Coancestry superposed on admixed populations yields measures of relatedness at individual-level resolution

Danfeng Chen[⋆] and John D. Storey[⋆†]

December 2024

[⋆]Lewis-Sigler Institute for Integrative Genomics, Princeton University, NJ 08544, USA

[†]Corresponding author: jstorey@princeton.edu

Abstract: The admixture model is widely applied to estimate and interpret population structure among individuals. Here we consider a "standard admixture" model that assumes the admixed populations are unrelated and also a generalized model, where the admixed populations themselves are related via coancestry (or covariance) of allele frequencies. The generalized model yields a potentially more realistic and substantially more flexible model that we call "super admixture". This super admixture model provides a one-to-one mapping in terms of probability moments with the population-level kinship model, the latter of which is a general model of genome-wide relatedness and structure based on identity-by-descent. We introduce a method to estimate the super admixture model that is based on method of moments, does not rely on likelihoods, is computationally efficient, and scales to massive sample sizes. We apply the method to several human data sets and show that the admixed populations are indeed substantially related, implying the proposed method captures a new and important component of evolutionary history and structure in the admixture model. We show that the fitted super admixture model estimates relatedness between all pairs of individuals at a resolution similar to the kinship model. The super admixture model therefore provides a tractable, forward generating probabilistic model of complex structure and relatedness that should be useful in a variety of scenarios.

Keywords: admixture, coancestry, kinship, population structure

Contents

1 Introduction

Populations are structured when genotype frequencies do not follow Hardy-Weinberg proportions. This may be due to several factors, including finite population sizes, migration, and genetic drift [1, 2]. Our goal here is to develop a framework and estimation method of a forward generating probability process that captures the observed genetic structure and relatedness among a set of individuals in a population-based study.

The framework is based on covarying allele frequencies among populations [3] and individuals [4], which we will refer to as *coancestry* [3–5]. The data underlying the proposed method are single nucleotide polymorphism (SNP) genotypes measured throughout the genome on a set of individuals. The aim is to formulate and estimate a model of the underlying process that leads to individual-specific allele frequencies (IAFs), which are parameters consisting of possibly distinct allele frequencies for every individual-SNP pair. IAFs have been formulated in previous work [6, 7] and they are the estimation target in several established admixture methods [8–10], a genome-wide association test for structured populations [11], and a test of structural Hardy-Weinberg equilibrium [12].

A joint probability distribution of the IAFs under a neutral model has been developed that yields covariances for all pairs of IAFs, parameterized by ancestral allele frequencies and coancestry parameters [4, 5]. This model produces a one-to-one mapping with the kinship parameters from the *identity-by-descent* model [13, 14], excluding close familial genetic relationships. This coancestry model therefore captures pairwise individual-level structure and relatedness equivalent to the kinship model. However, similarly to the kinship model, the coancestry parameterization is in terms of expected values, variances, and covariances of the IAFs and genotypes. It does not explicitly define a forward-generating probability model of IAFs.

Admixture models have been explored as a possible way to define such a forward-generating probability model [4, 5]. The products of an admixture model are individual-specific admixture proportions and population-specific allele frequencies. The IAFs are modeled as a weighted average of these *antecedent population allele frequencies* by the *individual-specific admixture proportions*. Several methods treat the admixture proportions and antecedent population allele frequencies as unknown parameters without explicitly making any assumptions about their random distributions [8–10]. Other methods place a prior probability distribution on them for Bayesian modeling fitting purposes [15–17]; however, these Bayesian methods do not include these prior distributions as an inference target.

In considering a model of random antecedent population allele frequencies, one could assume that the allele frequencies are independently generated among all antecedent populations based on a common set of parameters (e.g., independent draws from the Balding-Nichols distribution [18]). We will call this assumption the "standard admixture" model. However, this standard admixture model may be overly restrictive; rather, one could implement a coancestry model of the antecedent allele frequencies according to pairwise covariances [4, 5]. We will call this model the "super admixture" model, as coancestry (or covariance) is superposed on the admixed antecedent populations. Fig. 1 displays a schematic of these models.

Here, we develop a method that estimates the parameters in the super admixture model, which includes the standard admixture model as a special case. The method is based on method of moments estimation and geometric considerations, so it does not make assumptions about the probability distributions of the parameters and it does not involve costly likelihood maximization computations. Likelihood maximization is the most common approach used in fitting the admixture models [8, 9, 15–17], but we build from a recently proposed distribution-free moment-based method, called ALStructure, that only uses linear projections and geometric constraints on parameters to estimate the model [10]. ALStructure performs favorably to likelihood based methods (even in achieving a high likelihood) and can be tractably scaled to massive data sets. Our proposed super admixture method complements this framework and has similar advantages.

We establish super admixture through computational studies and analyses of data sets, including the human genome diversity panel (HGDP) [19], the 1000 genomes project (TGP) [20], the Human Origins study (HO) [21], and a study on individuals with Inadian ancestry (IND) [22]. We show on all of these data sets that the super admixture method is capable of capturing the same relatedness and structure as a model-free individual-level coancestry estimator [4], whereas the standard admixture model does not. We demonstrate that the framework can generate bootstrap genotypes that retain the structure seen in the human studies. For example, Fig. 2 shows these results on the HO study. We show that the coancestry among antecedent populations estimated by super admixture yields new insights and visualizations of structure previously unavailable, for example, Fig. 3 on the HO study. We develop and perform a statistical test to demonstrate on the studies that coancestry among the admixed antecedent populations is statistically different from zero to an high degree of significance.

Our proposed framework makes several contributions: (i) a distribution-free framework that can account for arbitrarily complex relationships among the admixed antecedent populations in the admixture model; (ii) admixture-based estimation of individual-level pairwise coancestry at a resolution equivalent to general, model-free coancestry and kinship; (iii) a partitioning of the super admixture model into evolutionary, genealogical, and statistical sampling components; and (iv) a tractable algorithm to form bootstrap samples of genotypes from the estimated evolutionary process.

2 Super admixture framework

Here, we first introduce the data and models, and then we detail the proposed framework. We describe how the framework is used to estimate the super admixture model, generate parameters and data from the model, and perform a hypothesis test of the standard versus super admixture models.

2.1 Coancestry

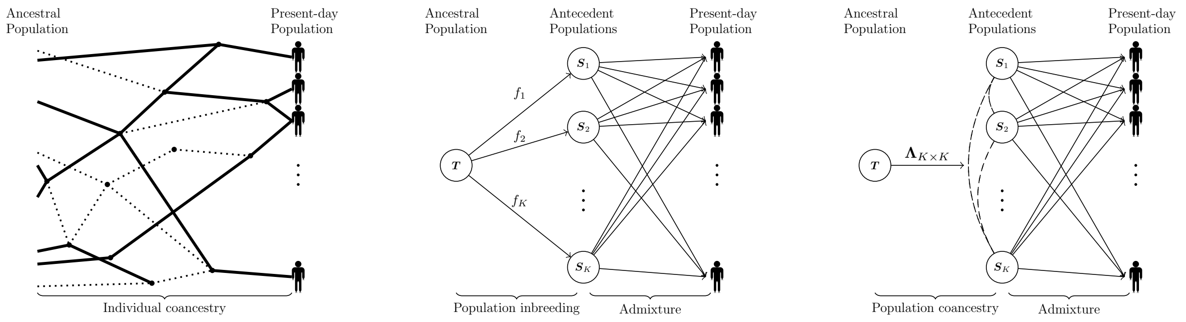
We assume that m SNPs are measured on n individuals. The genotype measurements are denoted by x_{ij} for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. For each SNP, one of the alleles is counted as a 0 and the other as a 1, implying that the SNP genotypes are $x_{ij} \in \{0, 1, 2\}$ where $x_{ij} = 0$ is homozygous for the 0 allele, $x_{ij} = 1$ is a heterozygote, and $x_{ij} = 2$ is homozygous for the 1 allele. We assume that $\mathbb{E}[x_{ij}|\pi_{ij}] = 2\pi_{ij}$ for IAF π_{ij}. This IAF parameterization allows each individual-SNP pair to possibly have a distinct allele frequency. The classical scenario where there is one allele frequency per SNP is a special case where $\pi_{i1} = \pi_{i2} = \cdots = \pi_{in}$. The conditional expected value $\mathbb{E}[x_{ij}|\pi_{ij}] = 2\pi_{ij}$ also allows for the IAFs π_{ij} to be random parameters, which we assume here.

We utilize an existing coancestry model where the IAFs are random parameters with respect to some ancestral population T that is common to all n individuals [4, 5]. This is a neutral model where

$$\mathbb{E}[\pi_{ij}|T] = a_i \tag{1}$$

$$\mathbb{C}[\pi_{ij}, \pi_{ik}|T] = a_i(1 - a_i)\theta_{jk} \tag{2}$$

for $i = 1, \ldots, m$ and $j, k = 1, \ldots, n$.The parameter a_i is the ancestral allele frequency in T for SNP i and $0 \leq \theta_{jk} \leq 1$ is the coancestry for individuals j and k with respect to T. (Note that the a_i and θ_{jk} parameters depend on T and could be different if conditioning on a different common ancestral population.) The coancestry model we utilize also makes the



(A) Coancestry model (B) Standard admixture model (C) Super admixture model

Figure 1: Graphical representations of the coancestry model, the standard admixture model, and the super admixture model. (A) In the coancestry model, individuals in the present-day population are connected by a complex genealogy. (B) In the standard admixture model, the arrows connecting T with $S_1, \ldots, S_K$ reflect that the antecedent populations evolved independently from T. Arrows connecting $S_1, \ldots, S_K$ with individuals in the present-day population reflect that these individuals were admixed from independent antecedent populations. (C) In the super admixture model, dashed lines connecting all pairs of antecedent populations reflect that antecedent populations have coancestry parameterized by $\mathbf{\Lambda}$. Arrows connecting $S_1, \ldots, S_K$ with individuals in the present-day population reflect that these individuals were admixed from covarying antecedent populations.

assumption used in previous work [4, 5, 7–12] that

$$x_{ij}|\pi_{ij} \sim \text{Binomial}(2, \pi_{ij})$$

where the x_{ij} are jointly independent. Under this model, it follows that

$$\mathbb{C}[x_{ij}, x_{ik}|T] = \begin{cases} 2a_i(1-a_i)\left(1+\theta_{jj}\right) & j = k, \\ 4a_i\left(1-a_i\right)\theta_{jk} & j \neq k. \end{cases}$$

A one-to-one mapping exists with the identity-by-descent kinship model (often used in GWAS methods), denoted by ϕ_{jk}, by matching variances and covariances [4, 5]. The parameters map so that

$$\theta_{jk} = \begin{cases} 2\phi_{jk} - 1 & \text{if} \quad j = k, \\ \phi_{jk} & \text{if} \quad j \neq k. \end{cases} \tag{3}$$

When $\min_{jk} \theta_{jk} = 0$, then T is the most recent common ancestral population [4]. The full set of parameters is denoted by the $n \times n$ symmetric matrix $\mathbf{\Theta}$ with (j,k) entry θ_{jk}.

2.2 Admixture models

General admixture

We first describe a general formulation of the admixture model, of which standard and super admixture are special cases. There are K populations $S_1, S_2, \ldots, S_K$ descended from T that precede the present day population, which we refer to as “antecedent populations”. While T has allele frequencies $a_1, a_2, \ldots, a_m$, antecedent population S_u has allele frequencies $p_{1u}, p_{2u}, \ldots, p_{mu}$ for $u = 1, 2, \ldots, K$. The allele frequencies $\{p_{iu}\}$ are random parameters from a distribution parameterized by $\{a_i\}$ plus other possible parameters that characterize the evolutionary process from T to S_u.

For each individual j, there is a genealogical process from population T to the present day population. This is captured by a random K-vector $q_{1j}, q_{2j}, \ldots, q_{Kj}$ of admixture proportions, where $0 \leq q_{uj} \leq 1$ and $\sum_{u=1}^{K} q_{uj} = 1$. The parameter q_{uj} is the proportion of the individual j randomly descended from S_u. Therefore, the IAFs are such that

$$\pi_{ij} = \sum_{u=1}^{K} p_{iu} q_{uj}. \tag{4}$$

We collect the antecedent population allele frequencies into the $m \times K$ matrix $\boldsymbol{P}$ and the admixture proportions into the $K \times n$ matrix $\boldsymbol{Q}$, it follows that

$$\boldsymbol{\Pi} = \boldsymbol{P}\boldsymbol{Q},$$

where $\boldsymbol{\Pi}$ is an $m \times n$ matrix with (i, j) entry π_{ij}.

Standard admixture

We define the standard admixture model to be the case where the antecedent allele frequencies are independently distributed. Specifically, in this model p_{iu} is a random parameter with mean a_i and variance $a_i(1 - a_i)f_u$. The standard admixture model is defined as follows for $i = 1, 2, \ldots, m$ and $u = 1, 2, \ldots, K$.

Standard Admixture:

$p_{i1}, p_{i2}, \ldots, p_{iK}$ are jointly *independent*

$\mathbb{E}[p_{iu}|T] = a_i$

$\mathbb{V}[p_{iu}|T] = a_i(1 - a_i)f_u$

Under this parameterization, a_i is the ancestral allele frequency in T and f_u is the inbreeding coefficient or F_{ST} of antecedent population S_u with respect to T. Since the $\{p_{iu}\}$ are jointly independent, there is no coancestry among antecedent populations and there is no dependence among loci.

One well-known distribution that could be utilized here is the Balding-Nichols (BN) distribution [18] with parameters a_i and f_u:

$$p_{iu} \sim \text{Beta}\left(\frac{1-f_u}{f_u}a_i, \frac{1-f_u}{f_u}(1-a_i)\right). \tag{5}$$

We will write this re-parameterized Beta distribution as $\text{BN}(a_i, f_u)$. This achieves the expected value and variance of the standard admixture definition. The Balding-Nichols distribution is often used to generate allele frequencies for a set of populations to achieve desired expected allele frequencies and F_{ST} values. This distribution has been discussed as useful for generating antecedent allele frequencies in the standard admixture model [4–7, 23].

Super admixture

The super admixture model extends the standard admixture model in that it includes a covariance among antecedent population allele frequencies, which we refer to as population-level coancestry. While we denoted individual-level coancestry by θ_{jk}, we will denote population-level coancestry by λ_{uv} for $u, v = 1, 2, \ldots, K$ where $0 \leq \lambda_{uv} \leq 1$. We collect these values into the $K \times K$ symmetric coancestry matrix $\boldsymbol{\Lambda}$. The super admixture model is defined as follows for $i = 1, 2, \ldots, m$ and $u, v = 1, 2, \ldots, K$.

$$\begin{aligned}
&\text{Super Admixture:}\\
&p_{i1}, p_{i2}, \ldots, p_{iK} \text{ are jointly } \textit{dependent}\\
&\mathbb{E}[p_{iu}|T] = a_i\\
&\mathbb{V}[p_{iu}|T] = a_i(1-a_i)\lambda_{uu}\\
&\mathbb{C}[p_{iu}, p_{iv}|T] = a_i(1-a_i)\lambda_{uv}
\end{aligned} \tag{6}$$

In this model we assume that allele frequencies between loci are independent, so the random K-vectors $(p_{h1}, p_{h2}, \ldots, p_{hK})$ and $(p_{i1}, p_{i2}, \ldots, p_{iK})$ are independent for $h \neq i$. Thus, a potential generalization of the super admixture model is to include dependence among loci. Otherwise, the super admixture model is general in that it allows for the full range of coancestry values among antecedent populations.

Forward generating probability process

We now describe the super admixture model as a forward generating probability process. Suppose that the admixture proportions $\boldsymbol{Q}$ are drawn from some probability distribution $\mathcal{Q}$. Then, for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$:

$$\begin{aligned}
(p_{i1}, p_{i2}, \ldots, p_{iK}) &\sim (\boldsymbol{a}, \boldsymbol{\Lambda}) \\
(q_{1j}, q_{2j}, \ldots, q_{Kj}) &\sim \mathcal{Q} \\
\pi_{ij} &= \sum_{u=1}^{K} p_{iu} q_{uj} \\
x_{ij} | \pi_{ij} &\sim \text{Binomial}(2, \pi_{ij})
\end{aligned}$$

The joint probability of all random quantities can be factored as follows:

$$\mathbb{P}(\boldsymbol{X}, \boldsymbol{Q}, \boldsymbol{P} | T, \mathcal{Q}) = \mathbb{P}(\boldsymbol{P}|T)\mathbb{P}(\boldsymbol{Q}|\mathcal{Q})\mathbb{P}(\boldsymbol{X}|\boldsymbol{P}, \boldsymbol{Q}).$$

One interpretation of this is that $\mathbb{P}(\boldsymbol{P}|T)$ represents evolutionary sampling, $\mathbb{P}(\boldsymbol{Q}|\mathcal{Q})$ represents genealogical sampling, and $\mathbb{P}(\boldsymbol{X}|\boldsymbol{P}, \boldsymbol{Q})$ represents statistical sampling.

Individual-level coancestry in the admixture models

Recall that in the covariance model, the covariance of two IAFs for a given SNP is $\mathbb{C}[\pi_{ij}, \pi_{ik}|T] = a_i(1 - a_i)\theta_{jk}$, shown in Eq. (2). Conditioning on the admixture proportions $\boldsymbol{Q}$, which are ancillary to allele frequencies, this covariance under the super admixture model is, for $j, k = 1, 2, \ldots, n$,

$$\begin{aligned}
\mathbb{C}\left[\pi_{ij}, \pi_{ik} | \boldsymbol{Q}, T\right] &= \mathbb{C}\left[\sum_{u=1}^{K} p_{iu} q_{uj}, \sum_{v=1}^{K} p_{iv} q_{vk} \middle| \boldsymbol{Q}, T\right] \\
&= \sum_{u=1}^{K} \sum_{v=1}^{K} q_{uj} q_{vk} \mathbb{C}\left[p_{iu}, p_{iv} | T\right] \\
&= a_i(1 - a_i) \sum_{u=1}^{K} \sum_{v=1}^{K} q_{uj} q_{vk} \lambda_{uv}. \qquad (7)
\end{aligned}$$

By setting the covariance from Eq. (2) equal to Eq. (7), it follows that under the super admixture model the individual-level coancestry is the following.

Super Admixture Individual-level Coancestry:

$$\theta_{jk} = \sum_{u=1}^{K} \sum_{v=1}^{K} q_{uj} q_{vk} \lambda_{uv} \tag{8}$$

In the standard admixture model, $\mathbb{V}[p_{iu}|T] = a_i(1-a_i)f_u$, whereas in the super admixture model $\mathbb{V}[p_{iu}|T] = a_i(1-a_i)\lambda_{uu}$. If we set $f_u = \lambda_{uu}$, the difference between the standard and super admixture models is therefore that in the standard model, $\lambda_{uv} = 0$ for $u \neq v$. To work with a single notation, we will therefore write λ_{uu} in place of f_u for the standard admixture model. The coancestry in this model is as follows.

Standard Admixture Individual-level Coancestry:

$$\theta_{jk} = \sum_{u=1}^{K} q_{uj} q_{uk} \lambda_{uu} \tag{9}$$

$$\lambda_{uv} = 0 \text{ for } u \neq v$$

Considering all pairs of individuals simultaneously, the individual-level coancestry matrix $\boldsymbol{\Theta}$ can be written in terms of the antecedent population-level coancestry $\boldsymbol{\Lambda}$ and the admixture proportions $\boldsymbol{Q}$ as

$$\boldsymbol{\Theta} = \boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q},$$

which is an important relationship we utilize to estimate $\boldsymbol{\Lambda}$.

2.3 Estimating coancestry among antecedent populations

Here, we propose a method to estimate the antecedent population-level coancestry $\boldsymbol{\Lambda}$ under the super admixture model, with the standard admixture model estimate as a special case. The rationale is to leverage the relationship, $\boldsymbol{\Theta} = \boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}$. Given values for $\boldsymbol{\Theta}$ and $\boldsymbol{Q}$, we identify values of $\boldsymbol{\Lambda}$ that make $\boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}$ close to $\boldsymbol{\Theta}$, while obeying the geometic constraints of $\boldsymbol{\Lambda}$ (i.e., $0 \leq \lambda_{uv} \leq 1$ and $\lambda_{uv} = \lambda_{vu}$).

Given values for $\boldsymbol{\Theta}$ and $\boldsymbol{Q}$, we formulate the problem of the estimating the antecedent population-level coancestry $\boldsymbol{\Lambda}$ under the super admixture model as follows.

Problem 1.

$$\min_{\mathbf{\Lambda}\in\mathbb{R}^{K\times K}} \|\mathbf{\Theta} - \boldsymbol{Q}'\mathbf{\Lambda}\boldsymbol{Q}\|_F^2$$

$$\text{subject to: } 0 \le \lambda_{uv} \le 1 \text{ and } \lambda_{uv} = \lambda_{vu}$$

$$\text{for } u, v = 1, \ldots, K$$

where $\|\cdot\|_F$ represents the Frobenius norm defined in Appendix A.1. We utilize the proximal forward-backward (PFB) method [24] to solve this optimization problem, resulting in Algorithm 1 for solving Problem 1. Every sequence of $(\mathbf{\Lambda}_t)_{t\in\mathbb{N}}$ generated from this algorithm is guaranteed to converge to a solution of the corresponding problem. The PFB method and how to employ it to our setting are detailed in Appendix B.2. The performance of Algorithm 1 is demonstrated in Appendix C.

Algorithm 1: Estimating $\mathbf{\Lambda}$ for the super admixture model given $\mathbf{\Theta}$ and $\boldsymbol{Q}$

```
input: Coancestry matrix $\mathbf{\Theta}$ and admixture proportions matrix $\boldsymbol{Q}$
let $L = \sigma^4_{\max}(\boldsymbol{Q})$
let $\mathbf{\Lambda}_0 \leftarrow (\boldsymbol{Q}\boldsymbol{Q}')^{-1}\mathbf{\Theta}(\boldsymbol{Q}\boldsymbol{Q}')^{-1}$
for $t = 1, 2, \ldots$ do
    $\boldsymbol{G} \leftarrow 2\boldsymbol{Q}(\boldsymbol{Q}'\mathbf{\Lambda}_{t-1}\boldsymbol{Q} - \mathbf{\Theta})\boldsymbol{Q}'$
    $\mathbf{\Lambda}^* \leftarrow \mathbf{\Lambda}_{t-1} - \frac{1}{L}\boldsymbol{G}$
    $\mathbf{\Lambda}_t = \{\lambda_{uv,t}\}$ where $\lambda_{uv,t} = \max(0, \min(1, \lambda^*_{uv}))$
return $\mathbf{\Lambda}_t$
```

$\sigma_{\max}(\cdot)$ denotes the maximum singular value (Appendix A.1).

