Notation

Throughout the Supplementary Material we will use the following notation:

M Number of diallelic sites

K Number of ancestral populations

R Relatedness coefficient with $R = (k_0, k_1, k_2)$ where k_i is the proportion of the genome where individuals $\mathcal{A}$ and $\mathcal{B}$ share i alleles IBD

$X^{\mathcal{A}}$ NGS data from individual $\mathcal{A}$, $X^{\mathcal{A}} = (X_1^{\mathcal{A}}, X_2^{\mathcal{A}}, ..., X_M^{\mathcal{A}})$

$X^{\mathcal{B}}$ NGS data from individual $\mathcal{B}$, $X^{\mathcal{B}} = (X_1^{\mathcal{B}}, X_2^{\mathcal{B}}, ..., X_M^{\mathcal{B}})$

F Matrix of ancestral allele frequencies for K populations for M sites $F = (F_1, F, ..., F_M)$ with $F_j = (F_{j1}, F_{j2}, ..., F_{jK})$

$\Phi^{\mathcal{A}}$ Paired ancestry proportions for individual $\mathcal{A}$ with paired ancestry from K different populations. $\Phi^{\mathcal{A}}$ contains K^2 entries, $\Phi^{\mathcal{A}} = (\phi_{11}^{\mathcal{A}}, \phi_{12}^{\mathcal{A}}, \phi_{21}^{\mathcal{A}}, ..., \phi_{KK}^{\mathcal{A}})$, where $\phi_{a_1,a_2}^{\mathcal{A}}$ denotes the proportion of the genome where the first allele is from population a_1 and the second allele is from population a_2.

$G^{\mathcal{A}}$ Latent variable of the ordered genotypes from individual $\mathcal{A}$, $G^{\mathcal{A}} = (G_1^{\mathcal{A}}, G_2^{\mathcal{A}}, ..., G_M^{\mathcal{A}})$ each with realization $g^{\mathcal{A}} \in \{0,1\}^2$

$G^{\mathcal{B}}$ Latent variable of the ordered genotypes from individual $\mathcal{B}$, $G^{\mathcal{B}} = (G_1^{\mathcal{B}}, G_2^{\mathcal{B}}, ..., G_M^{\mathcal{B}})$ each with realization $g^{\mathcal{B}} \in \{0,1\}^2$

Z Latent variable of the IBD state of allele 1 and 2, $Z = (Z_1, Z_2, ..., Z_M)$ each with realization $z = (z_1, z_2) \in \{0,1\}^2$

A Latent variable of the ancestral state of the two individuals' pair of alleles $A = (A_1, A_2, ..., A_M)$ each with realization $a = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}}, a_1^{\mathcal{B}}, a_2^{\mathcal{B}}) \in \{1, .., K\}^4$

$A^{\mathcal{A}}$ Latent variable of the ancestral states of individual $\mathcal{A}$'s pair of alleles $A^{\mathcal{A}} = (A_1^{\mathcal{A}}, A_2^{\mathcal{A}}, ..., A_M^{\mathcal{A}})$ each with realization $a^{\mathcal{A}} = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}}) \in \{1, .., K\}^2$

Derivation of the likelihood function

We want to estimate relatedness coefficients, $R = (k_0, k_1, k_2)$, for two individuals, $\mathcal{A}$ and $\mathcal{B}$, with ancestry from K different populations using a maximum likelihood approach. $R = (k_0, k_1, k_2)$ denote the fractions of the genome where the pair of individuals share 0, 1, and 2 alleles IBD ($\sum k_i = 1$) and in the following we derive the likelihood function that we use to estimate these.

For each individual we assume we have NGS data from M variable diallelic sites, $X^{\mathcal{A}} = (X_1^{\mathcal{A}}, X_2^{\mathcal{A}}, ..., X_M^{\mathcal{A}})$ and $X^{\mathcal{B}} = (X_1^{\mathcal{B}}, X_2^{\mathcal{B}}, ..., X_M^{\mathcal{B}})$. We let $\Phi^{\mathcal{A}} = (\phi_{11}^{\mathcal{A}}, \phi_{12}^{\mathcal{A}}, \phi_{21}^{\mathcal{A}}, ..., \phi_{KK}^{\mathcal{A}})$ and $\Phi^{\mathcal{B}} = (\phi_{11}^{\mathcal{B}}, \phi_{12}^{\mathcal{B}}, \phi_{21}^{\mathcal{B}}, ..., \phi_{KK}^{\mathcal{B}})$ denote the paired ancestry proportions of the two individuals and F denote the allele frequencies of the K ancestral populations for each of the M sites. We assume that both Φ and F are known and for simplicity we have not included them in the likelihood function. For each site, the genotypes of the two individuals are unobserved. Therefore, we include ordered genotypes for both individuals, $G^{\mathcal{A}}$ and $G^{\mathcal{B}}$, as latent variables in the model weighted by their probability. The likelihood function based on NGS data can thus be written

$$\begin{aligned}
P(X^{\mathcal{A}}, X^{\mathcal{B}} | R) &= \prod_{j=1}^{M} P(X_j^{\mathcal{A}}, X_j^{\mathcal{B}} | R) \\
&= \prod_{j=1}^{M} \sum_{g^{\mathcal{A}}, g^{\mathcal{B}} \in \{0,1\}^2} P(X_j^{\mathcal{A}}, X_j^{\mathcal{B}}, G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}} | R) \\
&= \prod_{j=1}^{M} \sum_{g^{\mathcal{A}}, g^{\mathcal{B}} \in \{0,1\}^2} P(X_j^{\mathcal{A}}, X_j^{\mathcal{B}} | G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}, R) P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}} | R) \\
&= \prod_{j=1}^{M} \sum_{g^{\mathcal{A}}, g^{\mathcal{B}} \in \{0,1\}^2} P(X_j^{\mathcal{A}}, X_j^{\mathcal{B}} | G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}) P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}} | R) \\
&= \prod_{j=1}^{M} \sum_{g^{\mathcal{A}}, g^{\mathcal{B}} \in \{0,1\}^2} P(X_j^{\mathcal{A}} | G_j^{\mathcal{A}} = g^{\mathcal{A}}) P(X_j^{\mathcal{B}} | G_j^{\mathcal{B}} = g^{\mathcal{B}}) P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}} | R).
\end{aligned}$$

Here we assumed that the sites are independent and that the probability of observing the NGS data is independent between the two individuals conditioned on the genotype being known. We end up with a product of the genotype likelihoods of the two individuals and the probability of the genotypes given the relatedness coefficients. The latter is the same likelihood that is used by relateAdmix Moltke and Albrechtsen (2013).

Building on this model, we can introduce the ancestry and the IBD state in the model. For each site, the four alleles' IBD status and ancestry are also not observed. Therefore, we also include these as latent variables in the model weighted by their probability. Specifically, we let $z = (z_1, z_2)$ denote IBD status, where z_1 indicates whether allele 1 of individuals $\mathcal{A}$ and $\mathcal{B}$ are IBD and z_2 indicates whether allele

2 in individuals $\mathcal{A}$ and $\mathcal{B}$ are IBD. Also, we let $a = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}}, a_1^{\mathcal{B}}, a_2^{\mathcal{B}})$ denote the unobserved ancestral populations of the two individuals' alleles. When including z and a the likelihood becomes

$$P(X^{\mathcal{A}}, X^{\mathcal{B}}|R) = \prod_{j=1}^{M} \sum_{z \in \{0,1\}^2} \sum_{a \in \{1,..,K\}^4} \sum_{g^{\mathcal{A}}, g^{\mathcal{B}} \in \{0,1\}^2} P(X_j^{\mathcal{A}}|G_j^{\mathcal{A}} = g^{\mathcal{A}}) P(X_j^{\mathcal{B}}|G_j^{\mathcal{B}} = g^{\mathcal{B}})$$
$$P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}, A_j = a, Z_j = z|R).$$

The last term can be be rewritten to

$$P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}, A_j = a, Z_j = z|R)$$
$$= P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}|A_j = a, Z_j = z, R) P(A_j = a, Z_j = z|R)$$
$$= P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}|A_j = a, Z_j = z) P(A_j = a, Z_j = z|R)$$
$$= P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}|A_j = a, Z_j = z) P(A_j = a|Z_j = z, R) P(Z_j = z|R)$$
$$= P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}|A_j = a, Z_j = z) P(A_j = a|Z_j = z) P(Z_j = z|R).$$

