Forensic Science International: Genetics 28 (2017) 178–187

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Research paper

The factor of 10 in forensic DNA match probabilities

Simone Gittelson[a,*], Tamyra R. Moretti[b], Anthony J. Onorato[b], Bruce Budowle[c,d], Bruce S. Weir[e], John Buckleton[e,f]

[a] National Institute of Standards and Technology, 100 Bureau Drive, MS 8980 Gaithersburg, MD 20899, USA
[b] DNA Support Unit, Federal Bureau of Investigation Laboratory, 2501 Investigation Parkway, Quantico, VA 22135, USA
[c] Center for Human Identification, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA
[d] Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia
[e] Department of Biostatistics, University of Washington, Seattle, WA 98195, USA
[f] ESR Ltd, Private Bag 92021, Auckland 1142, New Zealand

ARTICLE INFO

Article history:
Received 8 November 2016
Received in revised form 9 February 2017
Accepted 13 February 2017
Available online 16 February 2017

Keywords:
Subpopulations
Allele frequency
Database
Weight of evidence

ABSTRACT

An update was performed of the classic experiments that led to the view that profile probability assignments are usually within a factor of 10 of each other. The data used in this study consist of 15 Identifiler loci collected from a wide range of forensic populations. Following Budowle et al. [1], the terms cognate and non-cognate are used. The cognate database is the database from which the profiles are simulated. The profile probability assignment was usually larger in the cognate database. In 44%–65% of the cases, the profile probability for 15 loci in the non-cognate database was within a factor of 10 of the profile probability in the cognate database. This proportion was between 60% and 80% when the FBI and NIST data were used as the non-cognate databases. A second experiment compared the match probability assignment using a generalised database and recommendation 4.2 from NRC II (the 4.2 assignment) with a proxy for the matching proportion developed using subpopulation allele frequencies and the product rule. The findings support that the 4.2 assignment has a large conservative bias. These results are in agreement with previous research results.

Published by Elsevier Ireland Ltd.

1. Introduction

Modern DNA multiplexes are capable of developing a profile of human DNA at more than 15 loci. Forensic evidence associated with a match of a DNA profile from a scene and a person of interest is usually presented with an associated assessment of the weight of evidence. Note that we use the term profile probability for the probability of a profile, and the term match probability for the probability of a second copy of a profile given that a first copy has been observed [2]. The weight of evidence requires assigning a match probability, expected to be higher than the profile probability.

It is not currently possible to assign multilocus match probabilities based on the results of direct sampling, because it is likely that potential genotypes at multiple loci will be unobserved in any sample of practical size. Nevertheless, the field of forensic genetics can exploit allele proportions at a single locus and population genetic models to assign probabilities for full profiles. In spite of that, DNA probabilities produced by various models are currently not capable of being compared with sample proportions (i.e., sample proportions for a profile for profile probabilities, and sample proportions of pairs of profiles that are the same for match probabilities) by direct means.

In the early days of forensic modelling, core assumptions were based on the expectations of Hardy-Weinberg and linkage equilibrium, collectively termed the product rule. These concepts were examined by independence testing [3,4]. However, it became apparent that independence testing on the datasets available did not have the power to find departures from independence of the size that was plausible for human populations [5–7], and some considerable debate culminated in two National Research Council reports [8,9].

The match probability depends on the allele proportions [10]. Early efforts to assess the robustness of match probability assignments compared assignments made in different ways and with different databases to inform the allele proportions. These analyses led to the conclusion that the error induced by ignoring subpopulation effects may be of the order of a factor of 10 [11–15]. In practice, DNA analysts have used the results and conclusions of these experiments for the past twenty years to testify that the

* Corresponding author.
E-mail address: simone.gittelson@nist.gov (S. Gittelson).

http://dx.doi.org/10.1016/j.fsigen.2017.02.007
1872-4973/Published by Elsevier Ireland Ltd.

match probability can vary at most by a factor of 10 if a different database is used for informing the allele proportions. For example, if the match probability was in the order of magnitude of one in one million, the analyst would testify that using a different database would not produce a value larger than one in one hundred thousand, nor smaller than one in ten million. In the past two decades, this factor of 10 has become enshrined as a rule of thumb in forensic testimony.

