Received: 5 October 2020 | Revised: 7 February 2021 | Accepted: 10 February 2021

DOI: 10.1002/ajpa.24261

RESEARCH ARTICLE

American Journal of PHYSICAL ANTHROPOLOGY WILEY

Skin deep: The decoupling of genetic admixture levels from phenotypes that differed between source populations

Jaehee Kim[1] | Michael D. Edge[2] | Amy Goldberg[3] | Noah A. Rosenberg[1]

[1]Department of Biology, Stanford University, Stanford, California, USA

[2]Department of Biological Sciences, University of Southern California, Los Angeles, California, USA

[3]Department of Evolutionary Anthropology, Duke University, Durham, North Carolina, USA

Correspondence
Noah A. Rosenberg, Department of Biology, Stanford University, Stanford, CA 94305, USA.
Email: noahr@stanford.edu

Funding information
National Human Genome Research Institute, Grant/Award Number: HG005855; National Institute of General Medical Sciences, Grant/Award Numbers: GM130050, GM133481

Abstract

Objectives: In genetic admixture processes, source groups for an admixed population possess distinct patterns of genotype and phenotype at the onset of admixture. Particularly in the context of recent and ongoing admixture, such differences are sometimes taken to serve as markers of ancestry for individuals—that is, phenotypes initially associated with the ancestral background in one source population are assumed to continue to reflect ancestry in that population. Such phenotypes might possess ongoing significance in social categorizations of individuals, owing in part to perceived continuing correlations with ancestry. However, genotypes or phenotypes initially associated with ancestry in one specific source population have been seen to decouple from overall admixture levels, so that they no longer serve as proxies for genetic ancestry. Here, we aim to develop an understanding of the joint dynamics of admixture levels and phenotype distributions in an admixed population.

Methods: We devise a mechanistic model, consisting of an admixture model, a quantitative trait model, and a mating model. We analyze the behavior of the mechanistic model in relation to the model parameters.

Results: We find that it is possible for the decoupling of genetic ancestry and phenotype to proceed quickly, and that it occurs faster if the phenotype is driven by fewer loci. Positive assortative mating attenuates the process of dissociation relative to a scenario in which mating is random with respect to genetic admixture and with respect to phenotype.

Conclusions: The mechanistic framework suggests that in an admixed population, a trait that initially differed between source populations might serve as a reliable proxy for ancestry for only a short time, especially if the trait is determined by few loci. It follows that a social categorization based on such a trait is increasingly uninformative about genetic ancestry and about other traits that differed between source populations at the onset of admixture.

KEYWORDS
admixture, assortative mating, mechanistic model, population genetics

1 | INTRODUCTION

During intraspecific admixture processes, two or more long-separated populations merge to form a new admixed population. Viewed from a population-genetic standpoint, in an admixture process, distributions of genetic and phenotypic variation in the source populations combine to produce new distributions in the admixed group. The first generations after the onset of admixture generate transient dynamics

whose features are distinctive in relation to populations that are not admixed or for which admixture occurred only in the distant past (Gravel, 2012; Verdu & Rosenberg, 2011).

We seek to examine an aspect of emerging admixed populations. For admixed individuals, measurements of specific genotypes and phenotypes that differ in frequency or distribution between source populations can often provide reasonable estimates of individual levels of genetic ancestry in the particular source populations (Devillard et al., 2014; Parra et al., 1998; Shriver et al., 1997; Trigo et al., 2014). For some human phenotypes, such measurements might even be regarded by societies or admixed individuals themselves as proxies for overall genetic ancestry (Algee-Hewitt, 2016; Parra et al., 2004; Ruiz-Linares et al., 2014).

However, genotypes or phenotypes initially associated with ancestry in one source population at the start of an admixture process can decouple from overall admixture levels, so that they no longer serve as tight proxies for ancestry (Beleza et al., 2013; Durso et al., 2014; Leite et al., 2011; Magalhães da Silva et al., 2014; Parra et al., 2003, 2004; Pimenta et al., 2006; Ruiz-Linares et al., 2014). In human genetics, consider skin pigmentation and eye color, observable traits for which the phenotypic distribution differs substantially between sub-Saharan African and European populations. In the Cape Verdean admixed population, descended from European and West African sources, measurements of skin pigmentation and eye color are correlated with sub-Saharan African genetic ancestry (Beleza et al., 2013). At the same time, the correlations between phenotype and ancestry are imperfect; many individuals with a high proportion of sub-Saharan African genetic ancestry have skin pigmentation and eye color traits in a range more typical of individuals with higher European genetic ancestry, and vice versa. Similar patterns of incomplete correlation with overall genetic ancestry hold for genotypes that underlie these phenotypes (Beleza et al., 2013).

How does ancestry level decouple from genotype and phenotype in an admixed population? In humans, Parra et al. (2003) proposed one scenario for this decoupling, using an example of assortative mating by a phenotype correlated with ancestry in Brazil. They commented that in Brazil, assortative mating depends in part on *color*, a phenotypic measure based to a large extent on skin pigmentation. In their proposed hypothesis, in a population descended from source groups with substantially different skin pigmentation distributions (say, sub-Saharan Africans and Europeans), similarity according to a phenotype correlated with genetic ancestry (say, *color*) increases the probability that a pair is a mating pair. Mating probabilities for pairs of individuals are more closely related to the phenotype than to overall sub-Saharan African or European genetic admixture levels per se. Whereas in the early generations of such a process, the phenotype would strongly reflect genetic ancestry, after a sufficient length of time with assortative mating by the phenotype, phenotypic variation would be maintained, but with similar genetic ancestry distributions for individuals with substantially different phenotype (Figure 1). Only at genes associated with the phenotype and their nearby linked genomic regions would genetic ancestry and the phenotype be associated.

Could genetic ancestry in an admixed population become almost entirely decoupled from the phenotypes that differ between its source populations? This scenario would eliminate any connection between visible phenotypic markers of genetic ancestry and the genetic ancestry itself; the phenotype of an individual for a trait such as skin pigmentation would reveal little information about the genetic ancestry of molecular characters in the individual—other than for skin pigmentation genes and their closest genomic neighbors—nor about the total genomic ancestry of the individual.

To gain an understanding of the decoupling that can occur between phenotype and admixture, we develop a mechanistic model describing the joint dynamics of admixture levels and phenotype distributions in an admixed group. The approach includes a quantitative-genetic model that relates a phenotype to underlying loci that affect its trait value. We consider three forms of mating. First, individuals might mate randomly, independently of the overall admixture level. Second, individuals might assort by a phenotype that is initially correlated with the admixture level, but that is not identical to it. Third, individuals might assort by the admixture level itself. This latter case is meant to approximate situations in which correlated ancestry has been detected across mating pairs in admixed populations (Risch et al., 2009; Zou et al., 2015), potentially reflecting assortative mating by multidimensional phenotypes tightly correlated with admixture. Under the model, we explore the relationship between admixture level and phenotype over time, studying the effect of the mating model and the genetic architecture of the phenotype.

2 | MODEL

2.1 | Population model

Our mechanistic admixture model closely follows the model of Verdu, Goldberg, and Rosenberg (Goldberg et al., 2014; Goldberg & Rosenberg, 2015; Verdu & Rosenberg, 2011), building on earlier related models (Ewens & Spielman, 1995; Guo et al., 2005; Long, 1991). We start with individuals in each of two isolated source populations, S_1 and S_2. At the founding of an admixed population ($g = 0$), a founding parental pool H_0^{par} is formed, containing fraction $s_{1,0}$ from population S_1 and $s_{2,0}$ from population S_2. That is, a random individual in H_0^{par} originates from population S_1 with probability $s_{1,0}$ and from S_2 with probability $s_{2,0}$. This choice requires $s_{1,0}+s_{2,0} = 1$ and $0 \le s_{1,0},\ s_{2,0} \le 1$. The individuals in the founding parental pool mate according to a mating model and produce generation $g = 1$ of admixed offspring (H_1).

In subsequent generations ($g \ge 1$), in forming an admixed population H_{g+1} at generation $g + 1$, three populations contribute to its parental pool H_g^{par}: the source populations (S_1 and S_2) and the admixed population (H_g) of the previous generation, with fractional contributions $s_{1,g}$, $s_{2,g}$, and h_g, respectively. Here, $s_{1,g}$, $s_{2,g}$, and h_g represent probabilities for a random individual in H_g^{par} to originate from populations S_1, S_2, and H_g, with constraints $s_{1,g}+s_{2,g}+h_g = 1$ and

