Assortative mating by population of origin in a mechanistic model of admixture

Amy Goldberg[1,2,*], **Ananya Rastogi**[3], **Noah A Rosenberg**[2]

[1]Dept. of Evolutionary Anthropology, Duke University, Durham, NC, USA

[2]Dept. of Biology, Stanford University, Stanford, CA, USA

[3]Dept. of Systems Immunology & Braunschweig Integrated Centre of Systems Biology, Helmholtz Centre for Infection Research, Braunschweig, Germany

Abstract

Populations whose mating pairs have levels of similarity in phenotypes or genotypes that differ systematically from the level expected under random mating are described as experiencing assortative mating. Excess similarity in mating pairs is termed positive assortative mating, and excess dissimilarity is negative assortative mating. In humans, empirical studies suggest that mating pairs from various admixed populations—whose ancestry derives from two or more source populations—possess correlated ancestry components that indicate the occurrence of positive assortative mating on the basis of ancestry. Generalizing a two-sex mechanistic admixture model, we devise a model of one form of ancestry-assortative mating that occurs through preferential mating based on source population. Under the model, we study the moments of the admixture fraction distribution for different assumptions about mating preferences, including both positive and negative assortative mating by population. We demonstrate that whereas the mean admixture under assortative mating is equivalent to that of a corresponding randomly mating population, the variance of admixture depends on the level and direction of assortative mating. We consider two special cases of assortative mating by population: first, a single admixture event, and second, constant contributions to the admixed population over time. In contrast to standard settings in which positive assortment increases variation within a population, certain assortative mating scenarios allow the variance of admixture to decrease relative to a corresponding randomly mating population: with the three populations we consider, the variance-increasing effect of positive assortative mating within a population might be overwhelmed by a variance-decreasing effect emerging from mating preferences involving other pairs of populations. The effect of assortative mating is smaller on the X chromosome than on the autosomes because inheritance of the X in males depends only on the mother's ancestry, not on the mating pair. Because the variance of admixture is informative about the timing of admixture and possibly about sex-biased admixture contributions, the effects of assortative mating are important to consider in inferring features of

---

*Corresponding author. amy.goldberg@duke.edu.

population history from distributions of admixture values. Our model provides a framework to quantitatively study assortative mating under flexible scenarios of admixture over time.

1 Introduction

Mechanistic models describing the dynamics of admixture among two or more populations have proven informative for understanding the processes that underlie patterns of admixture in admixed populations. Such models have examined a variety of phenomena, including sex-biased admixture (Goldberg *et al.*, 2014; Goldberg and Rosenberg, 2015), the interaction of admixture with migration and population size (Long, 1991; Pool and Nielsen, 2009), the consequences of spatial structure for ancestry patterns (Wang *et al.*, 2011; Sedghifar *et al.*, 2015), hybrid incompatibilities and epistasis (Lindtke and Buerkle, 2015; Schumer *et al.*, 2015; Schumer and Brandvain, 2016; Sedghifar *et al.*, 2016), and the effect of admixture on linkage disequilibrium (Chakraborty and Weiss, 1988; Pfaff *et al.*, 2001; Gravel, 2012; Loh *et al.*, 2013; Liang and Nielsen, 2014; Ni *et al.*, 2016; Zaitlen *et al.*, 2017).

Recently, a general family of admixture models has introduced a framework for examining the effects of an ongoing admixture process on the properties of admixture levels in populations. Verdu and Rosenberg (2011) devised a mechanistic model of admixture allowing for flexible admixture contributions over time, studying the dynamics of the distribution of ancestry within an admixed population. Goldberg *et al.* (2014) and Goldberg and Rosenberg (2015) extended this model to allow for varying contributions from males and females, evaluating the consequences of sex-biased admixture for autosomal and X-chromosomal ancestry. Gravel (2012) derived the distribution of autosomal ancestry and ancestry-tract lengths under a related model allowing for multiple waves of migration. These studies demonstrate the potential of mechanistic admixture models for accommodating complex admixture histories.

One phenomenon of recent interest in the study of admixed populations is assortative mating, in which admixture levels of individuals in a population system that includes an admixed population and its source populations influence the formation of mating pairs. Empirical analyses of mating pairs in admixed populations have documented nonrandom pairings with respect to admixture levels (Risch *et al.*, 2009; Sebro *et al.*, 2010; Zaitlen *et al.*, 2017). In particular, Zou *et al.* (2015) found that in multiple Latino populations, spouse pairs are correlated in their genomic ancestry patterns. Assortative mating by admixture has also been suggested as a cause of geographic structure in patterns of genetic and phenotypic variation in admixed populations (Ruiz-Linares *et al.*, 2014), and as an explanation for empirically observed differences between phenotype-based ancestry classifications and genomic ancestry (Parra *et al.*, 2003).

A particular form of assortative mating by admixture can be more specifically characterized as assortative mating by source population (Ritchie *et al.*, 1989; Howard, 1993; Duenez-Guzman *et al.*, 2009; Melo *et al.*, 2009; Ruiz-Linares *et al.*, 2014; Schumer *et al.*, 2017). In this type of assortative mating by admixture, different groups—source populations and admixed populations, for example—come to have distinct geographic locations, trait preferences, host preferences, or in the case of human populations, social identities (Qian,

1997; Jacobson *et al.*, 2004). Individual ancestry levels within the admixed group are less important for mating choices than the group membership itself. Such a pattern can occur in hybrid zones; for example, Melo *et al.* (2009) found that certain *Heliconius* hybrids show distinct wing color patterns, and during mating, hybrids prefer the hybrid patterns over the patterns displayed in the parental populations.

Positive assortative mating by source population can take place, in which individuals from the source populations preferentially mate within the source groups, while individuals from the admixed group preferentially mate within the admixed group (Figure 1b). Alternatively, negative assortative mating by source population can also take place—for example, during major migration events—in which pairings might be more likely to involve individuals from different source populations (Figure 1c).

Here, we extend mechanistic models describing the evolution of admixture levels in an admixed population to accommodate assortative mating by source population. In Section 2, we develop the mechanistic admixture model, allowing mating pairs to vary in their probability of occurrence based on individual populations of origin. In Sections 3 and 4, we derive recursive expressions for the moments of autosomal and X-chromosomal admixture, respectively, as functions of sex-specific contributions and the properties of assortative mating. Next, in Sections 5 and 6, we analyze the behavior of the moments of the admixture fraction distribution for specific cases of the admixture model with constant contributions over time. We conclude with a discussion in Section 7.

2   Model

We follow the notation of Goldberg and Rosenberg (2015), extending earlier mechanistic admixture models to consider an admixed population that preferentially mates based on population of origin. That is, individuals from each of the populations—source populations S_1 and S_2, and the hybrid population, H—preferentially mate with individuals of the other sex based on population of origin. We consider the case of two source populations, so that individuals come from one of three populations: S_1, H, or S_2. Individuals from the source populations may contribute to pool of possible parents for the next generation of H, but individuals from H cannot contribute to the source populations, and each source population does not contribute to the other source population.

Following Goldberg and Rosenberg (2015), we define the parameters $s_{1,g}^{\delta}$ and $s_{2,g}^{\delta}$ as the fractions of individuals of sex δ at generation g that trace to source populations S_1 and S_2, respectively, in the previous generation. For the admixed population H, h_g^{δ} is the corresponding parameter. We maintain relations between parameter sets from Goldberg *et al.* (2014, eqs. 1–6). Specifically, female (superscript *f*) and male (superscript *m*) contributions must each sum to 1,

$$s_{1,g}^{f} + h_g^{f} + s_{2,g}^{f} = s_{1,g}^{m} + h_g^{m} + s_{2,g}^{m} = 1. \tag{1}$$

Additionally, the total contribution from a source population is the average of the sex-specific contributions from that source,

$$s_{1,g} = \left(s_{1,g}^f + s_{1,g}^m\right)/2 \tag{2}$$

$$s_{2,g} = \left(s_{2,g}^f + s_{2,g}^m\right)/2 \tag{3}$$

$$h_g = \left(h_g^f + h_g^m\right)/2 \tag{4}$$

For a randomly mating population, the probability that an individual in the admixed population has a specific parental pairing is the product of the associated female and male contribution parameters. For example, the probability that an individual in the admixed population in generation g has a female parent from S_1 and a male parent from H is $s_{1,g-1}^f h_{g-1}^m$. To consider deviations from random mating, we define a new parameter $c_{ij,g}$ as the difference between the probability that a mating pair in generation g contains a female from population i and a male from population j (Table 1) and the corresponding probability in a randomly mating population with the same contribution parameters. Here, $i, j \in \{1, h, 2\}$ for populations S_1, H, and S_2, respectively.

The parameters $c_{ij,g}$ govern the strength and direction of assortative mating. We assume that the assortative mating preference is constant over time after the founding of the admixed population. Therefore, we have two sets of parameters: $c_{ij,0}$ for the founding generation, and c_{ij} for all further generations. Because the sum of the probabilities for all parental pairings must be 1, we have

$$0 = c_{11,0} + c_{12,0} + c_{21,0} + c_{22,0} \tag{5}$$

$$0 = c_{11} + c_{1h} + c_{12} = c_{h1} + c_{hh} + c_{h2} = c_{21} + c_{2h} + c_{22} \tag{6}$$

$$0 = c_{11} + c_{h1} + c_{21} = c_{1h} + c_{hh} + c_{2h} = c_{12} + c_{h2} + c_{22}. \tag{7}$$

The values of the c_{ij} are bounded such that the probability of each given parental pairing (Table 1) takes its values in the interval $[0, 1]$, and such that each probability is no greater than the probability of one of its constituent components. For example, if $c_{11} > 0$, then $c_{11} \le \min\left(s_{1,g-1}^f, s_{1,g-1}^m\right) - s_{1,g-1}^f s_{1,g-1}^m$. If $c_{11} < 0$, then $|c_{11}| \le s_{1,g-1}^f s_{1,g-1}^m$. The value of c_{ij} is necessarily zero if the female contribution from population i or the male contribution from population j is zero.

For positive assortative mating in population i, with $i \in \{S_1, H, S_2\}$, individuals in population i are more likely to mate with individuals from their own population than from other populations, so that $c_{ii} > 0$ and $c_{ii} > c_{ij}$ and $c_{ii} > c_{ji}$ for each $j \ne i$. If individuals from a source population are least likely to mate with individuals from the other source population, then we also have $c_{12} < 0$ and $c_{21} < 0$. Similarly, for negative assortative mating, individuals

in each population are less likely to mate with individuals from their own population than with individuals from other populations, so that $c_{ii} < 0$ for all i and $c_{ii} < c_{ij}$ and $c_{ii} < c_{ji}$ for all i and $j \neq i$. If individuals from a source population are most likely to mate with individuals from the other source, then $c_{12} > 0$ and $c_{21} > 0$.

We study the random variable $H_{\alpha, g, \delta}^{\gamma}$ describing the ancestry proportion or admixture fraction from source population α at generation g for a chromosomal type γ in a random individual of sex δ in the admixed population. The chromosomal type is indicated by γ, and it can be either autosomal, A, or X-chromosomal, X. For the autosomes, $H_{\alpha, g, f}^{A}$ and $H_{\alpha, g, m}^{A}$ are identically distributed (Goldberg et al., 2014). For the X chromosome, the distribution of $H_{\alpha, g, f}^{X}$ depends on both female and male admixture distributions in generation $g - 1$, as females inherit one X chromosome from each parent. For $H_{\alpha, g, m}^{X}$, the distribution depends only on the female contributions, as males inherit a single X chromosome from their mothers. Note that it is permitted for an individual in the admixed population to have admixture fraction 0 or 1 if all its ancestors trace to a single source population.

Based on these inheritance patterns, we can write the values for the autosomal and X-chromosomal admixture fractions of an individual randomly chosen from the admixed population given one of nine possible sets of parents, L, along with the probability that an individual has that set of parents (Table 1). The probability of a parental pairing is a function of the sex-specific contributions from the populations and the assortative mating parameters c_{ij}. If $c_{ij,0} = 0$ and $c_{ij} = 0$ for all parental pairings, with j not necessarily distinct from i, this model reduces to the random-mating model of Goldberg and Rosenberg (2015).

3   Moments of the autosomal admixture fraction

3.1   Expectation

Under the model, following Goldberg et al. (2014, eq. 12), we can use the law of total expectation to write a recursion for the expected value of the admixture fraction from source population 1 for a random individual of sex δ sampled from the admixed population in generation g as a function of conditional expectations for all possible parental pairs L. As $H_{1, g, f}^{A}$ and $H_{1, g, m}^{A}$ are identically distributed, we write expressions for sex δ when considering autosomal admixture, understanding that δ takes on the same value, f or m, throughout Section 3. Using the values from Table 1, for the first generation, in which neither parent is from population H, we have

$$
\mathbb{E}\left[H_{1, 1, \delta}^{A}\right] = \left(s_{1,0}^{f} s_{1,0}^{m} + c_{11,0}\right)\mathbb{E}[1] + \left(s_{1,0}^{f} s_{2,0}^{m} + s_{2,0}^{f} s_{1,0}^{m} + c_{12,0} + c_{21,0}\right)\mathbb{E}\left[\frac{1}{2}\right] \\
+ \left(s_{2,0}^{f} s_{2,0}^{m} + c_{22,0}\right)\mathbb{E}[0] .
$$

(8)

For all subsequent generations, $g \geq 2$, we have

$$\mathbb{E}\left[H^A_{1,g,\delta}\right] = \left(s^f_{1,g-1}s^m_{1,g-1} + c_{11}\right)\mathbb{E}[1] + \left(s^f_{1,g-1}h^m_{g-1} + c_{1h}\right)\mathbb{E}\left[\frac{1 + H^A_{1,g-1,m}}{2}\right]$$

$$+ \left(h^f_{g-1}s^m_{1,g-1} + c_{h1}\right)\mathbb{E}\left[\frac{1 + H^A_{1,g-1,f}}{2}\right] +$$

$$\left(s^f_{1,g-1}s^m_{2,g-1} + s^f_{2,g-1}s^m_{1,g-1} + c_{12} + c_{21}\right)\mathbb{E}\left[\frac{1}{2}\right]$$

$$+ \left(h^f_{g-1}h^m_{g-1} + c_{hh}\right)\mathbb{E}\left[\frac{H^A_{1,g-1,f} + H^A_{1,g-1,m}}{2}\right]$$

$$+ \left(h^f_{g-1}s^m_{2,g-1} + c_{h2}\right)\mathbb{E}\left[\frac{H^A_{1,g-1,f}}{2}\right] + \left(s^f_{2,g-1}h^m_{g-1} + c_{2h}\right)\mathbb{E}\left[\frac{H^A_{1,g-1,m}}{2}\right]$$

$$+ \left(s^f_{2,g-1}s^m_{2,g-1} + c_{22}\right)\mathbb{E}[0] . \tag{9}$$

Using eqs. (1)–(4), we can simplify eq. (8) to give, for $g = 1$,

$$\mathbb{E}\left[H^A_{1,1,\delta}\right] = s_{1,0} + c_{11,0} + \frac{c_{12,0} + c_{21,0}}{2} .$$

For $g \geq 2$, recalling that $H^A_{1,g,f}$ and $H^A_{1,g,m}$ are identically distributed, eq. (9) becomes

$$\mathbb{E}\left[H^A_{1,g,\delta}\right] = s_{1,g-1} + c_{11} + \frac{c_{12} + c_{1h} + c_{h1} + c_{21}}{2} + \left(h_{g-1} + \frac{c_{1h}}{2} + \frac{c_{h1}}{2} + c_{hh} + \frac{c_{h2}}{2} + \frac{c_{2h}}{2}\right)\mathbb{E}\left[H^A_{1,g-1,\delta}\right] .$$

Applying eqs. (5)–(7), for the expected value of autosomal admixture in a random individual of sex δ from the admixed population sampled at $g = 1$, we have

$$\mathbb{E}\left[H^A_{1,1,\delta}\right] = s_{1,0} . \tag{10}$$

For subsequent generations $g \geq 2$, we have

$$\mathbb{E}\left[H^A_{1,g,\delta}\right] = s_{1,g-1} + h_{g-1}\mathbb{E}\left[H^A_{1,g-1,\delta}\right] . \tag{11}$$

Eqs. (10) and (11) are the same as were found under a randomly mating population by Goldberg et al. (2014, eqs. 17 and 19). That is, positive or negative assortative mating according to population of origin in an admixed population does not affect the expectation of autosomal admixture. The concordance of mean ancestry under assortative and random mating makes sense because the conditions in eqs. (5)–(7) maintain the ancestry proportions from generation to generation in a manner that does not permit natural selection favoring one or another ancestral membership. The ancestries of the parents contributing to the next generation are the same in both scenarios; assortative mating only changes the probability distribution of the parental pairing.

