an average of 2.25 generations for 20 generations, and 4.01 generations for 50 generations. As a result, we observe parameter-inference bias of a similar magnitude under models for which a correlation in ancestry is no longer observed (**Figure S24**).

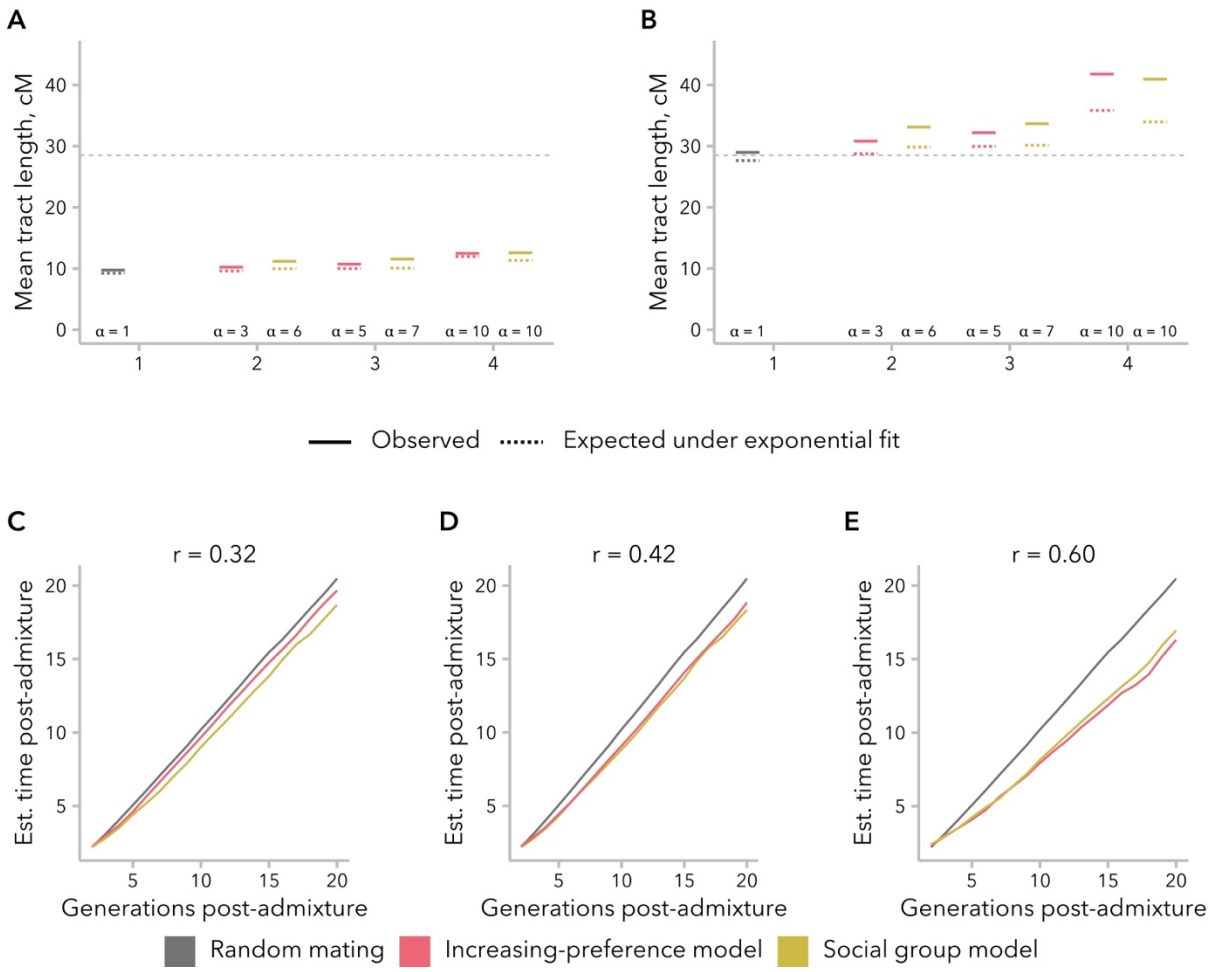

Figure 7. Non-random mating led to underestimation of the time since admixture. For each simulation, an exponential model was fit to the local-ancestry tract length distribution, with decay rate $\lambda = g \times m$ (g: generations since admixture; m: admixture contribution). **(A)** The observed local-ancestry tract length at generation 20 was similar to the mean tract length expected under the exponential distribution ($\frac{1}{\lambda}$), indicating a good fit to the data. The dashed line indicates the expected mean tract length for the true values of g and m. **(B)** The observed 95th percentile of the local-ancestry tract length distribution was longer than expected under the exponential fit: $Q_{95} = \frac{-\ln(1-0.95)}{\lambda}$, particularly when the correlation between mates is high. The dashed line indicates the expected 95th percentile tract length for the true values of g and m. **(C)** Estimates for the time since admixture based on the exponential fit to the local-ancestry tract distribution were similar between the increasing-preference and social group models. As expected, time since admixture was underestimated when mating was biased, and the discrepancy was greater when the correlation in ancestry between mates was greater.

The effects of continuous migration on the correlation in ancestry between mates differed between models and values of α

Realistic models of human admixture likely include more complex dynamics than a single pulse of admixture followed by complete isolation from the source populations, as we have modeled here. To begin to explore the behavior of the increasing-preference and social group models under more complex demographic scenarios, we examined the trajectory of the correlation in ancestry between mates over time in a scenario with continuous migration, wherein 1% of the population in each generation was replaced with migrants from the source populations. Relative to the single-pulse admixture scenario, we might expect that the correlation between mates would be decreased (for the same generation and α) because new migrants have fewer potential mates with similar ancestry proportion