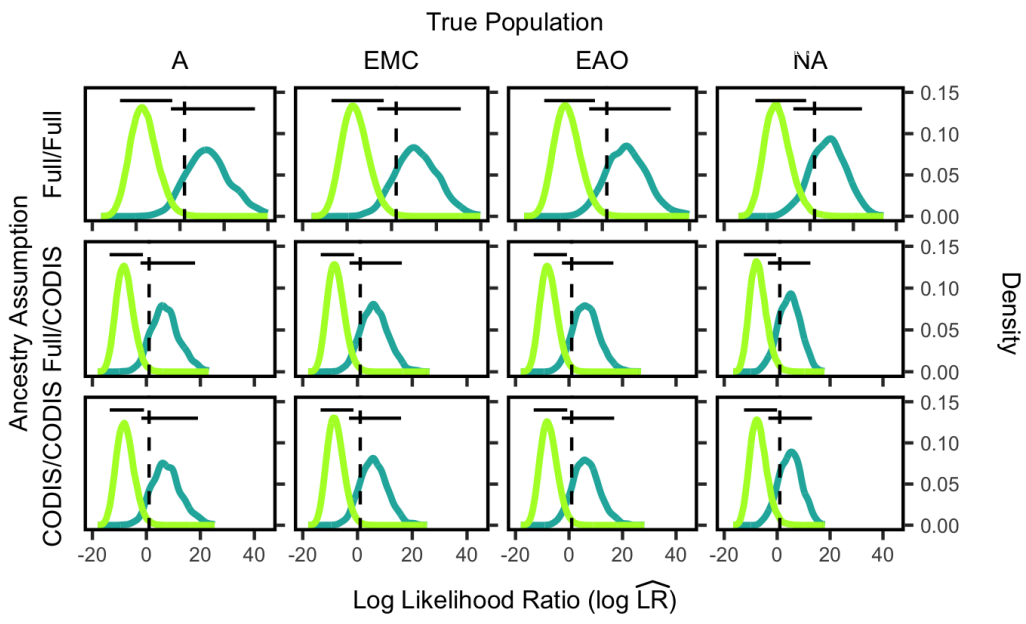
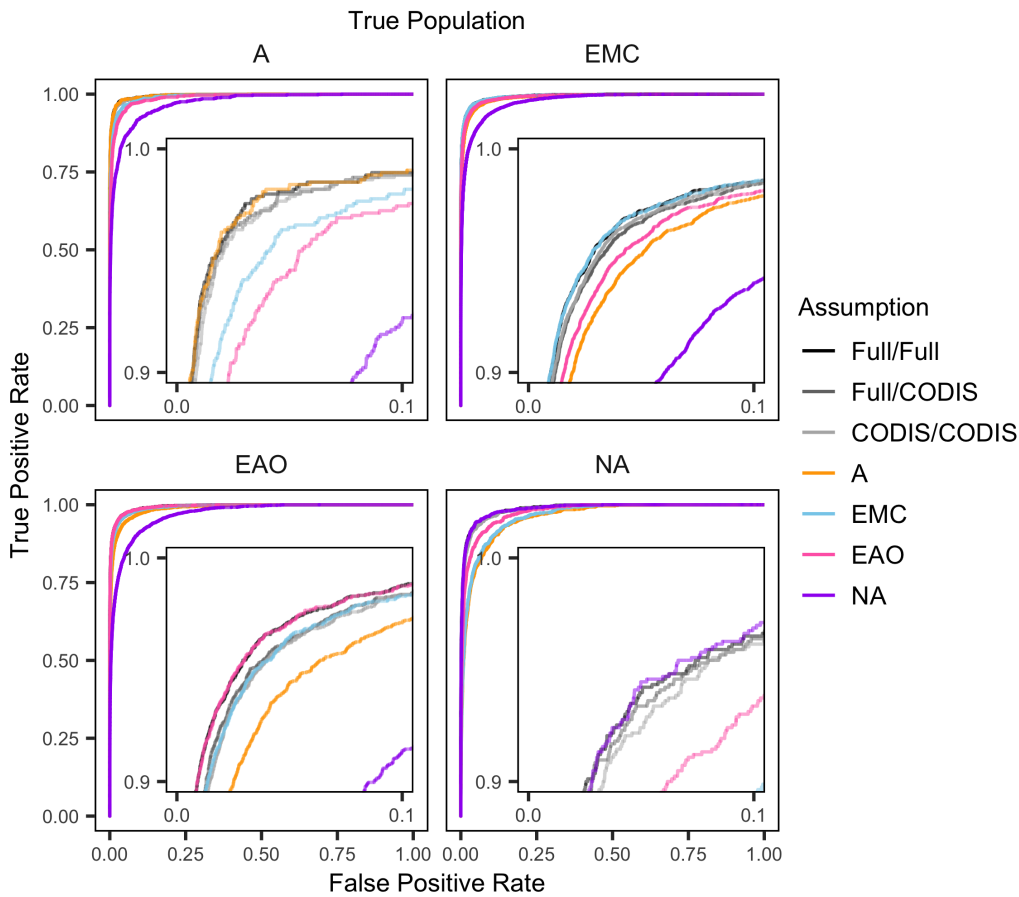