3.2 Higher moments

Next, we write a general recursion for the higher moments of the autosomal admixture fraction from population S_1 in a randomly chosen individual of sex δ from the admixed population. For moments $k \geq 1$, similar to eqs. 20 and 21 of Goldberg *et al.* (2014) in generation $g = 1$, we obtain

$$
\left(H^A_{1,1,\delta}\right)^k = \begin{cases} 1^k & \text{if } L = S_1S_1, \text{ with } \mathbb{P}[L = S_1S_1] = s^f_{1,0}s^m_{1,0} + c_{11,0} \\ \left(\dfrac{1}{2}\right)^k & \text{if } L = S_1S_2, \text{ with } \mathbb{P}[L = S_1S_2] = s^f_{1,0}s^m_{2,0} + c_{12,0} \\ \left(\dfrac{1}{2}\right)^k & \text{if } L = S_2S_1, \text{ with } \mathbb{P}[L = S_2S_1] = s^f_{2,0}s^m_{1,0} + c_{21,0} \\ 0^k & \text{if } L = S_2S_2, \text{ with } \mathbb{P}[L = S_2S_2] = s^f_{2,0}s^m_{2,0} + c_{22,0}. \end{cases}
\tag{12}
$$

For $g \geq 2$, we have

$$\left(H_{1,g,\delta}^{A}\right)^{k} =$$

$$
\begin{cases}
1^{k} & \text{if } L = S_1 S_1, \text{ with } \mathbb{P}[L = S_1 S_1] = s_{1,g-1}^{f} s_{1,g-1}^{m} + c_{11} \\[2ex]
\left(\dfrac{1 + H_{1,g-1,m}^{A}}{2}\right)^{k} & \text{if } L = S_1 H, \text{ with } \mathbb{P}[L = S_1 H] = s_{1,g-1}^{f} h_{g-1}^{m} + c_{1h} \\[2ex]
\left(\dfrac{1}{2}\right)^{k} & \text{if } L = S_1 S_2, \text{ with } \mathbb{P}[L = S_1 S_2] = s_{1,g-1}^{f} s_{2,g-1}^{m} + c_{12} \\[2ex]
\left(\dfrac{1 + H_{1,g-1,f}^{A}}{2}\right)^{k} & \text{if } L = H S_1, \text{ with } \mathbb{P}[L = H S_1] = h_{g-1}^{f} s_{1,g-1}^{m} + c_{h1} \\[2ex]
\left(\dfrac{H_{1,g-1,f}^{A} + H_{1,g-1,m}^{A}}{2}\right)^{k} & \text{if } L = H H, \text{ with } \mathbb{P}[L = H H] = h_{g-1}^{f} h_{g-1}^{m} + c_{hh} \\[2ex]
\left(\dfrac{H_{1,g-1,f}^{A}}{2}\right)^{k} & \text{if } L = H S_2, \text{ with } \mathbb{P}[L = H S_2] = h_{g-1}^{f} s_{2,g-1}^{m} + c_{h2} \\[2ex]
\left(\dfrac{1}{2}\right)^{k} & \text{if } L = S_2 S_1, \text{ with } \mathbb{P}[L = S_2 S_1] = s_{2,g-1}^{f} s_{1,g-1}^{m} + c_{21} \\[2ex]
\left(\dfrac{H_{1,g-1,m}^{A}}{2}\right)^{k} & \text{if } L = S_2 H, \text{ with } \mathbb{P}[L = S_2 H] = s_{2,g-1}^{f} h_{g-1}^{m} + c_{2h} \\[2ex]
0^{k} & \text{if } L = S_2 S_2, \text{ with } \mathbb{P}[L = S_2 S_2] = s_{2,g-1}^{f} s_{2,g-1}^{m} + c_{22}.
\end{cases}
$$

(13)

For $k = 1$, eqs. (12) and (13) simplify to produce the recursion in Table 1. Using the law of total expectation, and following our calculation for the expected value of autosomal admixture, we can write an expression for the higher moments of autosomal admixture by summing the conditional values for autosomal admixture given the parental pairings over all possible sets of parental source populations. For the first generation, $g = 1$, we have

$$
\mathbb{E}\left[\left(H_{1,1,\delta}^{A}\right)^{k}\right] = \left(s_{1,0}^{f}s_{1,0}^{m} + c_{11,0}\right)\mathbb{E}\left[1^{k}\right] + \left(s_{1,0}^{f}s_{2,0}^{m} + s_{2,0}^{f}s_{1,0}^{m} + c_{12,0} + c_{21,0}\right)\mathbb{E}
$$
$$
\left[\left(\frac{1}{2}\right)^{k}\right] + \left(s_{2,0}^{f}s_{2,0}^{m} + c_{22,0}\right)\mathbb{E}\left[0^{k}\right]. \tag{14}
$$

For $g \geq 2$, we have

$$
\begin{aligned}
\mathbb{E}\left[\left(H_{1,g,\delta}^{A}\right)^{k}\right] &= \left(s_{1,g-1}^{f}s_{1,g-1}^{m} + c_{11}\right)\mathbb{E}\left[1^{k}\right] \\
&+ \left(s_{1,g-1}^{f}h_{g-1}^{m} + c_{1h}\right)\mathbb{E}\left[\left(\frac{1 + H_{1,g-1,m}^{A}}{2}\right)^{k}\right] \\
&+ \left(h_{g-1}^{f}s_{1,g-1}^{m} + c_{h1}\right)\mathbb{E}\left[\left(\frac{1 + H_{1,g-1,f}^{A}}{2}\right)^{k}\right] \\
&+ \left(s_{1,g-1}^{f}s_{2,g-1}^{m} + s_{2,g-1}^{f}s_{1,g-1}^{m} + c_{12} + c_{21}\right)\mathbb{E}\left[\left(\frac{1}{2}\right)^{k}\right] \\
&+ \left(h_{g-1}^{f}h_{g-1}^{m} + c_{hh}\right)\mathbb{E}\left[\left(\frac{H_{1,g-1,f}^{A} + H_{1,g-1,m}^{A}}{2}\right)^{k}\right] \\
&+ \left(h_{g-1}^{f}s_{2,g-1}^{m} + c_{h2}\right)\mathbb{E}\left[\left(\frac{H_{1,g-1,f}^{A}}{2}\right)^{k}\right] + \left(s_{2,g-1}^{f}h_{g-1}^{m} + c_{2h}\right)\mathbb{E}\left[\left(\frac{H_{1,g-1,m}^{A}}{2}\right)^{k}\right] \\
&+ \left(s_{2,g-1}^{f}s_{2,g-1}^{m} + c_{22}\right)\mathbb{E}\left[0^{k}\right].
\end{aligned} \tag{15}
$$

Recalling that $H_{1,g-1,f}^{A}$ and $H_{1,g-1,m}^{A}$ are conditionally independent given $H_{1,g-2,f}^{A}$ and $H_{1,g-2,m}^{A}$ (Goldberg *et al.*, 2014), we use the binomial theorem to simplify eqs. (14) and (15). For $g = 1$, we have

$$
\mathbb{E}\left[\left(H_{1,1,\delta}^{A}\right)^{k}\right] = s_{1,0}^{f}s_{1,0}^{m} + c_{11,0} + \frac{s_{1,0}^{f}s_{2,0}^{m} + s_{2,0}^{f}s_{1,0}^{m} + c_{12,0} + c_{21,0}}{2^{k}}. \tag{16}
$$

For $g \geq 2$, we have

$$
\begin{aligned}
\mathbb{E}\left[\left(H_{1,g,\delta}^{A}\right)^{k}\right] &= s_{1,g-1}^{f}s_{1,g-1}^{m} + c_{11} \\
&+ \frac{s_{1,g-1}^{f}s_{2,g-1}^{m} + s_{2,g-1}^{f}s_{1,g-1}^{m} + c_{12} + c_{21} + c_{1h} + c_{h1}}{2^{k}} \\
&+ \frac{s_{1,g-1}^{f}h_{g-1}^{m} + h_{g-1}^{f}s_{1,g-1}^{m} + c_{1h} + c_{h1}}{2^{k}}\left(\sum_{i=0}^{k}\binom{k}{i}\mathbb{E}\left[\left(H_{1,g-1,\delta}^{A}\right)^{i}\right]\right) \\
&+ \frac{h_{g-1}^{f}h_{g-1}^{m} + c_{hh}}{2^{k}}\left(\sum_{i=0}^{k}\binom{k}{i}\mathbb{E}\left[\left(H_{1,g-1,\delta}^{A}\right)^{k-i}\right]\mathbb{E}\left[\left(H_{1,g-1,\delta}^{A}\right)^{i}\right]\right) \\
&+ \frac{s_{2,g-1}^{f}h_{g-1}^{m} + h_{g-1}^{f}s_{2,g-1}^{m} + c_{h2} + c_{2h}}{2^{k}}\mathbb{E}\left[\left(H_{1,g-1,\delta}^{A}\right)^{k}\right].
\end{aligned} \tag{17}
$$

For $k = 1$, eqs. (16) and (17) reduce to eqs. (10) and (11).

The recursions for the higher moments of the autosomal admixture fraction distribution follow the corresponding expressions for a randomly mating population, but with additional terms that are linear in the coefficients for the mating preferences. Setting the assortative mating parameters $c_{ij,0}$ and c_{ij} equal to 0 for all i and j, eqs. (16) and (17) reduce to the recursion for the moments of admixture for a corresponding randomly mating population, eqs. 24 and 26 from Goldberg *et al.* (2014).

3.3 Variance

Using eqs. (16) and (17) for $k = 2$, we can write expressions for the second moment of autosomal admixture. Recalling eqs. (1)–(4), for $g = 1$, we have

$$\mathbb{E}\left[\left(H_{1,1,\delta}^A\right)^2\right] = \frac{s_{1,0}^f\left(1 + s_{1,0}^m\right) + s_{1,0}^m\left(1 + s_{1,0}^f\right) + 2c_{11,0}}{4}. \tag{18}$$

For $g \geq 2$, because $H_{1,g-1,f}^A$ and $H_{1,g-1,m}^A$ are identically distributed, we have

$$\begin{aligned}
\mathbb{E}\left[\left(H_{1,g,\delta}^A\right)^2\right] &= \frac{s_{1,g-1}^f\left(1 + s_{1,g-1}^m\right) + s_{1,g-1}^m\left(1 + s_{1,g-1}^f\right) + 2c_{11}}{4} \\
&+ \frac{s_{1,g-1}^f h_{g-1}^m + h_{g-1}^f s_{1,g-1}^m + c_{1h} + c_{h1}}{2}\mathbb{E}\left[H_{1,g-1,\delta}^A\right] \\
&+ \frac{h_{g-1}^f h_{g-1}^m + c_{hh}}{2}\left(\mathbb{E}\left[H_{1,g-1,\delta}^A\right]\right)^2 + \frac{h_{g-1}^f + h_{g-1}^m}{4}\mathbb{E}\left[\left(H_{1,g-1,\delta}^A\right)^2\right].
\end{aligned} \tag{19}$$

Using the definition of the variance $\mathbb{V}\left[H_{1,g,\delta}^A\right] = \mathbb{E}\left[\left(H_{1,g,\delta}^A\right)^2\right] - \left(\mathbb{E}\left[H_{1,g,\delta}^A\right]\right)^2$, with the expressions for the expected value of autosomal admixture, eqs. (10) and (11), and the second moment, eqs. (18) and (19), we can write an expression for the variance of autosomal admixture. For $g = 1$, we have

$$\mathbb{V}\left[H_{1,1,\delta}^A\right] = \frac{s_{1,0}^f\left(1 - s_{1,0}^f\right) + s_{1,0}^m\left(1 - s_{1,0}^m\right) + 2c_{11,0}}{4}. \tag{20}$$

For all subsequent generations, $g \geq 2$, we have

$$\begin{aligned}
\mathbb{V}\left[H_{1,g,\delta}^A\right] &= \frac{s_{1,g-1}^f\left(1 - s_{1,g-1}^f\right) + s_{1,g-1}^m\left(1 - s_{1,g-1}^m\right) + 2c_{11}}{4} \\
&+ \frac{c_{1h} + c_{h1} - s_{1,g-1}^f h_{g-1}^f - s_{1,g-1}^m h_{g-1}^m}{2}\mathbb{E}\left[H_{1,g-1,\delta}^A\right] \\
&+ \frac{h_{g-1}^f\left(1 - h_{g-1}^f\right) + h_{g-1}^m\left(1 - h_{g-1}^m\right) + 2c_{hh}}{4}\left(\mathbb{E}\left[H_{1,g-1,\delta}^A\right]\right)^2 \\
&+ \frac{h_{g-1}^f + h_{g-1}^m}{4}\mathbb{V}\left[H_{1,g-1,\delta}^A\right].
\end{aligned} \tag{21}$$

For a randomly mating population, with $c_{ij,0} = 0$ and $c_{ij,g} = 0$ for all i and j, eqs. (20) and (21) reduce to eqs. 32 and 33 from Goldberg *et al.* (2014). Defining the variance of

autosomal admixture in a randomly mating population at $g = 1$ and g 2, eqs. 32 and 33 from Goldberg *et al.* (2014), as $\mathbb{V}_{RM}[H^A_{1,1,\delta}]$ and $\mathbb{V}_{RM}[H^A_{1,g,\delta}]$, respectively, we can rewrite eqs. (20) and (21). For $g = 1$, we have

$$\mathbb{V}[H^A_{1,1,\delta}] = \frac{c_{11,0}}{2} + \mathbb{V}_{RM}[H^A_{1,1,\delta}].$$

(22)

For g 2, recalling eqs. (6) and (7), we have

$$\mathbb{V}[H^A_{1,g,\delta}] = \frac{c_{11}}{2} + \left(\frac{c_{1h} + c_{h1}}{2}\right)\mathbb{E}[H^A_{1,g-1,\delta}] + \frac{c_{hh}}{2}\left(\mathbb{E}[H^A_{1,g-1,\delta}]\right)^2 + \mathbb{V}_{RM}[H^A_{1,g,\delta}].$$

(23)

Eqs. (22) and (23) indicate that, in a single generation, assortative mating changes the variance of admixture compared to a randomly mating population. The variance of autosomal admixture contains information about assortative mating because it depends on the assortative mating parameters c_{ij} and it differs from the expression for a randomly mating population with the same contributions for the source populations.

4   Moments of the X-chromosomal admixture fraction

4.1   Expectation

Following the same approach as in the corresponding derivation for the autosomal admixture fraction, we can use the law of total expectation to write recursions for the moments of X-chromosomal admixture. Because the distribution of admixture differs for females and males on the X chromosome (Table 1), we follow Goldberg and Rosenberg (2015) and write coupled expressions for the moments of female and male X-chromosomal admixture fractions.

For X-chromosomal admixture in a randomly chosen female from the admixed population, the recursive expressions for the moments of admixture match the expressions for autosomal admixture presented in Section 3, exchanging $\gamma = A$ for $\gamma = X$ in the superscript. However, X-chromosomal admixture is not identically distributed in females and in males from the admixed population. Therefore, certain expressions for the moments of admixture cannot be reduced as they were for the autosomes.

Using eqs. (1)–(7), we can use the values from Table 1 to write expressions for the expectation of X-chromosomal admixture. For X-chromosomal admixture in a female from the admixed population, the recursive expressions for the expectation of admixture are the same as for autosomal admixture (eqs. (10) and (11)), exchanging $\gamma = A$ for $\gamma = X$ in the superscript. For X-chromosomal admixture in males, for $g = 1$, we have

$$\mathbb{E}[H^X_{1,1,m}] = S^f_{1,0} + c_{11,0} + c_{12,0} = S^f_{1,0}.$$

(24)

For g 2, we have

$$\mathbb{E}\left[H^X_{1,g,m}\right] = s^f_{1,g-1} + c_{11} + c_{1h} + c_{12} + \left(h^f_{g-1} + c_{h1} + c_{hh} + c_{h2}\right)\mathbb{E}\left[H^X_{1,g-1,f}\right]$$
$$= s^f_{1,g-1} + h^f_{g-1}\mathbb{E}\left[H^X_{1,g-1,f}\right]. \tag{25}$$

As was the case for the autosomes, the expectation of X-chromosomal admixture in eqs. (24) and (25) is the same under assortative mating as the expected value for a randomly mating population (Goldberg and Rosenberg, 2015, eqs. 3–4).