to choose from compared to individuals born in the admixed population. Furthermore, in each successive generation, variance in ancestry proportion decreases for the admixed population (**Figure 4**), meaning that new migrants have increasingly dissimilar ancestry proportion to the average potential mate. On the other hand, mating events between two migrants from the same source population will increase the correlation coefficient.

We found that the balance between these countervailing effects on the correlation between mates differed between models. Under the increasing-preference model, the correlation coefficient initially decreased over time, similar to the scenario with single-pulse admixture. However, for simulations with $\alpha \geq 4$, there was a subsequent increase in correlation over time, resulting in a greater correlation coefficient at generation 20 in the scenario with migration than the one without migration (**Figure 8A**). As noted above, the increasing-preference model compensates for decreasing variance by increasing mate selectiveness over successive generations (**Figure 5B**). Consistent with this explanation for the difference between the single-pulse admixture and continuous migration simulations, we observed that the proportion of mating events between migrants increased over time in these simulations (**Figure 8C**). Additionally, under the other two ancestry-similarity models, which do not compensate for the decreased variance over time, we observed that the prevalence of mating events between migrants was constant and that the correlation between mates was similar with or without migration for a given generation and α (**Figure S25; Figure S26; Figure S27**).

In contrast, the social group model neither takes into consideration similarity in ancestry proportion, nor increases the strength of mating bias over time. Thus, for simulations under this model, new migrants are equally likely to mate with anyone within their social group regardless of ancestry proportion; as expected, we observed a constant prevalence in mating events between migrants (**Figure 8D**). Additionally, because variance in ancestry decreases over time, migrants are more and more dissimilar from potential mates. As a result, continuous migration always decreased the correlation in ancestry between mates under this model relative to the single-pulse admixture scenario, controlling for generation and α (**Figure 8C**).