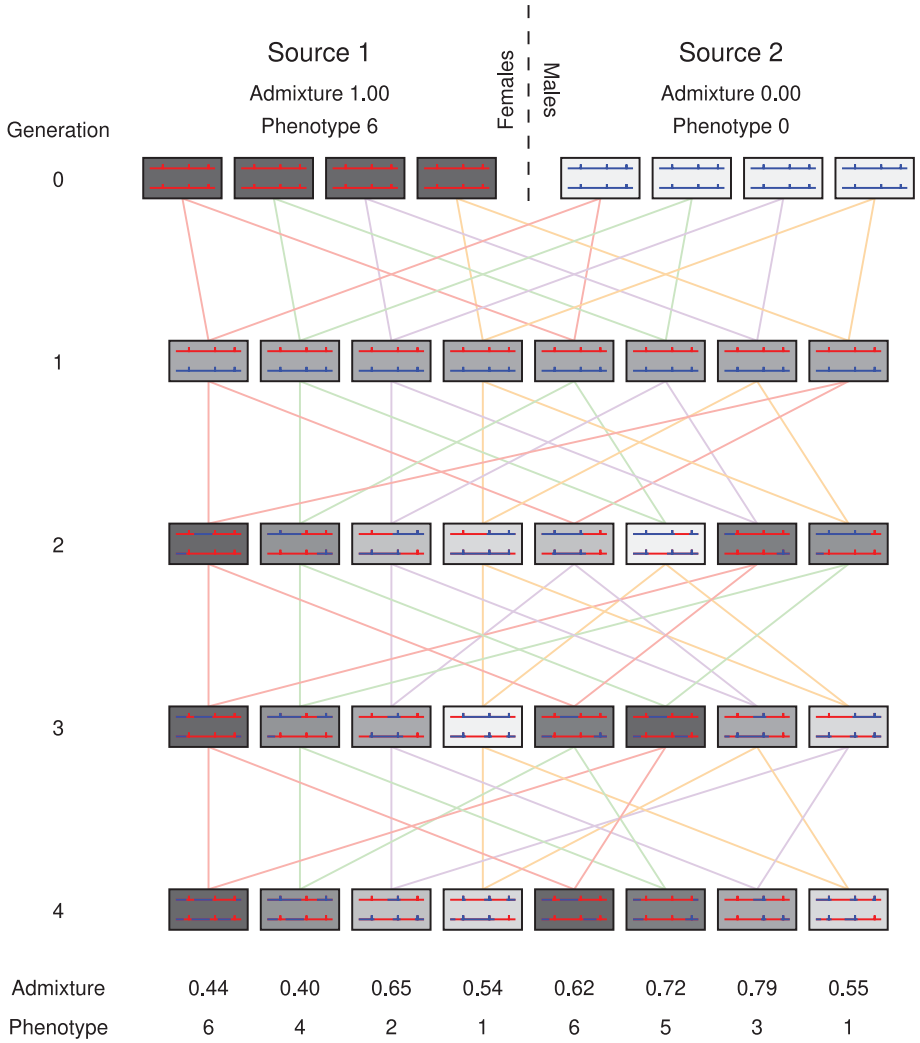

FIGURE 1 A schematic of an admixture process with positive assortative mating by a phenotype initially correlated with admixture levels. In generation 0, an admixture process begins with females from one population (source 1, left) and males from another (source 2, right). For a quantitative phenotype, source population 1 begins with a high trait value of 6 and source population 2 has a low trait value of 0. Three loci contribute additively to the genetic architecture of the phenotype; each allele derived from source population 1 contributes a value of 1 to the phenotype. The phenotype is represented by the shading of a box. Individuals are depicted as pairs of chromosomes with the ancestral sources of those chromosomes; short vertical lines along the chromosome indicate the three loci that contribute to the phenotype. After generation 1, positive assortative mating by phenotype proceeds in the admixed population. Lines connecting generations are displayed in four colors, representing four mating pairs. Initially, in generation 2, a strong correlation exists between admixture and phenotype ($r = 0.96$). By generation 4, however, owing to recombination events that stochastically dissociate the trait loci from the overall genetic admixture, the genetic admixture has been decoupled from the phenotype, so that some of the individuals with the highest trait values have among the lowest admixture coefficients for source population 1, and the correlation between phenotype and overall genetic admixture has dissipated ($r = -0.09$)

$0 \leq s_{1,g}, s_{2,g}, h_g \leq 1$. Offspring from matings in the parental pool H_g^{par} define the admixed population H_{g+1}. A schematic appears in Figure 2.

The total admixture fraction represents the proportion of the genome of an individual originating from a specific ancestral population, S_1 or S_2. We denote an individual's admixture fraction from a source population S_1 at generation g by $H_{A,g}$, with the A indicating consideration of autosomal genetic loci. Given a pair of individuals with admixture fractions $H_{A,g}^{(1)}$ and $H_{A,g}^{(2)}$, the ancestry of their offspring is deterministically set to the mean of the admixture fractions of the parents: $H_{A,g+1} = \frac{1}{2}\left[H_{A,g}^{(1)} + H_{A,g}^{(2)}\right]$. The possible values for the admixture fraction at generation g, representing possible values for the fraction of genealogical ancestors g generations ago who were in the source S_1, are $0, 1/2^g, 2/2^g, \ldots, (2^g - 1)/2^g, 1$.

2.2 | Quantitative trait model

To model a phenotype, we adopt the approach of Edge & Rosenberg (2015a, 2015b). Each individual is diploid, and k biallelic autosomal loci, each with the same effect size, additively determine the value of a quantitative trait. At each trait locus, we denote the allelic type more prevalent in S_1 than in S_2 as allelic type "1," and the

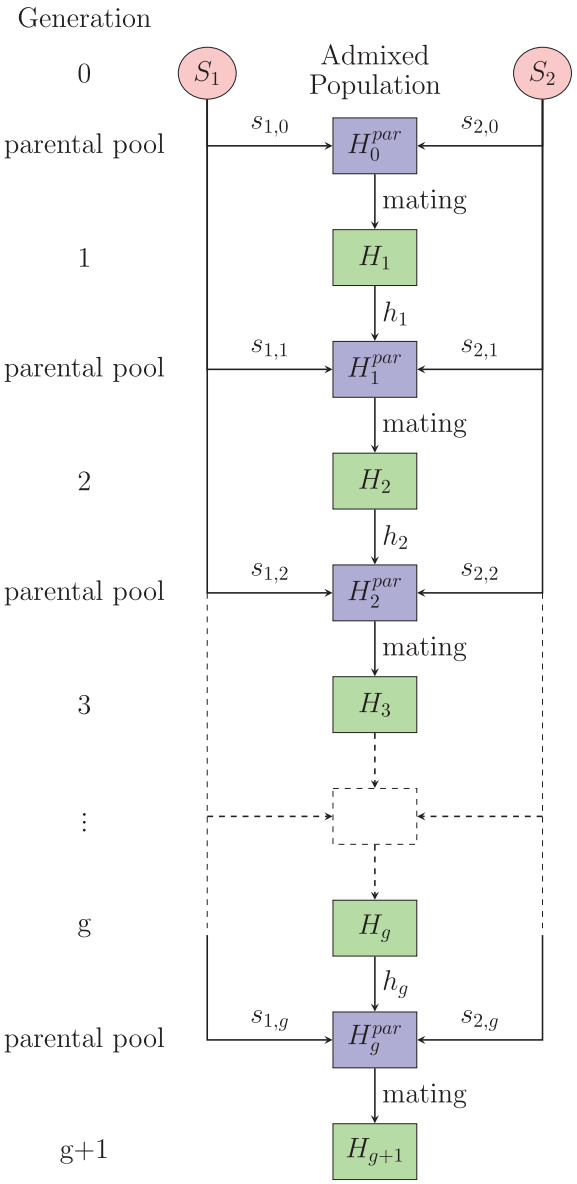

FIGURE 2 A schematic diagram of the admixture process. At the founding of the population ($g = 0$), two isolated source populations produce the first generation of an admixed population (H_1). In the subsequent generations ($g \geq 1$), populations from S_1, S_2, and H_g provide a parental pool H_g^{par} at generation g from which the admixed population H_{g+1} at generation $g + 1$ is produced. Fractional contributions from three populations in forming the parental pool are $s_{1,g}$, $s_{2,g}$, and h_g, respectively. Individuals in the parental pool mate based on mating models described in Section 2.3

	Locus							
	1	2	3	4	5	6	7	8
L_{i1}	1	0	0	1	1	1	0	1
L_{i2}	0	0	0	1	1	1	1	1
X_i	1	1	0	0	0	1	1	0
Contribution to T	1	0	2	0	0	2	1	0
	$T = 6$							

FIGURE 3 An example of the quantitative trait model. Here, a diploid individual with $k = 8$ trait loci is shown. At each locus i, an allele L_{ij} contributes to the overall trait value if and only if $L_{ij} = X_i$, where X_i is a variable indicating which of two alleles, "0" or "1," increases the trait value. The total trait value of an individual equals the number of alleles satisfying $L_{ij} = X_i$ across the k trait loci. In this example, the individual has $T = 6$

other allelic type as "0." The choice is arbitrary if the allele frequency is the same in the two populations. A diploid individual's genotype at locus i, $1 \leq i \leq k$, and allele j, $j = 1$ or 2, is represented by a random indicator variable L_{ij} : $L_{ij} = 1$ if the allele has type "1" and $L_{ij} = 0$ if it has type "0."

Let M be a random variable representing an individual's population membership, S_1 or S_2, and define allele frequencies for allelic type "1" at each locus given the membership: $P(L_{ij} = 1 | M = S_1) = p_i$ and $P(L_{ij} = 1 | M = S_2) = q_i$. Here, j can be either 1 or 2. By definition of allelic type "1," $0 \leq q_i \leq p_i \leq 1$.

An individual's trait value is determined by a sum of contributions across loci. At each locus, we denote an allele that increases the trait value by "+" and the other allele by "–." Whether the "1" type or the "0" type is the "+" allele at locus i is specified by a random variable X_i (Edge & Rosenberg, 2015a, 2015b): $X_i = 1$ if allelic type "1" is the "+" allele at locus i, and $X_i = 0$ if allelic type "0" is the "+" allele at locus i.

For a given set of values $\{X_1, X_2, \cdots, X_k\}$ for k quantitative trait loci, the total trait value T for a diploid individual is equal to the total number of "+" alleles carried by the individual, or $T|_{\{X_1, X_2, \cdots, X_k\}} = \left[\sum_{\{i:X_i=1\}} \sum_{j=1}^{2} L_{ij}\right] + \left[\sum_{\{i:X_i=0\}} \sum_{j=1}^{2} (1 - L_{ij})\right]$. This quantity takes values in {0, 1, ..., 2k}. An example of the quantitative trait model appears in Figure 3.

We consider an idealized case in which the number of "1" alleles correlates perfectly with trait value: $P(X_i = 1) = 1$ and $P(X_i = 0) = 0$ for all $i = 1, 2, \ldots, k$, so that allelic type "1" is the "+" allele and type "0" is the "–" allele for all trait loci. Because we define "1" to be the more frequent allelic type in the source population S_1, individuals from S_1 are more likely than are those from S_2 to have a large trait value. This idealized scenario considers a case in which the phenotype differs systematically between populations 1 and 2, and is depicted in Figure 1. The idealized case is instructive owing to its simplicity; a more complex scenario is considered in the Supporting Information.

2.3 | Mating model

We consider three mating models: (a) random mating, in which the probability that a pair consisting of a male and a female is a mating pair does not depend on the phenotypes or ancestries of the individuals; (b) assortative mating by admixture, in which this probability depends on their ancestries; and (c) assortative mating by phenotype, in which it depends on their trait values. For completeness, we include

negative assortative mating in describing our model, but our simulations focus on positive assortative mating.

In each generation g, the parental pool H_g^{par} contains $2N$ individuals, N female and N male. The admixture fraction from the source population S_1 and trait value of a female individual i are denoted by $H_{A,g}^{(i),f}$ and $T_g^{(i),f}$, respectively. Analogous quantities for a male j are $H_{A,g}^{(j),m}$ and $T_g^{(j),m}$.

We construct an $N \times N$ mating matrix M. Entry m_{ij} gives the probability that if a mating pair is selected at random, female i and male j are chosen. N mating pairs are drawn with replacement, with pair (i, j) given weight m_{ij}. An individual can be drawn multiple times, appearing in more than one mating pair.

The full specification of the m_{ij} is described in Appendix A for the three mating models, in terms of a parameter c. The value of c is 0 for random mating; increasing $|c|$ increases the strength of assortative mating, with $c > 0$ corresponding to positive assortative mating and $c < 0$ to negative assortative mating.

2.4 | Expectation and variance of the admixture fraction

To interpret our simulations of admixture dynamics, we will use results on the mean and variance of the admixture fraction in the admixed population. Let $H_{A,g}$ be a random variable representing the admixture fraction of an individual chosen at random in the admixed population H_g at generation $g \geq 1$. We denote by $\left(H_{A,g}^{f,p}, H_{A,g}^{m,p}\right)$ the admixture fractions of the members of a mating pair chosen at random from the parental pool H_g^{par} in generation $g \geq 0$; the superscript p denotes that the individual is from the parental pool.