To estimate all components of the super admixture model, one needs estimates of the $n\times n$ individual-level coancestry matrix $\mathbf{\Theta}$, the $K\times K$ antecedent population-level coancestry matrix $\mathbf{\Lambda}$, the $m \times K$ matrix of antecedent population allele frequencies $\boldsymbol{P}$, and the $K \times n$ matrix of admixture proportions $\boldsymbol{Q}$. There exists a wide range of methods for estimating $\boldsymbol{P}$ and $\boldsymbol{Q}$ [9, 10, 15–17, 25]. Here, we utilize the ALStructure method [10], which implements method of moments and geometric constraints to estimate $\boldsymbol{Q}$ similarly to our approach here. In that method, a linear basis of $\boldsymbol{Q}$ is determined from $\boldsymbol{X}$ that has theoretical guarantees to span the true basis as the number of SNPs m grows large. A projection-based estimate $\hat{\mathbf{\Pi}}$ of the IAFs is also formed. The quantity $\|\hat{\mathbf{\Pi}} - \boldsymbol{Q}\boldsymbol{P}\|_F$ is then algorithmically minimized through geometrically constrained alternating least squares to yield estimates $\hat{\boldsymbol{Q}}$ and $\hat{\boldsymbol{P}}$.

We utilize the structural Hardy-Weinberg (sHWE) test [12] for determining the number of antecedent populations K, as outlined in that work. The approach is to consider a range of K values to test the assumption that $x_{ij}|\pi_{ij} \sim \text{Binomial}(2, \pi_{ij})$ based on the estimates $\hat{\pi}_{ij}$ and a goodness-of-fit statistic with a parametric bootstrap null distribution; K is then parsimoniously chosen to satisfy this modeling assumption from a genome-wide perspective. A method of moments estimator of $\boldsymbol{\Theta}$ was derived in ref. [4], where it was shown to have favorable properties and is consistent for the true values under certain assumptions. We denote this Ochoa-Storey (OS) estimate by $\hat{\boldsymbol{\Theta}}^{\text{OS}}$ and review its details in Appendix B.1. If one has alternative ways to estimate $\boldsymbol{\Theta}$ and $\boldsymbol{Q}$, and to determine K, then those can be used within our framework as well.

Algorithm 2: Estimating $\boldsymbol{\Lambda}$ for the super admixture model given $\boldsymbol{X}$

```
  input: Genotype matrix $\boldsymbol{X}$
1 calculate the OS estimate of individual-level coancestry $\hat{\boldsymbol{\Theta}}^{\text{OS}}$
2 choose $K$ from the structural Hardy-Weinberg (sHWE) goodness of fit procedure
3 calculate the estimate $\hat{\boldsymbol{Q}}$ for $K$ via the ALStructure method
4 calculate the estimate $\hat{\boldsymbol{\Lambda}}^{\text{sup}}$ by applying Algorithm 1 with inputs $\hat{\boldsymbol{\Theta}}^{\text{OS}}$ and $\hat{\boldsymbol{Q}}$
5 return $\hat{\boldsymbol{\Lambda}}^{\text{sup}}$
```

Note that one can further calculate a corresponding estimate for individual-level coancestry by

$$\hat{\boldsymbol{\Theta}}^{\text{sup}} = \hat{\boldsymbol{Q}}' \hat{\boldsymbol{\Lambda}}^{\text{sup}} \hat{\boldsymbol{Q}},$$

which can be compared to $\hat{\boldsymbol{\Theta}}^{\text{OS}}$ in order to aid in model fit assessment.

We can estimate $\boldsymbol{\Lambda}$ under the standard admixture model by modifying the constraints in Problem 1. This leads to Problem B.1 and Algorithm B.2 described in Appendix B.2. Algorithm 2 can then be used to form the estimate $\hat{\boldsymbol{\Lambda}}^{\text{std}}$ under the standard admixture model with Algorithm 1 in Line 4 replaced by Algorithm B.2. The corresponding estimate for individual-level coancestry can be calculated as $\hat{\boldsymbol{\Theta}}^{\text{std}} = \hat{\boldsymbol{Q}}' \hat{\boldsymbol{\Lambda}}^{\text{std}} \hat{\boldsymbol{Q}}$. The performance of Algorithm B.2 is also demonstrated in Appendix C.

2.4 Simulating antecedent population allele frequencies

We now introduce a method to generate antecedent population allele frequencies with given coancestry $\boldsymbol{\Lambda}$. We noted above in Eq. (5) that for the standard admixture model, one way to generate allele frequencies $p_{i1}, p_{i2}, \ldots, p_{iK}$ is via independent realizations from the Balding-Nichols (BN) distribution: $p_{iu} \sim \text{BN}(a_i, \lambda_{uu})$ for $u = 1, 2, \ldots, K$. As there is no default approach to extending this to the super admixture case, we propose a method here called "double-admixture". The main idea of the method is that we form two layers of allele frequencies: the first layer is composed of independent draws from the BN distribution, and the second layer mixes these to create $p_{i1}, p_{i2}, \ldots, p_{iK}$ with coancestry $\boldsymbol{\Lambda}$.

Let S be the number of components that will be mixed, $\boldsymbol{W}$ be the $S \times K$ matrix of mixture proportions, and $\boldsymbol{\Gamma}$ an $S \times S$ diagonal matrix. The entries of $\boldsymbol{W}$ are w_{su} where $0 \leq w_{su} \leq 1$ and $\sum_{s=1}^{S} w_{su} = 1$ for $u = 1, 2, \ldots, K$. The diagonal values of $\boldsymbol{\Gamma}$ are represented by γ_s where $0 \leq \gamma_s \leq 1$, and all other values are 0. Suppose that for $i = 1, \ldots, m$ we generate

$$z_{is} \sim \mathrm{BN}(a_i, \gamma_s)$$

independently for $s = 1, \ldots, S$, and we then set

$$p_{iu} = \sum_{s=1}^{S} z_{is} w_{su}$$

for $u = 1, \ldots, K$. It can be verified that

$$\begin{aligned} \mathbb{E}[p_{iu}] &= a_i \quad u = 1, \ldots, K \\ \mathbb{C}[p_{iu}, p_{iv}] &= a_i(1 - a_i) \sum_{s=1}^{S} w_{su} w_{sv} \gamma_s \end{aligned}$$

for $u, v = 1, 2, \ldots, K$. By matching these equations with Eq. (6), one can see that if

$$\lambda_{uv} = \sum_{s=1}^{S} w_{su} w_{sv} \gamma_s \tag{10}$$

then $p_{i1}, p_{i2}, \ldots, p_{iK}$ has coancestry $\boldsymbol{\Lambda}$ as desired. In matrix terms, Eq. (10) is equivalent to

$$\boldsymbol{\Lambda} = \boldsymbol{W}' \boldsymbol{\Gamma} \boldsymbol{W}. \tag{11}$$

Therefore, the double-admixture method is based on the following optimization problem.

Problem 2.

$$\min_{\boldsymbol{W},\boldsymbol{\Gamma}} \|\boldsymbol{\Lambda} - \boldsymbol{W}'\boldsymbol{\Gamma}\boldsymbol{W}\|_F^2$$

$$\begin{aligned} \text{subject to: } & 0 \leq w_{su} \leq 1, \sum_{s=1}^{S} w_{su} = 1 \\ & \epsilon \leq \gamma_s \leq 1 - \epsilon \text{ for small } \epsilon > 0 \\ & \text{for } u = 1, 2, \ldots, K; s = 1, 2, \ldots, S \end{aligned}$$

Algorithm 3: Calculating $\boldsymbol{W}$ and $\boldsymbol{\Gamma}$ in the double-admixture method

```
input: Antecedent populations coancestry $\boldsymbol{\Lambda}$, number of BN distributions $S$, step
       size parameters $\tau_1$ and $\tau_2$, and a small positive number $\epsilon$
let $\boldsymbol{\Gamma}_0$ be an $S \times S$ diagonal matrix with diagonal elements drawn independently
  from Uniform$(0, 1)$
let $\boldsymbol{W}_0$ be an $S \times K$ matrix whose columns $(w_{1u}, w_{2u}, \ldots, w_{Su})'$ are drawn
  independently from Dirichlet$(\mathbf{1})$
for $t = 1, 2, \ldots$ do
    $L_1 \leftarrow 4(\|\boldsymbol{\Lambda}\|_2 \|\boldsymbol{\Gamma}_{t-1}\|_2 + 3K\|\boldsymbol{\Gamma}_{t-1}\|_2^2)$
    $\boldsymbol{G}_1 \leftarrow -4\boldsymbol{\Gamma}_{t-1}\boldsymbol{W}_{t-1}(\boldsymbol{\Lambda} - \boldsymbol{W}'_{t-1}\boldsymbol{\Gamma}_{t-1}\boldsymbol{W}_{t-1})$
    $\boldsymbol{W}^* \leftarrow \boldsymbol{W}_{t-1} - \frac{1}{\tau_1 L_1}\boldsymbol{G}_1$
    for $u = 1, \ldots, K$ do
        $\boldsymbol{w}_{u,t} \leftarrow \mathcal{P}_\Delta(\boldsymbol{w}^*_u)$ where $\boldsymbol{w}_{u,t}$ and $\boldsymbol{w}^*_u$ are the corresponding columns of $\boldsymbol{W}_t$
          and $\boldsymbol{W}^*$
    $L_2 \leftarrow 2\|\boldsymbol{W}_t\|_2^4$
    $\boldsymbol{G}_2 \leftarrow -2\boldsymbol{W}_t(\boldsymbol{\Lambda} - \boldsymbol{W}'_t\boldsymbol{\Gamma}_{t-1}\boldsymbol{W}_t)\boldsymbol{W}'_t$
    $\boldsymbol{\Gamma}^* \leftarrow \boldsymbol{\Gamma}_{t-1} - \frac{1}{\tau_2 L_2}\boldsymbol{G}_2$
    $\gamma_{s,t} = \max(\epsilon, \min(1 - \epsilon, \gamma^*_{ss}))$ for $s = 1, 2, \ldots, S$
    $\boldsymbol{\Gamma}_t = \text{diag}(\gamma_{1,t}, \gamma_{2,t}, \ldots, \gamma_{S,t})$
return $\boldsymbol{\Gamma}_t$ and $\boldsymbol{W}_t$
```

Here, we set $S = 2K$,$\tau_1 = \tau_2 = 1.1$, $\epsilon = 0.01$; user should investigate their choices.
$\| \cdot \|_2$ denotes the spectral norm and $\mathcal{P}_\Delta$ denotes projection onto the unit simplex (Appendix A.1).

We adapted the proximal alternating linearized minimization (PALM) method [26] to solve Problem 2, resulting in Algorithm 3 for calculating the parameters in the double-admixture method. Every sequence $(\boldsymbol{W}_t, \boldsymbol{\Gamma}_t)_{t\in\mathbb{N}}$ generated from Algorithm 3 is guaranteed to converge to a critical point. Integrating Algorithm 3 with the generative steps for p_{iu} described above, Algorithm 4 simulates antecedent population allele frequencies with the desired coancestry. In Appendix B.3, the PALM method is briefly introduced and the convergence of Algorithm 3 is proved.

Algorithm 4: The double-admixture algorithm for simulating $\boldsymbol{P}$

```
input: Ancestral allele frequencies $\boldsymbol{a}$, coancestry among antecedent populations $\boldsymbol{\Lambda}$,
       other input arguments for Algorithm 3
calculate $\hat{\boldsymbol{\Gamma}}$ and $\hat{\boldsymbol{W}}$ using Algorithm 3
for $i = 1, \ldots, m$ do
    generate $z_{is} \sim \mathrm{BN}(a_i, \hat{\gamma}_s)$ independently for $s = 1, 2, \ldots, S$
    set $p_{iu} \leftarrow \sum_{s=1}^{S} z_{is}\hat{w}_{su}$ for $u = 1, 2, \ldots, K$
return $\boldsymbol{P}$
```

One possible drawback of the double-admixture method is that the approach relies on the existence of $\boldsymbol{W}$ and $\boldsymbol{\Gamma}$ so that $\boldsymbol{\Lambda} = \boldsymbol{W}'\boldsymbol{\Gamma}\boldsymbol{W}$. We do not currently have a theoretical guarantee for such $\boldsymbol{W}$ and $\boldsymbol{\Gamma}$ (although one may exist since S can be made large). Therefore, we provide a complementary method in Appendix B.4, the NORmal To Anything (NORTA) approach [27], serving as a tool for simulating $\boldsymbol{P}$ when the double-admixture method is not applicable. It should be noted that the double-admixture method solves the optimization one time for the entire process so that its running time is independent of the number of loci m. In contrast, the NORTA method has to solve $K \times (K-1)/2$ root-finding problems per locus and therefore has a complexity of $\mathcal{O}(K^2 m)$, rendering it significantly more time consuming. The performances of the double-admixture and NORTA methods are demonstrated in Appendix C.

Note that if we set $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}$ for a diagonal standard admixture $\boldsymbol{\Lambda}$ and $\boldsymbol{W} = \boldsymbol{I}_K$ (where $\boldsymbol{I}_K$ is the $K \times K$ identity matrix), then the double-admixture method reduces to the BN sampling from Eq. (5), which produces valid antecedent population frequencies for the standard admixture model. From this observation, the double-admixture method can be viewed as a generalization of BN sampling.

2.5 Generating bootstrap datasets from realistic population structures

By utilizing the double-admixture method, we implemented the following algorithm to simulate genotypes from the super admixture model, shown in Algorithm 5. We assessed whether Algorithm 5 generates genotypes that satisfy the moment constraints imposed by the super admixture model in Appendix C. Algorithm 5 is especially useful when inputs $\boldsymbol{a}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{Q}$ reflect real populations. When these parameters are unavailable one can utilize an admixture method to estimate $\boldsymbol{Q}$ and the method proposed here to estimate $\boldsymbol{\Lambda}$. The ancestral allele frequencies can be estimated with simple sample means. We outline Algorithm 6, with the ALStructure algorithm for estimation of $\boldsymbol{Q}$ and the super admixture algorithm for estimation of $\boldsymbol{\Lambda}$.

Algorithm 5: Generating genotypes $\boldsymbol{X}$ from the super admixture model

```
   input: Ancestral allele frequencies $\boldsymbol{a}$, antecedent populations coancestry $\boldsymbol{\Lambda}$, and
          admixture proportions $\boldsymbol{Q}$
1  generate $\boldsymbol{P}$ using Algorithm 4
2  let $\boldsymbol{\Pi} = \boldsymbol{PQ}$
3  let $\boldsymbol{X} = \{x_{ij}\}$ by generating $x_{ij}|\pi_{ij} \sim \text{Binom}(2, \pi_{ij})$ independently for $i = 1, 2, \dots, m$
     and $j = 1, 2, \dots, n$
4  return $\boldsymbol{X}$
```

Line 1 can also be completed with the NORTA method, Algorithm B.4.

Algorithm 6: Generating bootstrap genotypes $\boldsymbol{X}^*$ from observed genotypes $\boldsymbol{X}$

```
   input: Genotype matrix $\boldsymbol{X}$
1  let $\hat{\boldsymbol{a}} = \{\hat{a}_i\}$ where $\hat{a}_i = \frac{1}{2n}\sum_{j=1}^{n} x_{ij}$ for $i = 1, 2, \dots, m$
2  obtain $\hat{\boldsymbol{\Lambda}}^{\text{sup}}$ and $\hat{\boldsymbol{Q}}$ from Algorithm 2 with input $\boldsymbol{X}$
3  generate $\boldsymbol{P}^*$ using Algorithm 4 with inputs $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{\Lambda}}^{\text{sup}}$
4  let $\boldsymbol{\Pi}^* = \boldsymbol{P}^*\hat{\boldsymbol{Q}}$
5  let $\boldsymbol{X}^* = \{x^*_{ij}\}$ by generating $x^*_{ij}|\pi^*_{ij} \sim \text{Binom}(2, \pi^*_{ij})$ independently for
     $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$
6  return $\boldsymbol{X}^*$
```

$\hat{\boldsymbol{\Lambda}}^{\text{sup}}$ can be replaced with $\hat{\boldsymbol{\Lambda}}^{\text{std}}$ in Line 2, in which case the BN sampling from Eq. (5) is used in Line 3. Line 3 can also be completed with the NORTA method, Algorithm B.4, if using $\hat{\boldsymbol{\Lambda}}^{\text{sup}}$.

We note that Algorithm 6 is a semi-parametric bootstrap simulation; Line 3 is semi-parametric, $\boldsymbol{\Pi}^*$ is semi-parametric because $\hat{\boldsymbol{Q}}$ is nonparametric, and Line 5 is parametric. The output $\boldsymbol{X}^*$ can be interpreted as a bootstrap replication of $\boldsymbol{X}$, where the population structure in $\boldsymbol{X}^*$ recapitulates the structure in $\boldsymbol{X}$. The process that the bootstrap method recapitulates is not just resampled genotypes for a fixed matrix of estimated IAFs. Rather, the antecedent population allele frequencies are resampled, also leading to resampled IAFs, so both evolutionary and statistical resampling occur.

2.6 Significance test of coancestry among antecedent populations

Here, we develop a hypothesis test of the standard admixture model (null) versus the super admixture model (alternative). We show below that on real data sets the test results are highly significant against the null in favor of the alternative. In terms of model parameters, the test is defined as follows:

$$H_0 : \max\left(\{\lambda_{uv}\}_{u\neq v}\right) = 0 \text{ (standard admixture model)}$$
$$H_1 : \max\left(\{\lambda_{uv}\}_{u\neq v}\right) > 0 \text{ (super admixture model)}$$

A straightforward test-statistic is $U = \|\hat{\mathbf{\Lambda}}^{\text{sup}} - \hat{\mathbf{\Lambda}}^{\text{std}}\|_F$. The larger U is, the more evidence there is against the null hypothesis in favor of the alternative hypothesis. In order to calculate a p-value for this test-statistic, we need to know the distribution of U when the null hypothesis is true. To this end, we adapt the bootstrap method of Algorithm 6, leading to Algorithm 7.

Algorithm 7: Hypothesis test of no coancestry among antecedent populations

```
input: Genotype matrix $\boldsymbol{X}$ and number of bootstrap replications $B$
calculate $\hat{a}_i = \frac{1}{2n}\sum_{j=1}^{n} x_{ij}$ for $i = 1, 2, \dots, m$
calculate estimates $\hat{\mathbf{\Lambda}}^{\text{std}}$, $\hat{\mathbf{\Lambda}}^{\text{sup}}$, and $\hat{\boldsymbol{Q}}$ by Algorithm 2 with input $\boldsymbol{X}$
calculate the observed test-statistic $U = \|\hat{\mathbf{\Lambda}}^{\text{sup}} - \hat{\mathbf{\Lambda}}^{\text{std}}\|_F$
for $b = 1, 2, \dots, B$ do
    generate $p^*_{iu} \sim \text{BN}(\hat{a}_i, \hat{\lambda}^{\text{std}}_{uu})$ independently and let $\boldsymbol{P}^* = \{p^*_{iu}\}$ for $i = 1, 2, \dots, m$
      and $u = 1, 2, \dots, K$
    let $\mathbf{\Pi}^* = \boldsymbol{P}^*\hat{\boldsymbol{Q}}$
    let $\boldsymbol{X}^* = \{x^*_{ij}\}$ by generating $x^*_{ij}|\pi^*_{ij} \sim \text{Binom}(2, \pi^*_{ij})$ independently for
      $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$
    calculate estimates $\hat{\mathbf{\Lambda}}^{\text{std}*}$ and $\hat{\mathbf{\Lambda}}^{\text{sup}*}$ by Algorithm 2 with input $\boldsymbol{X}^*$
    calculate the bootstrap null test-statistic $U^{*(b)} = \|\hat{\mathbf{\Lambda}}^{\text{sup}*} - \hat{\mathbf{\Lambda}}^{\text{std}*}\|_F$
return $p$-value $= \frac{1}{B}\sum_{b=1}^{B} 1(U^{*(b)} \geq U)$
```

To evaluate the validity of the proposed test, we performed this hypothesis testing on various simulation designs (Appendix C). Our simulations show that the test produces valid p-values, which are conservative (Fig. C.4), meaning the test has a maximum type I error rate less than or equal to the nominal level of the test. On real data sets analyzed below, these p-values are small, so the conservative behavior that we observe in simulations does not appear to be relevant for populations with nontrivial levels of structure.

3 Analysis of human studies

We applied the super admixture framework to four published studies: the human genome diversity panel (HGDP) [19], the 1000 genomes project (TGP) [20], the Human Origins study (HO) [21], and a study on individuals with Indian ancestry (IND) [22]. Within the TGP study, we also analyzed a subset of admixed populations with American ancestry, denoted by AMR. While HGDP, TGP, and HO are sampled from ancestries throughout the world, the IND and AMR data sets are regionally sampled. This yielded five data sets that collectively represent a range of population structures and study designs. Discussions of the results on HO, AMR, and IND are in the main text, while HGDP and TGP are in Appendix D.

3.1 Calculations

We processed the data sets and performed quality control checks to produce a genotype matrix $\boldsymbol{X}$ for each as the starting point of our analysis (Appendix D.1). We next applied Algorithm 2 to $\boldsymbol{X}$ to obtain $\hat{\boldsymbol{\Lambda}}^{\text{sup}}$ and $\hat{\boldsymbol{\Lambda}}^{\text{std}}$, the estimates of antecedent population coancestry for the super admixture and standard admixture models, respectively. We also calculated their corresponding individual-level coancestry estimates $\hat{\boldsymbol{\Theta}}^{\text{sup}}$ and $\hat{\boldsymbol{\Theta}}^{\text{std}}$. As a part of Algorithm 2, we calculated the appropriate number of antecedent populations K using the structural Hardy-Weinberg method [12] (detailed in Appendix D.6). The values of K ranged from $K = 11$ for HO to $K = 3$ for AMR, which are consistent with earlier work [10, 12, 28]. Also, in Algorithm 2 we calculated estimates of the admixture proportion matrices $\hat{\boldsymbol{Q}}$ using the ALStructure method [10].

To evaluate the accuracy of $\hat{\boldsymbol{\Theta}}^{\text{sup}}$ and $\hat{\boldsymbol{\Theta}}^{\text{std}}$, we computed the OS estimate [4] of individual-level coancestry $\hat{\boldsymbol{\Theta}}^{\text{OS}}$ on each data set. The OS estimate of $\boldsymbol{\Theta}$ is based on general assumptions and is a consistent estimator for arbitrary population structures under the appropriate conditions. Since $\hat{\boldsymbol{\Theta}}^{\text{OS}}$ makes no assumptions about the distributions of the IAFs or coancestry parameters, it serves as a benchmark for our methods [1], allowing us to observe if the super admixture or standard admixture models lose information about individual-level coancestry relative to OS. As shown in Table D.1, the Frobenius-based distances from $\hat{\boldsymbol{\Theta}}^{\text{sup}}$ to $\hat{\boldsymbol{\Theta}}^{\text{OS}}$ are about 10 to 40 times smaller than those from $\hat{\boldsymbol{\Theta}}^{\text{std}}$ to $\hat{\boldsymbol{\Theta}}^{\text{OS}}$. The distance from $\hat{\boldsymbol{\Theta}}^{\text{sup}}$ to $\hat{\boldsymbol{\Theta}}^{\text{OS}}$ is smaller than is arguably practically relevant, meaning that $\hat{\boldsymbol{\Theta}}^{\text{sup}}$ achieves the resolution of $\hat{\boldsymbol{\Theta}}^{\text{OS}}$ for practical purposes.

