TITLE

Decreased accuracy of forensic DNA mixture analysis for groups with lower genetic diversity

AUTHORS AND AFFILIATIONS

Maria Flores 1,2☆; Cara Ly 1☆; Evan Ho 1☆; Niquo Ceberio 1☆; Kamillah Felix 1; Hannah Mariko Thorner 3; Miguel Guardado 4; Matt Paunovich 1; Chris Godek 5; Carina Kalaydjian 5; Rori Rohlfs 1,6*

1) San Francisco State University; Department of Biology; San Francisco, CA, 94132, USA
2) University of California, Los Angeles; Department of Molecular, Cell and Developmental Biology; Los Angeles, CA, 90095, USA
3) George Washington University; Department of Forensic Sciences - Forensic Molecular Biology; Washington, DC, 20007, USA
4) University of California, San Francisco; Biological and Medical Informatics Graduate Program; San Francisco CA, 94143, USA.
5) San Francisco State University; Department of Mathematics; San Francisco, CA, 94132, USA
6) University of Oregon; Department of Data Science; Eugene, OR, 97403, USA

☆These authors contributed equally
* Corresponding author rori@uoregon.edu

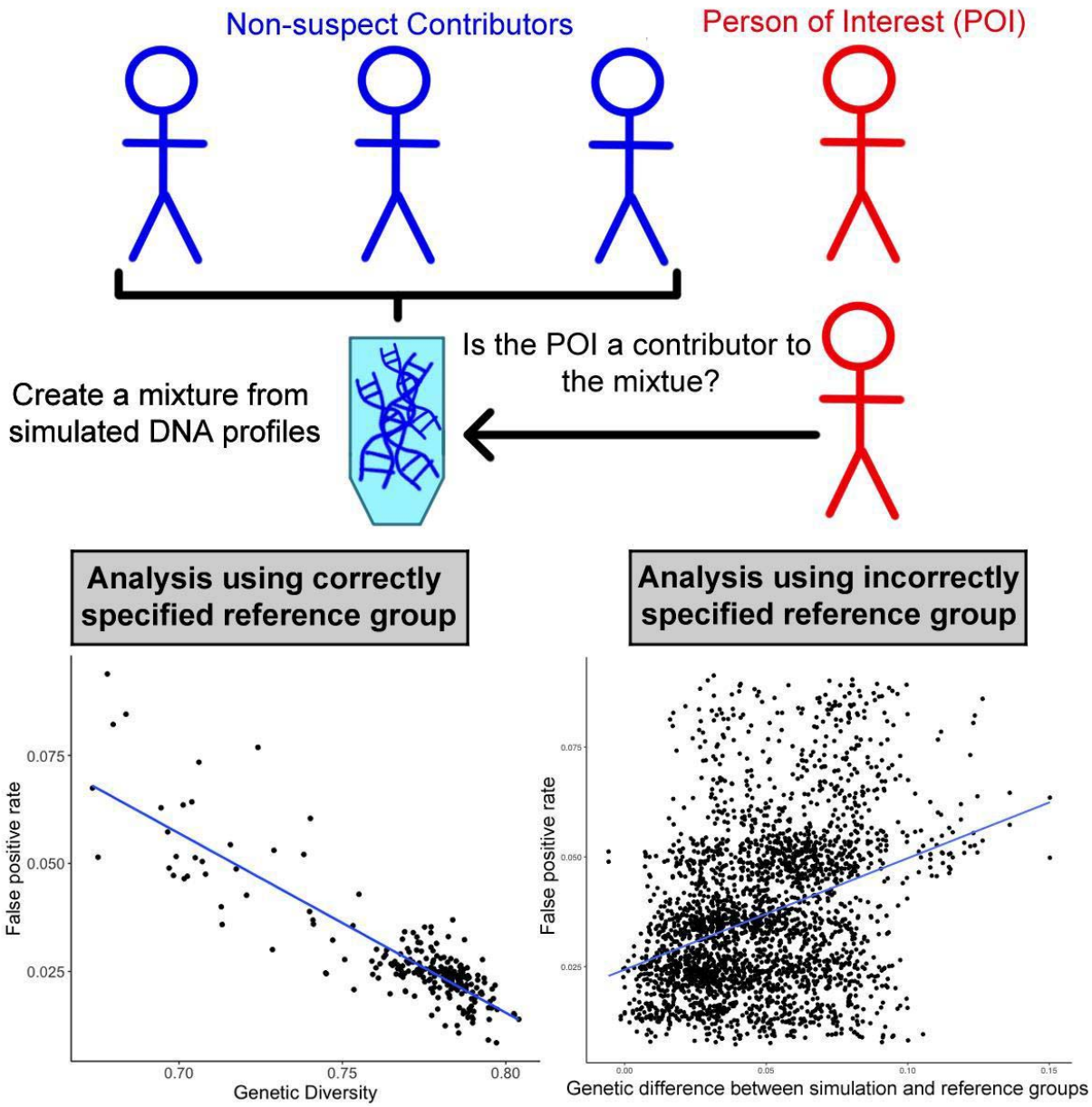
SUMMARY

Forensic investigation of DNA samples from multiple contributors has become commonplace. These complex analyses use statistical frameworks accounting for multiple levels of uncertainty in allelic contributions from different individuals, particularly for samples containing few molecules of DNA. These methods have been thoroughly tested along some axes of variation, but less attention has been paid to accuracy across human genetic variation. Here, we quantify the accuracy of DNA mixture analysis over 244 human groups. We find higher false inclusion rates for mixtures with more contributors, and for groups with lower genetic diversity. Even for two-contributor mixtures where one contributor is known and the reference group is correctly specified, false inclusion rates are 1e-5 or higher for 56 out of 244 groups. This means that, depending on multiple testing, some false inclusions may be expected. These false positives could be lessened with more selective and conservative use of DNA mixture analysis.

HIGHLIGHTS

1. Groups with lower genetic diversity have higher mixture analysis false positive rates.
2. Analyses with mis-specified references have somewhat higher false positive rates.
3. Mixture analysis accuracy decreases with more mixture contributors.

GRAPHICAL ABSTRACT



INTRODUCTION

Routine forensic DNA analysis now investigates DNA mixtures (samples with contributions from multiple individuals), including low-template samples based on

miniscule amounts of biological material[1–3]. For example, we see an increase in the analysis of touch-based DNA evidence, samples with especially few molecules which are difficult to analyze[4,5]. Analysis of DNA mixtures is complicated by overlapping alleles between individuals, unknown numbers of contributors, stutter, allelic drop-out, and allelic drop-in[6,7], and is exacerbated for low-template samples.

A number of statistical methods have been developed to analyze these complicated mixture data, including 'semi-continuous' methods based on the presence or absence of alleles, and 'continuous' methods that consider peak heights and stutter[5]. The output of either class of methods is a likelihood ratio (LR) calculated by dividing the probability of observing the mixed DNA profile under a hypothesis that a person of interest (POI) contributed, versus the probability under a hypothesis that the POI did not contribute[5]. There is significant variability across software packages in the transparency of their methodological approaches, as well as in access to their source code[8]. Studies have shown that with ideal data, semi-continuous methods approximate continuous methods[9].

The results of low-template DNA mixture analyses can have a deciding impact on the outcome of an investigation, therefore accuracy is of the utmost importance. Accordingly, a large number of studies have investigated the accuracy of low-template mixture analysis[5]. Particular attention has been paid to developing approaches for validation[10–12], as well as accuracy with regards to the variation in results and their interpretation across molecular amounts and ratios[13–18], model choice (including number of contributors)[18–21], and laboratories [6,12,22,23].

While these studies have defined a scope of reliable DNA mixture analysis under a set of important parameters, they have not considered variability of accuracy along population genetic diversity. Estimation of the frequency of a forensic DNA profile requires reference allele frequency distributions, which themselves are estimated based on a reference group of individuals. For high-quality single-source forensic samples with full genotyping across CODIS loci, estimated profile frequencies may vary by several orders of magnitude depending on the subject's genetic background and the (possibly mis-specified) reference group, yet the profile frequencies are so miniscule that they still produce decisively informative LRs[24–29]. However, in the more complex context of low-template DNA mixture analysis, variation in contributors' and POI's genetic backgrounds, as well as in the reference groups used to estimate the profile frequencies, may impact the outcome of forensic analyses.

In a different complex forensic technology of familial identification, accuracy is lower for groups with lower genetic diversity, and when allele frequencies are mis-specified[30,31]. In low-template mixture analysis, LR calculations are impacted by the reference allele frequencies, even when the allele frequencies are based on different subsets of the same database[32]. Estimates of the number of contributors are less reliable for groups with lower genetic diversity[33]. Variation in LRs has been observed across broadly defined groups (all with relatively high genetic diversity)[32,34,35]. These results all suggest variability in the accuracy of mixture analysis across genetic backgrounds.

We need to quantify the accuracy of low-template DNA mixture analysis across diverse genetic backgrounds that reflect realistic human genetic variation, and for mixtures with higher numbers of contributors as they are increasingly analyzed in case work[13,36]. Additionally, in casework, the appropriate reference allele frequency distribution is unknown. So, we must also estimate the accuracy of DNA mixture analysis under realistic reference allele frequency misspecification. Typically, analysts perform separate LR calculations with a few reference allele frequency distributions and combine the results in an *ad hoc* procedure[37]. In some cases, none of these reference groups are representative of the genetic background of the contributors to the mixture. Quantification of the accuracy of low-template DNA mixture analysis across population genetic diversity is urgently needed as the technique becomes increasingly ubiquitous, particularly in property crime investigations, raising the probability of coincidental POI inclusions simply due to multiple testing[4,13,36].

In this study, we used the free and open source R package, Forensim[38], to simulate and evaluate DNA mixtures based on allele frequency distributions from diverse groups. With this approach, we quantify the accuracy of DNA mixture analysis across groups, including with mis-specified reference groups.

RESULTS

Genetic diversity varies across groups

Genetic diversity was calculated using allele frequency data obtained from the aggregate study[46] for each group (Figure 1, Figure S1). We observe genetic diversities ranging from 0.67 to 0.80 over the 244 groups (Figure 1). The distribution of genetic diversities for the subset of 54 groups used for mis-specified reference analysis (see section *Accuracy of DNA mixture analysis decreases with mis-specified reference allele frequency distribution*) has the same range (Figure S1).