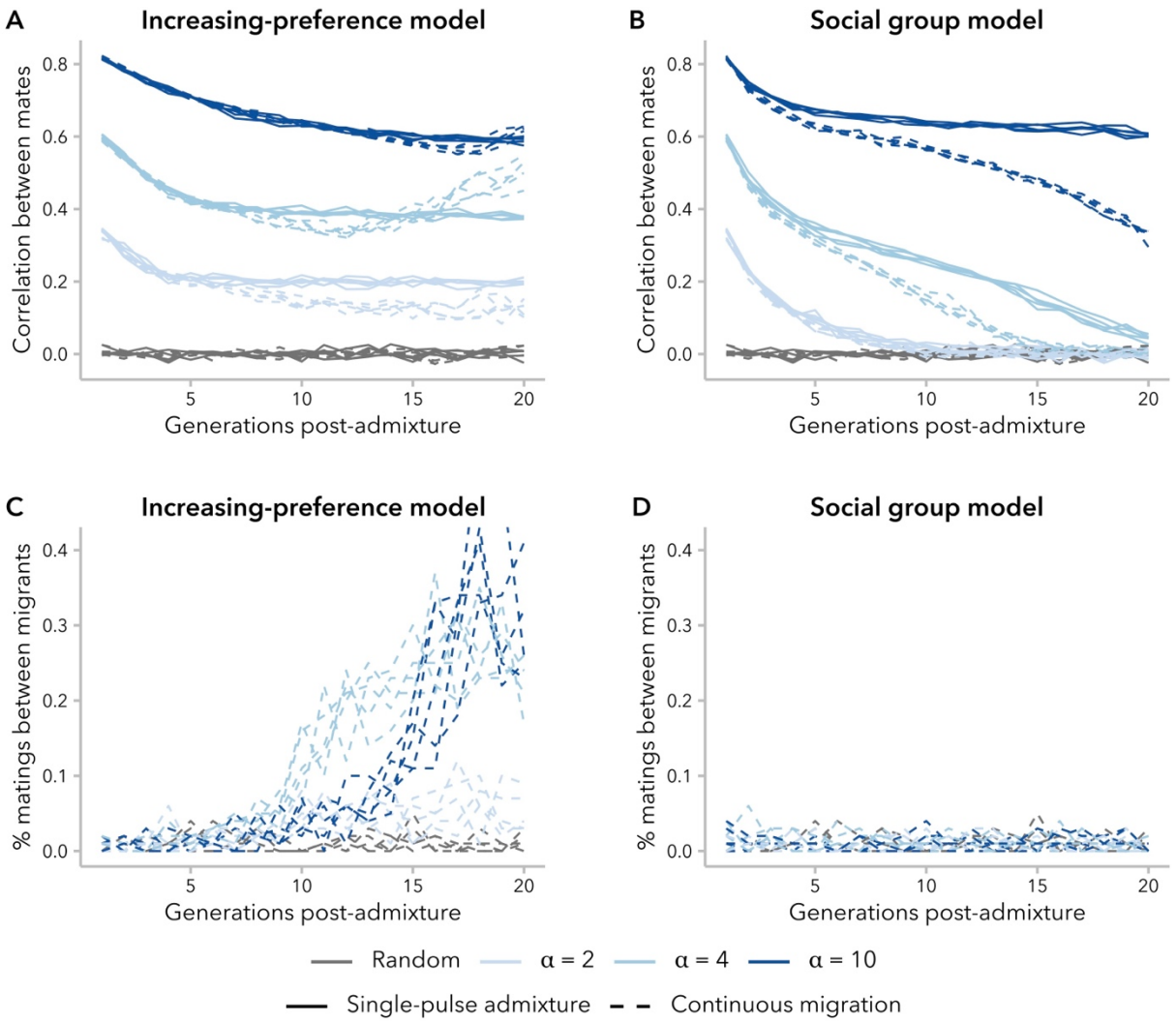

Figure 8. The effects of continuous migration differed between mate-choice models, driven by the percent of mating events between new migrants. **(A, B)** Correlation in ancestry between mates increased over time in simulations under the increasing-preference model with continuous migration, for some α. This behavior was not observed in any simulations without migration (see **Figure 2**) or under the social group model with continuous migration. **(C)** Under the increasing-preference model with continuous migration, mating between migrants was increasingly prevalent over time, and far more frequent than expected under random mating. **(D)** Mating between migrants under the social group model with continuous migration did not occur more often than expected by chance.

Discussion

While it is appreciated that humans, like individuals in many other natural populations, do not choose mates at random, population-genetic theory and methods to account for assortative mating in empirical data remain largely underdeveloped. In the context of ancestry-assortative mating, extension of existing theory on how assortative mating shapes the expected distribution of phenotypes in a population (WRIGHT 1921; NORRIS *et al.* 2019; KIM *et al.* 2021; BORDER *et al.* 2022; MURALIDHAR *et al.* 2022; HORWITZ *et al.* 2023) is made difficult by ambiguity about what the relevant phenotype is. Furthermore, global ancestry proportion is an unusual quantitative phenotype, in that its trait value is determined by every locus in the genome, meaning that the entire null distribution of any parameter of interest is likely to be affected. Simulations of ancestry-assortative mating that accurately recapitulate key summaries of empirical data are crucial to unraveling these impacts. Here, we considered two related prerequisite questions: first, are there multiple mechanisms of mate choice compatible with the observed correlation in global ancestry proportion in human populations? Second, does the choice of a particular mathematical function for defining biased mating meaningfully impact the conclusions drawn from the resulting simulations? We compared four models, including one that considers social groups rather than quantitative similarity in ancestry as the mechanism of mate choice, to better understand how we ought to think about modeling and correcting for ancestry-assortative mating.