The parental pool H_g^{par}, from which admixed population H_{g+1} is formed, consists of populations S_1, S_2, and H_g, with fractional contributions $s_{1,g}$, $s_{2,g}$, and h_g, respectively (Section 2.1 and Figure 2). Each population (S_1, S_2, H_g) has equally many males and females, each constant at N. Each individual has the same expected number of offspring, and no sex bias by the population of origin exists in parental pairings (Section 2.3); $H_{A,g}^{f,p}$ and $H_{A,g}^{m,p}$ are identically distributed. The quantity $r_{H_A,g} = \text{Cor}\left[H_{A,g}^{f,p}, H_{A,g}^{m,p}\right]$ gives the correlation of admixture fractions in a mating pair.

In Appendix B, we derive a relationship between the variance of the admixture fraction and the correlation in admixture levels for members of mating pairs. For a special case of a single admixture event in which source populations S_1 and S_2 do not contribute to the admixed population after its founding ($s_{1,g} = s_{2,g} = 0$ and $h_g = 1$ for all $g \geq 1$), Appendix B shows that the expectation of the admixture fraction stays constant in time (Equation (B5)), and that the variance reduces to a simple formula (Equation (B6)):

$$\text{Var}[H_{A,g+1}] = \frac{1}{2}(1 + r_{H_A,g})\text{Var}[H_{A,g}]. \quad (1)$$

Under random mating in an infinite population with no ongoing source contributions, with $r_{H_A,g} = 0$ for all $g \geq 0$, Equation (1) reduces to $\text{Var}[H_{A,g}] = s_{1,0}(1 - s_{1,0})/2^g$ (Verdu & Rosenberg, 2011). Equation (1) was also derived by Zaitlen et al. (2017), under different assumptions (notably, mating correlation $r_{H_A,g} = r$ constant in g).

3 | SIMULATION

3.1 | Simulation procedure

Having specified the populations of interest, the properties of trait values in the populations, and the mating probabilities for pairs of individuals, we now describe how we simulate populations under the model. At the first time step ($g = 0$), $s_{1,0}N$ and $s_{2,0}N$ males are randomly generated without replacement from the source populations S_1 and S_2, respectively, with $s_{1,0} + s_{2,0} = 1$. The corresponding numbers of females $s_{1,0}N$ and $s_{2,0}N$ are randomly generated without replacement from S_1 and S_2, respectively, contributing to the founding parental pool H_0^{par} of $2N$ individuals, with N males and N females. All individuals in S_1 have admixture fraction 1, and all individuals in S_2 have admixture fraction 0, by definition. For each individual in S_1 and S_2, genotypes at each of k quantitative trait loci are then randomly generated based on prespecified allele frequencies p_i and q_i.

We assume fixed differences between source populations at all trait loci, so $p_i = 1$ and $q_i = 0$ for all k loci. Each individual in S_1 has allele "1" at all trait loci, and each individual in S_2 instead has allele "0." In subsequent generations, allele "1" can be traced back to S_1, and allele "0" to S_2 (Figure 2). This choice for the p_i and q_i models a case in which trait-influencing alleles are initially entirely predictive of ancestry and vice versa. An alternative choice of the p_i and q_i appears in the Supporting Information.

As described in Appendix A, we compute an $N \times N$ mating matrix M. Considering all N^2 potential mating pairs, we randomly draw N mating pairs with replacement from the parental pool, weighting mate choices by mating probabilities in M. Each mating pair produces two offspring, one male and one female, to maintain constant population size for the offspring generation: N males, N females. An offspring admixture fraction is then assigned as the mean of its parental admixture fractions. Assuming no linkage disequilibrium and no mutation, the offspring genotype at the trait loci is then determined by independently selecting at each locus one random allele from one parent and one from the other. The $2N$ offspring individuals form the admixed population H_1 at generation 1.

In subsequent generations $g \geq 1$, we randomly select without replacement $s_{1,g}N$, $s_{2,g}N$, and h_gN males and $s_{1,g}N$, $s_{2,g}N$, and h_gN females from S_1, S_2, and H_g, respectively, forming a gth generation parental pool H_g^{par} of $2N$ individuals, consisting of N males and N females. The procedure to generate the offspring population H_{g+1} from H_g^{par} is the same as the procedure for generating H_1 from H_0^{par}.

Throughout the simulation, we keep the population size parameter N constant at 1000 for computational efficiency. The admixed population size (N) need not be identical to the source population sizes. For each set of parameters, (k, p_1, p_2, …, p_k, q_1, q_2, …, q_k, X_1, X_2, …, X_k, c, $s_{1,0}$, $s_{1,g}$, $s_{2,g}$), we proceeded to $G = 40$ generations, with 100 independent trajectories for each parameter set. We then averaged statistics of interest over the 100 trajectories.

3.2 | Base case

We start with an idealized base case that is instructive for characterizing model behavior. We then consider increasingly complex cases to explore the effects of the parameters.

First, we specify the parameters involving the population model (Section 2.1). We assume an equal influx from each source population at founding $g = 0$: $s_{1,0} = 0.5$, $s_{2,0} = 1 - s_{1,0} = 0.5$. We also assume no additional contributions from the source populations in the subsequent generations, $s_{1,g} = s_{2,g} = 0$, and $h_g = 1 - s_{1,g} - s_{2,g} = 1$ for all $g \geq 1$.

Next, we choose parameter values for the quantitative trait model (Section 2.2). We consider $k = 10$ trait loci. Across the k loci, all "1" alleles come from the source population S_1 and all "0" alleles come from S_2: $p_i = 1$ and $q_i = 0$ for all $i = 1, 2, \ldots, k$. For each locus i contributing to the quantitative trait, we define "1" to be the "+" allele and "0" to be the "−" allele: $X_i = 1$ for all $i = 1, 2, \ldots, k$.

For the mating model (Section 2.3), we set the assortative mating strength to $c = 0.5$.

3.3 | Statistics measured

In each simulated admixed population, in each generation g, we computed the correlation between admixture fraction and trait (Cor[H_A, T]), variance of the admixture fraction in the population (Var[H_A]), and variance of the trait value in the population (Var[T]). In the Results, we discuss how these statistics of interest change as we modify the simulation parameters.

4 | RESULTS

We first examine the base case to understand the general behavior of the model. Next, we study the effects of the assortative mating strength and the number of loci contributing to the quantitative trait. We also examine two additional model features—choosing allele frequencies in the source populations according to genetic drift from a shared ancestral population, rather than assuming fixed differences between source populations, and decoupling alleles that are more common in a specific source population ("1" alleles) and trait-increasing alleles ("+" alleles). These latter changes produce similar results, dampening larger effects seen with the base case; we describe them in the Supporting Information.

4.1 | Base case

4.1.1 | Correlation between ancestry and phenotype (Cor[H_A, T])

In the base case, each individual from S_1 has admixture fraction $H_A = 1$ and trait value $T = k$, and each individual from S_2 has $H_A = 0$ and $T = 0$. Therefore, in the founding parental pool H_0^{par}, admixture fraction and trait value are perfectly correlated: Cor[H_A, T] = 1. Subsequently, however, the correlation starts to decouple, as illustrated in Figure 1. With all parameters of the population model and quantitative trait model fixed, the decay in the correlation Cor[H_A, T] depends on the mating model.

The correlation is compared under the three mating models using base-case parameters in Figure 4e. Irrespective of the mating model, the founding parental pool has a perfect correlation. Even if the population starts with perfect correlation between admixture fractions and trait values, however, then random mating rapidly decouples them (red curve). The correlation decreases below 0.5 in 6 generations of random mating (Cor[H_A, T] = 0.490). After $g = 20$ generations, it is 0.137, and it is near zero at $g = 40$ (−0.003).

Compared to random mating, positive assortative mating slows the decoupling of admixture fractions and trait values. Assortative mating by phenotype (green curve in Figure 4e) maintains the correlation longer than assortative mating by admixture fraction (blue curve in Figure 4e). It takes 11 generations under assortative mating by phenotype for the correlation to drop below 0.5 (Cor[H_A, T] = 0.490), and 10 generations under assortative mating by admixture (Cor[H_A, T] = 0.443). Across the 40 generations we simulated, Cor[H_A, T] is consistently higher under assortative mating by phenotype than under assortative mating by admixture fraction. The correlation decreases to 0.227 at $g = 20$ and 0.043 at $g = 40$ under assortative mating by phenotype. The corresponding values under assortative mating by admixture are 0.065 at $g = 20$ and 0.009 at $g = 40$, both considerably lower than under assortative mating by phenotype.

In comparison with random mating, both assortative mating models have higher probabilities for matings within source populations, and thus, the proportion of individuals produced in the admixed population at $g = 1$ that are genetically admixed is smaller (blue and green lines in the marginal plots for H_A in Figure S1A). Over time, as displayed in Figure S1, random mating pulls individuals away from the source populations, pushing the H_A and T distributions toward the mean values rapidly. Both assortative mating models maintain individuals with H_A and T values near the source population values for longer, and thus, they retain a higher correlation Cor[H_A, T] than random mating.

The difference in the correlation Cor[H_A, T] between the two assortative mating models arises from the difference between the variance of admixture, Var[H_A], and the variance of the trait value, Var[T]. The covariance Cov[H_A, T] is similar under the two models. Given the similar covariance,

$$\frac{\text{Cor}[H_A,T]_{\text{p}}}{\text{Cor}[H_A,T]_{\text{g}}} \approx \sqrt{\frac{\text{Var}[H_A]_{\text{g}}}{\text{Var}[H_A]_{\text{p}}} \times \frac{\text{Var}[T]_{\text{g}}}{\text{Var}[T]_{\text{p}}}},$$

where "g" and "p" indicate the property on which mating pairs assort: genetic admixture or phenotype. As we will show, both assortative mating models increase both variances compared with random mating, particularly for the property on which mating assorts: $\text{Var}[H_A]_{\text{g}} > \text{Var}[H_A]_{\text{p}}$ and $\text{Var}[T]_{\text{p}} > \text{Var}[T]_{\text{g}}$. We will see that the