However, the data used in the experiments leading to this factor of 10 [11–15] was obtained in the early and mid-90s, on a different type of marker and with fewer loci than in common use today. The results herein update the expected variation based on short tandem repeat (STR) loci.

The product rule itself has largely been replaced by a model based on NRC II recommendations 4.1 for profile probabilities and 4.2 for match probabilities [9]. NRC II recommendation 4.1 assigns the profile probability (denoted $\hat{q}_{dl}$in this paper, with the "^" indicating that this quantity is assigned) for locus l and database d as:

$$\hat{q}_{dl} = \begin{cases} \hat{p}_{dli}^2 + \hat{p}_{dli}(1-\hat{p}_{dli})F & \text{for the homozygote } a_i a_i \text{ at locus } l \text{ in database d} \\ 2\hat{p}_{dli}\hat{p}_{dlj} & \text{for the heterozygote } a_i a_j \text{ at locus } l \text{ in database d} \end{cases}, \quad (1)$$

where $\hat{p}_{dli}$ is the assigned allele proportion of allele a_i of locus l based on the data in database d, and F is the inbreeding coefficient. This recommendation does not describe match probabilities, yet is widely used in the United States for assigning the weight of evidence. NRC II recommendation 4.2 assigns the match probability (denoted $\hat{m}_{dl}$in this paper, with the "^" indicating that this quantity is assigned) for locus l and database d as:

$$\hat{m}_{dl} = \begin{cases} \frac{(2\theta+(1-\theta)\hat{p}_{dli})(3\theta+(1-\theta)\hat{p}_{dli})}{(1+\theta)(1+2\theta)} & \text{for the homozygote } a_i a_i \text{ at locus } l \text{ in database } d \\ \frac{2(\theta+(1-\theta)\hat{p}_{dli})(\theta+(1-\theta)\hat{p}_{dlj})}{(1+\theta)(1+2\theta)} & \text{for the heterozygote } a_i a_j \text{ at locus } l \text{ in database } d \end{cases}, \quad (2)$$

where θ is the co-ancestry coefficient. NRC II recommendation 4.2 is a model for match probabilities proposed by Balding and Nichols [16]. The probabilistic genotyping software programs we are aware of, and of which we have knowledge about the implemented population genetic model, use recommendation 4.2 [17–20] or similar variations of it [21,22].

2. Methods

This study used databases containing the 15 Identifiler loci with $N_d \geq 100$, where N_d is the number of individuals in database d. There was one exception: the NIST Asian dataset [23] was included in this study even though it has only 97 individuals because knowledge on how this dataset behaves is important for the forensic science community that uses the NIST databases. All samples typed for 15 loci with the Identifiler kit (Applied Biosystems, San Franscisco, CA) were used, whether these loci had been typed with the Identifiler multiplex or any other multiplex. The data were obtained from the literature. Each population was assigned to one of four ethnic clusters: African, Caucasian, Asian or Hispanic. Supplementary Table 1 (mmc1) lists the populations used, the database sizes, the associated references and the ethnic cluster assignments.

Ideally, the experiments in this study would be conducted with real profiles from each of the subpopulations. Unfortunately, these data were unavailable, so we simulated the profiles. A profile was simulated by independently drawing two alleles from each locus based on the empirical allele frequency data of a database. Thus, at each locus a particular allele was drawn from the set of possible alleles at that locus with a probability corresponding to that allele's empirical relative frequency in the database of interest. For example, in the imaginary case of two possible alleles at a locus, say allele A and allele B, with empirical relative frequencies of 0.4 and 0.6 for alleles A and B, respectively, two alleles would be drawn independently with probabilities of 0.4 for drawing an A and 0.6 for drawing a B. In this simulation, the alleles drawn at one locus are assumed to be independent of the alleles drawn at each of the other loci.

Following Budowle et al. [1], the terms cognate and non-cognate are used. The cognate database describes the database for which the profiles are simulated. A non-cognate database describes a database different from the one for which the profiles are simulated.