After rewriting the last term, the likelihood will finally become:

$$P(X^{\mathcal{A}}, X^{\mathcal{B}}|R) = \prod_{j=1}^{M} \sum_{z \in \{0,1\}^2} \sum_{a \in \{1,..,K\}^4} \sum_{g^{\mathcal{A}}, g^{\mathcal{B}} \in \{0,1\}^2} P(X_j^{\mathcal{A}}|G_j^{\mathcal{A}} = g^{\mathcal{A}}) P(X_j^{\mathcal{B}}|G_j^{\mathcal{B}} = g^{\mathcal{B}})$$
$$P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}|A_j = a, Z_j = z) P(A_j = a|Z_j = z) P(Z_j = z|R)$$

where $P(X_j^{\mathcal{A}}|G_j^{\mathcal{A}} = g^{\mathcal{A}})$ and $P(X_j^{\mathcal{B}}|G_j^{\mathcal{B}} = g^{\mathcal{B}})$ are GLs and they can be calculated as described in Section Calculation of genotype likelihoods. The factor $P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}|A_j = a, Z_j = z)$ is described in Table S1, $P(A_j = a|Z_j = z)$ is described in Table S3, and $P(Z_j = z|R)$ are given by:

$$P(Z_j|R) = \begin{cases} k_0 & if \ \ Z_j = (0,0) \\ \frac{k_1}{2} & if \ \ Z_j = (0,1) \\ \frac{k_1}{2} & if \ \ Z_j = (1,0) \\ k_2 & if \ \ Z_j = (1,1) \end{cases}$$

EM algorithm

The relatedness coefficients $R = (k_0, k_1, k_2)$ are estimated by finding the maximum likelihood (ML) using an EM algorithm. The expectation of the relatedness for each site j is:

$$P(Z_j = z|X_j^{\mathcal{A}}, X_j^{\mathcal{B}}, R) = \frac{P(X_j^{\mathcal{A}}, X_j^{\mathcal{B}}|Z_j = z) P(Z_j = z|R)}{P(X_j^{\mathcal{A}}, X_j^{\mathcal{B}}|R)},$$

where,

$$P(X_j^{\mathcal{A}}, X_j^{\mathcal{B}}|Z_j = z) = \sum_{g^{\mathcal{A}}, g^{\mathcal{B}}} P(X_j^{\mathcal{A}}, X_j^{\mathcal{B}}, G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}|Z_j = z)$$
$$= \sum_{g^{\mathcal{A}}, g^{\mathcal{B}}} P(X_j^{\mathcal{A}}, X_j^{\mathcal{B}}|G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}, Z_j = z) P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}|Z_j = z)$$
$$= \sum_{g^{\mathcal{A}}, g^{\mathcal{B}}} P(X_j^{\mathcal{A}}, X_j^{\mathcal{B}}|G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}) P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}|Z_j = z)$$
$$= \sum_{g^{\mathcal{A}}, g^{\mathcal{B}}} P(X_j^{\mathcal{A}}|G_j^{\mathcal{A}} = g^{\mathcal{A}}) P(X_j^{\mathcal{B}}|G_j^{\mathcal{B}} = g^{\mathcal{B}}) P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}|Z_j = z)$$
$$= \sum_{a} \sum_{g^{\mathcal{A}}, g^{\mathcal{B}}} P(X_j^{\mathcal{A}}|G_j^{\mathcal{A}} = g^{\mathcal{A}}) P(X_j^{\mathcal{B}}|G_j^{\mathcal{B}} = g^{\mathcal{B}}) P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}, A_j = a|Z_j = z)$$
$$= \sum_{a} \sum_{g^{\mathcal{A}}, g^{\mathcal{B}}} P(X_j^{\mathcal{A}}|G_j^{\mathcal{A}} = g^{\mathcal{A}}) P(X_j^{\mathcal{B}}|G_j^{\mathcal{B}} = g^{\mathcal{B}}) P(G_j^{\mathcal{A}} = g^{\mathcal{A}}, G_j^{\mathcal{B}} = g^{\mathcal{B}}|A_j = a, Z_j = z)$$
$$P(A_j = a|Z_j = z).$$

The derived likelihoods can be found in Table S1 and S3. The initial starting parameters, $R^{(0)} = (k_0^{(0)}, k_1^{(0)}, k_2^{(0)})$, of the EM algorithm are random. For each ith iteration of the EM algorithm the parameters are updated to $R^{(i+1)}$, using the current parameters ($R^{(i)}$):

$$k_0^{(i+1)} = \frac{\sum_{j=1}^{M} P(Z_j = (0,0)|X_j^{\mathcal{A}}, X_j^{\mathcal{B}}, R^{(i)})}{M},$$

$$k_1^{(i+1)} = \frac{\sum_{j=1}^{M} P(Z_j = (1,0)|X_j^{\mathcal{A}}, X_j^{\mathcal{B}}, R^{(i)}) + P(Z_j = (0,1)|X_j^{\mathcal{A}}, X_j^{\mathcal{B}}, R^{(i)})}{M},$$

$$k_2^{(i+1)} = \frac{\sum_{j=1}^{M} P(Z_j = (1,1)|X_j^{\mathcal{A}}, X_j^{\mathcal{B}}, R^{(i)})}{M}.$$

The EM algorithm will continue to update the relatedness coefficients until the algorithm converges. The algorithm converges when the root sum squared difference between $R^{(i)}$ and $R^{(i+1)}$ is less than $1e^{-6}$.

Estimation of paired ancestries

A recent admixture event, such as an offspring of parents from different population, will cause a dependency of the ancestry of the offspring's two alleles. At each site in the genome the offspring will have exactly one allele from each of the two populations. In order to accommodate this assumption we estimate paired ancestry for each individual. For individual $\mathcal{A}$ with ancestry from K populations, we define a vector $\Phi^{\mathcal{A}}$ containing K^2 entries, $\Phi^{\mathcal{A}} = (\phi_{11}^{\mathcal{A}}, \phi_{12}^{\mathcal{A}}, \phi_{21}^{\mathcal{A}}, ..., \phi_{KK}^{\mathcal{A}})$, where $\phi_{a_1 a_2}^{\mathcal{A}}$ gives the proportion of sites from the individual's genome where the first allele has ancestry from ancestral population a_1 and the second one from a_2. Although the pairs of ancestries are ordered we assume that they are symmetric i.e. $\phi_{a_1 a_2}^{\mathcal{A}} = \phi_{a_2 a_1}^{\mathcal{A}}$. We assume that we know, or that we have previously estimated, the ancestral allele frequencies in each population, F. The likelihood of observing the sequencing data $X^{\mathcal{A}}$ for a single individual, M sites and K ancestral populations is then

$$P(X^{\mathcal{A}}|F, \Phi^{\mathcal{A}}) = \prod_{j=1}^{M} P(X_j^{\mathcal{A}}|F_j, \Phi^{\mathcal{A}}) = \prod_{j=1}^{M} \sum_{a_1^{\mathcal{A}}=1}^{K} \sum_{a_2^{\mathcal{A}}=a_1^{\mathcal{A}}}^{K} P(X_j^{\mathcal{A}}|A_j^{\mathcal{A}} = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}}), F_j) P(A_j^{\mathcal{A}} = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}})|\Phi^{\mathcal{A}}),$$

where the probability of a paired ancestry is directly specified by $\Phi^{\mathcal{A}}$,

$$P(A_j^{\mathcal{A}} = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}})|\Phi^{\mathcal{A}}) = \phi_{a_1^{\mathcal{A}} a_2^{\mathcal{A}}}^{\mathcal{A}}.$$

