A General Mechanistic Model for Admixture Histories of Hybrid Populations

Paul Verdu[1] and Noah A. Rosenberg
Department of Biology, Stanford University, Stanford, California 94305

ABSTRACT Admixed populations have been used for inferring migrations, detecting natural selection, and finding disease genes. These applications often use a simple statistical model of admixture rather than a modeling perspective that incorporates a more realistic history of the admixture process. Here, we develop a general model of admixture that mechanistically accounts for complex historical admixture processes. We consider two source populations contributing to the ancestry of a hybrid population, potentially with variable contributions across generations. For a random individual in the hybrid population at a given point in time, we study the fraction of genetic admixture originating from a specific one of the source populations by computing its moments as functions of time and of introgression parameters. We show that very different admixture processes can produce identical mean admixture proportions, but that such processes produce different values for the variance of the admixture proportion. When introgression parameters from each source population are constant over time, the long-term limit of the expectation of the admixture proportion depends only on the ratio of the introgression parameters. The variance of admixture decreases quickly over time after the source populations stop contributing to the hybrid population, but remains substantial when the contributions are ongoing. Our approach will facilitate the understanding of admixture mechanisms, illustrating how the moments of the distribution of admixture proportions can be informative about the historical admixture processes contributing to the genetic diversity of hybrid populations.

EXCHANGES of genes between two or more mutually isolated populations can result in new admixed or hybrid populations. For nearly 80 years, statistical models have been used to estimate the proportions of the genetic ancestry of an admixed population that are derived from the various parental source populations (Bernstein 1931; Roberts and Hiorns 1965; Long and Smouse 1983; Long 1991; Chakraborty *et al.* 1992; Bertorelle and Excoffier 1998; Pritchard *et al.* 2000; Chikhi *et al.* 2001; Wang 2003; Tang *et al.* 2005) and, more recently, to determine the probable ancestral origins of chromosomal segments within individual genomes (Ungerer *et al.* 1998; Falush *et al.* 2003; Hoggart *et al.* 2004; Patterson *et al.* 2004; Baird 2006; Tang *et al.* 2006; Sankararaman *et al.* 2008; Bercovici and Geiger 2009; Price *et al.* 2009). Admixed human populations have been employed in assessing patterns of migration and genetic structure (Parra *et al.* 2001; Seldin *et al.* 2007; Wang *et al.* 2008; Silva-Zolezzi *et al.* 2009), detecting natural selection (Workman *et al.* 1963; Cavalli-Sforza and Bodmer 1971; Chakraborty and Weiss 1988; Tang *et al.* 2007; Oleksyk *et al.* 2010; Lohmueller *et al.* 2011), and identifying phenotypically important genes through admixture-mapping strategies (McKeigue 1998, 2005; Halder and Shriver 2003; Reich and Patterson 2005; Smith and O'Brien 2005; Buerkle and Lexer 2008; Seldin *et al.* 2011).

Many recent methods consider admixed populations as statistical combinations of the source populations, treating allele frequencies in a hybrid population as linear combinations of allele frequencies in the source groups. While this perspective is informative in diverse applications for describing the current structure of admixed populations, it does not mechanistically account for the inherent complexity of admixture processes. In the case of humans, throughout history, previously isolated populations have come into contact through colonization waves, forced displacements, and population migrations. Moreover, admixture processes have often been influenced by sociocultural rules on intermarriage in contexts of ethnic conflict or discrimination, slavery, and clan or caste systems. Such complex histories of

Copyright © 2011 by the Genetics Society of America
doi: 10.1534/genetics.111.132787
Manuscript received July 14, 2011; accepted for publication September 23, 2011
Available freely online through the author-supported open access option.
Supporting information is available online at http://www.genetics.org/content/suppl/2011/10/03/genetics.111.132787.DC1.
[1]Corresponding author: Department of Biology, Stanford University, 371 Serra Mall, Stanford, CA 94305. E-mail: verdu@stanford.edu

social behaviors have produced a variety of patterns of genetic variation in different admixed groups (Parra *et al.* 1998, 2003; Bonilla *et al.* 2005; Bedoya *et al.* 2006; Chaix *et al.* 2007; Wang *et al.* 2008; Halder *et al.* 2009; Tishkoff *et al.* 2009; Verdu *et al.* 2009; Bryc *et al.* 2010).

Mechanistic perspectives that seek to describe the history of admixture processes through time rather than estimating admixture proportions from the source populations in descriptive statistical models have been part of some recent studies of admixture (Briscoe *et al.* 1994; Stephens *et al.* 1994; Pfaff *et al.* 2001), and they have been used to make theoretical predictions of admixture proportions as well as of Wright's fixation index F_{ST} and statistics measuring linkage disequilibrium (Chakraborty and Weiss 1988; Long 1991; Guo *et al.* 2005). Most of these approaches have relied on models with a relatively simple dynamic considering a single admixture event between populations, rather than on models that investigate a more complex history of admixture processes.

Ewens and Spielman (1995) proposed a mechanistic admixture model that incorporated multiple admixture events involving multiple source populations. This model has been used primarily to evaluate the influence of population subdivision and admixture on the performance of the transmission-disequilibrium test (Ewens and Spielman 1995) and to examine linkage disequilibrium statistics (Guo *et al.* 2005). However, complex mechanistic models have not been used to directly determine the influence of admixture histories on the admixture patterns of hybrid populations.

In this article, expanding on the models of Ewens and Spielman (1995) and Guo *et al.* (2005), we develop a general mechanistic model of a historical admixture process. We first introduce the model, the most general form of which considers m source populations that contribute to the ancestry of a hybrid population. We treat the fraction of genetic admixture in the hybrid population originating from a specific source population as a random variable, whose distribution we study over time in the $m = 2$ case. We next examine the expectation, variance, and higher moments of the admixture fraction as functions of time and of the introgression parameters, and we consider in detail a special case in which admixture is constant across generations. Finally, we conclude with a discussion of the implications of the work for empirical studies of admixture.

The Model

We describe a version of our mechanistic admixture model in which the number of source populations is two. The generalization to m source populations is straightforward, and we provide it in supporting information, File S1, Figure S1, and Table S1.

Define population H ("hybrid") as a population consisting of immigrant individuals from two isolated source populations, S_1 and S_2, and hybrid individuals who have ancestors from both S_1 and S_2. The hybrid population can be viewed as having a separate location or status from S_1 and S_2, so that individuals within H can interbreed with each other and with new immigrants that come from the source populations.

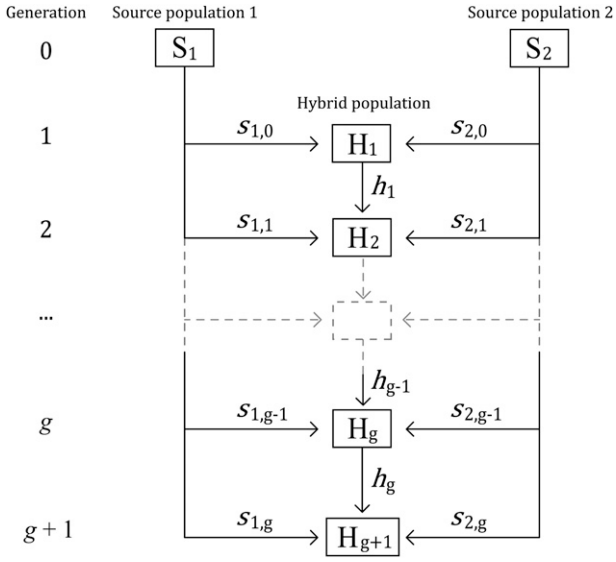

Figure 1 Diagram of a mechanistic model of admixture involving two isolated source populations.

We let $s_{1,g}$, $s_{2,g}$, and h_g be the fractional contributions of populations S_1, S_2, and H to the hybrid population H at generation $g + 1$. That is, for a randomly chosen individual in H at generation $g + 1$, the probabilities that a randomly chosen parent of the individual derives from populations S_1, S_2, and H are $s_{1,g}$, $s_{2,g}$, and h_g, respectively. These probabilities can differ in different generations, but for all $g \geq 0$, the parameters $s_{1,g}$, $s_{2,g}$, and h_g have values that are ≥ 0 and ≤ 1, such that $s_{1,g} + s_{2,g} + h_g = 1$. At generation 0, the hybrid population is not yet formed. Therefore, $h_0 = 0$ and $s_{1,0} + s_{2,0} = 1$. Hence, considering the period through generation g, in addition to g itself, this model has $2g - 1$ independent parameters: one introgression proportion in the first generation and two introgression proportions in each of the next $g - 1$ generations. A diagram of the model appears in Figure 1.

Admixture fractions for a random individual in the hybrid population

We focus on a key quantity in admixed populations, namely the fraction of admixture from one of the source populations for a random individual in H at a randomly chosen locus. This fraction represents the proportion of the genome of a randomly chosen individual in H that ultimately traces to a specific source.

We indicate the possible sources for the (unordered) parents of an individual in H by S_1S_1, S_1S_2, S_1H, S_2H, HH,

Table 1 Possible pairs of parents for a random individual in the hybrid population *H* at generation *g*, and their probabilities

Populations of origin of the parents of a random individual in population *H* at generation $g \geq 1$	Probability
S_1 and S_1	$s_{1,g-1}^2$
S_1 and H (or H and S_1)	$2s_{1,g-1}h_{g-1}$
S_1 and S_2 (or S_2 and S_1)	$2s_{1,g-1}s_{2,g-1}$
H and H	h_{g-1}^2
S_2 and H (or H and S_2)	$2s_{2,g-1}h_{g-1}$
S_2 and S_2	$s_{2,g-1}^2$

Note that at generation 0, $h_0 = 0$ because the hybrid population is not yet formed.

and S_2S_2. An individual in generation $g \geq 1$ has one of several possible types of parents, each with some probability dependent on the parameters $s_{1,g-1}$, $s_{2,g-1}$, and h_{g-1} (Table 1). If the parents have different ancestries, we do not distinguish the order of the two parents, so that, for example, "S_1H" does not convey which specific parent is from population S_1 and which is from H.

Let Y be a random variable indicating the source populations of the parents of a random individual in H. Let $H_{1,g}$ be the admixture fraction from source population S_1 for a random individual in population H at a random locus at generation g. Because at generation 0, the hybrid population is not yet formed, $h_0 = 0$, and $H_{1,0}$ is not defined. Using Table 1, we can write a recursion relation to calculate $H_{1,g}$ for all $g \geq 1$. For the first generation ($g = 1$), we have

$$H_{1,1} = \begin{cases} 1 \text{ if } Y = S_1S_1, \text{with } P[Y = S_1S_1] = s_{1,0}^2 \\ \frac{1}{2} \text{ if } Y = S_1S_2, \text{with } P[Y = S_1S_2] = 2s_{1,0}s_{2,0} \\ 0 \text{ if } Y = S_2S_2, \text{with } P[Y = S_2S_2] = s_{2,0}^2. \end{cases} \tag{1}$$

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

$$H_{1,g} = \begin{cases} 1 & \text{if } Y = S_1S_1, \text{ with } P[Y = S_1S_1] = s_{1,g-1}^2 \\ \frac{H_{1,g-1}+1}{2} & \text{if } Y = S_1H, \text{ with } P[Y = S_1H] = 2s_{1,g-1}h_{g-1} \\ \frac{1}{2} & \text{if } Y = S_1S_2, \text{ with } P[Y = S_1S_2] = 2s_{1,g-1}s_{2,g-1} \\ \frac{H_{1,g-1}^{(1)} + H_{1,g-1}^{(2)}}{2} & \text{if } Y = HH, \text{ with } P[Y = HH] = h_{g-1}^2 \\ \frac{H_{1,g-1}}{2} & \text{if } Y = S_2H, \text{ with } P[Y = S_2H] = 2s_{2,g-1}h_{g-1} \\ 0 & \text{if } Y = S_2S_2, \text{with } P[Y = S_2S_2] = s_{2,g-1}^2. \end{cases} \tag{2}$$

Here, $H_{1,g-1}^{(1)}$ and $H_{1,g-1}^{(2)}$ are fractions of ancestry from source population S_1 for the two parents of a hybrid individual at generation g with $Y = HH$. We use the superscripts (1) and (2) only to indicate that $H_{1,g-1}^{(1)}$ and $H_{1,g-1}^{(2)}$ are separate independent and identically distributed (IID) random variables, so that if an individual in population H at generation g has two parents from H, the admixture fraction is distributed as the mean of the admixture fractions for two IID random individuals from H in the previous generation.

