Autosomal admixture levels are informative about sex bias in admixed populations

Amy Goldberg¹,*
Paul Verdu²
Noah A Rosenberg¹

¹ Department of Biology, Stanford University, Stanford, CA, 94305-5020 USA
² CNRS-MNHN-Université Paris Diderot, UMR7206 Ecoanthropology and Ethnobiology, Paris, France

June 23, 2014

Abstract. Sex-biased admixture has been observed in a wide variety of admixed populations. Genetic variation in sex chromosomes and ratios of quantities computed from sex chromosomes and autosomes have often been examined in order to infer patterns of sex-biased admixture, typically using statistical approaches that do not mechanistically model the complexity of a sex-specific history of admixture. Here, expanding on a model of Verdu & Rosenberg (2011) that did not include sex specificity, we develop a model that mechanistically examines sex-specific admixture histories. Under the model, multiple source populations contribute to an admixed population, potentially with their male and female contributions varying over time. In an admixed population descended from two source groups, we derive the moments of the distribution of the autosomal admixture fraction from a specific source population as a function of sex-specific introgression parameters and time. Considering admixture processes that are constant in time, we demonstrate that surprisingly, although the mean autosomal admixture fraction from a specific source population does not reveal a sex bias in the admixture history, the variance of autosomal admixture is informative about sex bias. Specifically, the long-term variance decreases as the sex bias from a contributing source population increases. This result can be viewed as analogous to the reduction in effective population size for populations with an unequal number of breeding males and females. Our approach can contribute to methods for inference of the history of complex sex-biased admixture processes by enabling consideration of the effect of sex-biased admixture on autosomal DNA.

*Corresponding author. Phone: 650-724-5122. Fax: 650-724-5114. Email: agoldb@stanford.edu
Introduction

Populations often experience sex-biased demographic processes, in which males and females contributing to the gene pool of a population are drawn from source groups in different proportions, owing to patterns of inbreeding avoidance, dispersal, and mating practices (Pusey 1987; Lawson Handley & Perrin 2007). In humans, sex-biased demography has had a particular effect on admixed populations, populations that have often been founded or influenced by periods of colonization and forced migration involving an initial or continuing admixture process (Mesa et al. 2000; Seielstad 2000; Wilkins & Marlowe 2006; Tremblay & Vezina 2010; Heyer et al. 2012).

Genetic signatures of sex-biased admixture have been empirically investigated in a variety of human populations. In the Americas, these include African American, Latino, and Native American populations (Bolnick et al. 2006; Wang et al. 2008; Stefflova et al. 2009; Tishkoff et al. 2009; Bryc et al. 2010a,b; Moreno-Estrada et al. 2013; Verdu et al. 2014). Sex-biased admixture and migration have also been examined in populations throughout Asia (Oota et al. 2001; Wen et al. 2004; Chaix et al. 2007; Ségurel et al. 2008; Chaubey et al. 2011; Pemberton et al. 2012; Pijpe et al. 2013), Austronesia (Kayser et al. 2003, 2006, 2008; Cox et al. 2010; Lansing et al. 2011) and Africa (Wood et al. 2005; Tishkoff et al. 2007; Berniell-Lee et al. 2008; Beleza et al. 2013; Petersen et al. 2013; Verdu et al. 2013).

Sex-specific admixture and migration processes have typically been studied using comparisons of the Y chromosome, which is paternally inherited, and the mitochondrial genome, inherited maternally (Seielstad et al. 1998; Oota et al. 2001; Wood et al. 2005; Bolnick et al. 2006; Gunnarsdóttir et al. 2011; Lacan et al. 2011). More recently, as the Y chromosome and mitochondrial genome each represent single non-recombining loci that provide an incomplete genomic perspective, sex-biased admixture has been examined by comparisons of autosomal DNA to the X chromosome (Lind et al. 2007; Wang et al. 2008; Bryc et al. 2010a,b; Cox et al. 2010; Beleza et al. 2013; Verdu et al. 2013).

The Y–mitochondrial and X–autosomal frameworks are both sensible, as both involve comparisons of two types of loci that follow different modes of inheritance in males and females. What has not been clear, however, is that autosomal data, which have not typically been viewed as the most informative loci for studies of sex-specific processes, can carry information about sex-biased admixture, even in the absence of a comparison with other components of the genome.

We demonstrate this surprising result through an extension of a mechanistic model for the admixture history of a hybrid population. In a diploid autosomal framework, Verdu & Rosenberg (2011) examined contributions
of multiple source populations that varied through time, without considering sex specificity. Here, expanding on the model of Verdu & Rosenberg (2011), we develop a model that mechanistically considers sex-specific admixture histories in which multiple source populations contribute to the admixed population, potentially with varying female and male contributions across generations (Fig. 1). In an admixed population descended from two source populations, we derive the moments of the distribution of the fraction of autosomal admixture from a specific source population, as a function of sex-specific admixture parameters and time. We analyze the behavior of the model, considering admixture processes that are constant in time, and we show that the moments contain information about the sex bias.

The model

Several studies have described mechanistic models of admixture (Chakraborty & Weiss 1988; Long 1991; Ewens & Spielman 1995; Guo et al. 2005; Verdu & Rosenberg 2011; Gravel 2012; Jin et al. 2013). We follow the notation and style of the model of Verdu & Rosenberg (2011), studying a hybrid population, $H$, which consists of immigrant individuals from $M$ isolated source populations and hybrid individuals who have ancestors from two or more source populations. The source populations are labeled $S_{\alpha}$, for $\alpha$ from 1 to $M$. We focus on the case of $M = 2$.

We define the parameters $s_{\alpha,g-1}$ and $h_{g-1}$ as the contributions from source populations $S_{\alpha}$ and $H$, respectively, to the gene pool of the hybrid population $H$ at the next generation, $g$. That is, for a randomly chosen individual at generation $g$, the probabilities that a randomly chosen parent of the individual derives from $S_{\alpha}$ and $H$ are $s_{\alpha,g-1}$ and $h_{g-1}$, respectively. We define the sex-specific parameter $s_{\alpha,g-1}^\delta$, for $\delta \in \{f, m\}$, as the probability that the type-$\delta$ parent of a randomly chosen individual from the hybrid population at generation $g$ is from source population $S_{\alpha}$. Similarly, $h_{g-1}^\delta$ is the probability that the type-$\delta$ parent of a randomly chosen individual in $H$ at generation $g$ is from $H$ itself. We consider a two-sex model, using $f$ for female and $m$ for male. Thus, because each individual has one parent of each type, female and male, we have

$$s_{\alpha,g-1} = \frac{(s_{\alpha,g-1}^f + s_{\alpha,g-1}^m)}{2}$$

$$h_{g-1} = \frac{(h_{g-1}^f + h_{g-1}^m)}{2}.$$  

(1) 

(2)

The contributions to the next generation of the three source populations ($S_1, S_2, H$) sum to one:

$$s_{1,g-1} + s_{2,g-1} + h_{g-1} = 1.$$  

(3)
Similarly, the female and male contributions to the next generation separately sum to one,

\[ s_{1,g-1}^f + s_{2,g-1}^f + h_{g-1}^f = s_{1,g-1}^m + s_{2,g-1}^m + h_{g-1}^m = 1. \]  

(4)

At the first generation, \( g = 1 \), the hybrid population has not previously existed; therefore,

\[ h_0 = h_0^f = h_0^m = 0 \]  

(5)

\[ s_{1,0} + s_{2,0} = s_{1,0}^f + s_{2,0}^f = s_{1,0}^m + s_{2,0}^m = 1. \]  

(6)

The first generation has two independent parameters, \( s_{1,0}^f \) and \( s_{1,0}^m \). Each subsequent generation contributes four additional independent parameters \( (s_{1,g-1}^f, s_{2,g-1}^f, s_{1,g-1}^m, s_{2,g-1}^m) \), and considering the first \( g \) generations, there are \( 4g - 2 \) independent parameters. The model is discrete in time and assumes non-overlapping generations.

Our model allows us to consider complex sex-biased admixture processes by allowing uneven sex-specific contributions from each source population at each generation. It reduces to the model of Verdu & Rosenberg (2011) when the sex-specific contributions are equal within a source population, that is, if for each \( g \), \( s_{1,g-1}^f = s_{1,g-1}^m \) and \( s_{2,g-1}^f = s_{2,g-1}^m \). We perform similar computations to those of Verdu & Rosenberg (2011), illustrating that in certain cases, our results reduce to those obtained when sex specificity is not considered.

We let \( L \) be a random variable indicating the source populations of the parents of a random individual from the hybrid population, \( H \). \( L \) takes its values from the set of all possible ordered parental combinations, \( \{S_1S_1, S_1S_2, S_2S_1, S_2S_2, H, HS_1, HS_2, S_1H, S_2H, S_1S_2\} \), listing the female parent first. We assume random mating in the hybrid population at each generation, so that the probability that an offspring has a particular pair of source populations for his or her parents is simply the product of the probabilities for having the female and male parents (Table 1).

We define the fraction of admixture, the random variable \( H_{\alpha,g,\delta} \), as the probability that an autosomal genetic locus in a random individual of sex \( \delta \) from the hybrid population in generation \( g \) ultimately originates from source population \( \alpha \). The sex-specific fractions of admixture are related to the total fraction of admixture \( H_{\alpha,g} \) from source population \( \alpha \) in generation \( g \) by \( H_{\alpha,g} = (H_{\alpha,g,f} + H_{\alpha,g,m})/2 \).

Under the model, we derive expressions for the moments of the fraction of admixture. Autosomal DNA is inherited non-sex-specifically and from both parents; therefore, female and male offspring have identical distributions of admixture, and \( H_{\alpha,g,f} \) and \( H_{\alpha,g,m} \) are identically distributed. Each of these quantities depends on both the female fraction and the male fraction of admixture in the previous generation, but
conditional on the previous generation (that is, on $H_{s,g-1,f}$ and $H_{s,g-1,m}$), they are independent. For our two-population model, we consider the non-sex-specific fraction of admixture, $H_{1,g,\delta}$, treating $\delta$ here as representing either $f$ or $m$, (but retaining the same meaning throughout). The quantity $H_{1,g,\delta}$ depends on both sex-specific fractions of admixture from the previous generation, $H_{1,g-1,f}$ and $H_{1,g-1,m}$.