All of the experiments applied the following equations:

1) A "5/2N" adjustment on the allele proportions, $\hat{p}_{dli}$, of rare alleles [9]:

$$\hat{p}_{dli} = \max\left(\frac{x_{dli}}{2N_d}, \frac{5}{2N_d}\right),$$

where

x $_{dli}$ is the count of allele a_i at locus l in database d.

2) If match probabilities were assigned, multiplication of the match probabilities of the individual loci, $\hat{m}_{dl}$, to obtain the match probability of all 15 Identifiler loci, $\hat{m}_d$:

$$\hat{m}_d = \prod_l \hat{m}_{dl},$$

and if profile probabilities were assigned, multiplication of the profile probabilities of the individual loci, $\hat{q}_{dl}$, to obtain the profile probability of all loci, $\hat{q}_d$:

$$\hat{q}_d = \prod_l \hat{q}_{dl},$$

2.1. Experiment 1: the effect of using a database from a different subpopulation

Four hundred profiles were simulated[1] using allele frequencies from the cognate database for Identifiler loci: 100 for four loci (D8S1179, D3S1358, D19S433 and D5S818), 100 for eight loci (D8S1179, D3S1358, D19S433, D5S818, D21S11, TH01, vWA and D13S317), 100 for 12 loci (D8S1179, D3S1358, D19S433, D5S818, D21S11, TH01, vWA, D13S317, D7S820, D16S539, TPOX and D18S51) and 100 for all 15 Identifiler loci.

The objective of Experiment 1 is to examine the effect of using a database from a different subpopulation when applying the prevalent approach in the United States. The prevalent approach

[1] Ideally, this experiment would be conducted with real profiles from each of the subpopulations. Unfortunately, these data were unavailable.

in the United States is not to assign match probabilities, but profile probabilities using NRC II recommendation 4.1. In view of this objective, profile probabilities, $\hat{q}_{dl}$, were assigned at each locus according to NRC II recommendation 4.1 (see Eq. (1)), with the inbreeding coefficient set to 0.01.

To quantify the effect of using a different database (i.e., a non-cognate database) for assigning the profile probability, the ratio of the assigned profile probability in the non-cognate database, $\hat{q}_{\overline{c}}$, to the assigned profile probability in the cognate database, $\hat{q}_c$, was calculated. This value is ratio $\hat{s}$:

$$\hat{s} = \frac{\hat{q}_{\overline{c}}}{\hat{q}_c},$$

This experiment was performed for three types of scenarios:

(a) Subpopulations

This scenario examines all possible combinations of cognate and non-cognate databases within an ethnic cluster. For each ethnic cluster, the cognate database is each of the databases in that cluster in turn. For each cognate database, the non-cognate database is each of the other databases in the same cluster. The list of databases for each cluster is given in Supplementary Table 1 (mmc1).

(b) Federal Bureau of Investigation (FBI) and National Institute of Standards and Technology (NIST) databases

This scenario examines the performance of the Federal Bureau of Investigation (FBI) [24] and National Institute of Standards and Technology (NIST) [23] databases with regard to profiles simulated in other databases. For each ethnic cluster, the cognate database is each of the databases in that cluster except for the FBI and NIST databases of that ethnic group. For each cognate database, the non-cognate databases are the FBI and NIST databases of the corresponding ethnic group. For the Hispanic cluster, the results are given for two FBI Hispanic databases (SE and SW Hispanic), as well as for a database created by combining these two data sets (combined Hispanic).

(c) FBI Hispanic databases

This scenario examines the performance of the FBI SE Hispanic, the FBI SW Hispanic and the FBI combined Hispanic databases for profiles simulated in the SE Hispanic and SW Hispanic databases. First, the cognate database is the FBI SW Hispanic database, and the non-cognate databases are the FBI SE Hispanic database and the FBI combined Hispanic database. Second, the cognate database is the FBI SE Hispanic database, and the non-cognate databases are the FBI SW database and the FBI combined Hispanic database. In this experiment, 8000 profiles were simulated from the cognate database: 2000 for the 4 loci, 2000 for the 8 loci, 2000 for the 12 loci and 2000 for all 15 Identifiler loci.