The probability of the NGS data $X^{\mathcal{A}}$ at site j given that site has paired ancestry $A_j^{\mathcal{A}} = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}})$ is

$$P(X_j^{\mathcal{A}}|A_j^{\mathcal{A}} = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}}), F_j) = \sum_{g^{\mathcal{A}} \in \{0,1\}^2} P(G_j^{\mathcal{A}} = g^{\mathcal{A}}|A_j = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}}), F_j) P(X_j^{\mathcal{A}}|G_j^{\mathcal{A}} = g^{\mathcal{A}}),$$

where

$$P(G_j^{\mathcal{A}} = g^{\mathcal{A}}|A_j^{\mathcal{A}} = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}}), F_j) = \begin{cases} (1 - F_{ja_1^{\mathcal{A}}})(1 - F_{ja_2^{\mathcal{A}}}) & if \quad g^{\mathcal{A}} = (0,0) \\ F_{ja_1^{\mathcal{A}}}(1 - F_{ja_2^{\mathcal{A}}}) & if \quad g^{\mathcal{A}} = (1,0) \\ (1 - F_{ja_1^{\mathcal{A}}})F_{ja_2^{\mathcal{A}}} & if \quad g^{\mathcal{A}} = (0,1) \\ F_{ja_1^{\mathcal{A}}}F_{ja_2^{\mathcal{A}}} & if \quad g^{\mathcal{A}} = (1,1) \end{cases}.$$

EM for Φ In the following $\Phi^{\mathcal{A}}$ and $X^{\mathcal{A}}$ refer respectively to the paired ancestry proportions and the sequencing data for a single individual $\mathcal{A}$. We use an expectation maximization (EM) algorithm to obtain a maximum likelihood estimate $\Phi^{\mathcal{A}}$. We denote the ith iteration of the estimated pairwise ancestry proportion $\Phi^{(i)}$. Based on this estimate, the expectation of the total number of sites with paired ancestries $A^{\mathcal{A}} = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}})$ is given by

$$\mathbb{E}[A^{\mathcal{A}} = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}})|X^{\mathcal{A}}, F, \Phi^{(i)}] = \sum_{j=1}^{M} P(A_j^{\mathcal{A}} = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}})|X_j^{\mathcal{A}}, F_j, \Phi^{(i)}) =$$

$$\sum_{j=1}^{M} \frac{P(X_j^{\mathcal{A}}|A_j^{\mathcal{A}} = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}}), F_j) P(A_j^{\mathcal{A}} = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}})|\Phi^{(i)})}{\sum_{a_1'=1}^{K} \sum_{a_2'=a_1'}^{K} P(X_j^{\mathcal{A}}|A^{\mathcal{A}} = (a_1', a_2'), F_j) P(A^{\mathcal{A}} = (a_1', a_2')|\Phi^{(i)})}. \tag{1}$$

Then we can update the estimate of Φ by calculating each of its entries as

$$\phi_{a_1 a_2}^{(i+1)} = \frac{\mathbb{E}[A^{\mathcal{A}} = (a_1^{\mathcal{A}}, a_2^{\mathcal{A}})|X^{\mathcal{A}}, F, \Phi^{(i)}]}{\sum_{a_1'=1}^{K} \sum_{a_2'=a_1'}^{K} \mathbb{E}[A^{\mathcal{A}} = (a_1', a_2')|X^{\mathcal{A}}, F, \Phi^{(i)}]}. \tag{2}$$

Finally, we calculate the new expectation given $\Phi^{(i+1)}$ with equation (1) and use the expectation to calculate a new estimate with equation (2) iterating until convergence. A proof that this is an EM for a similar case can by found in Skotte *et al.* (2013).