[1] Note also that the OS estimate of $\boldsymbol{\Theta}$ is equal to the OS estimate of kinship, $\boldsymbol{\Phi}$, except for the diagonal elements where $\hat{\theta}_{jk}^{\text{OS}} = 2\hat{\phi}_{jk}^{\text{OS}} - 1$, as shown in Eq. (3).

We carried out Algorithm 7 to perform a hypothesis test of the standard admixture model versus the super admixture model for all five datasets, with $B = 1000$ bootstrap iterations. For all data sets, no bootstrap null test-statistic was equal to or greater than the observed test-statistic, so p-value < 0.001 for all data sets. The bootstrap null test-statistics and observed test-statistic for all data sets are shown in Fig. D.9.

We applied Algorithm 6 to generate bootstrap replications $\boldsymbol{X}^*$ from each data set's genotype matrix $\boldsymbol{X}$. We applied the double-admixture method (Algorithm 4) and the NORTA method (Algorithm B.5) to include the performance of both methods. We computed the OS estimate $\hat{\boldsymbol{\Theta}}^{\text{OS}*}$ of individual-level coancestry for each $\boldsymbol{X}^*$.

3.2 Visualizing results

We firstly visualized the results by making heatmaps of individual-level coancestry estimates $\hat{\boldsymbol{\Theta}}^{\text{OS}}$, $\hat{\boldsymbol{\Theta}}^{\text{sup}}$, and $\hat{\boldsymbol{\Theta}}^{\text{std}}$. We also made heatmaps of $\hat{\boldsymbol{\Theta}}^{\text{OS}*}$ from bootstrap resampled genotypes using both the double-admixture and NORTA methods for generating antecedent population allele frequencies. These are displayed as follows: HO – Fig. 2, AMR – Fig. 4, IND – Fig. 6, HGDP – Fig. D.5, and TGP – Fig. D.7. It can be seen that for all data sets, $\hat{\boldsymbol{\Theta}}^{\text{OS}}$ and $\hat{\boldsymbol{\Theta}}^{\text{sup}}$ are qualitatively equivalent, which is quantitatively supported by Table D.1 showing they are very close. The estimates $\hat{\boldsymbol{\Theta}}^{\text{OS}*}$ from the two bootstrap methods are also qualitatively equivalent to $\hat{\boldsymbol{\Theta}}^{\text{OS}}$ and $\hat{\boldsymbol{\Theta}}^{\text{sup}}$. Finally, it can be seen that the standard admixture coancestry estimate $\hat{\boldsymbol{\Theta}}^{\text{std}}$ is not close to the other estimates, further indicating the standard admixture model is not sufficient for these data sets.

We secondly visualized the results by building on the standard colored stacked bar plots of $\hat{\boldsymbol{Q}}$ displaying the admixture proportions of the K antecedent populations for the individuals. In our case, we have additional information, which is the estimated antecedent population coancestry matrix $\hat{\boldsymbol{\Lambda}}^{\text{sup}}$ from the super admixture model. This matrix gives additional information about the relationship among the antecedent populations that we would like to visualize. The first way we visualized $\hat{\boldsymbol{\Lambda}}^{\text{sup}}$ was create a heatmap of its values. We then constructed a dendogram built from $\hat{\boldsymbol{\Lambda}}^{\text{sup}}$ that is displayed above the stacked bar plot. This gives the user insight into the relatedness of the antecedent populations and connect them to the stacked bar plots. To this end, we calculated a distance matrix $\boldsymbol{D}$ from $\hat{\boldsymbol{\Lambda}}^{\text{sup}}$

according to:

$$d_{uv} = \begin{cases} 0 & \text{if } u = v \\ \max(\hat{\mathbf{\Lambda}}^{\text{sup}}) - \hat{\lambda}^{\text{sup}}_{uv} & \text{if } u \neq v. \end{cases}$$

We then applied the standard agglomerative clustering method to $\boldsymbol{D}$ using “weighted pair group method with arithmetic mean” (WPGMA) to obtain a dendrogram. These are displayed in the data sets as follows: HO – Fig. 3, AMR – Fig. 5, IND – Fig. 7, HGDP – Fig. D.6, and TGP – Fig. D.8.

3.3 Human Origins (HO) study

The Human Origins datasets (HO) consists of 2124 individuals from 170 sub-subpopulations grouped into 11 subpopulations. We observed the estimated individual-level coancestry agrees with current knowledge of early human migrations [29–32]. In Fig. 2, we observed the first major split between Sub-Saharan Africa and North Africa. This split reflects the divergence between Sub-Saharan Africans and the rest of human populations resulting from an out-of-Africa migration around 50-60 kya. Another split occurred between South Asia and East Asia, revealing the separation between West Eurasians and East Asians around 40-45 kya. Among the East Asia clade, we identified that the Oceanians have highest coancestry within and lowest coancestry between other subpopulations, consistent with the theory that Oceanians split earliest from the rest of East Asians.

The coancestry among antecedent populations is also compatible with early human dispersals (Fig. 3). Specifically, in the dendrogram plot of the antecedent population coancestry (Fig. 3B), we note that the first branch split individuals from Sub-Saharan Africa represented by the antecedent populations S_1 and S_2 from individuals outside of Sub-Saharan Africa represented by the other antecedent populations. Individuals outside of Sub-Saharan Africa further branched off into two lineages: the West Eurasians represented by antecedent populations S_3, S_4 and S_5, and the East Asians represented by antecedent populations S_6 - S_{11}. Then the Oceanians represented by the antecedent population S_9 split off from the majority of East Asian ancestry, while the latter further diverged into present-day Asians (antecedent populations S_6, S_7, S_8) and present-day Americans (antecedent populations S_{10} and S_{11}).

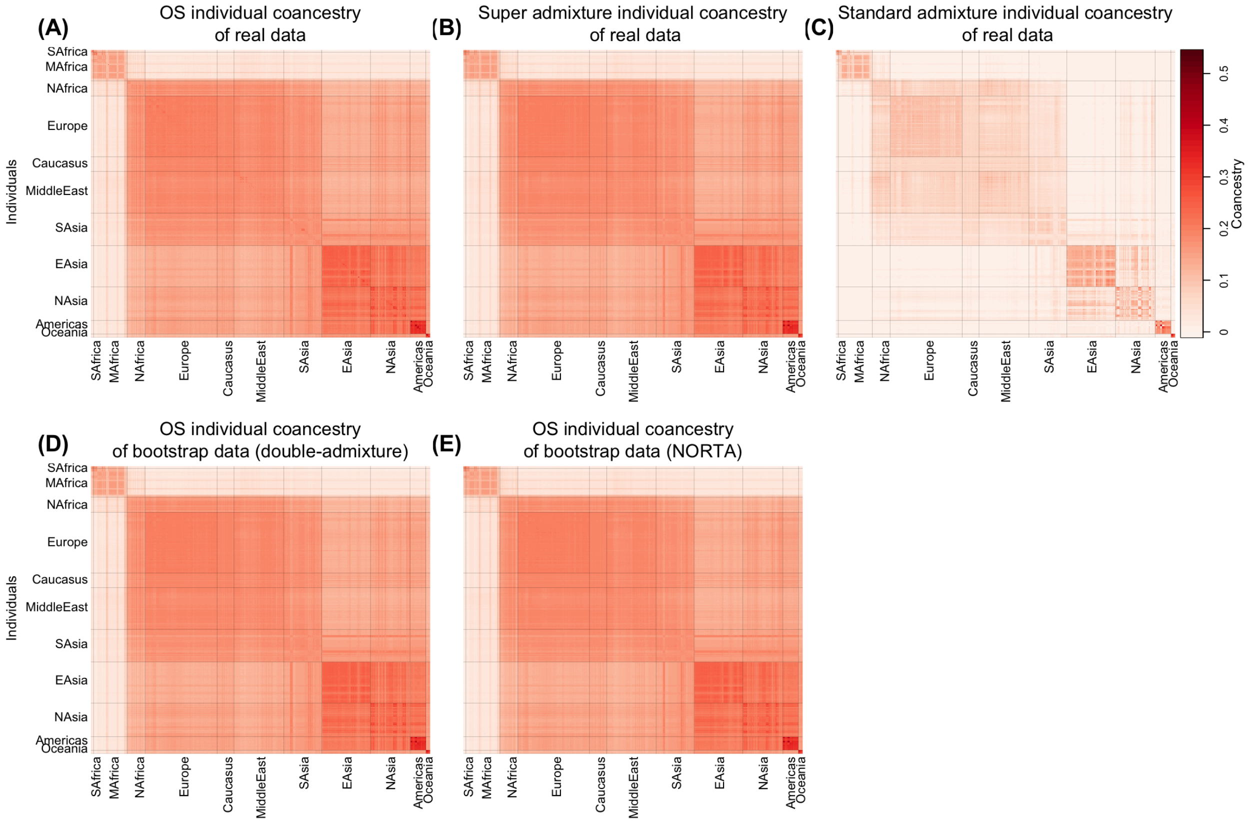


Figure 2: Heatmaps of individual-level coancestry estimates in the HO data set.

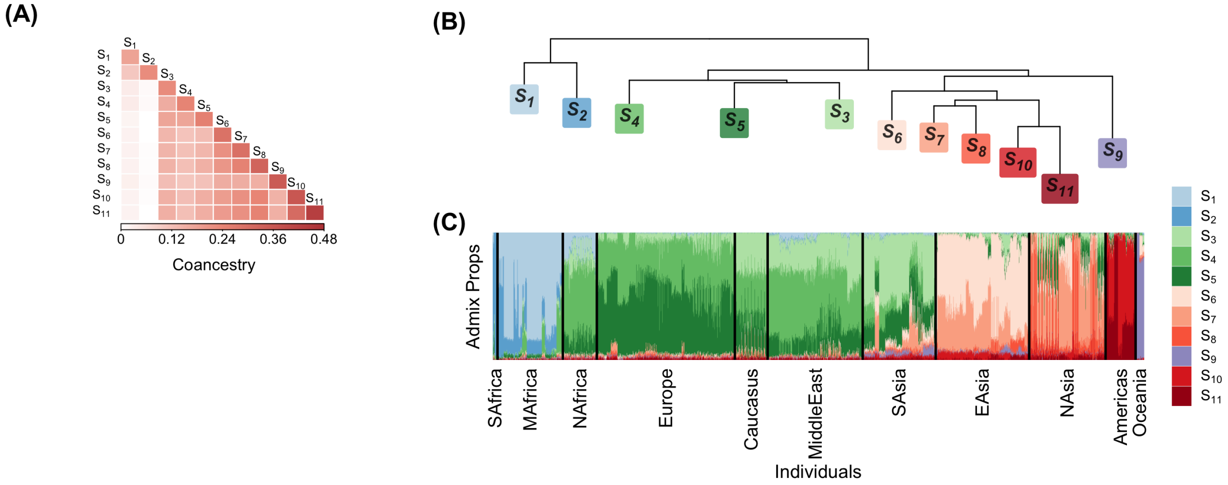


Figure 3: (A) Heatmap of antecedent population coancestry estimates in the HO data set. (B) Dendrogram representation of the antecedent population coancestry estimates. (C) Stacked bar plot of admixture proportions.

3.4 Admixed individuals (AMR) from the 1000 Genomes Project (TGP)

The AMR subset of TGP has 353 individuals from four regions (Mexican-American (MXL): 65, Puerto Rican (PUR): 104, Colombian (CLM): 97, Peruvian (PEL): 87). The individual-level coancestry plot (Fig. 4) revealed that this dataset does not have a discrete population structure. Instead, the coancestry changes smoothly over individuals, indicating wide-ranging historical admixture events. This is consistent with the AMR population descending from European, Native American, and Sub-Saharan African ancestries during the post-Columbian era [33, 34].

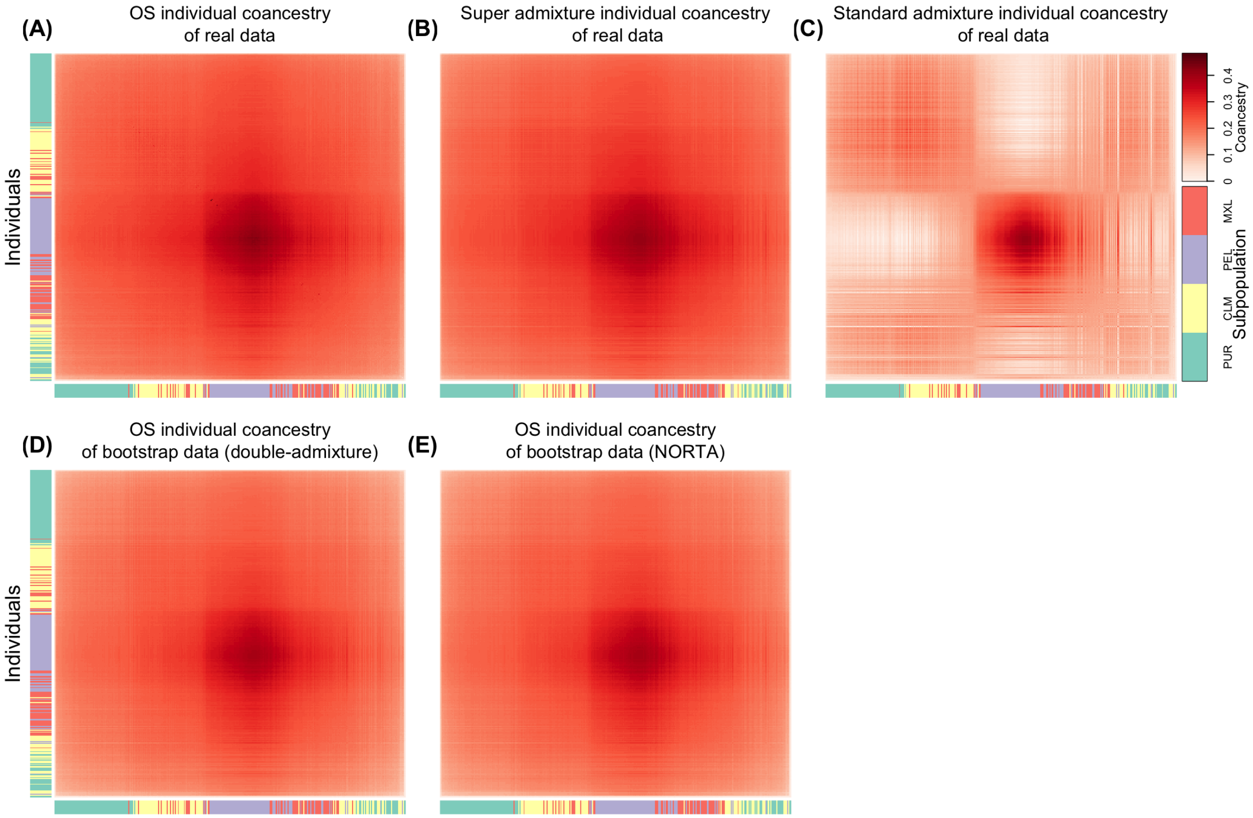


Figure 4: Heatmaps of individual-level coancestry estimates in the AMR data set.

In the analysis of the coancestry among antecedent populations (Fig. 5), we identified three major sources of ancestry: Sub-Saharan African ancestry represented by the antecedent population S_1, West Eurasian ancestry represented by the antecedent population S_2, and Native American ancestry represented by the antecedent population S_3. The first split occurred between Sub-Saharan Africans (S_1) and individuals outside of Sub-Saharan Africa (S_2 and S_3), and the second split between the West Eurasians (S_2) and the Native Americans

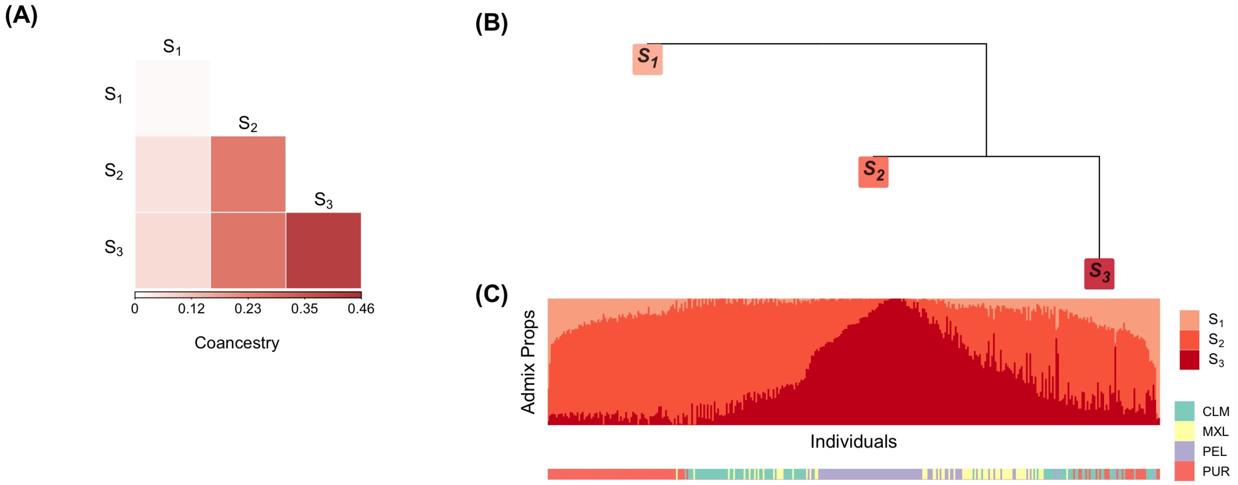


Figure 5: (A) Heatmap of antecedent population coancestry estimates in AMR. (B) Dendrogram representation of the antecedent population coancestry estimates. (C) Stacked bar plot of admixture proportions.

(S_3). We also noted that the Puerto Ricans contain the highest amount of Sub-Saharan African ancestry; the Peruvians have the highest proportion of Native American ancestry; the Colombians and the Mexican-Americans display extensive variation in in their admixture proportions of European and Native American ancestry. Our observations were confirmed by previous analyses of AMR populations [28, 33, 34].

3.5 Indian (IND) study

We combined the mainland Indians from the IND study with the Central/South Asia and the East Asia populations from HGDP to study the relationship between present-day Indians and other populations in Asia. Our merged data set consists of 298 mainland Indians from fou linguistic groups (Indo-European (IE): 92, Dravidian: 53, Austro-Asiatic (AA): 79, Tibeto-Burman (TB): 74), together with 190 Central/South Asians and 210 East Asians from HGDP. Previous analyses of South Asian populations have shown that the Indo-European speakers show a considerable amount of the Western Eurasian relatedness and are ancestrally close to Central Asians. The Austro-Asiatic speakers and the Tibeton-Burman speakers were mixed from East Asian ancestry. The Tibeton-Burman speakers generally have significant genomic proportions derived from East Asian ancestry so that some Tibeton-Burman speakers can be difficult to distinguish from East Asian populations based on genome-wide measures of relatedness. Consistent with these findings [22, 35, 36], we observe a split between Indo-European speakers and the rest of mainland Indians in the heatmap of individual-level coancestry (Fig. 6). The Indo-European speakers and the Central/South Asians of HGDP have relatively similar levels of coancestry. The second split occurred between the Austro-Asiatic speakers and the Tibeto-Burman speakers. The Tibeto-Burman speakers and East Asians of HGDP have relatively similar levels of coancestry.

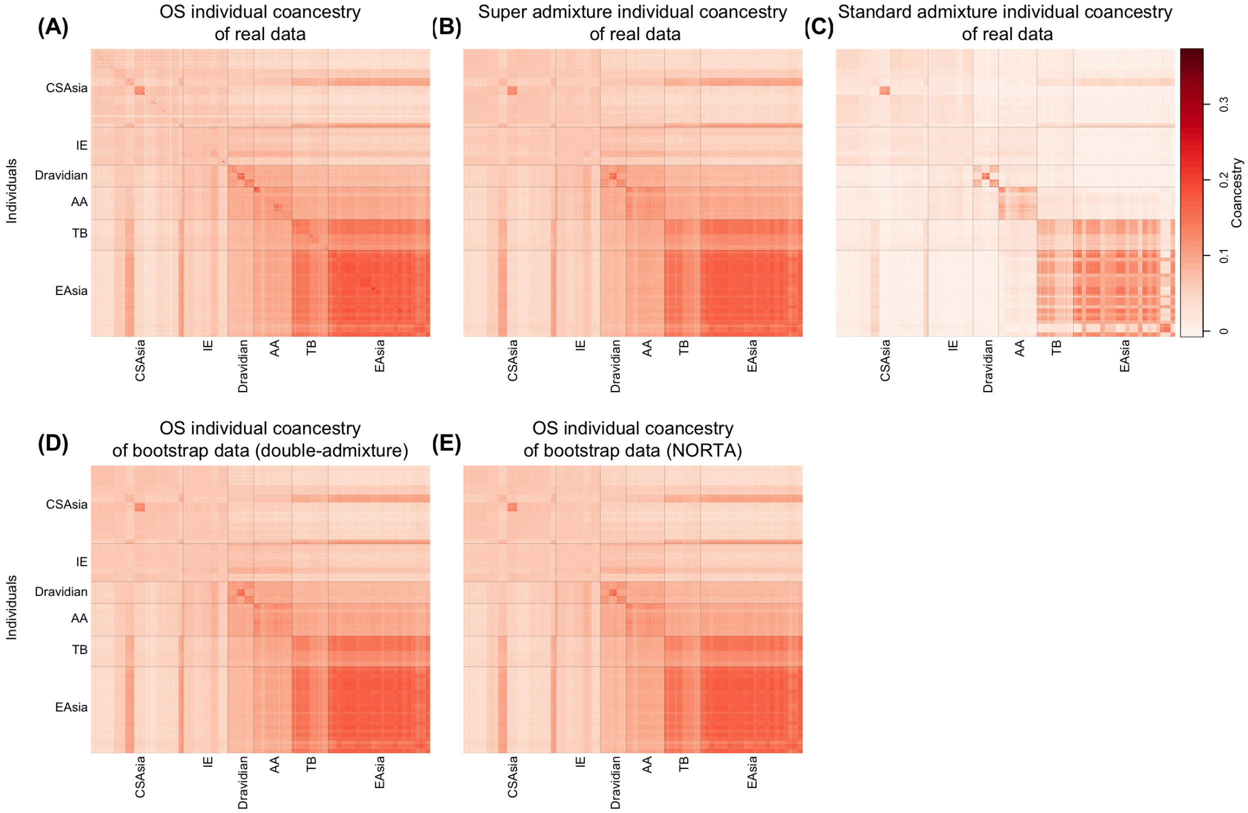


Figure 6: Heatmap of individual-level coancestry estimates in the merged data set of mainland Indians from IND, and Central/South Asians and East Asians from HGDP.

Our analysis reveals that there are three major branches of antecedent populations for this dataset (Fig. 7). The branch of antecedent populations S_1 and S_2 is most prevalent in Central/South Asians of HGDP and Indo-European speakers, suggesting this branch was at least partially derived from a West Eurasian source. The branch of the antecedent populations S_3, S_4 and S_5 is widespread in Dravidian speakers and Austro-Asiatic speakers, indicating it is relevant to South Indian ancestry and Austro-Asiatic speaker ancestry. The third branch of the antecedent populations S_6 and S_7 likely represents East Asian ancestry due to its high prevalence in the Tibeto-Burman speakers and East Asians of HGDP.