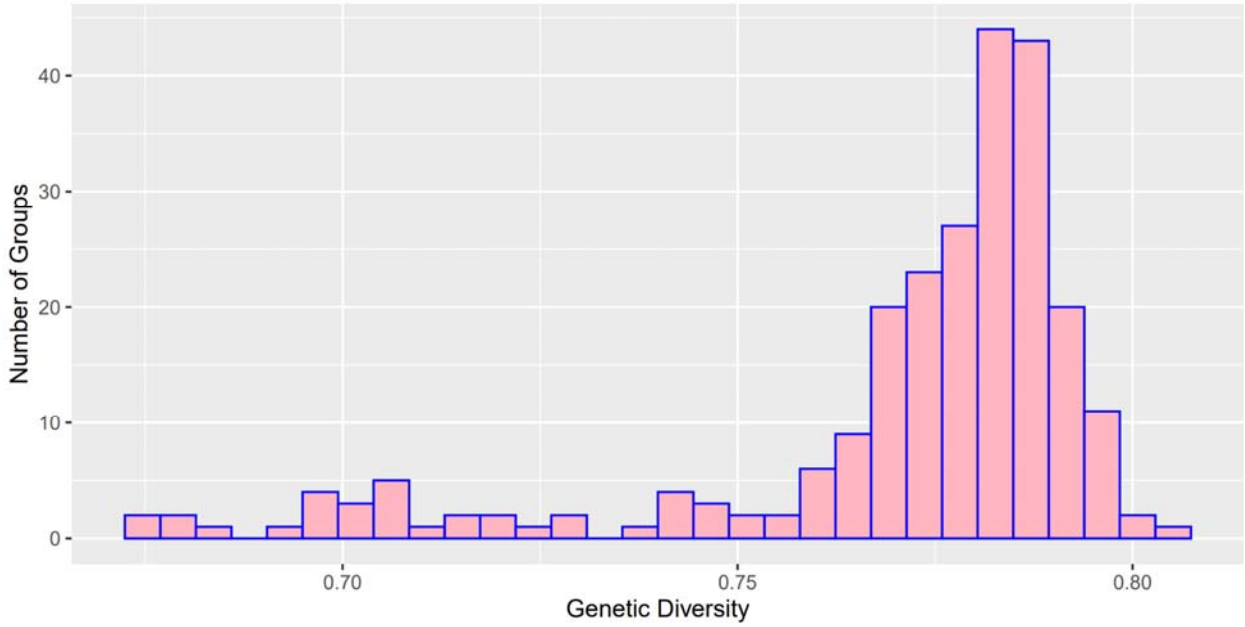


Figure 1: Distribution of genetic diversities over groups. Genetic diversity for the 244 groups analyzed for the accuracy of low-template DNA mixture analysis with the correctly specified reference. See also Figure S1.

Accuracy of DNA mixture analyses decreases with genetic diversity

We estimated the accuracy of determining if a POI contributed to a DNA mixture with 2-6 contributors, over 244 groups, including eight FBI references[39]. In these analyses, the correct group was used as a reference. As expected, false positive rates increase with the number of contributors to mixtures (Figure 2, Figure S2, and Figure S3). False positive rates as high as 0.09 were observed for six-contributor mixtures for a low diversity (0.68) group (Figure 2). Even for mixtures with just two contributors, for 56 out of 244 groups we observe false positive rates above 1e-5 (max of 1.6e-4), including 4 of the 8 FBI reference groups examined (Figure S3).

Higher false positive rates were observed for groups with lower genetic diversity (r^2=-0.73, -0.87, -0.90, -0.90, -0.87 for 2 through 6 contributors respectively, $p<2.2e-16$ for each Pearson correlation test) (Figure 3). In other words, the accuracy of DNA mixture analysis decreases when the contributors to the mixture are from groups with comparatively low genetic diversity.

Our estimations of power were consistently 1.0 across groups and numbers of contributors except for 35 instances where power was estimated at 0.99998 and 0.99999 (Table S1).

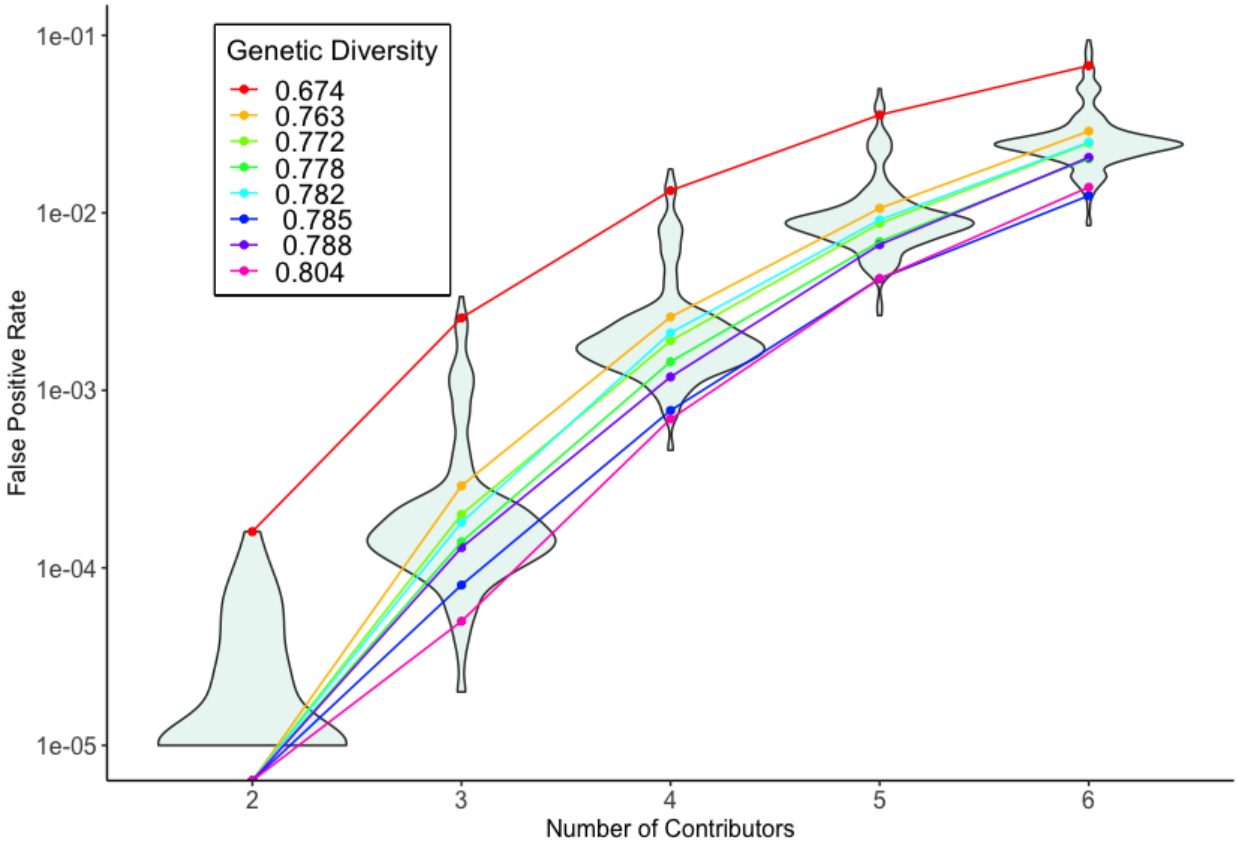


Figure 2: False positive rates when the reference group is correctly specified. This plot displays the distributions of false positive rates for 244 groups over 100,000 simulations per number of contributors to a DNA mixture. False positive rates are shown for some groups, representing quantiles according to genetic diversity. Note that there are 188 groups without false positive observations for two-contributor mixtures. Since the y-axis is log scale, those values are not plotted. See also Figures S2 and S3.

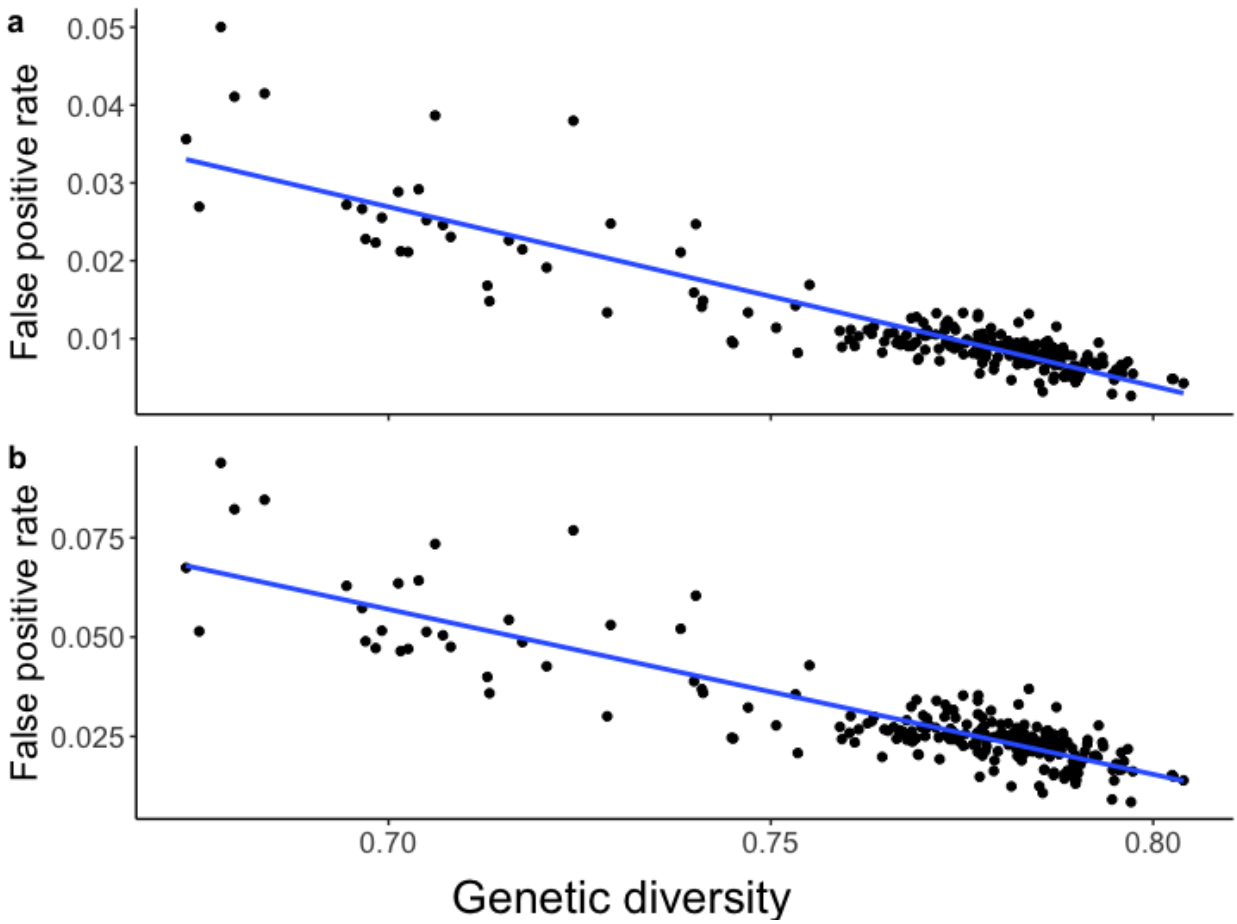


Figure 3: False positive rates versus genetic diversity. Group false positive rates over genetic diversity is shown for mixtures with (a) 5 contributors and (b) 6 contributors. The blue line shows the linear regression between variables.