As mentioned, NGSremix handles admixture by estimating paired ancestry proportions. We have also implemented a version of NGSremix without paired ancestry where we assume the ancestral populations to be discrete. The paired ancestry proportions result in less variance when the relatedness is high, see Supplementary Material Figure S5.
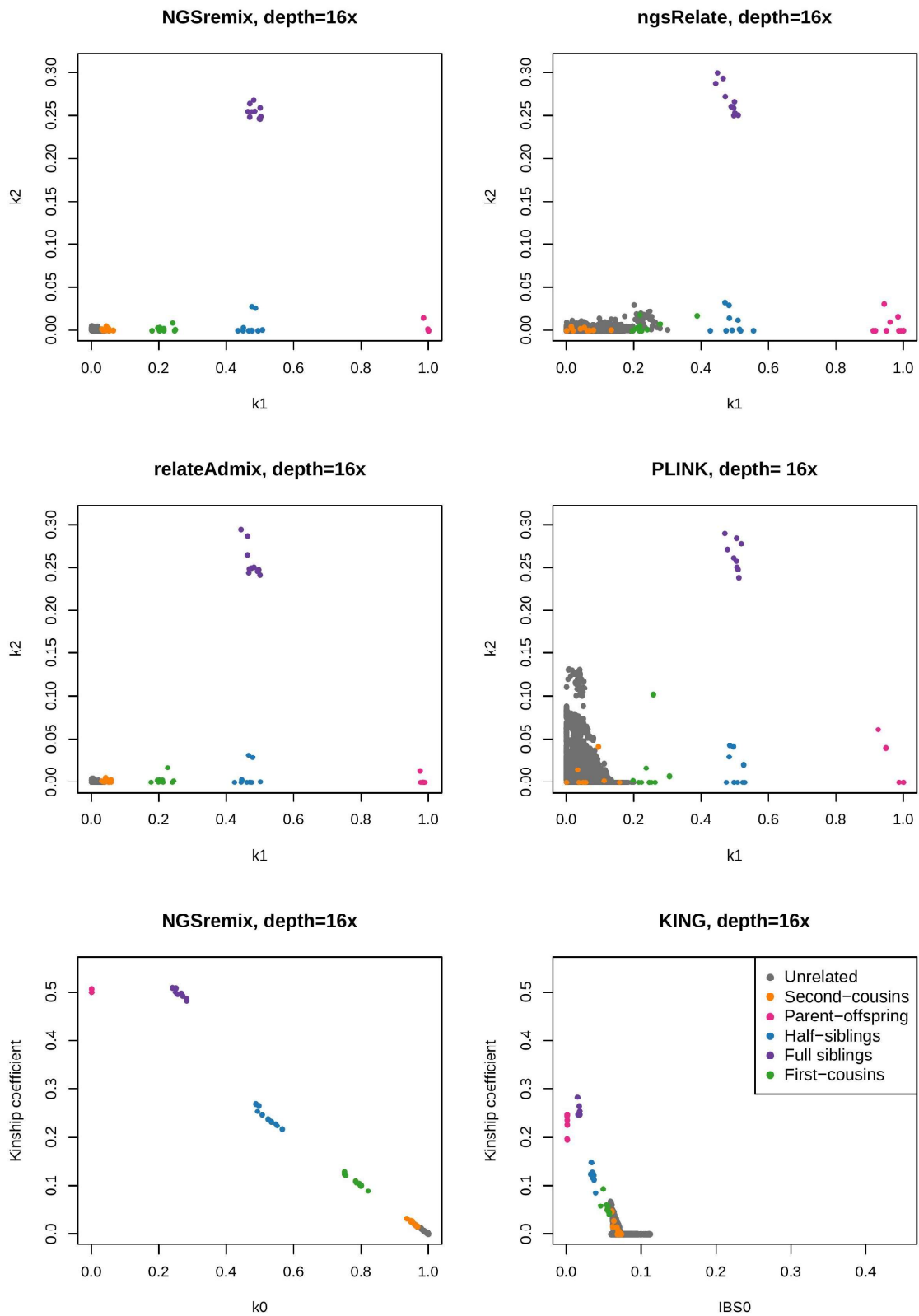
Figure S1 Estimated k_1 and k_2 or kinship for 120 simulated admixed individuals with 6 relationship types (10 of each type of related pairs) and an average depth of 16x. Plots of estimated R values are visualized for NGSremix, ngsRelate, reateAdmix, and PLINK. Plots of the kinship coefficient for NGSremix and KING are provided since KING only estimates kinship
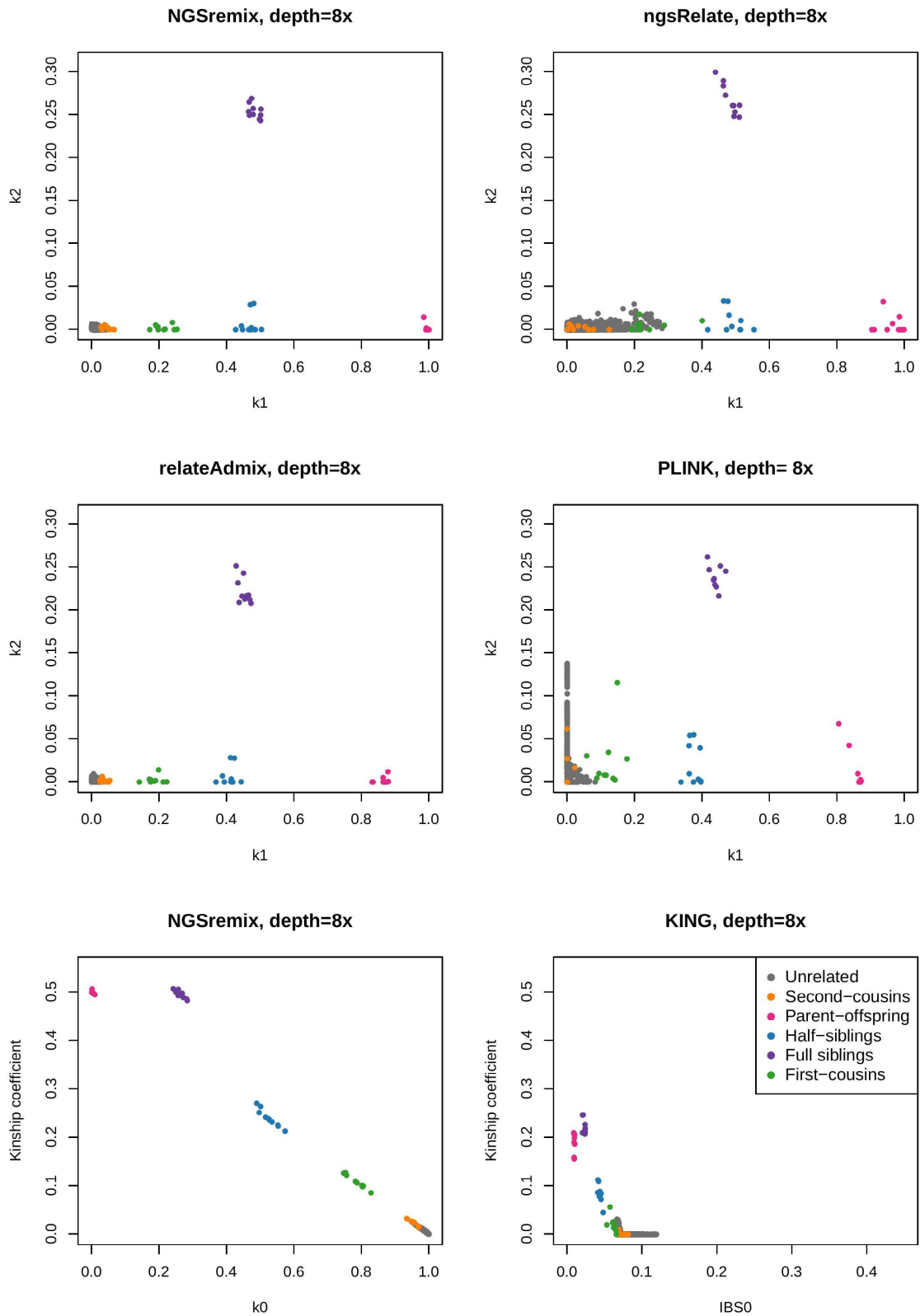
Figure S2 Estimated k_1 and k_2 or kinship for 120 simulated admixed individuals with 6 relationship types (10 of each type of related pairs) and an average depth of 8x. Plots of estimated R values are visualized for NGSremix, ngsRelate, reateAdmix, and PLINK. Plots of the kinship coefficient for NGSremix and KING are provided since KING only estimates kinship
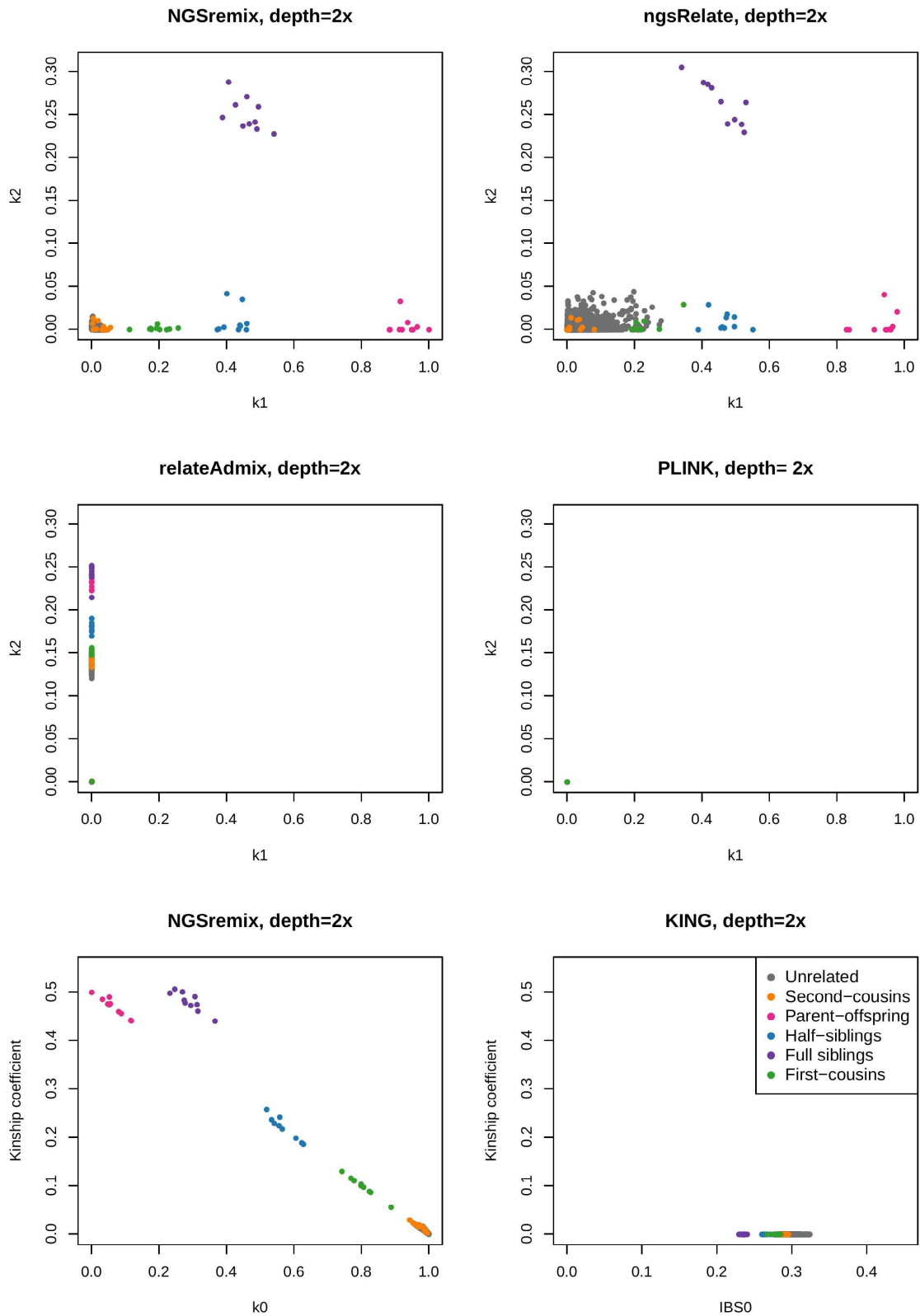
Figure S3 Estimated k_1 and k_2 or kinship for 120 simulated admixed individuals with 6 relationship types (10 of each type of related pairs) and an average depth of 2x. Plots of estimated R values are visualized for NGSremix, ngsRelate, reateAdmix, and PLINK. Plots of the kinship coefficient for NGSremix and KING are provided since KING only estimates kinship
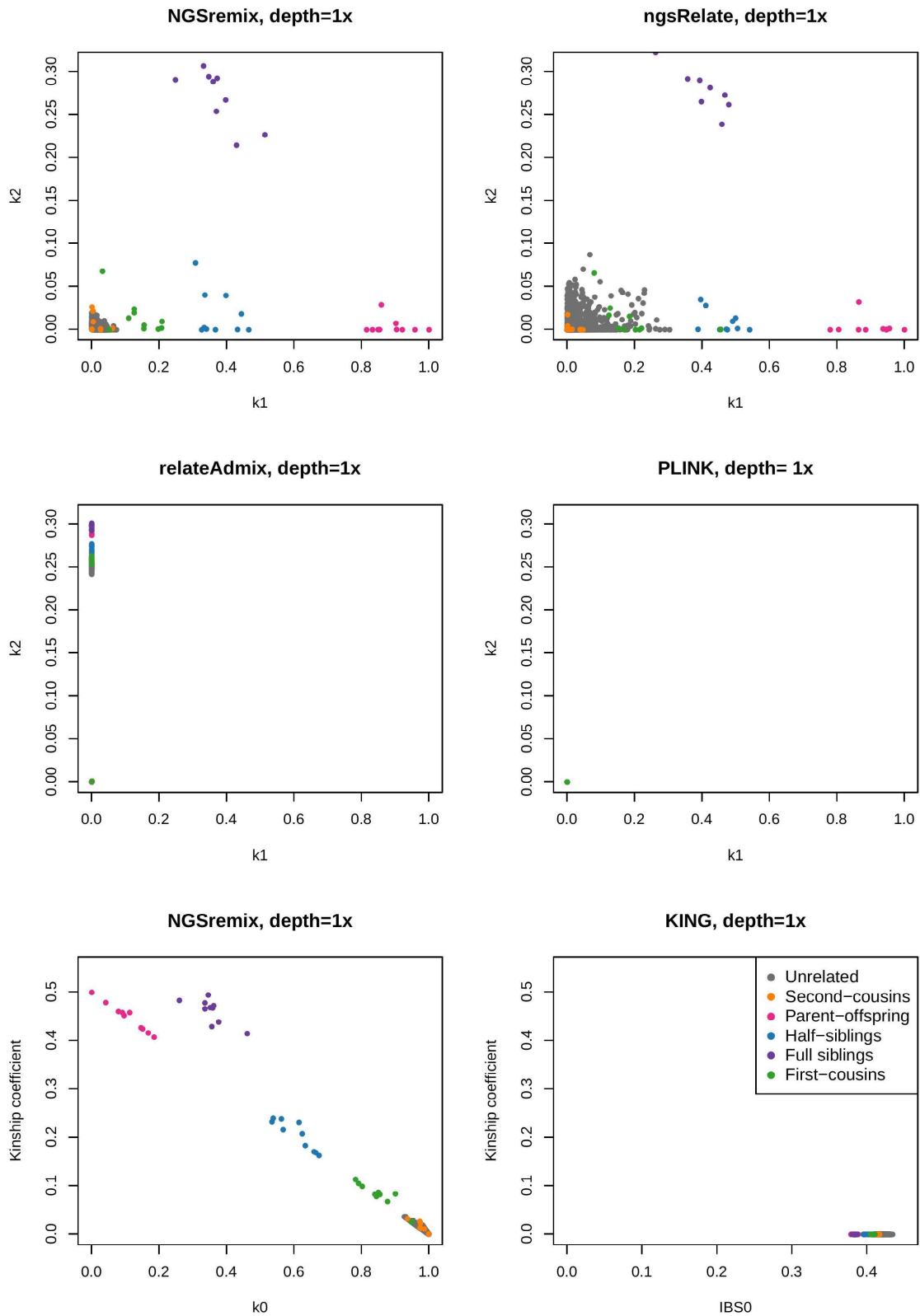
Figure S4 Estimated k_1 and k_2 or kinship for 120 simulated admixed individuals with 6 relationship types (10 of each type of related pairs) and an average depth of 1x. Plots of estimated R values are visualized for NGSremix, ngsRelate, reateAdmix, and PLINK. Plots of the kinship coefficient for NGSremix and KING are provided since KING only estimates kinship