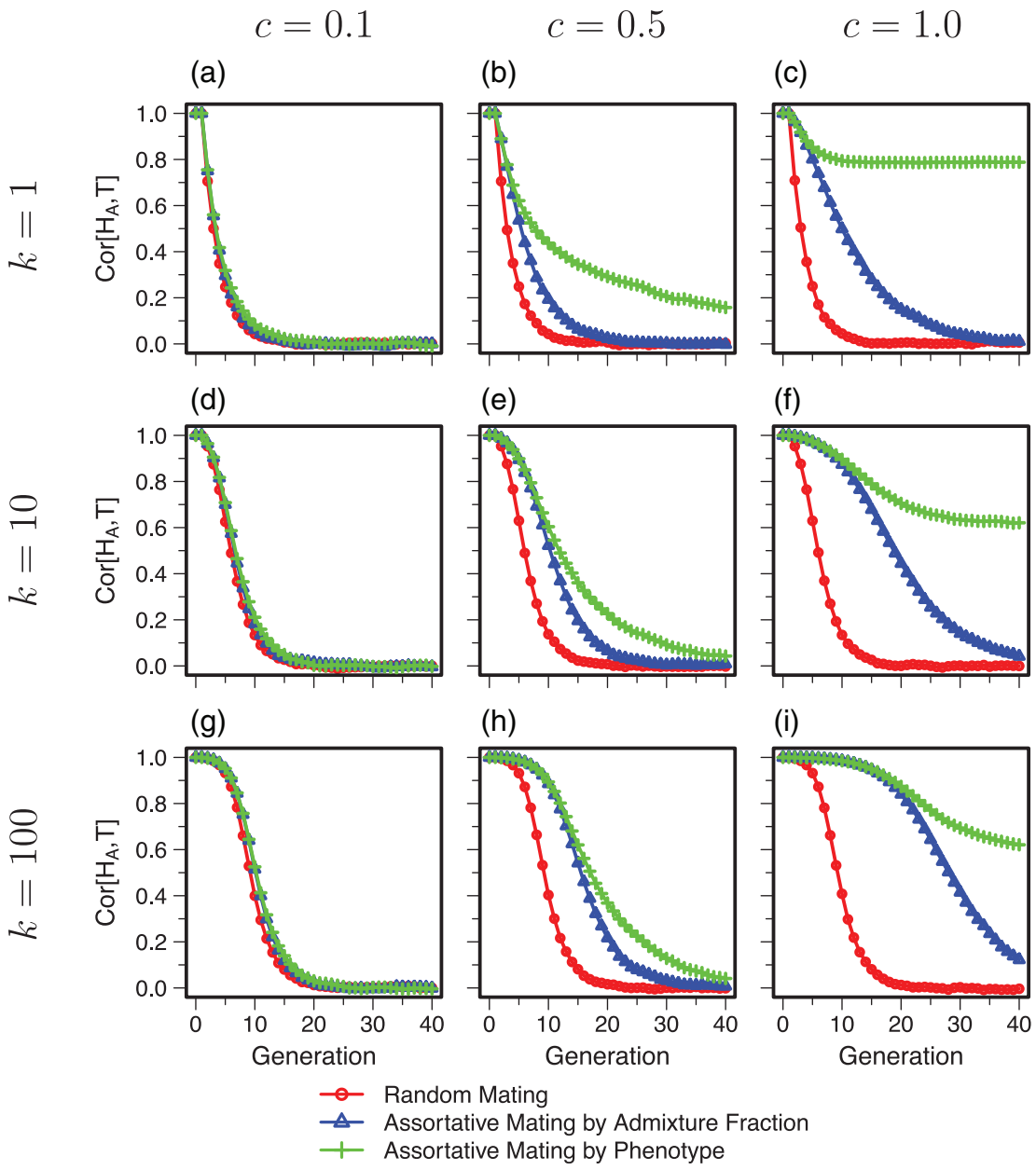

FIGURE 4 Correlation between admixture fraction and quantitative trait value (Cor[H_A, T]) as a function of time. All parameter values in panel (e) follow the base case; the number of quantitative trait loci k and the assortative mating strength c vary across panels. In each panel, for a given (k, c) pair, for each mating scheme, the mean of 100 simulated trajectories is plotted. The red, blue, and green curves represent results from random mating, assortative mating by admixture fraction, and assortative mating by phenotype, respectively. (a) k = 1, c = 0.1. (b) k = 1, c = 0.5. (c) k = 1, c = 1.0. (D) k = 10, c = 0.1. (E) k = 10, c = 0.5. (f) k = 10, c = 1.0. (g) k = 100, c = 0.1. (h) k = 100, c = 0.5. (i) k = 100, c = 1.0

increase in the variance of admixture due to assortative mating by admixture fraction exceeds that in the variance of the trait value due to assortative mating by trait:

$$\frac{\mathrm{Var}[H_A]_g}{\mathrm{Var}[H_A]_p} > \frac{\mathrm{Var}[T]_p}{\mathrm{Var}[T]_g}.$$

This result leads to a higher correlation Cor[H_A, T] under assortative mating by trait compared with that under assortative mating by admixture fraction.

4.1.2 | Variance of admixture and phenotype (Var[H_A] and Var[T])

Each individual in S_1 has admixture fraction 1, and each individual in S_2 has admixture 0. In the founding parental pool, Var[H_A] = 0.250 for all three mating models. As discussed in Section 2.4, the variance of the admixture fraction can be understood in relation to the correlation Cor$\left[H^f_{A,g}, H^m_{A,g}\right]$ of admixture fractions of members of mating pairs. Figure S2 shows this correlation for the

simulations of Figure 4, and Figure S3 shows the analogous correlation Cor$\left[T_g^f, T_g^m\right]$ of trait values.

Figure 5e then shows the variance of the admixture fraction under the three mating models, for the same simulations from Figure 4e with the base case parameters. The Var[H_A] curves in Figure 5e under the three mating models follow Equation (B6), using the time-varying $r_{H_A,g}$ in Figure S2.

Among the three mating models, the variance of admixture Var[H_A] decreases fastest for random mating. After one generation, Var[H_A] falls in half (0.125), and it continues to decrease monotonically by half. After 40 generations, it is 2.118×10^{-13}. The distribution of the admixture fraction concentrates around $H_A = \frac{1}{2}$ at each generation. Because the offspring admixture fraction is the mean of those of its parents, without additional influx from the source populations after the founding event, random mating rapidly drives the admixture fraction away from extreme values (0 or 1) toward the mean value of the parental pool ($\frac{1}{2}$).

Under assortative mating by admixture, pairs with similar admixture fractions have higher mating probabilities than under random mating. The fraction of offspring that are admixed is smaller than under random mating, and the admixture fraction distribution remains close to the extreme values (0 or 1) for longer (Figure S1). Hence, Var[H_A] is larger under assortative mating by admixture fraction (Figure 5e). Without influx from the source populations, Var[H_A] eventually decreases to zero, but the decrease is slower than for random mating. Var[H_A] = 0.184 after one generation of assortative mating by admixture fraction, and Var[H_A] = 1.118×10^{-8} after 40 generations. This result can also be seen in Equation (B6). From generation g to $g + 1$, Var[H_A]$_g$ decreases by a factor of $(1 + r_{H_A,g})/2$. With positive assortative mating by admixture ($r_{H_A,g} > 0$), Var[H_A] in

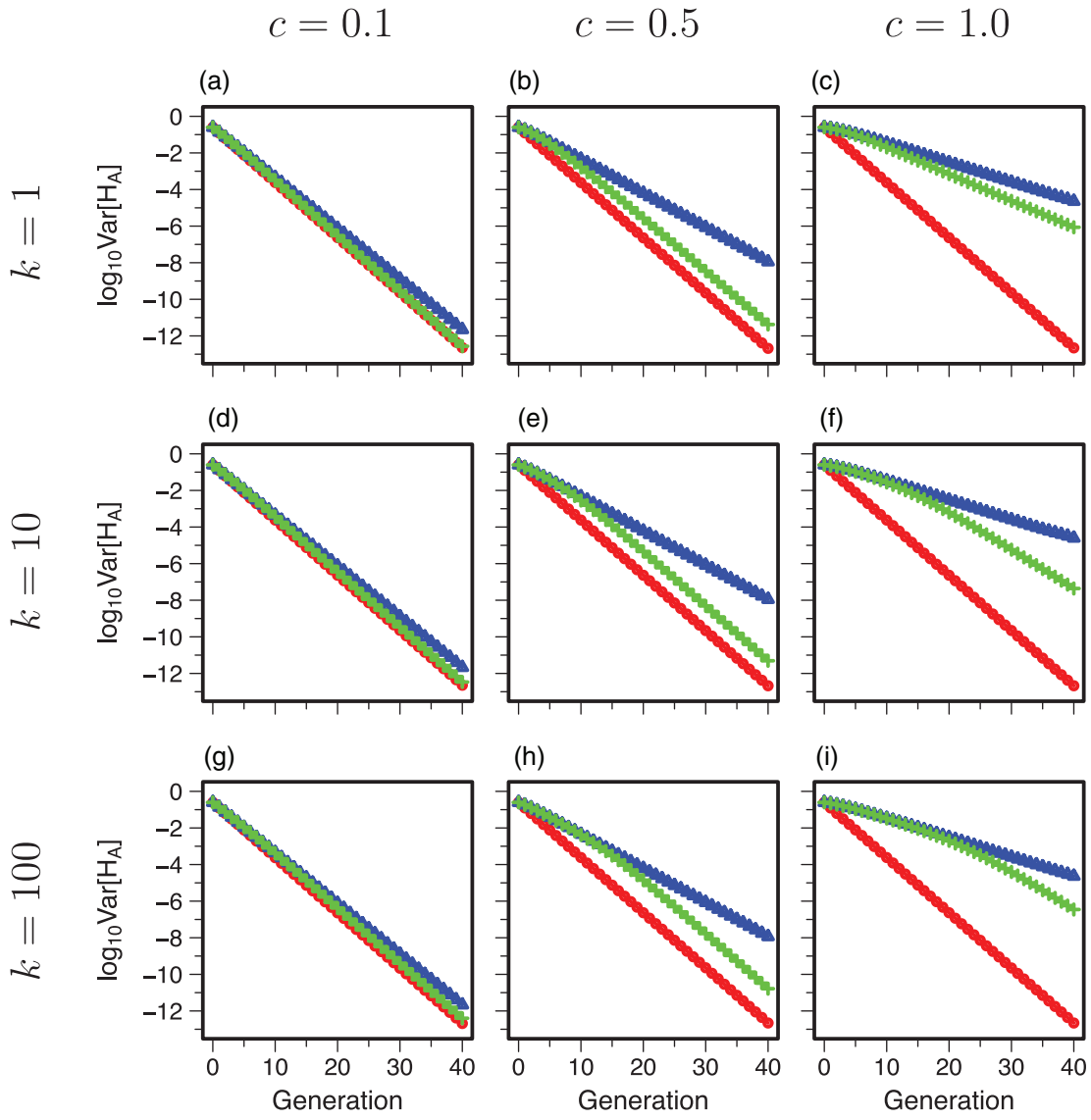

FIGURE 5 Variance of admixture fraction (Var[H_A]) as a function of time. The simulations shown are the same ones from Figure 4. (a) k = 1, c = 0.1. (b) k = 1, c = 0.5. (c) k = 1, c = 1.0. (d) k = 10, c = 0.1. (e) k = 10, c = 0.5. (f) k = 10, c = 1.0. (g) k = 100, c = 0.1. (h) k = 100, c = 0.5. (i) k = 100, c = 1.0. Colors and symbols follow Figure 4. The y-axis is plotted on a logarithmic scale

the next generation is increased compared with the case of random mating ($r_{H_A,g} = 0$).

Under assortative mating by phenotype, Var[H_A] = 0.183 after one generation of assortative mating by phenotype, and Var[H_A] = 4.910×10^{-12} after 40 generations. For the first few generations ($g < 5$), because Cor[H_A, T] is high, the correlation between the admixture fractions in mating pairs, and thus Var[H_A], is similar under the two assortative mating models, as shown in the comparison of the green and blue curves in Figure S2 and Figure 5. However, because the admixture fraction and phenotype decouple over time, mating assortatively by phenotype results in lower $r_{H_A,g}$ than mating assortatively by admixture fraction. In accord with Equation (B6), assortative mating by phenotype produces faster decay in Var[H_A] with its lower $r_{H_A,g}$ at each generation than assortative mating by admixture fraction.