We turn first to assumptions commonly made by statistical genetics methods (ZAITLEN *et al.* 2017; HUANG *et al.* 2023; PFENNIG AND LACHANCE 2023), which focus on the observed empirical correlation in ancestry between mates: namely, that this correlation is constant over time and that parameters of interest are comparable between empirical data and simulated data with the same correlation coefficient. In our simulations, we find that correlation coefficients are not stable over time and decay toward zero under most models (**Figure 2**). Thus, assuming that bias in mating is constant over time, the correlation between spouses observed in a contemporary sample is likely smaller than it was in previous generations. In addition, we find that multiple models can produce the same correlation between mates while differing in the underlying mating structure (**Figure 3**), average difference in ancestry between mates (**Figure 5**), and the right tail of the local-ancestry tract length distribution (**Figure 6**). However, we find that the effects of these differences between models on estimated time since admixture are likely to be small, likely because the effects on mean tract length are similar (**Figure 7**).

Prior studies that have modeled biased-mate choice to develop theory about assortative mating (KIM *et al.* 2021; MURALIDHAR *et al.* 2022) have, sometimes implicitly, assumed that these models are sufficient to generate a correlation in ancestry. We find that this assumption is not true for all models, bias parameters, and generations. Specifically, variance in ancestry proportion plays an essential role in determining whether a correlation between mates can be observed and, furthermore, whether mating is effectively random (**Figure 4; Figure 5**). This approach has also prioritized the role of individual mate-choice preference for similarity in ancestry ($|x_1 - x_i|$). In humans particularly, there is a wide array of social science research to support the role of sociological factors in mating outcomes. To wit, we found that a simplistic model imposing a barrier to mating between two social groups was sufficient to generate signatures of ancestry-assortative mating that could be observed for 20-50 generations post-admixture. To our knowledge, this type of model has not been used before to model ancestry-assortative mating in admixed populations

but may be representative of how social categories like race, ethnicity, and socioeconomic status influence spouse choice.

The four models that we consider in the present study are not the only possible models, but rather represent two classes of model worthy of further exploration: ancestry-similarity and social group models. For ancestry-similarity models, we have demonstrated that not all variants of this model type can sustain sufficiently high levels of variance to continue to observe positive correlations in ancestry (at least, without additional in-migration events). The increasing-preference model compensates for this decay in variance by increasing the choosiness of individuals. While this is effective in maintaining the correlation in ancestry, the current approach is likely making individuals unrealistically attuned to small differences between mates (**Figure S2B**). This overcompensation is heightened in the continuous-migration scenario, leading to unexpected (and undesirable) behavior (**Figure 8**). Future development of ancestry-similarity models should consider alternative approaches to tune the selectiveness of individuals over time. For instance, we might want to model a constant degree of bias (*i.e.*, $\max(\psi_i)\,/\min\,(\psi_i)$) over time.

In including the social group model in this study, we aimed only to demonstrate that social barriers can mediate ancestry-assortative mating, even when individuals do not directly use ancestry information to make mating decisions. To that end, we designed a proof-of-concept version of this model class, which is likely overly simplistic for drawing conclusions about real-world human populations. Future development of this class could include greater social complexity; for instance, more than two social groups, alternative rules for how individuals "inherit" their social group membership, and asymmetric barriers between groups. Each of these added layers could be implemented in different ways. For example, in a simulation with three groups, barriers might be more permeable between some pairs of groups than others.

A growing body of research suggests that patterns of assortative mating is not stable over time (*e.g.*, (MARE 1991; SUNDE *et al.* 2024)). Under an ancestry-similarity model, this could reflect changes in preference over time, potentially to account for decreasing differences between potential mates (as discussed above). Under a social group model, this could reflect changes in social mobility or acceptance of inter-group mating. Thus, modeling this type of change over time would be an interesting future direction for both classes of model, although it remains outside the scope of the present work.

Our results lead us to make several recommendations for future empirical studies of ancestry-assortative mating in humans. First, a correlation in ancestry between spouses can reflect multiple mechanisms of mate choice and should not be interpreted as unequivocal evidence of a preference for mates with similar ancestry proportion. Second, there are many different population structures that can give rise to the same Pearson correlation coefficient, and visualization of these structures may help to disambiguate these possibilities. In particular, rather than the more commonly used dot-plot, we suggest that density plots be used as an additional, often more informative, visualization tool. Third, difference in ancestry between mates ($\overline{\Delta}_{obs}$) was informative about differences between simulations with similar correlation coefficients. Importantly, this metric should only be interpreted relative to permutations of the data because its value is constrained by the variance in ancestry present in the dataset.