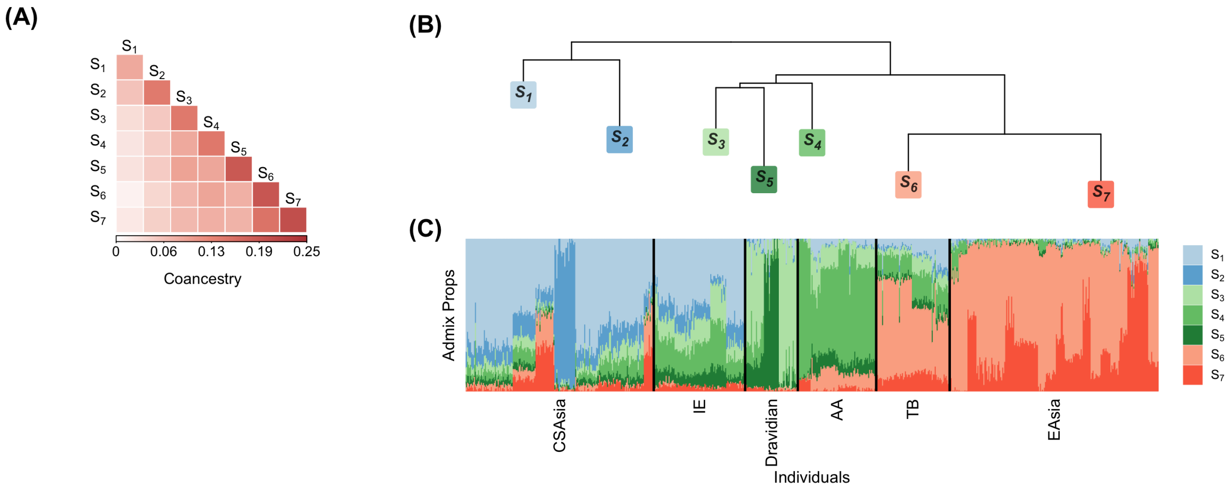


Figure 7: Heatmap of antecedent population coancestry estimates in the merged data set of mainland Indians from IND, and Central/South Asians and East Asians from HGDP. (B) Dendrogram representation of the antecedent population coancestry estimates. (C) Stacked bar plot of admixture proportions.

4 Discussion

The super admixture framework is an extension of the highly used admixture model. It superposes coancestry among the admixed antecedent populations. It provides a forward generating probability process that encompasses random evolutionary, genealogical, and statistical sampling processes. The antecedent populations are modeled to have an arbitrarily complex coancestry. This allows the generation of individual-specific allele frequencies (IAFs) that capture complex population structures and permit the estimation of individual-level coancestry that is at the resolution of general individual-level coancestry and kinship estimators for arbitrarily complex structures.

There are numerous parameters estimated from genome-wide genotype data that relate to structure, such as coancestry, inbreeding, and F_{ST}. When traits are included, one often estimates parameters in the context of genome-wide association studies [23, 37], genome-wide heritability [38–40] and polygenic risk scores [41, 42]. There does not exist a straightforward, general method for quantifying uncertainty among these various estimates. Within our framework, we have shown how to perform a bootstrap resampling method that randomly generates new genetic data that recapitulate population structure observed in real data. This bootstrap method may provide a way to formulate general methods for quantifying uncertainty in genome-wide genotype studies.

We developed a hypothesis test where one can test the standard versus super admixture model on real data. When we applied it to the five data sets analyzed here, all of them were highly significant in rejecting the standard admixture model in favor of the super admixture model. The individual-level coancestry estimates from the super admixture model also agreed with the general coancestry estimate, whereas the standard admixture individual-level coancestry estimates did not.

The stacked bar plot visualization of admixture proportions among individuals is ubiquitous in analyzing population structure. We showed here how the estimated antecedent population coancestry can be plotted with the stacked bar plot to visualize the relationship among the antecedent populations in conjunction with the bar plot. The admixture proportions among individuals are then interpretable in terms of the evolutionary history of the antecedent populations. We demonstrated this visualization on five data sets and showed how it agreed with known results on these human populations.

Understanding population structure in humans is one of the central problems in modern genetics. We demonstrated that the proposed super admixture framework is a powerful tool for learning admixed population coancestry, improving the analysis of genetic data from structured populations, bridging admixture with individual-level coancestry and kinship, and simulating new data reflecting a structured population. We anticipate that the super admixture framework will be useful in analyzing complex population structure in future applications.

Resources

The `superadmixture` software package is available at `https://github.com/StoreyLab/superadmixture`. The results in this paper can be reproduced with code available at `https://github.com/StoreyLab/superadmixture-manuscript-analysis`.

Acknowledgments

This work was supported in part by US National Institutes of Health grant R01 HG006448.

Appendices

A Supplementary theory

A.1 Mathematical notation and definitions

Definition 1. Given a vector $\boldsymbol{x} \in \mathbb{R}^n$, the ℓ_p norm is defined as:

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

Definition 2. Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, the maximum singular value of $\boldsymbol{A}$ is denoted as $\sigma_{\max}(\boldsymbol{A})$. The maximum and minimum eigenvalues of $\boldsymbol{A}$ are denoted as $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$.

Definition 3. Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, the Frobenius norm of $\boldsymbol{A}$ is defined by:

$$\|\boldsymbol{A}\|_F = \sqrt{\operatorname{tr}(\boldsymbol{A}'\boldsymbol{A})} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}.$$

Definition 4. The induced matrix norm $\|\boldsymbol{A}\|_{a,b}$ is defined as:

$$\|\boldsymbol{A}\|_{a,b} = \sup\{\|\boldsymbol{A}\boldsymbol{x}\|_b : \|\boldsymbol{x}\|_a = 1\}.$$

When $a = b = 2$, the induced matrix norm is the spectral norm:

$$\|\boldsymbol{A}\|_2 \equiv \|\boldsymbol{A}\|_{2,2} = \sqrt{\lambda_{\max}(\boldsymbol{A}'\boldsymbol{A})} = \sigma_{\max}(\boldsymbol{A}).$$

When $a = b = 1$, the induced matrix norm is the maximum absolute column sum of the matrix:

$$\|\boldsymbol{A}\|_1 \equiv \|\boldsymbol{A}\|_{1,1} = \max_j \sum_{i=1}^{m} |a_{ij}|.$$

Definition 5. The proximal operator $\operatorname{prox}_f(\boldsymbol{x})$ is defined as

$$\operatorname{prox}_f(\boldsymbol{x}) = \underset{\boldsymbol{u}\in\mathbb{R}^n}{\arg\min}\, f(\boldsymbol{u}) + \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{x}\|_2^2.$$

Let $f(\boldsymbol{x})$ be an indicator function defined as

$$1_C(\boldsymbol{x}) = \begin{cases} 0 & \boldsymbol{x} \in C \\ \infty & \boldsymbol{x} \notin C \end{cases}$$

where C is a nonempty subset $\mathbb{R}^n$. Let $\mathcal{P}_C$ denote the "projection onto C operator". Then

$$\operatorname{prox}_f(\boldsymbol{x}) = \arg\min_{\boldsymbol{u} \in C} \|\boldsymbol{u} - \boldsymbol{x}\|_2^2 = \mathcal{P}_C(\boldsymbol{x}).$$

The n-dimensional unit simplex is defined as

$$\Delta = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = 1, 0 \leq x_i \leq 1 \right\}.$$

We define the projection onto the unit simplex operator $\mathcal{P}_\Delta$ as

$$\mathcal{P}_\Delta(\boldsymbol{x}) = \arg\min_{\boldsymbol{u} \in \Delta} \|\boldsymbol{u} - \boldsymbol{x}\|_2^2.$$

Definition 6. A subset S of $\mathbb{R}^n$ is a real semi-algebraic set if there exists a finite number of real polynomial functions g_{ij}, h_{ij}: $\mathbb{R}^n \to \mathbb{R}$ such that

$$S = \cup_{j=1}^{p} \cap_{i=1}^{q} \{\boldsymbol{u} : \mathbb{R}^n : g_{ij}(\boldsymbol{u}) = 0 \text{ and } h_{ij}(\boldsymbol{u}) < 0\}.$$

Definition 7. A function $f : \mathbb{R}^n \to (-\infty, \infty]$ is called semi-algebraic if its graph

$$\{(\boldsymbol{u}, t) \in \mathbb{R}^{n+1} : f(\boldsymbol{u}) = t\}$$

is a semi-algebraic subset of $\mathbb{R}^{n+1}$.

A.2 Lemmas supporting the algorithms

Lemma 1. The definition of the induced matrix norm $\|\boldsymbol{A}\|_{a,b}$ implies for any $\boldsymbol{x} \in \mathbb{R}^n$, the sub-additivity holds:

$$\|\boldsymbol{A}\boldsymbol{x}\|_b \leq \|\boldsymbol{A}\|_{a,b} \|\boldsymbol{x}\|_a.$$

Lemma 2. Given matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n}$,

$$\sigma_{\min}(\boldsymbol{A}) \|\boldsymbol{B}\|_F \leq \|\boldsymbol{A}\boldsymbol{B}\|_F \leq \sigma_{\max}(\boldsymbol{A}) \|\boldsymbol{B}\|_F = \|\boldsymbol{A}\|_2 \|\boldsymbol{B}\|_F.$$

Lemma 3. Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$,

$$\frac{1}{m}\|\boldsymbol{A}\|_1 \leq \|\boldsymbol{A}\|_2 \leq \sqrt{n}\|\boldsymbol{A}\|_1.$$

Lemma 4. Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper and closed function. If f is semi-algebraic then it satisfies the Kurdyka-Lojasiewicz (KL) property at any point of $\mathrm{dom}(f)$ [26].

B Supplementary methods

B.1 Estimating individual-level pairwise coancestry

The OS estimate [4] begins with a measurement for allele matching between each pair of individuals:

$$A_{jk} = \frac{1}{m}\sum_{i=1}^{m}(x_{ij}-1)(x_{ik}-1)-1.$$

They prove that the expectation of A_{jk} is

$$\mathbb{E}[A_{jk}] = \begin{cases} \frac{(\theta_{jj}-1)\nu}{2} & j = k \\ (\theta_{jk}-1)\nu & j \neq k \end{cases}$$

where $\nu = \frac{4}{m}\sum_{i=1}^{m} a_i(1-a_i)$. Let $\underline{\theta}$ denote $\min_{j,k}\theta_{jk}$. The Ochoa-Storey (OS) coancestry estimate utilizes $\underline{\theta} = 0$, which sets the reference population T to the most recent common ancestral (MRCA) population. When $\underline{\theta} = 0$, then $\nu = -\underline{A}$ where $\underline{A} = \min_{j,k}\mathbb{E}[A_{jk}]$. Then OS estimator of coancestry is

$$\hat{\theta}^{\text{OS}}_{jk} = \begin{cases} 1 - \frac{2A_{jj}}{\underline{A}} & j = k \\ 1 - \frac{A_{jk}}{\underline{A}} & j \neq k \end{cases}.$$

In general, $-\underline{A} = (\underline{\theta}-1)\nu$. One can extend the OS estimator of coancestry for general values of $\underline{\theta}$ as follows:

$$\hat{\theta}^{\text{OS}}_{jk} = \begin{cases} 1 - \frac{2(\underline{\theta}-1)A_{jj}}{\underline{A}} & j = k \\ 1 - \frac{(\underline{\theta}-1)A_{jk}}{\underline{A}} & j \neq k \end{cases}.$$

B.2 Estimating coancestry among antecedent populations

Proximal Forward-Backward (PFB) algorithm. Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a be proper and closed function, let $h : \mathbb{R}^n \to (-\infty, +\infty)$ be convex and differentiable with a L-Lipschitz continuous gradient ∇h, i.e.,

$$\|\nabla h(\boldsymbol{x}_2) - \nabla h(\boldsymbol{x}_1)\|_2 \leq L\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_2 \quad \forall(\boldsymbol{x}_1, \boldsymbol{x}_2),$$

where $L \in (0, \infty)$. Suppose that $f(\boldsymbol{x}) + h(\boldsymbol{x}) \to \infty$ as $\|\boldsymbol{x}\|_2 \to \infty$. The problem is to identify:

$$\arg\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}) + h(\boldsymbol{x}).$$

It has been shown that this problem can be solved by the PFB algorithm [24]. Every sequence $(\boldsymbol{x}_t)_{t \in \mathbb{N}}$ generated by the following constant-step forward-backward algorithm converges to a solution to the problem.

Algorithm B.1: The constant-step forward-backward algorithm

```
Initialize $\boldsymbol{x}_0$
for $t = 1, 2, \ldots$ do
    $\boldsymbol{x}^* \leftarrow \boldsymbol{x}_{t-1} - \frac{1}{L}\nabla h(\boldsymbol{x}_{t-1})$
    $\boldsymbol{x}_t \leftarrow \text{prox}_{L^{-1}f}(\boldsymbol{x}^*)$
return $\boldsymbol{x}_t$
```

prox(·) denotes the proximal operator (Appendix A.1).

Solving Problem 1 by PFB. Problem 1 is equivalent to the problem of finding the minimizer of $f(\boldsymbol{\Lambda}) + h(\boldsymbol{\Lambda})$ where

$$f(\boldsymbol{\Lambda}) = \begin{cases} 0 & \boldsymbol{\Lambda} \text{ is symmetric and } 0 \leq \lambda_{uv} \leq 1, u, v = 1, \ldots, K \\ \infty & \text{otherwise} \end{cases}$$
$$h(\boldsymbol{\Lambda}) = \|\boldsymbol{\Theta} - \boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}\|_F^2.$$

The function f is proper and closed because $\text{dom}(f)$ is nonempty and closed. The function h is differentiable with a continuous gradient $\nabla h = -2\boldsymbol{Q}(\boldsymbol{\Theta} - \boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q})\boldsymbol{Q}'$. ∇h is Lipschitz continuous with $L = \sigma_{\max}^4(\boldsymbol{Q})$.

Proof. We note that

$$\begin{aligned} \|\nabla h(\boldsymbol{\Lambda}_2) - \nabla h(\boldsymbol{\Lambda}_1)\|_F &= \| - 2\boldsymbol{Q}(\boldsymbol{\Theta} - \boldsymbol{Q}'\boldsymbol{\Lambda}_2\boldsymbol{Q})\boldsymbol{Q}' + -2\boldsymbol{Q}(\boldsymbol{\Theta} - \boldsymbol{Q}'\boldsymbol{\Lambda}_1\boldsymbol{Q})\boldsymbol{Q}'\|_F \\ &= 2\|\boldsymbol{Q}\boldsymbol{Q}'(\boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_1)\boldsymbol{Q}\boldsymbol{Q}'\|_F \leq 2\sigma_{\max}^2(\boldsymbol{Q}\boldsymbol{Q}')\|\boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_1\|_F \quad (\textit{Lemma } 2) \\ &= 2\sigma_{\max}^4(\boldsymbol{Q})\|\boldsymbol{\Lambda}_2 - \boldsymbol{\Lambda}_1\|_F \end{aligned}$$

□

Therefore, we can employ the PFB algorithm to solve Problem 1. The proximal operator

$\text{prox}_{L^{-1}f}(\boldsymbol{\Lambda})$ can be calculated as $\text{prox}_{L^{-1}f}(\boldsymbol{\Lambda}) = \mathcal{P}_{\text{dom}(f)}(\boldsymbol{\Lambda})$. This implies

$$\{\text{prox}_{L^{-1}f}(\boldsymbol{\Lambda})\}_{uv} = \max(0, \min(\lambda_{uv}, 1)).$$

This leads to Algorithm 1, where we have now proved that every sequence $(\boldsymbol{\Lambda}_t)_{t\in\mathbb{N}}$ converges to a solution.

Solving Problem B.1 by PFB. We can formulate the estimation of coancestry among antecedent populations under the standard admixture model as follows.

Problem B.1.

$$\begin{aligned}
&\min_{\boldsymbol{\Lambda}\in\mathbb{R}^{K\times K}} \|\boldsymbol{\Theta} - \boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}\|_F^2 \\
\text{subject to: } & 0 \leq \lambda_{uu} \leq 1 \\
& \lambda_{uv} = 0 \quad \forall u \neq v \\
& u, v = 1, 2, \ldots, K
\end{aligned}$$

It is straightforward to see that Problem B.1 is identical to identifying the minimizer of $f(\boldsymbol{\Lambda}) + h(\boldsymbol{\Lambda})$ where

$$\begin{aligned}
f(\boldsymbol{\Lambda}) &= \begin{cases} 0 & 0 \leq \lambda_{uu} \leq 1; \lambda_{uv} = 0, \forall u \neq v; u, v = 1, \ldots, K \\ \infty & \text{otherwise} \end{cases}, \\
h(\boldsymbol{\Lambda}) &= \|\boldsymbol{\Theta} - \boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}\|_F^2.
\end{aligned}$$

The function f is proper and closed and h is differentiable with a continuous gradient $\nabla h = -2\boldsymbol{Q}(\boldsymbol{\Theta} - \boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q})\boldsymbol{Q}'$. The gradient ∇h is Lipschitz continuous with $L = \sigma_{\max}^4(\boldsymbol{Q})$. By Appendix A.1, $\text{prox}_{L^{-1}f}(\boldsymbol{\Lambda}) = \mathcal{P}_{\text{dom}(f)}(\boldsymbol{\Lambda})$. This implies:

$$\{\text{prox}_{L^{-1}f}(\boldsymbol{\Lambda})\}_{uv} = \begin{cases} \max(0, \min(\lambda_{uu}, 1)) & u = v \\ 0 & u \neq v \end{cases}.$$

This leads to Algorithm B.2, where every sequence $(\boldsymbol{\Lambda}_t)_{t\in\mathbb{N}}$ converges to a solution.

Algorithm B.2: Estimating $\boldsymbol{\Lambda}$ for the standard admixture model given $\boldsymbol{\Theta}$ and $\boldsymbol{Q}$

```
input: Coancestry matrix $\boldsymbol{\Theta}$ and admixture proportions matrix $\boldsymbol{Q}$
let $L = \sigma^4_{\max}(\boldsymbol{Q})$
let $\boldsymbol{\Lambda}_0 \leftarrow (\boldsymbol{QQ}')^{-1}\boldsymbol{\Theta}(\boldsymbol{QQ}')^{-1}$
for $t = 1, 2, \ldots$ do
    $\boldsymbol{G} \leftarrow 2\boldsymbol{Q}(\boldsymbol{Q}'\boldsymbol{\Lambda}_{t-1}\boldsymbol{Q} - \boldsymbol{\Theta})\boldsymbol{Q}'$
    $\boldsymbol{\Lambda}^* \leftarrow \boldsymbol{\Lambda}_{t-1} - \frac{1}{L}\boldsymbol{G}$
    $\boldsymbol{\Lambda}_t = \{\lambda_{uv,t}\}$ where
        $\lambda_{uv,t} = \begin{cases} \max(0, \min(1, \lambda^*_{uu})), & \text{if } u = v \\ 0, & \text{if } u \neq v \end{cases}$
return $\boldsymbol{\Lambda}_t$
```

$\sigma_{\max}(\cdot)$ denotes the maximum singular value (Appendix A.1).

B.3 Estimating Parameters in the Double-Admixture Model

Proximal Alternating Linearized Minimization (PALM) agloritihm. Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ and $g : \mathbb{R}^m \to (-\infty, +\infty)$ be closed functions. Let $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a continuously differentiable function. The problem is to find a solution to:

$$\underset{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{y}\in\mathbb{R}^m}{\arg\min} \; \Psi(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}) + g(\boldsymbol{y}) + h(\boldsymbol{x}, \boldsymbol{y})$$

over all $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^n \times \mathbb{R}^m$. It has been shown that this problem can be solved by the Proximal Alternating Linearized Minimization (PALM) algorithm. Assume that:

(i) $\inf_{\mathbb{R}^n \times \mathbb{R}^m} \Psi > -\infty$, $\inf_{\mathbb{R}^n} f > -\infty$ and $\inf_{\mathbb{R}^m} g > -\infty$.

(ii) Ψ is a Kurdyka-Lojasiewicz function (see Appendix A.1).

(iii) h is twice continuously differentiable.

(iv) There exists convex and compact sets $\mathcal{C}_{\boldsymbol{x}}$ and $\mathcal{C}_{\boldsymbol{y}}$ such that $\boldsymbol{x}_t \in \mathcal{C}_{\boldsymbol{x}}$ and $\boldsymbol{y}_t \in \mathcal{C}_{\boldsymbol{y}}$ for all $t \in \mathbb{N}$.

(v) For any fixed $\boldsymbol{y}$ the partial gradient $\nabla_{\boldsymbol{x}} h(\boldsymbol{x}, \boldsymbol{y})$ is Lipschitz continuous with moduli $L_1(\boldsymbol{y})$ over the domain $\mathcal{C}_{\boldsymbol{x}}$, that is

$$\|\nabla_{\boldsymbol{x}} h(\boldsymbol{x}_1, \boldsymbol{y}) - \nabla_{\boldsymbol{x}} h(\boldsymbol{x}_2, \boldsymbol{y})\|_2 \leq L_1(\boldsymbol{y})\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_2, \quad \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C}_{\boldsymbol{x}}.$$

Likewise, for any fixed $\boldsymbol{x}$ the partial gradient $\nabla_{\boldsymbol{y}} h(\boldsymbol{x}, \boldsymbol{y})$ is Lipschitz continuous with moduli $L_2(\boldsymbol{x})$ over the domain $\mathcal{C}_{\boldsymbol{y}}$.

(vi) For $i = 1, 2$ there exists $\lambda_i^-, \lambda_i^+ > 0$ such that:

$$
\begin{aligned}
\inf\{L_1(\boldsymbol{y}_t) : t \in \mathbb{N}\} &\geq \lambda_1^- \\
\inf\{L_2(\boldsymbol{x}_t) : t \in \mathbb{N}\} &\geq \lambda_2^- \\
\sup\{L_1(\boldsymbol{y}_t) : t \in \mathbb{N}\} &\leq \lambda_1^+ \\
\sup\{L_2(\boldsymbol{x}_t) : t \in \mathbb{N}\} &\leq \lambda_2^+
\end{aligned}
$$

We note that these assumptions are not exactly the same as the assumptions specified in ref. [26]. We modified the original assumptions to align the PALM algorithm to our setting. Following the proof provided in ref. [26], one can show that if these assumptions are met, the sequence $(\boldsymbol{x}_t, \boldsymbol{y}_t)_{t\in\mathbb{N}}$ generated by Algorithm B.3 converges to a critical point of Ψ.

Algorithm B.3: The general PALM algorithm

```
Initialization: start with any $\boldsymbol{x}_0 \in \mathcal{C}_{\boldsymbol{x}}$ and $\boldsymbol{y}_0 \in \mathcal{C}_{\boldsymbol{y}}$
for $t = 1, 2, \ldots$ do
    take $\tau_1 > 1$ and set $c = \tau_1 L_1(\boldsymbol{y}_{t-1})$ and compute
        $\boldsymbol{x}_t = \text{prox}_c^f \left(\boldsymbol{x}_{t-1} - \frac{1}{c}\nabla_{\boldsymbol{x}} h(\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1})\right)$
    take $\tau_1 > 1$ and set $d = \tau_2 L_2(\boldsymbol{x}_t)$ and compute
        $\boldsymbol{y}_t = \text{prox}_d^g \left(\boldsymbol{y}_{t-1} - \frac{1}{d}\nabla_{\boldsymbol{y}} h(\boldsymbol{x}_t, \boldsymbol{y}_{t-1})\right)$
```

prox(·) denotes the proximal operator (Appendix A.1).