Experiment 1 examines how the assigned profile probability varies when one uses allele frequency data from a subpopulation different from the subpopulation from which the profile alleles were drawn. This experiment approximates an update of the work underlying the factor of 10 espoused in the NRC II report.

However, Budowle et al. [1] pointed out that this methodology can lead to biases that make a profile appear to be more common in the cognate database. This bias increases as the database size decreases. To examine the extent of this bias on the results of Experiment 1, an additional test was performed to quantify the maximum effects of the bias alone:

(d) Extent of bias due to the methodology

Ten times 100 individuals were simulated using the frequency data of an existing database (NIST's US Caucasian dataset was used). These formed 10 new small databases, each with $N_d = 100$, which is the smallest database size used in Experiment 1 with the exception of the NIST Asian dataset ($N_d = 97$). Then cognate and non-cognate profile probabilities were assigned for each set of 100 individuals using the allele frequency data in the 10 newly created databases. Again, this simulation was done for 4, 8, 12 and all 15 Identifiler loci. This experiment was repeated for a database size of 97 and a database size of 250 to examine the effect of the database size on this bias.

Experiment 1, however, does not address the performance of match probabilities required for assessing the weight of evidence. More specifically, it does not answer the question of how well the match probability assignment performs with allele frequency data of a general population, where this general population is actually composed of many subpopulations. Experiment 2 addresses this question.

2.2. Experiment 2: the effect of using NRC II recommendation 4.2 with the FBI and NIST allele proportions

This experiment is more informative of realistic situations than Experiment 1. The situation investigated is where one uses an FBI or a NIST database for the ethnic cluster of interest (e.g., the FBI African American allele frequency data) to assign a match probability. Given that these FBI and NIST databases are used as representative of the general US population, and that the US population actually consists of many subpopulations, NRC II recommendation 4.2 (the 4.2 assignment) was used in conjunction with the FBI or NIST database to assign a match probability. While the matching proportion is sought, it is never known. Studies suggest that the use of subpopulation allele frequencies and the product rule give an estimate with a low bias [25,26] of the match frequency in that subpopulation. This estimate is the standard against which the 4.2 assignment will be compared.

Hence the cognate databases in this experiment are defined as each of the databases in an ethnic cluster, except for the general FBI and NIST databases for that ethnic group. For the cognate databases, match probabilities, $\hat{m}_{dl}$, were assigned using the product rule:

$$\hat{m}_{dl} = \begin{cases} \hat{p}_{dli}^2 & \text{for the homozygote } a_i a_i \text{ at locus } l \text{ in database } d \\ 2\hat{p}_{dli}\hat{p}_{dlj} & \text{for the heterozygote } a_i a_j \text{ at locus } l \text{ in database } d \end{cases}.$$

Budowle et al. [1] have shown that choosing a profile from a database has a mild tendency to bias the estimate for that database upwards. To generate profiles without regard to allele frequencies and better mimic the situation, profiles were simulated using allele frequencies from all the cognate databases within an ethnic cluster combined by simple amalgamation of the allele counts. A total of 40,000 profiles were simulated in this way from each of the four ethnic clusters (African, Asian, Caucasian, and Hispanic) for Identifiler loci: 10,000 for four loci (D8S1179, D3S1358, D19S433 and D5S818), 10,000 for eight loci (D8S1179, D3S1358, D19S433, D5S818, D21S11, TH01, vWA and D13S317), 10,000 for 12 loci (D8S1179, D3S1358, D19S433, D5S818, D21S11, TH01, vWA, D13S317, D7S820, D16S539, TPOX and D18S51) and 10,000 for all 15 Identifiler loci.

The non-cognate databases are defined as the general FBI or NIST databases for that ethnic group. For the non-cognate database, match probabilities, $\hat{m}_{dl}$, were assigned using NRC II recommendation 4.2 (see Eq. (2)), with the co-ancestry coefficient set to 0.01.

The ratio of the assigned match probability in the non-cognate databases, $\hat{m}_{\overline{c}}$, to the assigned match probability in the cognate database, $\hat{m}_c$, $\hat{r} = \frac{\hat{m}_{\overline{c}}}{\hat{m}_c}$, is calculated.