As with any simulation study, our results are limited by what we have elected to model or not model. Of particular note, we have focused on a single pulse of admixture, turning to a continuous-migration scenario only to highlight how one of our models behaves unexpectedly. More realistic models of human populations almost certainly involve much more complex migration dynamics, including multiple pulses of migration, differing contributions from source populations, changes in population size, asymmetries in mate preferences and in the ability of individuals to enact their mate choices, and multi-way admixture. Inclusion of at least some of these additional factors into future models is likely to impact the results; however, each additional factor drastically expands the range of possible implementations and parameter values. As such, these more complex models are likely to be more useful when targeted to matching the parameters of a particular population of interest whose history is well understood and less well suited to a general exploration of parameter space.

Data Availability

All scripts used for simulations, analyses, and figures is available on GitHub at: https://github.com/agoldberglab/ancestry-assortative-mating-simulation.

Acknowledgments

We thank Joshua Schraiber (ORCiD: 0000-0002-7912-2195), Michael M. Hoffman (ORCiD: 0000-0002-4517-1562), John Barton (ORCiD: 0000-0003-1467-421X), Shyamalika Gopalan (ORCiD: 0000-0002-2608-8472), and members of the Goldberg lab for helpful discussions.

Funding

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (R35 GM133481 to A.G. and R35 GM146926 to Z.A.S.).

References

Arauna, L. R., J. Bergstedt, J. Choin, J. Mendoza-Revilla, C. Harmant *et al.*, 2022 The genomic landscape of contemporary western Remote Oceanians. Curr Biol 32**:** 4565-4575 e4566.

Avadhanam, S., and A. L. Williams, 2022 Simultaneous inference of parental admixture proportions and admixture times from unphased local ancestry calls. Am J Hum Genet 109**:** 1405-1420.

Beleza, S., J. Campos, J. Lopes, Araujo, II, A. Hoppfer Almada *et al.*, 2012 The admixture structure and genetic variation of the archipelago of Cape Verde and its implications for admixture mapping studies. PLoS One 7**:** e51103.

Blossfeld, H.-P., 2009 Educational assortative marriage in comparative perspective. Annual Review of Sociology 35**:** 513-530.

Bonilla, C., M. D. Shriver, E. J. Parra, A. Jones and J. R. Fernandez, 2004 Ancestral proportions and their association with skin pigmentation and bone mineral density in Puerto Rican women from New York city. Hum Genet 115**:** 57-68.

Border, R., G. Athanasiadis, A. Buil, A. J. Schork, N. Cai *et al.*, 2022 Cross-trait assortative mating is widespread and inflates genetic correlation estimates. Science 378**:** 754-761.

Browning, J. R., 1951 Anti-miscegenation laws in the united states. Duke Bar Journal 1**:** 26-41.

Burger, R., and K. A. Schneider, 2006 Intraspecific competitive divergence and convergence under assortative mating. Am Nat 167**:** 190-205.

Burley, N., 1983 The meaning of assortative mating. Ethology and Sociobiology 4**:** 191-203.

Buss, D. M., and M. Barnes, 1986 Preferences in human mate selection. Journal of Personality and Social Psychology 50**:** 559-570.

Chiappori, P.-A., 2020 The theory and empirics of the marriage market. Annual Review of Economics 12**:** 547-578.

De La Mare, J. K., and A. J. Lee, 2023 Assortative preferences for personality and online dating apps: Individuals prefer profiles similar to themselves on agreeableness, openness, and extraversion. Personality and Individual Differences 208**:** 112185.

Dieckmann, U., and M. Doebeli, 1999 On the origin of species by sympatric speciation. Nature 400**:** 354-357.

Domingue, B. W., J. Fletcher, D. Conley and J. D. Boardman, 2014 Genetic and educational assortative mating among US adults. Proc Natl Acad Sci U S A 111**:** 7996-8000.

Fibla, J., I. Maceda, M. Laplana, M. Guerrero, M. M. Alvarez *et al.*, 2022 The power of geohistorical boundaries for modeling the genetic background of human populations: The case of the rural catalan Pyrenees. Front Genet 13**:** 1100440.