For the variance of the trait value Var[T], by definition of T, all individuals in S_1 and S_2 have trait values 20 and 0, respectively. Therefore, in the founding parental pool, H_0^{par}, noting that S_1 and S_2 each have 1000 individuals, Var[T] has the same constant value $\frac{2,000}{1,999} \cdot 10^2 = 100.050$ irrespective of the mating model. Figure 6e displays Var[T], which decreases most rapidly under random mating, falling by half (50.025) in one generation, and approaching a steady-state value ≈4.957 after 13 generations. Opposite to what was seen for Var[H_A], however, assortative mating by trait retains Var[T] higher for longer than assortative mating by admixture fraction. Similar to the case with Var[H_A] under assortative mating by admixture fraction, assortative mating by trait keeps the trait values near extremes for longer than the other two mating models.

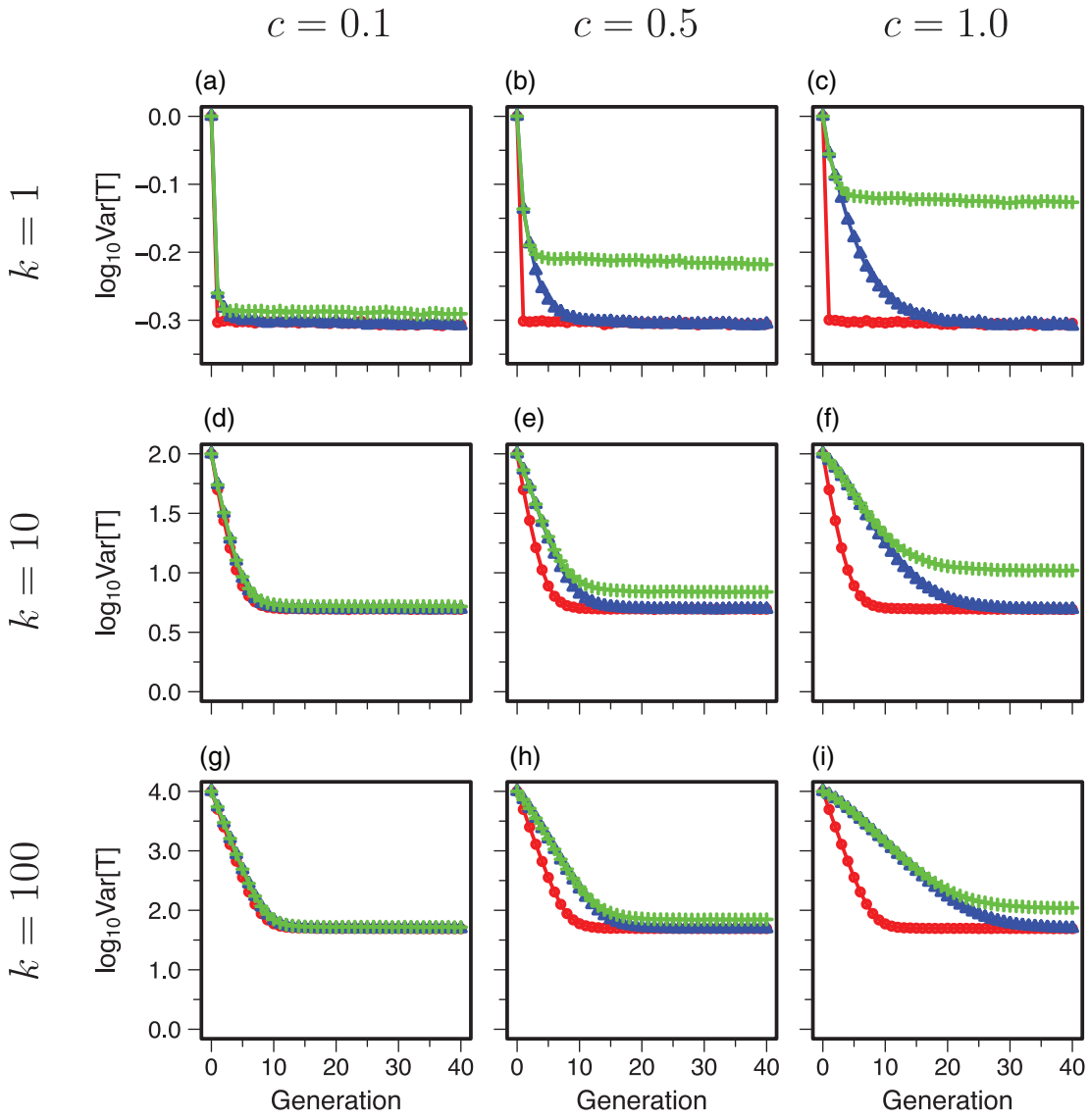

FIGURE 6 Variance of the phenotype (Var[T]) as a function of time. The simulations shown are the same ones from Figure 4. (a) $k = 1$, $c = 0.1$. (b) $k = 1$, $c = 0.5$. (c) $k = 1$, $c = 1.0$. (d) $k = 10$, $c = 0.1$. (e) $k = 10$, $c = 0.5$. (f) $k = 10$, $c = 1.0$. (g) $k = 100$, $c = 0.1$. (h) $k = 100$, $c = 0.5$. (i) $k = 100$, $c = 1.0$. Colors and symbols follow Figure 4. The y-axis is plotted on a logarithmic scale

Having examined the behavior of Cor[H_A, T], Var[H_A], and Var[T] in the base case, we now explore the effect on these quantities of the assortative mating strength c and the number of trait loci k.

4.2 | Assortative mating strength (c)

4.2.1 | Cor[H_A, T]

Each row of Figure 4 illustrates the influence of the assortative mating strength c on Cor[H_A, T] with a fixed number of trait loci k, and each column depicts the effect of the number of loci k on Cor[H_A, T] with fixed assortative mating strength c. All parameters other than c and k are held constant at the base case values.

With different assortative mating strengths and numbers of trait loci, c = 0.1, 0.5, 1.0 and k = 1, 10, 100, the qualitative behavior of Cor[H_A, T] remains the same as in the base case. As before, we observe decay in Cor[H_A, T] under all three mating models, with random mating decoupling ancestry and trait values the most rapidly. Cor[H_A, T] remains higher for longer under assortative mating by phenotype than under assortative mating by admixture fraction. The rate of decay and the degree to which the patterns differ across the three mating models depend on the assortative mating strength and the number of loci.

If assortative mating is weak (c = 0.1 in Figure 4a,d,g), then Cor[H_A, T] under assortative mating by admixture and by phenotype closely follow that under random mating. This pattern is seen irrespective of the number of loci. Note that in the limit of c = 0, the assortative mating and random mating models are identical because the mating function in Equation (A2) becomes a constant, the same for all three mating models.

Comparing panels within rows of Figure 4, results from random mating are identical, as c does not affect the random mating model. Under both assortative mating models, however, Cor[H_A, T] increases with c. In an extreme case of complete assortment ($c \to \infty$), the correlation would stay constant at 1: only identical individuals mate, so that an initial correlation between admixture and phenotype persists unchanged.

The difference among the three models increases with the assortative mating strength given a fixed number of trait loci. The difference is the greatest if k = 1 and c = 1.0 (Figure 4c). Even after 40 generations, assortative mating by trait retains a high correlation at 0.788, whereas the corresponding values under random mating and assortative mating by admixture are 0.006 and 0.010, respectively.

4.2.2 | Var[H_A] and Var[T]

The plots of Var[H_A] in Figure 5 and Var[T] in Figure 6 consider the same simulations that appear for Cor[H_A, T] in Figure 4. As is seen in classical work (Crow & Felsenstein, 1968; Crow & Kimura, 1970; Felsenstein, 1981), compared with random mating, assortative mating increases the variance of the property on which assortment takes place. Thus, the variance of the admixture fraction is increased to a greater extent under mating by admixture fraction than under mating by phenotype. Similarly, the variance of the phenotype is increased to a greater extent under mating by phenotype than under mating by admixture fraction. Both types of assortative mating increase both Var[H_A] and Var[T] compared with random mating.

The variance-increasing effect of assortative mating is visible across panels within each row. For low assortative mating strength (c = 0.1), panels a, d, and g in Figures 5 and 6 depict minimal differences in Var[H_A] and Var[T] between mating models. As c increases, for a given number of loci, Figures 5 and 6 display increased differences between random and assortative mating, with maximal separation at the largest c simulated, c = 1 (panels c, f, i). The random mating model is unaffected by c, as seen with Cor[H_A, T].

4.3 | Number of trait loci (k)

4.3.1 | Cor[H_A, T]

A comparison of panels within columns of Figure 4 shows that under random mating, with more loci associated with the phenotype, the ancestry–phenotype correlation is higher and stays high for longer: it takes longer for H_A and T to become decoupled. Under random mating, the correlation falls below 0.5 at g = 3 if k = 1, g = 6 if k = 10, and g = 10 if k = 100, independent of the assortative mating strength.

As the number of loci increases, results from the models with assortative mating by phenotype and by admixture become similar. If (c, k) = (1,100) (Figure 4i), then it takes 24 generations for Cor[H_A, T] values under the two models to differ by more than 0.1. Corresponding times for (c, k) = (1, 1) (Figure 4c) and (c, k) = (1, 10) (Figure 4f) are g = 6 and g = 15, respectively. Recall that the admixture fraction represents the probability that a random allele at a random autosomal genetic locus originates from the source population S_1, assuming infinitely many loci. In the $k \to \infty$ limit, with the whole genome contributing to the trait, the assortative mating models by admixture and by phenotype would behave very similarly.

4.3.2 | Var[H_A] and Var[T]

Comparing panels within columns in Figure 5, for a given assortative mating strength, Var[H_A] under assortative mating by admixture follows the same curve irrespective of the number of loci. Because the mating probability is independent of trait values if mating assortatively by admixture, k has no effect.

As in the base case, both assortative mating models have higher Var[H_A] and Var[T] than random mating. Assortative mating by admixture fraction has greater Var[H_A] than assortative mating by trait at each generation; for Var[T], assortative mating by trait has greater values. As was seen with Cor[H_A, T], for Var[H_A] and Var[T], the difference between random mating and both assortative mating models

increases with k, and the difference between the two assortative mating models diminishes as k increases.

5 | DISCUSSION

We have devised a mechanistic model of a quantitative phenotype in an admixed population, studying it in relation to loci affecting its trait value and to mate choice. The admixture level and the phenotype are examined using a discrete-time recursion that describes evolution in the admixed population. We have considered the correlation between ancestry and phenotype in the admixed population under three mating models: random mating, assortative mating by admixture fraction, and assortative mating by phenotype.