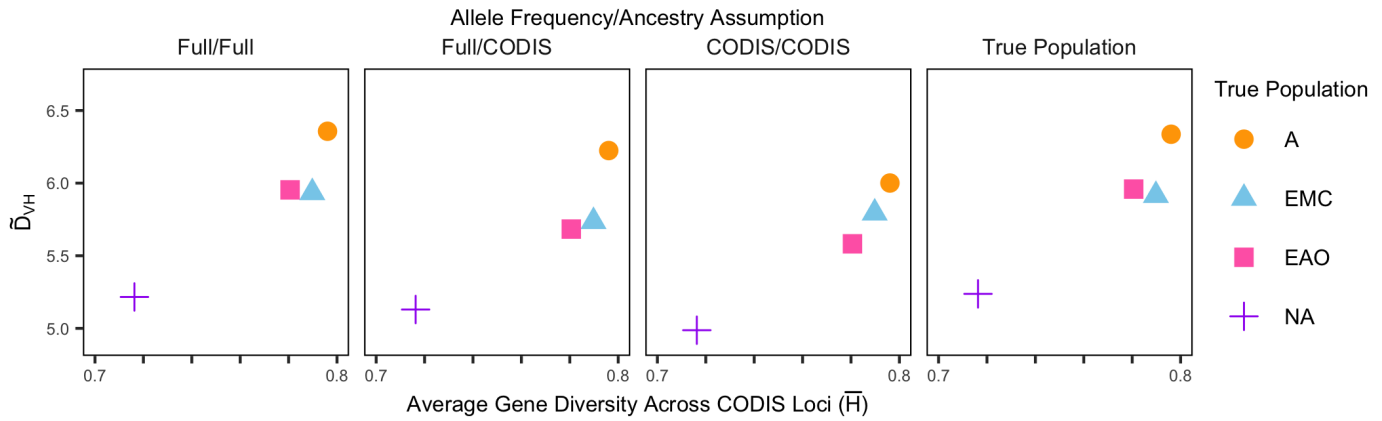
**Figure S1** Log likelihood ratio ( $\log \widehat{LR}$ ) distributions for siblings and unrelated individuals by population group for allele frequencies chosen by the *Predefined-Population* method, with coancestry assumption  $\theta = 0$  (compare with Figure 1). Each plot shows the  $\widehat{LR}$  distributions for unrelated individuals in light green and true siblings in dark green, with each  $\widehat{LR}$  calculated from Equation 4. The dashed vertical lines indicate  $\widehat{LR} = 1$ . The horizontal lines show the central 95% of each distribution. Each distribution in the A column consists of  $94 \times (940 - 10)$  points and  $94 \times 10$  points for the unrelated pairs and related pairs, respectively. Each distribution in the EMC column consists of  $532 \times (5320 - 10)$  and  $532 \times 10$  pairs, respectively. Each distribution in the EAO column consists of  $269 \times (2690 - 10)$  and  $269 \times 10$  pairs, respectively. Each distribution in the NA column consists of  $83 \times (830 - 10)$  and  $83 \times 10$  pairs, respectively. A, African; EMC, European, Middle Eastern, and Central/South Asian; EAO, East Asian and Oceanian; NA, Native American.



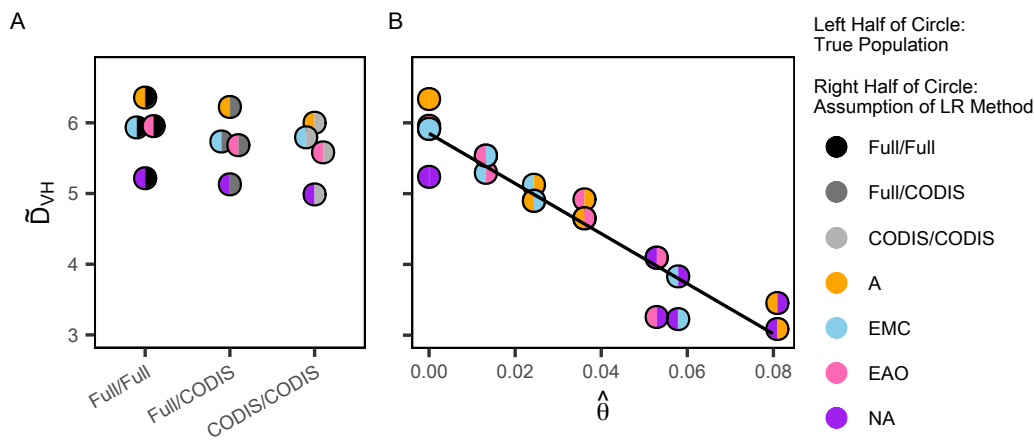
**Figure S2** Log likelihood ratio ( $\log \widehat{LR}$ ) distributions for siblings and unrelated individuals by population group for allele frequencies chosen by the *Ancestry-Estimation* method, with coancestry assumption  $\theta = 0$  (compare with Figure 3). The labels on the left side indicate the scenario assumed, either Full/Full, Full/CODIS, or CODIS/CODIS. The figure design otherwise follows Figure S1.