Fisher, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the Royal Society of Edinburgh 52**:** 399-433.

Fogel, A. S., E. M. McLean, J. B. Gordon, E. A. Archie, J. Tung *et al.*, 2021 Genetic ancestry predicts male-female affiliation in a natural baboon hybrid zone. Anim Behav 180**:** 249-268.

Funk, E. R., N. A. Mason, S. Palsson, T. Albrecht, J. A. Johnson *et al.*, 2021 A supergene underlies linked variation in color and morphology in a holarctic songbird. Nat Commun 12**:** 6833.

Goldberg, A., A. Rastogi and N. A. Rosenberg, 2020 Assortative mating by population of origin in a mechanistic model of admixture. Theor Popul Biol 134**:** 129-146.

Gravel, S., 2012 Population genetics models of local ancestry. Genetics 191**:** 607-619.

Greenwood, J., N. Guner, G. Kocharkov and C. Santos, 2014 Marry your like: Assortative mating and income inequality. American Economic Review 104**:** 348-353.

Haller, B. C., J. Galloway, J. Kelleher, P. W. Messer and P. L. Ralph, 2019 Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. Mol Ecol Resour 19**:** 552-566.

Haller, B. C., and P. W. Messer, 2023 SLiM 4: Multispecies eco-evolutionary modeling. Am Nat 201**:** E127-E139.

Hamid, I., K. L. Korunes, S. Beleza and A. Goldberg, 2021 Rapid adaptation to malaria facilitated by admixture in the human population of cabo verde. eLife 10.

Hellenthal, G., G. B. J. Busby, G. Band, J. F. Wilson, C. Capelli *et al.*, 2014 A genetic atlas of human admixture history. Science 343**:** 747-751.

Henz, U., and C. Mills, 2017 Social class origin and assortative mating in Britain, 1949–2010. Sociology 52**:** 1217-1236.

Horwitz, T. B., J. V. Balbona, K. N. Paulich and M. C. Keller, 2023 Evidence of correlations between human partners based on systematic reviews and meta-analyses of 22 traits and UK biobank analysis of 133 traits. Nat Hum Behav 7**:** 1568-1583.

Howe, L. J., T. Battram, T. T. Morris, F. P. Hartwig, G. Hemani *et al.*, 2021 Assortative mating and within-spouse pair comparisons. PLoS Genet 17**:** e1009883.

Huang, J., S. Basu, M. D. Shriver and A. A. Zaidi, 2023 Interpreting SNP heritability in admixed populations. bioRxiv.

Jiang, Y., D. I. Bolnick and M. Kirkpatrick, 2013 Assortative mating in animals. Am Nat 181**:** E125-138.

Kalmijn, M., 1998 Intermarriage and homogamy: Causes, patterns, trends. Annu Rev Sociol 24**:** 395-421.

Keddy-Hector, A. C., 1992 Mate choice in non-human primates. American Zoologist 32**:** 62-70.

Kim, J., M. D. Edge, A. Goldberg and N. A. Rosenberg, 2021 Skin deep: The decoupling of genetic admixture levels from phenotypes that differed between source populations. Am J Phys Anthropol 175**:** 406-421.

Kirkpatrick, M., 1982 Sexual selection and the evolution of female choice. Evolution 36**:** 1-12.

Klinkova, E., J. K. Hodges, K. Fuhrmann, T. de Jong and M. Heistermann, 2005 Male dominance rank, female mate choice and male mating and reproductive success in captive chimpanzees. International Journal of Primatology 26**:** 357-484.

Kondrashov, A. S., 1983 Multilocus model of sympatric speciation I. One character. Theoretical Population Biology 24**:** 121-135.

Kopp, M., M. R. Servedio, T. C. Mendelson, R. J. Safran, R. L. Rodriguez *et al.*, 2018 Mechanisms of assortative mating in speciation with gene flow: Connecting theory and empirical research. Am Nat 191**:** 1-20.

Korunes, K. L., G. B. Soares-Souza, K. Bobrek, H. Tang, Araujo, II *et al.*, 2022 Sex-biased admixture and assortative mating shape genetic variation and influence demographic inference in admixed Cabo Verdeans. G3 (Bethesda) 12.

Lande, R., 1981 Models of speciation by sexual selection on polygenic traits. Proc Natl Acad Sci U S A 78**:** 3721-3725.

Liang, M., and R. Nielsen, 2014 Understanding admixture fractions. bioRxiv**:** 008078.