Solving Problem 2 by PALM. Problem 2 is identical to identifying the minimizer of $f(\boldsymbol{W}) + g(\boldsymbol{\Gamma}) + h(\boldsymbol{W}, \boldsymbol{\Gamma})$ where

$$
f(\boldsymbol{W}) = \begin{cases} 0 & \boldsymbol{W} \in \mathbb{R}^{S\times K} : 0 \leq w_{su} \leq 1, \sum_{s=1}^{S} w_{su} = 1 \\ \infty & \text{otherwise} \end{cases},
$$

$$
g(\boldsymbol{\Gamma}) = \begin{cases} 0 & \boldsymbol{\Gamma} \in \mathbb{R}^{S\times S} : \epsilon \leq \gamma_{ss} \leq 1 - \epsilon; \gamma_{ss'} = 0 \ \forall s \neq s' \\ \infty & \text{otherwise} \end{cases},
$$

and h is defined as $h(\boldsymbol{W}, \boldsymbol{\Gamma}) = \|\boldsymbol{\Lambda} - \boldsymbol{W}'\boldsymbol{\Gamma}\boldsymbol{W}\|_F^2$. Define $\mathcal{C}_{\boldsymbol{W}} = \{\boldsymbol{W} \in \mathbb{R}^{S\times K} : w_{su} \geq 0, \sum_{s=1}^{S} w_{su} = 1\}$. Define $\mathcal{C}_{\boldsymbol{\Gamma}} = \{\boldsymbol{\Gamma} \in \mathbb{R}^{S\times S} : \epsilon \leq \gamma_{ss} \leq 1 - \epsilon; \gamma_{ss'} = 0 \ \forall s \neq s'\}$. We note that both functions $\boldsymbol{W} \to \nabla_{\boldsymbol{W}} h(\boldsymbol{W}, \boldsymbol{\Gamma})$ and $\boldsymbol{\Gamma} \to \nabla_{\boldsymbol{\Gamma}} h(\boldsymbol{W}, \boldsymbol{\Gamma})$ are continuous. Indeed,

$$\nabla_{\boldsymbol{W}} h(\boldsymbol{W}, \boldsymbol{\Gamma}) = -4\boldsymbol{\Gamma}\boldsymbol{W}(\boldsymbol{\Lambda} - \boldsymbol{W}'\boldsymbol{\Gamma}\boldsymbol{W}), \tag{B.1}$$

$$\nabla_{\boldsymbol{\Gamma}} h(\boldsymbol{W}, \boldsymbol{\Gamma}) = -2\boldsymbol{W}(\boldsymbol{\Lambda} - \boldsymbol{W}'\boldsymbol{\Gamma}\boldsymbol{W})\boldsymbol{W}'. \tag{B.2}$$

For all $\boldsymbol{W}_1, \boldsymbol{W}_2 \in \mathcal{C}_{\boldsymbol{W}}$,

$$\|\nabla_{\boldsymbol{W}} h(\boldsymbol{W}_1, \boldsymbol{\Gamma}) - \nabla_{\boldsymbol{W}} h(\boldsymbol{W}_2, \boldsymbol{\Gamma})\|_F \leq 4(\|\boldsymbol{\Lambda}\|_2\|\boldsymbol{\Gamma}\|_2 + 3K\|\boldsymbol{\Gamma}\|_2^2)\|\boldsymbol{W}_1 - \boldsymbol{W}_2\|_F. \tag{B.3}$$

For all $\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2 \in \mathcal{C}_{\boldsymbol{\Gamma}}$,

$$\|\nabla_{\boldsymbol{\Gamma}} h(\boldsymbol{W}, \boldsymbol{\Gamma}_1) - \nabla_{\boldsymbol{\Gamma}} h(\boldsymbol{W}, \boldsymbol{\Gamma}_2)\|_F \leq 2\|\boldsymbol{W}\|_2^4\|\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2\|_F. \tag{B.4}$$

Eqs. (B.3) and (B.4) are proved in the following paragraphs. By Appendix A.1, $\text{prox}_c^f(\boldsymbol{W}) = \mathcal{P}_{\text{dom}(f)}(\boldsymbol{W})$ and $\text{prox}_c^f(\boldsymbol{\Gamma}) = \mathcal{P}_{\text{dom}(g)}(\boldsymbol{\Gamma})$. This implies

$$\{\text{prox}_d^g(\boldsymbol{\Gamma})\}_{ss'} = \begin{cases} \min(\epsilon, \max(\gamma_{ss}, 1-\epsilon)) & s = s' \\ 0 & \text{otherwise} \end{cases} \tag{B.5}$$

and

$$\text{prox}_c^f(\boldsymbol{W}) = \begin{pmatrix} \mathcal{P}_\Delta(\boldsymbol{w}_1) & \ldots & \mathcal{P}_\Delta(\boldsymbol{w}_K) \end{pmatrix}, \tag{B.6}$$

where $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K$ are columns of $\boldsymbol{W}$. Applying Eqs. (B.1) to (B.6) to Algorithm B.3, we arrive at Algorithm 3 for solving Problem 2.

Proving the convergence of Algorithm 3. To prove the convergence, we need to show all assumptions of PALM hold. It is obvious that the assumptions (i), (iii) and (iv) hold.

Proof of assumption (ii): Ψ is a KL function. By Lemma 4, we note that the objective function H is a real polynomial function, hence semi-algebraic. For the indicator function f, we observe that the domain of f is defined by $\text{dom}(f) = \cap_{u=1}^K\{\boldsymbol{W} \in \mathbb{R}^{S\times K} : \boldsymbol{w}_u'\boldsymbol{1} = 1 \text{ and } \boldsymbol{w}_u \geq \boldsymbol{0}\}$. Hence, $\text{dom}(f)$ is a semi-algebraic set, so f is a semi-algebraic function. For the indicator function g, we observe that the domain of g is defined by $\text{dom}(g) = \cap_{s=1}^S\{\boldsymbol{\Gamma} \in \mathbb{R}^{S\times S} : \epsilon \leq \boldsymbol{e}_s'\boldsymbol{\Gamma}\boldsymbol{e}_s \leq 1-\epsilon\}$, where $\boldsymbol{e}_1 = (1, 0, \ldots, 0)$, …, $\boldsymbol{e}_S = (0, 0, \ldots, 1)$. Hence, $\text{dom}(g)$ is a semi-algebraic set, so g is a semi-algebraic function. Thus, $\Psi = f+g+H$ is a semi-algebraic function, and Ψ satisfies the KL property of any point of its domain.

Proof of assumption (v): To prove the assumption (v), we are to show Eqs. (B.3) and (B.4). We note that for all $\boldsymbol{W} \in \mathcal{C}_{\boldsymbol{W}}$, by the definition of the induced matrix norm and Lemma 3, we have $\|\boldsymbol{W}\|_2 \leq \sqrt{K}\|\boldsymbol{W}\|_1 = \sqrt{K}$. For all $\boldsymbol{W}_1, \boldsymbol{W}_2 \in \mathcal{C}_{\boldsymbol{W}}$:

$$
\begin{aligned}
&\|\nabla_{\boldsymbol{W}} H(\boldsymbol{W}_1, \boldsymbol{\Gamma}) - \nabla_{\boldsymbol{W}} H(\boldsymbol{W}_2, \boldsymbol{\Gamma})\|_F \\
=&\| - 4\boldsymbol{\Gamma}\boldsymbol{W}_1(\boldsymbol{\Lambda} - \boldsymbol{W}_1'\boldsymbol{\Gamma}\boldsymbol{W}_1) + 4\boldsymbol{\Gamma}\boldsymbol{W}_2(\boldsymbol{\Lambda} - \boldsymbol{W}_2'\boldsymbol{\Gamma}\boldsymbol{W}_2)\|_F \\
\leq& 4\|\boldsymbol{\Gamma}(\boldsymbol{W}_1 - \boldsymbol{W}_2)\boldsymbol{\Lambda}\|_F + 4\|\boldsymbol{\Gamma}\boldsymbol{W}_1\boldsymbol{W}_1'\boldsymbol{\Gamma}\boldsymbol{W}_1 - \boldsymbol{\Gamma}\boldsymbol{W}_2\boldsymbol{W}_2'\boldsymbol{\Gamma}\boldsymbol{W}_2\|_F \\
\leq& 4\|\boldsymbol{\Gamma}\|_2\|\boldsymbol{\Lambda}\|_2\|\boldsymbol{W}_1 - \boldsymbol{W}_2\|_F + 4\underbrace{\|\boldsymbol{\Gamma}\boldsymbol{W}_1\boldsymbol{W}_1'\boldsymbol{\Gamma}\boldsymbol{W}_1 - \boldsymbol{\Gamma}\boldsymbol{W}_2\boldsymbol{W}_2'\boldsymbol{\Gamma}\boldsymbol{W}_2\|_F}_{*} \quad \text{(by Lemma 2)}
\end{aligned}
$$

$$
\begin{aligned}
(*) &= \|\boldsymbol{\Gamma}\boldsymbol{W}_1\boldsymbol{W}_1'\boldsymbol{\Gamma}\boldsymbol{W}_1 - \boldsymbol{\Gamma}\boldsymbol{W}_2\boldsymbol{W}_2'\boldsymbol{\Gamma}\boldsymbol{W}_2\|_F \\
&= \|\boldsymbol{\Gamma}\boldsymbol{W}_1\boldsymbol{W}_1'\boldsymbol{\Gamma}\boldsymbol{W}_1 - \boldsymbol{\Gamma}\boldsymbol{W}_2\boldsymbol{W}_1'\boldsymbol{\Gamma}\boldsymbol{W}_1 \\
&\quad + \boldsymbol{\Gamma}\boldsymbol{W}_2\boldsymbol{W}_1'\boldsymbol{\Gamma}\boldsymbol{W}_1 - \boldsymbol{\Gamma}\boldsymbol{W}_2\boldsymbol{W}_2'\boldsymbol{\Gamma}\boldsymbol{W}_1 + \boldsymbol{\Gamma}\boldsymbol{W}_2\boldsymbol{W}_2'\boldsymbol{\Gamma}\boldsymbol{W}_1 - \boldsymbol{\Gamma}\boldsymbol{W}_2\boldsymbol{W}_2'\boldsymbol{\Gamma}\boldsymbol{W}_2\|_F \\
&\leq \|\boldsymbol{\Gamma}(\boldsymbol{W}_1 - \boldsymbol{W}_2)\boldsymbol{W}_1'\boldsymbol{\Gamma}\boldsymbol{W}_1\|_F + \|\boldsymbol{\Gamma}\boldsymbol{W}_2(\boldsymbol{W}_1 - \boldsymbol{W}_2)'\boldsymbol{\Gamma}\boldsymbol{W}_1\|_F + \|\boldsymbol{\Gamma}\boldsymbol{W}_2\boldsymbol{W}_2'\boldsymbol{\Gamma}(\boldsymbol{W}_1 - \boldsymbol{W}_2)\|_F \\
&\leq \|\boldsymbol{\Gamma}\|_2^2(\|\boldsymbol{W}_1\|_2^2 + \|\boldsymbol{W}_1\|_2\|\boldsymbol{W}_2\|_2 + \|\boldsymbol{W}_2\|_2^2)\|\boldsymbol{W}_1 - \boldsymbol{W}_2\|_F \quad \text{(by Lemma 2)} \\
&\leq 3K\|\boldsymbol{\Gamma}\|_2^2\|\boldsymbol{W}_1 - \boldsymbol{W}_2\|_F
\end{aligned}
$$

Therefore,

$$
\|\nabla_{\boldsymbol{W}} H(\boldsymbol{W}_1, \boldsymbol{\Gamma}) - \nabla_{\boldsymbol{W}} H(\boldsymbol{W}_2, \boldsymbol{\Gamma})\|_F \leq 4(\|\boldsymbol{\Lambda}\|_2\|\boldsymbol{\Gamma}\|_2 + 3K\|\boldsymbol{\Gamma}\|_2^2)\|\boldsymbol{W}_1 - \boldsymbol{W}_2\|_F.
$$

For all $\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2 \in \mathbb{R}^{K\times K}$,

$$
\begin{aligned}
&\|\nabla_{\boldsymbol{\Gamma}} H(\boldsymbol{W}, \boldsymbol{\Gamma}_1) - \nabla_{\boldsymbol{\Gamma}} H(\boldsymbol{W}, \boldsymbol{\Gamma}_2)\|_F \\
&= \| - 2\boldsymbol{W}(\boldsymbol{\Lambda} - \boldsymbol{W}'\boldsymbol{\Gamma}_1\boldsymbol{W})\boldsymbol{W}' + 2\boldsymbol{W}(\boldsymbol{\Lambda} - \boldsymbol{W}'\boldsymbol{\Gamma}_2\boldsymbol{W})\boldsymbol{W}'\|_F \\
&= 2\|\boldsymbol{W}\boldsymbol{W}'(\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)\boldsymbol{W}\boldsymbol{W}'\|_F \\
&\leq 2\|\boldsymbol{W}\boldsymbol{W}'\|_2^2\|\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2\|_F \quad \text{(by Lemma 2)} \\
&\leq 2\|\boldsymbol{W}\|_2^4\|\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2\|_F. \quad \text{(by Lemma 1)}
\end{aligned}
$$

Proof of assumption (vi): Since $\boldsymbol{W}_t \in \mathcal{C}_{\boldsymbol{W}}$ and $\boldsymbol{\Gamma}_t \in \mathcal{C}_{\boldsymbol{\Gamma}}$ for all $t \in \mathbb{N}$, and $\mathcal{C}_{\boldsymbol{W}}$ and $\mathcal{C}_{\boldsymbol{\Gamma}}$ are compact sets, $L_1(\boldsymbol{\Gamma}) = 4(\|\boldsymbol{\Lambda}\|_2\|\boldsymbol{\Gamma}\|_2 + 3K\|\boldsymbol{\Gamma}\|_2^2)$ and $L_2(\boldsymbol{W}) = 2\|\boldsymbol{W}\|_2^4$ are bounded. By Lemma 3, $\|\boldsymbol{W}\|_2 \geq \frac{1}{\sqrt{S}}\|\boldsymbol{W}\|_1 = \frac{1}{\sqrt{S}}$ and $\|\boldsymbol{\Gamma}\|_2 \geq \frac{1}{\sqrt{S}}\|\boldsymbol{\Gamma}\|_1 \geq \frac{\epsilon}{\sqrt{S}}$. Therefore, $2\|\boldsymbol{W}_t\|_2^4 \geq \frac{2}{S^2}$ and $4(\|\boldsymbol{\Lambda}\|_2\|\boldsymbol{\Gamma}_t\|_2 + 3K\|\boldsymbol{\Gamma}_t\|_2^2) \geq 4(\epsilon\|\boldsymbol{\Lambda}\|_2/\sqrt{S} + 3K\epsilon^2/S)$ for all $t \in \mathbb{N}$.

B.4 Simulating antecedent population coancestry through NORmal To Anything (NORTA)

NORmal To Anything (NORTA) algorithm. The NORmal To Anything (NORTA) algorithm is a transformation-based method for generating random vectors with given marginal distributions and a given covariance matrix. The goal of the NORTA method is to define a K-dimensional random vector $\boldsymbol{X}$ with the following properties:

(i) $X_u \sim F_u$, $u = 1, \dots, K$, where $\{F_u\}$ are marginal cumulative distribution functions,

(ii) $\mathbb{C}(\boldsymbol{X}) = \boldsymbol{\Sigma}_{\boldsymbol{X}} = \{\sigma_{uv,\boldsymbol{X}}\}$.

The NORTA algorithm represents $\boldsymbol{X}$ as a transformation of a K-dimensional, standard multivariate normal vector $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_K)'$ with covariance matrix $\mathbb{C}(\boldsymbol{Z}) = \boldsymbol{\Sigma}_{\boldsymbol{Z}} = \{\sigma_{uv,\boldsymbol{Z}}\}$.

Algorithm B.4: NORTA algorithm

```
input : Marginal cumulative distribution functions $F_u$ for $u = 1, 2, \dots, K$, and the
        covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$
let $\boldsymbol{\Sigma}_{\boldsymbol{Z}} = \{\sigma_{uv,\boldsymbol{Z}}\}$ be an $K \times K$ identity matrix $\boldsymbol{I}$
for $u = 1, \dots, K-1$ do
    for $v = u+1, \dots, K$ do
        find $\sigma^*$ such that $\mathbb{E}[X_u X_v] = \mathbb{E}[F_u^{-1}(\Phi(Z_u))F_v^{-1}(\Phi(Z_v))]$
        let $\sigma_{uv,\boldsymbol{Z}} = \sigma_{vu,\boldsymbol{Z}} = \sigma^*$
simulate $\boldsymbol{Z}$ from the standard multivariate norm distribution with the covariance
  matrix $\boldsymbol{\Sigma}_{\boldsymbol{Z}}$
let $\boldsymbol{X}$ be a transformation of $\boldsymbol{Z}$ where $X_u = F_u^{-1}(\Phi(Z_u))$ for $u = 1, \dots, K$.
return $\boldsymbol{X}$
```

Φ is the univariate Normal$(0, 1)$ cumulative distribution function (cdf); $F_u^{-1}(t) = \inf\{x : F_u(x) \geq t\}$ denotes the inverse cdf.

Note that $\mathbb{E}[X_u], \mathbb{E}[X_v]$, $\mathbb{V}(X_u)$ and $\mathbb{V}(X_v)$ are determined by F_u and F_v, implying $\mathbb{E}[X_u X_v]$ is determined by F_u and F_v. Let φ_σ denote the standard bivariate normal probability density distribution with the correlation (and also covariance in this case) σ. Then

$$\mathbb{E}[F_u^{-1}(\Phi(Z_u))F_v^{-1}(\Phi(Z_v))] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{X_u}^{-1}(\Phi(z_u))F_{X_v}^{-1}(\Phi(z_v))\varphi_\sigma(z_u, z_v)dz_u dz_v. \quad \text{(B.7)}$$

Determining σ to yield the desired covariance is equivalent to solving the root of the function

$$\begin{aligned} g(\sigma) &= \mathbb{E}[F_u^{-1}(\Phi(Z_u))F_v^{-1}(\Phi(Z_v))] - \mathbb{E}[X_uX_v] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_{X_u}^{-1}(\Phi(z_u))F_{X_v}^{-1}(\Phi(z_v))\varphi_\sigma(z_u,z_v)dz_udz_v - \mathbb{E}[X_uX_v]. \end{aligned}$$

We will denote this root by σ^* where $g(\sigma^*) = 0$.

Applying NORTA to generate coancestry among antecedent populations. We applied the NORTA algorithm with $\boldsymbol{X} = \boldsymbol{p}_i$, $F_u = \mathrm{BN}(a_i, \lambda_{uu})$, and $\sigma_{uv,\boldsymbol{X}} \equiv a_i(1-a_i)\lambda_{uv}$ for $i \in 1, \ldots, m$. In this scenario,

$$\mathbb{E}[X_uX_v] = a_i^2 + a_i(1-a_i)\lambda_{uv}$$

We note that there is no closed form solution for σ^*. We adopted the Newton method to perform a numerical search for σ^*. Let

$$g(\sigma; a_i, \lambda_{uv}) = \mathbb{E}[F_u^{-1}(\Phi(Z_u))F_v^{-1}(\Phi(Z_v))] - a_i^2 - a_i(1-a_i)\lambda_{uv}.$$

It follows that

$$g'(\sigma; a_i, \lambda_{uv}, F_u, F_v) = \mathbb{E}\left[F_u^{-1}(\Phi(Z_u))F_v^{-1}(\Phi(Z_v))\left(\frac{\sigma}{1-\sigma^2} + \frac{z_uz_v - \sigma(z_u^2 - \sigma z_uz_v + z_v^2)}{(1-\sigma^2)^2}\right)\right].$$

The Newton iteration for finding σ^* is then given by

$$\sigma \leftarrow \sigma - \frac{g(\sigma; a_i, \lambda_{uv}, F_u, F_v)}{g'(\sigma; a_i, \lambda_{uv}, F_u, F_v)}.$$

We calculated $g(\sigma; a_i, \lambda_{uv}, F_u, F_v)$ and $g'(\sigma; a_i, \lambda_{uv}, F_u, F_v)$ via numeric integration, leading to Algorithm B.5 for simulating antecedent population allele frequencies with the desired coancestry.

C Supplementary numerical results

C.1 Generating $\boldsymbol{\Lambda}$

To simulate $\boldsymbol{\Lambda}$ under the super admixture model, we simulated a $K \times K$ matrix $\boldsymbol{A}$ with elements drawn independently from $\mathrm{Uniform}(0, 0.3)$. We then set $\boldsymbol{\Lambda} = \boldsymbol{A}'\boldsymbol{A}$. To simulate $\boldsymbol{\Lambda}$

Algorithm B.5: NORTA algorithm for simulating $\boldsymbol{P}$

```
input: Ancestral allele frequencies $\boldsymbol{a}$ and coancestry among antecedent populations $\boldsymbol{\Lambda}$
for $i = 1, \ldots, m$ do
    let $\boldsymbol{\Sigma_Z}$ be an $K \times K$ identity matrix $\boldsymbol{I}$
    for $u = 1, \ldots, K-1$ do
        for $v = u+1, \ldots, K$ do
            let $\sigma = \lambda_{uv}/\sqrt{\lambda_{uu}\lambda_{vv}}$
            while not converged do
                let $\sigma \leftarrow \sigma - \frac{g(\sigma; a_i, \lambda_{uv}, F_u, F_v)}{g'(\sigma; a_i, \lambda_{uv}, F_u, F_v)}$
            let $\sigma_{uv,\boldsymbol{Z}} = \sigma_{vu,\boldsymbol{Z}} = \sigma$
    generate $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma_Z})$
    let $\boldsymbol{p}_i$ be a $K$-dimensional vector where $p_{iu} = F_u^{-1}(\Phi(z_u))$ for $u = 1, \ldots, K$
return $\boldsymbol{P}$
```

Φ is the univariate Normal$(0, 1)$ cumulative distribution function (cdf); $F_u = \text{BN}(a_i, \lambda_{uu})$; $F_u^{-1}(t) = \inf\{x : F_u(x) \geq t\}$ denotes the inverse cdf of F_u.

under the standard admixture model, we let $\boldsymbol{\Lambda}$ be a diagonal matrix whose diagonal elements are generated independently from Uniform$(0, 1)$. For both scenarios, we varied $K = 3, 6, 9$ and sampled 100 instances of $\boldsymbol{\Lambda}$ for each K. Our simulation resulted in 300 instances of $\boldsymbol{\Lambda}$ under the standard admixture model and 300 instances of $\boldsymbol{\Lambda}$ under the super admixture model.