Equations 1 and 2 allow us to analyze the behavior of the admixture fraction from a source population for a random individual in the hybrid population, as a function of the time g and the parameters $s_{1,i}$, $s_{2,i}$, and h_i for $i = 1, 2, \ldots, g-1$. Under our model, the set of possible values of $H_{1,g}$ is $Q_g = \{0, 1/2^g, \ldots, (2^g-1)/2^g, 1\}$. Using Equations 1 and 2, we can show that for a value q in the set Q_g, the probability $P(H_{1,g} = q)$ that a random individual in the hybrid population at generation g has admixture fraction q can be computed using the following recursion relation (*Appendix*). For the first generation ($g = 1$), we have $Q_1 = \{0, 1/2, 1\}$ and

$$P(H_{1,1} = q) = \begin{cases} s_{1,0}^2 & \text{if } q = 1 \\ 2s_{1,0}s_{2,0} & \text{if } q = \frac{1}{2} \\ s_{2,0}^2 & \text{if } q = 0. \end{cases} \tag{3}$$

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

$$\begin{aligned} P(H_{1,g} = q) = h_{g-1}^2 \sum_{r=0}^{2^{g-1}} \left[P\left(H_{1,g-1} = \frac{r}{2^{g-1}}\right) P\left(H_{1,g-1} = \frac{2^g q - r}{2^{g-1}}\right) \right] \\ + \quad 2s_{2,g-1}h_{g-1}P(H_{1,g-1} = 2q) \\ + \quad 2s_{1,g-1}h_{g-1}P(H_{1,g-1} = 2q-1) + I_g(q), \end{aligned} \tag{4}$$

where the function I_g is defined for all values of q in Q_g and equals

$$I_g(q) = \begin{cases} s_{1,g-1}^2 & \text{if } q = 1 \\ 2s_{1,g-1}s_{2,g-1} & \text{if } q = \frac{1}{2} \\ s_{2,g-1}^2 & \text{if } q = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

$P(H_{1,g} = q)$ is zero when q is not in Q_g.

We can use Equations 3–5 to examine the evolution of the distribution of $H_{1,g}$ across generations. For five scenarios in which the admixture process is constant after the founding of population H ($s_{1,g} = s_1$ and $s_{2,g} = s_2$ for all $g \geq 1$), Figure 2 plots the complete set of values of $P(H_{1,g})$ for the first six generations.

In Figure 2A, we consider a scenario in which the hybrid population H is founded with equal contributions from source populations S_1 and S_2 ($s_{1,0} = s_{2,0} = \frac{1}{2}$), which do not subsequently contribute to H ($s_{1,g} = s_{2,g} = 0$ for all $g \geq 1$). We can see that the probability $P(H_{1,g})$for a random individual in H to exhibit a given fraction of admixture from S_1 is distributed symmetrically around $\frac{1}{2}$ at each generation, with a single mode at $H_{1,g} = \frac{1}{2}$ for each of the first six generations. This pattern arises from the fact that after a symmetric founding event, in the absence of immigration, no new input enters the admixed population from either source, and the distribution remains symmetric.

Figure 2B considers an admixture process with the same starting conditions as in the previous case ($s_{1,0} = s_{2,0} = \frac{1}{2}$), in which the subsequent contributions from the source

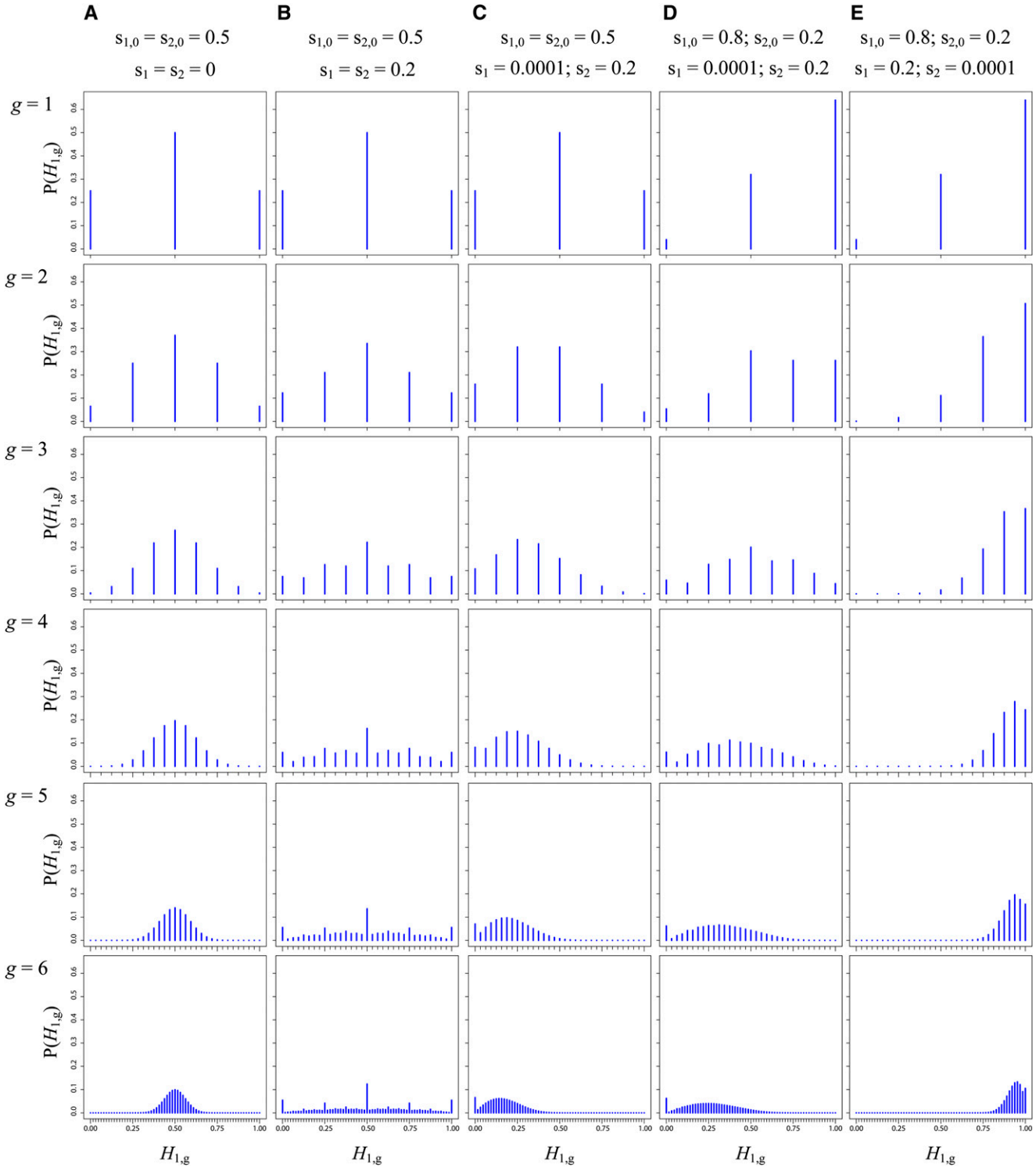

Figure 2 Probability distribution of the admixture fraction from source population S_1 [$P(H_{1,g})$, Equations 3–5] for a random individual in the hybrid population at each of several points in time. Rows correspond to distributions of $P(H_{1,g})$ in five scenarios for each of a series of values of g. Columns correspond to five constant admixture processes ($s_{1,g} = s_1$ and $s_{2,g} = s_2$ for all $g \geq 1$). (A) Population H is founded at generation 0 with equal proportions from source populations S_1 and S_2 ($s_{1,0} = s_{2,0} = 0.5$). Subsequently, both source populations do not contribute to H ($s_1 = s_2 = 0$). (B) Population H is founded with equal proportions from S_1 and S_2 ($s_{1,0} = s_{2,0} = 0.5$). Subsequently, the source populations contribute equally to H ($s_1 = s_2 = 0.2$). (C) Population H is founded with equal proportions from S_1 and S_2 ($s_{1,0} = s_{2,0} = 0.5$). Subsequently, the source populations contribute unequally to H, with S_2 contributing more than S_1 ($s_1 = 0.0001$, $s_2 = 0.2$). (D) Population H is founded at generation 0 with a greater contribution from S_1 than S_2 ($s_{1,0} = 0.8$, $s_{2,0} = 0.2$). Subsequently, the source populations contribute unequally to population H, with S_2 contributing more than S_1 ($s_1 = 0.0001$, $s_2 = 0.2$). (E) Population H is founded at generation 0 with a greater contribution from S_1 than S_2 ($s_{1,0} = 0.8$, $s_{2,0} = 0.2$). Subsequently, the source populations contribute unequally to population H, with S_1 contributing more than S_2 ($s_1 = 0.2$, $s_2 = 0.0001$).

populations S_1 and S_2 to the hybrid population H are symmetric and constant across generations as before, but nonzero ($s_{1,g} = s_{2,g} \neq 0$ for all $g \geq 1$). In this case, because at each generation, the two source populations make equal contributions, the distribution of $H_{1,g}$ continues to be symmetric around $\frac{1}{2}$. Instead of being unimodal as in the previous case, however, it is now multimodal. This multimodality arises from the fact that in a scenario with continuing gene flow from the sources, new modes arise as the new immigrants mate with individuals whose admixture fractions lie near preexisting modes.

In Figure 2C, we consider an admixture process with a symmetric founding of population H as before ($s_{1,0} = s_{2,0} = \frac{1}{2}$), in which the subsequent contributions from the source populations S_1 and S_2 are nonzero and constant across generations, but with S_2 contributing more than S_1 at each generation ($0 \neq s_{1,g} \ll s_{2,g}$ for all $g \geq 1$). In this case, the distribution of $H_{1,g}$ is no longer symmetric around $\frac{1}{2}$. Instead, it is shifted toward smaller values after the founding of the hybrid population H. This pattern arises from the fact that in a scenario with continuing gene flow in which at each generation, many more individuals immigrate into H from S_2 than from S_1, matings between new immigrants and admixed individuals are more likely to occur with immigrants from S_2 than with immigrants from S_1. Thus, after the symmetric founding of population H, the probability of randomly drawing an individual in H with a high fraction of admixture from population S_1 is lower than the probability of drawing an individual with a low fraction of admixture from S_1.

Figure 2D considers an admixture process in which population S_1 contributes more than population S_2 to the founding of population H ($s_{1,0} > s_{2,0} \neq 0$), but with the same subsequent constant admixture process as in Figure 2C ($0 \neq s_{1,g} \ll s_{2,g}$ for all $g \geq 1$). In this case, the distribution of $H_{1,g}$ is no longer symmetric around $\frac{1}{2}$ at generation 1, but is shifted toward higher values of the admixture fraction from source population S_1. Nevertheless, as in Figure 2C, the distribution of $H_{1,g}$ shifts toward zero in the subsequent generations. As in Figure 2C, in each generation, admixed individuals in population H are more likely to mate with new immigrants from S_2 than with new immigrants from S_1.

Finally, in Figure 2E, we consider a process in which the source population S_1 contributes more than population S_2 to the hybrid population not only in the founding of population H ($s_{1,0} > s_{2,0} \neq 0$) but also in each subsequent generation ($0 \neq s_{1,g} \gg s_{2,g}$ for all $g \geq 1$). In this case, the distribution of $H_{1,g}$ is shifted toward high values of the admixture fraction from population S_1. Unlike in Figure 2, C and D, in Figure 2E, an individual in population H is more likely to mate with a new immigrant from S_1 than with a new immigrant from S_2 at each generation following the founding of population H. Thus, unlike in Figure 2, C and D, generation after generation, the probability of randomly drawing an individual in population H with a high fraction of admixture from S_1 is higher than that of drawing an individual with a low fraction of admixture from S_1.

This collection of scenarios illustrates three main points. First, if contributions to the admixed population occur only in the first generation, then the long-term level of admixture continues to reflect the initial conditions. Second, the same starting conditions can lead to quite different long-term patterns, depending on the subsequent contributions to the hybrid population. Third, with constant contributions at each generation, the starting conditions influence the speed with which the distribution of admixture tends toward its long-term distribution, but do not predict the qualitative form of this distribution.

Moments of the admixture fraction for a random individual in the hybrid population

Analysis of the moments of the distribution of admixture as a function of time g can provide a way of understanding features of the distribution and its determinants in the historical admixture process itself. We can utilize the recursion in Equations 1 and 2 to obtain recursions for the expectation, variance, and higher moments of $H_{1,g}$ as functions of g and $s_{1,i}$, $s_{2,i}$, and h_i, for $i = 1, 2, \ldots, g-1$. We first obtain a recursion for the expectation $E[H_{1,g}]$. Next, we generalize the method used for finding the expectation, and we obtain a recursion for the kth moment, $E[H_{1,g}^k]$. Using the case of $k = 2$, we obtain a recursion for the variance $V[H_{1,g}]$.