Loh, P. R., M. Lipson, N. Patterson, P. Moorjani, J. K. Pickrell *et al.*, 2013 Inferring admixture histories of human populations using linkage disequilibrium. Genetics 193**:** 1233-1254.

Luo, S., and E. C. Klohnen, 2005 Assortative mating and marital quality in newlyweds: A couple-centered approach. J Pers Soc Psychol 88**:** 304-326.

Mare, R. D., 1991 Five decades of educational assortative mating. American Sociological Review 56**:** 15-32.

Mas Sandoval, A., S. Mathieson and M. Fumagalli, 2023 The genomic footprint of social stratification in admixing American populations. eLife 12**:** 2022.2011.2016.516754.

Melo, M. C., C. Salazar, C. D. Jiggins and M. Linares, 2009 Assortative mating preferences among hybrids offers a route to hybrid speciation. Evolution 63**:** 1660-1665.

Mooney, J. A., L. Agranat-Tamir, J. K. Pritchard and N. A. Rosenberg, 2023 On the number of genealogical ancestors tracing to the source groups of an admixed population. Genetics 224.

Moorjani, P., N. Patterson, J. N. Hirschhorn, A. Keinan, L. Hao *et al.*, 2011 The history of african gene flow into southern europeans, levantines, and jews. PLoS Genet 7**:** e1001373.

Muralidhar, P., G. Coop and C. Veller, 2022 Assortative mating enhances postzygotic barriers to gene flow via ancestry bundling. Proc Natl Acad Sci U S A 119**:** e2122179119.

Nagoshi, C. T., R. C. Johnson and G. P. Danko, 1990 Assortative mating for cultural identification as indicated by language use. Behav Genet 20**:** 23-31.

Nagylaki, T., 1978 The correlation between relatives with assortative mating. Ann Hum Genet 42**:** 131-137.

Natola, L., S. S. Seneviratne and D. Irwin, 2022 Population genomics of an emergent tri-species hybrid zone. Mol Ecol 31**:** 5356-5367.

Norris, E. T., L. Rishishwar, L. Wang, A. B. Conley, A. T. Chande *et al.*, 2019 Assortative mating on ancestry-variant traits in admixed latin American populations. Front Genet 10**:** 359.

Otto, S. P., M. R. Servedio and S. L. Nuismer, 2008 Frequency-dependent selection and the evolution of assortative mating. Genetics 179**:** 2091-2112.

Parra, F. C., R. C. Amado, J. R. Lambertucci, J. Rocha, C. M. Antunes *et al.*, 2003 Color and genomic ancestry in Brazilians. Proc Natl Acad Sci U S A 100**:** 177-182.

Pennings, P. S., M. Kopp, G. Meszena, U. Dieckmann and J. Hermisson, 2008 An analytically tractable model for competitive speciation. Am Nat 171**:** E44-71.

Pfennig, A., and J. Lachance, 2023 Challenges of accurately estimating sex-biased admixture from X chromosomal and autosomal ancestry proportions. Am J Hum Genet 110**:** 359-367.

Powell, D. L., B. M. Moran, B. Y. Kim, S. M. Banerjee, S. M. Aguillon *et al.*, 2021 Two new hybrid populations expand the swordtail hybridization model system. Evolution 75**:** 2524-2539.

Rettelbach, A., M. Kopp, U. Dieckmann and J. Hermisson, 2013 Three modes of adaptive speciation in spatially structured populations. Am Nat 182**:** E215-234.

Risch, N., S. Choudhry, M. Via, A. Basu, R. Sebro *et al.*, 2009 Ancestry-related assortative mating in Latino populations. Genome Biol 10**:** R132.

Robinson, C. E., H. Thyagarajan and A. K. Chippindale, 2023 Evolution of reproductive isolation in a long-term evolution experiment with Drosophila melanogaster: 30 years of divergent life-history selection. Evolution 77**:** 1756-1768.

Robinson, M. R., A. Kleinman, M. Graff, A. A. E. Vinkhuyzen, D. Couper *et al.*, 2017 Genetic evidence of assortative mating in humans. Nature Human Behaviour 1**:** 0016.

Schumer, M., D. L. Powell, P. J. Delclos, M. Squire, R. Cui *et al.*, 2017 Assortative mating and persistent reproductive isolation in hybrids. Proc Natl Acad Sci U S A 114**:** 10936-10941.