4.2 Higher moments

Following the derivation for the autosomes and using Table 1, we can write general coupled recursions for the higher moments of the X-chromosomal admixture fraction from population S_1 in a randomly chosen female and male from the admixed population. As was true for the expectation, the recursion for the X-chromosomal admixture fraction in a female from the admixed population for $k \geq 1$ is the same as the recursion for autosomal admixture in eqs. (12) and (13), exchanging the superscript A for X. For X-chromosomal admixture in a male, we have, for $g = 1$,

$$\left(H^X_{1,1,m}\right)^k = \begin{cases} 1^k & \text{if } L = S_1S_1, \text{ with } \mathbb{P}[L = S_1S_1] = s^f_{1,0}s^m_{1,0} + c_{11,0} \\ 1^k & \text{if } L = S_1S_2, \text{ with } \mathbb{P}[L = S_1S_2] = s_{1,0}s^m_{2,0} + c_{12,0} \\ 0^k & \text{if } L = S_2S_1, \text{ with } \mathbb{P}[L = S_2S_1] = s_{2,0}s^m_{1,0} + c_{21,0} \\ 0^k & \text{if } L = S_2S_2, \text{ with } \mathbb{P}[L = S_2S_2] = s^2_{2,0}s_{2,0} + c_{22,0} \end{cases}$$

For $g \geq 2$, we have

$$\left(H^X_{1,g,\delta}\right)^k = \begin{cases} 1^k & \text{if } L = S_1S_1, \text{ with } \mathbb{P}[L = S_1S_1] = s^f_{1,g-1}s^m_{1,g-1} + c_{11,g} \\ 1^k & \text{if } L = S_1H, \text{ with } \mathbb{P}[L = S_1H] = s^f_{1,g-1}h^m_{g-1} + c_{1h,g} \\ 1^k & \text{if } L = S_1S_2, \text{ with } \mathbb{P}[L = S_1S_2] = s^f_{1,g-1}s^m_{2,g-1} + c_{12,g} \\ \left(H^X_{1,g-1,f}\right)^k & \text{if } L = HS_1, \text{ with } \mathbb{P}[L = HS_1] = h^f_{g-1}s^m_{1,g-1} + c_{h1,g} \\ \left(H^X_{1,g-1,f}\right)^k & \text{if } L = HH, \text{ with } \mathbb{P}[L = HH] = h^f_{g-1}h^m_{g-1} + c_{hh,g} \\ \left(H^X_{1,g-1,f}\right)^k & \text{if } L = HS_2, \text{ with } \mathbb{P}[L = HS_2] = h^f_{g-1}s^m_{2,g-1} + c_{h2,g} \\ 0^k & \text{if } L = S_2S_1, \text{ with } \mathbb{P}[L = S_2S_1] = s^f_{2,g-1}s^m_{1,g-1} + c_{21,g} \\ 0^k & \text{if } L = S_2H, \text{ with } \mathbb{P}[L = S_2H] = s^f_{2,g-1}h^m_{g-1} + c_{2h,g} \\ 0^k & \text{if } L = S_2S_2, \text{ with } \mathbb{P}[L = S_2S_2] = s^f_{2,g-1}s^m_{2,g-1} + c_{22,g}. \end{cases}$$

Using the law of total expectation and the binomial theorem, and following the derivation of the autosomal admixture moments, we can write simplified coupled expressions for the moments of X-chromosomal admixture, separately considering a female and a male from the admixed population for $k \geq 1$. For $g = 1$, the recursion for the kth moment of X-

chromosomal admixture in a female is equal to the corresponding recursion in the case of autosomal admixture, eq. (16). For a male in $g = 1$, we have

$$\mathbb{E}\left[\left(H_{1,1,m}^X\right)^k\right] = s_{1,0}^f.$$ (26)

For $g \geq 2$, using the conditional independence of $H_{1,g-1,f}^X$ and $H_{1,g-1,m}^X$ given $H_{1,g-2,f}^X$ and $H_{1,g-2,m}^X$, we have

$$
\begin{aligned}
\mathbb{E}\left[\left(H_{1,g,f}^X\right)^k\right] = {} & s_{1,g-1}^f s_{1,g-1}^m + c_{11} \\
& + \frac{s_{1,g-1}^f s_{2,g-1}^m + s_{2,g-1}^f s_{1,g-1}^m + c_{12} + c_{21} + c_{1h} + c_{h1}}{2^k} \\
& + \frac{h_{g-1}^f s_{1,g-1}^m + c_{h1}}{2^k}\left(\sum_{i=0}^k \binom{k}{i}\mathbb{E}\left[\left(H_{1,g-1,f}^X\right)^i\right]\right) \\
& + \frac{s_{1,g-1}^f h_{g-1}^m + c_{1h}}{2^k}\left(\sum_{i=0}^k \binom{k}{i}\mathbb{E}\left[\left(H_{1,g-1,m}^X\right)^i\right]\right) \\
& + \frac{h_{g-1}^f h_{g-1}^m + c_{hh}}{2^k}\left(\sum_{i=0}^k \binom{k}{i}\mathbb{E}\left[\left(H_{1,g-1,f}^X\right)^{k-i}\right]\mathbb{E}\left[\left(H_{1,g-1,m}^X\right)^i\right]\right) \\
& + \frac{h_{g-1}^f s_{2,g-1}^m + c_{h2}}{2^k}\mathbb{E}\left[\left(H_{1,g-1,f}^X\right)^k\right] + \frac{s_{2,g-1}^f h_{g-1}^m + c_{2h}}{2^k}\mathbb{E}\left[\left(H_{1,g-1,m}^X\right)^k\right],
\end{aligned}
$$ (27)

$$\mathbb{E}\left[\left(H_{1,g,m}^X\right)^k\right] = s_{1,g-1}^f + h_{g-1}^f\mathbb{E}\left[\left(H_{1,g-1,f}^X\right)^k\right].$$ (28)

Unlike for the autosomes, the female and male admixture fractions are not identically distributed or conditionally independent, so the dependence on both $H_{1,g-1,f}^X$ and $H_{1,g-1,m}^X$ in eq. (27) cannot be further reduced. For $k \geq 2$, moments of the X-chromosomal admixture fraction depend on the assortative mating parameters, $c_{ij,0}$ and c_{ij}. However, conditional on $H_{1,g-1,f}^X$, moments of the X-chromosomal admixture fraction sampled in a male from the admixed population do not depend on the assortative mating parameters. Because a single copy of the X chromosome is inherited from mother to son, the distribution of admixture for male X chromosomes in a given generation is affected only by the origin of the mother and not by the probabilities of parental pairings in Table 1.

4.3 Variance

For $k = 2$, we can use eqs. (26)–(28) to write coupled expressions for the second moment of X-chromosomal admixture in a randomly selected female and male from the admixed population. Recalling eqs. (1)–(4), for $g = 1$, the second moment of X-chromosomal admixture in a female is the same as the expression for autosomal admixture in eq. (18), substituting superscript X in place of A. For the second moment of X-chromosomal admixture in males in $g = 1$, we have

$$\mathbb{E}\left[\left(H_{1,1,m}^{X}\right)^{2}\right] = s_{1,0}^{f}.$$

For $g \geq 2$, we have

$$\mathbb{E}\left[\left(H_{1,g,f}^{X}\right)^{2}\right] = \frac{s_{1,g-1}^{f}\left(1 + s_{1,g-1}^{m}\right) + s_{1,g-1}^{m}\left(1 + s_{1,g-1}^{f}\right) + 2c_{11}}{4} + \frac{h_{g-1}^{f}s_{1,g-1}^{m} + c_{h1}}{2}\mathbb{E}\left[H_{1,g-1,f}^{X}\right]$$
$$+ \frac{s_{1,g-1}^{f}h_{g-1}^{m} + c_{1h}}{2}\mathbb{E}\left[H_{1,g-1,m}^{X}\right] + \frac{h_{g-1}^{f}h_{g-1}^{m} + c_{hh}}{2}\mathbb{E}\left[H_{1,g-1,f}^{X}\right]\mathbb{E}\left[H_{1,g-1,m}^{X}\right]$$
$$+ \frac{h_{g-1}^{f}}{4}\mathbb{E}\left[\left(H_{1,g-1,f}^{X}\right)^{2}\right] + \frac{h_{g-1}^{m}}{4}\mathbb{E}\left[\left(H_{1,g-1,m}^{X}\right)^{2}\right],$$

$$\mathbb{E}\left[\left(H_{1,g,m}^{X}\right)^{2}\right] = s_{1,g-1}^{f} + h_{g-1}^{f}\mathbb{E}\left[\left(H_{1,g-1,f}^{X}\right)^{2}\right].$$

Following our derivation of the variance of autosomal admixture in Section 3.3, we can write the variance of X-chromosomal admixture in a female and in a male from the admixed population using the expected values and second moments of X-chromosomal admixture. For $g = 1$, we have

$$\mathbb{V}\left[H_{1,1,f}^{X}\right] = \frac{s_{1,0}^{f}\left(1 - s_{1,0}^{f}\right) + s_{1,0}^{m}\left(1 - s_{1,0}^{m}\right) + 2c_{11,0}}{4}, \tag{29}$$

$$\mathbb{V}\left[H_{1,1,m}^{X}\right] = s_{1,0}^{f}\left(1 - s_{1,0}^{f}\right). \tag{30}$$

For $g \geq 2$, we have

$$\mathbb{V}\left[H_{1,g,f}^{X}\right] = \frac{s_{1,g-1}^{f}\left(1 - s_{1,g-1}^{f}\right) + s_{1,g-1}^{m}\left(1 - s_{1,g-1}^{m}\right) + 2c_{11}}{4}$$
$$+ \frac{c_{h1} - s_{1,g-1}^{f}h_{g-1}^{f}}{2}\mathbb{E}\left[H_{1,g-1,f}^{X}\right] + \frac{c_{1h} - s_{1,g-1}^{m}h_{g-1}^{m}}{2}\mathbb{E}\left[H_{1,g-1,m}^{X}\right]$$
$$+ \frac{c_{hh}}{2}\mathbb{E}\left[H_{1,g-1,f}^{X}\right]\mathbb{E}\left[H_{1,g-1,m}^{X}\right] + \frac{h_{g-1}^{f}\left(1 - h_{g-1}^{f}\right)}{4}\left(\mathbb{E}\left[H_{1,g-1,f}^{X}\right]\right)^{2} \tag{31}$$
$$+ \frac{h_{g-1}^{m}\left(1 - h_{g-1}^{m}\right)}{4}\left(\mathbb{E}\left[H_{1,g-1,m}^{X}\right]\right)^{2} + \frac{h_{g-1}^{f}}{4}\mathbb{V}\left[H_{1,g-1,f}^{X}\right]$$
$$+ \frac{h_{g-1}^{m}}{4}\mathbb{V}\left[H_{1,g-1,m}^{X}\right],$$

$$\mathbb{V}\left[H_{1,g,m}^{X}\right] = s_{1,g-1}^{f}\left(1 - s_{1,g-1}^{f}\right) - 2s_{1,g-1}^{f}h_{g-1}^{f}\mathbb{E}\left[H_{1,g-1,f}^{X}\right]$$
$$+ h_{g-1}^{f}\left(1 - h_{g-1}^{f}\right)\left(\mathbb{E}\left[H_{1,g-1,f}^{X}\right]\right)^{2} + h_{g-1}^{f}\mathbb{V}\left[H_{1,g-1,f}^{X}\right]. \tag{32}$$

We denote the variance of X-chromosomal admixture under random mating, in a random female and male from the admixed population, by $\mathbb{V}_{RM}\left[H_{1,g,f}^X\right]$ and $\mathbb{V}_{RM}\left[H_{1,g,m}^X\right]$, respectively. Setting all $c_{ij,0}$ and c_{ij} parameters to zero, for $g = 1$, we have

$$\mathbb{V}_{RM}\left[H_{1,1,f}^X\right] = \frac{s_{1,0}^f\left(1 - s_{1,0}^f\right) + s_{1,0}^m\left(1 - s_{1,0}^m\right)}{4}, \tag{33}$$

$$\mathbb{V}_{RM}\left[H_{1,1,m}^X\right] = s_{1,0}^f\left(1 - s_{1,0}^f\right). \tag{34}$$

For $g \geq 2$, we have

$$\begin{aligned}
\mathbb{V}_{RM}\left[H_{1,g,f}^X\right] = {} & \frac{s_{1,g-1}^f\left(1 - s_{1,g-1}^f\right) + s_{1,g-1}^m\left(1 - s_{1,g-1}^m\right)}{4} \\
& - \frac{s_{1,g-1}^f h_{g-1}^f}{2}\mathbb{E}\left[H_{1,g-1,f}^X\right] - \frac{s_{1,g-1}^m h_{g-1}^m}{2}\mathbb{E}\left[H_{1,g-1,m}^X\right] \\
& + \frac{h_{g-1}^f\left(1 - h_{g-1}^f\right)}{4}\left(\mathbb{E}\left[H_{1,g-1,f}^X\right]\right)^2 + \frac{h_{g-1}^m\left(1 - h_{g-1}^m\right)}{4}\left(\mathbb{E}\left[H_{1,g-1,m}^X\right]\right)^2 \\
& + \frac{h_{g-1}^f}{4}\mathbb{V}\left[H_{1,g-1,f}^X\right] + \frac{h_{g-1}^m}{4}\mathbb{V}\left[H_{1,g-1,m}^X\right],
\end{aligned} \tag{35}$$

$$\begin{aligned}
\mathbb{V}_{RM}\left[H_{1,g,m}^X\right] = {} & s_{1,g-1}^f\left(1 - s_{1,g-1}^f\right) - 2s_{1,g-1}^f h_{g-1}^f\mathbb{E}\left[H_{1,g-1,f}^X\right] \\
& + h_{g-1}^f\left(1 - h_{g-1}^f\right)\left(\mathbb{E}\left[H_{1,g-1,f}^X\right]\right)^2 + h_{g-1}^f\mathbb{V}\left[H_{1,g-1,f}^X\right].
\end{aligned} \tag{36}$$

Using eqs. (33)–(36), we can rewrite the variances in eqs. (29)–(32) as functions of the variance under a similar randomly mating population. For $g = 1$, we have

$$\mathbb{V}\left[H_{1,1,f}^X\right] = \frac{c_{11,0}}{2} + \mathbb{V}_{RM}\left[H_{1,1,f}^X\right], \tag{37}$$

$$\mathbb{V}\left[H_{1,1,m}^X\right] = \mathbb{V}_{RM}\left[H_{1,1,m}^X\right]. \tag{38}$$

For $g \geq 2$, conditional on $H_{1,g-1,f}^X$ and $H_{1,g-1,m}^X$, and recalling eqs. (6) and (7), we have

$$\begin{aligned}
\mathbb{V}\left[H_{1,g,f}^X\right] = {} & \frac{c_{11}}{2} + \frac{c_{h1}}{2}\mathbb{E}\left[H_{1,g-1,f}^X\right] + \frac{c_{1h}}{2}\mathbb{E}\left[H_{1,g-1,m}^X\right] \\
& + \frac{c_{hh}}{2}\mathbb{E}\left[H_{1,g-1,f}^X\right]\mathbb{E}\left[H_{1,g-1,m}^X\right] + \mathbb{V}_{RM}\left[H_{1,g,f}^X\right],
\end{aligned} \tag{39}$$

$$\mathbb{V}\left[H_{1,g,m}^X\right] = \mathbb{V}_{RM}\left[H_{1,g,m}^X\right]. \tag{40}$$

In contrast to the expectation, the variance of X-chromosomal admixture depends on the assortative mating parameters c_{11}, c_{h1}, c_{1h}, and c_{hh}. Conditional on the female and male variances of X-chromosomal admixture in generation $g-1$, the variance of X-chromosomal admixture in a male sampled in generation g is equivalent to that in a corresponding randomly mating population.

5   Special case: a single admixture event

To analyze the behavior of the model, we study the moments of the distribution of autosomal and X-chromosomal admixture for two special cases of constant admixture processes over time. First, in Section 5, we consider a case with no contributions from the sources after the initial founding. That is, for $g \geq 2$, we set $s_{1,g-1}^f = s_{1,g-1}^m = s_{2,g-1}^f = s_{2,g-1}^m = 0$ and $h_{g-1}^f = h_{g-1}^m = 1$. Because only the admixed population contributes after the first generation, all assortative mating parameters for further generations equal 0, $c_{ij} = 0$.

Next, in Section 6, we examine the special case of constant non-zero contributions over time. In Sections 5 and 6, we assume that the assortative mating process is the same for both sexes. That is, $c_{ij,0} = c_{ji,0}$ and $c_{ij} = c_{ji}$. The behavior of the expectation of autosomal and X-chromosomal admixture is equivalent to that seen in a randomly mating population, as studied by Goldberg *et al.* (2014) and Goldberg and Rosenberg (2015). Therefore, as in Goldberg *et al.* (2014), we focus on the variance, on which the assortative mating parameters have an impact. In Section 5, we consider the variances of autosomal and X-chromosomal admixture as functions of $s_{1,0}^f, s_{1,0}^m, c_{11,0}$, and g. We focus on the dependence of the variance of admixture on the assortative mating parameter $c_{11,0}$. Note that the variance of admixture under random mating ($c_{11,0} = 0$) in a single-admixture scenario was studied in detail for autosomes by Goldberg *et al.* (2014); for completeness, we include analogous calculations for the X chromosome in Appendix A.

5.1   Autosomes

Under this model of a single admixture event, we can write an exact expression for the variance of autosomal admixture. Using the equation for the variance of autosomal admixture in eq. (20), we observe that the variance for a single admixture event can be written as a geometric sequence with ratio $\frac{1}{2}$. For $g \geq 1$, we have

$$\mathbb{V}\left[H_{1,g,\delta}^A\right] = \frac{s_{1,0}^f\left(1 - s_{1,0}^f\right) + s_{1,0}^m\left(1 - s_{1,0}^m\right) + 2c_{11,0}}{2^{g+1}}. \tag{41}$$

Recalling the variance of autosomal admixture for a randomly mating population produced by a single admixture event from Goldberg *et al.* (2014, eq. 35), we rewrite eq. (41) as a function of $\mathbb{V}_{RM}\left[H_{1,g,\delta}^A\right]$,

$$\mathbb{V}\left[H_{1,g,\delta}^A\right] = \mathbb{V}_{RM}\left[H_{1,g,\delta}^A\right] + \frac{c_{11,0}}{2^g}. \tag{42}$$

As is true in a randomly-mating population, the long-term limit of the variance over time is zero, with the effect of assortative mating being multiplied by $\frac{1}{2}$ each generation, and with the distribution of admixture narrowing around the same mean value as in a randomly mating population, $s_{1,0}$ (Goldberg *et al.*, 2014, eq. 35).