Expectation of $H_{1,g}$: Using the law of total expectation, we can obtain an expression for the expectation $E[H_{1,g}]$ as a function of conditional expectations for different possible pairs of parents Y for a random individual in population H at generation g:

$$E\left[H_{1,g}\right] = E_Y\left[E\left[H_{1,g}|Y\right]\right] = \sum P(Y=y)E\left[H_{1,g}|Y=y\right]. \quad (6)$$

$$y \in \left\{ \begin{array}{l} S_1S_1 \\ S_1H \\ S_1S_2 \\ HH \\ S_2H \\ S_2S_2 \end{array} \right\}$$

For the first generation, because parents cannot derive from population H, we have

$$\begin{aligned} E\left[H_{1,1}\right] &= P(Y=S_1S_1)E\left[H_{1,1}\,|\,Y=S_1S_1\right] \\ &\quad + P(Y=S_1S_2)E\left[H_{1,1}\,|\,Y=S_1S_2\right] \\ &\quad + P(Y=S_2S_2)E\left[H_{1,1}\,|\,Y=S_2S_2\right]. \end{aligned} \quad (7)$$

Using Equations 1 and 2,

$$\begin{aligned} E\left[H_{1,1}\right] &= s_{1,0}^2E[1] \\ &\quad + 2s_{1,0}s_{2,0}E\left[\frac{1}{2}\right] \\ &\quad + s_{2,0}^2E[0], \end{aligned} \quad (8)$$

and for all subsequent generations ($g \geq 2$),

$$\begin{aligned} E[H_{1,g}] = s_{1,g-1}^2 E[1] \\ + \quad 2s_{1,g-1}h_{g-1}E\left[\frac{H_{1,g-1}+1}{2}\right] \\ + \quad 2s_{1,g-1}s_{2,g-1}E\left[\frac{1}{2}\right] \\ + \quad h_{g-1}^2 E\left[\frac{H_{1,g-1}^{(1)}+H_{1,g-1}^{(2)}}{2}\right] \\ + \quad 2s_{2,g-1}h_{g-1}E\left[\frac{H_{1,g-1}}{2}\right] \\ + \quad s_{2,g-1}^2 E[0]. \end{aligned} \tag{9}$$

Recalling that for all $g \geq 0$, $s_{1,g} + s_{2,g} + h_g = 1$, $h_0 = 0$, and for all $g \geq 2$, $H_{1,g-1}^{(1)}$ and $H_{1,g-1}^{(2)}$ are IID random variables, we can simplify the recursion expression. For $g = 1$,

$$E[H_{1,1}] = s_{1,0}, \tag{10}$$

and for all subsequent generations ($g \geq 2$), we have

$$E[H_{1,g}] = s_{1,g-1} + h_{g-1}E[H_{1,g-1}]. \tag{11}$$

This result demonstrates that for a random individual in the hybrid population H, the expectation of the admixture fraction from population S_1 in one generation is a linear function of the corresponding expectation in the previous generation.

Moments of $H_{1,g}$: Using a similar computation to that employed in obtaining the recursion for the expected admixture, we can write recursions for higher moments of the admixture fraction ($E[H_{1,g}^k]$, for each $k \geq 1$). For the first generation ($g = 1$), we have for $k \geq 1$,

$$H_{1,1}^k = \begin{cases} 1^k & \text{if } Y = S_1S_1, \text{ with } P[Y = S_1S_1] = s_{1,0}^2 \\ \left(\frac{1}{2}\right)^k & \text{if } Y = S_1S_2, \text{ with } P[Y = S_1S_2] = 2s_{1,0}s_{2,0} \\ 0^k & \text{if } Y = S_2S_2, \text{ with } P[Y = S_2S_2] = s_{2,0}^2. \end{cases} \tag{12}$$

For all $g \geq 2$, we have

$$H_{1,g}^k = \begin{cases} 1^k & \text{if } Y = S_1S_1, \text{ with } P[Y = S_1S_1] = s_{1,g-1}^2 \\ \left(\frac{H_{1,g-1}+1}{2}\right)^k & \text{if } Y = S_1H, \text{ with } P[Y = S_1H] = 2s_{1,g-1}h_{g-1} \\ \left(\frac{1}{2}\right)^k & \text{if } Y = S_1S_2, \text{ with } P[Y = S_1S_2] = 2s_{1,g-1}s_{2,g-1} \\ \left(\frac{H_{1,g-1}^{(1)}+H_{1,g-1}^{(2)}}{2}\right)^k & \text{if } Y = HH, \text{ with } P[Y = HH] = h_{g-1}^2 \\ \left(\frac{H_{1,g-1}}{2}\right)^k & \text{if } Y = S_2H, \text{ with } P[Y = S_2H] = 2s_{2,g-1}h_{g-1} \\ 0^k & \text{if } Y = S_2S_2, \text{ with } P[Y = S_2S_2] = s_{2,g-1}^2, \end{cases} \tag{13}$$

where $H_{1,g-1}^{(1)}$ and $H_{1,g-1}^{(2)}$ represent IID random variables for the fractions of ancestry from source population S_1 for two hybrid individuals in generation $g - 1$.

Using the law of total expectation, for $k \geq 1$, we have for the first generation ($g = 1$)

$$\begin{aligned} E\left[H_{1,1}^k\right] = s_{1,0}^2 E[1^k] \\ + \quad 2s_{1,0}s_{2,0}E\left[\left(\frac{1}{2}\right)^k\right] \\ + \quad s_{2,0}^2 E[0^k]. \end{aligned} \tag{14}$$

For $g \geq 2$, we have

$$\begin{aligned} E\left[H_{1,g}^k\right] = s_{1,g-1}^2 E[1^k] \\ + \quad 2s_{1,g-1}h_{g-1}E\left[\left(\frac{H_{1,g-1}+1}{2}\right)^k\right] \\ + \quad 2s_{1,g-1}s_{2,g-1}E\left[\left(\frac{1}{2}\right)^k\right] \\ + \quad h_{g-1}^2 E\left[\left(\frac{H_{1,g-1}^{(1)}+H_{1,g-1}^{(2)}}{2}\right)^k\right] \\ + \quad 2s_{2,g-1}h_{g-1}E\left[\left(\frac{H_{1,g-1}}{2}\right)^k\right] \\ + \quad s_{2,g-1}^2 E[0^k]. \end{aligned} \tag{15}$$

Recalling that for all $g \geq 0$, $s_{1,g} + s_{2,g} + h_g = 1$, $h_0 = 0$, and for all $g \geq 2$, $H_{1,g-1}^{(1)}$ and $H_{1,g-1}^{(2)}$ are IID random variables, we can use the binomial theorem to obtain a simplified recursion for the moments of $H_{1,g}$. For the first generation, we have

$$E\left[H_{1,1}^k\right] = s_{1,0}^2 + \frac{s_{1,0}s_{2,0}}{2^{k-1}}. \tag{16}$$

For $g \geq 2$,

$$\begin{aligned} E\left[H_{1,g}^k\right] = s_{1,g-1}^2 + \frac{s_{1,g-1}h_{g-1}}{2^{k-1}}\left(\sum_{i=0}^{k}\binom{k}{i}E\left[H_{1,g-1}^i\right]\right) + \frac{s_{1,g-1}s_{2,g-1}}{2^{k-1}} \\ + \quad \frac{h_{g-1}^2}{2^k}\left(\sum_{i=0}^{k}\binom{k}{i}E\left[H_{1,g-1}^i\right]E\left[H_{1,g-1}^{k-i}\right]\right) + \frac{s_{2,g-1}h_{g-1}}{2^{k-1}}E\left[H_{1,g-1}^k\right]. \end{aligned} \tag{17}$$

Note that by simplifying Equation 16 with $k = 1$, we obtain for the first generation

$$E[H_{1,1}] = s_{1,0}(s_{1,0} + s_{2,0}) = s_{1,0}, \tag{18}$$

which matches Equation 10. Simplifying Equation 17 with $k = 1$ using the fact that $s_{1,g} + s_{2,g} + h_g = 1$ for all $g \geq 0$, for all subsequent generations ($g \geq 2$), we obtain

$$E[H_{1,g}] = s_{1,g-1} + h_{g-1}E[H_{1,g-1}], \tag{19}$$

which matches Equation 11.

Variance of $H_{1,g}$: When $k = 2$, Equations 16 and 17 provide a recursion relation for the second moment of $H_{1,g}$. For the first generation, because $s_{1,0} + s_{2,0} = 1$, we have

$$E\left[H_{1,1}^2\right] = s_{1,0}^2 + \frac{s_{1,0}s_{2,0}}{2} = \frac{s_{1,0}\left(s_{1,0}+1\right)}{2}. \tag{20}$$

For subsequent generations ($g \geq 2$), because $s_{1,g} + s_{2,g} + h_g = 1$ for all $g \geq 0$, we obtain

$$\begin{aligned} E\left[H_{1,g}^2\right] &= s_{1,g-1}^2 + \frac{s_{1,g-1}s_{2,g-1}}{2} + \frac{s_{1,g-1}h_{g-1}}{2}\left(1 + 2E\left[H_{1,g-1}\right] + E\left[H_{1,g-1}^2\right]\right) \\ &\quad + \frac{h_{g-1}^2}{2}\left(E\left[H_{1,g-1}^2\right] + \left(E\left[H_{1,g-1}\right]\right)^2\right) + \frac{s_{2,g-1}h_{g-1}}{2}E\left[H_{1,g-1}^2\right] \\ &= \frac{s_{1,g-1}\left(s_{1,g-1}+1\right)}{2} + h_{g-1}\left(s_{1,g-1} + \frac{h_{g-1}}{2}E\left[H_{1,g-1}\right]\right)E\left[H_{1,g-1}\right] \\ &\quad + \frac{h_{g-1}}{2}E\left[H_{1,g-1}^2\right]. \end{aligned} \tag{21}$$

With the relationship $V[H_{1,g}] = E[H_{1,g}^2] - (E[H_{1,g}])^2$, and using Equations 10, 11, 20, and 21, we obtain a recursion for the variance of $H_{1,g}$. For the first generation ($g = 1$), we have

$$V\left[H_{1,1}\right] = \frac{s_{1,0}\left(1-s_{1,0}\right)}{2}, \tag{22}$$

and for $g \geq 2$,

$$\begin{aligned} V\left[H_{1,g}\right] &= \frac{s_{1,g-1}\left(1-s_{1,g-1}\right)}{2} - s_{1,g-1}h_{g-1}E\left[H_{1,g-1}\right] \\ &\quad + \frac{h_{g-1}\left(1-h_{g-1}\right)}{2}\left(E\left[H_{1,g-1}\right]\right)^2 + \frac{h_{g-1}}{2}V\left[H_{1,g-1}\right]. \end{aligned} \tag{23}$$

This recursion for the variance of the admixture fraction utilizes the variance in the previous generation, along with the expectation in the previous generation and its square.

Special Case: Constant Admixture after the Founding of the Hybrid Population

Using our recursions for the moments of the admixture fraction $H_{1,g}$, we can examine particular cases in which $s_{1,g}$, $s_{2,g}$, and h_g are specified. Here we consider a special case that reflects a constant process in which admixture occurs in the same way from one generation to the next after the founding of the hybrid population. In this section, we specify that for all $g \geq 1$, all introgression parameters are constant in time after the founding of population H ($s_{1,g} = s_1$, $s_{2,g} = s_2$, and $h_g = h$ for all $g \geq 1$). We first consider a case in which no admixture from source populations S_1 and S_2 occurs after the founding of the hybrid population.

A single admixture event

Suppose that source populations S_1 and S_2 do not contribute to the hybrid population after its founding ($s_1 = s_2 = 0$, and $h = 1$). As before, because at generation 0 the hybrid population is not yet formed, we specify that $h_0 = 0$ and $s_{1,0} + s_{2,0} = 1$, with $s_{1,0}$ and $s_{2,0}$ both taking values in (0, 1).

Expectation of $H_{1,g}$: Under this scenario, we can simplify Equations 10 and 11 for the expected admixture from population S_1. Because $s_1 = s_2 = 0$ and $h = 1$, for all $g \geq 1$, we have

$$E\left[H_{1,g}\right] = s_{1,0}. \tag{24}$$

When admixture occurs only in the initial generation, the expected admixture fraction for a random individual in the hybrid population at any generation depends only on the initial contribution from source population S_1.

Variance of $H_{1,g}$: Using Equations 22 and 23, $V[H_{1,g}]$ follows the recursion relation of a geometric sequence with ratio 1/2 and initial value $V[H_{1,1}] = s_{1,0}(1 - s_{1,0})/2$. Therefore, for $g \geq 1$,

$$V\left[H_{1,g}\right] = \frac{s_{1,0}\left(1-s_{1,0}\right)}{2^g}. \tag{25}$$

The variance decreases monotonically as a function of g and is smaller when the initial contribution $s_{1,0}$ from source population S_1 is farther away from $\frac{1}{2}$.