C.2 Generating $\boldsymbol{Q}$

We adopted the spatial model from ref. [4] to generate admixture proportions reflecting real data. This model represents the admixture process as a diffusion on a one-dimensional geography. It assumes K independent populations equally spaced at positions $x_0, x_0 + 1, \ldots, x_0 + K - 1$ on an infinite line. If all populations begin to diffuse at time $t = 0$ at the same diffusion rate, then population u will be distributed as a Gaussian with mean $\mu_u = x_0 + u - 1$ and standard deviation σ. Therefore, under the spatial model an individual j sampled at the position j will have the admixture proportions shown as follows:

$$\boldsymbol{q}_j = (q_{1j}, q_{2j}, \ldots, q_{Kj})' = \left(\frac{\mathcal{N}(j; \mu_1, \sigma^2)}{\sum_{u=1}^{K} \mathcal{N}(j; \mu_u, \sigma^2)}, \quad \frac{\mathcal{N}(j; \mu_2, \sigma^2)}{\sum_{u=1}^{K} \mathcal{N}(j; \mu_u, \sigma^2)}, \quad \ldots, \quad \frac{\mathcal{N}(j; \mu_K, \sigma^2)}{\sum_{u=1}^{K} \mathcal{N}(j; \mu_u, \sigma^2)} \right)$$

where $\mathcal{N}(; \mu, \sigma^2)$ denotes a Normal(μ, σ^2) distribution. We chose $\sigma^2 = 0.5$ and $n = 2000$ in our simulations.

C.3 Evaluating algorithms for estimating coancestry among antecedent populations

To evaluate Algorithms 1 and B.2, we generated 300 unique combinations of $(\boldsymbol{\Lambda}, \boldsymbol{Q})$ under the standard admixture model and 300 unique combinations of $(\boldsymbol{\Lambda}, \boldsymbol{Q})$ under the super admixture model. For each pair of $(\boldsymbol{\Lambda}, \boldsymbol{Q})$, we calculated the corresponding $\boldsymbol{\Theta} = \boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}$. If the super admixture model is assumed, we applied Algorithm 1 with a random initial matrix $\boldsymbol{\Lambda}_0$ to estimate $\boldsymbol{\Lambda}$. If the standard admixture model is assumed, we applied Algorithm B.2 with a random initial matrix $\boldsymbol{\Lambda}_0$ to estimate $\boldsymbol{\Lambda}$. We recorded values of $\boldsymbol{\Lambda}_t$ per iteration. We quantified the differences between $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}_t$ by $\frac{\|\boldsymbol{\Lambda}_t - \boldsymbol{\Lambda}\|_F}{\|\boldsymbol{\Lambda}\|_F}$, and visualized the change over iterations in Fig. C.1. We validated that both algorithms are capable of generating a sequence of $\boldsymbol{\Lambda}_t$ such that the difference between $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}_t$ decreases as $t \to \infty$.

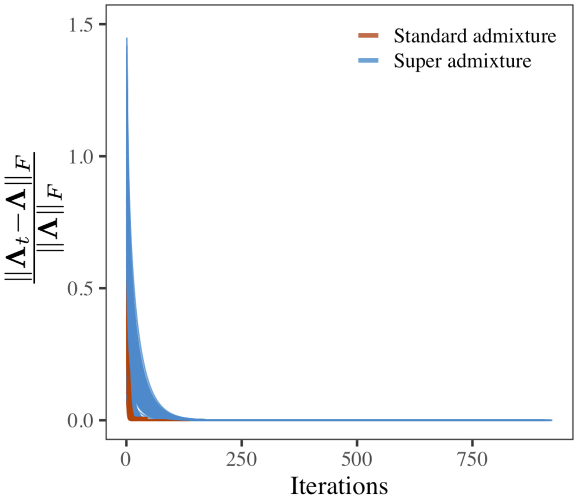


Figure C.1: The convergence of Algorithms 1 and B.2.

C.4 Evaluating algorithms for antecedent population allele frequencies

To evaluate Algorithms 4 and B.5,, we assessed whether they could generate allele frequencies that satisfy the moments of the super admixture model:

$$\mathbb{E}[p_{iu}|T] = a_i$$
$$\mathbb{V}[p_{iu}|T] = a_i(1 - a_i)\lambda_{uu}$$
$$\mathbb{C}[p_{iu}, p_{iv}|T] = a_i(1 - a_i)\lambda_{uv}$$

To achieve this, we generated 300 unique combinations of $(a, \boldsymbol{\Lambda})$ where a is a scalar and is simulated from Uniform$(0, 1)$; $\boldsymbol{\Lambda}$ assumes the super admixture model and is simulated as previously described. For each pair of $(a, \boldsymbol{\Lambda})$, we generated $B = 100,000$ replications of the n-vector allele frequenecies $\boldsymbol{p}^{(b)}$ from the double-admixture method (Algorithm 4) or from the NORTA method (Algorithm B.5). Then we calculated the empirical mean and the empirical covariance matrix as $\hat{\boldsymbol{a}} = \frac{1}{B}\sum_{b=1}^{B} \boldsymbol{p}^{(b)}$ and $\hat{\boldsymbol{C}} = \frac{1}{B}\sum_{b=1}^{B}(\boldsymbol{p}^{(b)} - \hat{\boldsymbol{a}})(\boldsymbol{p}^{(b)} - \hat{\boldsymbol{a}})\prime$, respectively. We measured the differences between empirical moments and the desired moments by $\|\hat{\boldsymbol{a}} - \boldsymbol{a}\|_2/\|\boldsymbol{a}\|_2$ and $\|\hat{\boldsymbol{C}} - \boldsymbol{C}\|_F/\|\boldsymbol{C}\|_F$, where $\boldsymbol{a}$ denotes a $K \times 1$ dimensional vector whose entries are all equal to a and $\boldsymbol{C} = a(1 - a)\boldsymbol{\Lambda}$. We found that $\|\hat{\boldsymbol{a}} - \boldsymbol{a}\|_2/\|\boldsymbol{a}\|_2$ is generally less than 0.02 and $\|\hat{\boldsymbol{C}} - \boldsymbol{C}\|_F/\|\boldsymbol{C}\|_F$ is generally less than 0.04 for both algorithms (Fig. C.2). These findings confirmed the performance of both algorithms.

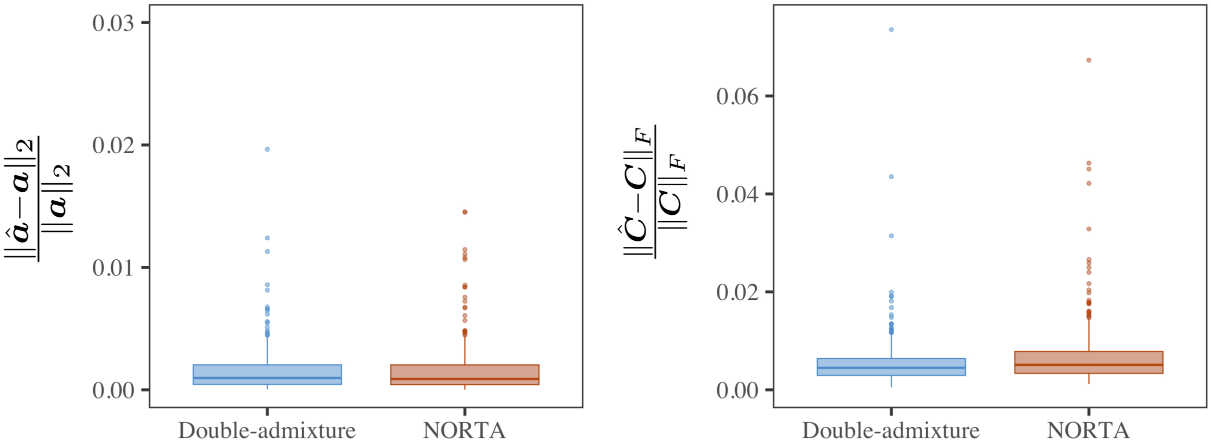


Figure C.2: Box-plots of $\frac{\|\hat{\boldsymbol{a}}-\boldsymbol{a}\|_2}{\|\boldsymbol{a}\|_2}$ and $\frac{\|\hat{\boldsymbol{C}}-\boldsymbol{C}\|_F}{\|\boldsymbol{C}\|_F}$ across 300 simulations.

C.5 Evaluating the algorithm for generating genotypes from the super admixture model

To evaluate Algorithm 5, we assessed whether this algorithm is capable of generating genotypes that satisfy the moment constraints imposed by the super admixture model. More specifically, we examined if the estimated individual-level coancestry agrees with $\boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}$ and if the estimated coancestry among antecedent populations agrees with $\boldsymbol{\Lambda}$.

To check these, we generated 300 unique combinations of $(\boldsymbol{\Lambda}, \boldsymbol{Q})$ under the super admixture model as previously described. For each pair of $(\boldsymbol{\Lambda}, \boldsymbol{Q})$, we (i) simulated ancestral allele frequencies $\boldsymbol{a}$ $(m = 500,000)$ by generating each a_i independently from Uniform$(0, 1)$, (ii) generated genotypes $\boldsymbol{X}$ using Algorithm 5, (iii) estimated the individual-level coancestry

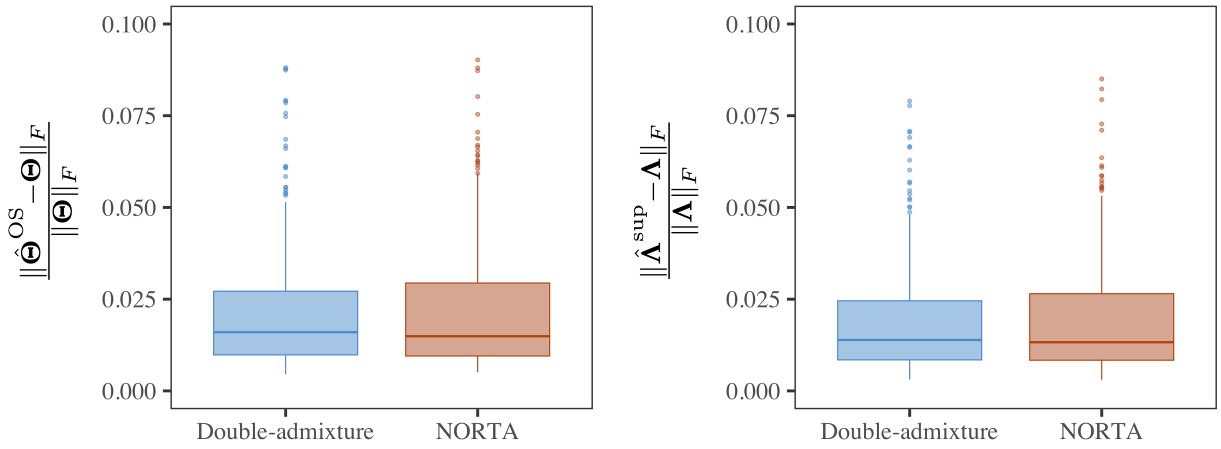


Figure C.3: Box-plots of $\frac{\|\hat{\boldsymbol{\Theta}}^{\text{OS}}-\boldsymbol{\Theta}\|_F}{\|\boldsymbol{\Theta}\|_F}$ and $\frac{\|\hat{\boldsymbol{\Lambda}}^{\text{sup}}-\boldsymbol{\Lambda}\|_F}{\|\boldsymbol{\Lambda}\|_F}$ across 300 simulations.

$\hat{\boldsymbol{\Theta}}^{\text{OS}}$ and (iv) applied Algorithm 1 to estimate $\hat{\boldsymbol{\Lambda}}^{\text{sup}}$ with $\hat{\boldsymbol{\Theta}}^{\text{OS}}$ and $\boldsymbol{Q}$ as inputs. In (ii), a matrix of antecedent population allele frequencies $\boldsymbol{P}$ is generated at an intermediate step. We used both the double-admixture method and the NORTA method for generating $\boldsymbol{P}$ to compare their performances. For (iii), the OS estimate utilizes the minimum pairwise coancestry equal to 0, which might not hold here. We used the strategy described in Appendix B.1 to adapt the OS estimate to our simulation. We assessed the agreement between $\hat{\boldsymbol{\Theta}}^{\text{OS}}$ and $\boldsymbol{\Theta}$ and between $\hat{\boldsymbol{\Lambda}}^{\text{sup}}$ and $\boldsymbol{\Lambda}$ by $\|\hat{\boldsymbol{\Theta}}^{\text{OS}} - \boldsymbol{\Theta}\|_F/\|\boldsymbol{\Theta}\|_F$ and $\|\hat{\boldsymbol{\Lambda}}^{\text{sup}} - \boldsymbol{\Lambda}\|_F/\|\boldsymbol{\Lambda}\|_F$, respectively. In Fig. C.3, we observed the majority of $\|\hat{\boldsymbol{\Theta}}^{\text{OS}} - \boldsymbol{\Theta}\|_F/\|\boldsymbol{\Theta}\|_F$ and $\|\hat{\boldsymbol{\Lambda}}^{\text{sup}} - \boldsymbol{\Lambda}\|_F/\|\boldsymbol{\Lambda}\|_F$ are less than 0.05 regardless of the method used for simulating $\boldsymbol{P}$. These observations demonstrated our simulated genotypes satisfied the desired moment constraints.

C.6 Null p-value distribution of the hypothesis test of standard admixture versus super admixture

Recall the hypothesis test of the standard admixture model (null) versus the super admixture model (alternative):

$$H_0 : \max\left(\{\lambda_{uv}\}_{u \neq v}\right) = 0 \text{ (standard admixture model)}$$
$$H_1 : \max\left(\{\lambda_{uv}\}_{u \neq v}\right) > 0 \text{ (super admixture model)}$$

To check whether the true null hypothesis p-values calculated by Algorithm 7 are stochastically greater than or equal to the Uniform$(0, 1)$ distribution, we generated 300 unique combinations of $(\boldsymbol{\Lambda}, \boldsymbol{Q})$ under the standard admixture model. For each pair of $(\boldsymbol{\Lambda}, \boldsymbol{Q})$, we (i) simulated ancestral allele frequencies $\boldsymbol{a}$ ($m = 500,000$) by generating each a_i independently from Uniform$(0, 1)$, (ii) generated genotypes $\boldsymbol{X}$ from the standard admixture model, (iii) applied Algorithm 7 to compute the p-values. We compared the empirical distribution of the p-values against Uniform$(0, 1)$. Fig. C.4 shows that our proposed method is conservative, meaning it has a maximum type I error probability less than or equal to the nominal level of the test.

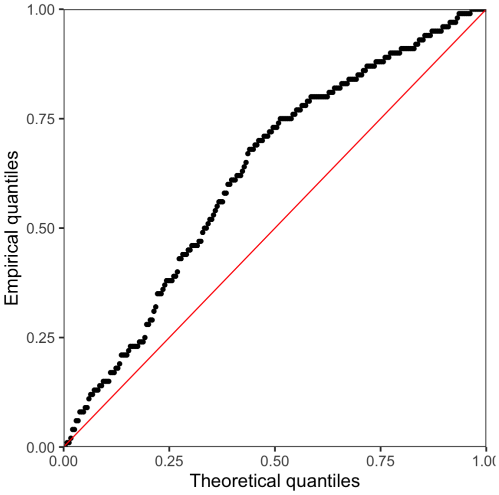


Figure C.4: Quantile-quantile plots of true null p-values vs. Uniform$(0, 1)$ distribution.

D Supplementary analyses of human studies

D.1 Data processing

HO data. We downloaded the publicly available main Human Origins dataset and the Pacific dataset. The downloaded datasets are in the Eigensoft package format. We converted them to the PLINK format with the Eigensoft `convertf` function. We then merged the main Human Origins dataset with the Pacific dataset. These datasets have non-overlapping individuals that were genotyped using the same microarray platform. We excluded individuals from singleton subpopulations and the ancient individuals from the Lapita-Vanuatu population. We excluded SNPs with minor allele frequency (MAF) less than 0.01. The final dataset has $486,981$ SNPs and 2124 individuals.

HGDP data. We downloaded the publicly available HGDP dataset. We preserved loci that (i) are autosomal, biallelic SNPs, (ii) have MAF ≥ 0.01 and (iii) are in approximate linkage equilibrium with each other (`PLINK --indep-pairwise 1000kb 0.3`). The final dataset has $997,431$ SNPs and 929 individuals.

TGP data. We downloaded the publicly available TGP dataset. We preserved loci that (i) are autosomal, biallelic SNPs, (ii) are variants in the Yoruba individuals, (iii) have MAF ≥ 0.05 and (iv) are in approximate linkage equilibrium with each other (`PLINK --indep-pairwise 1000kb 0.3`). The final dataset has $712,998$ SNPs and 2583 individuals.

AMR subset of TGP. We identified individuals in the TGP dataset marked as `AMR` to create the AMR subset of TGP. We preserved loci that (i) are autosomal, biallelic SNPs, (ii) are variants in the Yoruba individuals, (iii) have MAF ≥ 0.01 and (iv) are in approximate linkage equilibrium with each other (`PLINK --indep-pairwise 1000kb 0.3`). The final dataset has $555,145$ SNPs and 353 individuals.

IND data. We obtained the Indian dataset from the authors of ref. [22]. We merged this dataset with the Central/South Asia population and the East Asia population of HGDP. These datasets have non-overlapping individuals that were genotyped using the same microarray platform. We excluded SNPs with MAF < 0.01 and SNPs with MAF differences greater than 0.2 to resolve the allele flipping issue. We also excluded SNPs with missing rates in IND or in the HGDP subset greater than 0.005. We applied this filter to keep high quality variants. The final dataset has 221, 499 SNPs and 698 individuals.

D.2 HGDP study analysis

We observed good concordance between the individual-level coancestry of HGDP for the OS and super admixture estimates (Fig. D.5) and early human migrations [29–32]. We note that the earliest major split occurred between Africa and MiddleEast from an out-of-Africa migration around 50 to 60 kya, resulting in the divergence between Sub-Saharan Africans and the remaining human populations. Another major split occurred between Central / South Asia and East Asia, revealing the separation between West Eurasians and East Asians around 40 to 45 kya. Among the East Asia clade, the Oceanians have the highest within subpopulation coancestry and lowest between subpopulation coancestry, consistent with the theory that Oceanians split earliest from the remaining East Asians.

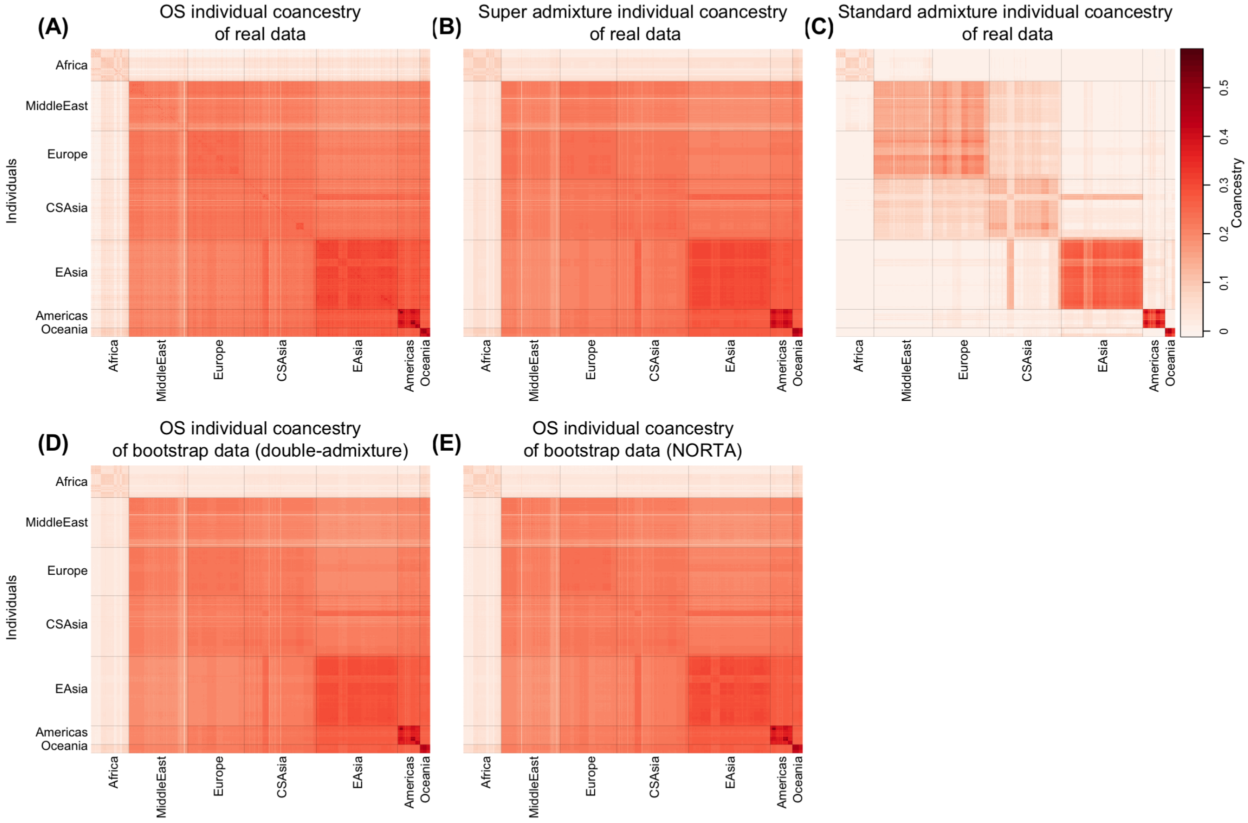


Figure D.5: Heatmap of individual-level coancestry estimates in HGDP.

We confirmed that the super admixture antecedent population coancestry estimates are also compatible with known early human dispersals (Fig. D.6). Specifically, in Fig. D.6B the deepest split occurred roughly between individuals from Sub-Saharan Africa represented by the antecedent populations S_1 and S_2 and individuals outside of Sub-Saharan Africa represented by the remaining antecedent populations. Individuals outside of Sub-Saharan Africa further branched into two lineages: the West Eurasians represented by antecedent populations S_3 and S_4, and the East Asians represented by antecedent populations S_5, S_6, and S_7. The Oceanians represented by S_7 split from the majority of ancestral East Asians, while the remaining East Asians further diverged into present-day Asians (S_5) and present-day Americans (S_6).