For positive assortative mating, with $c_{11,0} > 0$, the variance (eq. (42)) is larger than in a corresponding randomly mating population (Goldberg *et al.*, 2014, eq. 35). For negative assortative mating, $c_{11,0} < 0$, the variance is smaller (Figure 2a). The effect of the initial non-random mating on the population decreases monotonically each generation, as all individuals mate randomly within the admixed population, with no further contributions from the sources. In a given generation and given the contributions $s_{1,0}^f$ and $s_{1,0}^m$, the variance is directly related to $c_{11,0}$. Therefore, the maximal and minimal variance occur when $c_{11,0}$ is maximized and minimized, respectively.

Negative assortative mating also introduces a minimum in the variance of autosomal admixture, with $\mathbb{V}\left[H_{1,g,\delta}^A\right] = 0$, when the numerator of eq. (41) is zero, at $c_{11,0} = -s_{1,0}^f\left(1 - s_{1,0}^f\right)/2 - s_{1,0}^m\left(1 - s_{1,0}^m\right)/2$. Under this scenario of complete negative assortative mating, individuals mate only with individuals from the other source population.

5.2 X chromosome

Similarly to the case of autosomal admixture, we can write an exact solution for the variance of X-chromosomal admixture for a scenario with a single admixture event and no subsequent admixture. We follow the derivation of the expectation of X-chromosomal admixture in Goldberg and Rosenberg (2015, eqs. 6–11). Recalling eqs. (29) and (30), we rewrite the variance of X-chromosomal admixture under a scenario of a single admixture event as a coupled pair of recursions for $\mathbb{V}\left[H_{1,g,f}^X\right]$ and $\mathbb{V}\left[H_{1,g,m}^X\right]$.

For $g = 1$,

$$\mathbb{V}\left[H_{1,1,f}^X\right] = \frac{s_{1,0}^f\left(1 - s_{1,0}^f\right) + s_{1,0}^m\left(1 - s_{1,0}^m\right) + 2c_{11,0}}{4},$$

$$\mathbb{V}\left[H_{1,1,m}^X\right] = s_{1,0}^f\left(1 - s_{1,0}^f\right).$$

For $g \geq 2$,

$$\mathbb{V}\left[H_{1,g,f}^X\right] = \frac{1}{4}\left(\mathbb{V}\left[H_{1,g-1,f}^X\right] + \mathbb{V}\left[H_{1,g-1,m}^X\right]\right),$$

$$\mathbb{V}\left[H_{1,g,m}^X\right] = \mathbb{V}\left[H_{1,g-1,f}^X\right]. \tag{43}$$

Next, for $g \geq 3$, we can rewrite the variance of X-chromosomal admixture in a female as a two-generation recursion,

$$\mathbb{V}\left[H_{1,g,f}^{X}\right] = \frac{\mathbb{V}\left[H_{1,g-1,f}^{X}\right] + \mathbb{V}\left[H_{1,g-2,f}^{X}\right]}{4}.$$

The variance of X-chromosomal admixture in a female is similar in form to the expectation seen in Goldberg and Rosenberg (2015, eq. 8), and we can take an analogous approach to solving the recursion. The recursion for the variance includes a factor of 4 in each generation; therefore, the closed-form expression contains a factor of 4^g. We define $y_g = 4^g \mathbb{V}\left[H_{1,g,f}^{X}\right]$, and for $g \geq 3$, we have

$$\frac{y_g}{4^g} = \frac{y_{g-1}/4^{g-1} + y_{g-2}/4^{g-2}}{4}.$$

Multiplying both sides by 4^g, we have a recursion $y_g = y_{g-1} + 4y_{g-2}$, with

$$y_1 = \mathbb{V}\left[H_{1,1,f}^{X}\right] = \frac{s_{1,0}^{f}\left(1 - s_{1,0}^{f}\right) + s_{1,0}^{m}\left(1 - s_{1,0}^{m}\right) + 2c_{11,0}}{4}$$

$$y_2 = \frac{\mathbb{V}\left[H_{1,1,f}^{X}\right] + \mathbb{V}\left[H_{1,1,m}^{X}\right]}{4} = \frac{5s_{1,0}^{f}\left(1 - s_{1,0}^{f}\right) + s_{1,0}^{m}\left(1 - s_{1,0}^{m}\right) + 2c_{11,0}}{16}.$$

Next, for $g \geq 3$, we write y_g as $A_g s_{1,0}^{f}\left(1 - s_{1,0}^{f}\right) + B_g s_{1,0}^{m}\left(1 - s_{1,0}^{m}\right) + 2B_g c_{11,0}$, where

$$A_g = A_{g-1} + 4A_{g-2}$$

$$B_g = B_{g-1} + 4B_{g-2}$$

and $A_1 = 1$, $A_2 = 5$, $B_1 = 1$, and $B_2 = 1$. Calculating further values of B_g, we note that $B_g = A_{g-1}$. We can therefore rewrite $y_g = A_g s_{1,0}^{f}\left(1 - s_{1,0}^{f}\right) + A_{g-1} s_{1,0}^{m}\left(1 - s_{1,0}^{m}\right) + 2A_{g-1}c_{11,0}$, leading to

$$\mathbb{V}\left[H_{1,g,f}^{X}\right] = \frac{A_g s_{1,0}^{f}\left(1 - s_{1,0}^{f}\right) + A_{g-1}\left[s_{1,0}^{m}\left(1 - s_{1,0}^{m}\right) + 2c_{11,0}\right]}{4^g}. \tag{44}$$

The sequence A_g is reminiscent of a similar recursively defined sequence that appears in the X-chromosomal expectation for a single admixture event in Goldberg and Rosenberg (2015). A_g is a Lucas sequence (OEIS A006131), which can be written in closed form by using its generating function, $a(x) = 1/(1 - x - 4x^2)$. For $g \geq 1$, we have

$$A_g = \frac{\left(\frac{1 + \sqrt{17}}{2}\right)^{g+1} - \left(\frac{1 - \sqrt{17}}{2}\right)^{g+1}}{\sqrt{17}}.$$

For a single admixture event, the variance of the admixture fraction in both males and females (eqs. (43) and (44)) decays to zero over time, as 4^g grows faster than A_g. The effect of assortative mating on the X-chromosomal variance decreases at a slower rate—$c_{11,0}$ accumulates a factor of $\frac{1 + \sqrt{17}}{8}$ each generation—than for the autosomal variance, in which $c_{11,0}$ is multiplied by a factor of $\frac{1}{2}$ each generation (eq. (42)).

For specified contribution parameters $s_{1,0}^f$ and $s_{1,0}^m$, given g, the maximum and minimum of the variance of X-chromosomal admixture as functions of $c_{11,0}$ occur at the same values of $c_{11,0}$ as the maximum and minimum of the autosomal variance: the maximum and minimum of $c_{11,0}$.

Figures 2–4 analyze the behavior of the variances of autosomal and X-chromosomal admixture for the case of a single admixture event. Figure 2 plots $\mathbb{V}\left[H_{1,g,\delta}^A\right]$, $\mathbb{V}\left[H_{1,g,f}^X\right]$, and $\mathbb{V}\left[H_{1,g,m}^X\right]$ in relation to the time since admixture. For positive assortative mating, we observe that for both the autosomes (Fig. 2a) and the X chromosome (Figs. 2b and 2c), the variance of admixture is greater relative to corresponding randomly mating populations with the same contribution parameters. For negative assortative mating, the variance is smaller.

Figure 3 notes a special case of the minimum of the autosomal variance at constant $s_{1,0}^f = s_{1,0}^m = 0.5$ and $c_{11,0} = -s_{1,0}^f\left(1 - s_{1,0}^f\right)/2 - s_{1,0}^m\left(1 - s_{1,0}^m\right)/2 = -0.25$. Using eq. (41), $\mathbb{V}\left[H_{1,g,\delta}^A\right] = 0$, even though this set of contributions ($s_{1,0}^f = s_{1,0}^m = 0.5$) is a maximum of the variance for a randomly mating population (Goldberg *et al.*, 2014).

Figure 4 plots the variance of admixture over the permissible range of $c_{11,0}$, [−0.25, 0.25], for each value of g in [1, 6] when $s_{1,0}^f = s_{1,0}^m = 0.5$. The variance of admixture increases linearly with $c_{11,0}$. That is, positive values of $c_{11,0}$ produce a higher variance of admixture than that of a randomly mating population ($c_{11,0} = 0$), and negative values of $c_{11,0}$ produce a lower variance. The autosomal variance decreases monotonically with g for each value of $c_{11,0}$. The X-chromosomal variance is not monotonic as g increases for a specific value of $c_{11,0}$, as can be seen from the existence of points of intersection of the lines (Fig. 4b), and as is described for the expectation when $c_{11,0} = 0$ by Goldberg and Rosenberg (2015). For the autosomes, the minimum observed in Figure 4a across values of $c_{11,0}$ with $S_{1,0}^f$ and $S_{1,0}^m$ held constant occurs at $c_{11,0} = -0.25$, the minimum depicted in Figure 3, with $\mathbb{V}\left[H_{1,g,\delta}^A\right] = 0$ for all g. The male X-chromosomal admixture in $g = 1$ is 0 and does not depend on $c_{11,0}$ (Fig. 4c).

For all initial contributions $s_{1,0}^f$ and $s_{1,0}^m$, and assortative mating parameter values $c_{11,0}$, the limit of the autosomal and X-chromosomal variances as $g \to \infty$ equals zero. That is, for admixture processes with no contributions after the founding, in both assortative and randomly mating populations, the distribution of admixture narrows around the mean.

6 Special case: constant contributions over time

Next, we consider the special case of constant, non-zero contributions from the source populations to the admixed population over time, after its initial founding. As in Goldberg *et al.* (2014) and Goldberg and Rosenberg (2015), we choose the sex-specific parameters to have values that are constant in time, $\left(s_{1,g}^f, s_{1,g}^m, s_{2,g}^f, s_{2,g}^m\right) = \left(s_1^f, s_1^m, s_2^f, s_2^m\right)$ for all g 1. Sex-specific contributions take their range in [0, 1]; however, the total contributions from each source, s_1 and s_2, take their range in (0, 1). We maintain separate parameters for the founding contributions, $s_{1,0}^f, s_{1,0}^m, s_{2,0}^f, s_{2,0}^m$. Here, we analyze the limiting behavior of the variance of autosomal and X-chromosomal admixture in an assortatively mating population under this scenario of constant contributions from the source populations to the admixed population in each generation after founding.

6.1 Limit of the variance of admixture over time

Appendix B uses the closed-form expressions for the variance of autosomal admixture in a randomly mating population, derived by Goldberg *et al.* (2014), to write a closed-form expression for the variance of autosomal admixture in eqs. (20) and (21). Taking the limit as $g \to \infty$ of the expression in eq. (57) and assuming $c_{ij} = c_{ji}$, we have

$$\lim_{g \to \infty} \mathbb{V}\left[H_{1,g,\delta}^A\right] = \frac{c_{11}}{2} + c_{1h}\left(\frac{s_1}{1-h}\right) + \frac{c_{hh}}{2}\left(\frac{s_1}{1-h}\right)^2 + \lim_{g \to \infty} \mathbb{V}_{RM}\left[H_{1,g,\delta}^A\right], \qquad (45)$$

where $\lim_{g \to \infty} \mathbb{V}_{RM}\left[H_{1,g,\delta}^A\right]$ is given by Goldberg *et al.* (2014, eq. 53) and is reproduced in Appendix B, eq. (58).

Next, we consider the variance of X-chromosomal admixture. Using a generating function approach, Appendix C calculates the limits of $\mathbb{V}_{RM}\left[H_{1,g,f}^X\right]$ and of $\mathbb{V}_{RM}\left[H_{1,g,m}^X\right]$ (eqs. (71) and (72)) as $g \to \infty$. Recalling eqs. (37)–(40) and using eq. (73) for the definition of P_3, we have

$$\begin{aligned}
\lim_{g \to \infty} \mathbb{V}\left[H_{1,g,f}^X\right] &= \frac{c_{11}}{2} + \frac{c_{1h}}{2}\left(P_3 + s_1^f + h^f P_3\right) + \frac{c_{hh}}{2}P_3\left(s_1^f + h^f P_3\right) \\
&+ \lim_{g \to \infty} \mathbb{V}_{RM}\left[H_{1,g,f}^X\right]
\end{aligned} \qquad (46)$$

$$\lim_{g \to \infty} \mathbb{V}\left[H_{1,g,m}^X\right] = s_1^f\left(1 - s_1^f\right) - 2s_1^f h^f P_3 + h^f\left(1 - h^f\right)P_3^2 + h^f \lim_{g \to \infty} \mathbb{V}\left[H_{1,g,f}^X\right]. \qquad (47)$$

Figure 5 plots the variance of autosomal and X-chromosomal admixture over time for various values of the assortative mating parameters, both positive and negative, for constant contributions from the source populations over time. In these plots of the variance of admixture over time, we hold $c_{11} = c_{22}$ and $c_{1h} = 0$, and by consequence, $c_{h1} = c_{h2} = c_{2h} = 0$. This set of values for the c_{ij} can be interpreted as representing (1) equal levels of within-population preference for the two source populations, and (2) for each source, the mating probability between hybrid and source individuals is the same as in a randomly mating population.

These limiting variances often differ from those of corresponding randomly mating populations with the same contributions. For the autosomes (Fig. 5a) and the X chromosome (Figs. 5b,c), the variance has a linear relationship with c_{ii}, with positive values increasing the variance and negative values decreasing it. The limits of the autosomal and X-chromosomal variances in admixture depend on the ongoing sex-specific contributions from the source populations, $s_1^f, s_1^m, s_2^f, s_2^m$, and on the assortative mating parameters c_{ij}. The limits do not depend on the contributions or assortative mating parameters in the founding generation, $s_{1,0}^f, s_{1,0}^m, s_{2,0}^f, s_{2,0}^m, c_{11,0}$.

6.2 Positive assortative mating can increase or decrease the variance relative to a corresponding randomly mating population

We next consider the effect of assortative mating on the variance of admixture compared to a corresponding randomly mating population with the same contribution parameters. We demonstrate that the behavior of the variance depends on the nature of the assortative mating parameters. That is, positive assortative mating can either increase or decrease the variance relative to a corresponding randomly mating population.

First, we consider assortative mating in which preference for one's own population is the largest of the three preferences, and, for source populations, preference for the other source population is the smallest of the three. Additionally, all three populations, S_1, H, S_2, experience the same direction of mating preference. For positive assortative mating, this scenario amounts to $c_{11} > c_{1h} > c_{12}$, $c_{22} > c_{h2} > c_{12}$ and $c_{hh} > 0$. For negative assortative mating, we have $c_{11} < c_{1h} < c_{12}$, $c_{22} < c_{h2} < c_{12}$, and $c_{hh} < 0$. The variance of X-chromosomal admixture in males depends on the contribution parameters and the limiting admixture in females, but not on the assortative mating parameters beyond their role in female admixture (eq. (47)). Therefore, we consider the variances only of autosomal admixture and X-chromosomal admixture in females.

6.2.1 Positive assortative mating for each population—Here, we show that for positive assortative mating, eq. (45) can be written as the sum of a positive quantity and the limit of the variance of admixture for autosomes under random mating. That is,

$$\lim_{g \to \infty} \left(\mathbb{V}\left[H_{1,g,\delta}^A \right] - \mathbb{V}_{RM}\left[H_{1,g,\delta}^A \right] \right) = \frac{c_{11}}{2} + c_{1h}\left(\frac{s_1}{1-h} \right) + \frac{c_{hh}}{2}\left(\frac{s_1}{1-h} \right)^2 \geq 0.$$

We recall that the quantity $\frac{s_1}{1-h}$ is the limit of the expectation of autosomal ancestry, and therefore takes its value in $(0, 1)$ (Verdu and Rosenberg, 2011, eq. 31). We can rewrite the right-hand side of eq. (48) as

$$f^A(c_{11}, c_{1h}, c_{hh}) = \frac{c_{11}}{2} + c_{1h}D_1 + \frac{c_{hh}}{2}D_1^2.$$

where D_1 is a given constant in $(0, 1)$. Under this assortative mating scenario, we can rewrite eq. (6) to give a range for c_{1h} of $(-\frac{c_{11}}{2}, c_{11})$ and, by our definition, we have $c_{11}, c_{hh} > 0$ for positive assortment. Because f is linear in c_{1h}, its minimum in terms of c_{1h} occurs at the boundary of the range of c_{1h}. Substituting for c_{1h} its lower bound, $-\frac{c_{11}}{2}$, we have

$$f^A(c_{11}, c_{1h}, c_{hh}) = \frac{c_{11}}{2} - \frac{c_{11}}{2}D_1 + \frac{c_{hh}}{2}D_1^2 = \frac{c_{11}}{2}(1 - D_1) + \frac{c_{hh}}{2}D_1^2. \tag{48}$$

This quantity is positive because D_1 takes its values in $(0, 1)$, and $c_{11}, c_{hh} > 0$. Therefore, we have demonstrated that this scenario of positive assortative mating increases the variance of autosomal admixture relative to a randomly mating population with the same contribution parameters.