The scenario in Figure 2A, in which $s_{1,0} = s_{2,0} = \frac{1}{2}$ and $s_{1,g} = s_{2,g} = 0$ for all $g \geq 1$, provides an example of the setting considered here. In Figure 2A, the distribution of the admixture fraction for a random individual in H becomes increasingly concentrated near $\frac{1}{2}$ as time progresses. As predicted by Equation 24, the mean admixture is constant over time with a value of $\frac{1}{2}$. As predicted by Equation 25, the variance decreases over time; it eventually approaches zero, so that the admixture fraction for a random individual approaches the mean. This phenomenon can be attributed to the fact that except during the founding event, each mating in the population involves two individuals from the hybrid population itself; no new source of admixture draws the admixture fraction toward extreme values of 0 or 1. Thus, with admixture values equal to the mean of those of their parents, offspring individuals are likely to have intermediate admixture within the unit interval.

It is noteworthy that if admixture occurs in a single event, then Equations 24 and 25 provide a basis for estimating the time of the event from the observed mean and variance of admixture. Given mean M and variance V (with $V \neq 0$), Equations 24 and 25 yield

$$g = \frac{\ln\left[M(1-M)\right] - \ln V}{\ln 2}. \tag{26}$$

It can be seen from Equation 26 that for a fixed mean, a smaller variance indicates a larger value of g and therefore a longer time since admixture, and for a fixed variance,

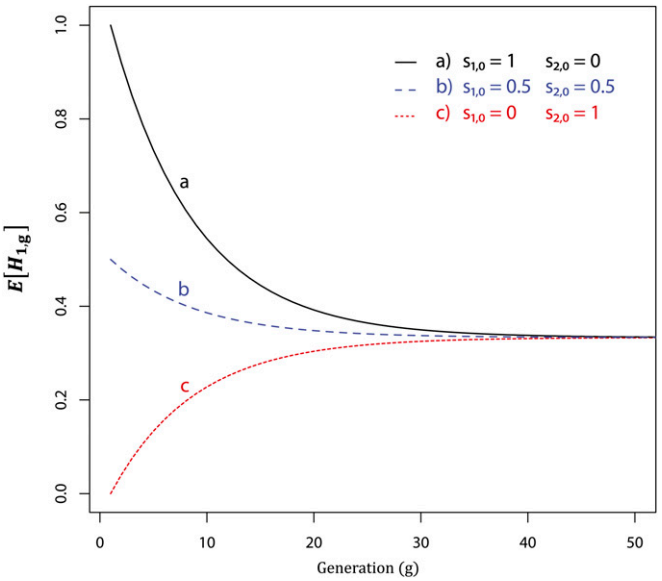

Figure 3 Founding effect: expectation of the admixture fraction from source population S_1 ($E[H_{1,g}]$, Equation 30) for a random individual in the hybrid population H, when the admixture process is constant over time. Three scenarios are shown: population H founded exclusively by source population S_1, population H founded by both source populations S_1 and S_2 in equal proportions, and population H founded exclusively by source population S_2. The subsequent admixture process is identical among three scenarios with different starting conditions and is constant over time after the founding of population H at the first generation: $s_1 = 0.04$ and $s_2 = 0.08$. For all three scenarios, using Equation 31, the long-term limit of the admixture proportion from population S_1 is 1/3.

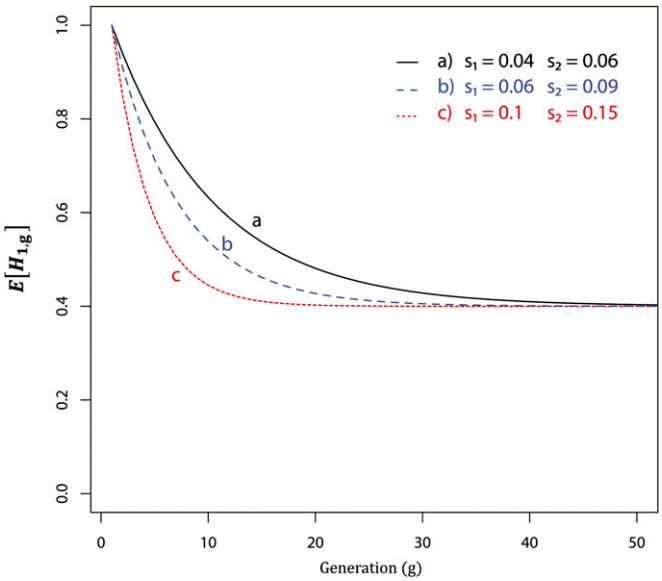

Figure 4 Ratio effect: expectation of the admixture fraction from source population S_1 ($E[H_{1,g}]$, Equation 30) for a random individual in the hybrid population H, when the admixture process is constant over time. Three scenarios are shown with the same initial founding event for population H ($s_{1,0} = 1$, $s_{2,0} = 0$). After this founding event, the three scenarios have different proportions of descent from source populations S_1 and S_2 in H (constant at each generation), but with the same ratio $s_1/s_2 = 2/3$. For all three scenarios, using Equation 31, the long-term limit of the admixture proportion from population S_1 is 2/5.

a smaller value of $M(1 - M)$ indicates a shorter time since admixture.

Nonzero combined contribution from the source populations at each generation

In this section, we consider values of s_1 and s_2 in [0, 1] and values of h in (0, 1). As before, because at generation 0 the hybrid population is not yet formed, $h_0 = 0$ and $s_{1,0} + s_{2,0} = 1$. This set of assumptions corresponds to a process with a nonzero combined contribution of populations S_1 and S_2 to H in each generation ($s_1 + s_2 \neq 0$ because $h \neq 1$), although we do allow one or the other contribution to be zero ($s_1 = 0$ and $s_2 \neq 0$ or $s_1 \neq 0$ and $s_2 = 0$). The contribution of population H to itself in each generation is nonzero ($h \neq 0$).

Expectation of $H_{1,g}$: Applying Equations 10 and 11, the recursion relation for $E[H_{1,g}]$ can be simplified. For the first generation ($g = 1$), we have

$$E[H_{1,1}] = s_{1,0}. \tag{27}$$

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

$$E[H_{1,g}] = s_1 + hE[H_{1,g-1}]. \tag{28}$$

This equation is a nonhomogeneous first-order recurrence of the form

$$E[H_{1,g}] = \psi(g) + \lambda E[H_{1,g-1}], \tag{29}$$

with initial condition $E[H_{1,1}] = s_{1,0}$, where $\psi(g) = s_1$ and $\lambda = h$. Because we consider an admixture process that is constant from one generation to the next and we assume $h \neq 0$ and $h \neq 1$, we can apply Theorem 3.1.2 of Cull *et al.* (2005) to Equation 29 to obtain the unique solution for $E[H_{1,g}]$:

$$E[H_{1,g}] = \begin{cases} s_{1,0}, & g = 1 \\ s_{1,0}h^{g-1} + s_1\dfrac{1-h^{g-1}}{1-h}, & g \geq 2. \end{cases} \tag{30}$$

Figures 3 and 4 illustrate the expected admixture fraction as a function of g under constant admixture, as determined in Equation 30. In Figure 3, we can see that in three admixture scenarios with different parameter values for the founding of the hybrid population H, but with identical introgression parameters constant in the subsequent generations, the expected admixture fraction from the source population S_1 approaches the same long-term limit. Moreover, in Figure 4, considering three scenarios with identical founding parameter values ($s_{1,0}$ and $s_{2,0}$), but different values for the introgression parameters s_1 and s_2 in the subsequent generations with identical ratios, s_1/s_2, the expected admixture fraction also approaches the same long-term limit.

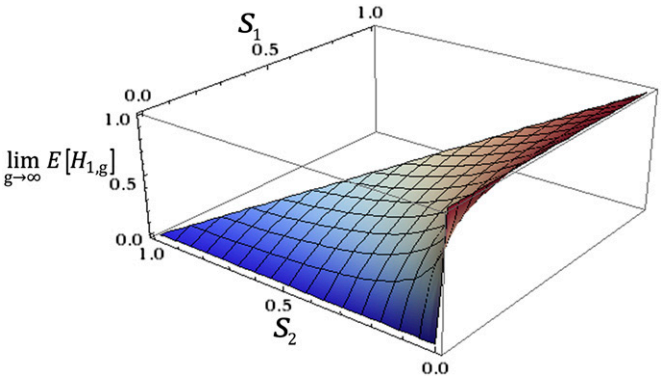

Figure 5 Long-term limit of the expectation of the admixture fraction from source population S_1 ($\lim_{g\to\infty} E[H_{1,g}]$, Equation 31) as a function of the introgression parameters s_1 and s_2 when the admixture process is constant over time.

Using Equation 30 and the relation $s_1 + s_2 + h = 1$ with $h \in (0, 1)$, we can compute the long-term limit of $E[H_{1,g}]$ as $g \to \infty$:

$$\lim_{g\to\infty} E\left[H_{1,g}\right] = \frac{s_1}{1-h} = \frac{s_1}{s_1+s_2}. \tag{31}$$

Equation 31 demonstrates that the starting conditions ($s_{1,0}$ and $s_{2,0}$) for the founding of the hybrid population H do not influence the long-term limiting expectation, as observed in Figure 3. The limiting expected admixture in Equation 31 can be rewritten as $1 - 1/(1 + s_1/s_2)$, showing that the limiting expectation is determined only by the ratio of the constant contributions from populations S_1 and S_2, as observed in Figure 4.

Using Equation 31, we can plot the long-term limit of the expected admixture fraction from source population S_1 as a function of the introgression parameters s_1 and s_2 (Figure 5). When the admixture process is constant over time, for a given value of s_2, the long-term expectation of the admixture fraction from the source population S_1 increases monotonically with s_1. Because the long-term limit depends only on the ratio s_1/s_2, different introgression proportions as well as different founding scenarios for population H can lead, in the long-term, to the same expected admixture fractions in H.

Variance of $H_{1,g}$: When the admixture process is constant across generations, we can employ the same methods used for obtaining the expectation of $H_{1,g}$ to obtain a solution for $E[H^2_{1,g}]$. In this case, for the first generation ($g = 1$), Equation 20 gives

$$E\left[H^2_{1,1}\right] = \frac{s_{1,0}\left(s_{1,0}+1\right)}{2}. \tag{32}$$

For $g \geq 2$, Equation 21 gives

$$E\left[H^2_{1,g}\right] = \frac{s_1(s_1+1)}{2} + s_1 h E\left[H_{1,g-1}\right] + \frac{h^2}{2}E\left[H_{1,g-1}\right]^2 + \frac{h}{2}E\left[H^2_{1,g-1}\right]. \tag{33}$$

As was true in the case of Equation 28, this equation is a nonhomogeneous first-order recurrence with the form

$$E\left[H^2_{1,g}\right] = \psi(g) + \lambda E\left[H^2_{1,g-1}\right]. \tag{34}$$

Here, the initial condition is $E[H^2_{1,1}] = s_{1,0}(s_{1,0}+1)/2$, $\lambda = h/2$, and for all $g \geq 2$,

$$\psi(g) = \frac{s_1(s_1+1)}{2} + s_1 h E\left[H_{1,g-1}\right] + \frac{h^2}{2}E\left[H_{1,g-1}\right]^2. \tag{35}$$

Using Equation 30, we can simplify Equation 35 for all $g \geq 2$, to obtain

$$\psi(g) = \frac{s_1(s_1+1)}{2} + s_1 h\left(s_{1,0}h^{g-2} + s_1\frac{1-h^{g-2}}{1-h}\right) + \frac{h^2}{2}\left(s_{1,0}h^{g-2} + s_1\frac{1-h^{g-2}}{1-h}\right)^2. \tag{36}$$

Because $h \neq 0$ and $h \neq 1$, Theorem 3.1.2 of Cull *et al.* (2005) applies in the same way as in the computation of $E[H_{1,g}]$, producing a unique solution for $E[H^2_{1,g}]$:

$$E\left[H^2_{1,g}\right] = \begin{cases} \frac{s_{1,0}\left(s_{1,0}+1\right)}{2}, & g = 1 \\ \frac{s_{1,0}\left(s_{1,0}+1\right)}{2}\left(\frac{h}{2}\right)^{g-1} & \\ \quad + \sum_{i=2}^{g}\left(\frac{s_1(s_1+1)}{2} + s_1 h\left(s_{1,0}h^{i-2} + s_1\frac{1-h^{i-2}}{1-h}\right)\right. & \\ \quad \left. + \frac{h^2}{2}\left(s_{1,0}h^{i-2} + s_1\frac{1-h^{i-2}}{1-h}\right)^2\right)\left(\frac{h}{2}\right)^{g-i}, & g \geq 2. \end{cases} \tag{37}$$