**Figure S3** Receiver-operating-characteristic (ROC) curves showing true positive rate as a function of false positive rate in assigning individuals as siblings, with coancestry assumption  $\theta = 0$  (compare with Figure 4). The plots are calculated from the distributions in Figures S1 and S2. Each curve for A uses  $94 \times 940$  pairs, each curve for EMC uses  $532 \times 5320$  pairs, each curve for EAO uses  $269 \times 2690$  pairs, and each curve for NA uses  $83 \times 830$  pairs. The inset panels show the detail at the upper left corner of each plot.



**Figure S4** The empirical distinguishability ( $\bar{D}_{VH}$ ) for siblings and unrelated individuals as a function of average gene diversity across the 13 CODIS loci,  $\bar{H}$ , with coancestry assumption  $\theta = 0$  (compare with Figure 5). Points are colored according to the true population group. Each panel considers a different pair of assumptions about allele frequencies and ancestry in computing the likelihood ratios, as shown in Figures S1 and S2.  $\bar{D}_{VH}$  is computed from Equation 5 and  $\bar{H}$  is computed from Equation 6.



**Figure S5** The empirical distinguishability ( $\tilde{D}_{VH}$ ) for siblings and unrelated individuals. (A) Distinguishability for the Full/Full, Full/CODIS, and CODIS/CODIS *Ancestry-Estimation* scenarios. (B) Distinguishability for the *Predefined-Population* scenarios as a function of the estimated coancestry coefficient,  $\hat{\theta}$ , for pairs of populations, one reporting the true population and the other reporting the source population for the allele frequencies. The  $\theta$  estimate of genetic distance between two populations is from the 13 CODIS loci only, as in the lower triangle of Table 3. The coancestry assumption is  $\theta = 0$  (compare with Figure 6).  $\hat{\theta} = 0$  indicates that the true population and source population are the same. The left color of each circle corresponds to the true population group, and the right color of each circle corresponds to the assumption used in the LR calculations. The Full/Full, Full/CODIS, and CODIS/CODIS cases are plotted separately in A for comparison with the *Predefined-Population* case with “correctly-specified” populations (single-color circles in B at  $\hat{\theta} = 0$ ).  $\tilde{D}_{VH}$  values are taken from Table 2. The equation of the regression line is  $\tilde{D}_{VH} = 5.851 - 35.416 \hat{\theta}$ .

■ **Table S1**  $\tilde{D}_{VH}$  of major population groups, assuming allele frequencies from each major population group for the *Predefined-Population* method, with coancestry assumption  $\theta = 0$  (compare with Table 1).  $\tilde{D}_{VH}$  values are calculated using Equation 5 from the distributions plotted in Figure S1. A, African; EMC, European, Middle Eastern, and Central/South Asian; EAO, East Asian and Oceanian; NA, Native American.

| Assumed Population | True Population |      |      |      |
|--------------------|-----------------|------|------|------|
|                    | A               | EMC  | EAO  | NA   |
| A                  | 6.34            | 5.13 | 4.92 | 3.09 |
| EMC                | 4.90            | 5.92 | 5.54 | 3.23 |
| EAO                | 4.65            | 5.29 | 5.96 | 4.10 |
| NA                 | 3.45            | 3.83 | 3.26 | 5.24 |

■ **Table S2**  $\tilde{D}_{VH}$  for both methods, *Predefined-Population* and *Ancestry-Estimation*, with coancestry assumption  $\theta = 0$  (compare with Table 2). **Full/Full**: Full-data allele frequencies and full-data ancestry proportions from STRUCTURE runs with 791 loci. **Full/CODIS**: Full-data allele frequencies from STRUCTURE runs with 791 loci and CODIS ancestry proportions from STRUCTURE runs with 13 CODIS loci. **CODIS/CODIS**: CODIS allele frequencies and CODIS ancestry proportions from STRUCTURE runs with 13 loci. **Best-Specified**: Allele-frequencies from the assumed population to which the individuals and siblings belong. **Second-Best-Specified**: The second-highest distinguishability value from each column of Table S1, assuming the allele frequencies from the second-best assumed population. **Third-Best-Specified**: The third-highest distinguishability value from each column of Table S1, assuming the allele frequencies from the third-best assumed population. **Fourth-Best-Specified**: The lowest distinguishability value from each column of Table S1, assuming the allele frequencies from the fourth-best assumed population.  $\tilde{D}_{VH}$  values are calculated using Equation 5 and the distributions plotted in Figures S1 and S2.