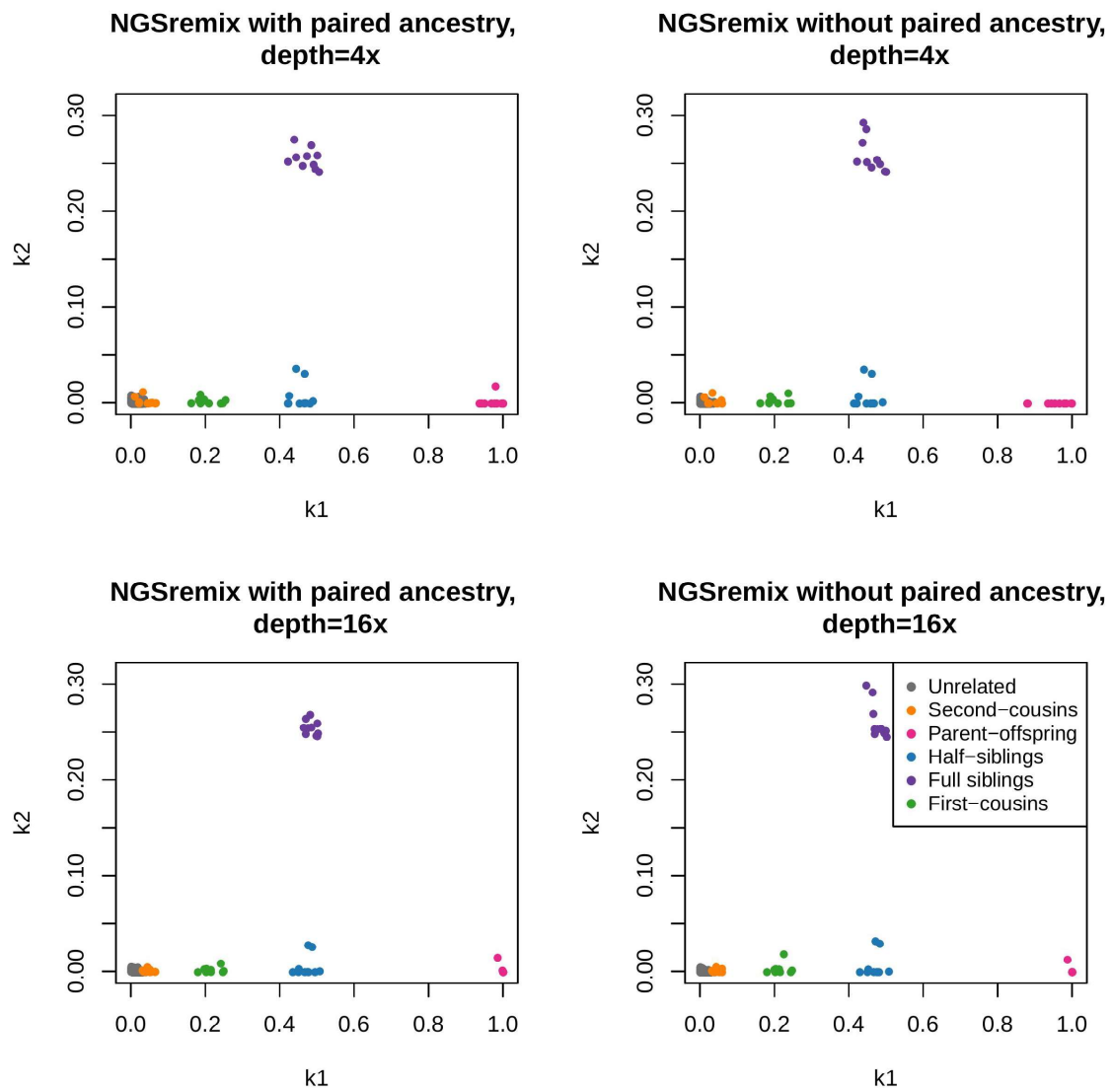
Figure S5 k_1 and k_2 values estimated using NGSremix with and without paired ancestry for 120 simulated admixed individuals and 6 relationship types (10 of each type of related pairs) and an average depth of 4x or 16x.
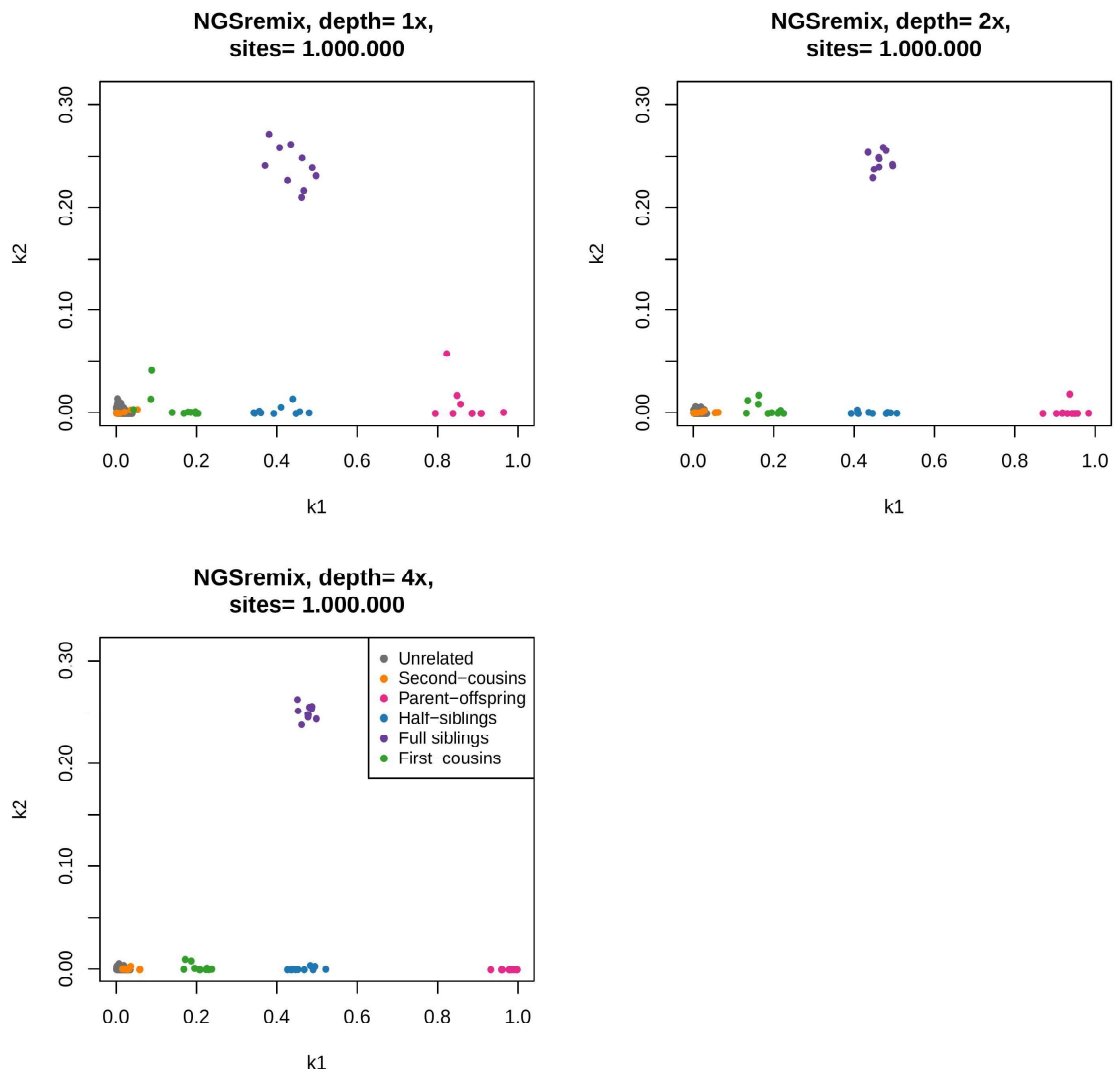
Figure S6 k_1 and k_2 values estimated using NGSremix for 120 simulated admixed individuals with 1.000.000 sites and 6 relationship types (10 of each type of related pairs) and an average depth of 1x, 2x, or 4x.
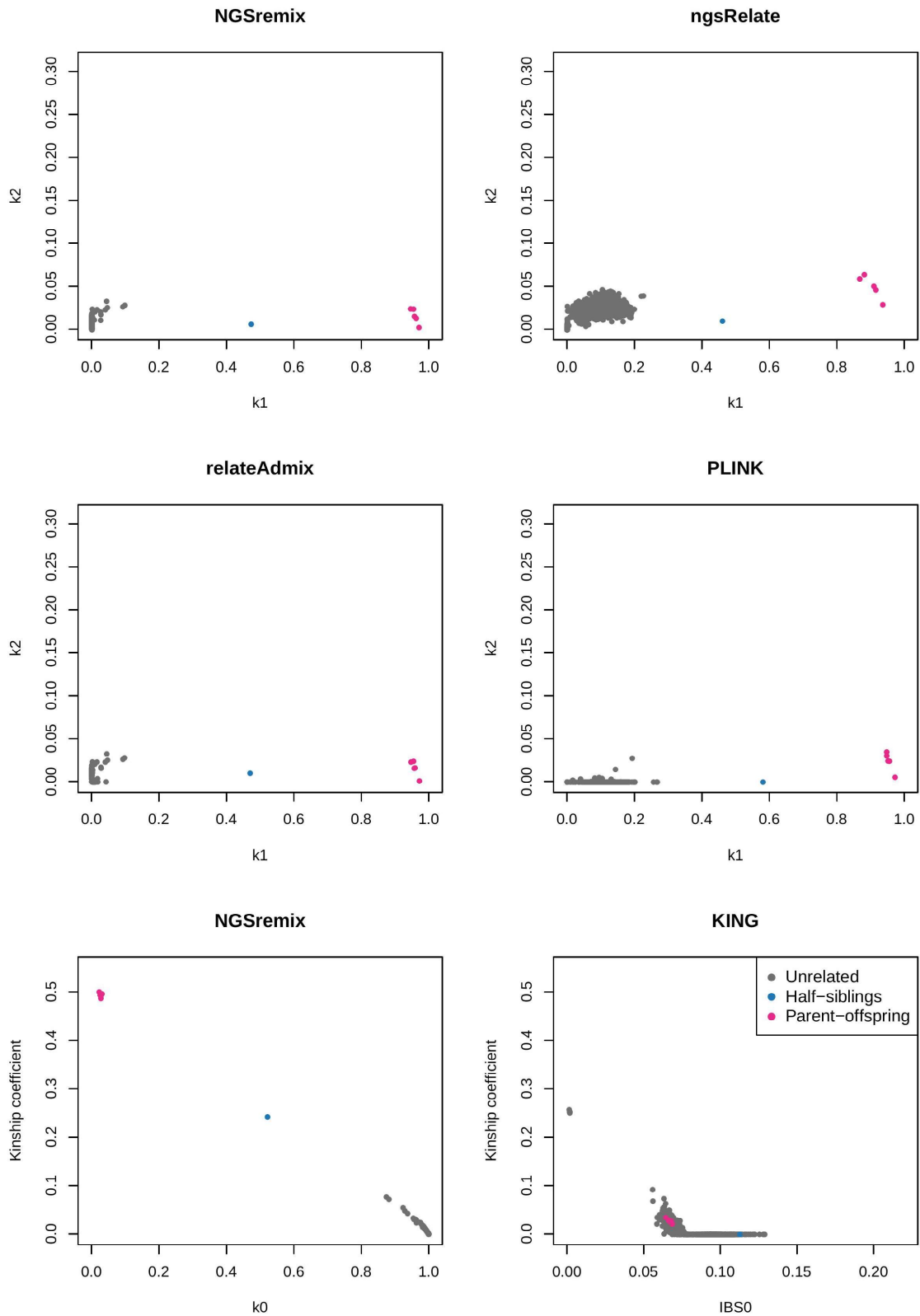
Figure S7 Inferred relatedness, k_1 and k_2 or kinship, for 61 admixed African Americans from the 1000 genomes project sequenced at high depth. The ancestry is modelled using the European, African, and Native American populations in the reference panel. Plots of estimated R values are visualized for NGSremix, ngsRelate, relateAdmix, and PLINK. Plots of the kinship coefficient for NGSremix and KING are provided since KING estimates kinship and not R.
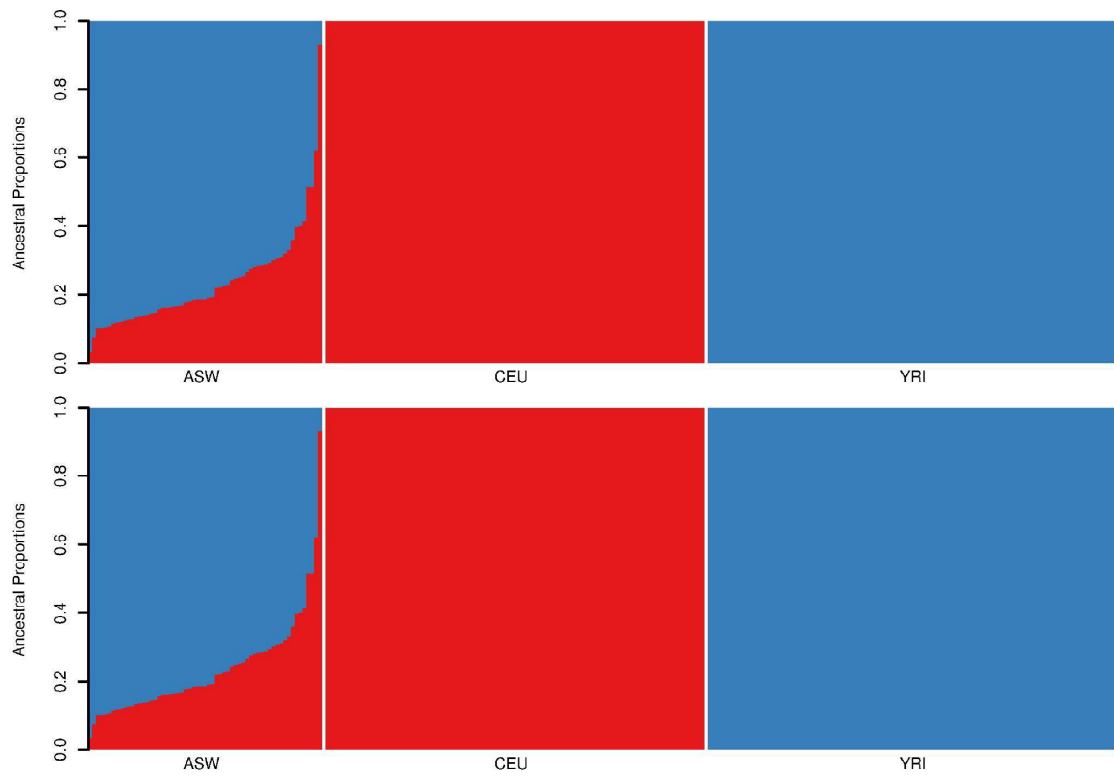
Figure S8 Top: Ancestry proportions for ASW, CEU, and YRI allowing for two ancestry components. **Bottom:** Ancestry proportions for ASW, CEU, MXL, PEL, and YRI allowing for three ancestry components.
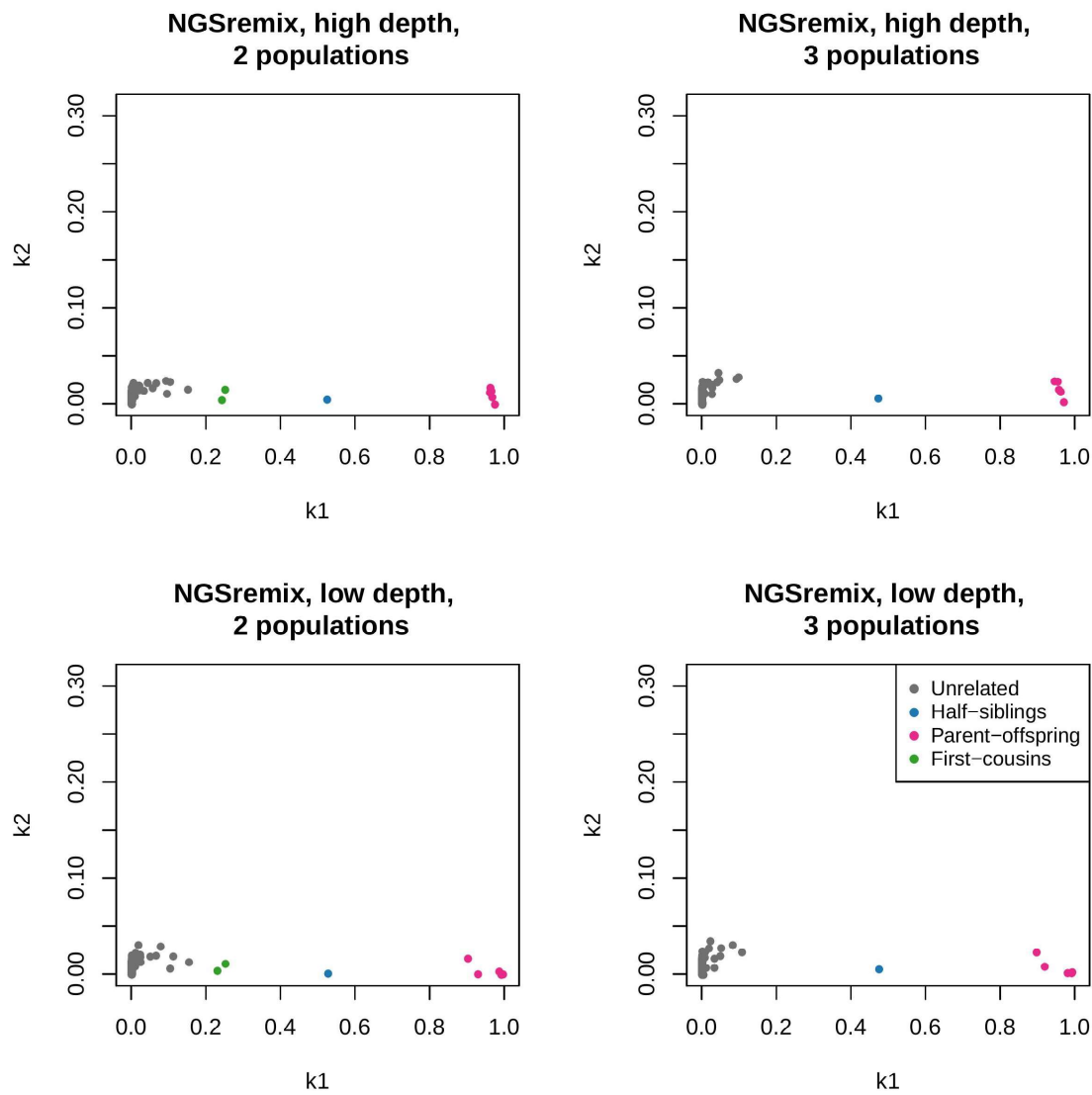
Figure S9 Inferred relatedness, k_1 and k_2 for all pairs of 61 admixed individuals from the ASW population from the 1000 genomes project based on the high quality genotype dataset and the low-depth NGS dataset using NGSremix. The ASW ancestry is modelled using European and African populations in the reference panel (plots to the left) and using European, African and Native American populations in the reference panel (plots to the right). The ancestry proportions are visualized in Supplementary Figure S8 .