Decomposing the summation and summing separate geometric series, we obtain

$$E\left[H^2_{1,g}\right] = \begin{cases} \frac{s_{1,0}\left(s_{1,0}+1\right)}{2}, & g = 1 \\ A_1 + A_2 h^{g-1} + \left(A_3 + A_4\sum_{i=1}^{g-1}(2h)^i\right)\left(\frac{h}{2}\right)^{g-1}, & g \geq 2, \end{cases} \tag{38}$$

where

$$A_1 = \frac{s_1\left((1-h)^2 + s_1\right)}{(2-h)(1-h)^2}, \tag{39}$$

$$A_2 = 2s_1\left(s_{1,0} - \frac{s_1}{1-h}\right)\left(\frac{1}{1-h}\right), \tag{40}$$

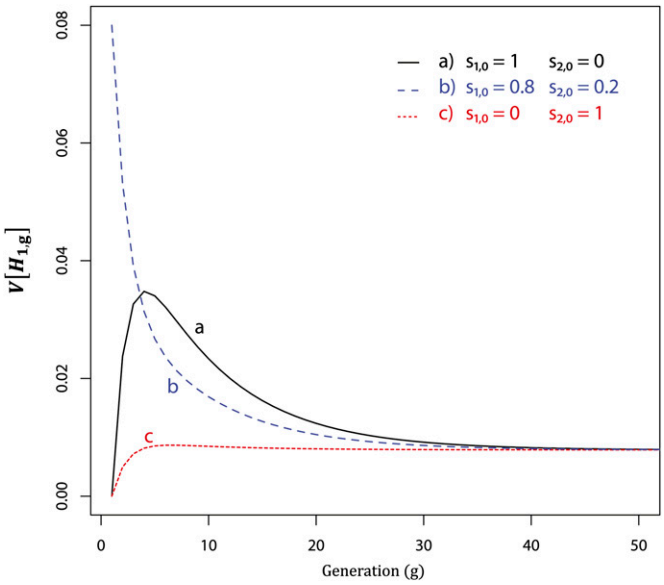

Figure 6 Founding effect: variance of the admixture fraction from source population S_1 ($V[H_{1,g}]$, Equation 43) for a random individual in the hybrid population H, when the admixture process is constant over time. Three scenarios are shown: population H founded exclusively by source population S_1, population H founded by both source populations S_1 and S_2 in unequal proportions, with population S_1 contributing more than population S_2 ($s_{1,0} = 0.8$, $s_{2,0} = 0.2$), and population H founded exclusively by source population S_2. The subsequent admixture process is identical among three scenarios with different starting conditions and is constant over time after the founding of population H at the first generation: $s_1 = 0.01$ and $s_2 = 0.05$. For all three scenarios, using Equation 47, the long-term limit of the variance of the admixture proportion from population S_1 is $5/636 \approx 0.008$.

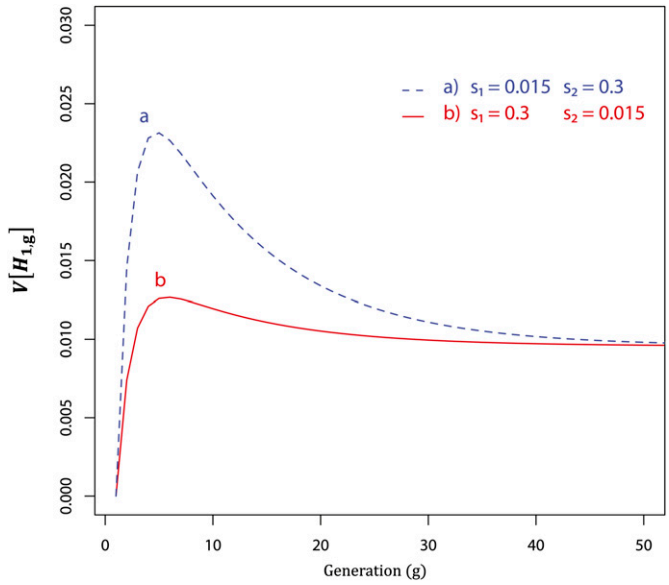

Figure 7 Opposite admixture processes: variance of the admixture fraction from source population S_1 ($V[H_{1,g}]$, Equation 43) for a random individual in the hybrid population H, when the admixture process is constant over time. Two scenarios are shown with the same initial founding event ($s_{1,0} = 1$, $s_{2,0} = 0$). After this founding event, the scenarios have opposite proportions of descent from source populations S_1 and S_2 in H, constant at each generation. For both scenarios, using Equation 47, the long-term limit of the admixture proportion from population S_1 is $60/5523 \approx 0.011$.

$$A_3 = \frac{2s_1(s_1 - s_{1,0}h)}{1-h} + \frac{s_1^2 h}{(1-h)^2} + \frac{s_{1,0}(s_{1,0}+1)}{2} - \frac{s_1(s_1+1)}{2-h} - 2s_1 s_{1,0}, \tag{41}$$

$$A_4 = \frac{1}{2}\left(s_{1,0}^2 + \frac{s_1}{1-h}\left(\frac{s_1}{1-h} - 2s_{1,0}\right)\right). \tag{42}$$

With the relationship $V[H_{1,g}] = E[H_{1,g}^2] - (E[H_{1,g}])^2$, and using Equations 30 and 38, we obtain the variance of $H_{1,g}$:

$$V[H_{1,g}] = \begin{cases} \frac{s_{1,0}(1-s_{1,0})}{2}, & g = 1 \\ A_1 + A_2 h^{g-1} + A_3\left(\frac{h}{2}\right)^{g-1} & \\ \quad + A_4\left(\frac{h}{2}\right)^{g-1}\sum_{i=1}^{g-1}(2h)^i - \left(s_{1,0}h^{g-1} + s_1\frac{1-h^{g-1}}{1-h}\right)^2, & g \geq 2. \end{cases} \tag{43}$$

We can simplify Equation 43 to obtain expressions for $V[H_{1,g}]$ without the summation $\sum_{i=1}^{g-1}(2h)^i$. For all values of h in (0, 1) with $h \neq \frac{1}{2}$, by summing the geometric series from Equation 43,

$$V[H_{1,g}] = \begin{cases} \frac{s_{1,0}(1-s_{1,0})}{2}, & g = 1 \\ A_1 + \left(A_2 - A_5 h^{g-1}\right)h^{g-1} + (A_3 + A_5)\left(\frac{h}{2}\right)^{g-1} & \\ \quad - \left(s_{1,0}h^{g-1} + s_1\frac{1-h^{g-1}}{1-h}\right)^2, & g \geq 2, \end{cases} \tag{44}$$

where $A_5 = 2hA_4/(1-2h)$. For $h = \frac{1}{2}$, Equation 43 gives

$$V[H_{1,g}] = \begin{cases} \frac{s_{1,0}(1-s_{1,0})}{2}, & g = 1 \\ \frac{2s_1(1-2s_1)}{3} + \left(\frac{2s_1(2s_1-1)}{3} + \frac{s_{1,0}(1-s_{1,0})}{2}\right. & \\ \quad \left. + \left(\frac{s_{1,0}^2}{2} + 2s_1(s_1 - s_{1,0})\right)\right. & \\ \quad \left. \times (g-1)\right)\left(\frac{1}{4}\right)^{g-1}, & g \geq 2. \end{cases} \tag{45}$$

Figures 6 and 7 illustrate the variance of the admixture fraction under the special case of constant admixture, computing Equation 43 for different sets of values of the introgression parameters. Figure 6 shows that in three scenarios with different founding parameter values ($s_{1,0}$ and $s_{2,0}$), because the admixture process is constant over time and identical among the scenarios, the variance of the admixture fraction from one of the source populations approaches the same long-term limit. In Figure 7, considering two admixture

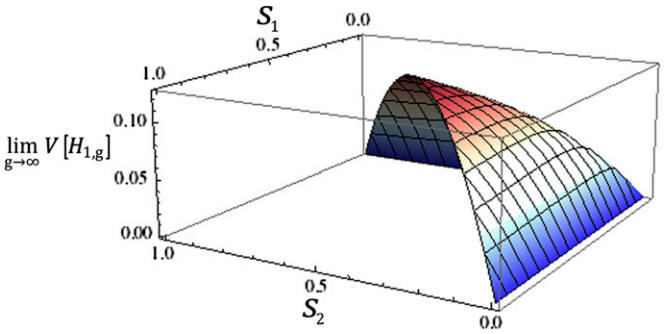

Figure 8 Long-term limit of the variance of the admixture fraction from source population S_1 ($\lim_{g\to\infty} V[H_{1,g}]$, Equation 47) as a function of the introgression parameters s_1 and s_2 when the admixture process is constant over time.

scenarios with identical founding events but opposite constant admixture processes, the variance of the admixture fraction also approaches the same limit.

In Figures 6 and 7, we can see that for some sets of values of $s_{1,0}$, s_1, and s_2, the variance of the admixture fraction from one of the source populations increases monotonically from the beginning of the admixture process until it reaches a maximal value and then decreases monotonically to its long-term limit. In these cases, at the beginning of the admixture process, the source populations introduce considerable variance to the distribution of the admixture fraction for a random individual in H. After a certain amount of time, the proportion of matings that involve members of the hybrid population H with similar admixture fractions increases, reducing the proportion of matings that generate offspring admixture fractions at opposite extremes. Additional matings then occur among individuals with similar admixture, ultimately decreasing the variance of the admixture fraction until $V[H_{1,g}]$ approaches its long-term limit.

Because $h \neq 0$ and $h \neq 1$, we can compute the long-term limit of $V[H_{1,g}]$ as $g \to \infty$ using Equation 43. We obtain, for all values of h in (0, 1),

$$\lim_{g\to\infty} V[H_{1,g}] = \frac{(s_1/(1-h))^2 + s_1}{2-h} - \left(\frac{s_1}{1-h}\right)^2 = \frac{s_1}{2-h}\left(1 - \frac{s_1}{1-h}\right). \quad (46)$$

The starting conditions do not influence the long-term limit, as observed in Figure 6.

Recalling that $s_1 + s_2 + h = 1$, an alternative representation for Equation 46 is

$$\lim_{g\to\infty} V[H_{1,g}] = \left(\frac{s_1}{s_1+s_2}\right)\left(\frac{s_2}{s_1+s_2}\right)\left(\frac{s_1+s_2}{1+s_1+s_2}\right). \quad (47)$$

It is possible to see from Equation 47 that if $s_1 + s_2$ is fixed, then the limiting variance is greater when both source populations contribute similarly to the hybrid population ($s_1 \approx s_2$) than when one source population contributes more than the other ($s_1 \gg s_2$ or $s_2 \gg s_1$). Additionally, for a fixed ratio s_1/s_2, the variance is greater when the combined contribution from both source populations, $s_1 + s_2$, is greater. This result is sensible, as continuing contributions from the source populations generate individuals with admixture fractions at opposite extremes, thereby increasing the variance of admixture fractions.

Using Equation 47, we can plot the long-term limit of the variance of admixture proportions as a function of s_1 and s_2 (Figure 8). Figure 8 illustrates that the long-term limit of $V[H_{1,g}]$ is greater when $s_1 = s_2$ and $s_1 + s_2 \approx 1$ (and thus $h \approx 0$). This scenario corresponds to an admixture process in which the admixed individuals in population H contribute little to the next generation, and the population H is largely founded anew at each generation from the source populations S_1 and S_2, with identical proportions.

When $s_1 + s_2 \to 0$, with s_1/s_2 held constant, $h \to 1$ and Equation 47 gives

$$\lim_{s_1+s_2\to 0} \lim_{g\to\infty} V[H_{1,g}] = 0. \quad (48)$$

This scenario corresponds to an admixture process in which populations S_1 and S_2 found the hybrid population H at the first generation and contribute little in subsequent generations. It tends toward the special case in which the source populations do not further contribute to the hybrid population after the founding event. The result in Equation 48 is consistent with the corresponding limit of Equation 25 for the case of no continuing admixture.

Considering our results on the expectation and variance of the admixture fraction together, although different admixture proportions that are constant and nonzero across generations can lead in the long-term to the same expected fraction of admixture, such parameter values can produce different variances. The long-term limiting expectation and variance do not depend on the conditions of the founding event of the hybrid population H; they depend only on the subsequent constant admixture process.

Discussion

Our study provides a theoretical framework for analyzing complex admixture processes that involve dynamic contributions of mutually isolated source populations to the ancestry of a hybrid population. Using our mechanistic approach, we have analytically derived recursions for the expectation, variance, and higher moments of the admixture fractions in a hybrid population. In the special case of constant admixture, we have solved the recursions and analyzed the behavior of the expectation and variance.

An important observable quantity that can be estimated in modern admixed populations and used for understanding historical aspects of the admixture process is the mean admixture fraction from a source population. For a hybrid population, this quantity provides a simple summary of its overall level of admixture. However, when a hybrid population is founded in a single admixture event, we have found that the mean admixture fraction is constant across generations and is therefore uninformative about the time since the founding of the hybrid population. When the source populations contribute in a constant manner to the hybrid population after the founding event, very different admixture processes can produce identical expected admixture fractions in the long-term.

