

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, \dots, S_m , and hybrid individuals who have ancestors that trace ultimately to S_1, S_2, \dots, S_m .

We let $s_{1,g}, \dots, s_{m,g}$, and h_g be the fractional contributions of populations S_1, \dots, 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, \dots, S_m , and H are $s_{1,g}, \dots, s_{m,g}$, and h_g respectively. For all $g \geq 0$, the parameters h_g and $s_{i,g}$ with $i \in \{1, \dots, m\}$ 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 $(i, j) \in \{1, \dots, m\}$, 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 $i \in \{1, \dots, m\}$, 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 $i \in \{1, \dots, m\}$, 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, $i \in \{1, \dots, m\}$, 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

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

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

$$H_{i,g} = \begin{cases} 1 & \text{if } Y = S_i S_i, \text{ with } P[Y = S_i S_i] = s_{i,g-1}^2 \\ \frac{H_{i,g-1} + 1}{2} & \text{if } Y = S_i H, \text{ with } P[Y = S_i H] = 2s_{i,g-1}h_{g-1} \\ \frac{1}{2} & \text{if } Y = S_i S_j, \text{ with } P[Y = S_i S_j] = 2s_{i,g-1}s_{j,g-1} \\ \frac{H_{i,g-1}^{(1)} + H_{i,g-1}^{(2)}}{2} & \text{if } Y = HH, \text{ with } P[Y = HH] = h_{g-1}^2 \\ \frac{H_{i,g-1}}{2} & \text{if } Y = S_j H, \text{ with } P[Y = S_j H] = 2s_{j,g-1}h_{g-1} \\ 0 & \text{if } Y = S_j S_j, \text{ with } P[Y = S_j S_j] = s_{j,g-1}^2 \\ 0 & \text{if } Y = S_j S_l, \text{ with } P[Y = S_j S_l] = 2s_{j,g-1}s_{l,g-1}. \end{cases} \quad (\text{S2})$$

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 eqs. 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: $s_{1,t}, \dots, s_{m,t}$ 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 k th 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 H. For, the first generation, because parents cannot derive from H itself, we have for any value of i from 1 to m ,

$$\begin{aligned} E[H_{i,1}] &= P(Y = S_i S_i) E[H_{i,1}|Y = S_i S_i] \\ &\quad + \sum_{\substack{j=1 \\ j \neq i}}^m P(Y = S_i S_j) E[H_{i,1}|Y = S_i S_j] \\ &\quad + \sum_{\substack{j=1 \\ j \neq i}}^m P(Y = S_j S_j) E[H_{i,1}|Y = S_j S_j] \\ &\quad + \sum_{\substack{j=1 \\ j \neq i}}^m \sum_{\substack{l=1 \\ l \neq i \\ l \neq j}}^m P(Y = S_j S_l) E[H_{i,1}|Y = S_j S_l]. \end{aligned} \quad (\text{S4})$$

As in the $m=2$ case, we use eqs. 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 ,

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

For all $g \geq 2$,

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

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

Moments of $H_{i,g}$, $i \in \{1, \dots, m\}$: 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 $i \in \{1, \dots, m\}$ ($E[H_{i,g}^k]$, for each $k \geq 1$). 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 ,

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

For all $g \geq 2$, we have

$$H_{i,g}^k = \begin{cases} 1^k & \text{if } Y = S_i S_i, \text{ with } P[Y = S_i S_i] = s_{i,g-1}^2 \\ \left(\frac{H_{i,g-1} + 1}{2}\right)^k & \text{if } Y = S_i H, \text{ with } P[Y = S_i H] = 2s_{i,g-1}h_{g-1} \\ \left(\frac{1}{2}\right)^k & \text{if } Y = S_i S_j, \text{ with } P[Y = S_i S_j] = 2s_{i,g-1}s_{j,g-1} \\ \left(\frac{H_{i,g-1}^{(1)} + H_{i,g-1}^{(2)}}{2}\right)^k & \text{if } Y = HH, \text{ with } P[Y = HH] = h_{g-1}^2 \\ \left(\frac{H_{i,g-1}}{2}\right)^k & \text{if } Y = S_j H, \text{ with } P[Y = S_j H] = 2s_{j,g-1}h_{g-1} \\ 0^k & \text{if } Y = S_j S_j, \text{ with } P[Y = S_j S_j] = s_{j,g-1}^2 \\ 0^k & \text{if } Y = S_j S_l, \text{ with } P[Y = S_j S_l] = 2s_{j,g-1}s_{l,g-1}, \end{cases} \quad (S8)$$