5.1 | Behavior of the model

Initially, ancestry and phenotype are coupled, as the source populations differ in phenotype. Random mating then decouples the correlation between ancestry and trait faster than is seen in both assortative mating models (Figure 4), and assortative mating by phenotype maintains the correlation to a greater extent than does assortative mating by admixture (Figure 4). Compared with random mating, in a similar manner to classic assortative mating models (Crow & Felsenstein, 1968; Crow & Kimura, 1970; Felsenstein, 1981), the assortative mating increases the population variance of the property on which the assortment is based (Figures 5 and 6). In fact, our Equation (1) multiplies the variance of admixture in a model without assortment (Verdu & Rosenberg, 2011) by a factor that increases with positive assortative mating.

Increasing the strength of assortative mating magnifies the difference among models in the speed at which the correlation declines (Figure 4). As the number of loci underlying the trait increases, the assortative mating models have increasingly similar trajectories. Assortative mating by admixture fraction affects all loci, whereas assortative mating by trait affects only trait loci and their genomic neighbors. Hence, with more trait loci, assortative mating by trait increasingly mimics assortative mating by admixture (Figure 4).

Differences between assortative and random mating are apparent under the idealized setting in which distinct alleles are fixed in the two source populations, and in which all alleles that are more frequent in the source population S_1 than S_2 are the trait-increasing alleles (Figure 4). In scenarios that relax these idealized assumptions, when the source populations have allele frequency differences that do not amount to fixed differences (Figures S4 and S5), generally similar qualitative patterns are observed.

5.2 | Applications and extensions

The focus of our simulations has been on understanding demographic phenomena, but the model is relevant to efforts to investigate determinants of disease traits in admixed human populations. For example, in admixture-mapping studies and studies of health disparities involving admixed populations, correlations of phenotypes and admixture levels are often computed (Gravlee et al., 2009; Non et al., 2012; Peralta et al., 2006; Tang et al., 2006). The mechanistic model can potentially provide insights into the way in which these correlations change over time in scenarios in which specific trait architectures are of interest.

In our assortative mating models, in each generation, we standardized admixture level $H_{A,g}$ and phenotype T_g in the mating function (Equation (A2)) to account for different scales in admixture fractions and phenotypes. With this choice, as the variance of admixture or phenotype decays, individuals can recognize progressively finer differences in admixture fraction or phenotype during mate choice. To relax this strong assumption about mate recognition, we have also examined a mating function in which the relative preference remains constant in time. Given assortative mating strength c, this choice reduces the effect of assortative mating compared with a time-varying scaling factor; however, qualitative patterns are similar (Figure S6).

We have examined a model with a single admixture event at the founding of the admixed population. Under this idealized model, even if the founding admixed population starts with a perfect correlation between admixture fraction and trait value, the correlation decreases over time and eventually approaches zero in the absence of further influx from source populations. In principle, the framework can account for continuous influx; if ancestry–trait correlation exists in source populations, then such influx would be expected to slow the decoupling between admixture and phenotype in the admixed populations under all three mating models, qualitatively maintaining their relative order in the rate of decoupling.

The model is potentially valuable beyond the human admixture context. The motivating scenario also applies in naturally occurring admixture when a single visible trait (e.g., coat color, flower color) is regarded as a marker for ancestry in an admixed group (e.g., Alberts & Altmann, 2001). Further, the analysis might be relevant to hybrid zones, where decoupling of traits from ancestry or each other is sometimes observed in populations of heterogeneous ancestry (e.g., den Hartog et al., 2008; Fuzessy et al., 2014). Insights into homogenization of ancestry and phenotypes in admixed populations can also be useful for hybrid speciation, in which sustained positive assortative mating in the admixed population can lead to its reproductive isolation (Abbott et al., 2013; Mallet, 2005). For non-human cases, phenomena of interest for model extensions include selective regimes that differ in the admixed and source populations, sex-biased mate choice, and assortative mating on socially learned behavior (e.g., Verzijden et al., 2012; Westerman et al., 2014).

5.3 | Limitations

Our framework can accommodate various genetic architectures and admixture assumptions, including disassortative mating; however, it has numerous limitations. First, the model does not include sex bias

during admixture, a common phenomenon in human admixture processes (Adhikari et al., 2017; Goldberg & Rosenberg, 2015; Micheletti et al., 2020; Wilkins, 2006). A recent genetic model does allow sex bias, but with no phenotype and with the assortative mating occurring by population membership rather than by admixture level itself (Goldberg et al., 2020).

We have also modeled individual admixture as the mean of parental admixture levels. Stochasticity during genetic transmission is not considered, nor is the finiteness of chromosomes; our approach amounts to assuming that the genome contains infinitely many independent segments. Thus, "genetic" admixture in our model measures idealized genealogical ancestry, assuming an equal mixture of ancestors from a specified number of generations in the past. This choice is reasonable in the early stages of an admixture process (Gravel, 2012), after which it will be informative to model the distinction between genealogical and genetic ancestry.

Our simulations give each mating pair equally many offspring, so that variance in reproductive success is not considered. More generally, the expected number of offspring is independent of genotype and phenotype, so that no natural selection occurs. Mating pairs are drawn with replacement, permitting individuals to be included in multiple mating pairs; thus, we allow a form of nonmonogamous mating. Although any particular mating pair is unlikely to be a pair of close relatives, such pairs, even sibs, are permissible.

We have examined only a univariate trait, and our trait model does not incorporate dominance, spatial positioning of trait loci along a genome, variable effect sizes across trait loci, epistasis, environmental effects and consequent varying heritability in the phenotype, or genotype-by-environment interaction. As assortative mating in humans often operates on sociocultural traits (Kalmijn, 1998; Rosenfeld, 2008; Schwartz, 2013; Watson et al., 2004), the latter pair of limitations might be particularly important for human data.

6 | CONCLUSIONS: CONSEQUENCES OF THE DECOUPLING OF ANCESTRY AND PHENOTYPE

In the children's story "The Sneetches," two sympatric populations of the titular fictional species differ by a polymorphic physical marking that appears in members of one but not the other population. The Star-Belly Sneetches possess a star-shaped abdominal marking; the Plain-Belly Sneetches do not (Geisel, 1961). Through a multistage process in which the mark is repeatedly added and removed from individual Sneetches, phenotypes of individuals are shuffled in relation to initial population membership, "Until neither the Plain nor the Star-Bellies knew/Whether this one was that one… or that one was this one/Or which one was what one… or what one was who."

Our study was motivated partly by a hypothesis of Parra et al. (2003) claiming that in humans, assortative mating by *color* in Brazil could eventually decouple *color* from ancestry, so that subpopulations with distinct *color* could eventually possess similar African ancestry. We have seen not only that assortative mating by a quantitative trait that differs between source populations can decouple the phenotype from the ancestry, but also that random mating can decouple the phenotype from ancestry even faster. In an admixed population with assortative mating that is affected by a visible genetically influenced phenotype (such as *color* in Parra et al. (2003)), mating by many other genetically influenced phenotypes is random or less strongly assortative. Thus, in an admixed population, we might expect that among traits to which genotypes contribute, those with little influence on mating behavior will decouple from ancestry most rapidly. Traits on which assortment does occur, such as *color* in the Brazil scenario, will be the slowest to decouple from ancestry—but under our model, they eventually will do so. Thus, in an admixed population, phenotypes that once reflected ancestry in the source populations might no longer be predictive of ancestry after sufficient time has passed.

The decoupling from the ancestry of genetically influenced phenotypes that initially differed between source populations is informative in relation to systems of social categorization that involve visible genetically influenced traits. Consider a setting in which differences among individuals in visible traits such as skin pigmentation contribute to differences in social categorizations, and in which social justifications for the categorization system rely on the assumption that the visible traits correlate with ancestry or with genetically influenced traits that are *not* visible. In an admixed population, after enough time, the simultaneous decoupling from the ancestry of genetically influenced phenotypes with an initial difference between source populations generates a scenario in which visible traits salient in social categorizations are decoupled from all other genetically influenced traits except those with a genetic basis in the same loci.

Among the many goals of mathematical modeling in population biology (Rosenberg, 2020; Servedio et al., 2014), two are (a) to characterize the determinants of a biological phenomenon, as we have done in discerning the effects of genetic and evolutionary parameters on the decoupling between admixture and phenotype; and (b) to mathematize and test the validity of predictions of a verbal model, as in our formalization of the decoupling model of Parra et al. (2003). Like a children's story, a mathematical model with these objectives obtains insights about the real world by disregarding much of its complexity, exploring—with some risk of oversimplification—key features of the real world in the context of the world that it creates.

In "The Sneetches," the abdominal marking is used by the characters as a signifier of group membership. It confers to individuals no other qualities, but a perception by the characters of its correlation with other traits is consequential for the way that they treat each other. At the story's end, after the phenotype has been reshuffled with respect to population of origin, the species remains phenotypically polymorphic, but the phenotype of an individual has become uninformative about "ancestry." The decoupling of "ancestry" and phenotype has led to a loss of social meaning for the phenotype, which no longer serves as a signifier of group membership. The characters no longer perceive a correlation between the marking and other traits, recognizing the lack of information that the marking contains about traits other than itself. In admixed populations, the

mathematical model suggests that the visible traits used in social categorizations may come to possess little or no ancestry information. These traits are then rendered informative about few genetically influenced traits other than themselves—so that the model provides a mechanistic explanation for the expression that such traits are "only skin deep."

ACKNOWLEDGMENTS

We thank S. Gravel and an anonymous reviewer for their comments. We acknowledge support from National Institutes of Health grants R01 HG005855, F32 GM130050, and R35 GM133481.

AUTHOR CONTRIBUTIONS

Jaehee Kim: Conceptualization; formal analysis; methodology; writing-original draft; writing-review & editing. **Michael Edge:** Conceptualization; methodology; writing-review & editing. **Amy Goldberg:** Conceptualization; methodology; writing-review & editing. **Noah Rosenberg:** Conceptualization; methodology; supervision; writing-review & editing.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

No data sets were generated or analyzed for this study. Code associated with the manuscript is available at https://github.com/jk2236/AssortativeMating.

ORCID

Jaehee Kim https://orcid.org/0000-0002-5210-2004
Noah A. Rosenberg https://orcid.org/0000-0002-1829-8664

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*, 229–246.

Adhikari, K., Chacón-Duque, J. C., Mendoza-Revilla, J., Fuentes-Guajardo, M., & Ruiz-Linares, A. (2017). The genetic diversity of the Americas. *Annual Review of Genomics and Human Genetics*, *18*, 277–296.

Alberts, S. C., & Altmann, J. (2001). Immigration and hybridization patterns of yellow and anubis baboons in and around Amboseli, Kenya. *American Journal of Primatology*, *53*, 139–154.

Algee-Hewitt, B. F. B. (2016). Population inference from contemporary American craniometrics. *American Journal of Physical Anthropology*, *160*, 604–624.