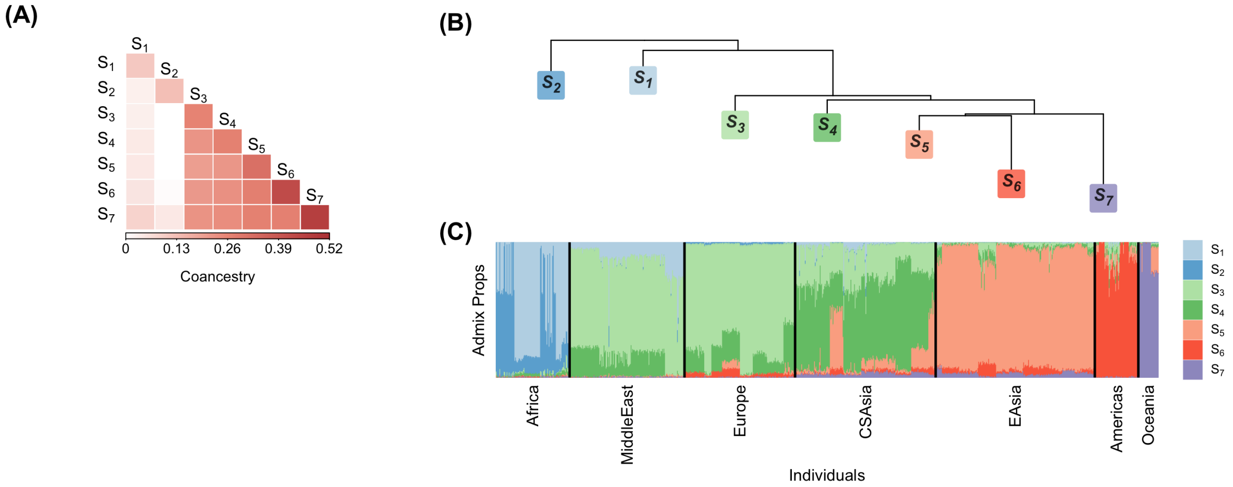


Figure D.6: Heatmap of antecedent population coancestry estimates in HGDP. (B) Dendrogram representation of the antecedent population coancestry estimates. (C) Stacked bar plot of admixture proportions.

D.3 TGP study analysis

We also observed good correspondence between the individual-level coancestry OS and super admixture estimates of the TGP data (Fig. D.7) and known early human migrations [29–32]. Similarly to our analysis of HO and HGDP, the earliest major split is between AFR and the other populations, which reflects the divergence between Sub-Saharan Africans and the remaining of human populations. Another split occurs between EUR and EAS, revealing the separation between West Eurasians and East Asians.

As in our analysis of HO and HGDP, there is agreement between the estimated antecedent population coancestry and existing results. In Fig. D.8B, the deepest split occurred roughly between individuals from Sub-Saharan Africa represented by antecedent population S_1 and individuals outside of Sub-Saharan Africa represented by the rest of the antecedent populations. We also noted the divergence between the Europeans represented by antecedent population S_3, and the Asians represented by antecedent populations S_2, S_4 and S_5. The Americans sampled in TGP appear to have a higher European ancestry compared to that in the HO and HGDP datasets.

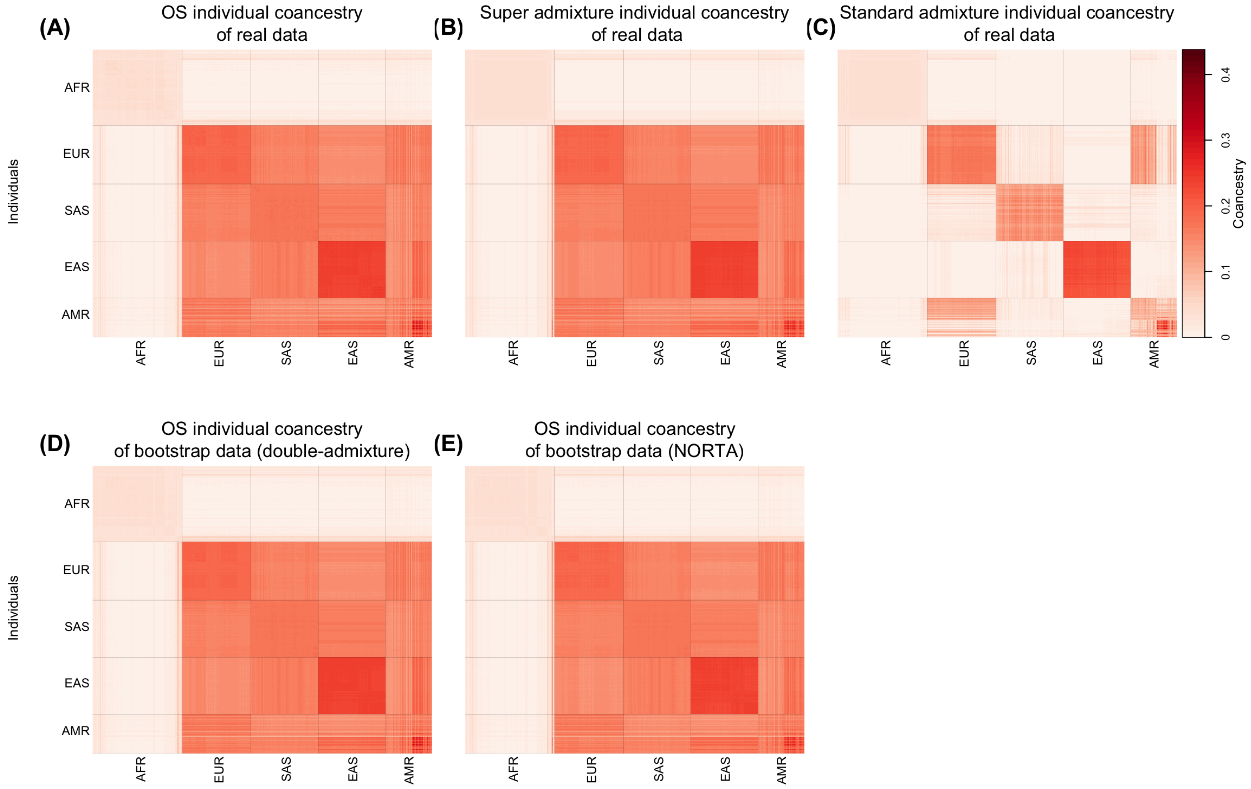


Figure D.7: Heatmap of individual-level coancestry estimates in TGP.

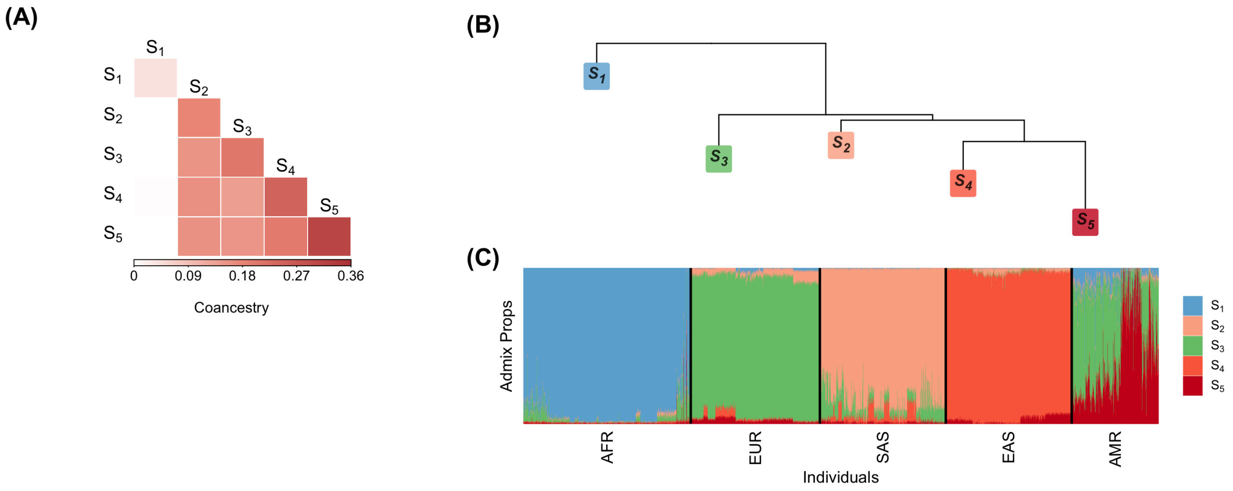


Figure D.8: Heatmap of antecedent population coancestry estimates in TGP. (B) Dendrogram representation of the antecedent population coancestry estimates. (C) Stacked bar plot of admixture proportions.

D.4 Confirming significant hypothesis tests of standard admixture versus super admixture in the human studies

We applied Algorithm 7 to each of the five human datasets to statistically evaluate the presence of coancestry among antecedent populations. Each panel of Fig. D.9 shows the distribution of the $B = 1000$ bootstrap null test-statistics for each dataset. The observed test-statistic U_{obs} for each data set is noted on the top-right of each panel. In each data set U_{obs} exceeded all bootstrap null test-statistics, implying p-value < 0.001 for each.

D.5 Comparing the individual-level coancestry estimates

In Table D.1, we compared $\hat{\boldsymbol{\Theta}}^{\text{sup}}$ and $\hat{\boldsymbol{\Theta}}^{\text{std}}$ to the general OS estimate of individual-level coancestry $\hat{\boldsymbol{\Theta}}^{\text{OS}}$ from ref. [4] on the five data sets. The super admixture coancestry estimate has about 10 to 40 times smaller distance to the OS estimate compared to the standard admixture estimate.

D.6 Selecting the number of antecedent populations

We utilized the structural Hardy-Weinberg (sHWE) framework [12] for determining the number of antecedent populations K, as outlined in that work. The approach considers a range of K values for a model of structure that results in estimated IAFs, which is the case for our framework. For each K, a hypothesis test is performed for each SNP of the assumption that $x_{ij}|\pi_{ij} \sim \text{Binomial}(2, \pi_{ij})$ $(j = 1, 2, \ldots, n)$ based on the estimates $\hat{\pi}_{ij}$ and a goodness-of-fit statistic with a parametric bootstrap null distribution. This results in m p-values per value of K.

As proposed in the sHWE framework, for each value of K, we (i) calculated the m sHWE p-values, (ii) binned the sHWE p-values into equal-sized bins (number of bins, $C = 150$), (iii) removed the first bin $[0, 1/C)$, and (iv) calculated the following negative entropy that measures how well the sHWE p-values follow the Uniform$(0, 1)$ distribution,

$$\sum_{c=2}^{C} f_c \log_{10} f_c,$$

where f_c denotes the proportion of p-values in the c-th bin.

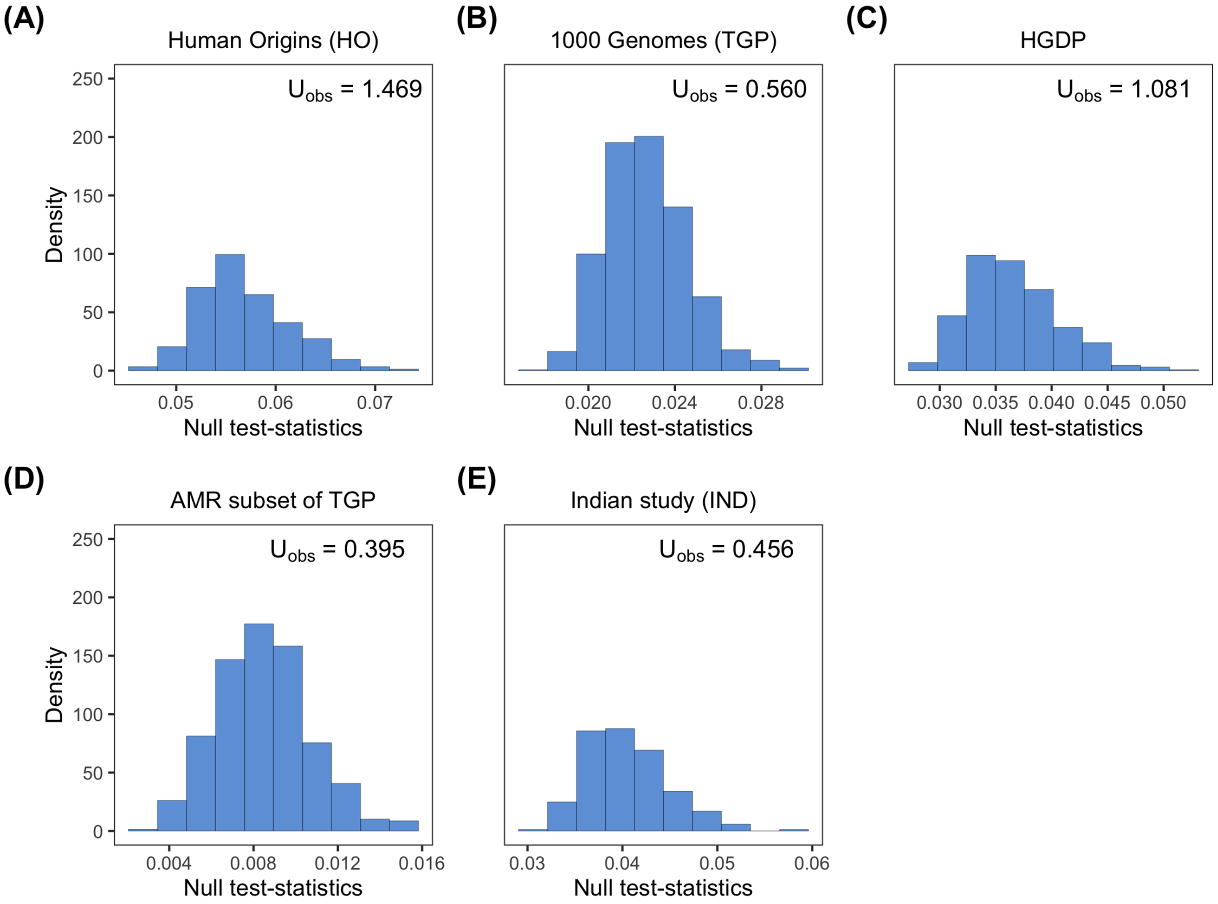


Figure D.9: Distributions of test-statistics for the hypothesis tests of standard admixture versus super admixture across all five human studies.

Table D.1: The distance between $\hat{\mathbf{\Theta}}^{\text{OS}}$ and $\hat{\mathbf{\Theta}}^{\text{sup}}$ and between $\hat{\mathbf{\Theta}}^{\text{OS}}$ and $\hat{\mathbf{\Theta}}^{\text{std}}$ for each of the five data sets.

	$\frac{1}{n}\|\hat{\mathbf{\Theta}}^{\text{sup}} - \hat{\mathbf{\Theta}}^{\text{OS}}\|_F$	$\frac{1}{n}\|\hat{\mathbf{\Theta}}^{\text{std}} - \hat{\mathbf{\Theta}}^{\text{OS}}\|_F$
HO	0.003	0.114
AMR	0.006	0.124
IND	0.006	0.052
HGDP	0.008	0.146
TGP	0.002	0.087

	$\|\hat{\mathbf{\Theta}}^{\text{sup}} - \hat{\mathbf{\Theta}}^{\text{OS}}\|_F/\|\hat{\mathbf{\Theta}}^{\text{OS}}\|_F$	$\|\hat{\mathbf{\Theta}}^{\text{std}} - \hat{\mathbf{\Theta}}^{\text{OS}}\|_F/\|\hat{\mathbf{\Theta}}^{\text{OS}}\|_F$
HO	0.021	0.751
AMR	0.023	0.507
IND	0.073	0.613
HGDP	0.040	0.724
TGP	0.017	0.700

We determined K to be the value achieving the minimum negative entropy. In the event of a plateau where the entropy is more or less the same over a range of K, we chose a small value of K, where the plateau began. This resulted in $K = 11$ for HO, $K = 7$ for HGDP, $K = 5$ for TGP, $K = 7$ for IND, and $K = 3$ for AMR.

In Fig. D.10, we observed a consistent decline in $\|\hat{\mathbf{\Theta}}^{\text{sup}} - \hat{\mathbf{\Theta}}^{\text{OS}}\|_F/\|\hat{\mathbf{\Theta}}^{\text{OS}}\|_F$ as K increased. In Fig. D.11, we noted an increase in $\|\hat{\mathbf{\Theta}}^{\text{std}} - \hat{\mathbf{\Theta}}^{\text{OS}}\|_F/\|\hat{\mathbf{\Theta}}^{\text{OS}}\|_F$ as K increased. This implies the standard admixture fit cannot be improved with a larger K. In the following two subsections, we show that the results for HO and IND are similar over a range of K close to the values that we selected.

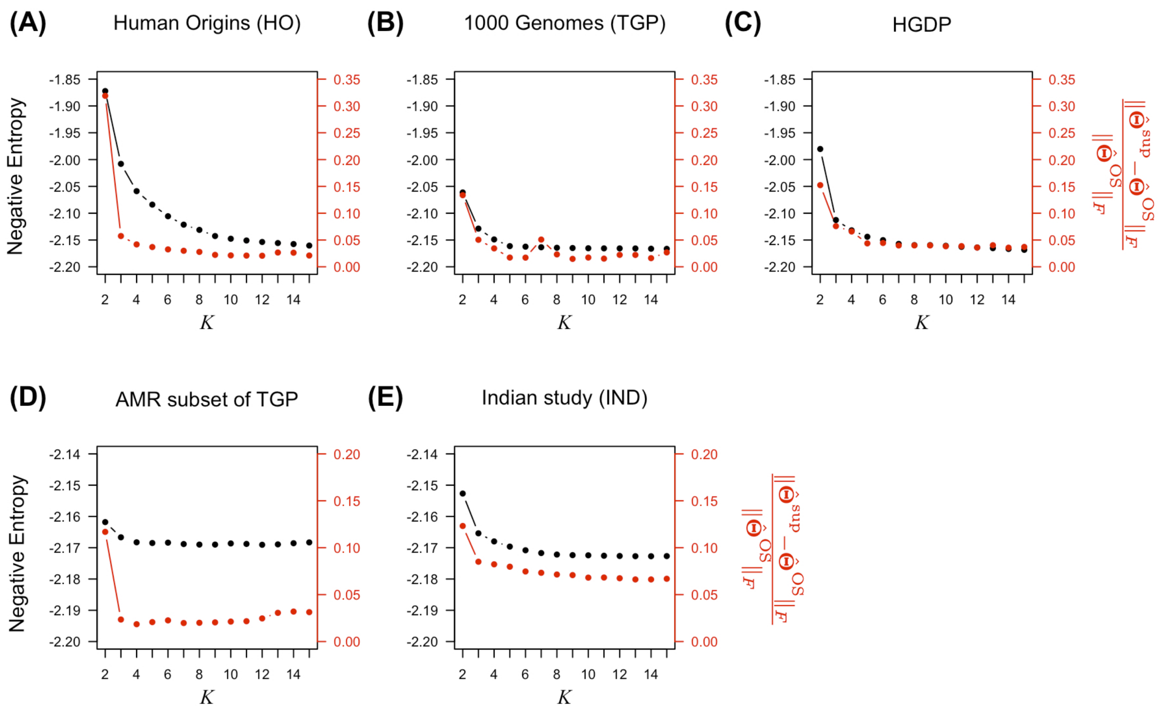


Figure D.10: The negative entropy and $\frac{\|\hat{\Theta}^{\text{sup}} - \hat{\Theta}^{\text{OS}}\|_F}{\|\hat{\Theta}^{\text{OS}}\|_F}$ across different numbers of antecedent populations.

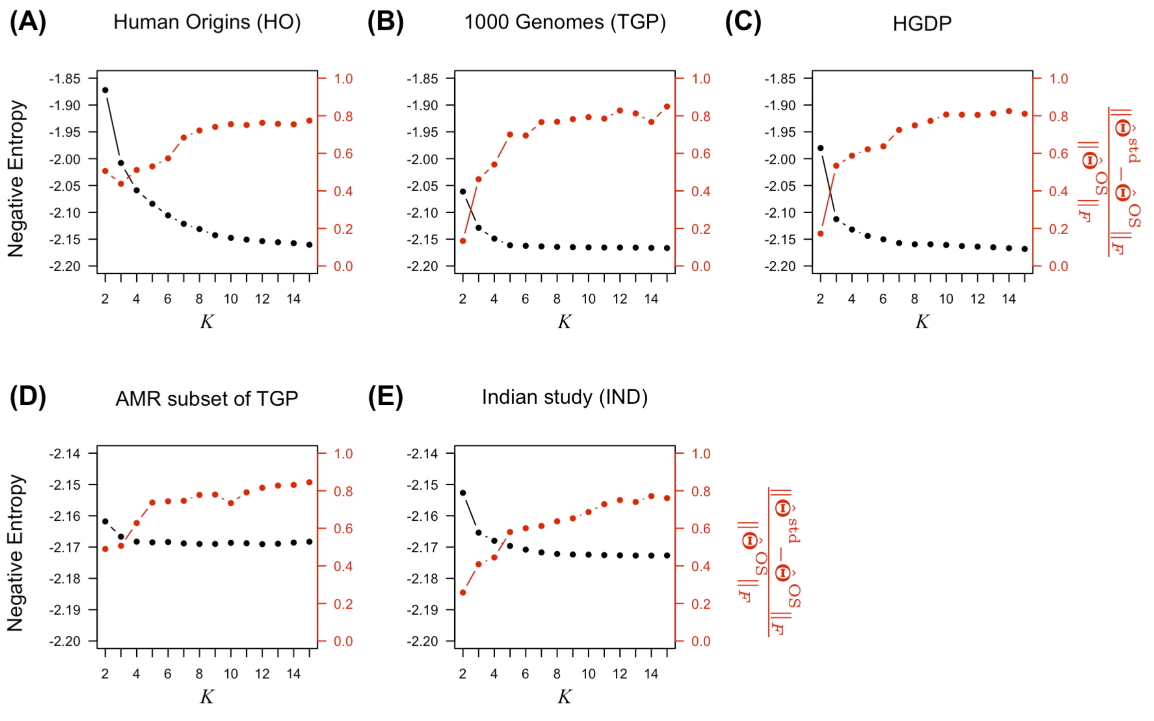


Figure D.11: The negative entropy and $\frac{\|\hat{\Theta}^{\text{std}} - \hat{\Theta}^{\text{OS}}\|_F}{\|\hat{\Theta}^{\text{OS}}\|_F}$ across different numbers of antecedent populations.

D.7 Analysis of HO over a range of antecedent population numbers

In our analysis of HO, we utilized $K = 11$ antecedent populations. We also analyzed the HO dataset for $K = 7, 8, 9, 10$ in this section to demonstrate the choice of K did not have a major impact on our results and conclusions. In Figs. D.12 to D.15, we observed the estimated antecedent population coancestries and the admixture proportions were consistent for $K = 7, 8, 9, 10$.

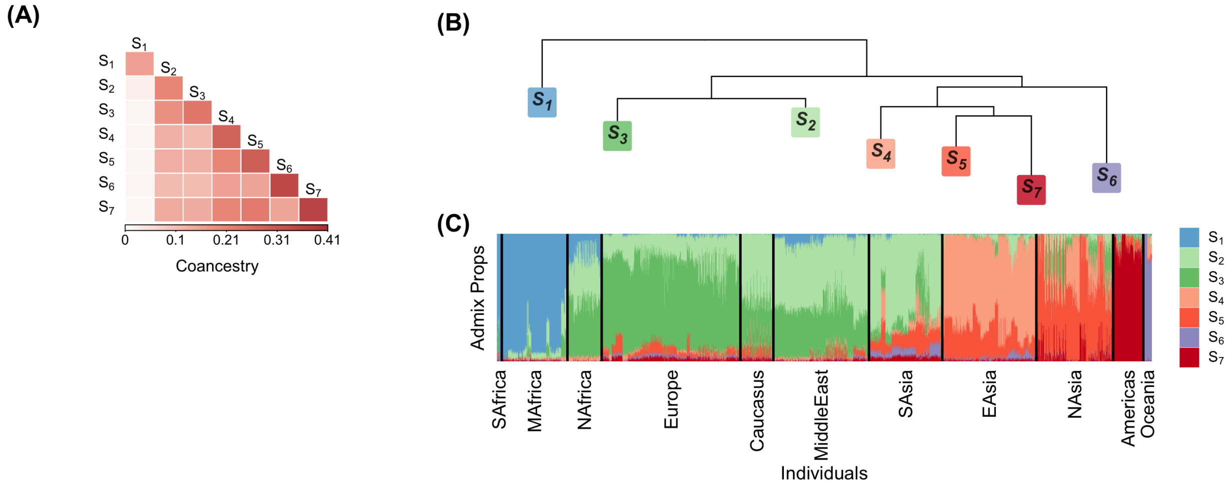


Figure D.12: (A) Heatmap of antecedent population coancestry estimates in HO with $K = 7$. (B) Dendrogram representation of the antecedent population coancestry estimates. (C) Stacked bar plot of admixture proportions.