Figure S10 Left: NGSremix based on GL. **Center**: Called genotypes based on maximal genotype likelihood. **Right**: Called genotypes using a 0.95 genotype likelihood threshold. The results are based on the low-depth NGS data from ASW population modelled with three ancestral populations.
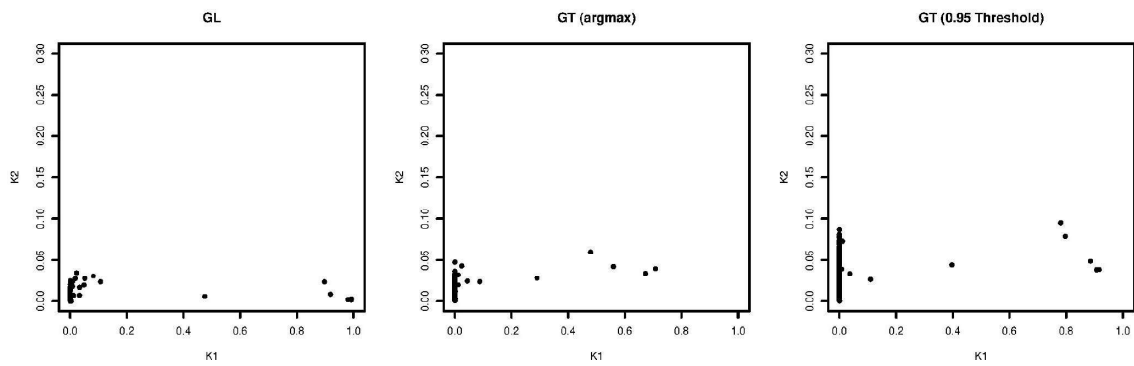
Supplementary tables