**Distribution of the admixture fraction from a specific source**

The definition of the model parameters and the values from Table 1 allow us to write a recursion relation for the fraction of admixture from source population 1 for a random individual of sex $\delta$ from the hybrid population at generation $g$, or $H_{1,g,\delta}$. For the first generation, $g = 1$, we have

$$H_{1,1,\delta} = \begin{cases} 1 & \text{if } L = S_1S_1, \text{with } P[L = S_1S_1] = s_{1,0}^f s_{1,0}^m \\ \frac{1}{2} & \text{if } L = S_1S_2, \text{with } P[L = S_1S_2] = s_{1,0}^f s_{2,0}^m \\ \frac{1}{2} & \text{if } L = S_2S_1, \text{with } P[L = S_2S_1] = s_{2,0}^f s_{1,0}^m \\ 0 & \text{if } L = S_2S_2, \text{with } P[L = S_2S_2] = s_{2,0}^f s_{2,0}^m. \end{cases}$$  \hspace{1cm} (7)

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

$$H_{1,g,\delta} = \begin{cases} 1 & \text{if } L = S_1S_1, \text{with } P[L = S_1S_1] = s_{1,g-1}^f s_{1,g-1}^m \\ \frac{1 + H_{1,g-1,m}}{2} & \text{if } L = S_1H, \text{with } P[L = S_1H] = s_{1,g-1}^f h_{g-1}^m \\ \frac{1}{2} & \text{if } L = S_1S_2, \text{with } P[L = S_1S_2] = s_{1,g-1}^f s_{2,g-1}^m \\ \frac{1 + H_{1,g-1,f}}{2} & \text{if } L = HS_1, \text{with } P[L = HS_1] = h_{g-1}^f s_{1,g-1}^m \\ \frac{H_{1,g-1,f} + H_{1,g-1,m}}{2} & \text{if } L = HH, \text{with } P[L = HH] = h_{g-1}^f h_{g-1}^m \\ \frac{H_{1,g-1,f}}{2} & \text{if } L = HS_2, \text{with } P[L = HS_2] = h_{g-1}^f s_{2,g-1}^m \\ \frac{1}{2} & \text{if } L = S_2S_1, \text{with } P[L = S_2S_1] = s_{2,g-1}^f s_{1,g-1}^m \\ \frac{H_{1,g-1,m}}{2} & \text{if } L = S_2H, \text{with } P[L = S_2H] = s_{2,g-1}^f h_{g-1}^m \\ 0 & \text{if } L = S_2S_2, \text{with } P[L = S_2S_2] = s_{2,g-1}^f s_{2,g-1}^m. \end{cases}$$  \hspace{1cm} (8)

Using eqs. (7) and (8), we can analyze the distribution of the fraction of admixture as a function of the time $g$ and the parameters $s_{1,g}^f, s_{1,g}^m, s_{2,g}^f, s_{2,g}^m$. Under our model, $H_{1,g,\delta}$ takes its values in $Q_g = \{0, 1/2^g, ..., 1 - 1/2^g, 1\}$. Therefore, using eqs. (7) and (8), and recalling that $H_{1,g,f}$ and $H_{1,g,m}$ are identically distributed, for a value $q$ in the set $Q_g$, we can compute the probability $P(H_{1,g,\delta} = q)$ that a random individual from the
hybrid population at generation $g$ has admixture fraction $q$. For $g = 1$, $Q_1 = \{0, \frac{1}{2}, 1\}$, and

$$P(H_{1,1,\delta} = q) = \begin{cases} s_{1,0}^f s_{1,0}^m & \text{if } q = 1 \\ s_{1,0}^f s_{2,0}^m + s_{2,0}^f s_{1,0}^m & \text{if } q = \frac{1}{2} \\ s_{2,0}^m s_{2,0}^m & \text{if } q = 0. \end{cases}$$ \quad (9)

For all subsequent generations, $g \geq 2$, for $q$ in $Q_g$,

$$P(H_{1,g,\delta} = q) = h_{g-1}^f h_{g-1}^m \sum_{r=0}^{2^{g-1}} \left[ P(H_{1,g-1,\delta_1} = \frac{r}{2^{g-1}}) P(H_{1,g-1,\delta_2} = \frac{2^g q - r}{2^{g-1}}) \right]$$

$$+ \left( s_{1,g-1}^f h_{g-1}^m + h_{g-1}^f s_{1,g-1}^m \right) P(H_{1,g-1,\delta} = 2q - 1)$$

$$+ \left( s_{2,g-1}^f h_{g-1}^m + h_{g-1}^f s_{2,g-1}^m \right) P(H_{1,g-1,\delta} = 2q) + I_g(q).$$ \quad (10)

The function $I_g$ is defined for all values of $q$ in $Q_g$, and is equal to

$$I_g(q) = \begin{cases} s_{1,g-1}^f s_{1,g-1}^m & \text{if } q = 1 \\ s_{1,g-1}^f s_{2,g-1}^m + s_{2,g-1}^f s_{1,g-1}^m & \text{if } q = \frac{1}{2} \\ s_{2,g-1}^f s_{2,g-1}^m & \text{if } q = 0 \\ 0 & \text{otherwise.} \end{cases}$$ \quad (11)

In eq. (10), we calculate the probability distribution of $H_{1,g,\delta}$ by taking a sum over all possible parental pairings at the previous generation that would lead to an admixture fraction $q$ at generation $g$. Only three values of $q$ allow for a history without a single hybrid ancestor—$q = 0$, $q = \frac{1}{2}$, and $q = 1$—producing the terms in eq. (11). When there is no sex bias and $s_{1,g-1}^f = s_{1,g-1}^m$, $h_{g-1}^f = h_{g-1}^m$, and $s_{2,g-1}^f = s_{2,g-1}^m$, eqs. (9)-(11) reduce to the corresponding eqs. (3)-(5) from Verdu & Rosenberg (2011).

Eqs. (9)-(11) can be used to analyze the behavior of the distribution of $H_{1,g-1,\delta}$ over time. In Figure 2, we consider constant admixture processes after the founding of the hybrid population ($s_{\alpha,g}^\delta = s_{\alpha}^\delta$ for each $\alpha \in \{1, 2\}$, $\delta \in \{f, m\}$, and $g \geq 1$), plotting $P(H_{1,g,\delta})$ for the first six generations, as computed recursively using eq. (10). In Figure 2A and 2B, we consider a hybrid population founded with equal contributions from source populations $S_1$ and $S_2$, but with no further contributions after $g = 1$. In both of these cases, the distribution of the autosomal admixture fraction contracts around the mean of $\frac{1}{2}$. However, whereas Figure 2A has equal contributions from each sex in the founding generation, Figure 2B has a large initial sex bias. We see that the width of the distribution is smaller with the sex-biased contributions, despite equality of the total contributions $s_{1,0}$ and $s_{2,0}$.
In Figures 2C-2E, we consider admixture scenarios in which the founding of the hybrid population is followed by constant contributions from the source populations over time, $s_1 = 0.1$ and $s_2 = 0.3$. Because the two source populations contribute after the founding, the distribution does not contract around the mean as in Figures 2A and 2B. Also, because the total contributions from $S_1$ and $S_2$ are unequal, the distribution of $H_{1,g,\delta}$ is no longer symmetrical. Rather, because the contribution from $S_2$ is greater, the distribution is shifted toward zero.

Figures 2C and 2D have the same continuing contributions for $g \geq 2$, with no sex bias in the founding generation for Figure 2C, and a large initial sex bias for Figure 2D. Despite different founding contributions, Figures 2C and 2D have similar distributions of $H_{1,g,\delta}$ after a few generations. In Figure 2E, the hybrid population is founded without a sex bias and with equal contributions from the two source populations. The total contributions $s_1$ and $s_2$ are the same as in Figures 2C and 2D, but unlike in Figures 2C and 2D, the continuing contributions are sex-biased, with $s_1^f \neq s_1^m$ and $s_2^f \neq s_2^m$. Even when $s_1$ and $s_2$ are held constant, the distribution of $H_{1,g,\delta}$ depends on the $s_2^\delta$. Notably, the probability of $H_{1,\delta,\delta} = 0$ drops from 0.157 in Figure 2C to 0.000 in Figure 2E. Similarly, the probability of $H_{1,\delta,\delta} = 1$ drops to zero in Figure 2E as well. With these reductions at the extremes, we see a rise in the probability of intermediate values for $H_{1,g,\delta}$.

**Expectation of the fraction of admixture**

Using the law of total expectation, we write the expectation of the fraction of admixture from source population 1 for a random individual of sex $\delta$ in population $H$ at generation $g$ as a function of conditional expectations for all possible pairs of parents $L$,

$$
\mathbb{E}[H_{1,g,\delta}] = \mathbb{E}_L[\mathbb{E}[H_{1,g,\delta}|L]] = \sum_{\ell \in \{S_1S_1, S_1H, S_1S_2, HS_1, HH, HS_2, S_2S_1, S_2H, S_2S_2\}} \mathbb{P}(L = \ell)\mathbb{E}[H_{1,g,\delta}|L = \ell].
$$

(12)
We can simplify this recursion relation. For \( g = 1 \),
\[
E[H_{1,1,\delta}] = P(L = S_1S_1)E[H_{1,1,\delta}|L = S_1S_1] \\
+ P(L = S_1S_2)E[H_{1,1,\delta}|L = S_1S_2] \\
+ P(L = S_2S_1)E[H_{1,1,\delta}|L = S_2S_1] \\
+ P(L = S_2S_2)E[H_{1,1,\delta}|L = S_2S_2].
\] (13)

For all subsequent generations, \( g \geq 2 \), we have
\[
E[H_{1,g,\delta}] = P(L = S_1S_1)E[H_{1,g,\delta}|L = S_1S_1] \\
+ P(L = S_1H)E[H_{1,g,\delta}|L = S_1H] \\
+ P(L = S_1S_2)E[H_{1,g,\delta}|L = S_1S_2] \\
+ P(L = HS_1)E[H_{1,g,\delta}|L = HS_1] \\
+ P(L = HH)E[H_{1,g,\delta}|L = HH] \\
+ P(L = HS_2)E[H_{1,g,\delta}|L = HS_2] \\
+ P(L = S_2S_1)E[H_{1,g,\delta}|L = S_2S_1] \\
+ P(L = S_2H)E[H_{1,g,\delta}|L = S_2H] \\
+ P(L = S_2S_2)E[H_{1,g,\delta}|L = S_2S_2].
\] (14)

Using eqs. (7) and (8), for the first generation, \( g = 1 \), we have
\[
E[H_{1,1,\delta}] = s_1^f s_{1,0}^m E[1] \\
+ \left( s_1^f s_{2,0}^m + s_2^f s_{1,0}^m \right) E\left[ \frac{1}{2} \right] \\
+ s_{2,0}^m E[0].
\] (15)

For all subsequent generations, \( g \geq 2 \), we have
\[
E[H_{1,g,\delta}] = s_{1,g-1}^f s_{1,g-1}^m E[1] \\
+ s_{1,g-1}^f h_{g-1}^m E\left[ \frac{1 + H_{1,g-1,m}}{2} \right] + h_{g-1}^f s_{1,g-1}^m E\left[ \frac{1 + H_{1,g-1,f}}{2} \right] \\
+ \left( s_{1,g-1}^f s_{2,g-1}^m + s_{2,g-1}^f s_{1,g-1}^m \right) E\left[ \frac{1}{2} \right] \\
+ h_{g-1}^f h_{g-1}^m E\left[ \frac{H_{1,g-1,f} + H_{1,g-1,m}}{2} \right] \\
+ s_{2,g-1}^f h_{g-1}^m E\left[ \frac{H_{1,g-1,m}}{2} \right] + h_{g-1}^f s_{2,g-1}^m E\left[ \frac{H_{1,g-1,f}}{2} \right] \\
+ s_{2,g-1}^f s_{2,g-1}^m E[0].
\] (16)
Recalling eqs. (1), (2), and (4), we can simplify the expectation of the fraction of admixture in a random individual of sex $\delta$ from the hybrid population. For $g = 1$, eq. (15) gives

$$\mathbb{E}[H_{1,1,\delta}] = s_{1,0},$$

(17)

the same expression found by Verdu & Rosenberg (2011, eq. 10). For $g \geq 2$, by eq. (16),

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

(18)

Because $H_{1,g,f}$ and $H_{1,g,m}$ are identically distributed, recalling eq. (2), we can simplify the expectation using

$$\mathbb{E}[H_{1,g,f}] = \mathbb{E}[H_{1,g,m}] = \mathbb{E}[H_{1,g,\delta}],$$

where $\delta$ is left as an unspecified sex ($f$ or $m$). For $g \geq 2$, the expectation of the fraction of admixture from source population 1 is

$$\mathbb{E}[H_{1,g,\delta}] = s_{1,g-1} + h_{g-1} \mathbb{E}[H_{1,g-1,\delta}].$$

(19)

We see in eqs. (17) and (19) that the expectation of the fraction of admixture for a random individual of sex $\delta$ from the hybrid population at generation $g$, $\mathbb{E}[H_{1,g,\delta}]$, depends on the total contributions of the source populations ($S_1, S_2, H$) at each generation, $s_{1,g-1}$ and $h_{g-1}$, and not on the sex-specific parameters, $s_{1,g-1}^f, s_{1,g-1}^m, h_{g-1}^f, h_{g-1}^m$. This recursion (eqs. (17) and (19)) is the same as in the non-sex-specific model of Verdu & Rosenberg (2011, eqs. 10 and 11).