Accuracy of DNA mixture analysis decreases with mis-specified reference allele frequency distribution

We went on to estimate the accuracy of DNA mixture analysis when the reference group is mis-specified, that is, when the genetic background of the mixture contributors is not the same as the reference group. For computational efficiency, we used a subset of 54 genetically distinct groups (see methods section *Subsetting groups for computational efficiency).*

We find that false positive rates vary with simulation group, such that simulation groups with lower genetic diversity tend to have higher false positive rates, across correct and mis-specified reference groups (Figure 4). The trend of higher false positive rates for groups with lower genetic diversity remains with the standard FBI reference groups (Figure S4). This is particularly noticeable when there are more contributors in the mixture. In addition, simulation groups with lower F_{ST} to each other also tend to have similar false positive rates (Figure S5).

DNA mixture analysis false positive rate varies with correctly and incorrectly specified reference groups as well. False positive rates are correlated with F_{ST} between the simulation and reference groups (r^2 = 0.34, p = 2.2e-16, Pearson correlation test) (Figure 5). That is, false positive rate is higher with increasingly inappropriate reference group. We further see that LRs are elevated when the mis-specified reference group has higher genetic diversity than the simulation group (Figure S6).

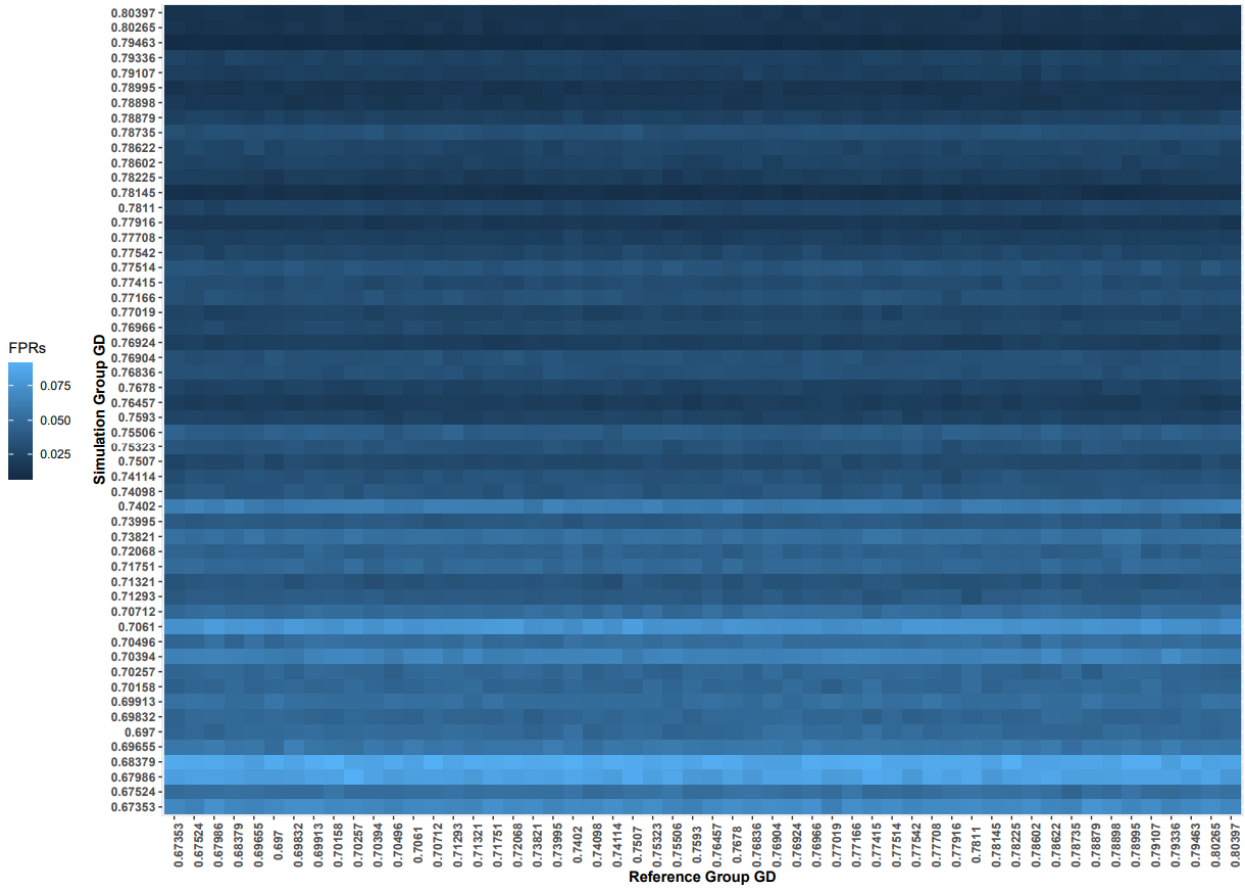


Figure 4: False positive identification rates with correctly and incorrectly specified reference groups. The groups are arranged by genetic diversity, as labeled on the axes. These false positive rates are for analyses of 6-contributor mixtures**.** Note that the diagonal represents correctly specified reference groups. See also Figures S4 and S5.

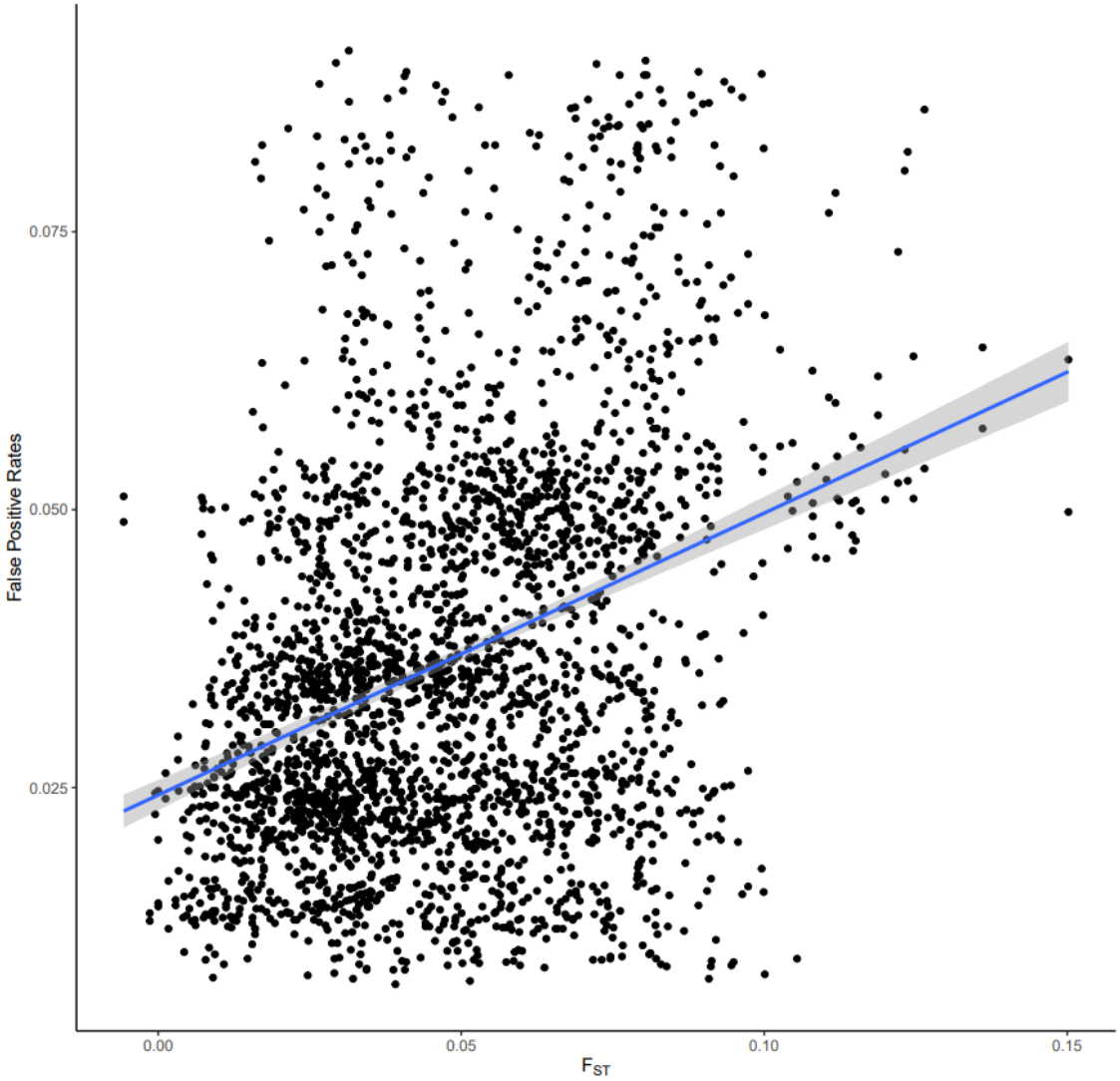


Figure 5: False positive identification rates versus F_{ST} between the simulation and reference groups. False positive rates of 6 contributor mixture analyses for each pair of reference and simulation groups are plotted against the F_{ST} between those groups. Each F_{ST} value has two false positive rates because one group serves as the simulation group while the other the reference group, then vice versa. The blue line is the line of best fit while the shadow represents the 95% confidence interval. See also Figures S6 and S7.

Power remained at 1.0 for all simulations with mis-specified reference groups. However, the distribution of log(LRs) varies for analyses of POI+ mixtures. When the reference group is correctly specified, LRs are lower for groups of lower genetic diversity (Figure S7). However, when the reference is mis-specified, the log(LR)s are inflated, particularly if a mixture of individuals from a low GD group are analyzed assuming a high GD group as reference (Figure S7).

DISCUSSION

Our results show that DNA mixture analysis has elevated false positive rates for groups with lower genetic diversity. False positive rates were 1e-5 or higher for 23% of groups, even for simple two-contributor mixtures where one contributor is known under both the prosecution and defense hypotheses. If the non-POI contributor is unknown, false positive rates are expected to increase. We observe higher false positive rates with the number of mixture contributors. This elevation is sometimes worsened when LRs are computed with an inappropriate reference allele frequency distribution. In practice the most accurate reference group cannot be known, so elevated FPRs are to be expected, particularly for members of groups with low genetic diversity.

The analyses here do not account for allelic drop-out or drop-in, prominent features of low-template DNA mixture analysis. With non-zero rates of drop-out and drop-in, accuracy will decrease[40,41]. These analyses further are based on simulations and evaluation with no co-ancestry (theta=0). Again, accuracy is expected to decrease with realistic non-zero co-ancestry[42]. Finally, in this analysis we only consider scenarios where non-POI contributors were known under the defense and prosecution hypotheses. If non-POI contributors are unknown, we expect accuracy to be lower. These simplifications in our analyses suggest that a more thorough analysis may produce even higher false positive rates.

Further studies of the accuracy of DNA mixture analysis could characterize reliability intersectionally over group genetic variation, as well as differences in mixture composition and data quality. Such studies would be based on broadly inclusive global estimates of allele frequency distributions, gathered with rigorous and consistent procedures for informed consent and data sharing[44].