Table S1 Calculation of $P(G_j^A = g^A, G_j^B = g^B | A_j = a, Z_j = z)$ for all possible combinations of g^A, g^B, z_1, z_2, and $a = (a_1^A, a_2^A, a_1^B, a_2^B)$. From the ancestral allele frequency matrix F we extract the allele frequency from site j of allele 0 from population k and define as $p_k = F_{jk}$. The allele frequency of allele 1 in population k we define as $q_k = 1 - p_k$.

$g^A\ g^B$	$z_1 = 0\ z_2 = 0$	$z_1 = 0\ z_2 = 1$	$z_1 = 1\ z_2 = 0$	$z_1 = 1\ z_2 = 1$
00 00	$p_{a_1^A} p_{a_2^A} p_{a_1^B} p_{a_2^B}$	$p_{a_1^A} p_{a_2^A} p_{a_1^B}$	$p_{a_1^A} p_{a_2^A} p_{a_2^B}$	$p_{a_1^A} p_{a_2^A}$
00 01	$p_{a_1^A} p_{a_2^A} p_{a_1^B} q_{a_2^B}$	0	$p_{a_1^A} p_{a_2^A} q_{a_2^B}$	0
00 10	$p_{a_1^A} p_{a_2^A} q_{a_1^B} p_{a_2^B}$	$p_{a_1^A} p_{a_2^A} q_{a_1^B}$	0	0
00 11	$p_{a_1^A} p_{a_2^A} q_{a_1^B} q_{a_2^B}$	0	0	0
01 10	$p_{a_1^A} q_{a_2^A} q_{a_1^B} p_{a_2^B}$	0	0	0
01 00	$p_{a_1^A} q_{a_2^A} p_{a_1^B} p_{a_2^B}$	0	$p_{a_1^A} q_{a_2^A} p_{a_2^B}$	0
01 01	$p_{a_1^A} q_{a_2^A} p_{a_1^B} q_{a_2^B}$	$p_{a_1^A} q_{a_2^A} p_{a_1^B}$	$p_{a_1^A} q_{a_2^A} q_{a_2^B}$	$p_{a_1^A} q_{a_2^A}$
01 11	$p_{a_1^A} q_{a_2^A} q_{a_1^B} q_{a_2^B}$	$p_{a_1^A} q_{a_2^A} q_{a_1^B}$	0	0
10 00	$q_{a_1^A} p_{a_2^A} p_{a_1^B} p_{a_2^B}$	$q_{a_1^A} p_{a_2^A} q_{a_1^B}$	0	0
10 01	$p_{a_1^A} q_{a_2^A} q_{a_1^B} p_{a_2^B}$	0	0	0
10 11	$q_{a_1^A} p_{a_2^A} q_{a_1^B} q_{a_2^B}$	0	$q_{a_1^A} p_{a_2^A} q_{a_2^B}$	0
10 10	$q_{a_1^A} p_{a_2^A} q_{a_1^B} p_{a_2^B}$	$q_{a_1^A} p_{a_2^A} q_{a_1^B}$	$q_{a_1^A} p_{a_2^A} p_{a_2^B}$	$q_{a_1^A} p_{a_2^A}$
11 00	$q_{a_1^A} q_{a_2^A} p_{a_1^B} p_{a_2^B}$	0	0	0
11 01	$q_{a_1^A} q_{a_2^A} p_{a_1^B} q_{a_2^B}$	$q_{a_1^A} q_{a_2^A} p_{a_1^B}$	0	0
11 10	$q_{a_1^A} q_{a_2^A} q_{a_1^B} p_{a_2^B}$	0	$q_{a_1^A} q_{a_2^A} p_{a_2^B}$	0
11 11	$q_{a_1^A} q_{a_2^A} q_{a_1^B} q_{a_2^B}$	$q_{a_1^A} q_{a_2^A} q_{a_1^B}$	$q_{a_1^A} q_{a_2^A} q_{a_2^B}$	$q_{a_1^A} q_{a_2^A}$