**Higher moments of the fraction of admixture**

We can write a general recursion for the higher moments of the fraction of admixture from population $S_1$ in a randomly chosen individual of sex $\delta$ from the hybrid population. For $k \geq 1$, in the first generation, $g = 1$, we have

$$H_{1,1,\delta}^k = \begin{cases} 
1^k & \text{if } L = S_1 S_1, \text{ with } P[L = S_1 S_1] = s_{1,0}^f s_{1,0}^m \\
(\frac{1}{2})^k & \text{if } L = S_1 S_2, \text{ with } P[L = S_1 S_2] = s_{1,0}^f s_{2,0}^m \\
(\frac{1}{2})^k & \text{if } L = S_2 S_1, \text{ with } P[L = S_2 S_1] = s_{2,0}^f s_{1,0}^m \\
0^k & \text{if } L = S_2 S_2, \text{ with } P[L = S_2 S_2] = s_{2,0}^f s_{2,0}^m. 
\end{cases}$$

(20)
Recalling eq. (3) and noting that $h_0 = 0$, we use the binomial theorem to simplify the recursion for the

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

$$H^k_{1,g,\delta} = \begin{cases} 1^k & \text{if } L = S_1S_1, \text{ with } P[L = S_1S_1] = s_{1,g-1}^f s_{1,g-1}^m \\
\left(\frac{1+H_{1,g-1,m}}{2}\right)^k & \text{if } L = S_1H, \text{ with } P[L = S_1H] = s_{1,g-1}^f h_{g-1}^m \\
\left(\frac{1}{2}\right)^k & \text{if } L = S_1S_2, \text{ with } P[L = S_1S_2] = s_{1,g-1}^f s_{2,g-1}^m \\
\left(\frac{1+H_{1,g-1,f}}{2}\right)^k & \text{if } L = HS_1, \text{ with } P[L = HS_1] = h_{g-1}^f s_{1,g-1}^m \\
\left(H_{1,g-1,f}+H_{1,g-1,m}\right)^k & \text{if } L = HH, \text{ with } P[L = HS_1] = h_{g-1}^f h_{g-1}^m \\
\left(\frac{1+H_{1,g-1,f}}{2}\right)^k & \text{if } L = HS_2, \text{ with } P[L = HS_2] = h_{g-1}^f s_{2,g-1}^m \\
\left(H_{1,g-1,f}\right)^k & \text{if } L = S_2S_1, \text{ with } P[L = S_2S_1] = s_{2,g-1}^f s_{1,g-1}^m \\
\left(\frac{H_{1,g-1,m}}{2}\right)^k & \text{if } L = S_2H, \text{ with } P[L = S_2H] = s_{2,g-1}^f h_{g-1}^m \\
0^k & \text{if } L = S_2S_2, \text{ with } P[L = S_2S_2] = s_{2,g-1}^f s_{2,g-1}^m. \quad (21)
\end{cases}
$$

As in the case of $k = 1$, we use the law of total expectation to write a recursion for higher moments of the distribution of the fraction of admixture for all $k \geq 1$. Using the values for the recursion for the fraction of admixture, eqs. (7) and (8), for the first generation, $g = 1$, we have

$$E[H^k_{1,1,\delta}] = s_{1,0}^f s_{1,0}^m E[1^k] + \left(s_{1,0}^f s_{2,0}^m + s_{2,0}^f s_{1,0}^m\right) E\left[\left(\frac{1}{2}\right)^k\right] + s_{2,0}^f s_{2,0}^m E[0^k]. \quad (22)
$$

For $g \geq 2$, we have

$$E[H^k_{1,g,\delta}] = s_{1,g-1}^f s_{1,g-1}^m E[1^k] + \left(s_{1,g-1}^f h_{g-1}^m + s_{2,g-1}^f s_{1,g-1}^m\right) E\left[\left(\frac{1+H_{1,g-1,m}}{2}\right)^k\right] + \left(s_{1,g-1}^f h_{g-1}^m + s_{2,g-1}^f s_{1,g-1}^m\right) E\left[\left(\frac{1}{2}\right)^k\right] + h_{g-1}^f h_{g-1}^m E\left[\left(H_{1,g-1,f}+H_{1,g-1,m}\right)^k\right] + s_{2,g-1}^f h_{g-1}^m E\left[\left(H_{1,g-1,m}\right)^k\right] + h_{g-1}^f h_{g-1}^m E\left[\left(H_{1,g-1,f}\right)^k\right] + s_{2,g-1}^f s_{2,g-1}^m E[0^k]. \quad (23)
$$

Recalling eq. (3) and noting that $h_0 = 0$, we use the binomial theorem to simplify the recursion for the
moments of $H_{1,g,\delta}$. For $g = 1$, we have

$$\mathbb{E}[H_{1,1,\delta}^k] = s_{1,0}^f s_{1,0}^m + \frac{s_{1,0}^f s_{2,0}^m + s_{2,0}^f s_{1,0}^m}{2^k}. \quad (24)$$

For $g \geq 2$, we have

$$\mathbb{E}[H_{1,g,\delta}^k] = s_{1,g-1}^f s_{1,g-1}^m + \frac{s_{1,g-1}^f s_{2,g-1}^m + s_{2,g-1}^f s_{1,g-1}^m}{2^k}
\quad + \frac{s_{1,g-1}^f h_{g-1}^m}{2^k} \left( \sum_{i=0}^{k} \binom{k}{i} \mathbb{E}[H_{1,g-1,i}^i] \right)
\quad + \frac{h_{g-1}^f s_{1,g-1}^m}{2^k} \left( \sum_{i=0}^{k} \binom{k}{i} \mathbb{E}[H_{1,g-1,i}^i] \right)
\quad + \frac{h_{g-1}^f h_{g-1}^m}{2^k} \mathbb{E}[H_{1,g-1,1}^k] + \frac{h_{g-1}^f s_{2,g-1}^m}{2^k} \mathbb{E}[H_{1,g-1,1}^k]. \quad (25)$$

Because $H_{1,g,f}$ and $H_{1,g,m}$ have the same distribution, we can simplify the $k$th moment of the distribution of the fraction of admixture from $S_1$, for $\delta \in \{f, m\}$, to give

$$\mathbb{E}[H_{1,g,\delta}^k] = s_{1,g-1}^f s_{1,g-1}^m + \frac{s_{1,g-1}^f s_{2,g-1}^m + s_{2,g-1}^f s_{1,g-1}^m}{2^k}
\quad + \frac{s_{1,g-1}^f h_{g-1}^m}{2^k} \left( \sum_{i=0}^{k} \binom{k}{i} \mathbb{E}[H_{1,g-1,i}^i] \right)
\quad + \frac{h_{g-1}^f s_{1,g-1}^m}{2^k} \left( \sum_{i=0}^{k} \binom{k}{i} \mathbb{E}[H_{1,g-1,i}^i] \right)
\quad + \frac{s_{2,g-1}^f h_{g-1}^m}{2^k} + \frac{h_{g-1}^f s_{2,g-1}^m}{2^k} \mathbb{E}[H_{1,g-1,1}^k]. \quad (26)$$

For $k = 1$, eqs. (24) and (26) should produce the expectation that we have already derived for $k = 1$. For $k = 1$, using eqs. (1), (2), and (4), eq. (24) gives

$$\mathbb{E}[H_{1,1,\delta}] = s_{1,0}^f s_{1,0}^m + \frac{s_{1,0}^f s_{2,0}^m + s_{2,0}^f s_{1,0}^m}{2} = s_{1,0}, \quad (27)$$

which matches eq. (17). For $g \geq 2$ and $k = 1$, eq. (26) gives

$$\mathbb{E}[H_{1,g,\delta}] = s_{1,g-1}^f s_{1,g-1}^m + \frac{s_{1,g-1}^f s_{2,g-1}^m + s_{2,g-1}^f s_{1,g-1}^m + s_{1,g-1}^f h_{g-1}^m + h_{g-1}^f s_{1,g-1}^m}{2}
\quad + \left( \frac{2 h_{g-1}^f h_{g-1}^m + s_{1,g-1}^f h_{g-1}^m + h_{g-1}^f s_{1,g-1}^m + h_{g-1}^f s_{2,g-1}^m + h_{g-1}^f h_{g-1}^m}{2} \right) \mathbb{E}[H_{1,g-1,\delta}], \quad (28)$$
which simplifies to match eq. (19). Finally, with equal contributions in each population from females and males, so that \( s_{1,g-1}^f = s_{1,g-1}^m = s_{1,g-1} \) and \( s_{2,g-1}^f = s_{2,g-1}^m = s_{2,g-1} \), eqs. (24) and (26) reduce to eqs. 16 and 17 from Verdu & Rosenberg (2011).

**Variance of the fraction of admixture**

When \( k = 2 \), eqs. (24) and (26) produce a recursion for the second moment of \( H_{1,g,\delta} \). Recalling eqs. (1)-(6), for \( g = 1 \), we have

\[
\mathbb{E}[H_{1,1,\delta}^2] = \frac{s_{1,1,0}^f (1 + s_{1,0}^m) + s_{1,1,0}^m (1 + s_{1,0}^f)}{4}. \tag{29}
\]

For \( g \geq 2 \), we have

\[
\mathbb{E}[H_{1,g,\delta}^2] = s_{1,g-1}^f s_{1,g-1}^m + \frac{s_{1,g-1}^f s_{2,g-1}^m + s_{2,g-1}^f s_{1,g-1}^m}{4}
+ \frac{s_{1,g-1}^f h_{g-1}^m}{4} (1 + 2\mathbb{E}[H_{1,g-1,m}] + \mathbb{E}[H_{1,g-1,m}^2])
+ \frac{h_{g-1}^f s_{1,g-1}^m}{4} (1 + 2\mathbb{E}[H_{1,g-1,f}] + \mathbb{E}[H_{1,g-1,f}^2])
+ \frac{h_{g-1}^f h_{g-1}^m}{4} (\mathbb{E}[H_{1,g-1,f}^2] + 2\mathbb{E}[H_{1,g-1,f}]\mathbb{E}[H_{1,g-1,m}] + \mathbb{E}[H_{1,g-1,m}^2])
+ \frac{s_{2,g-1}^f h_{g-1}^m}{4} \mathbb{E}[H_{1,g-1,m}] + \frac{h_{g-1}^f s_{2,g-1}^m}{4} \mathbb{E}[H_{1,g-1,f}]. \tag{30}
\]