As more investigations employ low-template DNA mixture analysis, including in cases involving individuals from groups of low genetic diversity, more false positive identifications will occur, potentially leading to wrongful convictions. This warrants consideration of strategies to prevent these false identifications.

A change in technical approach could decrease the chance of false positive identifications. For example, using a higher co-ancestry coefficient in the LR calculation, as suggested by Steele and Balding[34]. Still, further investigation is needed to determine the appropriate parameter value for a broader range of groups and for mixtures with more than two contributors. It may be useful to estimate the appropriate reference group(s) and use that in the LR calculations, as suggested for other forensic identification technologies[43]. Although note that a single reference group will not accurately reflect genetic backgrounds in mixtures where each individual may have a different most-appropriate group. If implemented, these changes in technical approach should be thoroughly examined for their impact on accuracy, particularly when using popularly applied DNA mixture analysis software[8].

A different strategy could be to limit the number of cases in which DNA mixture analysis is used. A selective process could filter out DNA mixtures most prone to analytical errors, for example those with more than three contributors, mixtures where some

individuals contribute very little DNA, or mixtures with high estimated drop-out rates. This could be combined with court presentations of the results including estimates of false positive rates.

LIMITATIONS OF THE STUDY

Our study is sensitive to the categorization schemes used to create groups, including ascertainment biases of what groups were considered, how those groups were defined, and the specific individual-level criteria for inclusion. Because of the opacity and practical limitations of these sampling practices, the resulting published allele frequency datasets imperfectly represent the groups they claim to represent to a certain degree unknowable to us. Thus, we refrain from drawing conclusions about specific social groups based on these results. Since our analysis uses these data, we must be aware that groups defined differently (perhaps by genetic ancestry) would produce different results.

Our study used one particular semi-continuous DNA mixture analysis method. While the trends of the results would likely be similar with other semi-continuous and continuous methods, this study did not specify the specific accuracy with other methods.

ACKNOWLEDGEMENTS

We are deeply indebted to the individuals whose genetic information contributed to publicly available data on which this project is based. We appreciate Dr Carolina Adam's thoughtful comments on the manuscript. This project, particularly R.V.R., M.G., and M.F., was supported by NIJ grant 2019-DU-BX-0028. M.P., M.F., and C.L. were supported by NIH grant T34-GM008574, N.C. was supported by NIH grant R25-GM059298, E.H. and K.F. were supported by the Genentech Foundation Scholars grant #G-7874540. M.G. is a recipient of a Howard Hughes Medical Institute Gilliam Fellowship, Achievement Award for College Scientists Foundation Scholarship, and a UCSF Discovery Fellows Program Award. M.F. is a recipient of the UCLA Eugene V. Cota Robles Fellowship. We gratefully acknowledge that the practices presented in "Equity in author order: a feminist laboratory's approach" were considered for the authorship of this work[45].

AUTHOR CONTRIBUTIONS

Conceptualization, R.R.; Methodology, R.R.; Software, M.F., C.L., E.H., N.C., K.F., M.G., M.P., C.G., C.K., and R.R.; Validation, M.F., C.L., E.H., N.C., K.F., M.G., C.G., and R.R.; Formal Analysis, M.F., C.L., E.H., K.F., C.G., and R.R.; Data Curation, M.F., C.L., E.H., N.C., and H.M.T.; Writing – Original Draft, M.F., E.H., N.C., K.F., H.M.T., and R.R.; Writing – Review & Editing, M.F., E.H., N.C., K.F., H.M.T., M.G., and R.R.; Visualization, E.H., N.C., K.F., and H.M.T.; Supervision, R.R.; Project Administration, R.R.; Funding Acquisition, R.R.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY STATEMENT

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. One or more of the authors of this paper self-identifies as a gender minority in their field of research. One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community. One or more of the authors of this paper self-identifies as living with a disability. One or more of the authors of this paper received support from a program designed to increase minority representation in their field of research. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

METHODS

Resource Availability

Lead Contact
Further information and requests for resources should be directed to and will be fulfilled by the lead contact Rori Rohlfs (rori@uoregon.edu).

Materials Availability
This study did not generate new unique reagents.

Data and Code Availability
This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table. All original code has been deposited at https://github.com/Low-Template-Analysis/LowTemplateAnalysis and is publicly available as of the date of publication. DOIs are listed in the key resources table. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Allele frequencies	This paper; Buckleton, J., *et al*.[46] and Steffen, C.R., *et al*.[39] data	https://github.com/Low-Template-Analysis/LowTemplateAnal

		ysis/tree/main/Allele%20Frequency%20Files 10.1016/j.fsigen.2016.03.004 10.1016/j.fsigen.2017.08.011
Software and algorithms		
R v3.6.1 - *Data Quality Control* R v4.1.3 - *Direct Analysis and Cross Analysis* R v4.2.2 - *Fst Analysis and AEH*	R Core Team[47]	https://www.R-project.org/
Python 3.8.8 -*Data Quality Control*	Python Software Foundation[48]	https://www.python.org/
Forensim	Haned, H.[38]	10.1016/j.fsigen.2010.03.017 https://forensim.r-forge.r-project.org/
Script to calculate average expected heterozygosity (AEH)	This paper	https://github.com/Low-Template-Analysis/LowTemplateAnalysis/tree/main/AEHFolder
Script to calculate the accuracy of mixture analyses with misspecified reference groups	This paper	https://github.com/Low-Template-Analysis/LowTemplateAnalysis/tree/main/CrossAnalysis
Script to calculate the accuracy of mixture analyses with correctly specified reference groups	This paper	https://github.com/Low-Template-Analysis/LowTemplateAnalysis/tree/main/DirectAnalysis
Script to estimate F_{st}	This paper; Rohlfs, R. *et al.*[30]	https://github.com/Low-Template-Analysis/LowTemplateAnalysis/tree/main/FSTFolder
Data Quality Control	This paper	https://github.com/Low-

Scripts		Template-Analysis/LowTemplateAnalysis/tree/main/GroupQualityControlFolder

Allele frequency distribution data

Our primary source for allele frequency distribution data is Buckleton *et al.*'s worldwide survey from 2016, henceforth referred to as "the aggregate study[46]." The aggregate study collected data from 250 source papers that quantified the allele frequency distributions for 466 groups. Across (and even within) source papers, these groups were defined based on heterogenous social non-genetic classification schemes including ethnicity[ex: 49–51], race[ex: 52–54], nationality[ex: 55–57], geography[ex: 58–60], religion[ex: 61], language[ex: 62–64], tribe[ex: 65–67], and caste[ex:68,69]. In many cases, more than one of these frameworks for group identities were described (ex: ethnicity, nationality, and geography). Specificity of sampling location also varied from facility[ex: 64], to city[ex: 70–72], to province/state[ex: 73–75], to country[ex: 76–78]. As a result of heterogeneity in classification schemes and specificity of sampling geography, groups described in the source papers have vastly different scales of resolution. Troublingly, sample sources were also collected from diverse sources including donors who gave informed consent[ex: 79–81], subjects of paternity tests[ex: 82–84], people arrested or convicted[ex: 85–87], collection by law enforcement[ex: 66,88,89], immigration casework[ex: 90,91], and blood banks[ex: 92–94]. The publication of these genetic samples from indirect sources, sometimes without clear informed consent, deserves its own investigation and analysis[95]. Since the data are already publicly available via two sources, in this study we use these accessible data to estimate the accuracy of forensic identification methods already in use.

The source studies defined groups based on divergent social constructs across the sociopolitical contexts in which the researchers operated. It is not clear that the individuals sampled form homogenous groups, nor that they are representative of the social labels used to describe the groups. These socially constructed groups are not distinct genetic entities or genetic ancestry groups[95]. While it is abundantly clear that race is a construct created by humans, rather than intrinsically defined by genetic variation[96–99], human genetic research, relevantly in forensics, often relies on models of humans divided into distinct panmictic groups, or admixtures of these panmictic groups[100–102] these models do not reflect the reality of continuous human genetic variation and sustain the invalid idea that race has a biological basis[103–105]. Misinterpretation of biological distinction between social groups is worsened when scientists construct, label, and investigate groups based on socially-defined characteristics[106–108]. This social method of grouping and labeling is used by the source and aggregate studies in their reports of allele frequency distributions, in many cases without transparent definition of group descriptors[109].

While the reported allele frequency distributions would be different if groups were alternatively defined, they do provide insight into one view of the range of variation in CODIS loci allele frequencies across humans. This perspective is useful as a first

attempt to explore variation in the accuracy of DNA mixture analysis over empirical human genetic variation. Still, because these groups are based on heterogenous social identities, yet are not obviously representative of those social identities, and do not describe genetic ancestry groups, it would be misleading and counter-productive to label them with those social identities for this genetic analysis[101,105,107]. Since previous analyses found that identification accuracy of another forensic technology (familial searching) is strongly correlated with group genetic diversity, in this study we identify groups by their genetic diversity. For transparency, the previously published group labels and their genetic diversities are listed in Table S2.

Data quality control

This analysis uses allele frequency data for the thirteen CODIS short tandem repeat (STR) loci historically used in forensic identification. These data were taken from the previously discussed aggregate study[46], as well as the 2015 FBI expanded dataset[39]. The aggregate study collected data from 250 papers that quantified the allele frequency distributions for 466 groups. With the addition of the FBI dataset, we had 475 distinct groups to consider.

Allele frequency data quality control

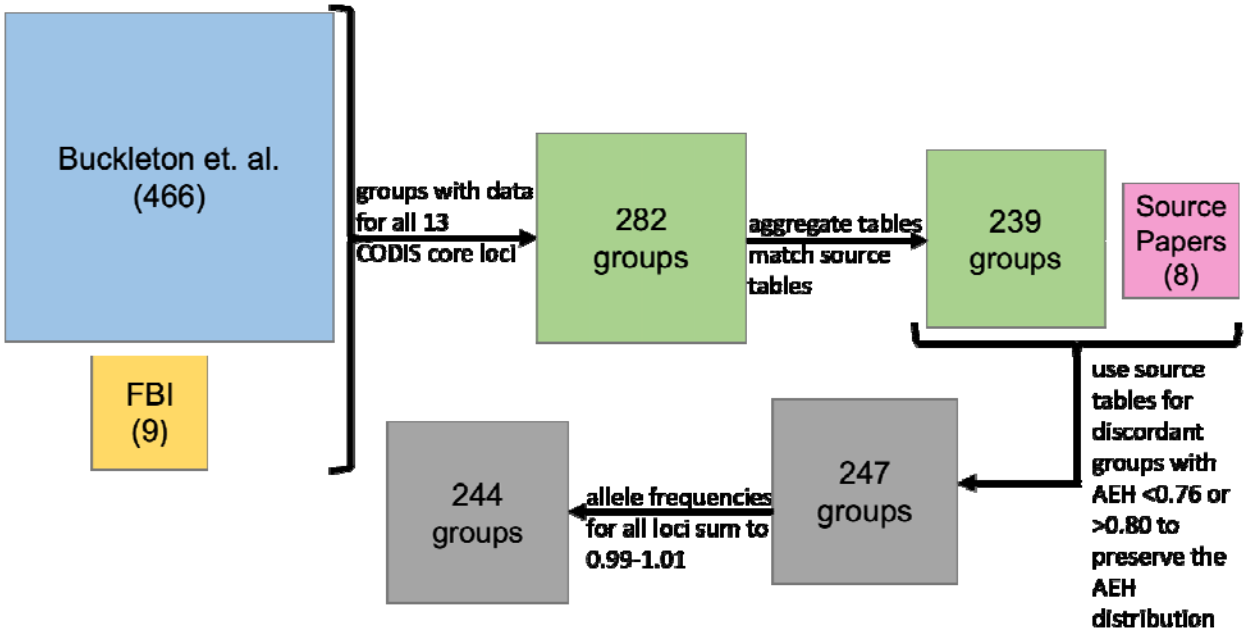


Figure 6: Allele frequency data quality control. Boxes reflect the number of groups retained after each quality control step. The analysis started with 475 groups from the aggregate study and the FBI expanded dataset. 1) The data were filtered for groups with allele frequencies for 13 original core CODIS loci. 2) Because the aggregate study obtained their allele frequency values from other source papers, we only used tables with values that reasonably matched those reported in the source papers. 3) To retain groups representing the full observed range of genetic diversity values, for the groups with unusually high or low genetic diversity values, which were removed by the previous

filter, we used the source paper data for groups with genetic diversity < 0.76 or genetic diversity > 0.80. 4) Finally, we filtered out groups where at least one locus had allele frequencies that did not sum to 1.0 (within .01).