The behavior of the variance of the admixture fraction is more complex than that of the expectation. First, the variance is not constant in time, and therefore it does contain information about the time since the founding event. Second, the limiting variance can differ quite substantially for processes with the same limiting expectation, with the limit depending on the magnitude of the ongoing contributions from the source populations. Third, a low variance is characteristic of an admixture process that occurred as a single event, whereas higher variance occurs when admixture is ongoing. These results suggest that in addition to the mean admixture, other easily measured quantities such as the variance and higher moments of the admixture fraction are likely to be informative, together with the mean, in statistical procedures for estimating the parameters of the historical admixture model that gives rise to a hybrid population.

Numerous statistical methods have been developed to estimate the admixture proportions from given source populations in hybrid populations using, for instance, maximum likelihood (Wang 2003; Tang *et al.* 2005; Alexander *et al.* 2009), least squares (Roberts and Hiorns 1965; Long and Smouse 1983), coalescence times (Bertorelle and Excoffier 1998), Bayesian approaches (Pritchard *et al.* 2000; Corander *et al.* 2003; Patterson *et al.* 2004), and principal components analysis (Paschou *et al.* 2007; McVean 2009; Bryc *et al.* 2010). Although many of these methods do estimate a composite parameter representing the time since initial admixture, they generally do not use a full mechanistic approach and have largely not tried to reconstruct the history of the admixture process.

Our model incorporates a general variation over time in the relative contributions of the source populations to the hybrid population. Owing to the potentially large number of parameters in a general case with arbitrary changes in admixture with time, it is unclear when the full history of admixture will be identifiable from genetic data. Indeed, as we have focused on the mean and variance of admixture in special cases of constant admixture processes, it is also uncertain how much information will be available for estimation from higher moments in a complex case with more parameters. However, our model is flexible enough to accommodate reductions in the number of parameters through assumptions of constant admixture over periods of many generations or over the entire history of the model. It is thus likely that identifiability can be achieved at least in some cases.

The initial theoretical framework that we have developed can be expanded to account for additional aspects of the admixture process for hybrid populations. For instance, in File S1, we extend the approach to consider m potential source populations, deriving general expressions for the moments of the random fraction of admixture originating from any specific one of the m source populations. However, we have not modeled sex-specific contributions from the source populations or assortative mating between hybrid individuals on the basis of their admixture fractions (Risch *et al.* 2009). Further, while we have considered the distribution of the admixture fraction across individuals in a hybrid population, we have studied admixture only pointwise in the genome, and we have not investigated variation in admixture across the genome of a random individual. The distribution of the length of chromosomal segments ultimately tracing to a particular source population, and other variables that could potentially be examined in a recombination-based model, could provide a useful additional set of quantities to consider beyond those available in our current formulation.

Finally, we have not accounted for genetic drift in the founding populations over the course of the admixture process, a phenomenon that can confound the accurate estimation of admixture proportions (Long 1991). In the future, all of these factors can be incorporated by extending our initial mechanistic admixture model. The various extensions will make it possible to draw more information from genetic data to shed light on the complex mechanisms underlying observed genetic variation in hybrid individuals and populations.

Acknowledgments

The authors thank Ethan Jewett and Mike DeGiorgio for useful discussions and comments. This work was supported in part by U.S. National Institutes of Health grants R01 GM081441 and R01 HG005855, by U.S. National Science Foundation grant BCS-1147534, and by the Burroughs Welcome Fund.

Literature Cited

Alexander, D. H., J. Novembre, and K. Lange, 2009 Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19: 1655–1664.

Baird, S. J. E., 2006 Phylogenetics: Fisher's markers of admixture. Heredity 97: 81–83.

Bedoya, G., P. Montoya, J. Garcia, I. Soto, S. Bourgeois *et al.*, 2006 Admixture dynamics in Hispanics: a shift in the nuclear

genetic ancestry of a South American population isolate. Proc. Natl. Acad. Sci. USA 103: 7234–7239.
Bercovici, S., and D. Geiger, 2009 Inferring ancestries efficiently in admixed populations with linkage disequilibrium. J. Comput. Biol. 16: 1141–1150.
Bernstein, F., 1931 Die geographische Verteilung der Bludgruppen und ihre anthropologische Bedeutung, pp. 227–243 in *Comitato Italiano per o Studio dei Problemi della Populazione*. Instituto Poligraphico dello Stato, Rome.
Bertorelle, G., and L. Excoffier, 1998 Inferring admixture proportions from molecular data. Mol. Biol. Evol. 15: 1298–1311.
Bonilla, C., G. Gutierrez, E. J. Parra, C. Kline, and M. D. Shriver, 2005 Admixture analysis of a rural population of the state of Guerrero, Mexico. Am. J. Phys. Anthropol. 128: 861–869.
Briscoe, D., J. C. Stephens, and S. J. O'Brien, 1994 Linkage disequilibrium in admixed populations—applications in gene-mapping. J. Hered. 85: 59–63.
Bryc, K., A. Auton, M. R. Nelson, J. R. Oksenberg, S. L. Hauser *et al.*, 2010 Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc. Natl. Acad. Sci. USA 107: 786–791.
Buerkle, C. A., and C. Lexer, 2008 Admixture as the basis for genetic mapping. Trends Ecol. Evol. 23: 686–694.
Cavalli-Sforza, L. L., and W. F. Bodmer, 1971 *The Genetics of Human Populations*. W. H. Freeman, San Francisco.
Chaix, R., L. Quintana-Murci, T. Hegay, M. F. Hammer, Z. Mobasher *et al.*, 2007 From social to genetic structures in central Asia. Curr. Biol. 17: 43–48.
Chakraborty, R., and K. M. Weiss, 1988 Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc. Natl. Acad. Sci. USA 85: 9119–9123.
Chakraborty, R., M. I. Kamboh, M. Nwankwo, and R. E. Ferrell, 1992 Caucasian genes in American blacks: new data. Am. J. Hum. Genet. 50: 145–155.
Chikhi, L., M. W. Bruford, and M. A. Beaumont, 2001 Estimation of admixture proportions: a likelihood-based approach using Markov chain Monte Carlo. Genetics 158: 1347–1362.
Corander, J., P. Waldmann, and M. J. Sillanpää, 2003 Bayesian analysis of genetic differentiation between populations. Genetics 163: 367–374.
Cull, P., M. Flahive, and R. Robson, 2005 *Difference Equations: From Rabbits to Chaos*. Springer-Verlag, New York.
Ewens, W. J., and R. S. Spielman, 1995 The transmission/disequilibrium test: history, subdivision, and admixture. Am. J. Hum. Genet. 57: 455–464.
Falush, D., M. Stephens, and J. K. Pritchard, 2003 Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567–1587.
Guo, W., W. K. Fung, N. Shi, and J. Guo, 2005 On the formula for admixture linkage disequilibrium. Hum. Hered. 60: 177–180.
Halder, I., and M. D. Shriver, 2003 Measuring and using admixture to study the genetics of complex diseases. Hum. Genomics 1: 52–62.
Halder, I., B.-Z. Yang, H. R. Kranzler, M. B. Stein, M. D. Shriver *et al.*, 2009 Measurement of admixture proportions and description of admixture structure in different U.S. populations. Hum. Mutat. 30: 1299–1309.
Hoggart, C. J., M. D. Shriver, R. A. Kittles, D. G. Clayton, and P. M. McKeigue, 2004 Design and analysis of admixture mapping studies. Am. J. Hum. Genet. 74: 965–978.
Lohmueller, K. E., C. D. Bustamante, and A. G. Clark, 2011 Detecting directional selection in the presence of recent admixture in African-Americans. Genetics 187: 823–835.
Long, J. C., 1991 The genetic structure of admixed populations. Genetics 127: 417–428.
Long, J. C., and P. E. Smouse, 1983 Intertribal gene flow between the Ye'cuana and Yanomama: genetic analysis of an admixed village. Am. J. Phys. Anthropol. 61: 411–422.
McKeigue, P. M., 1998 Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. Am. J. Hum. Genet. 63: 241–251.
McKeigue, P. M., 2005 Prospects for admixture mapping of complex traits. Am. J. Hum. Genet. 76: 1–7.
McVean, G., 2009 A genealogical interpretation of principal components analysis. PLoS Genet. 5: e1000686.
Oleksyk, T. K., M. W. Smith, and S. J. O'Brien, 2010 Genome-wide scans for footprints of natural selection. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365: 185–205.
Parra, E. J., A. Marcini, J. Akey, J. Martinson, M. A. Batzer *et al.*, 1998 Estimating African American admixture proportions by use of population-specific alleles. Am. J. Hum. Genet. 63: 1839–1851.
Parra, E. J., R. A. Kittles, G. Argyropoulos, C. L. Pfaff, K. Hiester *et al.*, 2001 Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. Am. J. Phys. Anthropol. 114: 18–29.
Parra, F. C., R. C. Amado, J. R. Lambertucci, J. Rocha, C. M. Antunes *et al.*, 2003 Color and genomic ancestry in Brazilians. Proc. Natl. Acad. Sci. USA 100: 177–182.
Paschou, P., E. Ziv, E. G. Burchard, S. Choudhry, W. Rodriguez-Cintron *et al.*, 2007 PCA-correlated SNPs for structure identification in worldwide human populations. PLoS Genet. 3: 1672–1686.
Patterson, N., N. Hattangadi, B. Lane, K. E. Lohmueller, D. A. Hafler *et al.*, 2004 Methods for high-density admixture mapping of disease genes. Am. J. Hum. Genet. 74: 979–1000.
Pfaff, C. L., E. J. Parra, C. Bonilla, K. Hiester, P. M. McKeigue *et al.*, 2001 Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am. J. Hum. Genet. 68: 198–207.
Price, A. L., A. Tandon, N. Patterson, K. C. Barnes, N. Rafaels *et al.*, 2009 Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLoS Genet. 5: e1000519.
Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics 155: 945–959.
Reich, D., and N. Patterson, 2005 Will admixture mapping work to find disease genes? Philos. Trans. R. Soc. Lond. B Biol. Sci. 360: 1605–1607.
Risch, N., S. Choudhry, M. Via, A. Basu, R. Sebro *et al.*, 2009 Ancestry-related assortative mating in Latino populations. Genome Biol. 10: R132.
Roberts, D. F., and R. W. Hiorns, 1965 Methods of analysis of the genetic composition of a hybrid population. Hum. Biol. 37: 38–43.
Sankararaman, S., S. Sridhar, G. Kimmel, and E. Halperin, 2008 Estimating local ancestry in admixed populations. Am. J. Hum. Genet. 82: 290–303.
Seldin, M. F., B. Pasaniuc, and A. L. Price, 2011 New approaches to disease mapping in admixed populations. Nat. Rev. Genet. 12: 523–528.
Seldin, M. F., C. Tian, R. Shigeta, H. R. Scherbarth, G. Silva *et al.*, 2007 Argentine population genetic structure: large variance in Amerindian contribution. Am. J. Phys. Anthropol. 132: 455–462.
Silva-Zolezzi, I., A. Hidalgo-Miranda, J. Estrada-Gil, J. C. Fernandez-Lopez, L. Uribe-Figueroa *et al.*, 2009 Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. Proc. Natl. Acad. Sci. USA 106: 8611–8616.

Smith, M. W., and S. J. O'Brien, 2005 Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. Nat. Rev. Genet. 6: 623–632.
Stephens, J. C., D. Briscoe, and S. J. O'Brien, 1994 Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am. J. Hum. Genet. 55: 809–824.
Tang, H., J. Peng, P. Wang, and N. J. Risch, 2005 Estimation of individual admixture: analytical and study design considerations. Genet. Epidemiol. 28: 289–301.
Tang, H., M. Coram, P. Wang, X. Zhu, and N. Risch, 2006 Reconstructing genetic ancestry blocks in admixed individuals. Am. J. Hum. Genet. 79: 1–12.
Tang, H., S. Choudhry, R. Mei, M. Morgan, W. Rodriguez-Cintron *et al.*, 2007 Recent genetic selection in the ancestral admixture of Puerto Ricans. Am. J. Hum. Genet. 81: 626–633.
Tishkoff, S. A., F. A. Reed, F. R. Friedlaender, C. Ehret, A. Ranciaro *et al.*, 2009 The genetic structure and history of Africans and African Americans. Science 324: 1035–1044.
Ungerer, M. C., S. J. E. Baird, J. Pan, and L. H. Rieseberg, 1998 Rapid hybrid speciation in wild sunflowers. Proc. Natl. Acad. Sci. USA 95: 11757–11762.
Verdu, P., F. Austerlitz, A. Estoup, R. Vitalis, M. Georges *et al.*, 2009 Origins and genetic diversity of pygmy hunter-gatherers from Western Central Africa. Curr. Biol. 19: 312–318.
Wang, J., 2003 Maximum-likelihood estimation of admixture proportions from genetic data. Genetics 164: 747–765.
Wang, S., N. Ray, W. Rojas, M. V. Parra, G. Bedoya *et al.*, 2008 Geographic patterns of genome admixture in Latin American Mestizos. PLoS Genet. 4: e1000037.
Workman, P. L., B. S. Blumberg, and A. J. Cooper, 1963 Selection, gene migration and polymorphic stability in a U. S. white and Negro population. Am. J. Hum. Genet. 15: 429–437.