Recalling that \( H_{1,g,f} \) and \( H_{1,g,m} \) are identically distributed, eq. (30) simplifies to give

\[
\mathbb{E}[H_{1,g,\delta}^2] = \frac{s_{1,g-1}^f (1 + s_{1,g-1}^m) + s_{1,g-1}^m (1 + s_{1,g-1}^f)}{4}
+ \frac{s_{1,g-1}^f h_{g-1}^m}{2} + \frac{h_{g-1}^f s_{1,g-1}^m}{2} \mathbb{E}[H_{1,g-1,\delta}] + \frac{h_{g-1}^f h_{g-1}^m}{2} (\mathbb{E}[H_{1,g-1,\delta}])^2
+ \frac{h_{g-1}^f + h_{g-1}^m}{4} \mathbb{E}[H_{1,g-1,\delta}]. \tag{31}
\]

Using the definition of the variance \( \mathbb{V}[H_{1,g,\delta}] = \mathbb{E}[H_{1,g,\delta}^2] - (\mathbb{E}[H_{1,g,\delta}])^2 \), and eqs. (17), (19), (29) and (31), for the first generation, for the variance of the fraction of admixture, we have

\[
\mathbb{V}[H_{1,1,\delta}] = \frac{s_{1,1,0}^f (1 - s_{1,0}^m) + s_{1,0}^m (1 - s_{1,1,0}^f)}{4}. \tag{32}
\]

For all subsequent generations, \( g \geq 2 \), we have

\[
\mathbb{V}[H_{1,g,\delta}] = \frac{s_{1,g-1}^f (1 - s_{1,g-1}^m) + s_{1,g-1}^m (1 - s_{1,g-1}^f)}{4}
- \frac{s_{1,g-1}^f h_{g-1}^m}{2} + \frac{s_{1,g-1}^m h_{g-1}^m}{2} \mathbb{E}[H_{1,g-1,\delta}]
+ \frac{h_{g-1}^f (1 - h_{g-1}^f) + h_{g-1}^m (1 - h_{g-1}^m)}{4} (\mathbb{E}[H_{1,g-1,\delta}])^2 + \frac{h_{g-1}^f + h_{g-1}^m}{4} \mathbb{V}[H_{1,g-1,\delta}]. \tag{33}
\]
With no sex bias, so that \( s_{1,g} = s_{1,g}^m = s_{1,g}^f \) and \( s_{2,g} = s_{2,g}^m = s_{2,g}^f \), eqs. (32) and (33) are equivalent to eqs. 22 and 23 from Verdu & Rosenberg (2011).

The recursion for the variance of the fraction of admixture of a random individual of sex \( \delta \) from the hybrid population is dependent on the variance from the previous generation, the expectation from the previous generation, and its square. By contrast with the expectation, the variance of the fraction of admixture depends on the sex-specific contributions from the source populations.

Eqs. (32) and (33) are invariant with respect to an exchange of all variables corresponding to males (superscript \( m \)) with those corresponding with females (superscript \( f \)). Thus, while the variance is affected by the sex-specific admixture contributions, it does not identify the direction of the bias. Despite the dependence of the variance of the autosomal fraction of admixture on sex-specific contributions, under the model, the symmetry demonstrates that autosomal DNA alone does not identify which sex contributes more to the hybrid population from a given source population. This result is reasonable given the non-sex-specific inheritance pattern of autosomal DNA.

Special case: a single admixture event

Using the recursions in eqs. (17), (19), (32), and (33), we can study specific cases in which the contributions are specified. We first consider the case in which the source populations \( S_1 \) and \( S_2 \) do not contribute to the hybrid population after its founding: \( s_{1,g}^f = s_{1,g}^m = s_{2,g}^f = s_{2,g}^m = 0 \), and \( h = (h_g^f + h_g^m)/2 = 1 \), for all \( g \geq 1 \). As before, at the first generation, the hybrid population is not yet formed, and \( h_0 = 0 \). Therefore, \( s_{1,0} + s_{2,0} = s_{1,0}^f + s_{1,0}^m = s_{2,0}^f + s_{2,0}^m = 1 \).

Under this scenario, we can derive the exact expectation and variance of the autosomal fraction of admixture of a random individual from the hybrid population. In the case of a single admixture event, the expectation of the admixture fraction is equal to the expectation at the first generation, because the further contributions are all zero. Using eq. (19), \( s_{1,g-1} = s_{2,g-1} = 0 \) for all \( g \geq 2 \). Therefore, from eq. (17), in the case of a single admixture event, for all \( g \geq 1 \),

\[
E[H_{1,g,\delta}] = s_{1,0}.
\]  

The expectation of the autosomal fraction of admixture from source population \( S_1 \) is constant over time, and it depends on the total—not the sex-specific—contribution from the source population \( S_1 \). As in the general case in eq. (19), for a single admixture event, a sex bias does not affect the expectation. Because the source populations provide no further contributions after the founding generation, unlike in the general
case, the mean admixture fraction does not change with time.

Using eqs. (32) and (33), because \( s_1^f = s_1^m = s_2^f = s_2^m = 0 \) for all \( g \geq 2 \), the variance of the fraction of admixture follows a geometric sequence with ratio \( \frac{1}{2} \). For all generations \( g \geq 1 \),

\[
\mathbb{V}[H_{1,g,\delta}] = \frac{s_{1,0}(1 - s_{1,0}^f) + s_{1,0}^m(1 - s_{1,0}^m)}{2^{g+1}}.
\]  

(35)

For \( s_{1,0}^f = s_{1,0}^m \), by eqs. (1) and (2), the variance matches eq. 25 of Verdu & Rosenberg (2011).

With a single admixture event, the variance decreases monotonically, and its limit is zero for all parameter values. Individuals from the hybrid population only mate within the population, decreasing the variance by a factor of two each generation. Thus, eq. (35) predicts that the distribution of the admixture fraction for a random individual in the hybrid population contracts around the mean, converging to a constant equal to the mean admixture from the first generation.

In eq. (35), considering all possible pairs \( (s_{1,0}^f, s_{1,0}^m) \), with each entry in \([0,1] \), the maximal \( \mathbb{V}[H_{1,g,\delta}] \) occurs at \( (s_{1,0}^f, s_{1,0}^m) = (\frac{1}{2}, \frac{1}{2}) \), a scenario with equal contributions from the two source populations, and no sex bias. At the maximum, the variance is \( \mathbb{V}[H_{1,g,\delta}] = \frac{1}{2^{g+2}} \). Four minima occur, at \( (s_{1,0}^f, s_{1,0}^m) = (0, 0), (0, 1), (1, 0), \) and \( (1, 1) \), cases in which all individuals in generation \( g = 1 \) have the same pair of source populations for their two parents, and in later generations, all individuals continue to have the same value of \( H_{1,g,\delta} \). In these cases, \( \mathbb{V}[H_{1,g,\delta}] = 0 \).

Figure 3 plots the variance in eq. (35) as a function of the sex-specific parameters \( s_{1,0}^f \) and \( s_{1,0}^m \) for three values of \( g \). For \( g = 1 \), a maximum of \( \mathbb{V}[H_{1,1,\delta}] = \frac{1}{8} \) occurs at \( (s_{1,0}^f, s_{1,0}^m) = (\frac{1}{2}, \frac{1}{2}) \), and a minimum, \( \mathbb{V}[H_{1,1,\delta}] = 0 \), at \( (s_{1,0}^f, s_{1,0}^m) = (0, 0), (0, 1), (1, 0), \) or \( (1, 1) \). After one generation of mixing within the hybrid population, with no further contributions from the source populations, the maximum and minima occur at the same values of \( (s_{1,0}^f, s_{1,0}^m) \), but the variance is halved (Fig. 3B). That is, for a given set of values \( (s_{1,0}^f, s_{1,0}^m) \), \( \mathbb{V}[H_{1,2,\delta}] =\mathbb{V}[H_{1,1,\delta}]/2 \). Similarly, for \( g = 8 \) in Figure 3C, \( \mathbb{V}[H_{1,8,\delta}] = \mathbb{V}[H_{1,1,\delta}]/2^7 \). By \( g = 8 \), the hybrid population is quite homogeneous in admixture, and the variance of the admixture fraction has decreased to near zero for all sets of founding parameters. Therefore, the admixture fraction distribution is close to constant, with \( H_{1,8,\delta} \approx s_{1,0} \).

We can analyze the dependence of the variance on the sex-specific parameters by considering constant total contributions \( s_{1,0} \) and allowing the sex-specific contributions to vary, constrained by eq. (1) so that \( 0 \leq s_{1,0}^f, s_{1,0}^m \leq \text{min}(1, 2s_{1,0}) \). Rewriting eq. (35) in terms of \( s_{1,0} \) and \( s_{1,0}^f \),

\[
\mathbb{V}[H_{1,g,\delta}] = \frac{s_{1,0}^f(2s_{1,0} - s_{1,0}^f) - s_{1,0}(2s_{1,0} - 1)}{2^g}.
\]  

(36)
From this expression, it is possible to observe that given a constant $s_{1,0}$ in $[0, 1]$, the maximal variance is produced when $s_{1,0}^f = s_{1,0}^m = s_{1,0}$. The minimal variance occurs when $(s_{1,0}^f, s_{1,0}^m) = (0, 2s_{1,0})$ or $(2s_{1,0}, 0)$ for $s_{1,0} \leq \frac{1}{2}$, or $(1, 2s_{1,0} - 1)$ or $(2s_{1,0} - 1, 1)$ for $s_{1,0} \geq \frac{1}{2}$. This minimum only takes the value $\mathbb{V}[H_{1,g,\delta}] = 0$ when $s_{1,0}$ equals 0, $\frac{1}{2}$, or 1.

For the specific case of $s_{1,0} = \frac{1}{2}$, the total contribution for which the maximal variance occurs in Figure 3, we illustrate the variance at several locations in the allowed range for $s_{1,0}^f$ and $s_{1,0}^m$ (Fig. 4). Four scenarios are plotted with the same total founding contribution from source population 1, $s_{1,0} = \frac{1}{2}$, but with different levels of sex bias. As the female and male contributions become increasingly different, the initial variance decreases. The largest variance for $s_{1,0} = \frac{1}{2}$ occurs at $s_{1,0}^f = s_{1,0}^m = s_{1,0}$, with no sex bias. The minimum occurs when males all come from one source population and females all from the other. In this extreme sex-biased case, the variance is zero constantly over time, as each individual has a male parent from one population and a female parent from the other, and an admixture fraction of $\frac{1}{2}$.

**Special case: constant nonzero contributions**

Next, we consider the case in which an initial admixture event founds the hybrid population, and is then followed by constant nonzero contributions from the source populations. After the founding, for each $g \geq 1$, all admixture parameters are constant in time: $s_{\alpha, g}^\delta = s_{\alpha}^\delta$ for each $\alpha \in \{1, 2\}$ and $\delta \in \{f, m\}$, and $h_{\delta}^g = h_{\delta}$ for each $\delta$. Thus, we have parameter values for the founding, and constant continuing admixture parameters $s_{1}^f, s_{1}^m, s_{2}^f, \text{ and } s_{2}^m$. Each parameter takes its value in $[0, 1]$, as do $s_1$ and $s_2$. By contrast, $h$ takes its value in $(0, 1)$. The case of $h = 1$ is a single admixture event, analyzed above. The $h = 0$ case is trivial because the hybrid population is re-founded at each generation, and the distribution of the admixture fraction thus depends only on the contribution in the previous generation. Therefore, we require $s_1 + s_2 \neq 0 \text{ and } s_1 + s_2 \neq 1$.