3. Results

3.1. Experiment 1: the effect of using a database from a different subpopulation

The results for Experiment 1 are presented as a series of graphs (Figs. 1–3 and Supplementary Fig. 1 (mmc7)). Each graph shows the ratio $\hat{s}$ vs $\hat{q}_c$ on a logarithmic scale. Values of $\log(\hat{s}) = 0$ indicate that the $\hat{q}_{\overline{c}}$ is the same as $\hat{q}_c$. Values of $\log(\hat{s}) > 0$ mean that $\hat{q}_{\overline{c}}$ is greater than $\hat{q}_c$. In this situation, the non-cognate database profile probability is larger than the cognate database profile probability. Values of $\log(\hat{s}) < 0$ mean that $\hat{q}_{\overline{c}}$ is smaller than $\hat{q}_c$. In this situation, the non-cognate database makes the profile look rarer than suggested by the value assigned by the cognate database. Note that the graphs are made up of hexagons rather than data points. The area of each hexagon is proportional to the number of data points plotted at each position in the graph.

Each graph plots data for 4, 8, 12 and 15 loci. This presentation creates four clusters of points within each graph. The profile probability assigned in the cognate database, $\hat{q}_c$, decreases as the number of loci increases, so that from left to right these clusters represent the data for 15 loci, 12 loci, 8 loci and 4 loci, respectively. As the profile probability in the cognate database increases (i.e., proceeding from left to right on the x-axis), the mean of $\hat{s}$ regresses towards 1 from below and the range for $\hat{s}$ contracts.