Communicating editor: Y. S. Song

Appendix

Here, we obtain the recursion relation for the distribution of the admixture fraction from source population S_1 in the hybrid population H at generation g (Equations 3–5). Using the definitions of random variables Y and $H_{1,g}$, we can obtain an expression for $P(H_{1,g} = q)$, where q lies in $Q_g = \{0,\ 1/2^g, \ldots,\ (2^g-1)/2^g,\ 1\}$, conditional on the different possible pairs of parents Y for a random individual in population H at generation g:

$$P\big(H_{1,g}=q\big)=\sum_{y\in\begin{Bmatrix} S_1S_1\\ S_1H\\ S_1S_2\\ HH\\ S_2H\\ S_2S_2\end{Bmatrix}} P(Y=y)P\big(H_{1,g}=q|Y=y\big). \tag{A1}$$

Using Equations 1 and 2, we can evaluate the conditional probabilities $P(H_{1,g} = q \mid Y = y)$ in Equation A1 in terms of the unconditional probabilities $P(H_{1,g-1})$ in the previous generation, where for all $g \geq 1$, $P(H_{1,g} = q) = 0$ when q is not in Q_g. For q in Q_g, we have

$$P\big(H_{1,g}=q \mid Y=S_1S_1\big)=\begin{cases}1 & \text{if } q=1\\ 0 & \text{otherwise,}\end{cases} \tag{A2}$$

$$P\big(H_{1,g}=q \mid Y=S_1H\big)=P\big(H_{1,g-1}=2q-1\big), \tag{A3}$$

$$P\big(H_{1,g}=q \mid Y=S_1S_2\big)=\begin{cases}1 & \text{if } q=\frac{1}{2}\\ 0 & \text{otherwise,}\end{cases} \tag{A4}$$

$$P\big(H_{1,g}=q \mid Y=HH\big)=\sum_{r=0}^{2^{g-1}}P\Big(H_{1,g-1}^{(1)}=\frac{r}{2^{g-1}}\cap H_{1,g-1}^{(2)}=\frac{2^gq-r}{2^{g-1}}\Big), \tag{A5}$$

$$P\big(H_{1,g}=q \mid Y=S_2H\big)=P\big(H_{1,g-1}=2q\big), \tag{A6}$$

$$P\big(H_{1,g}=q \mid Y=S_2S_2\big)=\begin{cases}1 & \text{if } q=0\\ 0 & \text{otherwise.}\end{cases} \tag{A7}$$

By inserting Equations A2–A7 into Equation A1, we obtain the following recursion relation for $P(H_{1,g} = q)$. For the first generation ($g = 1$), we have $Q_1 = \{0,\ \frac{1}{2},\ 1\}$ and

$$P\big(H_{1,1}=q\big)=\begin{cases}P(Y=S_1S_1) & \text{if } q=1\\ P(Y=S_1S_2) & \text{if } q=\frac{1}{2}\\ P(Y=S_2S_2) & \text{if } q=0.\end{cases} \tag{A8}$$

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

$$\begin{aligned}P\big(H_{1,g}=q\big)=\;&P(Y=S_1S_1)P\big(H_{1,g}=q \mid Y=S_1S_1\big)\\ &+\;P(Y=S_1H)P\big(H_{1,g-1}=2q-1\big)\\ &+\;P(Y=S_1S_2)P\big(H_{1,g}=q \mid Y=S_1S_2\big)\\ &+\;P(Y=HH)\\ &\times\;\sum_{r=0}^{2^{g-1}}P\Big(H_{1,g-1}^{(1)}=\frac{r}{2^{g-1}}\cap H_{1,g-1}^{(2)}=\frac{2^gq-r}{2^{g-1}}\Big)\\ &+\;P(Y=S_2H)P\big(H_{1,g-1}=2q\big)\\ &+\;P(Y=S_2S_2)P\big(H_{1,g}=q \mid Y=S_2S_2\big).\end{aligned} \tag{A9}$$

For any value of q in Q_g where $g \geq 1$, we can use Table 1 to evaluate Equations A8 and A9 in terms of parameters $s_{1,g-1}$, $s_{2,g-1}$, and h_{g-1}. Simplifying Equations A8 and A9 we obtain Equations 3–5.

GENETICS

Supporting Information
http://www.genetics.org/content/suppl/2011/10/03/genetics.111.132787.DC1

A General Mechanistic Model for Admixture Histories of Hybrid Populations

Paul Verdu and Noah A. Rosenberg

Copyright © 2011 by the Genetics Society of America
DOI: 10.1534/genetics.111.132787

File S1

Supporting Material for

A General Mechanistic Model for Admixture Histories of Hybrid Populations

Paul Verdu and Noah A. Rosenberg

In this supplement, we generalize the two-source-population admixture model considered in the main text to allow m source populations. For a random individual in the hybrid population H, we obtain recursion relations for the moments of the admixture fraction from any specific one of the source populations. The derivations are almost completely analogous to those for the m=2 case, and we show that the results for arbitrary m agree with corresponding results in the main text for m=2.

The General Model Considering m Potential Source Populations

Define population H ("hybrid") as a population consisting of immigrant individuals from m mutually isolated source populations, S_1, S_2, …, S_m, and hybrid individuals who have ancestors that trace ultimately to S_1, S_2, …, S_m.

We let $s_{1,g}$, ..., $s_{m,g}$, and h_g be the fractional contributions of populations S_1, …, S_m, and H to the hybrid population at generation $g+1$. That is, for a randomly chosen individual in H at generation $g+1$, the probabilities that a randomly chosen parent of the individual derives from population S_1, …, S_m, and H are $s_{1,g}$, ..., $s_{m,g}$, and h_g respectively. For all $g \geq 0$, the parameters h_g and $s_{i,g}$ with $i \in \{1, \ldots, m\}$ have values that are greater than or equal to 0 and less than or equal to 1, such that $h_g + \sum_{i=1}^{m} s_{i,g} = 1$. At generation 0, the hybrid population is not yet formed. Therefore, $h_0 = 0$ and $\sum_{i=1}^{m} s_{i,0} = 1$. Hence, considering the period through generation g, in addition to g itself, this model has $mg-1$ independent parameters: $m-1$ introgression proportions in the first generation and one introgression proportion from each of the m source populations in each of the next $g-1$ generations. A diagram of the model appears in Figure S1.

Admixture fractions for a random individual in the hybrid population

As in the m=2 case, we focus on the fraction of ancestry from a specific source population for a random individual in H at a random locus. This fraction represents the proportion of the genome of a randomly chosen individual in H that ultimately traces to the specified source.

We indicate the possible sources for the (unordered) parents of an individual in H by S_iS_j and S_iH, with $i, j \in \{1, \ldots, m\}$, and HH. An individual in H at generation $g \geq 1$ has one of several possible types of parents, each with some probability dependent on the parameters $s_{i,g-1}$ with $i \in \{1, \ldots, m\}$, as described in Table S1. If the parents have different ancestries, we do not distinguish the order of the two parents, so that, for example, "S_iH" does not convey which specific parent is from population S_i and which is from H.

Let Y be a random variable indicating the source populations of the parents of a random individual in H. Let $H_{i,g}$ be the admixture fraction from source population S_i with $i \in \{1, \ldots, m\}$, for a random individual in population H at a random locus at generation g. Because at generation 0, the hybrid population is not yet formed, $h_0 = 0$, and $H_{i,0}$ is not defined. Using Table S1, we can write a recursion relation to calculate $H_{i,g}$ for any one of the source populations $S_i, i \in \{1, \ldots, m\}$, and for all $g \geq 1$. For the first generation ($g = 1$), for any mutually distinct values of i, j, and l between 1 and m, we have

$$H_{i,1} = \begin{cases} 1 & \text{if } Y = S_iS_i, \text{with } P[Y = S_iS_i] = s_{i,0}^2 \\ \frac{1}{2} & \text{if } Y = S_iS_j, \text{with } P[Y = S_iS_j] = 2s_{i,0}s_{j,0} \\ 0 & \text{if } Y = S_jS_j, \text{with } P[Y = S_jS_j] = s_{j,0}^2 \\ 0 & \text{if } Y = S_jS_l, \text{with } P[Y = S_jS_l] = 2s_{j,0}s_{l,0}. \end{cases} \quad \text{(S1)}$$

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

$$H_{i,g} = \begin{cases} 1 & \text{if } Y = S_iS_i, \text{with } P[Y = S_iS_i] = s_{i,g-1}^2 \\ \frac{H_{i,g-1} + 1}{2} & \text{if } Y = S_iH, \text{with } P[Y = S_iH] = 2s_{i,g-1}h_{g-1} \\ \frac{1}{2} & \text{if } Y = S_iS_j, \text{with } P[Y = S_iS_j] = 2s_{i,g-1}s_{j,g-1} \\ \frac{H_{i,g-1}^{(1)} + H_{i,g-1}^{(2)}}{2} & \text{if } Y = HH, \text{with } P[Y = HH] = h_{g-1}^2 \\ \frac{H_{i,g-1}}{2} & \text{if } Y = S_jH, \text{with } P[Y = S_jH] = 2s_{j,g-1}h_{g-1} \\ 0 & \text{if } Y = S_jS_j, \text{with } P[Y = S_jS_j] = s_{j,g-1}^2 \\ 0 & \text{if } Y = S_jS_l, \text{with } P[Y = S_jS_l] = 2s_{j,g-1}s_{l,g-1}. \end{cases} \quad \text{(S2)}$$

Here, $H_{i,g-1}^{(1)}$ and $H_{i,g-1}^{(2)}$ are fractions of ancestry from source population S_i for the two parents of a hybrid individual at generation g with $Y = HH$. We use the superscripts (1) and (2) only to indicate that $H_{i,g-1}^{(1)}$ and $H_{i,g-1}^{(2)}$ are separate independent and identically distributed (IID) random variables, so that if an individual in population H at generation g has two parents from H, the admixture fraction is distributed as the mean of the admixture fractions for two IID random individuals from H in the previous generation.

Moments of the admixture fraction for a random individual in the hybrid population

Similarly to the m=2 case, we can utilize the recursion relation in Equations S1 and S2 to obtain recursions for the expectation, variance, and higher moments of $H_{i,g}$, $i \in \{1, \dots, m\}$, as functions of g and the proportions of descent in the hybrid population H: $s_{1,t}$, ..., $s_{m,t}$ and h_t, for $t = 1, 2, \dots, g-1$. We first obtain a recursion for the expectation $E[H_{i,g}]$, $i \in \{1, \dots, m\}$. Next, we generalize the method used for finding the expectation, and we obtain a recursion relation for the kth moment, $E[H_{i,g}^k]$. Using the case of $k = 2$, we obtain a recursion for the variance $V[H_{i,g}]$, $i \in \{1, \dots, m\}$.