Individually, however, $h_f$ and $h_m$ can each vary in $[0, 1]$, as long as they are not both zero or one.

The recursion for the expectation of the autosomal fraction of admixture, eqs. (17) and (19), is equivalent to that derived by Verdu & Rosenberg (2011). Therefore, the closed form of the expectation is equivalent as well. From Verdu & Rosenberg (2011) eq. 30, we have

$$
\mathbb{E}[H_{1,g,\delta}] = \begin{cases} 
s_{1,0}, & g = 1 \\
 s_{1,0}h_{\delta}^{g-1} + s_{1,0}^{1-h_{\delta}^{-1}}, & g \geq 2.
\end{cases}
$$

(37)

We can use the same method as Verdu & Rosenberg (2011) to simplify the second moment. Under the
special case of constant contributions across generations, for \( g = 1 \), eq. (29) gives

\[
E[H_{1,1,\delta}^2] = \frac{s^f_{1,0}(1 + s^m_{1,0}) + s^m_{1,0}(1 + s^f_{1,0})}{4}.
\] (38)

For \( g \geq 2 \), eq. (31) gives

\[
E[H_{1,g,\delta}^2] = \frac{s^f_{1}(1 + s^m_{1}) + s^m_{1}(1 + s^f_{1})}{4} + \frac{s^f_{2}h^m + h^f s^m_{1}}{2} E[H_{1,g-1,\delta}] + \frac{h^f h^m}{2} (E[H_{1,g-1,\delta}])^2 + \frac{h^f + h^m}{4} E[H_{1,g-1,\delta}].
\] (39)

Because this equation is a non-homogenous first-order recurrence with the form

\[
E[H_{1,g,\delta}^2] = \psi(g) + \lambda E[H_{1,g-1,\delta}],
\] (40)

we can use Theorem 3.1.2 of Cull et al. (2005) to solve for a unique solution for \( E[H_{1,g,\delta}^2] \), as in Verdu & Rosenberg (2011). For the initial condition, we have

\[
\alpha_0 = E[H_{1,1,\delta}^2] = \frac{s^f_{1,0}(1 + s^m_{1,0}) + s^m_{1,0}(1 + s^f_{1,0})}{4}.
\] (41)

We define \( \lambda = (h^f + h^m)/4 = h/2 \), and for all \( g \geq 2 \), we have

\[
\psi(g) = \frac{s^f_{1}(1 + s^m_{1}) + s^m_{1}(1 + s^f_{1})}{4} + \frac{s^f_{1}h^m + h^f s^m_{1}}{2} E[H_{1,g-1,\delta}] + \frac{h^f h^m}{2} (E[H_{1,g-1,\delta}])^2.
\] (42)

Using the expected admixture fraction from eq. (37), we can simplify eq. (42). For all \( g \geq 2 \),

\[
\psi(g) = \frac{s^f_{1}(1 + s^m_{1}) + s^m_{1}(1 + s^f_{1})}{4} + \frac{s^f_{1}h^m + h^f s^m_{1}}{2} \left( s_{1,0} h^{g-2} + s_1 \frac{1 - h^{g-2}}{1 - h} \right) + \frac{h^f h^m}{2} \left( s_{1,0} h^{g-2} + s_1 \frac{1 - h^{g-2}}{1 - h} \right)^2.
\] (43)

Therefore, using Theorem 3.1.2 of Cull et al. (2005), we have a unique solution for \( E[H_{1,g,\delta}^2] \):

\[
E[H_{1,g,\delta}^2] = \begin{cases} 
\alpha_0, & g = 1 \\
\alpha_0 \left( \frac{h}{2} \right)^{g-1} + \sum_{i=2}^{g} \left[ \frac{s^f_{1}(1 + s^m_{1}) + s^m_{1}(1 + s^f_{1})}{4} \right. \\
+ \frac{s^f_{1}h^m + h^f s^m_{1}}{2} \left( s_{1,0} h^{i-2} + s_1 \frac{1 - h^{i-2}}{1 - h} \right) \\
+ \frac{h^f h^m}{2} \left( s_{1,0} h^{i-2} + s_1 \frac{1 - h^{i-2}}{1 - h} \right)^2 \left( \frac{h}{2} \right)^{g-i}, & g \geq 2.
\end{cases}
\] (44)
Eq. (44) can be simplified by separating the sum and summing the resulting geometric series:

\[
E[H_{1,g,\delta}^2] = \begin{cases} 
\alpha_0, & g = 1 \\
A_1 + A_2 h^{g-1} + \left(A_3 + A_4 \sum_{i=1}^{g-1} (2h)^i \right) \left(\frac{h}{2}\right)^{g-1}, & g \geq 2,
\end{cases}
\]  
(45)

where \(\alpha_0\) is defined in eq. (41), and

\[
A_1 = \frac{s_1 + s_1^f s_1^m}{2-h} + \frac{s_1 (s_1^f h^m + h^f s_1^m)}{(1-h)(2-h)} + \frac{s_1^2 h^f h^m}{(1-h)^2(2-h)},
\]
(46)

\[
A_2 = \frac{1}{h} \left( s_1^f h^m + h^f s_1^m + \frac{2h}{1-h} s_1 \right) \left( s_{1,0} - \frac{s_1}{1-h} \right),
\]
(47)

\[
A_3 = \alpha_0 + \frac{h^f h^m s_{1,0}}{1-h} \left[ \frac{2}{h} \left( \frac{s_1}{1-h} - s_{1,0} \right) - \frac{s_1}{(1-h)(2-h)} \right] + \left( s_1^f h^m + h^f s_1^m \right) \left[ \frac{s_1 h}{1-h} - s_{1,0} - s_1 \right] - \frac{s_1 + s_1^f s_1^m}{2-h},
\]
(48)

\[
A_4 = \frac{h^f h^m}{2h^2} \left[ s_{1,0} + \frac{s_1}{1-h} \left( \frac{s_1}{1-h} - 2s_{1,0} \right) \right].
\]
(49)

When \(s_1^f = s_1^m\) and \(s_2^f = s_2^m\), \(A_1, A_2, A_3,\) and \(A_4\) are equal to the corresponding quantities in eqs. 39-42 of Verdu & Rosenberg (2011). Therefore, without sex bias, the closed form of the second moment of the admixture fraction, eq. 45, is equal to eq. 38 in Verdu & Rosenberg (2011).

Using the relation \(\forall[H_{1,g,\delta}] = E[H_{1,g,\delta}^2] - (E[H_{1,g,\delta}])^2\) and eqs. (37) and (45), for the variance of the autosomal fraction of admixture, we have

\[
\forall[H_{1,g,\delta}] = \begin{cases} 
\frac{s_{1,0} (1-s_{1,0}) + s_{1,0}^m (1-s_{1,0}^m)}{4}, & g = 1 \\
A_1 + A_2 h^{g-1} + \left[A_3 + A_4 \sum_{i=1}^{g-1} (2h)^i \right] \left(\frac{h}{2}\right)^{g-1} - \left(s_{1,0} h^{g-1} + s_1 \frac{1-h^{g-1}}{1-h} \right)^2, & g \geq 2.
\end{cases}
\]  
(50)

For \(h \neq \frac{1}{2}\), we have

\[
\forall[H_{1,g,\delta}] = \begin{cases} 
\frac{s_{1,0} (1-s_{1,0}) + s_{1,0}^m (1-s_{1,0}^m)}{4}, & g = 1 \\
A_1 + A_2 h^{g-1} + \left[A_3 + A_4 \left(\frac{2h-(2h)^g}{1-2h}\right) \right] \left(\frac{h}{2}\right)^{g-1} - \left(s_{1,0} h^{g-1} + s_1 \frac{1-h^{g-1}}{1-h} \right)^2, & g \geq 2.
\end{cases}
\]  
(51)

For \(h = \frac{1}{2}\), eq. (50) gives

\[
\forall[H_{1,g,\delta}] = \begin{cases} 
\frac{s_{1,0} (1-s_{1,0}) + s_{1,0}^m (1-s_{1,0}^m)}{4}, & g = 1 \\
A_1 + A_2 \left(\frac{1}{2}\right)^{g-1} + A_3 \left(\frac{1}{4}\right)^{g-1} + A_4 (g-1) \left(\frac{1}{4}\right)^{g-1} - \left[2s_1 + (s_{1,0} - 2s_1) \left(\frac{1}{2}\right)^{g-1} \right]^2, & g \geq 2.
\end{cases}
\]  
(52)

Eqs. (50)-(52) simplify to eqs. 43-45 of Verdu & Rosenberg (2011) when \(s_1^f = s_1^m\) and \(s_2^f = s_2^m\).
Limiting variance of admixture over time

Figure 5 illustrates the variance of the autosomal fraction of admixture as a function of \( g \) when the contributions from the source populations are constant over time, computed using eq. (50). The figure shows that if the continuing contributions are held constant, then the long-term limiting variance does not depend on the founding parameters. Unlike in the hybrid isolation case, under the scenario of constant, nonzero contributions from the source populations over time, \( h \neq 0 \) and \( h \neq 1 \), a nonzero limit is reached. Applying eq. (50), we have

\[
\lim_{g \to \infty} V[H_{1,g,\delta}] = A_1 - \left( \frac{s_1}{1-h} \right)^2,
\]

which does not depend on the founding parameters. The limit matches that of Verdu & Rosenberg (2011, eq. 46) in the absence of sex bias \( (s_1^f = s_1^m = s_1, h^f = h^m = h, \text{ and } s_2^f = s_2^m = s_2) \).

The maxima and minima of the limiting variance

Using eqs. (1), (2), (4) and (46), the limit in eq. (53) can be equivalently written in terms of the two female sex-specific contributions, \( s_1^f \) and \( s_2^f \), and the total contributions from the two source populations, \( s_1 \) and \( s_2 \). Considering admixture scenarios with constant \( s_1, s_2 \), with \( s_1 + s_2 \in (0,1] \), but allowing \( s_1^f \) and \( s_2^f \) to range over the closed unit interval, the limiting variance depends on two independent parameters, \( s_1^f, s_2^f \in [0,1] \), subject to the constraint in eq. (1):

\[
\lim_{g \to \infty} V[H_{1,g,\delta}] = -\frac{(s_1 s_2 - s_1^f s_1)^2 + s_1 s_2 (s_1 + s_2)}{(s_1 + s_2)^2 (1 + s_1 + s_2)}.
\]

Treating \( s_1 \) and \( s_2 \) as constants in [0,1], the critical points of eq. (54) are the same as those of

\[
f(s_1^f, s_2^f) = -(s_1^f s_2 - s_1^f s_1)^2.
\]

First we consider the maximum. Because \( f(s_1^f, s_2^f) \) is always negative or zero, the maximal variance given \( s_1 \) and \( s_2 \) occurs when \( f(s_1^f, s_2^f) = 0 \), which occurs on the line \( s_1^f s_2 = s_2^f s_1 \). Equivalently, recalling eq. (1), this line can be written as

\[
\frac{s_1^f}{s_2^f} = \frac{s_1^m}{s_2^m} = \frac{s_1}{s_2}.
\]

Eq. (56) has many solutions for \((s_1^f, s_1^m, s_2^f, s_2^m)\) given \( s_1 \) and \( s_2 \). One solution is \( (s_1^f, s_2^f) = (s_1, s_2) \), which by eq. (1) is equivalent to \( (s_1^m, s_2^m) = (s_1^m, s_2^m) \). Therefore, the limiting variance of the admixture fraction is maximized when there is no sex bias. Figure 6 plots two examples of the variance for constant \( s_1 \) and \( s_2 \), but increasingly different sex-specific contributions from the source populations. In both panels, the admixture history with no sex bias produces the greatest limit.
For fixed \( s_1 \) and \( s_2 \), however, the case without sex bias is not the only maximum of the limiting variance. Figure 7 plots the variance over time for four different admixture histories, each with the same total contributions \( s_1 \) and \( s_2 \), but quite different sex-specific contributions \((s_1^f, s_1^m, s_2^f, s_2^m)\). Each of the four scenarios plotted reaches the same limit because each provides a solution to eq. (56). Because \( f(s_1^f, s_2^f) = 0 \), eq. (54) depends only on the total contributions \( s_1 \) and \( s_2 \). For constant \( s_1 \) and \( s_2 \), any admixture history whose contributions solve \( f(s_1^f, s_2^f) = 0 \) has limiting variance

\[
\lim_{g \to \infty} V[H_{1,g,\delta}] = \frac{s_1 s_2}{(s_1 + s_2)(1 + s_1 + s_2)}.
\]

This maximal limiting variance depends on the total contributions from the source populations, but not on the sex-specific contributions; it is equivalent to eq. 47 of Verdu & Rosenberg (2011).