Schwartz, C. R., 2013 Trends and variation in assortative mating: Causes and consequences. Annual Review of Sociology 39**:** 451-470.

Sebro, R., T. J. Hoffman, C. Lange, J. J. Rogus and N. J. Risch, 2010 Testing for non-random mating: Evidence for ancestry-related assortative mating in the Framingham heart study. Genet Epidemiol 34**:** 674-679.

Sebro, R., G. M. Peloso, J. Dupuis and N. J. Risch, 2017 Structured mating: Patterns and implications. PLoS Genet 13**:** e1006655.

Sebro, R., and N. J. Risch, 2012 A brief note on the resemblance between relatives in the presence of population stratification. Heredity (Edinb) 108**:** 563-568.

Seger, J., 1985 Unifying genetic models for the evolution of female choice. Evolution 39**:** 1185-1193.

Setchell, J. M., and E. Jean Wickings, 2006 Mate choice in male mandrills (mandrillus sphinx). Ethology 112**:** 91-99.
Shriver, M. D., E. J. Parra, S. Dios, C. Bonilla, H. Norton *et al.*, 2003 Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet 112**:** 387-399.
Smadja, C. M., E. Loire, P. Caminade, D. Severac, M. Gautier *et al.*, 2022 Divergence of olfactory receptors associated with the evolution of assortative mating and reproductive isolation in mice. Peer Community Journal 2.
Smieja, M., and M. Stolarski, 2018 Assortative mating for emotional intelligence. Curr Psychol 37**:** 180-187.
Spear, M. L., A. Diaz-Papkovich, E. Ziv, J. M. Yracheta, S. Gravel *et al.*, 2020 Recent shifts in the genomic ancestry of Mexican Americans may alter the genetic architecture of biomedical traits. eLife 9.
Sunde, H. F., N. H. Eftedal, R. Cheesman, E. C. Corfield, T. H. Kleppesto *et al.*, 2024 Genetic similarity between relatives provides evidence on the presence and history of assortative mating. Nat Commun 15**:** 2641.
Tenesa, A., K. Rawlik, P. Navarro and O. Canela-Xandri, 2016 Genetic determination of height-mediated mate choice. Genome Biol 16**:** 269.
Torche, F., 2010 Educational assortative mating and economic inequality: A comparative analysis of three latin American countries. Demography 47**:** 481-502.
Tung, J., M. J. Charpentier, S. Mukherjee, J. Altmann and S. C. Alberts, 2012 Genetic effects on mating success and partner choice in a social mammal. Am Nat 180**:** 113-129.
Van Belle, S., A. Estrada, T. E. Ziegler and K. B. Strier, 2009 Sexual behavior across ovarian cycles in wild black howler monkeys (alouatta pigra): Male mate guarding and female mate choice. Am J Primatol 71**:** 153-164.
Veller, C., and G. M. Coop, 2024 Interpreting population- and family-based genome-wide association studies in the presence of confounding. PLoS Biol 22**:** e3002511.
Verdu, P., and N. A. Rosenberg, 2011 A general mechanistic model for admixture histories of hybrid populations. Genetics 189**:** 1413-1426.
Woodman, J. P., E. F. Cole, J. A. Firth, C. M. Perrins and B. C. Sheldon, 2023 Disentangling the causes of age-assortative mating in bird populations with contrasting life-history strategies. J Anim Ecol 92**:** 979-990.
Wright, S., 1921 Systems of mating. III. Assortative mating based on somatic resemblance. Genetics 6**:** 144-161.
Wright, S., 1950 Genetical structure of populations. Nature 166**:** 247-249.
Xie, Y., S. Cheng and X. Zhou, 2015 Assortative mating without assortative preference. Proc Natl Acad Sci U S A 112**:** 5974-5978.
Yamamoto, K., K. Sonehara, S. Namba, T. Konuma, H. Masuko *et al.*, 2023 Genetic footprints of assortative mating in the Japanese population. Nat Hum Behav 7**:** 65-73.
Zaitlen, N., S. Huntsman, D. Hu, M. Spear, C. Eng *et al.*, 2017 The effects of migration and assortative mating on admixture linkage disequilibrium. Genetics 205**:** 375-383.
Zou, J. Y., D. S. Park, E. G. Burchard, D. G. Torgerson, M. Pino-Yanes *et al.*, 2015 Genetic and socioeconomic study of mate choice in Latinos reveals novel assortment patterns. Proc Natl Acad Sci U S A 112**:** 13621-13626.