Similarly, for X-chromosomal ancestry in females, we can show that for positive assortative mating, eq. (46) can be written as the sum of a positive quantity and the limit of the variance of the female X-chromosomal admixture under random mating. That is,

$$\lim_{g \to \infty} \left(\mathbb{V}\left[H_{1, g, f}^X\right] - \mathbb{V}_{RM}\left[H_{1, g, f}^X\right] \right) = \frac{c_{11}}{2} + \frac{c_{1h}}{2}\left(P_3 + s_1^f + h^f P_3\right)$$
$$+ \frac{c_{hh}}{2}P_3\left(s_1^f + h^f P_3\right) \geq 0. \tag{49}$$

We recall that the quantities P_3 and $s_1^f + h^f P_3$ are the limits of the expectations of female X-chromosomal, and male X-chromosomal ancestry, respectively, and therefore take their values in $(0, 1)$ (Goldberg and Rosenberg, 2015, Appendix). We can rewrite eq. (49) as

$$f^X(c_{11}, c_{1h}, c_{hh}) = \frac{c_{11}}{2} + \frac{c_{1h}}{2}D_2 + \frac{c_{hh}}{2}D_3,$$

with $D_2 \in (0, 2)$ and $D_3 \in (0, 1)$. These upper bounds arise from the fact that $0 < P_3 < 1$ and $0 < s_1^f + h^f P_3 < 1$ (Goldberg and Rosenberg, 2015, Appendix). As in the cause for autosomal admixture above, f is linear in c_{1h}. Therefore, its minimum in terms of c_{1h} occurs at the boundary of the range of c_{1h}. Substituting for c_{1h} its lower bound, $-\frac{c_{11}}{2}$, we have

$$f^X(c_{11}, c_{1h}, c_{hh}) = \frac{c_{11}}{2} - \frac{c_{11}}{4}D_2 + \frac{c_{hh}}{2}D_3 = \frac{c_{11}}{2}\left(1 - \frac{D_2}{2}\right) + \frac{c_{hh}}{2}D_3 . \tag{50}$$

For this scenario of positive assortative mating, we can see that f is always positive because c_{11}, $c_{hh} > 0$ by definition, and $D_2 \in (0, 2)$ and $D_3 \in (0, 1)$.

6.2.2 Negative assortative mating for each population

A similar argument can be made for negative assortative mating. Under a scenario of intermediate preference for hybrids, we can rewrite eq. (6) to give a range for $c_{1h} \in \left(c_{11}, -\frac{c_{11}}{2}\right)$ and $c_{11}, c_{hh} < 0$ for negative assortment. To show that the maximum value f^A can take is less than zero, we use the upper bound of c_{1h}, which is equal to $-\frac{c_{11}}{2}$, and is a positive value because $c_{11} < 0$. Using this bound gives the same quantity seen in eq. (48); under negative assortative mating, f^A is negative because D takes its values in $(0, 1)$, and $c_{11}, c_{hh} < 0$. Next, considering female X-chromosomal admixture, we can see that f^X is negative for all parameter values under this scenario of negative assortative mating. Substituting the lower bound of $c_{1h} = -\frac{c_{11}}{2}$ into (49), we have the same quantity seen in (50). For negative assortative mating, we have $c_{11}, c_{hh} < 0$, and $D_2 \in (0, 2)$, $D_3 \in (0, 1)$. Therefore, f^X is negative for all parameter values in this negative assortative mating scenario.

In sum, when all populations experience positive assortative mating and the preference for one's own population is the strongest preference, we see that the variances of autosomal and X-chromsomal ancestry can be written as a positive term plus the corresponding variance for a randomly mating population. Similarly, for negative assortment, the variances in equations (45) and (46) can be written as a negative term plus the variance in a corresponding randomly mating population. That is, holding the contributions $\left(s_1^f, s_1^m, s_2^f, s_2^m\right)$ constant, positive assortative mating increases the variance of admixture and negative assortative mating decreases the variance relative to random mating.

6.2.3 Positive mating with respect to S_1

However, we can think of a more general type of assortative mating instead, allowing assortative mating to be positive in one population and random or negative in the other two populations. That is, we consider positive assortative mating with respect to population i as $c_{ii} > c_{ij}, c_{ji}$ for $i \neq j$, and negative assortative mating with respect to population i as $c_{ii} < c_{ij}, c_{ji}$ for $i \neq j$. Notably, these conditions are not mutually exclusive; positive assortative mating might occur in one population at the same time as negative assortative mating in another population. That is, in contrast to the assortative mating scheme above, we allow cases in which $c_{11} > 0$ while $c_{22} < 0$ within the definition of positive assortative mating *with respect to* S_1.

Under this type of positive assortative mating, we allow cases in which $c_{11} > c_{12} > c_{1h}$, and expand the range of c_{1h} to include $c_{1h} \in \left(-2c_{11}, -\frac{c_{11}}{2}\right)$ with $c_{11} > 0$. Therefore, eqs. (48) and (50) are no longer strictly positive, and they depend on the relative values of the assortative mating parameters as well as on the contribution parameters producing the expectation of

admixture. Similarly, for negative assortative mating, $c_{1h} \in \left(-\frac{c_{11}}{2}, -2c_{11}\right)$ with $c_{11}, c_{hh} < 0$, and eqs. (48) and (50) are not strictly negative.

Figure 6 plots an example of positive assortative mating for individuals from S_1 and H, but negative assortative mating in S_2. First, in green, we see an example with positive assortative mating in all populations, with $c_{11} = c_{hh} = 0.02$ and $c_{22} = 0.06$. In this scenario, the variance of admixture is higher than in the corresponding randomly mating population (black dashed line). Maintaining $c_{11} = c_{hh} = 0.02$, but now setting $c_{22} = -0.04$, the purple line plots a case in which positive assortative mating in S_1 is overwhelmed by negative assortative mating in S_2, producing a variance of admixture that is less than in the randomly mating population with the same contribution parameters.

7 Discussion

7.1 Summary of results

A recent series of papers has devised mechanistic models to consider admixture over time (Verdu and Rosenberg, 2011; Goldberg *et al.*, 2014; Goldberg and Rosenberg, 2015), assuming that the parents of a randomly chosen individual from the admixed population are chosen independently. We have extended these models to consider the effect on the distribution of admixture of preferential mating based on population of origin. Although the mean autosomal and X-chromosomal admixture fractions depend only on the sex-specific contributions from the source populations (eqs. (11) and (25)), we have found that the variances of admixture contain signatures of assortative mating (eqs. (21), (31), and (32)).

For both autosomes and X chromosomes, when all three populations experience positive assortative mating, the variance of admixture always increases relative to a randomly mating population with the same contributions (Fig. 5). Similarly, when all three populations experience negative assortative mating, the variance of admixture always decreases. Because an individual's autosomal admixture fraction is the average of its parents' admixture fractions, in the absence of new contributions from the source populations, random pairings of different admixture fractions decrease the variance of admixture every generation (Goldberg and Rosenberg, 2015). However, positive assortative mating decreases the proportion of matings that occur between different admixture fractions, with more low-low and high-high pairings maintaining a wide spread in the distribution of admixture. Similarly, negative assortative mating increases the proportion of matings that occur between pairs with different admixture fractions, thereby decreasing the spread of the admixture fraction distribution.

Assortative mating has a smaller effect on the variance of X-chromosomal admixture than on the variance of autosomal admixture. Conditional on the distribution of ancestry in the previous generation, the variance of X-chromosomal admixture in a male chosen at random from the admixed population is equivalent for randomly mating and assortatively mating populations (eq. (40)) because it is determined by the female X-chromosomal ancestry only, not by the ancestries of both members of a mating pair. Hence, the effect of assortative mating on the variance of X-chromosomal admixture is seen only in females (eq. (39)).

Our model is flexible, allowing assortative mating to occur in different directions and strengths in the three populations. That is, positive assortative mating with respect to S_1 can occur at the same time as random mating or negative assortative mating in H and S_2. We found that the effect of assortative mating on the variance of admixture depends on trade-offs in the preferences among the three populations. For both autosomes and the X chromosome, certain scenarios of assortative mating with different directions of mating preference in different populations change the direction of the effect of assortative mating on the variance of admixture (Fig. 6).

7.2 Comparisons to previous models of assortative mating

Focusing on admixed populations, we chose to model mating preferences by population of origin. This type of assortative mating captures scenarios in which the admixed population is a distinct group, and preferences follow population membership rather than any specific trait, ancestry, or genotype. Different populations might have different geographical locations, trait preferences, or host preferences. For example, scenarios of preference for or against new migrants compared to other admixed individuals have been hypothesized in swordfish, butterflies, and grasshoppers (Ritchie *et al.*, 1989; Howard, 1993; Duenez-Guzman *et al.*, 2009; Melo *et al.*, 2009; Schumer *et al.*, 2017). Similarly, in admixed human populations, language barriers or social identities may produce population-based preferences (Qian, 1997; Jacobson *et al.*, 2004; Ruiz-Linares *et al.*, 2014).

Our model of assortative mating by population recapitulates patterns seen in other assortative mating models in settings that do not consider admixture. Increased variance in a phenotype, with no change in the mean, has been observed in classic models of positive assortative mating by a single-locus genotype or a quantitative trait value without admixture (Jennings 1916, pgs. 66–68; Fisher 1918, pgs. 410–414; Wright 1921, pgs. 153–155; Crow and Kimura 1970, ch. 4). These studies demonstrate that under a model of assortative mating by genotype, positive assortative mating produces an increase in the proportion of homozygotes compared to heterozygotes, analogous to the excess proportion of admixture fractions equaling zero and one seen in our model. For the X chromosome, studies of the distribution of traits and the correlation of relatives under assortative mating by trait loci also find that assortative mating influences X chromosomes in females more than in males (Risch, 1979; Yengo and Visscher, 2018).

Empirical examples of the correlation in ancestry between members of mating pairs in admixed populations motivated Zaitlen *et al.* (2017) to model mating preferences as a fixed correlation in admixture fraction between mates in the admixed population. Extending a Wright-Fisher model, Zaitlen *et al.* (2017) found a similar inflation of the variance of admixture under positive assortative mating and a decrease in the variance of admixture with negative assortment. In their model, mating preferences have the same direction and strength for all individuals; therefore, their model does not allow for the cases we examine in which the variance of admixture decreases under positive assortative mating with respect to some populations and negative assortative mating in others.

7.3 Sex-specific assortative mating

For most of our analysis, we have assumed that mating preferences were symmetric with respect to sex, $c_{ij} = c_{ji}$. However, our general expressions, eqs. (23) and (39), for the autosomal and X-chromosomal variances, respectively, permit this assumption to be relaxed. One observation is that in eq. (23), the autosomal variance depends on the mean of c_{1h} and c_{h1}. In contrast, for the X chromosome in eq. (39), the variance is a function of the more complex $c_{h1}\mathbb{E}\left[H^X_{1, g-1, f}\right]$ and $c_{1h}\mathbb{E}\left[H^X_{1, g-1, m}\right]$. The impact of the sex-specific mating preferences is weighted by the sex-specific contributions in the previous population. The effect of relaxing the symmetry could potentially be investigated using similar numerical studies to those we have used for the symmetric case.

7.4 Sex-biased admixture and assortative mating

We also assumed equal female and male contributions from each of the two source populations, $s^f_{1, g} = s^m_{1, g}$ and $s^f_{2, g} = s^m_{2, g}$, for some of our analyses. However, sex bias and assortative mating may occur in the same population (Bryc *et al.*, 2015; Zou *et al.*, 2015). Our work does allow consideration of different sex-specific contributions from the sources. Goldberg and Rosenberg (2015) examined the effect of sex-biased admixture on the mean X-chromosomal and autosomal admixture in the case of random mating, $c_{ij} = 0$ for all i and j. When estimating sex bias levels from data, because assortative mating does not influence the mean admixture levels, inclusion of assortative mating would not change estimates of sex bias obtained from mean X-chromosomal and autosomal admixture levels.

7.5 Assortative mating can bias inference of the timing of admixture

The variance of admixture is informative about the timing of admixture and has been used to study the timing of admixture in human populations (Verdu and Rosenberg, 2011; Gravel, 2012; Goldberg *et al.*, 2014; Liang and Nielsen, 2014; Verdu *et al.*, 2014; Zaitlen *et al.*, 2017). Because assortative mating often changes the variance of admixture, estimating the timing of admixture without accounting for assortative mating may lead to bias in estimates of the timing of admixture.

For example, considering autosomal admixture under the case of a single admixture event, we can rewrite eq. (41) to consider how assortative mating affects this inference in the case of a single admixture event. For an observed variance of autosomal admixture V^A, we have

$$g = \frac{\ln\left[s^f_{1,0}\left(1 - s^f_{1,0}\right) + s^m_{1,0}\left(1 - s^m_{1,0}\right) + 2c_{11,0}\right] - \ln V^A}{\ln 2} - 1. \tag{51}$$

The estimated timing of admixture, g, is directly related to the assortative mating constant $c_{11,0}$. Therefore, fixing $s^f_{1,0}$, $s^m_{1,0}$, and V^A, positive assortative mating increases g relative to random mating. That is, failing to account for positive assortative mating makes admixture appear more recent than the true admixture event. This observation follows from the increase in variance under positive assortative mating compared to random mating: because the variance decreases over time, a fixed variance persisting due to positive assortative mating might be misinterpreted as a more recent admixture under random mating. Negative

assortative mating has the opposite effect, in that failing to account for negative assortative mating produces an overestimate of g relative to the true level.

Figure 7 plots estimates of g according to eq. (51) for varying levels of $c_{11,0}$, considering both positive and negative assortative mating, specifying $V^A = 0.0625$, and $s_{1,0}^f = s_{1,0}^m = 0.5$ under a single admixture event. For a randomly mating population, an autosomal variance of 0.0625 implies an admixture event 2 generations ago under a single-admixture model with $s_{1,0}^f = s_{1,0}^m = 0.5$. However, when assortative mating is allowed, the admixture event might have taken place in the current generation or up to 3 generations ago, with the estimated g monotonically increasing in $c_{11,0}$. When $c_{11,0}$ 0, the age of admixture is greater than in a randomly mating population with the same variance and contributions. Conversely, when $c_{11,0}$ 0, estimates of g are smaller than in a corresponding randomly mating population.

The variance-inflating effect of positive assortative mating on the admixture level is similar to that described by Zaitlen *et al.* (2017) for the magnitude of ancestry linkage disequilibrium. However, because different scenarios of assortative mating affect the variance of admixture in different directions, under our model, we show that failing to account for assortative mating can either underestimate or overestimate the timing of admixture, depending on whether assortative mating is positive or negative.

Acknowledgments

Appendices

A: Extrema of V[H1,g,δX] for a single admixture event in a randomly mating population

This appendix derives the extrema of $\mathbb{V}\left[H_{1,g,f}^X\right]$ and $\mathbb{V}\left[H_{1,g,m}^X\right]$ as functions of the contribution parameters $S_{1,0}^f$ and $S_{1,0}^m$ for a randomly mating population founded in a single admixture event with no further contributions. The corresponding extrema for the variance of autosomal admixture were previously analyzed by Goldberg *et al.* (2014). Here, we use the model of Goldberg and Rosenberg (2015) to study the variance of X-chromosomal admixture in a randomly mating population ($c_{11,0} = 0$). We perform this computation for completeness, to provide parallel results for the X-chromosomal case to those derived by Goldberg *et al.* (2014) for the autosomes.

Figure A1 plots the variance of X-chromosomal admixture in females using eq. (44) with $c_{11,0} = 0$, for multiple values of g. As g increases, the admixed population mixes with no further contributions from the source populations, and the variance decreases for all values of $S_{1,0}^f$ and $s_{1,0}^m$.

From eq. (44), at $c_{11,0} = 0$, the maximum of $\mathbb{V}\left[H_{1,g,f}^X\right]$ in terms for $s_{1,0}^f$ and $s_{1,0}^m$ occurs when $s_{1,0}^f\left(1 - s_{1,0}^f\right)$ and $s_{1,0}^m\left(1 - s_{1,0}^m\right)$ are separately maximized. That is, the maximum occurs when $s_{1,0}^f = s_{1,0}^m = s_{1,0} = \frac{1}{2}$.

For the minimal variance, which equals 0, equation (44) equals 0 at four points: $\left(s_{1,0}^f, s_{1,0}^m\right) = (0,0), (0,1), (1,0), (1,1)$. The locations of the global maximum and the four minima are the same as those seen for the autosomal variance by Goldberg *et al.* (2014). If only a single population contributes to the hybrid population, $\left(s_{1,0}^f, s_{1,0}^m\right) = (0,0)$ or $(1,1)$, then no variation exists in the population because all individuals have the same admixture fraction of 0 or 1. Similarly, when all males are from one source population and all females are from the other, so that $\left(s_{1,0}^f, s_{1,0}^m\right) = (0,1)$ or $(1,0)$, no variation exists because all individuals have an admixture fraction of $\frac{1}{2}$.