| Assumption                       | True Population |      |      |      |
|----------------------------------|-----------------|------|------|------|
|                                  | A               | EMC  | EAO  | NA   |
| Best-Specified Population        | 6.34            | 5.92 | 5.96 | 5.24 |
| Full/Full                        | 6.36            | 5.93 | 5.95 | 5.22 |
| Full/CODIS                       | 6.22            | 5.74 | 5.68 | 5.13 |
| CODIS/CODIS                      | 6.00            | 5.79 | 5.58 | 4.99 |
| Second-Best-Specified Population | 4.90            | 5.29 | 5.54 | 4.10 |
| Third-Best-Specified Population  | 4.65            | 5.13 | 4.92 | 3.23 |
| Fourth-Best-Specified Population | 3.45            | 3.83 | 3.26 | 3.09 |

■ **Table S3** Mean  $\bar{D}_{VH}$  for down-sampled analyses. We performed 10 replicate analyses, in each one choosing 83 individuals at random within each of the four ancestry groups (A, EMC, EAO, NA). This size reflects the size of the NA group, all of whose individuals were used in each of the 10 replicates. For each set of  $83 \times 4 = 332$  individuals, we performed STRUCTURE analyses with both the full 791 loci and the 13 CODIS loci, in each case conducting 10 replicate STRUCTURE runs following the same procedure as was used with the full set of 978 individuals (see [Materials and Methods](#)). The distinguishability values in the table were obtained with coancestry assumption  $\theta = 0.01$ . The table gives the mean distinguishability values across the 10 replicate down-sampled data sets. **Full/Full**: Full-data allele frequencies and full-data ancestry proportions from STRUCTURE runs with 791 loci. **Full/CODIS**: Full-data allele frequencies from STRUCTURE runs with 791 loci and CODIS ancestry proportions from STRUCTURE runs with 13 CODIS loci. **CODIS/CODIS**: CODIS allele frequencies and CODIS ancestry proportions from STRUCTURE runs with 13 loci. **Best-Specified**: Allele frequencies from the assumed population to which the individuals and siblings belong. **Second-Best-Specified**: The second-highest distinguishability value, assuming the allele frequencies from the second-best assumed population. **Third-Best-Specified**: The third-highest distinguishability value, assuming the allele frequencies from the third-best assumed population. **Fourth-Best-Specified**: The lowest distinguishability value, assuming the allele frequencies from the fourth-best assumed population.  $\bar{D}_{VH}$  values are calculated using Equation 5.

| Assumption                       | True Population |      |      |      |
|----------------------------------|-----------------|------|------|------|
|                                  | A               | EMC  | EAO  | NA   |
| Best-Specified Population        | 6.48            | 5.92 | 6.18 | 4.83 |
| Full/Full                        | 6.53            | 5.93 | 6.18 | 4.80 |
| Full/CODIS                       | 6.49            | 5.79 | 6.13 | 4.75 |
| CODIS/CODIS                      | 6.46            | 5.84 | 6.09 | 4.74 |
| Second-Best-Specified Population | 5.88            | 5.67 | 5.75 | 4.45 |
| Third-Best-Specified Population  | 5.78            | 5.18 | 5.55 | 4.17 |
| Fourth-Best-Specified Population | 5.03            | 4.46 | 4.89 | 4.03 |

■ **Table S4** Standard deviation of  $\bar{D}_{VH}$  for down-sampled analyses. The table gives the standard deviations of distinguishability values across 10 replicate down-sampled analyses whose mean distinguishability values appear in Table S3.

| Assumption                       | True Population |       |       |       |
|----------------------------------|-----------------|-------|-------|-------|
|                                  | A               | EMC   | EAO   | NA    |
| Best-Specified Population        | 0.268           | 0.349 | 0.273 | 0.317 |
| Full/Full                        | 0.252           | 0.350 | 0.252 | 0.309 |
| Full/CODIS                       | 0.215           | 0.316 | 0.235 | 0.205 |
| CODIS/CODIS                      | 0.204           | 0.329 | 0.243 | 0.259 |
| Second-Best-Specified Population | 0.228           | 0.292 | 0.233 | 0.178 |
| Third-Best-Specified Population  | 0.228           | 0.281 | 0.184 | 0.148 |
| Fourth-Best-Specified Population | 0.475           | 0.322 | 0.304 | 0.114 |