Thus far, we have considered the maximal limiting variance as a function of the sex-specific parameters given constant total contributions \( s_1 \) and \( s_2 \). We can also identify the values of \( s_1 \) and \( s_2 \) that maximize the limiting variance, considering all \( s_1, s_2 \in [0,1] \). For each choice of \( s_1 \) and \( s_2 \), the maximal variance over values of \( s_1^f \) and \( s_1^m \) is given by eq. (57). We can therefore find the \( s_1 \) and \( s_2 \) that maximize eq. (57). As demonstrated by Verdu & Rosenberg (2011), given \( s_1 + s_2 \), the maximal limiting variance occurs when \( s_1 = s_2 \). Over the range of possible choices for \( s_1 + s_2 \in (0,1) \), the maximum occurs when \( s_1 = s_2 = \frac{1}{2} \). Unlike in Verdu & Rosenberg (2011), however, this maximum requires the sex-specific contributions to solve \( f(s_1^f, s_2^f) = 0 \).

Interestingly, one of the minima of the limiting variance occurs when \( s_1 = s_2 = \frac{1}{2} \), but with \( f(s_1^f, s_2^f) \neq 0 \). Specifically, when \( s_1 = s_2 = \frac{1}{2} \), but all males come from one source population and all females from the other, \((s_1^f, s_1^m, s_2^f, s_2^m) = (1,0,0,1) \) or \((0,1,1,0) \), the limiting variance in eq. (54) is zero. In this case, \( L = S_1S_2 \) or \( L = S_2S_1 \) for every individual in the hybrid population. By eq. (8), the hybrid population is founded anew at each generation, with each individual having admixture fraction \( H_{1,g,\delta} = \frac{1}{2} \). Therefore, the population has zero variance.

More generally, given \( s_1 \) and \( s_2 \), the minimal limiting variance occurs when \((s_1^f, s_2^f) = (2s_1,0) \) or \((s_1^f, s_2^f) = (0,2s_2) \). Given \( s_1 \) and \( s_2 \), the limiting variance is minimized with respect to \( s_1^f \) and \( s_2^f \) when \( f(s_1^f, s_2^f) \) is smallest (eq. (54)). Because \( f \) is the negative of the square of a difference of products, it is greatest when one term is zero and the other is at its maximum, as at \((s_1^f, s_2^f) = (2s_1,0) \) or \((s_1^f, s_2^f) = (0,2s_2) \). These points represent the maximal sex bias for fixed \( s_1 \), \( s_2 \).

If we allow \( s_1 \) and \( s_2 \) to vary, because a variance is bounded below by zero, any set of parameters that produces zero variance is a minimum. In eq. (54) if either \( s_1 = 0 \) or \( s_2 = 0 \), then the limiting variance of
the admixture fraction is zero. When only one population contributes after the founding, in the limit, all ancestry in the hybrid population traces to that population.

Properties of the limiting variance

The limiting variance of the fraction of admixture over time in eq. (53) is a function of the sex-specific contributions from the hybrid population, \( h^f \) and \( h^m \), and source population 1, \( s^f_1 \) and \( s^m_1 \). Recalling eq. (4), the limiting variance is equivalently written as a function of the sex-specific contributions from source population 2, \( s^f_2 \) and \( s^m_2 \), and either source population 1 (eq. (54)), or the hybrid population. It can be viewed as a function of all six sex-specific parameters \( (s^f_1, s^m_1, s^f_2, s^m_2, h^f, h^m) \), four of which can be selected while assigning the other two by the constraint from eq. (4).

We can therefore analyze the behavior of the limiting variance as a function of two of the sex-specific parameters by specifying two other parameters, and allowing the final two parameters, one female and one male, to vary according to eq. (4). Of the four parameters we consider, using the constraint from eq. (4) separately in males and females, two must be male and two must be female. Because the variance is invariant with respect to exchanging the source populations or the sexes, the six-dimensional parameter space generates a number of symmetries. Figures 8-12 examine the five possible, non-redundant ways of choosing two populations and the corresponding male and female parameters from those populations, and holding two corresponding parameters fixed (either from the same sex in the two populations, or for males and females from one population) while allowing the other two to vary. Figure 13 then highlights an informative case that considers the limiting variance as a function of a male and a female parameter from different populations.

Each figure shows multiple contour plots of the limiting variance as a function of two sex-specific parameters, for fixed values of two other parameters. Three cases plot the limiting variance as a function of the female and male parameters from a given population, with the female and male contributions of another population specified. In two other cases, parameters for a single sex from two populations are plotted, specifying the contributions from the other sex for those populations.

By considering these parameter combinations, we can examine the dependence of the variance on sex-specific parameters and parameter interactions, as well as potential bounds on both the parameters and the variance. We highlight a number of symmetries in the limiting variance. The plots also illustrate the maxima and minima found in the previous section.
Properties of the limiting variance in terms of $s_1^f$ and $s_1^m$

In each panel in Figure 8, we consider the variance of the fraction of admixture as a function of $s_1^f$, the female contribution from $S_1$, on the x-axis, and $s_1^m$, the male contribution from $S_1$, on the y-axis, computed using eq. (53). We plot the variance for fixed $h^f$, the female contribution from $H$, and $h^m$, the male contribution from $H$. The domain for $s_1^f$ and $s_1^m$ is constrained by eq. (4), with $s_1^f$ taking values in $[0, 1 - h^f]$, and $s_1^m$ taking values in $[0, 1 - h^m]$.

The upper left plot in Figure 8 shows the variance as a function of $s_1^f$ and $s_1^m$, with $h^f = h^m = h = 0$. In this setting, the hybrid population is founded anew by the source populations each generation, and $s_1^f$ and $s_1^m$ both take values from the full domain $[0, 1]$. For $h^f = h^m = h = 0$, the maximal limiting variance is $\lim_{g \to \infty} \mathbb{V}[H_{1,g,s}] = \frac{1}{8}$, occurring when $s_1^f = s_1^m = s_1 = 0.5$. At this maximum, given eq. (4) and $h^f = h^m = 0$, we have $s_2^f = s_2^m = s_2 = 0.5$. Therefore, as in eq. (54), the maximal limiting variance occurs when female and male contributions from the source populations are equal, and the total contributions from the source populations are equal.

The minima of $\lim_{g \to \infty} \mathbb{V}[H_{1,g,s}] = 0$ occur at the four corners of the plot. At the origin, when $s_1^f = s_1^m = s_1 = 0$, the limiting variance is zero because only $S_2$ contributes to the hybrid population. Individuals in the hybrid population all have parents $L = S_2S_2$, and an admixture fraction of zero (eq. (8)). By exchanging $S_1$ for $S_2$, the case of $s_1^f = s_1^m = s_1 = 1$ is similar.

Additional minima occur at $(s_1^f, s_1^m) = (1, 0)$ or $(0, 1)$. Here, all males come from one source population and all females from the other. Therefore, all individuals at the next generation of the hybrid population have parents $L = S_1S_2$ or $L = S_2S_1$, and admixture fraction $\frac{1}{2}$ (eq. (8)).

For $h^f = h^m = h = 0$, the limiting variance is symmetrical over the line $s_1^m = s_1^f$, as a result of the symmetry between males and female in the variance (eq. (32)). Because the hybrid population provides no contribution and the variance of the fraction of admixture is symmetric with respect to source population, the variance is also symmetric over the lines $s_1^f = 0.5$ and $s_1^m = 0.5$.

The columns of Figure 8 consider increasing, fixed values for $h^f$, and the rows consider increasing, fixed values for $h^m$, both from $\{0, 0.25, 0.5, 0.75, 0.95\}$. All panels maintain the general shape of the limiting variance as a function of $s_1^f$ and $s_1^m$ seen for $h^f = h^m = 0$. However, as the domain for $s_1^f$ and $s_1^m$ shrinks with increasing $h^f$ and $h^m$, the location of the maximal variance changes across panels. In all cases, the maximum of the limiting variance occurs when $s_1^f$ and $s_1^m$ each lie at the midpoints of their respective domains, $s_1^f = (1 - h^f)/2$ and $s_1^m = (1 - h^m)/2$. The magnitude of the limiting variance at each maximum

21
decreases as its location moves away from $s^f_1 = s^m_1 = 0.5$.

For each panel, the minimum $\lim_{g \to \infty} \mathcal{V}[H_{1,g,s}] = 0$ occurs when $s^f_1$ and $s^m_1$ are either both zero or they lie at the maxima of their respective domains. In these cases, only one source population contributes to the hybrid population, and therefore, all individuals in the hybrid population have an admixture fraction from $S_1$ of either 0, when $s^f_1 = s^m_1 = 0$, or 1, when $s^f_1 = 1 - h^f$ and $s^m_1 = 1 - h^m$ ((8)). The limiting variance is no longer zero at the two corners of each contour plot where only one of $\{s^f_1, s^m_1\}$ is at the maximum of its domain; these corners, however, are minima of the variance conditional on the values of $s_1$ and $s_2$. In these cases, males all come from one source population and females from the other, producing a minimum of eq. (54) conditional on fixed $s_1$ and $s_2$.

As in the case of $h^f = h^m = 0$, each plot is symmetrical in reflecting over both the midpoint of the x-axis, $s^f_1 = (1 - h^f)/2$, and that of the y-axis, $s^m_1 = (1 - h^m)/2$. The limiting variance is symmetrical with respect to source population (eq. (54)), and this pair of reflections corresponds to an exchange of source populations. For $h^f = h^m = 0$, the line $s^f_1 = s^m_1$ generates circular contours, but as the contributions from the hybrid population increase, the contours become elliptical.

In Figure 8, plots on the diagonal have equal contributions from males and females in the hybrid population, $h^f = h^m = h$. For $h^f \neq h^m$, plots above the diagonal are equivalent to those below the diagonal with an exchange of female for male contributions, for both $s_1$ and $h$. For example, the plot with $h^f = 0.25$ and $h^m = 0.5$ is equivalent to the plot with $h^f = 0.5$ and $h^m = 0.25$ if the axes are also switched so that $s^m_1$ appears along the x-axis and $s^f_1$ is on the y-axis.