Our quality control procedure (Figure 6) aimed to produce a uniform and reliable dataset across groups including allele frequency data for the 13 core CODIS loci. First, we considered only the groups that had all 13 original core CODIS STRs, leaving us with 282 groups. Second, we cross referenced the allele frequency data reported by the aggregate study[46] to the source papers. We performed the comparison using the 'sdiff' bash function on .csv files containing the allele frequency data reported by the respective papers. We observed discrepancies between the two papers for 71 allele frequency distributions (Table S3). In one case, the cited source for a group did not reasonably match the file listed in the Supplement 2 of the aggregate study[46]. No alternative source could be found, so this group was dropped from further analysis. Additional discrepancies between the aggregate and source allele frequency distributions include: individual frequency value differences (rounding to different degrees of precision, typos, values assigned to different alleles, values assigned to different groups reported by the same source), allele differences (dropped rare alleles, getting rid of alleles with a > or < designation, mislabeled alleles), and loci differences (loci reported in the aggregate table not seen in the source table) (Table S3). We eliminated groups with such discrepancies from our analysis if their genetic diversity values were typical, defined as being between 0.760 and 0.800. Those discordant allele frequency tables for groups with extreme genetic diversity values (less than 0.760 or greater than 0.800) were replaced with the allele frequency tables from the source papers. We did this to preserve data from groups with unusual genetic diversity. This resulted in a total of 247 groups.

The final quality control step checked that the frequencies across alleles for a locus sum between 0.99 and 1.01. This eliminated two groups from the aggregate study[46] as well as one group from the FBI expanded dataset[39]. After all our quality control steps, we have a set of 244 groups in all. The scripts used to evaluate the allele frequency tables and the final data tables employed in this study are posed at https://github.com/Low-Template-Analysis/LowTemplateAnalysis.

Simulation approach to estimate accuracy of DNA mixture analyses for diverse groups

We used Forensim, a free open source R package[38], to first simulate forensic genetic profiles and mixtures, and then to compute likelihood ratios (LRs) to quantify the strength of evidence. These LRs are computed as $LR = \frac{P(G \mid H_p)}{P(G \mid H_d)}$ where G is the genotypes observed, H_p is the prosecution hypothesis that a POI contributed to the DNA mixture, and H_d is the defense hypothesis that a POI did not contribute to the DNA mixture, rather an additional unknown individual contributed[4,8]. When LR evaluates to < 1 it supports the defense hypothesis, and conversely, when it evaluates to > 1, it supports the prosecution hypothesis[38].

Our study examines the accuracy of low template mixture analysis assuming that the genotypes of all non-POI contributors in the simulated mixture are known under both the defense and prosecution hypotheses. While conditions for forensic mixture analysis may not typically be this ideal, it provides us an upper limit for estimating the accuracy.

In our early exploration of Forensim, we noticed that in very rare occasions it would return different LRs given the same inputs. We addressed this by calculating all LRs twice and recalculating if the difference between the two LRs were greater than 0.001.

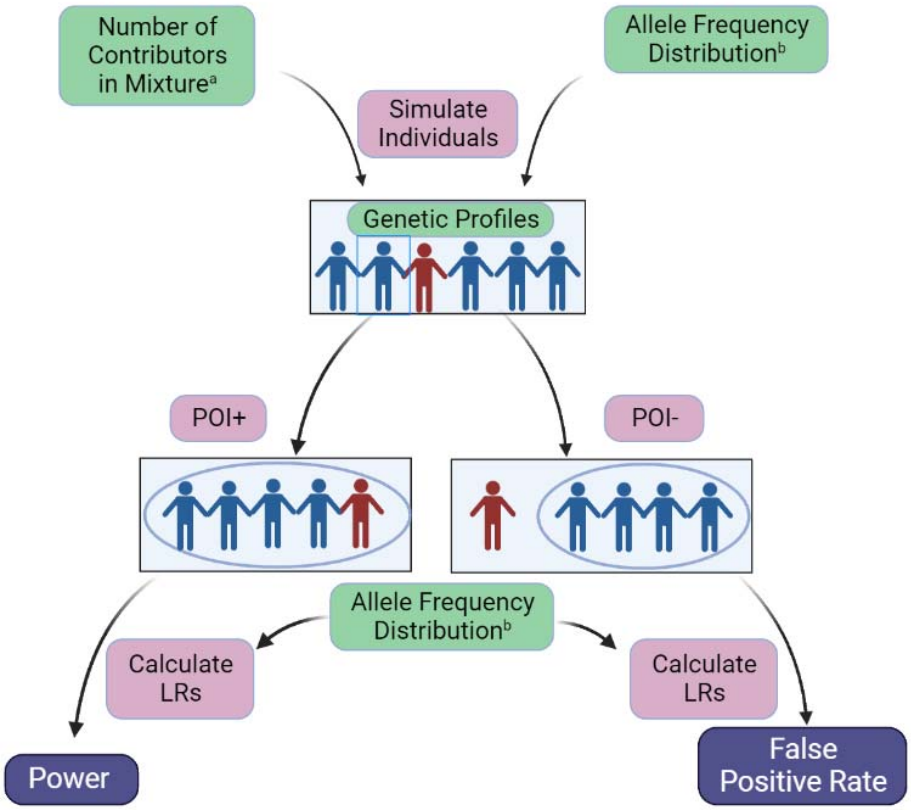


Figure 7: Simulating individuals, DNA mixtures, and calculating likelihood ratios with correctly specified reference groups. This schematic shows how we varied the number of contributors to a mixture from two to six individuals[a], as well as the allele frequency distribution to reflect each of the 244 reference groups[b]. With the resulting genetic profiles, we created two types of simulations: where the person of interest contributed (POI+), and those where the person of interest did not contribute (POI-). We then used the same allele frequency distribution from the simulations to calculate LRs. We analyzed the results from POI+ mixtures to estimate power, and results from POI- mixtures to estimate false positive rate.

We generated the individual DNA profiles and subsequent mixtures using a particular group's allele frequency distribution. Our analysis examined mixtures ranging from two to six contributors; we simulated both when the POI is a contributor (POI+) to estimate power, and when the POI is not a contributor (POI-) to estimate false positive rate. We repeated this analysis for each of the 244 groups. For each group, we simulated

100,000 mixtures, resulting in 1,000,000 LRs generated (two through six contributors, POI+ and POI-, 100,000 mixtures each) (Figure 7). We use a threshold of LR=1 to determine if a mixture is interpreted to contain DNA from the POI.

Simulation approach to estimate accuracy of DNA mixture analyses with mis-specified reference group

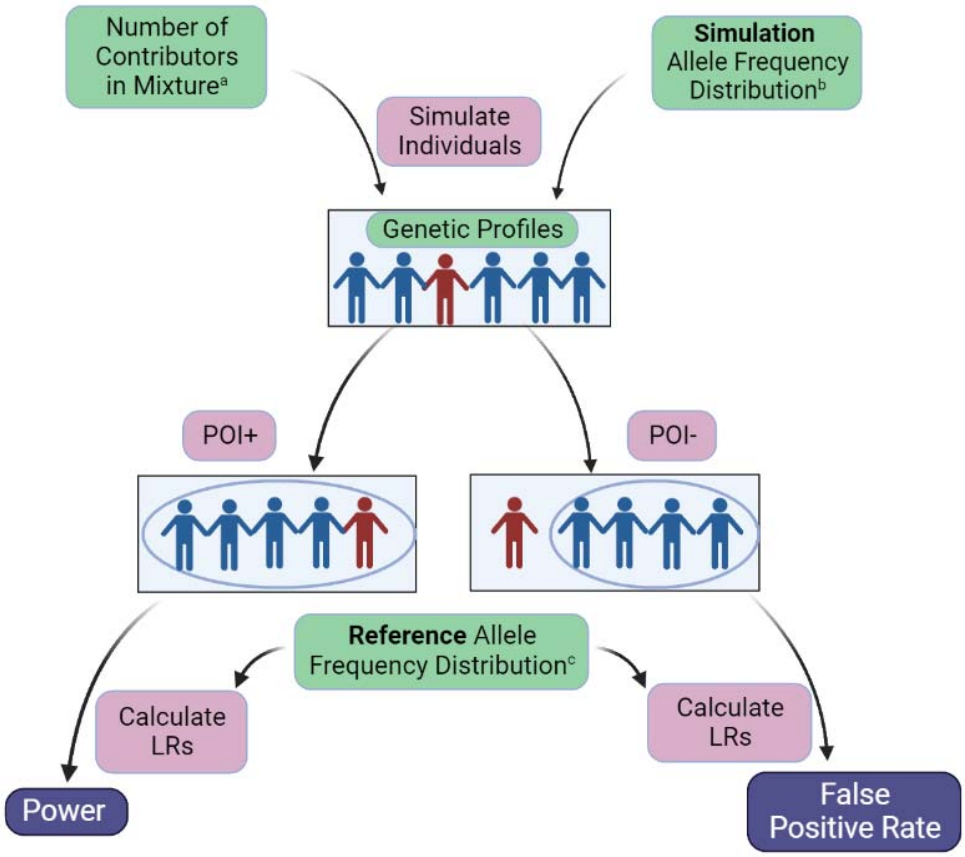


Figure 8: Simulating individuals, DNA mixtures, and calculating likelihood ratios with mis-specified reference groups. This schematic shows how we varied the number of contributors to a mixture from two to six individuals[a], as well as the simulation allele frequency distribution to reflect 54 groups[b]. Again, we created POI + and POI- mixtures. We then used a different reference allele frequency distribution[c] to calculate LRs. We analyzed the results from POI+ mixtures to estimate power, and results from POI- mixtures to estimate false positive rate.