Beleza, S., Johnson, N. A., Candille, S. I., Absher, D. M., Coram, M. A., Lopes, J., Campos, J., Araújo, I. I., Anderson, T. M., Vilhjálmsson, B. J., Nordborg, M., Correia e Silva, A., Shriver, M. D., Rocha, J., Barsh, G. S., & Tang, H. (2013). Genetic architecture of skin and eye color in an African-European admixed population. *PLoS Genetics*, *9*, e1003372.

Crow, J. F., & Felsenstein, J. (1968). The effect of assortative mating on the genetic composition of a population. *Eugenics Quarterly*, *15*, 85–97.

Crow, J. F., & Kimura, M. (1970). *An introduction to population genetics theory*. New York: Harper & Row.

den Hartog, P. M., Slabbekoorn, H., & ten Cate, C. (2008). Male territorial vocalizations and responses are decoupled in an avian hybrid zone. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *363*, 2879–2889.

Devillard, S., Jombart, T., Léger, F., Pontier, D., Say, L., & Ruette, S. (2014). How reliable are morphological and anatomical characters to distinguish European wildcats, domestic cats and their hybrids in France? *Journal of Zoological Systematics and Evolutionary Research*, *52*, 154–162.

Durso, D. F., Bydlowski, S. P., Hutz, M. H., Suarez-Kurtz, G., Magalhães, T. R., & Pena, S. D. J. (2014). Association of genetic variants with self-assessed color categories in Brazilians. *PLoS One*, *9*, e0083926.

Edge, M. D., & Rosenberg, N. A. (2015a). A general model of the relationship between the apportionment of human genetic diversity and the apportionment of human phenotypic diversity. *Human Biology*, *87*, 313–337.

Edge, M. D., & Rosenberg, N. A. (2015b). Implications of the apportionment of human genetic diversity for the apportionment of human phenotypic diversity. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *52*, 32–45.

Ewens, W. J., & Spielman, R. S. (1995). The transmission/disequilibrium test: History, subdivision, and admixture. *American Journal of Human Genetics*, *57*, 455–464.

Felsenstein, J. (1981). Continuous-genotype models and assortative mating. *Theoretical Population Biology*, *19*, 341–357.

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

Fuzessy, L. F., de Oliveira Silva, I., Malukiewicz, J., Silva, F. F. R., do Carmo Pônzio, M., Boere, V., & Ackermann, R. R. (2014). Morphological variation in wild marmosets (*Callithrix penicillata* and *C. geoffroyi*) and their hybrids. *Evolutionary Biology*, *41*, 480–493.

Geisel, T. S. (1961). *The Sneetches and other stories*. New York: Random House.

Goldberg, A., Rastogi, A., & Rosenberg, N. A. (2020). Assortative mating by population of origin in a mechanistic model of admixture. *Theoretical Population Biology*, *134*, 129–146.

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

Goldberg, A., Verdu, P., & Rosenberg, N. A. (2014). Autosomal admixture levels are informative about sex bias in admixed populations. *Genetics*, *198*, 1209–1229.

Gravel, S. (2012). Population genetics models of local ancestry. *Genetics*, *191*, 607–619.

Gravlee, C. C., Non, A. L., & Mulligan, C. J. (2009). Genetic ancestry, social classification, and racial inequalities in blood pressure in southeastern Puerto Rico. *PLoS One*, *4*, e0006821.

Guo, W., Fung, W. K., Shi, N., & Guo, J. (2005). On the formula for admixture linkage disequilibrium. *Human Heredity*, *60*, 177–180.

Jennings, H. S. (1916). The numerical results of diverse systems of breeding. *Genetics*, *1*, 53–89.

Kalmijn, M. (1998). Intermarriage and homogamy: Causes, patterns, trends. *Annual Review of Sociology*, *24*, 395–421.

Leite, T. K. M., Fonseca, R. M. C., de França, N. M., Parra, E. J., & Pereira, R. W. (2011). Genomic ancestry, self-reported "color" and quantitative measures of skin pigmentation in Brazilian admixed siblings. *PLoS One*, *6*, e0027162.

Long, J. C. (1991). The genetic structure of admixed populations. *Genetics*, *127*, 417–428.

Magalhães da Silva, T., Sandhya Rani, M. R., Nunes de Oliveira Costa, G., Figueiredo, M. A., Melo, P. S., Nascimento, J. F., Molyneaux, N. D.,

Barreto, M. L., Reis, M. G., Teixeira, M. G., & Blanton, R. E. (2014). The correlation between ancestry and color in two cities of Northeast Brazil with contrasting ethnic compositions. *European Journal of Human Genetics*, *23*, 984–989.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*, 229–237.

Micheletti, S. J., Bryc, K., Ancona Esselmann, S. G., Freyman, W. A., Moreno, M. E., Poznik, G. D., Shastri, A. J., Beleza, S., Mountain, J. L., Agee, M., Aslibekyan, S., Auton, A., Bell, R., Clark, S., Das, S., Elson, S., Fletez-Brant, K., Fontanillas, P., Gandhi, P., … Zare, A. (2020). Genetic consequences of the transatlantic slave trade in the Americas. *American Journal of Human Genetics*, *107*, 265–277.

Non, A. L., Gravlee, C. C., & Mulligan, C. J. (2012). Education, genetic ancestry, and blood pressure in African Americans and whites. *American Journal of Public Health*, *102*(8), 1559–1565.

Parra, E. J., Kittles, R. A., & Shriver, M. D. (2004). Implications of correlations between skin color and genetic ancestry for biomedical research. *Nature Genetics*, *36*, S54–S60.

Parra, E. J., Marcini, A., Akey, J., Martinson, J., Batzer, M. A., Cooper, R., Forrester, T., Allison, D. B., Deka, R., Ferrell, R. E., & Shriver, M. D. (1998). Estimating African American admixture proportions by use of population-specific alleles. *American Journal of Human Genetics*, *63*, 1839–1851.

Parra, F. C., Amado, R. C., Lambertucci, J. R., Rocha, J., Antunes, C. M., & Pena, S. D. J. (2003). Color and genomic ancestry in Brazilians. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 177–182.

Peralta, C. A., Ziv, E., Katz, R., Reiner, A., Burchard, E. G., Fried, L., Kwok, P. Y., Psaty, B., & Shlipak, M. (2006). African ancestry, socioeconomic status, and kidney function in elderly African Americans: A genetic admixture analysis. *Journal of the American Society of Nephrology*, *17*, 3491–3496.

Pimenta, J. R., Zuccherato, L. W., Debes, A. A., Maselli, L., Soares, R. P., Moura-Neto, R. S., Rocha, J., Bydlowski, S. P., & Pena, S. D. J. (2006). Color and genomic ancestry in Brazilians: A study with forensic microsatellites. *Human Heredity*, *62*, 190–195.

Risch, N., Choudhry, S., Via, M., Basu, A., Sebro, R., Eng, C., Beckman, K., Thyne, S., Chapela, R., Rodriguez-Santana, J. R., Rodriguez-Cintron, W., Avila, P. C., Ziv, E., & Gonzalez Burchard, E. (2009). Ancestry-related assortative mating in Latino populations. *Genome Biology*, *10*, R132.

Rosenberg, N. A. (2020). Fifty years of theoretical population biology. *Theoretical Population Biology*, *133*, 1–12.

Rosenfeld, M. J. (2008). Racial, educational and religious endogamy in the United States: A comparative historical perspective. *Social Forces*, *87*, 1–31.

Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., Fuentes, M., Pizarro, M., Everardo, P., de Avila, F., Gómez-Valdés, J., León-Mimila, P., Hunemeier, T., Ramallo, V., Silva de Cerqueira, C. C., Burley, M. W., Konca, E., de Oliveira, M. Z., Veronez, M. R., … Gonzalez-José, R. (2014). Admixture in Latin America: Geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genetics*, *10*, e1004572.

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

Servedio, M. R., Brandvain, Y., Dhole, S., Fitzpatrick, C. L., Goldberg, E. E., Stern, C. A., Van Cleve, J., & Yeh, J. (2014). Not just a theory—The utility of mathematical models in evolutionary biology. *PLoS Biology*, *12*, e1002017.

Shriver, M. D., Smith, M. W., Jin, L., Marcini, A., Akey, J. M., Deka, R., & Ferrell, R. E. (1997). Ethnic-affiliation estimation by use of population-specific DNA markers. *American Journal of Human Genetics*, *60*, 957–964.

Tang, H., Jorgenson, E., Gadde, M., Kardia, S. L. R., Rao, D. C., Zhu, X., Schork, N. J., Hanis, C. L., & Risch, N. (2006). Racial admixture and its impact on BMI and blood pressure in African and Mexican Americans. *Human Genetics*, *119*, 624–633.

Trigo, T. C., Tirelli, F. P., de Freitas, T. R. O., & Eizirik, E. (2014). Comparative assessment of genetic and morphological variation at an extensive hybrid zone between two wild cats in southern Brazil. *PLoS One*, *9*, e0108469.

Verdu, P., & Rosenberg, N. A. (2011). A general mechanistic model for admixture histories of hybrid populations. *Genetics*, *189*, 1413–1426.

Verzijden, M. N., ten Cate, C., Servedio, M. R., Kozak, G. M., Boughman, J. W., & Svensson, E. I. (2012). The impact of learning on sexual selection and speciation. *Trends in Ecology & Evolution*, *27*, 511–519.

Watson, D., Klohnen, E. C., Casillas, A., Nus Simms, E., Haig, J., & Berry, D. S. (2004). Match makers and deal breakers: Analyses of assortative mating in newlywed couples. *Journal of Personality*, *72*, 1029–1068.

Westerman, E. L., Chirathivat, N., Schyling, E., & Monteiro, A. (2014). Mate preference for a phenotypically plastic trait is learned, and may facilitate preference-phenotype matching. *Evolution*, *68*, 1661–1670.

Wilkins, J. F. (2006). Unraveling male and female histories from human genetic data. *Current Opinion in Genetics and Development*, *16*, 611–617.

Wright, S. (1921). Systems of mating. III. Assortative mating based on somatic resemblance. *Genetics*, *6*, 144–161.

Zaitlen, N., Huntsman, S., Hu, D., Spear, M., Eng, C., Oh, S. S., White, M. J., Mak, A., Davis, A., Meade, K., Brigino-Buenaventura, E., LeNoir, M. A., Bibbins-Domingo, K., Burchard, E. G., & Halperin, E. (2017). The effects of migration and assortative mating on admixture linkage disequilibrium. *Genetics*, *205*, 375–383.