Table S2 Computation time for methods applied to estimate relatedness on real low-depth sequencing data (individuals = 417, pairwise comparisons = 86736, sites = 255046), described in Section Performance assessment using real data. The no. of threads were set to 20 for all methods except for KING where this setting is not possible.

Methods	Wall time
NGSremix paired ancestry	2hr 58min 51sec
NGSremix not paired ancestry	3hr 36min 10sec
PLINK	4sec
ngsRelate	2hr 9min 14sec
relateAdmix	12min 50sec
KING	1sec

a	$z_1=0\ z_2=0$	$z_1=1\ z_2=0$	$z_1=0\ z_2=1$	$z_1=1\ z_2=1$
$a_1^{\mathcal{A}} \neq a_1^{\mathcal{B}}\quad a_2^{\mathcal{A}} \neq a_2^{\mathcal{B}}$	$\Phi^{\mathcal{A}}_{a_1^{\mathcal{A}},a_2^{\mathcal{A}}}\Phi^{\mathcal{B}}_{a_1^{\mathcal{B}},a_2^{\mathcal{B}}}$	0	0	0
$a_1^{\mathcal{A}} = a_1^{\mathcal{B}}\quad a_2^{\mathcal{A}} \neq a_2^{\mathcal{B}}$	$\Phi^{\mathcal{A}}_{a_1^{\mathcal{A}},a_2^{\mathcal{A}}}\Phi^{\mathcal{B}}_{a_1^{\mathcal{B}},a_2^{\mathcal{B}}}$	$\dfrac{\Phi^{\mathcal{A}}_{a_1^{\mathcal{A}},a_2^{\mathcal{A}}}\Phi^{\mathcal{B}}_{a_1^{\mathcal{B}},a_2^{\mathcal{B}}}}{\sum\limits_{x=1}^{K}\sum\limits_{y=1}^{K}\sum\limits_{k=1}^{K}\Phi^{\mathcal{A}}_{k,x}\Phi^{\mathcal{B}}_{k,y}}$	0	0
$a_1^{\mathcal{A}} \neq a_1^{\mathcal{B}}\quad a_2^{\mathcal{A}} = a_2^{\mathcal{B}}$	$\Phi^{\mathcal{A}}_{a_1^{\mathcal{A}},a_2^{\mathcal{A}}}\Phi^{\mathcal{B}}_{a_1^{\mathcal{B}},a_2^{\mathcal{B}}}$	0	$\dfrac{\Phi^{\mathcal{A}}_{a_1^{\mathcal{A}},a_2^{\mathcal{A}}}\Phi^{\mathcal{B}}_{a_1^{\mathcal{B}},a_2^{\mathcal{B}}}}{\sum\limits_{x=1}^{K}\sum\limits_{y=1}^{K}\sum\limits_{k=1}^{K}\Phi^{\mathcal{A}}_{x,k}\Phi^{\mathcal{B}}_{y,k}}$	0
$a_1^{\mathcal{A}} = a_1^{\mathcal{B}}\quad a_2^{\mathcal{A}} = a_2^{\mathcal{B}}$	$\Phi^{\mathcal{A}}_{a_1^{\mathcal{A}},a_2^{\mathcal{A}}}\Phi^{\mathcal{B}}_{a_1^{\mathcal{B}},a_2^{\mathcal{B}}}$	$\dfrac{\Phi^{\mathcal{A}}_{a_1^{\mathcal{A}},a_2^{\mathcal{A}}}\Phi^{\mathcal{B}}_{a_1^{\mathcal{B}},a_2^{\mathcal{B}}}}{\sum\limits_{x=1}^{K}\sum\limits_{y=1}^{K}\sum\limits_{k=1}^{K}\Phi^{\mathcal{A}}_{k,x}\Phi^{\mathcal{B}}_{k,y}}$	$\dfrac{\Phi^{\mathcal{A}}_{a_1^{\mathcal{A}},a_2^{\mathcal{A}}}\Phi^{\mathcal{B}}_{a_1^{\mathcal{B}},a_2^{\mathcal{B}}}}{\sum\limits_{x=1}^{K}\sum\limits_{y=1}^{K}\sum\limits_{k=1}^{K}\Phi^{\mathcal{A}}_{x,k}\Phi^{\mathcal{B}}_{y,k}}$	$\dfrac{\Phi^{\mathcal{A}}_{a_1^{\mathcal{A}},a_2^{\mathcal{A}}}\Phi^{\mathcal{B}}_{a_1^{\mathcal{B}},a_2^{\mathcal{B}}}}{\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\Phi^{\mathcal{A}}_{k_1,k_2}\Phi^{\mathcal{B}}_{k_1,k_2}}$