Expectation of $H_{i,g}$, $i \in \{1, \dots, m\}$: Using the law of total expectation, we can obtain an expression for the expectation $E[H_{i,g}]$, $i \in \{1, \dots, m\}$, as a function of conditional expectations for different possible pairs of parents Y for a random individual in population H at generation g:

$$E[H_{i,g}] = E_Y\left[E[H_{i,g}|Y]\right] = \sum_{y \in A} P(Y = y) E[H_{i,g} | Y = y]. \quad \text{(S3)}$$

The sum proceeds over the set A of all possible parental types for an individual in population H. For, the first generation, because parents cannot derive from population H itself, we have for any value of i from 1 to m,

$$
\begin{aligned}
E\left[H_{i,1}\right] = {} & P(Y = S_iS_i)E\left[H_{i,1}|Y = S_iS_i\right] \\
& + \sum_{\substack{j=1 \\ j\neq i}}^{m} P\left(Y = S_iS_j\right)E\left[H_{i,1}\middle|Y = S_iS_j\right] \\
& + \sum_{\substack{j=1 \\ j\neq i}}^{m} P\left(Y = S_jS_j\right)E\left[H_{i,1}\middle|Y = S_jS_j\right] \\
& + \sum_{\substack{j=1 \\ j\neq i}}^{m}\sum_{\substack{l=1 \\ l\neq i \\ l\neq j}}^{m} P\left(Y = S_jS_l\right)E\left[H_{i,1}\middle|Y = S_jS_l\right].
\end{aligned}
\tag{S4}
$$

As in the m=2 case, we use Equations S1 and S2 and recall that for all $g \geq 0$, $h_g + \sum_{i=1}^{m} s_{i,g} = 1$, $h_0 = 0$, and for all $g \geq 2$, $H_{1,g-1}^{(1)}$ and $H_{1,g-1}^{(2)}$ are IID random variables. We then obtain a recursion for the expectation of the admixture fraction. For $g = 1$, for any value of i from 1 to m,

$$
E\left[H_{i,1}\right] = s_{i,0}^2 + \sum_{\substack{j=1 \\ j\neq i}}^{m} s_{i,0}s_{j,0} = s_{i,0}\left(s_{i,0} + \sum_{\substack{j=1 \\ j\neq i}}^{m} s_{j,0}\right) = s_{i,0}. \tag{S5}
$$

For all $g \geq 2$,

$$
\begin{aligned}
E\left[H_{i,g}\right] &= s_{\mathrm{i},g-1}\left(s_{i,g-1} + h_{g-1} + \sum_{\substack{j=1 \\ j\neq i}}^{m} s_{j,g-1}\right) + h_{g-1}\left(s_{i,g-1} + h_{g-1} + \sum_{\substack{j=1 \\ j\neq i}}^{m} s_{\mathrm{j,g-1}}\right)E\left[H_{i,g-1}\right] \\
&= s_{i,g-1} + h_{g-1}E\left[H_{i,g-1}\right].
\end{aligned}
\tag{S6}
$$

Setting $i = 1$, Equations S5 and S6 match Equations 10 and 11 from the $m = 2$ case.

Moments of $H_{i,g}$, $i \in \{1, \ldots, m\}$: Using a similar computation to that employed in obtaining the recursion for the expected admixture, we can write recursions for higher moments of the admixture fraction from population S_i, with $i \in \{1, \ldots, m\}$ $\left(E\left[H_{i,g}^k\right]\right.$, for each $k \geq 1\left.\right)$. For the first generation ($g = 1$), we have for $k \geq 1$ and for any mutually distinct values of i, j, and l from 1 to m,

$$H_{i,1}^{k} = \begin{cases} 1^k & \text{if } Y = S_iS_i, \text{with } P[Y = S_iS_i] = s_{i,0}^2 \\ \left(\frac{1}{2}\right)^k & \text{if } Y = S_iS_j, \text{with } P\left[Y = S_iS_j\right] = 2s_{i,0}s_{j,0} \\ 0^k & \text{if } Y = S_jS_j, \text{with } P\left[Y = S_jS_j\right] = s_{j,0}^2 \\ 0^k & \text{if } Y = S_jS_l, \text{with } P\left[Y = S_jS_l\right] = 2s_{j,0}s_{l,0}. \end{cases} \quad \text{(S7)}$$

For all $g \geq 2$, we have

$$H_{i,g}^{k} = \begin{cases} 1^k & \text{if } Y = S_iS_i, \text{with } P[Y = S_iS_i] = s_{i,g-1}^2 \\ \left(\frac{H_{i,g-1}+1}{2}\right)^k & \text{if } Y = S_iH, \text{with } P[Y = S_iH] = 2s_{i,g-1}h_{g-1} \\ \left(\frac{1}{2}\right)^k & \text{if } Y = S_iS_j, \text{with } P\left[Y = S_iS_j\right] = 2s_{i,g-1}s_{j,g-1} \\ \left(\frac{H_{i,g-1}^{(1)}+H_{i,g-1}^{(2)}}{2}\right)^k & \text{if } Y = HH, \text{with } P[Y = HH] = h_{g-1}^2 \\ \left(\frac{H_{i,g-1}}{2}\right)^k & \text{if } Y = S_jH, \text{with } P\left[Y = S_jH\right] = 2s_{j,g-1}h_{g-1} \\ 0^k & \text{if } Y = S_jS_j, \text{with } P\left[Y = S_jS_j\right] = s_{j,g-1}^2 \\ 0^k & \text{if } Y = S_jS_l, \text{with } P\left[Y = S_jS_l\right] = 2s_{j,g-1}s_{l,g-1}\,, \end{cases} \quad \text{(S8)}$$

where $H_{i,g-1}^{(1)}$ and $H_{i,g-1}^{(2)}$ represent IID random variables for the fractions of ancestry from source population S_i for two hybrid individuals in generation $g-1$.

Using Equations S7 and S8 with the approach used previously to obtain the expectation of the admixture fraction from source population S_i, and using the binomial theorem, we obtain a recursion for $k \geq 1$. For $g = 1$, we have for $k \geq 1$ and any i from 1 to m,

$$E\left[H_{i,1}^{k}\right] = s_{i,0}^2 + \frac{s_{i,0}}{2^{k-1}} \sum_{\substack{j=1 \\ j \neq i}}^{m} s_{j,0}. \quad \text{(S9)}$$

For $g \geq 2$,

$$E\left[H_{i,g}^{k}\right] = s_{i,g-1}^{2} + \frac{s_{i,g-1}h_{g-1}}{2^{k-1}}\left(\sum_{r=0}^{k}\binom{k}{r}E\left[H_{i,g-1}^{r}\right]\right) + \frac{s_{i,g-1}}{2^{k-1}}\sum_{\substack{j=1\\ j\neq i}}^{m} s_{j,g-1}$$
$$+ \frac{h_{g-1}^{2}}{2^{k}}\left(\sum_{r=0}^{k}\binom{k}{r}E\left[H_{i,g-1}^{r}\right]E\left[H_{i,g-1}^{k-r}\right]\right) + \left(\frac{h_{g-1}}{2^{k-1}}\sum_{\substack{j=1\\ j\neq i}}^{m} s_{j,g-1}\right)E\left[H_{i,g-1}^{k}\right]. \quad \text{(S10)}$$

Recalling that for all $g \geq 0$, $h_g + \sum_{i=1}^{m} s_{i,g} = 1$, and that $h_0 = 0$, Equations S9 and S10 reduce to Equations S5 and S6 by setting $k = 1$. Moreover, by setting $i = 1$, Equations S9 and S10 match Equations 16 and 17 from the $m = 2$ case.

Variance of $H_{i,g}$, $i \in \{1, \ldots, m\}$: When $k = 2$, Equations S9 and S10 provide a recursion relation for the second moment of $H_{i,g}$, $i \in \{1, \ldots, m\}$. For the first generation, because $\sum_{i=1}^{m} s_{i,0} = 1$, we have for any value of i from 1 to m,

$$E\left[H_{i,1}^{2}\right] = \frac{s_{i,0}\left(s_{i,0}+1\right)}{2}. \quad \text{(S11)}$$

For subsequent generations ($g \geq 2$), because $h_g + \sum_{i=1}^{m} s_{i,g} = 1$, for $g \geq 0$, we obtain

$$E\left[H_{i,g}^{2}\right] = \frac{s_{i,g-1}\left(s_{i,g-1}+1\right)}{2} + h_{g-1}\left(s_{i,g-1} + \frac{h_{g-1}}{2}E\left[H_{i,g-1}\right]\right)E\left[H_{i,g-1}\right] + \frac{h_{g-1}}{2}E\left[H_{i,g-1}^{2}\right]. \quad \text{(S12)}$$

Setting $i = 1$, Equations S11 and S12 agree with Equations 20 and 21 from the $m = 2$ case.

With the relationship $V\left[H_{i,g}\right] = E\left[H_{i,g}^{2}\right] - \left(E\left[H_{i,g}\right]\right)^{2}$, $i \in \{1, \ldots, m\}$, and using Equations S5, S6, S11, and S12, we obtain a recursion for the variance of $H_{i,g}$. For the first generation ($g = 1$), we have for any i from 1 to m,

$$V\left[H_{i,1}\right] = \frac{s_{i,0}\left(1-s_{i,0}\right)}{2}, \quad \text{(S13)}$$

and for $g \geq 2$,

$$V\left[H_{i,g}\right] = \frac{s_{i,g-1}\left(1-s_{i,g-1}\right)}{2} - s_{i,g-1}h_{g-1}E\left[H_{i,g-1}\right]$$
$$+ \frac{h_{g-1}\left(1-h_{g-1}\right)}{2}\left(E\left[H_{i,g-1}\right]\right)^{2} + \frac{h_{g-1}}{2}V\left[H_{i,g-1}\right]. \quad \text{(S14)}$$

Setting $i = 1$, Equations S13 and S14 agree with Equations 22 and 23 from the $m = 2$ case.

Table S1 Possible pairs of parents for a random individual in the hybrid population H at generation g, and their probabilities.

Populations of origin of the parents of a random individual in population H at generation $g \geq 1$	Probability
S_i and S_i	$s_{i,g-1}^2$
S_i and H (or H and S_i)	$2s_{i,g-1}h_{g-1}$
S_i and S_j (or S_j and S_i)	$2s_{i,g-1}s_{j,g-1}$
H and H	h_{g-1}^2
S_j and H (or H and S_j)	$2s_{j,g-1}h_{g-1}$
S_j and S_j	$s_{j,g-1}^2$

Note that at generation 0, $h_0 = 0$ because the hybrid population is not yet formed. The indices i and j represent distinct populations.

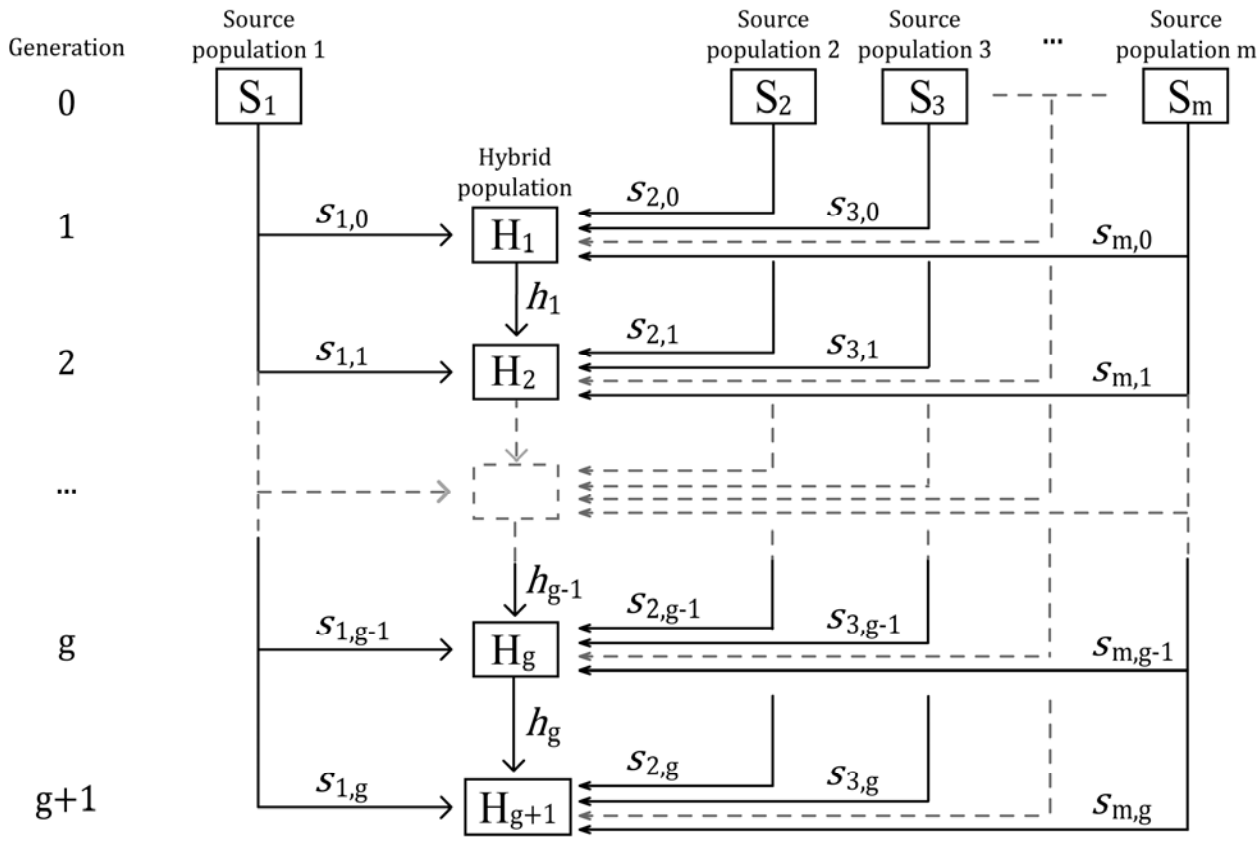

Figure S1 Diagram of a mechanistic model of admixture involving *m* isolated source populations.