To estimate the accuracy of DNA mixture analysis when an inappropriate reference allele frequency distribution was used, we performed the analysis where one allele frequency distribution is used to simulate individuals (simulation allele frequency distribution), while a different allele frequency distribution is used to compute the LR (the reference allele frequency distribution) (Figure 8). Apart from this, the structure of the analysis remains similar to the previous analysis. For each pair of simulation and reference groups, we varied the number of contributors from two to six individuals. We simulated POI+ and POI- mixtures to estimate power and false positive rate,

respectively. Again, we use a threshold of LR=1 to decide if a mixture is interpreted to contain DNA from the POI. Because this analysis is computationally intensive ($O(n^2)$ with the number of groups), it was done on 54 out of the 244 groups (see subsection "Subsetting groups for computational efficiency"). For each pair of those groups, we simulated 10,000 mixtures with one group as reference and the other for simulations, resulting in 100,000 LRs (two through six contributors, POI+ and POI-, 10,000 mixtures each) per group.

Subsetting groups for computational efficiency

It is computationally infeasible to vary both the simulation and reference groups with a large list of 244 groups. In order to address this, we trimmed down the number of groups used. We identified genetically similar groups based on their F_{ST} and removed one of them. This left us with a set of genetically diverse groups.

Specifically, we calculated F_{ST} between each pair of groups using the method by Weir and Cockerham[110], which accounts for groups of different sizes, and across multiple alleles and loci. We then iteratively identified the pair of groups with the lowest F_{ST} and randomly removed one of those two groups, repeating until all F_{ST} values exceed a threshold. For this analysis, we used a threshold F_{ST} value of 0.005.

In the end, 54 groups were left encompassing a genetic diversity range of 0.67 to 0.80 (Figure S1). Note that the FBI expanded groups[39] and the groups reported by Hill *et al.*[93] were not considered for filtering based on F_{ST} value because their allele frequency distributions are widely used by the field[39,93].

REFERENCES

1. Steele, C.D., and Balding, D.J. (2014). Statistical Evaluation of Forensic DNA Profile Evidence. Annu. Rev. Stat. Its Appl. *1*, 361–384. 10.1146/annurev-statistics-022513-115602.
2. Benschop, C.C.G., Haned, H., de Blaeij, T.J.P., Meulenbroek, A.J., and Sijen, T. (2012). Assessment of mock cases involving complex low template DNA mixtures: A descriptive study. Forensic Sci. Int. Genet. *6*, 697–707. 10.1016/j.fsigen.2012.04.007.
3. Bieber, F.R., Buckleton, J.S., Budowle, B., Butler, J.M., and Coble, M.D. (2016). Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion. BMC Genet. *17*, 125. 10.1186/s12863-016-0429-7.
4. Cavanaugh, S.E., and Bathrick, A.S. (2018). Direct PCR amplification of forensic touch and other challenging DNA samples: A review. Forensic Sci. Int. Genet. *32*, 40–49. 10.1016/j.fsigen.2017.10.005.
5. Coble, M.D., and Bright, J.-A. (2019). Probabilistic genotyping software: An overview. Forensic Sci. Int. Genet. *38*, 219–224. 10.1016/j.fsigen.2018.11.009.
6. Emily Rogers, Roman Aranda IV, Philippa M. Spencer, and Denise R. Myers (2022). DNA Mixture Study: Novel Metrics to Quantify the Intra- and Inter-Laboratory Variability in Forensic DNA Mixture Interpretation.

7. Gill, P., Haned, H., Bleka, O., Hansson, O., Dørum, G., and Egeland, T. (2015). Genotyping and interpretation of STR-DNA: Low-template, mixtures and database matches—Twenty years of research and development. Forensic Sci. Int. Genet. *18*, 100–117. 10.1016/j.fsigen.2015.03.014.
8. Edge, M.D., and Matthews, J.N. (2022). Open practices in our science and our courtrooms. Trends Genet. *38*, 113–115. 10.1016/j.tig.2021.09.010.
9. Balding, D.J. (2013). Evaluation of mixed-source, low-template DNA profiles in forensic science. Proc. Natl. Acad. Sci. *110*, 12241–12246. 10.1073/pnas.1219739110.
10. Haned, H., Gill, P., Lohmueller, K., Inman, K., and Rudin, N. (2016). Validation of probabilistic genotyping software for use in forensic DNA casework: Definitions and illustrations. Sci. Justice *56*, 104–108. 10.1016/j.scijus.2015.11.007.
11. Coble, M.D., Buckleton, J., Butler, J.M., Egeland, T., Fimmers, R., Gill, P., Gusmão, L., Guttman, B., Krawczak, M., Morling, N., et al. (2016). DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications. Forensic Sci. Int. Genet. *25*, 191–197. 10.1016/j.fsigen.2016.09.002.
12. The Scientific Working Group on DNA Analysis and Methods (SWGDAM) (2015). Scientific Working Group on DNA Analysis Methods Guidelines for the Validation of Probabilistic Genotyping Systems. 12.
13. Haned, H., Benschop, C.C.G., Gill, P.D., and Sijen, T. (2015). Complex DNA mixture analysis in a forensic context: Evaluating the probative value using a likelihood ratio model. Forensic Sci. Int. Genet. *16*, 17–25. 10.1016/j.fsigen.2014.11.014.
14. Bright, J.-A., Evett, I.W., Taylor, D., Curran, J.M., and Buckleton, J. (2015). A series of recommended tests when validating probabilistic DNA profile interpretation software. Forensic Sci. Int. Genet. *14*, 125–131. 10.1016/j.fsigen.2014.09.019.
15. You, Y., and Balding, D. (2019). A comparison of software for the evaluation of complex DNA profiles. Forensic Sci. Int. Genet. *40*, 114–119. 10.1016/j.fsigen.2019.02.014.
16. Taylor, D., Buckleton, J., and Evett, I. (2015). Testing likelihood ratios produced from complex DNA profiles. Forensic Sci. Int. Genet. *16*, 165–171. 10.1016/j.fsigen.2015.01.008.
17. Marsden, C.D., Rudin, N., Inman, K., and Lohmueller, K.E. (2016). An assessment of the information content of likelihood ratios derived from complex mixtures. Forensic Sci. Int. Genet. *22*, 64–72. 10.1016/j.fsigen.2016.01.008.
18. Bleka, Ø., Benschop, C.C.G., Storvik, G., and Gill, P. (2016). A comparative study of qualitative and quantitative models used to interpret complex STR DNA profiles. Forensic Sci. Int. Genet. *25*, 85–96. 10.1016/j.fsigen.2016.07.016.
19. Swaminathan, H., Qureshi, M.O., Grgicak, C.M., Duffy, K., and Lun, D.S. (2018). Four model variants within a continuous forensic DNA mixture interpretation framework: Effects on evidential inference and reporting. PLOS ONE *13*, e0207599. 10.1371/journal.pone.0207599.
20. Norsworthy, S., Lun, D.S., and Grgicak, C.M. (2018). Determining the number of contributors to DNA mixtures in the low-template regime: Exploring the impacts of

sampling and detection effects. Leg. Med. *32*, 1–8. 10.1016/j.legalmed.2018.02.001.
21. Alladio, E., Omedei, M., Cisana, S., D'Amico, G., Caneparo, D., Vincenti, M., and Garofano, P. (2018). DNA mixtures interpretation – A proof-of-concept multi-software comparison highlighting different probabilistic methods' performances on challenging samples. Forensic Sci. Int. Genet. *37*, 143–150. 10.1016/j.fsigen.2018.08.002.
22. Bright, J.-A., Richards, R., Kruijver, M., Kelly, H., McGovern, C., Magee, A., McWhorter, A., Ciecko, A., Peck, B., Baumgartner, C., et al. (2018). Internal validation of STRmix™ – A multi laboratory response to PCAST. Forensic Sci. Int. Genet. *34*, 11–24. 10.1016/j.fsigen.2018.01.003.
23. Butler, J.M., Kline, M.C., and Coble, M.D. (2018). NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): Variation observed and lessons learned. Forensic Sci. Int. Genet. *37*, 81–94. 10.1016/j.fsigen.2018.07.024.
24. Gill, P., and Evett, I. (1995). Population genetics of short tandem repeat (STR) loci. In Human Identification: The Use of DNA Markers Contemporary Issues in Genetics and Evolution., B. S. Weir, ed. (Springer Netherlands), pp. 69–87. 10.1007/978-0-306-46851-3_9.
25. Evett, I.W., Gill, P.D., Lambert, J.A., Oldroyd, N., Frazier, R., Watson, S., Panchal, S., Connolly, A., and Kimpton, C. (1997). Statistical analysis of data for three British ethnic groups from a new STR multiplex. Int. J. Legal Med. *110*, 5–9. 10.1007/BF02441017.
26. Evett, I.W., Gill, P.D., Scrange, J.K., and Weir, B.S. (1996). Establishing the robustness of short-tandem-repeat statistics for forensic applications. Am. J. Hum. Genet. *58*, 398–407.
27. Balding, D.J., and Nichols, R.A. (1997). Significant genetic correlations among Caucasians at forensic DNA loci. Heredity *78*, 583–589. 10.1038/hdy.1997.97.
28. Curran, J.M., Walsh, S.J., and Buckleton, J. (2007). Empirical testing of estimated DNA frequencies. Forensic Sci. Int. Genet. *1*, 267–272. 10.1016/j.fsigen.2007.06.004.
29. Weir, B.S. (2007). The rarity of DNA profiles. Ann. Appl. Stat. *1*. 10.1214/07-AOAS128.
30. Rohlfs, R.V., Fullerton, S.M., and Weir, B.S. (2012). Familial Identification: Population Structure and Relationship Distinguishability. PLoS Genet. *8*, e1002469. 10.1371/journal.pgen.1002469.
31. Rohlfs, R.V., Murphy, E., Song, Y.S., and Slatkin, M. (2013). The Influence of Relatives on the Efficiency and Error Rate of Familial Searching. PLoS ONE *8*, e70495. 10.1371/journal.pone.0070495.
32. Bright, J.-A., Stevenson, K.E., Curran, J.M., and Buckleton, J.S. (2015). The variability in likelihood ratios due to different mechanisms. Forensic Sci. Int. Genet. *14*, 187–190. 10.1016/j.fsigen.2014.10.013.
33. Coble, M.D., Bright, J.-A., Buckleton, J.S., and Curran, J.M. (2015). Uncertainty in the number of contributors in the proposed new CODIS set. Forensic Sci. Int. Genet. *19*, 207–211. 10.1016/j.fsigen.2015.07.005.
34. Steele, C.D., and Balding, D.J. (2014). Choice of population database for forensic DNA profile analysis. Sci. Justice *54*, 487–493. 10.1016/j.scijus.2014.10.004.
35. Adamowicz, M.S., Rambo, T.N., and Clarke, J.L. (2022). Internal Validation of MaSTR™ Probabilistic Genotyping Software for the Interpretation of 2–5 Person