Next, we can find the critical points of the X-chromosomal variance for a given value of $s_{1,0}$, permitting $s_{1,0}^f$ and $s_{1,0}^m$ to vary. Recalling eq. (2), we rewrite the variance of X-chromosomal admixture (eq. (44), with $c_{11,0} = 0$) as a function of $s_{1,0}^f$ and $s_{1,0}$. We have

$$\mathbb{V}\left[H_{1,g,f}^X\right] = \frac{A_g s_{1,0}^f\left(1 - s_{1,0}^f\right) + A_{g-1}\left[\left(2s_{1,0} - s_{1,0}^f\right)\left(1 - 2s_{1,0} + s_{1,0}^f\right)\right]}{4^g} \tag{52}$$

Setting the first derivative of eq. (52) with respect to $s_{1,0}^f$ equal to 0, we find that the critical point of the variance of X-chromosomal admixture occurs at

$$s_{1,0}^f = \frac{A_g + A_{g-1}\left(4s_{1,0} - 1\right)}{2A_g + 2A_{g-1}}. \tag{53}$$

This critical point is a maximum because the second derivative of eq. (52) is negative, $-2(A_g + A_g - 1)$.

Following the same procedure, we can write the variance of X-chromosomal admixture in females as a function of $S_{1,0}^m$ and $s_{1,0}$. We have

$$\mathbb{V}\left[H_{1,g,f}^X\right] = \frac{A_g\left(2s_{1,0} - s_{1,0}^m\right)\left(1 - 2s_{1,0} + s_{1,0}^m\right) + A_{g-1}s_{1,0}^m\left(1 - s_{1,0}^m\right)}{4^g},$$

producing a maximum at

$$s_{1,0}^m = \frac{A_{g-1} + A_g\left(4s_{1,0} - 1\right)}{2A_g + 2A_{g-1}}, \tag{54}$$

We also have the constraints that $s_{1,0}^f$ and $s_{1,0}^m$ take their values in [0, 1]. Eqs. (53) and (54) lie within this interval when both of the following pairs of inequalities hold:

$$\frac{A_{g-1} - A_g}{4A_{g-1}} \leq s_{1,0}^f \leq \frac{A_g + 3A_{g-1}}{4A_{g-1}} \tag{55}$$

$$\frac{A_g - A_{g-1}}{4A_g} \leq s_{1,0}^m \leq \frac{3A_g + A_{g-1}}{4A_g}. \tag{56}$$

Eq. (55) is always satisfied, as $A_g > A_g{-}1$, so that the left hand side of the inequality is negative and the right hand side exceeds 1. Therefore, the maximum always occurs when $S_{1,0}^f$ takes the value in eq. (53). Wheneq. (56) is satisfied, the maximum occurs when $s_{1,0}^m$ follows eq. (54), and when $s_{1,0}^m$ is outside the bounds in eq. (56), the maximum occurs at $s_{1,0}^m$ equal to 0 or 1.

Next, we find the minima of the variance of the female X-chromosomal admixture fraction with respect to $s_{1,0}^f$ and $s_{1,0}^m$ for specified $s_{1,0}$. Eqs. (53) and (54) are quadratic with a single critical point that is a maximum. Therefore, the minima of the variance of X-chromosomal admixture must lie along the boundary of the interval in which $s_{1,0}^f$ and $s_{1,0}^m$ take their values, [0, 1].

For $s_{1,0} \notin \{0, 1\}$, the minimum variance is on the boundary of parameter values, $0 \leq s_{1,0}^f, s_{1,0}^m \leq \min(1, 2s_{1,0})$. Specifically, for $0 \leq s_{1,0} \leq \frac{1}{2}$, the minimum variance occurs when $\left(s_{1,0}^f, s_{1,0}^m\right) = (0, 2s_{1,0})$. Using eq. (44), we see that because $A_g > A_g{-}1$, the variance of the female X-chromosomal admixture fraction is smaller when $\left(s_{1,0}^f, s_{1,0}^m\right) = (0, 2s_{1,0})$ than when $\left(s_{1,0}^f, s_{1,0}^m\right) = (2s_{1,0}, 0)$. Therefore, unlike for the autosomes, the minima are not symmetric for the X chromosome (Figure A1, for $g > 1$). Specifically, because A_g is monotonically increasing, variance depends to a greater extent on $s_{1,0}^f$ than on $S_{1,0}^m$, and the minimum variance for a specific $s_{1,0}$ occurs when $s_{1,0} = 0$, but not $s_{1,0}^m = 0$. Similarly, for $\frac{1}{2} \leq s_{1,0} \leq 1$, the minimum variance occurs when $\left(s_{1,0}^f, s_{1,0}^m\right) = (1, 1 - 2s_{1,0})$. Note that for $s_{1,0} = \frac{1}{2}$, both $\left(s_{1,0}^f, s_{1,0}^m\right) = (1, 0)$ and $\left(s_{1,0}^f, s_{1,0}^m\right) = (0, 1)$ are minima.
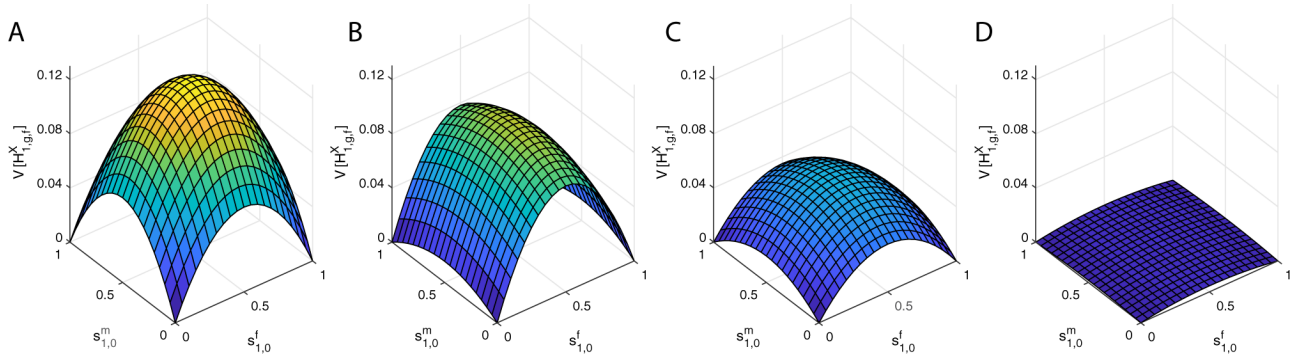
Appendix

Figure A1:
The variance of ancestry as a function of the sex-specific contributions for a single admixture event in a randomly mating population. (A) $g = 1$. (B) $g = 2$. (C) $g = 3$. (D) $g = 8$. In each panel, $\mathbb{V}\left[H^X_{1,g,f}\right]$ is plotted over the range of permissible values of $s^f_{1,0}$ and $S^m_{1,0}$, using eq. (44).

B: Closed-form expression and temporal limit for V[H1,g,δA] under constant contributions over time in an assortatively mating population

In this appendix, we use the closed form of the variance of autosomal admixture for a randomly mating population, derived by Goldberg *et al.* (2014), to write an expression for $\lim_{g \to \infty} \mathbb{V}\left[H^A_{1,g,\delta}\right]$ under assortative mating.

Goldberg *et al.* (2014) derived expressions for $\mathbb{E}\left[H^A_{1,g-1,\delta}\right]$ and $\mathbb{V}_{RM}\left[H^A_{1,g,\delta}\right]$ (eq. 37 and eqs. 51–52, respectively). We substitute these equations in eqs. (22) and (23) to produce a closed form for the variance under assortative mating by population. We have

$$\mathbb{V}\left[H^A_{1,g,\delta}\right] =$$

$$\begin{cases} \dfrac{c_{11,0}}{2} + \mathbb{V}_{RM}\left[H^A_{1,g,\delta}\right], & g = 1 \\[2ex] \dfrac{c_{11}}{2} + \dfrac{c_{1h} + c_{h1}}{2}\left(s_{1,0}h^{g-1} + s_1\dfrac{1-h^{g-1}}{1-h}\right) + \dfrac{c_{hh}}{2}\left(s_{1,0}h^{g-1} + s_1\dfrac{1-h^{g-1}}{1-h}\right)^2 + \mathbb{V}_{RM}\left[H^A_{1,g,\delta}\right], & g \geq 1, \end{cases}$$ (57)

where $\mathbb{V}_{RM}\left[H^A_{1,g,\delta}\right]$ is given by eqs. 51 and 52 from Goldberg *et al.* (2014).

For the limit as $g \to \infty$, we recall eq. 53 from Goldberg *et al.* (2014),

$$\lim_{g \to \infty} \mathbb{V}_{RM}\left[H^A_{1,g,\delta}\right] = \frac{s_1 + s^f_1 s^m_1}{2-h} + \frac{s_1\left(s^f_1 h^m + h^f s^m_1\right)}{(1-h)(2-h)} + \frac{s^2_1 h^f h^m}{(1-h)^2(2-h)} - \left(\frac{s_1}{1-h}\right)^2.$$

Therefore, taking the limit as $g \to \infty$ of the variance of autosomal admixture in eq. (57), we have

$$\lim_{g \to \infty} \mathbb{V}\left[H_{1,g,\delta}^{A}\right] = \frac{c_{11}}{2} + \frac{c_{1h} + c_{h1}}{2}\left(\frac{s_1}{1-h}\right) + \frac{c_{hh}}{2}\left(\frac{s_1}{1-h}\right)^2 + \lim_{g \to \infty} \mathbb{V}_{RM}\left[H_{1,g,\delta}^{A}\right], \quad (58)$$

Finally, assuming assortative mating is the same for the two sexes, $c_{ij} = c_{ji}$, we have eq. (45).

C: Closed form expression and temporal limit for V[H1,g,δX] under constant contributions over time in a randomly mating population

Here, we find an expression for the limit of the variance of X-chromosomal admixture as $g \to \infty$ in a randomly mating population. We follow the structure of Appendix 1 from Goldberg and Rosenberg (2015). Using eqs. 3–4 from Goldberg and Rosenberg (2015) and eqs. (33)–(35) above, we can rewrite the variance of the X-chromosomal admixture fraction in females for the case of constant admixture. For $g = 1$, we have equation (33). For $g = 2$, we have

$$\mathbb{V}_{RM}\left[H_{1,2,f}^{X}\right] = \frac{\left(h^m s_{1,0}^f + s_1^m\right)\left(1 - h^m s_{1,0}^f - s_1^m\right) + s_1^f}{4} + \frac{h^f\left(s_{1,0} + s_{1,0}^f s_{1,0}^m\right)}{8}$$
$$- \left(\frac{s_1^m + h^f s_{1,0}}{2}\right)^2. \tag{59}$$

For $g \geq 3$, we rewrite the female X-chromosomal variance as a second order recursion using equations (35) and (36). We have

$$\mathbb{V}_{RM}\left[H_{1,g,f}^{X}\right] = \frac{s_1^f\left(1 - s_1^f\right)\left(1 + h^m\right) + s_1^m\left(1 - s_1^m\right)}{4} - \frac{s_1^f h^f}{2}\mathbb{E}\left[H_{1,g-1,f}^{X}\right] - \frac{s_1^m h^m}{2}\mathbb{E}\left[H_{1,g-1,m}^{X}\right]$$
$$+ \frac{h^f\left(1 - h^f\right)}{4}\left(\mathbb{E}\left[H_{1,g-1,f}^{X}\right]\right)^2 + \frac{h^m\left(1 - h^m\right)}{4}\left(\mathbb{E}\left[H_{1,g-1,m}^{X}\right]\right)^2$$
$$+ \frac{h^m}{4}\left[-2s_1^f h^f \mathbb{E}\left[H_{1,g-2,f}^{X}\right] + h^f\left(1 - h^f\right)\left(\mathbb{E}\left[H_{1,g-2,f}^{X}\right]\right)^2\right] + \frac{h^f}{4}\mathbb{V}\left[H_{1,g-1,f}^{X}\right] + \frac{h^f h^m}{4}\mathbb{V}\left[H_{1,g-2,f}^{X}\right].$$

We simplify the notation by defining $z_g = \mathbb{V}\left[H_{1,g,f}^{X}\right]$. For $g \geq 3$, we have

$$z_g = d_{1,g} + d_2 z_{g-1} + d_3 z_{g-2}, \tag{60}$$

with

$$d_{1,g} = \frac{s_1^f\left(1 - s_1^f\right)\left(1 + h^m\right) + s_1^m\left(1 - s_1^m\right)}{4} - \frac{s_1^f h^f}{2}\mathbb{E}\left[H_{1,g-1,f}^{X}\right]$$
$$- \frac{s_1^m h^m}{2}\mathbb{E}\left[H_{1,g-1,m}^{X}\right] + \frac{h^f\left(1 - h^f\right)}{4}\left(\mathbb{E}\left[H_{1,g-1,f}^{X}\right]\right)^2$$
$$+ \frac{h^m(1 - h^m)}{4}\left(\mathbb{E}\left[H_{1,g-1,m}^{X}\right]\right)^2 + \frac{h^m}{4}\left[-2s_1^f h^f \mathbb{E}\left[H_{1,g-2,f}^{X}\right]\right.$$
$$\left. + h^f\left(1 - h^f\right)\left(\mathbb{E}\left[H_{1,g-2,f}^{X}\right]\right)^2\right], \tag{61}$$

$$d_2 = \frac{h^f}{4}, \tag{62}$$

$$d_3 = \frac{h^f h^m}{4}, \tag{63}$$

In eq. (61), $d_{1,g}$ depends on g through the expectation of X-chromosomal admixture. However, we can use closed-form solutions for $\mathbb{E}\left[H^X_{1,g,f}\right]$ and $\mathbb{E}\left[H^X_{1,g,m}\right]$ (Goldberg and Rosenberg, 2015, eqs. 17 and 18).

From eqs. (33) and (59), respectively, we have

$$z_1 = \frac{s^f_{1,0}\left(1 - s^f_{1,0}\right) + s^m_{1,0}\left(1 - s^m_{1,0}\right) + 2c_{11,0}}{4}$$

$$z_2 = \frac{\left(h^m s^f_{1,0} + s^m_1\right)\left(1 - h^m s^f_{1,0} - s^m_1\right) + s^f_1}{4} + \frac{h^f\left(s_{1,0} + s^f_{1,0}s^m_{1,0}\right)}{8} - \left(\frac{s^f_1 + h^f s_{1,0}}{2}\right)^2.$$

We define a generating function $Z(x) = \sum_{g=1}^{\infty} z_g x^g$ whose coefficients z_g represent the values of $\mathbb{V}\left[H^X_{1,g,f}\right]$ in each generation. As the admixed population does not yet exist in generation 0, $\mathbb{V}\left[H^X_{1,g,f}\right]$ and $Z(x)$ are undefined for $g = 0$. For convenience, we therefore work with $W(x) = Z(x)/x$, setting $w_g = z_{g+1}$ for g 0. We then have

$$W(x) = \sum_{g=0}^{\infty} z_{g+1}x^g = \sum_{g=0}^{\infty} w_g x^g, \tag{64}$$

and $w_g = d_{1,g} + d_2 w_{g-1} + d_3 w_{g-2}$ for g 2 by eq. (60). Using eq. (64), it follows that

$$\begin{aligned} W(x) &= w_0 + w_1 x + \sum_{g=2}^{\infty}\left(d_{1,g} + d_2 w_{g-1} + d_3 w_{g-2}\right)x^g \\ &= w_0 + w_1 x + d_{1,g}\sum_{g=2}^{\infty} x^g + d_2 x \sum_{g=2}^{\infty} w_{g-1}x^{g-1} + d_3 x^2 \sum_{g=2}^{\infty} w_{g-2}x^{g-2} \\ &= w_0 + w_1 x + d_{1,g}\left(\frac{x^2}{1-x}\right) + d_2 x[W(x) - w_0] + d_3 x^2 W(x). \end{aligned}$$

Solving for $W(x)$, we have

$$W(x) = \frac{(w_0 + w_1 x - d_2 w_0 x)(1 - x) + d_{1,g}x^2}{(1 - x)\left(1 - d_2 x - d_3 x^2\right)}. \tag{65}$$

We can decompose the expression in eq. (65), producing

$$W(x) = \frac{Q_{1,g}}{1 - r_1 x} + \frac{Q_{2,g}}{1 - r_2 x} + \frac{Q_{3,g}}{1 - x},$$

(66)

where r_1 and r_2 are reciprocals of the two roots of $1 - d_2 x - d_3 x^2$,

$$r_1 = \frac{-2 d_3}{d_2 + \sqrt{d_2^2 + 4 d_3}} = \frac{h^f + \sqrt{\left(h^f\right)^2 + 16 h^f h^m}}{8}$$

(67)

$$r_2 = \frac{-2 d_3}{d_2 - \sqrt{d_2^2 + 4 d_3}} = \frac{h^f - \sqrt{\left(h^f\right)^2 + 16 h^f h^m}}{8}.$$