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 eqs. 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 ,

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

For $g \geq 2$,

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

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

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

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

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

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

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

With the relationship $V[H_{i,g}] = E[H_{i,g}^2] - (E[H_{i,g}])^2$, $i \in \{1, \dots, m\}$, and using eqs. 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 ,

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

and for $g \geq 2$,

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

Setting $i = 1$, eqs. S13 and S14 agree with eqs. 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. Note that at generation 0, $h_0 = 0$ because the hybrid population is not yet formed. The indices i and j represent distinct populations.

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	$s_{j,g-1}^2$

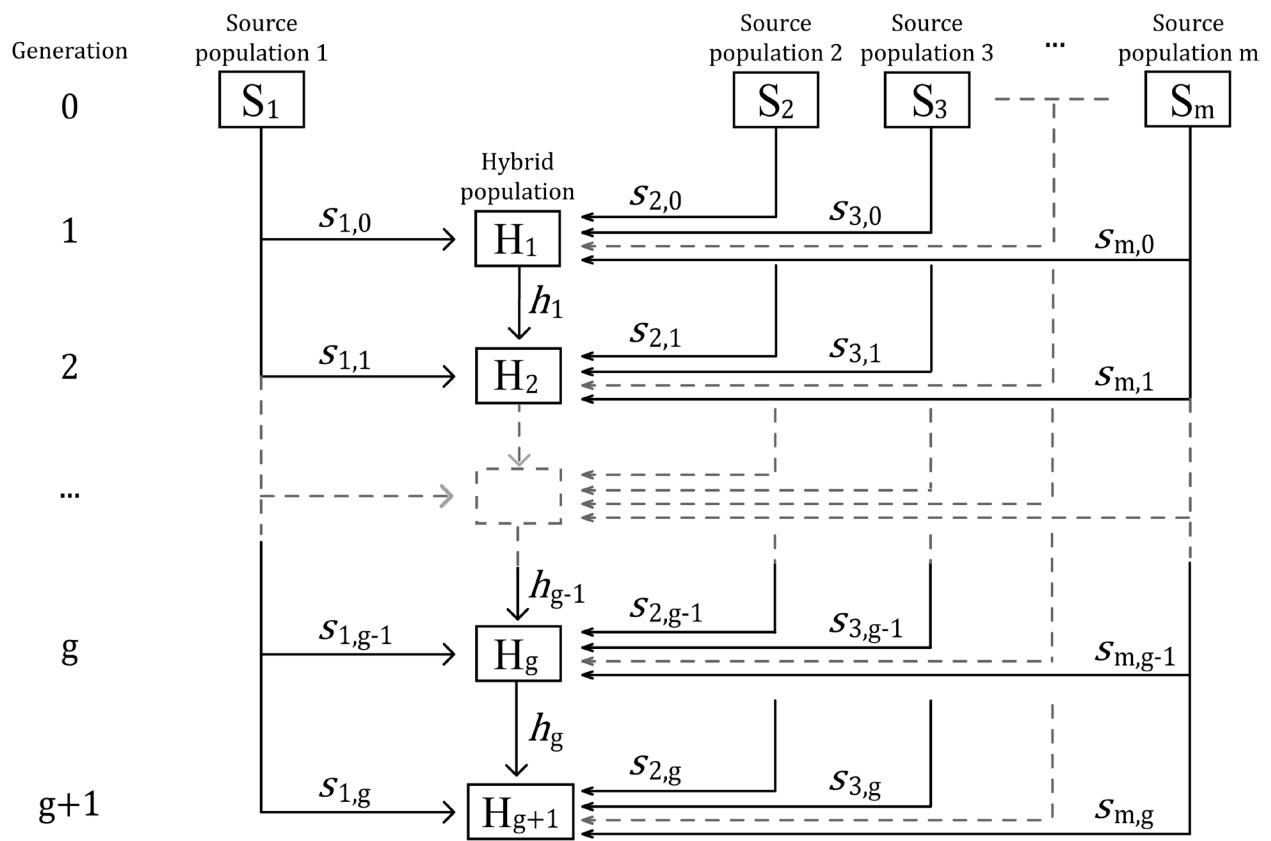


Figure S1 Diagram of a mechanistic model of admixture involving m isolated source populations.