Mixed DNA Profiles. Genes *13*, 1429. 10.3390/genes13081429.
36. Haddrill, P.R. (2021). Developments in forensic DNA analysis. Emerg. Top. Life Sci. *5*, 381–393. 10.1042/ETLS20200304.
37. Kelly, H., Bright, J., Coble, M.D., and Buckleton, J.S. (2020). A description of the likelihood ratios in the probabilistic genotyping software STRmix ™. WIREs Forensic Sci. *2*. 10.1002/wfs2.1377.
38. Haned, H. (2011). Forensim: An open-source initiative for the evaluation of statistical methods in forensic genetics. Forensic Sci. Int. Genet. *5*, 265–268. 10.1016/j.fsigen.2010.03.017.
39. Steffen, C.R., Coble, M.D., Gettings, K.B., and Vallone, P.M. (2017). Corrigendum to 'U.S. Population Data for 29 Autosomal STR Loci' [Forensic Sci. Int. Genet. 7 (2013) e82–e83]. Forensic Sci. Int. Genet. *31*, e36–e40. 10.1016/j.fsigen.2017.08.011.
40. Kalafut, T., Pugh, S., Gill, P., Abbas, S., Semaan, M., Mansour, I., Curran, J., Bright, J., Hicks, T., Wivell, R., et al. (2022). A mixed DNA profile controversy revisited. J. Forensic Sci. *67*, 128–135. 10.1111/1556-4029.14912.
41. Manabe, S., Morimoto, C., Hamano, Y., Fujimoto, S., and Tamaki, K. (2017). Development and validation of open-source software for DNA mixture interpretation based on a quantitative continuous model. PLOS ONE *12*, e0188183. 10.1371/journal.pone.0188183.
42. Kruijver, M., Kelly, H., Bright, J.-A., and Buckleton, J. (2022). Evaluating DNA Mixtures with Contributors from Different Populations Using Probabilistic Genotyping. Genes *14*, 40. 10.3390/genes14010040.
43. Fortier, A.L., Kim, J., and Rosenberg, N.A. (2020). Human-Genetic Ancestry Inference and False Positives in Forensic Familial Searching. G3 GenesGenomesGenetics *10*, 2893–2902. 10.1534/g3.120.401473.
44. Wallace, S.E., Kirby, E., and Knoppers, B.M. (2020). How Can We Not Waste Legacy Genomic Research Data? Front. Genet. *11*, 446. 10.3389/fgene.2020.00446.
45. Liboiron, M., Ammendolia, J., Winsor, K., Zahara, A., Bradshaw, H., Melvin, J., Mather, C., Dawe, N., Wells, E., Liboiron, F., et al. (2017). Equity in Author Order: A Feminist Laboratory's Approach. Catal. Fem. Theory Technoscience *3*, 1–17. 10.28968/cftt.v3i2.28850.
46. Buckleton, J., Curran, J., Goudet, J., Taylor, D., Thiery, A., and Weir, B.S. (2016). Population-specific F values for forensic STR markers: A worldwide survey. Forensic Sci. Int. Genet. *23*, 91–100. 10.1016/j.fsigen.2016.03.004.
47. R Core Team R: A language and environment for statistical computing.
48. Python Software Foundation Python Language Reference.
49. Deng, Y., Zhu, B., Yu, X., Li, Y., Fang, J., Xiong, X., Mu, H., Huang, Y., and Shi, X. (2007). Genetic polymorphisms of 15 STR loci of Chinese Dongxiang and Salar ethnic minority living in Qinghai Province of China. Leg. Med. *9*, 38–42. 10.1016/j.legalmed.2006.08.003.
50. Pérez-Miranda, A.M., Alfonso-Sánchez, M.A., Kalantar, A., Peña, J.A., Pancorbo, M.M. de, and Herrera, R.J. (2005). Allelic frequencies of 13 STR loci in autochthonous Basques from the province of Vizcaya (Spain). Forensic Sci. Int. *152*, 259–262. 10.1016/j.forsciint.2004.09.118.
51. Calzada, P., Suárez, I., García, S., Barrot, C., Sánchez, C., Ortega, M., Mas, J.,

Huguet, E., Corbella, J., and Gené, M. (2005). The Fang population of Equatorial Guinea characterised by 15 STR-PCR polymorphisms. Int. J. Legal Med. *119*, 107–110. 10.1007/s00414-004-0505-9.

52. Gross, A.M., and Budowle, B. (2006). Minnesota Population Data on 15 STR Loci Using the IdentifilerTM Kit. J. Forensic Sci. *51*, 1410–1413. 10.1111/j.1556-4029.2006.00280.x.
53. Bright, J.-A., Buckleton, J.S., and McGovern, C.E. (2010). Allele frequencies for the four major sub-populations of New Zealand for the 15 Identifiler loci. Forensic Sci. Int. Genet. *4*, e65–e66. 10.1016/j.fsigen.2009.05.008.
54. Lucassen, A., Ehlers, K., Grobler, P.J., and Shezi, A.L. (2014). Allele frequency data of 15 autosomal STR loci in four major population groups of South Africa. Int. J. Legal Med. *128*, 275–276. 10.1007/s00414-013-0898-4.
55. Melo, M.M., Carvalho, M., Lopes, V., Anjos, M.J., Serra, A., Vieira, D.N., Sequeiros, J., and Corte-Real, F. (2010). Genetic study of 15 STRs loci of Identifiler system in Angola population. Forensic Sci. Int. Genet. *4*, e153–e157. 10.1016/j.fsigen.2010.03.010.
56. Turrina, S., Ferrian, M., Caratti, S., and De Leo, D. (2014). Evaluation of genetic parameters of 22 autosomal STR loci (PowerPlex® Fusion System) in a population sample from Northern Italy. Int. J. Legal Med. *128*, 281–283. 10.1007/s00414-013-0934-4.
57. Gutiérrez, C.C., Loáisiga, J.C., Campos, L.G., Monterrosa, J.C., Morales, J., Inda, F.J., Uriarte, I., Yurrebaso, I., and García, O. (2011). Population genetic data for 15 STR loci (PowerPlex 16 kit) in Nicaragua. Forensic Sci. Int. Genet. *5*, 563–564. 10.1016/j.fsigen.2010.08.011.
58. Alshamali, F., Alkhayat, A.Q., Budowle, B., and Watson, N.D. (2005). STR population diversity in nine ethnic populations living in Dubai. Forensic Sci. Int. *152*, 267–279. 10.1016/j.forsciint.2004.09.133.
59. Rubi-Castellanos, R., Anaya-Palafox, M., Mena-Rojas, E., Bautista-España, D., Muñoz-Valle, J.F., and Rangel-Villalobos, H. (2009). Genetic data of 15 autosomal STRs (Identifiler kit) of three Mexican Mestizo population samples from the States of Jalisco (West), Puebla (Center), and Yucatan (Southeast). Forensic Sci. Int. Genet. *3*, e71–e76. 10.1016/j.fsigen.2008.07.006.
60. Pepinski, W., Niemcunowicz-Janica, A., Skawronska, M., Koc-Zorawska, E., Janica, J., and Soltyszewski, I. (2004). Allele distribution of 15 STR loci in a population sample of Byelorussian minority residing in the northeastern Poland. Forensic Sci. Int. *139*, 265–267. 10.1016/j.forsciint.2003.11.013.
61. Coudray, C., Guitard, E., El-Chennawi, F., Larrouy, G., and Dugoujon, J.-M. (2007). Allele frequencies of 15 short tandem repeats (STRs) in three Egyptian populations of different ethnic groups. Forensic Sci. Int. *169*, 260–265. 10.1016/j.forsciint.2006.03.017.
62. Gomes, V., Sánchez-Diz, P., Alves, C., Gomes, I., Amorim, A., Carracedo, Á., and Gusmão, L. (2009). Population data defined by 15 autosomal STR loci in Karamoja population (Uganda) using AmpF/STR Identifiler kit. Forensic Sci. Int. Genet. *3*, e55–e58. 10.1016/j.fsigen.2008.06.005.
63. Coudray, C., Guitard, E., Keyser-Tracqui, C., Melhaoui, M., Cherkaoui, M., Larrouy, G., and Dugoujon, J.-M. (2007). Population genetic data of 15 tetrameric

short tandem repeats (STRs) in Berbers from Morocco. Forensic Sci. Int. *167*, 81–86. 10.1016/j.forsciint.2006.01.003.

64. Tofanelli, S., Boschi, I., Bertoneri, S., Coia, V., Taglioli, L., Franceschi, M.G., Destro-Bisol, G., Pascali, V., and Paoli, G. (2003). Variation at 16 STR loci in Rwandans (Hutu) and implications on profile frequency estimation in Bantu-speakers. Int. J. Legal Med. *117*, 121–126. 10.1007/s00414-002-0349-0.
65. Di Cristofaro, J., Buhler, S., Temori, S.A., and Chiaroni, J. (2012). Genetic data of 15 STR loci in five populations from Afghanistan. Forensic Sci. Int. Genet. *6*, e44–e45. 10.1016/j.fsigen.2011.03.004.
66. Budowle, B., Shea, B., Niezgoda, S., and Chakraborty, R. (2001). CODIS STR loci data from 41 sample populations. J. Forensic Sci. *46*, 453–489.
67. Krithika, S., Trivedi, R., Kashyap, V.K., and Vasulu, T.S. (2007). Genotype Profile for Fifteen Tetranucleotide Repeat Loci in Two Tibeto-Burman Speaking Tribal Populations of Arunachal Pradesh, India. J. Forensic Sci. *52*, 239–241. 10.1111/j.1556-4029.2006.00347.x.
68. Ashma, R., and Kashyap, V.K. (2002). Genetic polymorphism at 15 STR loci among three important subpopulation of Bihar, India. Forensic Sci. Int. *130*, 58–62. 10.1016/S0379-0738(02)00346-8.
69. Sarkar, N., and Kashyap, V.K. (2002). Genetic diversity at two pentanucleotide STR and thirteen tetranucleotide STR loci by multiplex PCR in four predominant population groups of central India. Forensic Sci. Int. *128*, 196–201. 10.1016/S0379-0738(02)00194-9.
70. Egyed, B., Füredi, S., Angyal, M., Balogh, I., Kalmar, L., and Padar, Z. (2006). Analysis of the population heterogeneity in Hungary using fifteen forensically informative STR markers. Forensic Sci. Int. *158*, 244–249. 10.1016/j.forsciint.2005.07.004.
71. Dobashi, Y., Kido, A., Fujitani, N., Hara, M., Susukida, R., and Oya, M. (2005). STR data for the AmpFLSTR Identifiler loci in Bangladeshi and Indonesian populations. Leg. Med. *7*, 222–226. 10.1016/j.legalmed.2005.04.001.
72. Alves, C., Gusmão, L., López-Parra, A.M., Mesa, M.S., Amorim, A., and Arroyo-Pardo, E. (2005). STR allelic frequencies for an African population sample (Equatorial Guinea) using AmpFlSTR Identifiler and Powerplex 16 kits. Forensic Sci. Int. *148*, 239–242. 10.1016/j.forsciint.2004.05.007.
73. Yang, B., Wang, G., Liu, Y., and Yang, W. (2005). Population data for the AmpF STR® Identifiler™ PCR Amplification Kit in China Han in Jilin Province, China. Forensic Sci. Int. *151*, 293–297. 10.1016/j.forsciint.2005.02.001.
74. Marino, M., Sala, A., and Corach, D. (2006). Population genetic analysis of 15 autosomal STRs loci in the central region of Argentina. Forensic Sci. Int. *161*, 72–77. 10.1016/j.forsciint.2005.08.014.
75. Poiares, L. de A., Osorio, P. de S., Spanhol, F.A., Coltre, S.C., Rodenbusch, R., Branco, C.C., Pacheco, P.R., Mota-Vieira, L., Largura, A., Sandrini, F., et al. (2009). 15 STR loci frequencies in the population from Paraná, Southern Brazil. Forensic Sci. Int. Genet. *4*, e23–e24. 10.1016/j.fsigen.2009.01.005.
76. Sánchez-Diz, P., Acosta, M.A., Fonseca, D., Fernández, M., Gómez, Y., Jay, M., Alape, J., Lareu, M.V., Carracedo, A., and Restrepo, C.M. (2009). Population data on 15 autosomal STRs in a sample from Colombia. Forensic Sci. Int. Genet. *3*, e81–e82.