Figure 9 plots the limiting variance as a function of $s^f_1$ on the x-axis, and $s^m_1$ on the y-axis, as in Figure 8, but we now fix values of $s^f_2$ by column and $s^m_2$ by row using eqs. (54) and (1). The maxima and minima occur at the same parameter values found in Figure 8, but they appear in different locations on the plots. For example, in Figure 9, the global maximum across panels occurs in the plot with $s^f_2 = s^m_2 = 0.5$ specified, and $s^f_1 = s^m_1 = 0.5$. By eq. (4), this location implies $h^f = h^m = 0$, the plot that contains the maximal variance in Figure 8. In the upper left plot in Figure 9, the limiting variance is a constant zero for all $s^f_1$ and $s^m_1$, because $s_2 = 0$ (eq. (54)).

Whereas all panels in Figure 8 are symmetric in reflecting over the midpoints of both domains, in Figure 9, only the plots with $s^f_2 = s^m_2$ are symmetric over the line $s^m_1 = s^f_1$. However, the symmetry corresponding to transposing males and females is visible in that a plot above the diagonal and its corresponding plot below the diagonal are equivalent if the axes for $s^f_1$ and $s^m_1$ are switched.
Properties of the limiting variance in terms of $h^f$ and $h^m$

Similarly to Figure 8, Figure 10 considers the limit of the variance of the fraction of admixture over time as a function of the four variables $s^f_1$, $s^m_1$, $h^f$, and $h^m$ using eq. (53). In Figure 10, each plot shows $h^f$ on the x-axis and $h^m$ on the y-axis, for specified values of $s^f_1$ and $s^m_1$, with the domains of $h^f$ and $h^m$ constrained by eq. (4). As in Figure 8, the plots along the diagonal are the cases with $s^f_1 = s^m_1$, and there is a symmetry over this line of plots in that if the values of $s^f_1$ and $s^m_1$ are switched, then the plots will be equivalent with a transposition of the axes.

As in Figure 9, in the upper left plot in Figure 10, $s^f_1 = s^m_1 = s_1 = 0$, and the limiting variance is a constant zero. In Figure 10, the maximal variance occurs at the origin ($h^f = h^m = h = 0$) of the plot with $s^f_1 = s^m_1 = s_1 = 0.5$. As in Figure 8, at the maximum by eq. (4), $s^f_2 = s^m_2 = s_2 = 0.5$. In this case, females and males contribute equally. Both source populations contribute maximally to pull the distribution of the fraction of admixture toward the extremes of zero and one.

Because the limiting variance is symmetrical with respect to source population, and recalling eq. (4), each plot in Figure 10 is equivalent to a corresponding plot in Figure 9 reflected along both the x-axis and y-axis. For example, the plot in Figure 10 with $s^f_1 = 0.5$ and $s^m_1 = 0.25$, is equivalent to the Figure 9 plot with $s^f_2 = 0.5$ and $s^m_2 = 0.25$ if reflected on both the x- and y-axes.

Figures 8-10 illustrate that the global maximum of the limiting variance occurs when the two source populations contribute equally, the contributions from the two sexes are equal, and the hybrid population does not contribute to the next generation. As the parameters move from the location of the maximal limiting variance to the minimum, the variance monotonically decreases.

Properties of the limiting variance in terms of $s^f_1$, $h^f$, and $s^f_2$

Next we plot on the x- and y-axes two parameters of the same sex from different populations. Because the variance is invariant with respect to transposition of females and males, we consider only females without loss of generality. Specifically, in Figure 11 we plot the limiting variance as a function of $s^f_1$ on the x-axis and $h^f$ on the y-axis, for fixed values of $s^m_1$ and $h^m$. In Figure 12, we plot the limiting variance as a function of $s^f_1$ and $s^f_2$, for fixed $s^m_1$ and $h^m$. For both Figure 11 and Figure 12, the domains of $s^f_1$, $s^f_2$, and $h^f$ are constrained by eq. (4).

For Figure 11, the maximal limiting variance occurs in the plot with $s^m_1 = 0.5$ and $h^m = 0$, at $(s^f_1, h^f) = (0.5, 0)$. By eq. (4), this location is the same parameter set for the maximum in Figures 8-10. The maximum
in each plot occurs when \((s^f_1, h^f) = (0.5, 0)\), but the magnitude of the variance decreases with increased distance from the plot with fixed \(s^m_1 = 0.5\) and \(h^m = 0\). Similarly, within each plot, the limiting variance decreases with distance from \((s^f_1, h^f) = (0.5, 0)\).

In the first column of Figure 11, where \(s^m_1 = 0\), the line \(s^f_1 = 0\) produces zero variance because the hybrid population is homogenous, with only one source population contributing. Similarly, on the diagonal \(s^m_1 + h^m = 1\), as \(s^f_2 = s^m_2 = 0\) by eq. (4), the line \(s^f_1 + h^f = 1\) has minimal variance.

In Figure 12, because of the symmetry in sex in eq. (54), the plots above and below those where \(s^m_1 = s^m_2\) are equivalent with a transposition of axes. As in Figures 8-11, the maximal variance occurs in the plot with \(s^m_1 = s^m_2 = 0.5\) at \(s^f_1 = s^f_2 = 0.5\). Also, similar to Figure 11, in the first column, when \(s^m_1 = 0\), the line \(s^f_1 = 0\) is a minimum because no contributions trace to source population \(S_1\); in the first row, when \(s^m_2 = 0\), the line \(s^f_2 = 0\) is a minimum.

Analogous to the similarity between Figures 9 and 10, by eq. (4), each plot in Figure 12 is a transformation of a plot in Figure 11. For example, for the plot with \(s^m_1 = 0.25\) and \(h^m = 0.5\) specified in Figure 11, because the male contributions sum to one, this panel also specifies \(s^m_2 = 0.25\). Therefore, we can compare this plot to the plot with \(s^m_1 = s^m_2 = 0.25\) in Figure 12. Both show \(s^f_1\) on the x-axis, and using eq. (4), we can rewrite the y-axis in Figure 12 as \(s^f_2 = 1 - h^f\).

**Properties of the limiting variance in terms of non-corresponding parameters**

Finally we consider a case in which males from one population in \((S_1, S_2, H)\) are compared to females from a different population. While multiple parameter configurations are possible, we plot one that is particularly informative, providing a perspective on eq. (54) beyond the observations visible in Figures 8-12. Figure 13 plots the limit of the variance of the admixture fraction as a function of \(s^f_1\) on the x-axis and \(s^m_2\) on the y-axis, for fixed values of \(s^f_2\) and \(s^m_1\). We rewrite eq. (54) as a function of \(s^f_1, s^m_1, s^f_2,\) and \(s^m_2\) using eq. (1).

We have,

\[
\lim_{g \to \infty} V[H_{1,g,s}] = \frac{-2(s^f_1 s^m_2 - s^f_2 s^m_1)^2 + (s^f_1 + s^m_1)(s^f_2 + s^m_2)(s^f_1 + s^m_1 + s^f_2 + s^m_2)}{(s^f_1 + s^m_1 + s^f_2 + s^m_2)^2(2 + s^f_1 + s^m_1 + s^f_2 + s^m_2)}.
\]

The limit depends on products of sex-specific parameters, including \(s^f_1 s^m_2\), as can be seen in the shape of the contours in Figure 13, but not in the analogous plots in Figure 9.
Discussion

Our model demonstrates the potential utility of autosomal DNA in the study of sex-biased admixture histories. Under a framework in which admixture occurs over time, potentially with different male and female contributions from the source populations, we have derived recursive expressions for the expectation, variance, and higher moments of the fraction of autosomal admixture. For the special case of constant admixture over time, we have analyzed the behavior of the variance of the admixture fraction. Although the expectation of the autosomal admixture fraction is dependent only on the total contributions from the source populations, we have found that the variance of the autosomal admixture can be informative about sex-specific contributions. Specifically, for constant admixture over time, we have shown that the variance of the autosomal admixture fraction decreases as the male and female contributions become increasingly unequal.

That autosomal DNA can carry a signature of sex-biased admixture might at first appear counterintuitive, as unlike the sex chromosomes, autosomes are carried equally in both sexes. The phenomenon can, however, be understood by analogy with the well-known result that increasing sex bias decreases the effective size of populations (Wright 1931; Crow & Dennison 1988; Caballero 1994; Hartl & Clark 2007). In a computation of effective size using the coalescent, for example (Nordborg & Krone 2002; Ramachandran et al. 2008), the sex bias causes pairs of genetic lineages to be likely to find common ancestors more recently than in a non-sex-biased population, as the reduced chance of a coalescence in the sex that represents a larger fraction of the breeding population is outweighed by the greater chance of a coalescence in the less populous sex. In a similar manner, if admixture is sex-biased, because lineages are more likely to travel along paths through populations with the larger sex-specific contributions, then the variability of genealogical paths—and hence, the variance of the admixture fraction—is reduced compared to the non-sex-biased case.

Autosomal DNA, with its multitude of independent loci, potentially provides more information about the complex histories of hybrid populations, and the full autosomal genome might be less susceptible to selective pressures at individual loci than the sex chromosomes. To take advantage of autosomal information, many recent efforts to study sex-biased demography have focused on comparing autosomal DNA with the X chromosome (Ramachandran et al. 2004, 2008; Wilkins & Marlowe 2006; Hammer et al. 2008, 2010; Bustamante & Ramachandran 2009; Keinan et al. 2009; Casto et al. 2010; Emery et al. 2010; Keinan & Reich 2010; Labuda et al. 2010; Lambert et al. 2010; Gottipati et al. 2011; Heyer et al. 2012; Arbiza et al. 2014). Our study enhances the set of frameworks available for considering effects of admixture and sex bias on autosomal variation.
For a single admixture event, the expectation of the autosomal admixture fraction is constant in time and not dependent on sex-specific contributions. Unlike in the case of hybrid isolation, if constant nonzero contributions from the source populations occur over time, then the variance of the fraction of autosomal admixture reaches a nonzero limit, dependent on these continuing sex-specific admixture rates, but not on the founding contributions. In both scenarios, the variance can be informative about the magnitude of a sex bias in the admixture history of a hybrid population. For an arbitrary constant total contribution from a source population, the maximal variance occurs when there is no sex bias. The maximal variance across all allowable parameter values of the constant admixture model is seen when there is no sex bias, and equal contributions from both source populations, that is, \( s_f^1 = s_m^1 = s_f^2 = s_m^2 = 0.5 \). Two admixture histories minimize the variance of the autosomal admixture fraction. First, the variance is zero when only one source population contributes to the hybrid population, and either all hybrid individuals have an admixture fraction of 0 or they all have a fraction of 1. Second, the variance is zero if all males come from one source population and all females come from the other source population. In this scenario, all individuals in the hybrid population have an admixture fraction of \( \frac{1}{2} \).

While the variance of the autosomal admixture fraction suggests that autosomal DNA is informative about sex-biased admixture, the relationship between the variance and the sex-specific parameters is complex. We uncovered an interesting case in which quite different sex-specific histories can lead to the same variance over time (Fig. 7). The variance is in fact dependent on the product of multiple sex-specific parameters, but not on each parameter separately (Fig. 13). In particular, when \( s_f^1 s_m^2 = s_f^2 s_m^1 \), we demonstrate that the variance is maximized (eqs. (54)-(56)). Therefore, when the equality \( s_f^1 s_m^2 = s_f^2 s_m^1 \) holds, the limiting variance depends only on the total contributions from the source populations, \( s_1 \) and \( s_2 \) (eq. (57)). The symmetry arises from the non sex-specific inheritance of autosomal DNA.