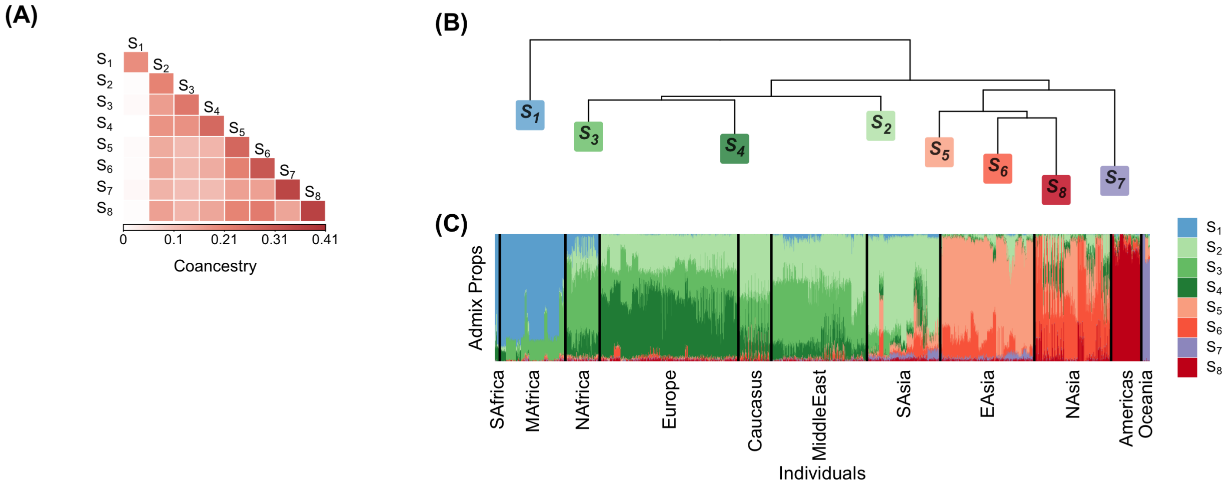


Figure D.13: (A) Heatmap of antecedent population coancestry estimates in HO with $K = 8$. (B) Dendrogram representation of the antecedent population coancestry estimates. (C) Stacked bar plot of admixture proportions.

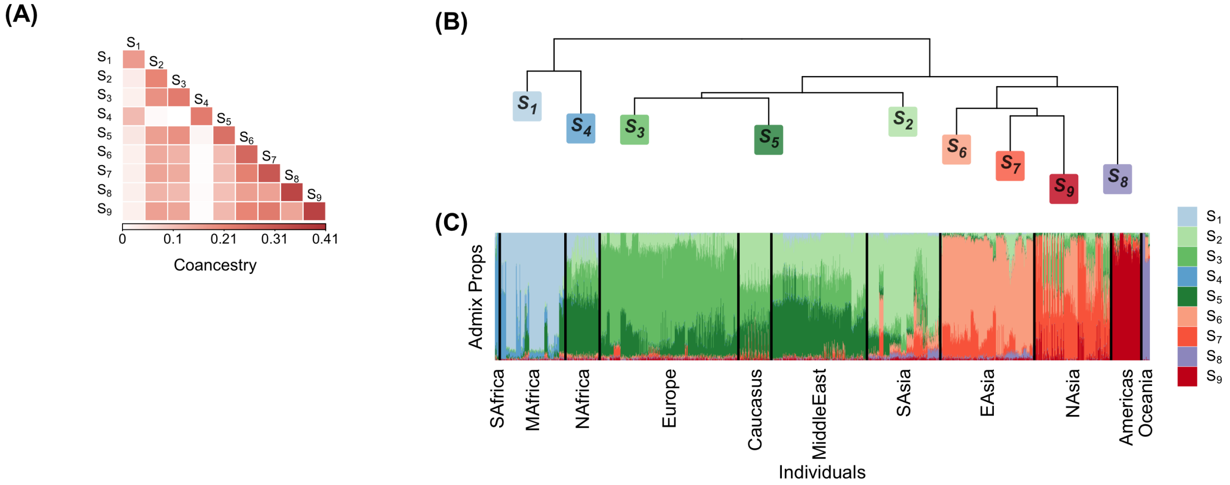


Figure D.14: (A) Heatmap of antecedent population coancestry estimates in HO with $K = 9$. (B) Dendrogram representation of the antecedent population coancestry estimates. (C) Stacked bar plot of admixture proportions.

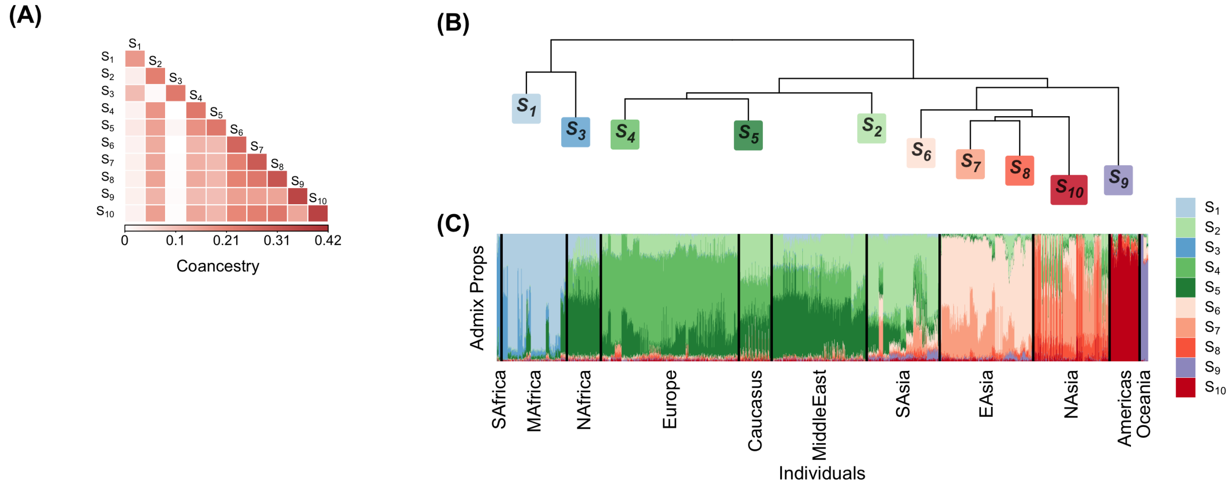


Figure D.15: (A) Heatmap of antecedent population coancestry estimates in HO with $K = 10$. (B) Dendrogram representation of the antecedent population coancestry estimates. (C) Stacked bar plot of admixture proportions.

D.8 Analysis of IND over a range of antecedent population numbers

In our analysis of IND, we utilized $K = 7$ antecedent populations. In Figs. D.16 to D.19, we observe the estimated antecedent population coancestries and the admixture proportions were consistent for $K = 3, 4, 5, 6$.

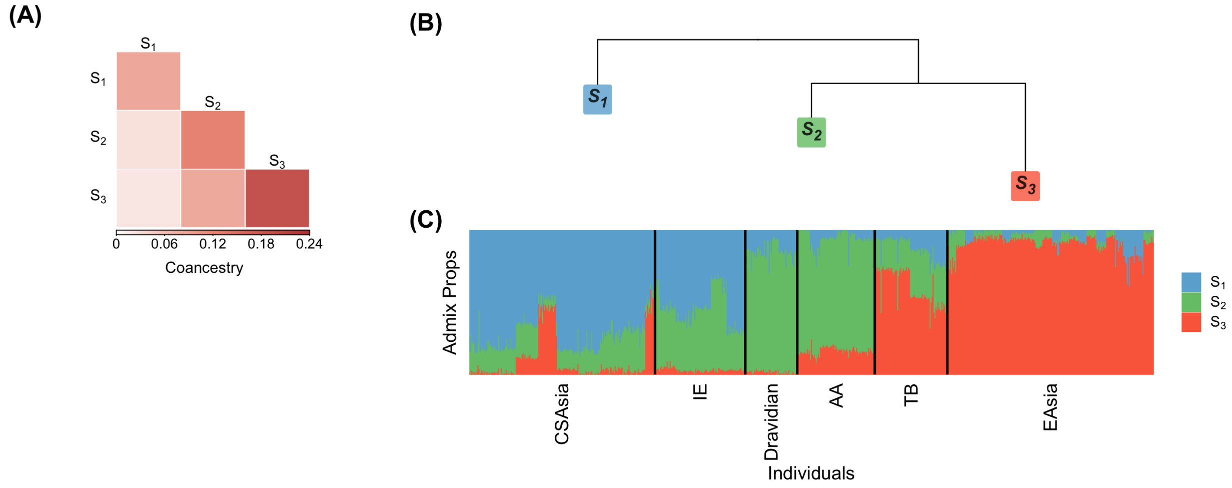


Figure D.16: (A) Heatmap of antecedent population coancestry estimates in the merged data sets of IND with Central/South Asians and East Asians of HGDP with $K = 3$. (B) Dendrogram representation of the antecedent population coancestry estimates. (C) Stacked bar plot of admixture proportions.

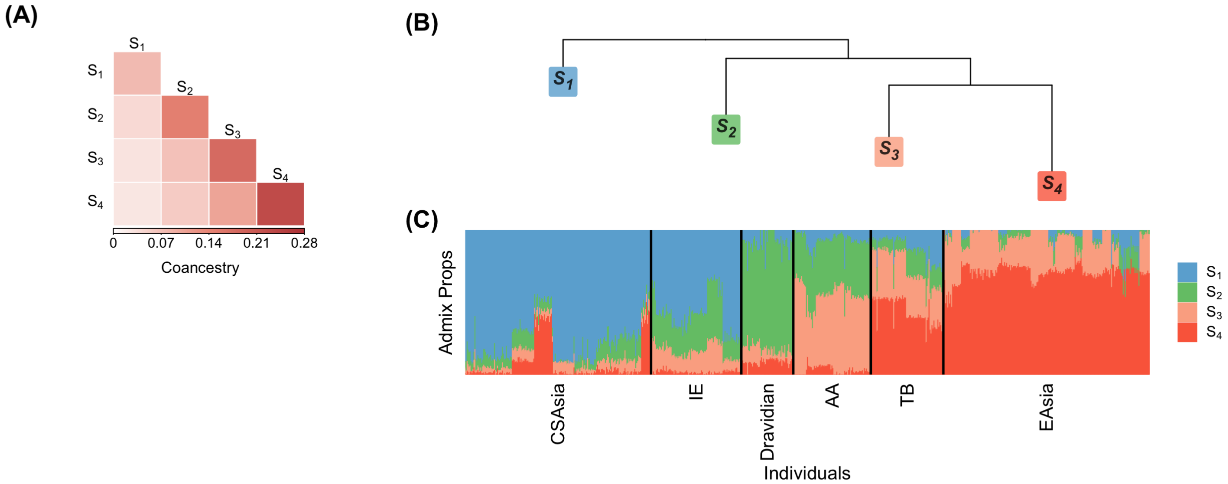


Figure D.17: (A) Heatmap of antecedent population coancestry estimates in the merged data sets of IND with Central/South Asians and East Asians of HGDP with $K = 4$. (B) Dendrogram representation of the antecedent population coancestry estimates. (C) Stacked bar plot of admixture proportions.

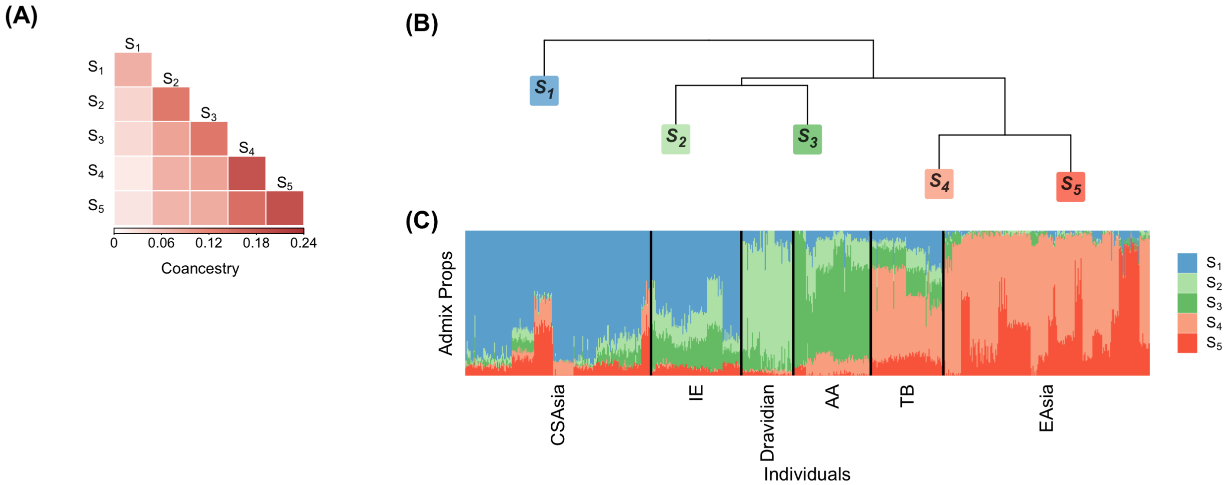


Figure D.18: (A) Heatmap of antecedent population coancestry estimates in the merged data sets of IND with Central/South Asians and East Asians of HGDP with $K = 5$. (B) Dendrogram representation of the antecedent population coancestry estimates. (C) Stacked bar plot of admixture proportions.

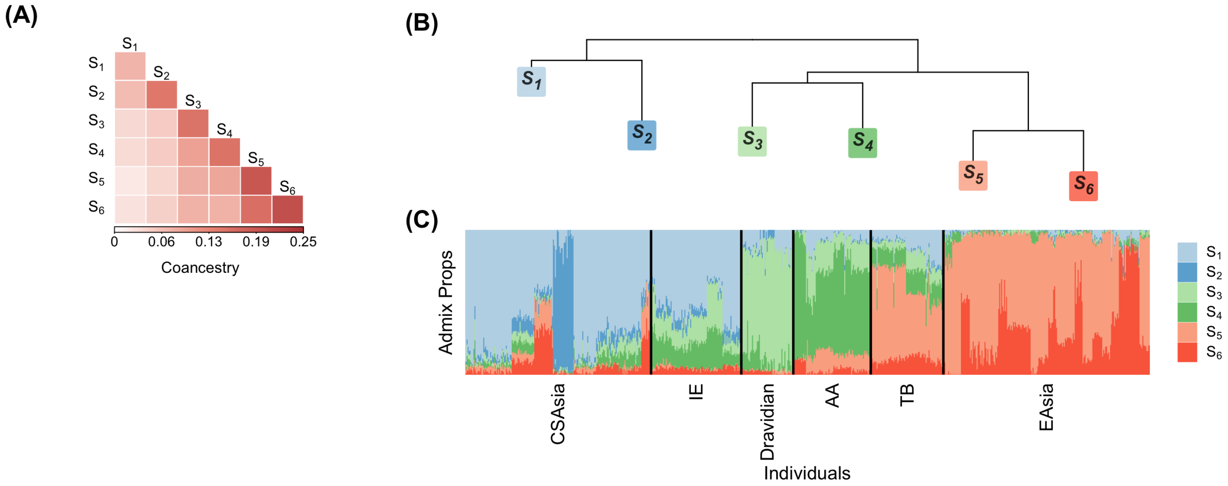


Figure D.19: (A) Heatmap of antecedent population coancestry estimates in the merged data sets of IND with Central/South Asians and East Asians of HGDP with $K = 6$. (B) Dendrogram representation of the antecedent population coancestry estimates. (C) Stacked bar plot of admixture proportions.

References

[1] M. Slatkin. “Gene flow and the geographic structure of natural populations”. *Science* 236(4803) (1987), pp. 787–792.

[2] A. J. Bohonak. “Dispersal, gene flow, and population structure”. *The Quarterly Review of Biology* 74(1) (1999), pp. 21–45.

[3] B. S. Weir and W. G. Hill. “Estimating F-statistics”. *Annual Review of Genetics* 36(1) (2002), pp. 721–750.

[4] A. Ochoa and J. D. Storey. “Estimating FST and kinship for arbitrary population structures”. *PLoS Genetics* 17(1) (2021), e1009241.

[5] A. Ochoa and J. D. Storey. “FST and kinship for arbitrary population structures I: Generalized definitions”. *bioRxiv* (2016), doi: 10.1101/083915.

[6] T. Thornton et al. “Estimating kinship in admixed populations”. *The American Journal of Human Genetics* 91(1) (2012). Publisher: Elsevier, pp. 122–138.

[7] W. Hao, M. Song, and J. D. Storey. “Probabilistic models of genetic variation in structured populations applied to global human studies”. *Bioinformatics* 32(5) (2016), pp. 713–721.

[8] H. Tang et al. “Estimation of individual admixture: Analytical and study design considerations”. *Genetic Epidemiology* 28(4) (2005), pp. 289–301.

[9] D. H. Alexander, J. Novembre, and K. Lange. “Fast model-based estimation of ancestry in unrelated individuals”. *Genome Research* 19(9) (2009), pp. 1655–1664.

[10] I. Cabreros and J. D. Storey. “A likelihood-free estimator of population structure bridging admixture models and principal components analysis”. *Genetics* 212(4) (2019), pp. 1009–1029.

[11] M. Song, W. Hao, and J. D. Storey. “Testing for genetic associations in arbitrarily structured populations”. *Nat Genet* 47(5) (2015), pp. 550–554.

[12] W. Hao and J. D. Storey. “Extending tests of Hardy–Weinberg equilibrium to structured populations”. *Genetics* 213(3) (2019), pp. 759–770.

[13] S. Wright. “The genetical structure of populations”. *Annals of Eugenics* 15(1) (1949), pp. 323–354.

[14] A. Jacquard. “Inbreeding: One word, several meanings”. *Theoretical Population Biology* 7(3) (1975), pp. 338–363.

[15] J. K. Pritchard, M. Stephens, and P. Donnelly. "Inference of population structure using multilocus genotype data". *Genetics* 155(2) (2000), pp. 945–959.

[16] A. Raj, M. Stephens, and J. K. Pritchard. "fastSTRUCTURE: Variational inference of population structure in large SNP data sets". *Genetics* 197(2) (2014), pp. 573–589.

[17] P. Gopalan et al. "Scaling probabilistic models of genetic variation to millions of humans". *Nature Genetics* 48(12) (2016), pp. 1587–1590.

[18] D. J. Balding and R. A. Nichols. "A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity". *Genetica* 96(1) (1995), pp. 3–12.

[19] L. L. Cavalli-Sforza. "The Human Genome Diversity Project: past, present and future". *Nature Reviews Genetics* 6(4) (2005), pp. 333–340.

[20] A. Auton and et al. "A global reference for human genetic variation". *Nature* 526(7571) (2015), pp. 68–74.

[21] I. Lazaridis et al. "Ancient human genomes suggest three ancestral populations for present-day Europeans". *Nature* 513(7518) (2014), pp. 409–413.

[22] A. Basu, N. Sarkar-Roy, and P. P. Majumder. "Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure". *Proceedings of the National Academy of Sciences* 113(6) (2016), pp. 1594–1599.

[23] A. L. Price et al. "Principal components analysis corrects for stratification in genome-wide association studies". *Nature Genetics* 38(8) (2006), pp. 904–909.

[24] P. L. Combettes and V. R. Wajs. "Signal recovery by proximal forward-backward splitting". *Multiscale Modeling & Simulation* 4(4) (2005), pp. 1168–1200.

[25] D. Falush, M. Stephens, and J. K. Pritchard. "Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies". *Genetics* 164(4) (2003), pp. 1567–1587.

[26] J. Bolte, S. Sabach, and M. Teboulle. "Proximal alternating linearized minimization for nonconvex and nonsmooth problems". *Mathematical Programming* 146(1) (2014), pp. 459–494.

[27] M. C. Cario and B. L. Nelson. "Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix". *Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL* (1997), pp. 1–19.

[28] A. Ochoa and J. D. Storey. "New kinship and FST estimates reveal higher levels of differentiation in the global human population". *bioRxiv* (2019), doi: 10.1101/653279.

[29] J. D. Wall. "Inferring human demographic histories of non-African populations from patterns of allele sharing". *The American Journal of Human Genetics* 100(5) (2017), pp. 766–772.

[30] M. Lipson and D. Reich. "A working model of the deep relationships of diverse modern human genetic lineages outside of Africa". *Molecular Biology and Evolution* 34(4) (2017), pp. 889–902.

[31] R. Nielsen et al. "Tracing the peopling of the world through genomics". *Nature* 541(7637) (2017), pp. 302–310.

[32] A. Bergström et al. "Origins of modern human ancestry". *Nature* 590(7845) (2021), pp. 229–237.

[33] K. Bryc et al. "Genome-wide patterns of population structure and admixture among Hispanic/Latino populations". *Proceedings of the National Academy of Sciences* 107 (supplement_2 2010), pp. 8954–8961.

[34] K. Adhikari et al. "Admixture in Latin America". *Current Opinion in Genetics & Development*. Genetics of human origin 41 (2016), pp. 106–114.

[35] P. de Barros Damgaard et al. "The first horse herders and the impact of early Bronze Age steppe expansions into Asia". *Science* 360(6396) (2018), eaar7711.

[36] V. M. Narasimhan et al. "The formation of human populations in South and Central Asia". *Science* 365(6457) (2019), eaat7487.

[37] W. Astle and D. J. Balding. "Population structure and cryptic relatedness in genetic association studies". *Statistical Science* 24(4) (2009), pp. 451–471.

[38] H. M. Kang et al. "Variance component model to account for sample structure in genome-wide association studies". *Nature Genetics* 42(4) (2010), pp. 348–354.

[39] J. Yang et al. "Common SNPs explain a large proportion of the heritability for human height". *Nature Genetics* 42(7) (2010), pp. 565–569.

[40] X. Zhou and M. Stephens. “Genome-wide efficient mixed-model analysis for association studies”. *Nature Genetics* 44(7) (2012), pp. 821–824.

[41] C. Márquez-Luna et al. “Multiethnic polygenic risk scores improve risk prediction in diverse populations”. *Genetic Epidemiology* 41(8) (2017), pp. 811–823.

[42] O. Weissbrod et al. “Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores”. *Nature Genetics* 54(4) (2022), pp. 450–458.