10.1016/j.fsigen.2008.08.002.
77. Park, J.-H., Hong, S.-B., Kim, J.-Y., Chong, Y., Han, S., Jeon, C.-H., and Ahn, H.-J. (2013). Genetic variation of 23 autosomal STR loci in Korean population. Forensic Sci. Int. Genet. *7*, e76–e77. 10.1016/j.fsigen.2012.10.005.
78. Westen, A.A., Kraaijenbrink, T., Robles de Medina, E.A., Harteveld, J., Willemse, P., Zuniga, S.B., van der Gaag, K.J., Weiler, N.E.C., Warnaar, J., Kayser, M., et al. (2014). Comparing six commercial autosomal STR kits in a large Dutch population sample. Forensic Sci. Int. Genet. *10*, 55–63. 10.1016/j.fsigen.2014.01.008.
79. Gonçalves, R., Jesus, J., Fernandes, A.T., and Brehm, A. (2002). Genetic profile of a multi-ethnic population from Guiné-Bissau (west African coast) using the new PowerPlex® 16 System kit. Forensic Sci. Int. *129*, 78–80. 10.1016/S0379-0738(02)00204-9.
80. Fridman, C., dos Santos, P.C.C., Kohler, P., Garcia, C.F., Lopez, L.F., Massad, E., and Gattás, G.J.F. (2008). Brazilian population profile of 15 STR markers. Forensic Sci. Int. Genet. *2*, e1–e4. 10.1016/j.fsigen.2007.09.003.
81. Xing, J., Xuan, J., Wang, B., Pang, H., Li, C., and Ding, M. (2011). Genetic polymorphism of 15 STR loci in a Manchu population in Northeast China. Forensic Sci. Int. Genet. *5*, e93–e95. 10.1016/j.fsigen.2010.12.006.
82. Decorte, R., Engelen, M., Larno, L., Nelissen, K., Gilissen, A., and Cassiman, J.-J. (2004). Belgian population data for 15 STR loci (AmpFlSTR® SGM Plus and AmpFlSTR™ profiler PCR amplification kit). Forensic Sci. Int. *139*, 211–213. 10.1016/j.forsciint.2003.11.007.
83. Borosky, A., Toscanini, U., Gómez, A., Parolín, M.L., Basso, N., and Vullo, C. (2014). Forensic population data for 20 STR loci in Argentina. Forensic Sci. Int. Genet. *13*, e20–e21. 10.1016/j.fsigen.2014.07.008.
84. Cassar, M., Farrugia, C., and Vidal, C. (2008). Allele frequencies of 14 STR loci in the population of Malta. Leg. Med. *10*, 153–156. 10.1016/j.legalmed.2007.09.004.
85. Stanciu, F., Popescu, O.R., and Stoian, I.M. (2009). Allele frequencies of 15 STR loci in Moldavia region (NE Romania). Forensic Sci. Int. Genet. *4*, e39–e40. 10.1016/j.fsigen.2009.02.008.
86. Ali, I., Rasheed, A., Saleem, M.M., Saeed, A., Rasheed, H., Nasih, A., Eeman, M., and Sameer, H. (2012). Allele frequencies of 15 STR loci using AmpF/STR Identifiler kit in the Maldivian population. Forensic Sci. Int. Genet. *6*, e136. 10.1016/j.fsigen.2012.01.007.
87. Budowle, B., Chidambaram, A., Strickland, L., Beheim, C.W., Taft, G.M., and Chakraborty, R. (2002). Population studies on three Native Alaska population groups using STR loci. Forensic Sci. Int. *129*, 51–57. 10.1016/S0379-0738(02)00209-8.
88. Eckhoff, C., Walsh, S.J., and Buckleton, J.S. (2007). Population data from sub-populations of the Northern Territory of Australia for 15 autosomal short tandem repeat (STR) loci. Forensic Sci. Int. *171*, 237–249. 10.1016/j.forsciint.2006.06.009.
89. Sanz, P., Prieto, V., Flores, I., Torres, Y., López-Soto, M., and Farfán, M.J. (2001). Population data of 13 STRS in southern Spain (Andalusia). Forensic Sci. Int. *119*, 113–115. 10.1016/S0379-0738(00)00422-9.
90. Branham, A., Wenk, R., and Chiafari, F. (2012). Allele frequencies of fifteen STR loci in U.S. immigrants from Haiti compared with African Americans and Afro-Caribbeans. Forensic Sci. Int. Genet. *6*, e3–e4. 10.1016/j.fsigen.2010.11.001.

91. Tillmar, A.O., Bäckström, G., and Montelius, K. (2009). Genetic variation of 15 autosomal STR loci in a Somali population. Forensic Sci. Int. Genet. *4*, e19–e20. 10.1016/j.fsigen.2009.01.004.
92. Wolfgramm, E. de V., Silva, B.C., da Costa Aguiar, V.R., Malta, F.S.V., de Castro, A.M., de Souza Ferreira, A.C., Prezoti, A.N.L., de Paula, F., and Louro, I.D. (2011). Genetic analysis of 15 autosomal and 12 Y-STR loci in the Espirito Santo State population, Brazil. Forensic Sci. Int. Genet. *5*, e41–e43. 10.1016/j.fsigen.2010.05.001.
93. Hill, C.R., Duewer, D.L., Kline, M.C., Coble, M.D., and Butler, J.M. (2013). U.S. population data for 29 autosomal STR loci. Forensic Sci. Int. Genet. *7*, e82–e83. 10.1016/j.fsigen.2012.12.004.
94. Goodwin, W., Scoular, C., and Linacre, A. (2001). 13 STR loci frequency data from a Scottish population. Forensic Sci. Int. *116*, 187–188. 10.1016/S0379-0738(00)00317-0.
95. Khan, A.T., Gogarten, S.M., McHugh, C.P., Stilp, A.M., Sofer, T., Bowers, M.L., Wong, Q., Cupples, L.A., Hidalgo, B., Johnson, A.D., et al. (2022). Recommendations on the use and reporting of race, ethnicity, and ancestry in genetic research: Experiences from the NHLBI TOPMed program. Cell Genomics *2*, 100155. 10.1016/j.xgen.2022.100155.
96. Fullwiley, D. (2011). 6. Can DNA “Witness” Race? Forensic Uses of an Imperfect Ancestry Testing Technology. In Race and the Genetic Revolution, S. Krimsky and K. Sloan, eds. (Columbia University Press). 10.7312/krim15696-007.
97. Saini, A. (2019). Superior: the return of race science (Beacon Press).
98. Lewontin, R.C. (1972). The Apportionment of Human Diversity. In Evolutionary Biology, T. Dobzhansky, M. K. Hecht, and W. C. Steere, eds. (Springer US), pp. 381–398. 10.1007/978-1-4684-9063-3_14.
99. Rosenberg, N.A., Edge, M.D., Pritchard, J.K., and Feldman, M.W. (2019). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. Evol. Med. Public Health *2019*, 26–34. 10.1093/emph/eoy036.
100. Fullwiley, D. (2007). Race and Genetics: Attempts to Define the Relationship. BioSocieties *2*, 221–237. 10.1017/S1745855207005625.
101. Committee on the Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research, Board on Health Sciences Policy, Committee on Population, Health and Medicine Division, Division of Behavioral and Social Sciences and Education, and National Academies of Sciences, Engineering, and Medicine (2023). Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field (National Academies Press) 10.17226/26902.
102. Balding, D.J., and Nichols, R.A. (1994). DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci. Int. *64*, 125–140. 10.1016/0379-0738(94)90222-4.
103. Fullwiley, D. (2014). The “Contemporary Synthesis”: When Politically Inclusive Genomic Science Relies on Biological Notions of Race. Isis *105*, 803–814. 10.1086/679427.
104. Dauda, B., Molina, S.J., Allen, D.S., Fuentes, A., Ghosh, N., Mauro, M., Neale, B.M., Panofsky, A., Sohail, M., Zhang, S.R., et al. (2023). Ancestry: How researchers use it and what they mean by it. Front. Genet. *14*, 1044555.

10.3389/fgene.2023.1044555.
105. Coop, G. (2022). Genetic similarity versus genetic ancestry groups as sample descriptors in human genetics. 10.48550/ARXIV.2207.11595.
106. Duster, T. (2003). Backdoor to eugenics 2nd ed. (Routledge).
107. Lewis, A.C.F., Molina, S.J., Appelbaum, P.S., Dauda, B., Di Rienzo, A., Fuentes, A., Fullerton, S.M., Garrison, N.A., Ghosh, N., Hammonds, E.M., et al. (2022). Getting genetic ancestry right for science and society. Science *376*, 250–252. 10.1126/science.abm7530.
108. Fujimura, J.H., Duster, T., and Rajagopalan, R. (2008). Introduction: Race, Genetics, and Disease: Questions of Evidence, Matters of Consequence. Soc. Stud. Sci. *38*, 643–656. 10.1177/0306312708091926.
109. Mauro, M., Allen, D.S., Dauda, B., Molina, S.J., Neale, B.M., and Lewis, A.C.F. (2022). A scoping review of guidelines for the use of race, ethnicity, and ancestry reveals widespread consensus but also points of ongoing disagreement. Am. J. Hum. Genet. *109*, 2110–2125. 10.1016/j.ajhg.2022.11.001.
110. Weir, B.S., and Cockerham, C.C. (1984). Estimating F-Statistics for the Analysis of Population Structure. Evolution *38*, 1358. 10.2307/2408641.