We have considered two scenarios, isolation of a hybrid population after its founding, and constant contributions from source populations to the hybrid population over time. While the admixture history of real hybrid populations is likely much more complex than these, our models can provide a starting point for statistical frameworks to estimate the parameters of mechanistic admixture models. It is noteworthy that although sex bias does influence autosomal variation, because autosomal DNA is not inherited sex-specifically, the sex that contributes more from a given source population cannot be identified with autosomal DNA alone. Because the X-chromosome follows a sex-specific mode of inheritance, consideration of the X-chromosome alongside autosomal data under the mechanistic model may be able to help differentiate between scenarios that produce the same variance with different choices of the sex with a greater contribution.
Acknowledgments. We thank Ethan Jewett and Michael D. Edge for useful discussions. We acknowledge support from a National Science Foundation Graduate Research Fellowship and from National Science Foundation grant BCS-1147534.

Bibliography


Figures

Figure 1: Schematic of the mechanistic model of admixture over time. Two source populations, $S_1$ and $S_2$, contribute both males and females 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 $s_{1,g}^f, s_{2,g}^f$, and $s_{1,g}^m, s_{2,g}^m$, for females and males respectively. $H_{\alpha,g,\delta}$ 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 $H$.

Figure 2: Probability distribution of the fraction of admixture from source population $S_1$, $\mathbb{P}(H_{1,g,\delta})$, for a random individual from the hybrid population for the first six generations (eqs. (9)-(11)). Each column corresponds to a specified admixture scenario, with constant contributions from the source populations over time after founding ($s_{\alpha,g}^\delta = s_\alpha^\delta$ for each $\alpha \in \{1, 2\}, \delta \in \{f, m\}$, and $g \geq 2$).

Figure 3: The variance of the fraction of admixture, $V[H_{1,g,\delta}]$, as a function of female and male contributions from source population $S_1$ in the first generation, $s_{1,0}^f$ and $s_{1,0}^m$, in the case of hybrid isolation. (A) $g = 1$. (B) $g = 2$. (C) $g = 8$. At each generation, the variance decreases toward zero by a factor of two. Considering all ($s_{1,0}^f, s_{1,0}^m$) in $[0, 1] \times [0, 1]$, the maximal variance occurs when $(s_{1,0}^f, s_{1,0}^m) = (\frac{1}{2}, \frac{1}{2})$, and the minimal variance occurs when $(s_{1,0}^f, s_{1,0}^m) = (0, 0), (0, 1), (1, 0), (1, 1)$. The variance is calculated using eq. (35).

Figure 4: The variance of the fraction of admixture, $V[H_{1,g,\delta}]$, when contributions from the source populations occur only in the founding generation and the total contribution from source population 1 is held constant at $s_{1,0} = \frac{1}{2}$. The limit of the variance of the fraction of admixture over time is zero for any choice of ($s_{1,0}^f, s_{1,0}^m$). The magnitude of the variance, calculated from eq. (35), is inversely related to the level of sex bias. For all four scenarios, $s_1 = s_2 = 0$ and $s_{1,0} = s_{2,0} = \frac{1}{2}$.

Figure 5: The variance of the fraction of admixture over time for constant, nonzero contributions from the source populations, with different levels of sex bias in the founding of the hybrid population, and constant, equal, and nonzero subsequent contributions from the source populations and sex for $g \geq 1$. In all cases, $s_{1}^f = s_{1}^m = s_{2}^f = s_{2}^m = 0.2$. The variance, calculated with eq. (50), reaches a nonzero limit $\lim_{g \to \infty} V[H_{1,g,\delta}] = 1/24$.

Figure 6: The variance of the fraction of admixture over time for constant, nonzero contributions from the source populations, with different levels of sex bias, but the same total contribution from the two source populations. (A) $s_{1,0}^f = s_{1,0}^m = 0.9$, and $s_1 = s_2 = 0.1$, (B) $s_{1,0}^f = s_{1,0}^m = 0.75$ and $s_1 = s_2 = 0.2$. The
variance reaches a nonzero limit when \( s^f_1, s^m_1, s^f_2, \) and \( s^m_2 \) are nonzero and constant over time. The variance is calculated using eq. (50).

**Figure 7:** The variance of the fraction of admixture over time for constant, nonzero contributions from the source populations, but multiple different ratios of female to male contributions. When the sex-specific parameters satisfy the equation \( s^f_1s^m_2 = s^m_1s^f_2 \), multiple different demographic scenarios have the same limiting variance of the admixture fraction. In all cases, \( \lim_{g \to \infty} \mathbb{V}[H_{1,g,\delta}] = \frac{4}{75} \). The variance is calculated using eq. (33). For all scenarios \( s^f_{1,0} = s^m_{1,0} = 0.95 \).

**Figure 8:** Contour plots of the limit of the variance of the fraction of admixture over time as a function of \( s^f_1 \) on the x-axis and \( s^m_1 \) on the y-axis for specified values of \( h^f \) by column and \( h^m \) by row. The domains of \( s^f_1 \) and \( s^m_1 \) are \([0, 1-h^f]\) and \([0, 1-h^m]\), respectively.

**Figure 9:** Contour plots of the limit of the variance of the fraction of admixture over time as a function of \( s^f_1 \) on the x-axis and \( s^m_1 \) on the y-axis for specified values of \( s^f_2 \) by column and \( s^m_2 \) by row. The domains of \( s^f_1 \) and \( s^m_1 \) are \([0, 1-s^f_2]\) and \([0, 1-s^m_2]\), respectively.

**Figure 10:** Contour plots of the limit of the variance of the fraction of admixture over time as a function of \( h^f \) on the x-axis and \( h^m \) on the y-axis for specified values of \( s^f_1 \) by column and \( s^m_1 \) by row. The domains of \( h^f \) and \( h^m \) are \([0, 1-s^f_1]\) and \([0, 1-s^m_1]\), respectively.

**Figure 11:** Contour plots of the limit of the variance of the fraction of admixture over time as a function of \( s^f_1 \) on the x-axis and \( h^f \) on the y-axis for specified values of \( s^m_1 \) by column and \( h^m \) by row. The domains of \( s^f_1 \) and \( h^f \) are bounded by the function \( s^f_1 + h^f = 1 \).

**Figure 12:** Contour plots of the limit of the variance of the fraction of admixture over time as a function of \( s^f_1 \) on the x-axis and \( s^f_2 \) on the y-axis for specified values of \( s^m_1 \) by column and \( s^m_2 \) by row. The domains of \( s^f_1 \) and \( s^f_2 \) are bounded by the function \( s^f_1 + s^f_2 = 1 \).

**Figure 13:** Contour plots of the limit of the variance of the fraction of admixture over time as a function of \( s^f_1 \) on the x-axis and \( s^m_2 \) on the y-axis for specified values of \( s^f_2 \) by column and \( s^m_1 \) by row. The domains of \( s^f_1 \) and \( s^m_2 \) are \([0, 1-s^f_2]\) and \([0, 1-s^m_1]\), respectively.

**Table 1:** The probabilities that an individual from the hybrid population at generation \( g \) has one of nine possible sets of parents from \( S_1, S_2 \) or \( H \), assuming random mating. The parameter \( s^\delta_{\alpha,g} \) is the probability that the parent of sex \( \delta \) for a randomly chosen individual from the hybrid population, at generation \( g \), is from the source population \( \alpha \). Similarly, the probability that this parent is from \( H \) is \( h^\delta_g \).
Figure 1
The figure shows the probability distribution of $P(H_{1,1}, \delta)$, $P(H_{1,2}, \delta)$, $P(H_{1,3}, \delta)$, $P(H_{1,4}, \delta)$, $P(H_{1,5}, \delta)$, and $P(H_{1,6}, \delta)$ for each of the six subfigures labeled A to E. The distribution is plotted across different values of $H_{1,g}$ and $\delta$.

The table below provides the values of the parameters $s_{1,0}^f$, $s_{2,0}^f$, $s_{1,0}^e$, $s_{2,0}^e$, $s_{1,0}^w$, and $s_{2,0}^w$ for each subfigure, with $g = 1$ to 6.

<table>
<thead>
<tr>
<th>Subfigure</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$[0.5 \ 0.5]$</td>
</tr>
<tr>
<td>B</td>
<td>$[0.95 \ 0.05]$</td>
</tr>
<tr>
<td>C</td>
<td>$[0.5 \ 0.5]$</td>
</tr>
<tr>
<td>D</td>
<td>$[0.95 \ 0.05]$</td>
</tr>
<tr>
<td>E</td>
<td>$[0.5 \ 0.5]$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subfigure</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$[0.5 \ 0.5]$</td>
</tr>
<tr>
<td>B</td>
<td>$[0.95 \ 0.95]$</td>
</tr>
<tr>
<td>C</td>
<td>$[0.5 \ 0.5]$</td>
</tr>
<tr>
<td>D</td>
<td>$[0.95 \ 0.95]$</td>
</tr>
<tr>
<td>E</td>
<td>$[0.5 \ 0.5]$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subfigure</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$[0.5 \ 0.5]$</td>
</tr>
<tr>
<td>B</td>
<td>$[0.5 \ 0.3]$</td>
</tr>
<tr>
<td>C</td>
<td>$[0.1 \ 0.3]$</td>
</tr>
<tr>
<td>D</td>
<td>$[0.1 \ 0.3]$</td>
</tr>
<tr>
<td>E</td>
<td>$[0.2 \ 0]$</td>
</tr>
</tbody>
</table>

Figure 2
Figure 3
Figure 4
Figure 5
Figure 6
Figure 7
Figure 8
Figure 9

---

Figure 9
Figure 10

$s^m_1 = 0.0$

$s^m_1 = 0.25$

$s^m_1 = 0.5$

$s^m_1 = 0.75$

$s^m_1 = 0.95$

$V[H_{1,g,\delta}]$
Figure 11
Figure 12
Figure 13
<table>
<thead>
<tr>
<th>Case</th>
<th>Female Parent’s Population</th>
<th>Male Parent’s Population</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$S_1$</td>
<td>$S_1$</td>
<td>$s_{1,g-1}^fs_{1,g-1}^m$</td>
</tr>
<tr>
<td>2</td>
<td>$S_1$</td>
<td>$H$</td>
<td>$s_{1,g-1}^f h_{g-1}^m$</td>
</tr>
<tr>
<td>3</td>
<td>$S_1$</td>
<td>$S_2$</td>
<td>$s_{1,g-1}^f s_{2,g-1}^m$</td>
</tr>
<tr>
<td>4</td>
<td>$H$</td>
<td>$S_1$</td>
<td>$h_{g-1}^f s_{1,g-1}^m$</td>
</tr>
<tr>
<td>5</td>
<td>$H$</td>
<td>$H$</td>
<td>$h_{g-1}^f h_{g-1}^m$</td>
</tr>
<tr>
<td>6</td>
<td>$H$</td>
<td>$S_2$</td>
<td>$h_{g-1}^f s_{2,g-1}^m$</td>
</tr>
<tr>
<td>7</td>
<td>$S_2$</td>
<td>$S_1$</td>
<td>$s_{2,g-1}^f s_{1,g-1}^m$</td>
</tr>
<tr>
<td>8</td>
<td>$S_2$</td>
<td>$H$</td>
<td>$s_{2,g-1}^f h_{g-1}^m$</td>
</tr>
<tr>
<td>9</td>
<td>$S_2$</td>
<td>$S_2$</td>
<td>$s_{2,g-1}^f s_{2,g-1}^m$</td>
</tr>
</tbody>
</table>