(68)

Setting eq. (65) equal to eq. (66), we have

$$\begin{aligned}
Q_{1,g} &= \frac{(w_0 r_1 + w_1 - d_2 w_0)(r_1 - 1) + d_{1,g}}{(1 - r_1)(r_2 - r_1)} \\
Q_{2,g} &= \frac{(w_0 r_2 + w_1 - d_2 w_0)(r_2 - 1) + d_{1,g}}{(1 - r_2)(r_1 - r_2)} \\
Q_{3,g} &= \frac{d_{1,g}}{(1 - r_1)(1 - r_2)}.
\end{aligned}$$

(69)

The Taylor expansion of eq. (66) around $x = 0$ then gives

$$\begin{aligned}
W(x) &= Q_{1,g} \sum_{g=0}^{\infty} r_1^g x^g + Q_{2,g} \sum_{g=0}^{\infty} r_2^g x^g + Q_{3,g} \sum_{g=0}^{\infty} x^g, \\
&= \sum_{q=0}^{\infty} \left(Q_{1,g} r_1^g + Q_{2,g} r_2^g + Q_{3,g} \right) x^g.
\end{aligned}$$

Therefore, for $g \geq 0$, $w_g = Q_{1,g} r_1^g + Q_{2,g} r_2^g + Q_{3,g}$, and the closed-form expression for the X-chromosomal female mean admixture fraction in generation $g \geq 1$, $\mathbb{V}\left[H_{1,g,f}^X\right]$, is

$$z_g = Q_{1,g} r_1^{g+1} + Q_{2,g} r_2^{g+1} + Q_{3,g}.$$

(70)

Because for $h^f, h^m \in [0, 1]$, r_1 monotonically increases in h^f and h^m and r_2 monotonically decreases, the maxima and minima of r_1 and r_2 over permissible values of h^f and h^m occur at the boundaries of the closed interval $[0, 1]$. Using eqs. (67) and (68), we have $r_1 \in \left[0, \frac{1 + \sqrt{17}}{8}\right]$ and $r_2 \in \left[\frac{1 - \sqrt{17}}{8}, 0\right]$; we exclude $(r_1, r_2) = \left(\frac{1 + \sqrt{17}}{8}, \frac{1 - \sqrt{17}}{8}\right)$, as h^f and h^m cannot both be 1. Because $|r_1|, |r_2| < 1$, the variance of the female X-chromosomal admixture fraction in eq. (70) approaches a limit as $g \to \infty$. Using eqs. (70) and (36), we can find expressions for the limit of the variance of the X-chromosomal admixture fractions,

$$\lim_{g \to \infty} \mathbb{V}_{RM}\left[H^X_{1,g,f}\right] = Q_3 \tag{71}$$

$$\lim_{g \to \infty} \mathbb{V}_{RM}\left[H^X_{1,g,m}\right] = s^f_{1,g-1}\left(1 - s^f_{1,g-1}\right) - 2s^f_{1,g-1}h^f_{g-1}P_3 \\ + h^f_{g-1}\left(1 - h^f_{g-1}\right)P_3^2 + h^f_{g-1}Q_3, \tag{72}$$

where P_3 is the limit of the mean female X-chromosomal admixture fraction (Goldberg and Rosenberg, 2015, eq. 19), and $Q_3 = \lim_{g\to\infty} Q_{3,g}$. Rewriting P_3 here, we have

$$P_3 = \frac{s^f_1 + s^m_1 + s^f_1 h^m}{s^f_1 + s^m_1 + s^f_2 + s^m_2 + h^m\left(s^f_1 + s^f_2\right)}. \tag{73}$$

Q_3 is obtained by substituting $\mathbb{E}\left[H^X_{1,g,f}\right]$ and $\mathbb{E}\left[H^X_{1,g,m}\right]$ with their limits, P_3 and $s^f_1 + h^f P_3$, respectively, into eq. (69). Using eqs. (61)–(63) to simplify, we have

$$Q_3 \\ = \frac{s^f_1\left(1 - s^f_1\right) + s^m_1(1 - s^m_1) + s^f_1 h^m\left(1 - s^f_1 h^m\right) - 2s^f_1 h^f\left(1 + h^m + \left(h^m\right)^2\right)P_3 + \left[h^f\left(1 - h^f\right) + h^f h^m\left(1 - h^f h^m\right)\right]P_3^2}{4(1 - r_1)(1 - r_2)}.$$

References

Bryc K, Durand EY, Macpherson JM, Reich D, and Mountain JL, 2015 The genetic ancestry of African Americans, Latinos, and European Americans across the United States. American Journal of Human Genetics 96: 37–53. [PubMed: 25529636]

Chakraborty R and Weiss KM, 1988 Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proceedings of the National Academy of Sciences 85: 9119–9123.

Crow JF and Kimura M, 1970 An Introduction to Population Genetics Theory. Minneapolis: Burgess.

Duenez-Guzman EA, Mavárez J, Vose MD, and Gavrilets S, 2009 Case studies and mathematical models of ecological speciation. 4. Hybrid speciation in butterflies in a jungle. Evolution 63: 2611–2626. [PubMed: 19545268]

Fisher RA, 1918 The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb pp. 399–433.

Goldberg A and Rosenberg NA, 2015 Beyond 2/3 and 1/3: the complex signatures of sex-biased admixture on the X chromosome. Genetics 201: 263–279. [PubMed: 26209245]

Goldberg A, Verdu P, and Rosenberg NA, 2014 Autosomal admixture levels are informative about sex bias in admixed populations. Genetics 198: 1209–1229. [PubMed: 25194159]

Gravel S, 2012 Population genetics models of local ancestry. Genetics 191: 607–619. [PubMed: 22491189]

Howard DJ, 1993 Reinforcement: origin, dynamics, and fate of an evolutionary hypothesis In: Hybrid Zones and the Evolutionary Process, Edited by Harrison R. Oxford University Press, New York.

Jacobson CK, Amoateng AY, and Heaton TB, 2004 Inter-racial marriages in South Africa. Journal of Comparative Family Studies 35: 443–458.

Jennings HS, 1916 The numerical results of diverse systems of breeding. Genetics 1: 53–89. [PubMed: 17245851]

Liang M and Nielsen R, 2014 The lengths of admixture tracts. Genetics 197: 953–967. [PubMed: 24770332]

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

Lindtke D and Buerkle CA, 2015 The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. Evolution 69: 1987–2004. [PubMed: 26174368]

Loh P-R, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, and Berger B, 2013 Inferring admixture histories of human populations using linkage disequilibrium. Genetics 193: 1233–1254. [PubMed: 23410830]

Long JC, 1991 The genetic structure of admixed populations. Genetics 127: 417–428. [PubMed: 2004712]

Melo MC, Salazar C, Jiggins CD, and Linares M, 2009 Assortative mating preferences among hybrids offers a route to hybrid speciation. Evolution 63: 1660–1665. [PubMed: 19492995]

Ni X, Yang X, Guo W, Yuan K, Zhou Y, Ma Z, and Xu S, 2016 Length distribution of ancestral tracks under a general admixture model and its applications in population history inference. Scientific Reports 6: 20048. [PubMed: 26818889]

Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, and Pena SDJ, 2003 Color and genomic ancestry in brazilians. Proceedings of the National Academy of Sciences 100: 177–182.

Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E, and Shriver MD, 2001 Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. American Journal of Human Genetics 68: 198–207. [PubMed: 11112661]

Pool JE and Nielsen R, 2009 Inference of historical changes in migration rate from the lengths of migrant tracts. Genetics 181: 711–719. [PubMed: 19087958]

Qian Z, 1997 Breaking the racial barriers: Variations in interracial marriage between 1980 and 1990. Demography 34: 263–276. [PubMed: 9169282]

Risch H, 1979 The correlation between relatives under assortative malting for an X-linked and autosomal trait. Annals of Human Genetics 43: 151–165. [PubMed: 525974]

Risch N, Choudhry S, Via M, Basu A, Sebro R, Eng C, Beckman K, Thyne S, Chapela R, Rodriguez-Santana JR, et al., 2009 Ancestry-related assortative mating in Latino populations. Genome Biology 10: R132. [PubMed: 19930545]

Ritchie MG, Butlin RK, and Hewitt GM, 1989 Assortative mating across a hybrid zone in Chorthippus parallelus (Orthoptera: Acrididae). Journal of Evolutionary Biology 2: 339–352.

Ruiz-Linares A, Adhikari K, Acuña-Alonzo V, Quinto-Sanchez M, Jaramillo C, Arias W, Fuentes M, Pizarro M, Everardo P, de Avila F, et al., 2014 Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. PLoS Genetics 10: e1004572. [PubMed: 25254375]

Schumer M and Brandvain Y, 2016 Determining epistatic selection in admixed populations. Molecular Ecology 25: 2577–2591. [PubMed: 27061282]

Schumer M, Cui R, Rosenthal GG, and Andolfatto P, 2015 Reproductive isolation of hybrid populations driven by genetic incompatibilities. PLoS Genetics 11: e1005041. [PubMed: 25768654]

Schumer M, Powell DL, Delclós PJ, Squire M, Cui R, Andolfatto P, and Rosenthal GG, 2017 Assortative mating and persistent reproductive isolation in hybrids. Proceedings of the National Academy of Sciences 114: 10936–10941.

Sebro R, Hoffman TJ, Lange C, Rogus JJ, and Risch NJ, 2010 Testing for non-random mating: evidence for ancestry-related assortative mating in the Framingham heart study. Genetic Epidemiology 34: 674–679. [PubMed: 20842694]

Sedghifar A, Brandvain Y, and Ralph P, 2016 Beyond clines: lineages and haplotype blocks in hybrid zones. Molecular Ecology 25: 2559–2576. [PubMed: 27148805]

Sedghifar A, Brandvain Y, Ralph P, and Coop G, 2015 The spatial mixing of genomes in secondary contact zones. Genetics 201: 243–261. [PubMed: 26205988]

Verdu P, Pemberton TJ, Laurent R, Kemp BM, Gonzalez-Oliver A, Gorodezky C, Hughes CE, Shattuck MR, Petzelt B, Mitchell J, et al., 2014 Patterns of admixture and population structure in native populations of northwest north america. PLoS Genetics 10: e1004530. [PubMed: 25122539]

Verdu P and Rosenberg NA, 2011 A general mechanistic model for admixture histories of hybrid populations. Genetics 189: 1413–1426. [PubMed: 21968194]

Wang L, Luzynski K, Pool JE, Janoušek V, Dufková P, Vysko ilová MM, Teeter KC, Nachman MW, Munclinger P, Macholán M, et al., 2011 Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. Molecular Ecology 20: 2985–3000. [PubMed: 21668551]

Wright S, 1921 Assortative mating based on somatic resemblance. Genetics 6: 144–161. [PubMed: 17245960]

Yengo L and Visscher PM, 2018 Assortative mating on complex traits revisited: Double first cousins and the X-chromosome. Theoretical Population Biology 124: 51–60. [PubMed: 30316741]

Zaitlen N, Huntsman S, Hu D, Spear M, Eng C, Oh SS, White MJ, Mak A, Davis A, Meade K, et al., 2017 The effects of migration and assortative mating on admixture linkage disequilibrium. Genetics 205: 375–383. [PubMed: 27879348]

Zou JY, Park DS, Burchard EG, Torgerson DG, Pino-Yanes M, Song YS, Sankararaman S, Halperin E, and Zaitlen N, 2015 Genetic and socioeconomic study of mate choice in Latinos reveals novel assortment patterns. Proceedings of the National Academy of Sciences 112: 13621–13626.
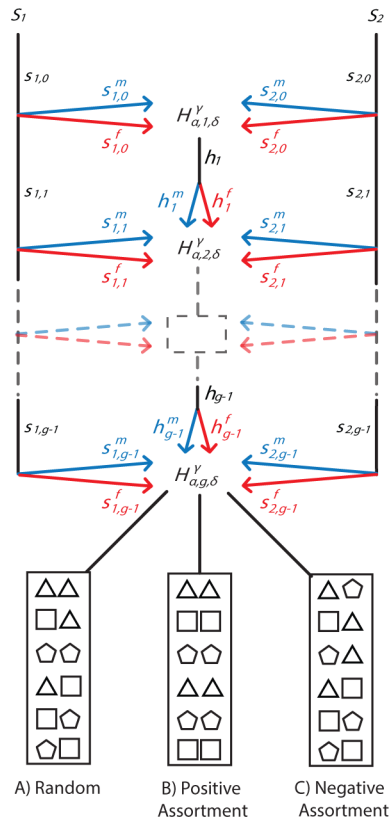
Figure 1:

Schematic of the mechanistic model of assortative mating by population. Two source populations, S_1 and S_2, contribute females and males to the next generation of the hybrid population H, potentially with time-varying proportions. The fractional contributions of the source populations and the hybrid population to the next generation g are $s_{1,g}$, $s_{2,g}$, and h_g, respectively. Sex-specific contributions from the populations are $s^f_{1,g}$, $s^f_{2,q}$ and h^f_g, and $S^m_{1,g}$, $S^m_{2,g}$ and h^m_g, for females and males, respectively. $H^\gamma_{\alpha,g,\delta}$ represents the fraction of admixture from source population $\alpha \in \{1, 2\}$ in generation g for a random individual of sex $\delta \in \{f, m\}$ in population H for chromosomal type $\gamma \in \{A, X\}$. Within the admixed population, at every generation, parents from generation $g-1$ pair according to one of three mating models. Individuals from S_1 are represented by triangles, S_2 by pentagons, and H by squares. (A) Random mating. The probability of a pairing is given by the product of the proportional contributions of the two populations. (B) Positive assortative mating. Individuals are more likely to mate with individuals from their own population. (C) Negative assortative mating. Individuals are more likely to mate with individuals from a different population. In each panel, a mating pair is indicated by a pair of adjacent symbols. Each panel considers the same values for the contributions from the three populations to generation $g+1$.
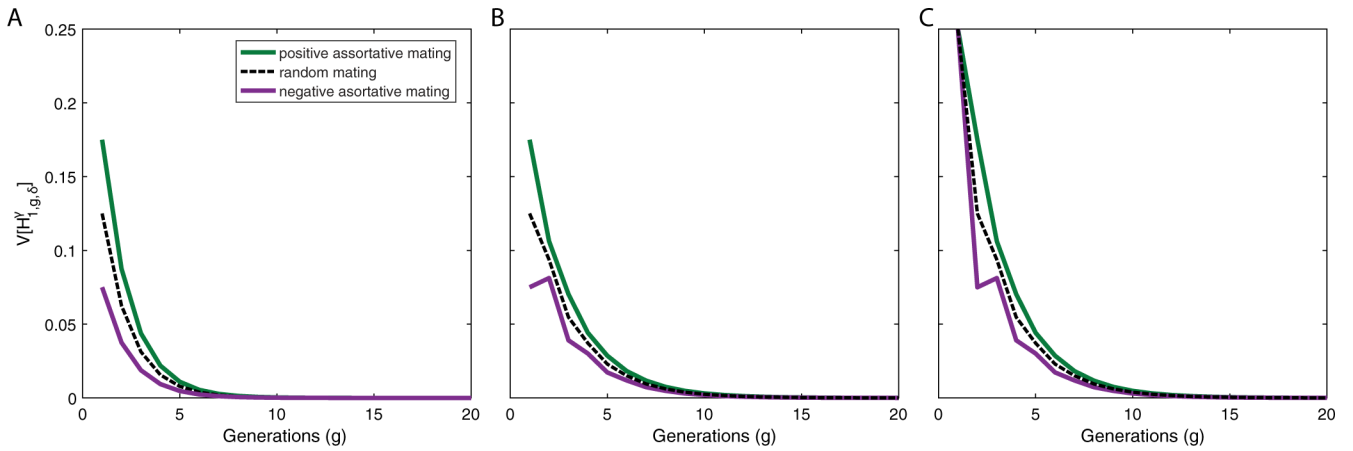
Figure 2:

Variance of ancestry under a single-admixture scenario with assortative mating in the founding generation. (A) Autosomes, $\mathbb{V}\left[H^A_{1,g,\delta}\right]$. (B) X chromosomes in a female, $\mathbb{V}\left[H^X_{1,g,f}\right]$. (C) X chromosomes in a male, $\mathbb{V}\left[H^X_{1,g,m}\right]$. Each panel considers positive assortative mating (green), $c_{11,0} = 0.1$, negative assortative mating (purple), $c_{11,0} = -0.1$, and random mating (black), $c_{11,0} = 0$. The plots use eqs. (41), (43), and (44) with $s^f_{1,0} = s^m_{1,0} = 0.5$. For positive assortment, the variance is higher than for random mating. For negative assortative mating, the variance is lower.
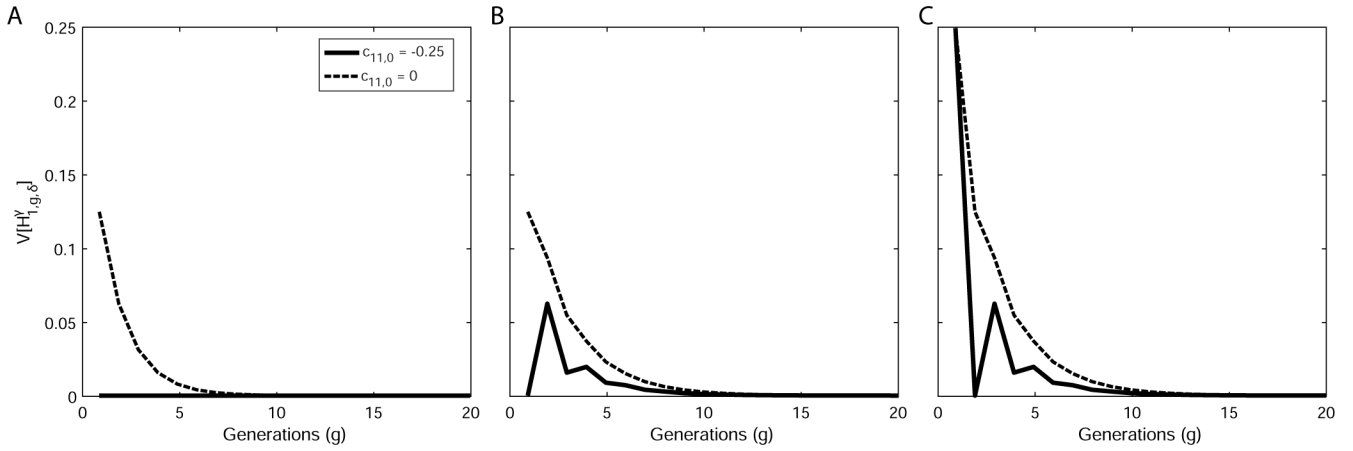
Figure 3:
Special case of the variance of ancestry under a single-admixture scenario with negative assortative mating in the founding generation. (A) Autosomes $\mathbb{V}\left[H^A_{1,g,\delta}\right]$. (B) X chromosomes in a female, $\mathbb{V}\left[H^X_{1,g,f}\right]$. (C) X chromosomes in a male, $\mathbb{V}\left[H^X_{1,g,m}\right]$. In each panel, the variance is plotted for negative assortative mating, $c_{11,0} = -0.25$, using eqs. (41), (43), and (44) with $s^f_{1,0} = s^m_{1,0} = 0.5$. The plot highlights a special case in which the variance of autosomal admixture is zero because the numerator of eq. (41) is zero, even though $s^f_{1,0} = s^m_{1,0} = 0.5$ is a maximum of the variance with respect to the sex-specific contributions in a randomly mating population. The $c_{11,0} = 0$ case is copied from Figure 2.
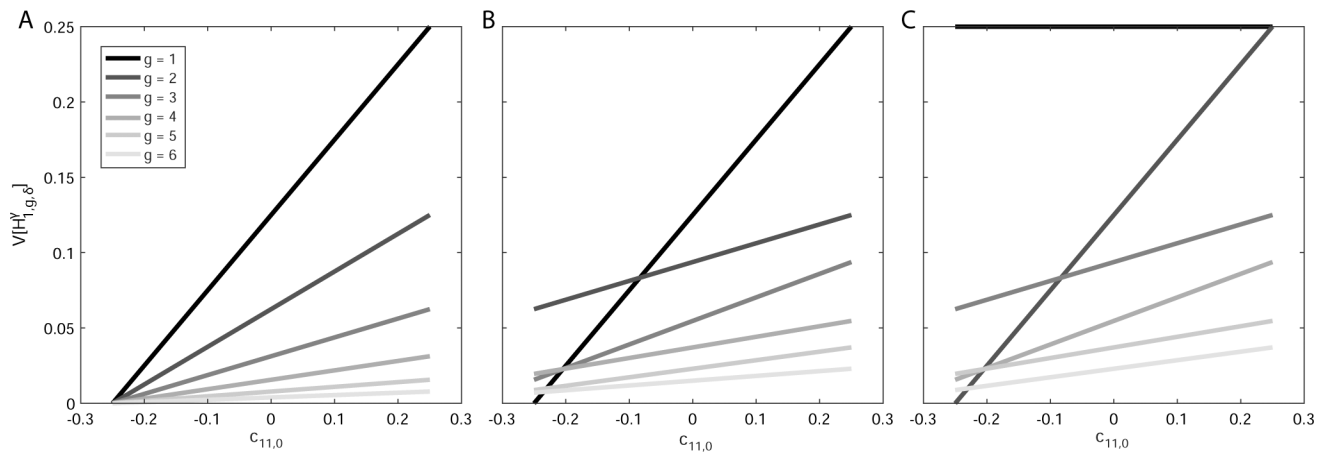
Figure 4:
Variance of ancestry under a single-admixture scenario as a function of the assortative mating parameter $c_{11,0}$. (A) Autosomes, $\mathbb{V}\left[H^A_{1,g,\delta}\right]$. (B) X chromosomes in a female, $\mathbb{V}\left[H^X_{1,g,f}\right]$. (C) X chromosomes in a male, $\mathbb{V}\left[H^X_{1,g,m}\right]$. In each panel, the variance is plotted over the range of $c_{11,0}$, $[-0.25, 0.25]$, for values of g between 1 and 6 (from darkest to lightest gray), with $s^f_{1,0} = s^m_{1,0} = 0.5$, using eqs. (41), (43), and (44).
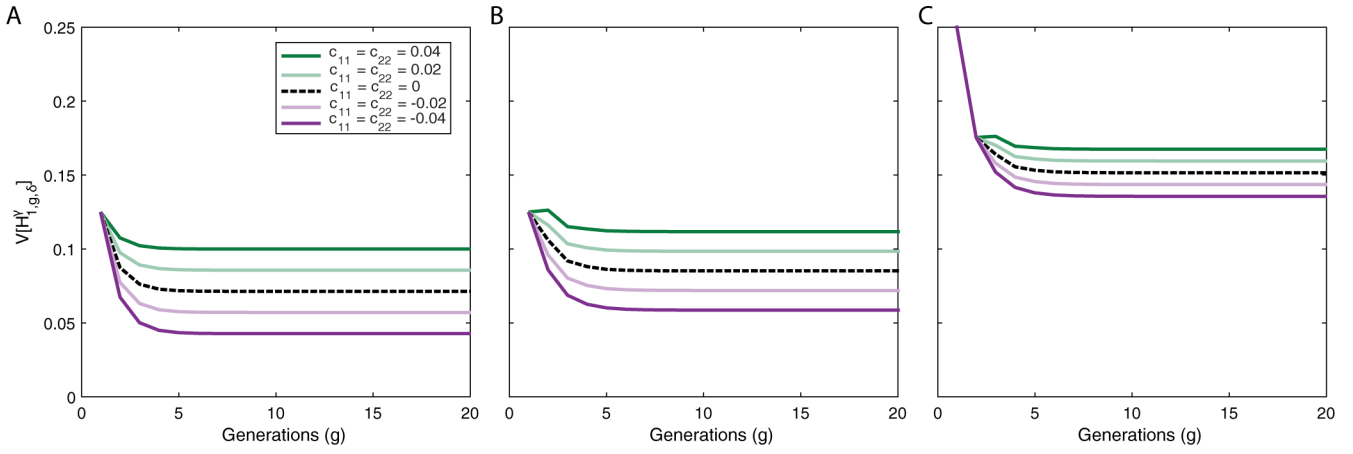
Figure 5:
Variance of ancestry under a constant-admixture scenario with assortative mating in producing $g \geq 2$. (A) Autosomes, $\mathbb{V}\left[H_{1,g,\delta}^{A}\right]$. (B) X chromosomes in a female, $\mathbb{V}\left[H_{1,g,f}^{X}\right]$. (C) X chromosomes in a male, $\mathbb{V}\left[H_{1,g,m}^{X}\right]$. Each panel considers positive assortative mating (greens), negative assortative mating (purples), and random mating (black dashed). The plots use eqs. (20), (21), and (29)–(32). We fix $c_{11} = c_{22}$ and $c_{1h} = 0$, with $s_1^f = s_1^m = s_2^f = s_2^m = 0.2$. For initial conditions, $s_{1,0}^f = s_{1,0}^m = s_{2,0}^f = s_{2,0}^m = 0.5$ and $c_{11,0} = 0$. Positive assortative mating increases the variance relative to a randomly mating population, whereas negative assortative mating decreases it.
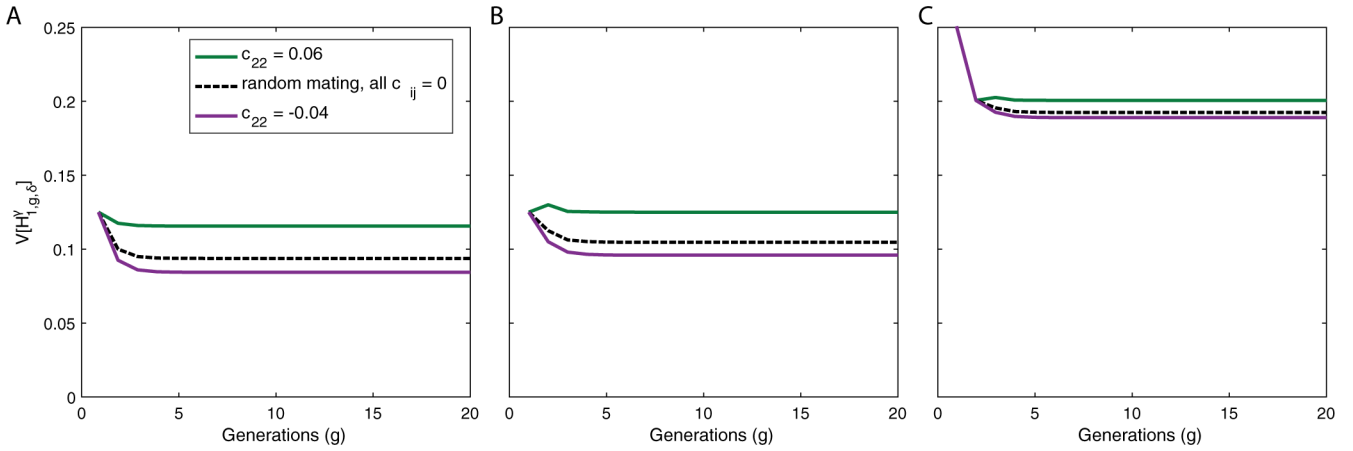
Figure 6:

Example of the variance of ancestry under a constant-admixture scenario with assortative mating in producing $g \geq 2$ when populations are allowed to differ in the direction of their mating preference. (A) Autosomes, $\mathbb{V}\left[H^A_{1,g,\delta}\right]$. (B) X chromosomes in a female, $\mathbb{V}\left[H^X_{1,g,f}\right]$. (C) X chromosomes in a male, $\mathbb{V}\left[H^X_{1,g,m}\right]$. In all scenarios, populations S_1 and H experience positive assortative mating, $c_{11}, c_{hh} = 0.02$. Green curves show a scenario in which S_2 also experiences positive assortative mating, $c_{22} = 0.06$; purple curves show a scenario in which S_2 experiences negative assortative mating, $c_{22} = -0.04$. The scenario of random mating with the same contribution parameters is the black dashed line. The plots use eqs. (20), (21), and (29)–(32). We set $s^f_1 = s^m_1 = s^f_2 = s^m_2 = 0.2$, and use initial conditions $s^f_{1,0} = s^m_{1,0} = s^f_{2,0} = s^m_{2,0} = 0.5$ and $c_{11,0} = 0$. In this example, negative assortative mating in S_2 offsets the positive assortative mating in S_1 and H, producing lower variances of admixture in S_1, $\mathbb{V}\left[H^\gamma_{1,g,\delta}\right]$, in the assortatively mating population than in the randomly mating population.
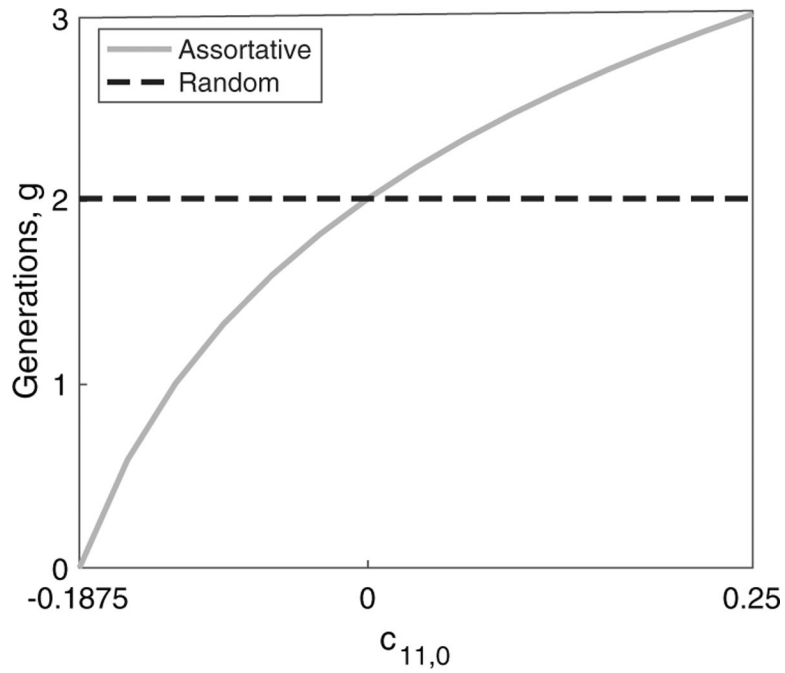
Figure 7:
Timing of admixture under assortative mating. The number of generations since admixture, g, for a single admixture event is plotted as a function of $c_{11,0}$ using the variance in autosomal admixture, eq. (51), with $V^A = 0.0625$ and $s_{1,0}^s = s_{1,0}^m = 0.5$. The plot traverses the range of possible $c_{11,0}$ values for $g \geq 0$. Recalling that the values of the $c_{ij,0}$ are bounded such that the probability of each given parental pairing takes its values in the interval [0, 1], and such that each probability is no greater than the probability of one of its constituent components, for $c_{11,0} > 0$, we have $c_{11,0} \leq \min\left(s_{1,0}^f, s_{1,0}^m\right) - s_{1,0}^f s_{1,0}^m$. In this case, $c_{11,0} \leq 0.25$. For $c_{11,0} < 0$, we have $c_{11,0} \geq s_{1,0}^f s_{1,0}^m$. In this case, we have $c_{11,0} \geq -0.25$. However, g cannot be negative; therefore, in this case, we truncate the domain at the value of $c_{11,0}$ that produces $g = 0$. The value of g for a randomly mating population with the same contributions and variance is shown in the dashed line.

Table 1:

Recursions for autosomal and X-chromosomal admixture. The table shows the probability of parental pairings and the admixture fractions for each of the nine possible pairings for a randomly chosen female and a randomly chosen male from the admixed population at generation g.

Case	Female parent's population	Male parent's population	Probability	Admixture in males		Admixture in females	
				$H^A_{1,g,m}$	$H^X_{1,g,m}$	$H^A_{1,g,f}$	$H^X_{1,g,f}$
1	S_1	S_1	$s^f_{1,g-1}s^m_{1,g-1}+c_{11,g}$	1	1	1	1
2	S_1	H	$s^f_{1,g-1}h^m_{g-1}+c_{1h,g}$	$\dfrac{1+H^A_{1,g-1,m}}{2}$	1	$\dfrac{1+H^A_{1,g-1,m}}{2}$	$\dfrac{1+H^X_{1,g-}}{2}$
3	S_1	S_2	$s^f_{1,g-1}s^m_{2,g-1}+c_{12,g}$	½	1	½	½
4	H	S_1	$h^f_{g-1}s^m_{1,g-1}+c_{h1,g}$	$\dfrac{1+H^A_{1,g-1,f}}{2}$	$H^X_{1,g-1,f}$	$\dfrac{1+H^A_{1,g-1,f}}{2}$	$\dfrac{1+H^X_{1,g-}}{2}$
5	H	H	$h^f_{g-1}h^m_{g-1}+c_{hh,g}$	$\dfrac{H^A_{1,g-1,f}+H^A_{1,g-1,m}}{2}$	$H^X_{1,g-1,f}$	$\dfrac{H^A_{1,g-1,f}+H^A_{1,g-1,m}}{2}$	$\dfrac{H^X_{1,g-1,f}+H^X_{1}}{2}$
6	H	S_2	$h^f_{g-1}s^m_{2,g-1}+c_{h2,g}$	$\dfrac{H^A_{1,9-1,f}}{2}$	$H^X_{1,g-1,f}$	$\dfrac{H^A_{1,g-1,f}}{2}$	$\dfrac{H^X_{1,g-1,}}{2}$
7	S_2	S_1	$s^f_{2,g-1}s^m_{1,g-1}+c_{21,g}$	½	0	½	½
8	S_2	H	$s^f_{2,g-1}h^m_{g-1}+c_{2h,g}$	$\dfrac{H^A_{1,g-1,m}}{2}$	0	$\dfrac{H^A_{1,g-1,m}}{2}$	$\dfrac{H^X_{1,g-1,}}{2}$
9	S_2	S_2	$s^f_{2,g-1}s^m_{2,g-1}+c_{22,g}$	0	0	0	0