Zou, J. Y., Park, D. S., Burchard, E. G., Torgerson, D. G., Pino-Yanes, M., Song, Y. S., Sankararaman, S., Halperin, E., & Zaitlen, N. (2015). Genetic and socioeconomic study of mate choice in Latinos reveals novel assortment patterns. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 13621–13626.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Kim J, Edge MD, Goldberg A, Rosenberg NA. Skin deep: The decoupling of genetic admixture levels from phenotypes that differed between source populations. *Am J Phys Anthropol*. 2021;175:406–421. https://doi.org/10.1002/ajpa.24261

APPENDIX A.: MATING MODEL DETAILS

This appendix describes the construction of the mating matrix M under random mating, assortative mating by admixture, and assortative mating by phenotype, where M is written

$$\begin{array}{c} \\ \left(H_{A,g}^{(1),f},T_g^{(1),f}\right) \\ \left(H_{A,g}^{(2),f},T_g^{(2),f}\right) \\ \vdots \\ \left(H_{A,g}^{(N-1),f},T_g^{(N-1),f}\right) \\ \left(H_{A,g}^{(N),f},T_g^{(N),f}\right) \end{array} \begin{array}{c} \begin{array}{ccccc} \left(H_{A,g}^{(1),m},T_g^{(1),m}\right) & \left(H_{A,g}^{(2),m},T_g^{(2),m}\right) & \cdots & \left(H_{A,g}^{(N-1),m},T_g^{(N-1),m}\right) & \left(H_{A,g}^{(N),m},T_g^{(N),m}\right) \end{array} \\ \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1(N-1)} & m_{1N} \\ m_{21} & m_{22} & \cdots & m_{2(N-1)} & m_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ m_{(N-1)1} & m_{(N-1)2} & \cdots & m_{(N-1)(N-1)} & m_{(N-1)N} \\ m_{N1} & m_{N2} & \cdots & m_{N(N-1)} & m_{NN} \end{pmatrix} \end{array}. \tag{A1}$$

With no selection or sex bias, each individual in the population has the same expected number of offspring irrespective of ancestry or phenotype. We assume that the expected number of offspring of an individual is proportional to the expected number of matings of the individual, the sum of matrix entries across all mates available for an individual. Thus, the equal-offspring requirement translates into an assumption of equal row sums for females that are in turn equal to equal column sums for males in M. Note that this equal-offspring assumption independent of ancestry and phenotype accords with a standard property of assortative mating models that assortative mating on its own does not alter allele frequencies over time (Crow & Kimura, 1970; Fisher, 1918; Goldberg et al., 2020; Jennings, 1916; Wright, 1921; Zaitlen et al., 2017).

A.1. | Three mating models

We write the mating probability in Equation (A1) as $m_{ij} = \alpha_{ij}\psi\left(H_{A,g}^{(i),f},T_g^{(i),f},H_{A,g}^{(j),m},T_g^{(j),m}\right)$ for female i and male j, where function ψ quantifies the dependence of m_{ij} on the ancestry and trait values of pair (i, j); α_{ij} is a normalization constant that is specific to a pair (i, j) and that enforces constant row and column sums. Because the sum of all entries in mating matrix M is 1, the row and column sums each equal $1/N$.

In random mating, the mating probability is independent of individual ancestry and trait values, so that $\psi\left(H_{A,g}^{(i),f},T_g^{(i),f},H_{A,g}^{(j),m},T_g^{(j),m}\right)$ is constant across all i and j, and m_{ij} has the same value for all mating pairs. Therefore, for all pairs (i, j), each taken from {1, 2, …, N}, $m_{ij} = \alpha_{ij} = \alpha$ for a constant $\alpha = 1/N^2$.

In assortative mating by admixture, the mating probability depends only on the ancestries of potential mates and not on the phenotypes: $\psi\left(H_{A,g}^{(i),f},T_g^{(i),f},H_{A,g}^{(j),m},T_g^{(j),m}\right) = \psi\left(H_{A,g}^{(i),f},H_{A,g}^{(j),m}\right)$ and $m_{ij} = \alpha_{ij}\psi\left(H_{A,g}^{(i),f},H_{A,g}^{(j),m}\right)$. For positive assortment, the mating function ψ has higher values if two individuals have similar ancestries and lower values as the ancestries become more different. For example, in complete assortment, ψ is 1 if the two input parameters have the same value and 0 if the values differ. For negative assortment, ψ instead increases with the difference between the ancestries of the members of a mating pair.

In assortative mating by phenotype, the mating probability depends only on the trait values of potential mates and not on the ancestries: $\psi\left(H_{A,g}^{(i),f},T_g^{(i),f},H_{A,g}^{(j),m},T_g^{(j),m}\right) = \psi\left(T_g^{(i),f},T_g^{(j),m}\right)$ and $m_{ij} = \alpha_{ij}\psi\left(T_g^{(i),f},T_g^{(j),m}\right)$. The qualitative requirements for the function ψ are the same as with assortative mating by admixture, but with the trait values of the mating pair as arguments instead of the ancestries.

We adopt the following form for the mating function:

$$\psi\left(X_g^{(i),f},X_g^{(j),m}\right) = e^{\frac{-c\left|X_g^{(i),f}-X_g^{(j),m}\right|}{\sigma_{X_g}}}. \tag{A2}$$

The finite constant c quantifies the assortative mating strength. Given two values $\left(X_g^{(i),f},X_g^{(j),m}\right)$, where $X_g = H_{A,g}$ or $X_g = T_g$, increasing c lowers the mating probability, producing stronger positive assortative mating. For $c > 0$, ψ has value 1 if potential mates have the same ancestry (or phenotype), decreasing exponentially with increasing difference between individuals; $c < 0$ indicates negative assortative mating, where pairs with different ancestry (or phenotype) have the highest probability of mating.

At generation g, admixture fraction $H_{A,g}$ takes values in {0, $1/2^g$, $2/2^g$, …, $(2^g - 1)/2^g$, 1} (Section 2.1), and phenotype T_g takes values in {0, 1, …, $2k$} (Section 2.2). To compare mating schemes, we consider variables standardized by dividing X_g ($H_{A,g}$ or T_g) by its standard deviation σ_{X_g} based on its distribution in H_g^{par} at generation g. For the unstandardized variables, because T_g takes a higher value than $H_{A,g}$, the effect of assortative mating by phenotype at the same assortative mating strength c is artificially inflated compared with the effect of assortative mating by admixture fraction.

With mating function ψ, in mating matrix M, the sum across potential mates of the matrix entries for a random individual in the parental pool must be $1/N$. To obtain the normalizing coefficients α_{ij}, we use procedures from numerical optimization, as described in the Supporting Information.

A.2. | Simulating the mating models

To calculate the mating matrix M in our simulations, using the mating function in Equation (A2), we compute an unnormalized $N \times N$ mating matrix $\tilde{M}$, with one matrix entry for each pair containing a female and a male from the parental pool. Diagonal entries of $\tilde{M}$ equal 1; off-diagonal entries (i, j) are $\tilde{M} = e^{-c\Delta_{i,j}}$ for male i and female j, where $\Delta_{i,j} = \left|H_{A,g}^{(i),f} - H_{A,g}^{(j),m}\right| / \sigma_{H_{A,g}}$ for assortative mating by admixture and $\Delta_{i,j} = \left|T_g^{(i),f} - T_g^{(j),m}\right| / \sigma_{T_g}$ for assortative mating by trait. For random mating, $c = 0$ and all entries equal 1. To produce matrix M, $\tilde{M}$ is normalized as described in the Supporting Information.

APPENDIX B.: EVALUATING THE VARIANCE OF THE ADMIXTURE FRACTION

This appendix derives Equation (1). Let the random variable Y indicate the population membership of a random individual in H_g^{par}, the parental pool in generation g for generation $g+1$. Then

$$H_{A,g}^{f,p}, H_{A,g}^{m,p} = \begin{cases} H_{A,g} \text{ with } P(Y = H_g) = h_g \\ 1 \text{ with } P(Y = S_1) = s_{1,g} \\ 0 \text{ with } P(Y = S_2) = s_{2,g}. \end{cases} \tag{B1}$$

For the expectation of admixture in the parental pool, we have

$$E\left[H_{A,g}^{f,p}\right] = E\left[H_{A,g}^{m,p}\right] = E_Y\left[E[H_{A,g}^{f,p}|Y]\right] = \sum_{y \in \{S_1, S_2, H_g\}} P(Y = y) E\left[H_{A,g}^{f,p} | Y = y\right]$$

$$= s_{1,g} + h_g E[H_{A,g}] = s_{1,g} + h_g \mu_g. \tag{B2}$$

As a consequence of Equation (B2), with $\mu_g = E[H_{A,g}]$, we have

$$E\left[\left(H_{A,g}^{f,p}\right)^2\right] = E\left[\left(H_{A,g}^{m,p}\right)^2\right] = s_{1,g} + h_g E\left[H_{A,g}^2\right] = s_{1,g} + h_g \mu_g^2 + h_g \text{Var}[H_{A,g}] \tag{B3}$$

$$\text{Var}\left[H_{A,g}^{f,p}\right] = \text{Var}\left[H_{A,g}^{m,p}\right] = s_{1,g} + h_g \mu_g^2 + h_g \text{Var}[H_{A,g}] - \left(s_{1,g} + h_g \mu_g\right)^2. \tag{B4}$$

An offspring individual has admixture fraction deterministically set to the mean of those of the parents:

$$E[H_{A,g+1}] = E\left[\frac{1}{2}\left(H_{A,g}^{f,p} + H_{A,g}^{m,p}\right)\right] = s_{1,g} + h_g E[H_{A,g}] = s_{1,g} + h_g \mu_g. \tag{B5}$$

We obtain the recursion for the variance of the admixture fraction over a single generation as follows:

$$\begin{aligned} \text{Var}[H_{A,g+1}] &= E\left[H_{A,g+1}^2\right] - (E[H_{A,g+1}])^2 \\ &= \frac{1}{4} E\left[\left(H_{A,g}^{f,p} + H_{A,g}^{m,p}\right)\left(H_{A,g}^{f,p} + H_{A,g}^{m,p}\right)\right] - (E[H_{A,g+1}])^2 \\ &= \frac{1}{2}\left(E\left[\left(H_{A,g}^{f,p}\right)^2\right] + E\left[H_{A,g}^{f,p} H_{A,g}^{m,p}\right]\right) - (E[H_{A,g+1}])^2 \\ &= \frac{1}{2}\left(E\left[\left(H_{A,g}^{f,p}\right)^2\right] + \text{Cor}\left[H_{A,g}^{f,p}, H_{A,g}^{m,p}\right] \text{Var}\left[H_{A,g}^{f,p}\right] + \left(E\left[H_{A,g}^{f,p}\right]\right)^2\right) - (E[H_{A,g+1}])^2. \\ &= \frac{1}{2}(1 + r_{H_A,g})\left[h_g \text{Var}[H_{A,g}] + \mu_g^2 h_g (1 - h_g) - 2\mu_g h_g s_{1,g} + s_{1,g}(1 - s_{1,g})\right], \end{aligned} \tag{B6}$$

where $r_{H_A,g} = \text{Cor}\left[H_{A,g}^{f,p}, H_{A,g}^{m,p}\right]$ denotes the correlation of the admixture fractions in a mating pair. The last step is obtained from Equations (B2)–(B5). The time-varying $r_{H_A,g}$ value in general depends on the parameters of the population model, the quantitative trait model, and the mating model. However, if $s_{1,g} = s_{2,g} = 0$ and $h_g = 1$, then we obtain Equation (1) from Equation (B6).