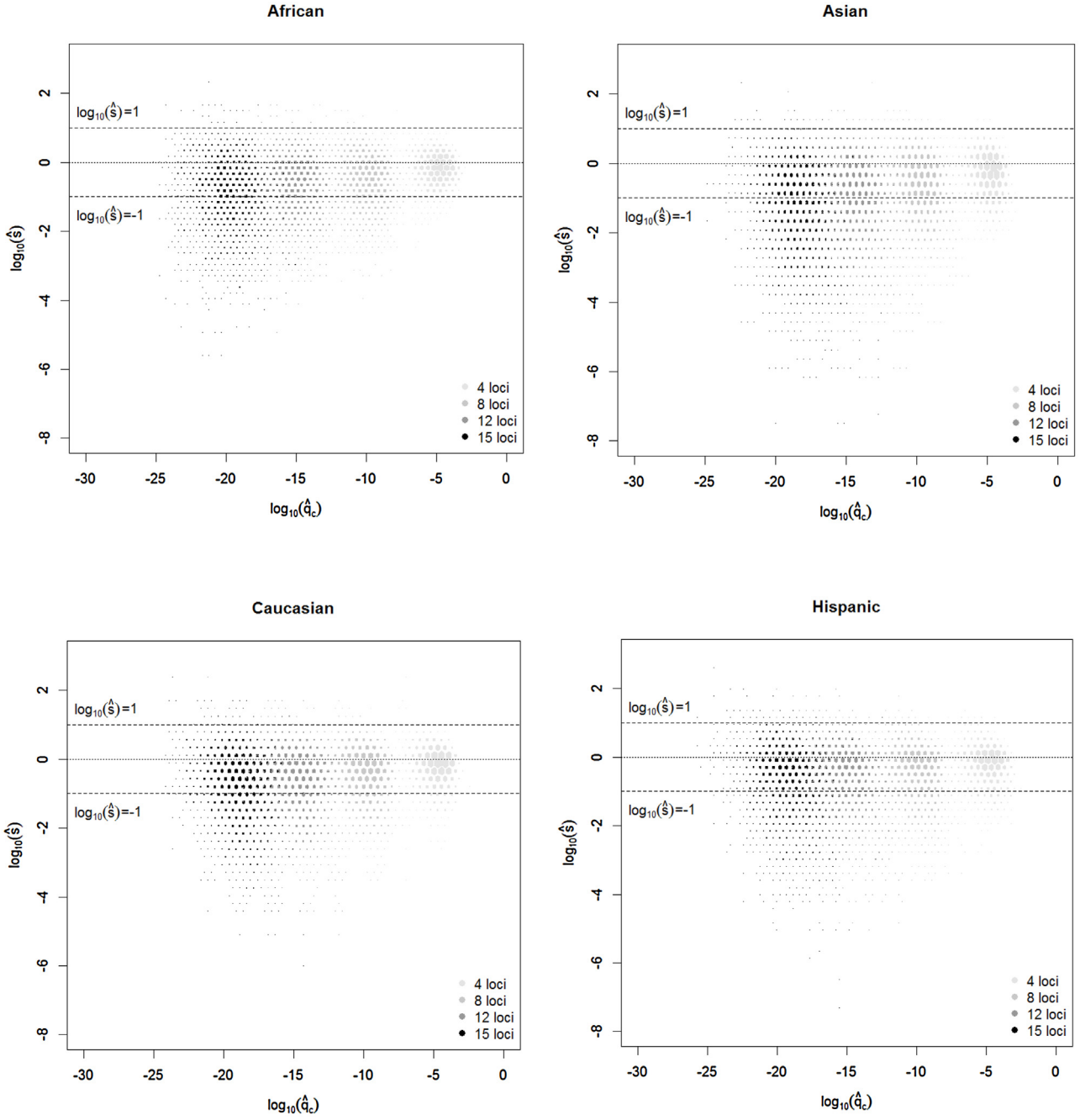

Fig. 1. Results of Experiment 1a. Each database within a cluster serves as the cognate database to simulate 400 profiles. Ratio $\hat{s}$ is obtained for each non-cognate database within the cluster. The graphs show the plots of the log of $\hat{s}$ in function of the log of $\hat{q}_c$ for the African, Asian, Caucasian and Hispanic clusters.

For each of the graphs in Figs. 1–3 and Supplementary Fig. 1 (mmc7) and for each number of loci used to produce the data in the graph, three numbers are presented in Supplementary Tables 2 through 4 (mmc2, mmc3, mmc4):

1 the proportion of values where the assigned non-cognate profile probability, $\hat{q}_{\overline{c}}$, represents a rarer profile than the assigned cognate profile probability, $\hat{q}_c$, that is, the proportion where $\log(\hat{s}) < 0$
2 the proportion of values where the assigned non-cognate profile probability, $\hat{q}_{\overline{c}}$, differs from the assigned cognate profile probability, $\hat{q}_c$, by a factor of 10 or less
3 the proportion of values where the assigned non-cognate profile probability, $\hat{q}_{\overline{c}}$, represents a profile rarer than the assigned cognate profile probability, $\hat{q}_c$, by a factor greater than 10, that is, the proportion where $\log(\hat{s}) < -1$

Generally, the range of values obtained for $\log(\hat{s})$ increases as $\hat{q}_c$ decreases, so this range increases as the number of loci increases. Hence, the proportion of ratios, $\hat{s}$, within a range from 0.1 to 10 decreases as the number of loci used to assign the profile probability increases (Supplementary Tables 2 through 4, sheet 2). One also notes that the proportion of $\hat{q}_{\overline{c}}$ indicating a rarer profile than $\hat{q}_c$ increases as the number of loci increases and $\hat{q}_c$ decreases (Supplementary Tables 2 through 4, sheet 1).

3.1.1. Experiment 1a: subpopulations

Fig. 1 shows the results of Experiment 1a as four graphs, one for each ethnic cluster. In each graph, every database within that ethnic cluster is used as the cognate database, and for each cognate database, every other database within that cluster is used as the non-cognate database. All ratios $\hat{s}$ are presented on the graph. The largest variation of $\hat{s}$ is observed for the Asian ethnic group where the $\log(\hat{s})$ values range from −10 to 4. This range means that using a non-cognate database can make a profile look up to 10^{10} times rarer than what it would be using the cognate database, and up to 10^4 times more common than what it would be using the cognate database. The variation is smaller for the other ethnic groups: the $\log(\hat{s})$ values range from −8 to 4 for the Caucasian ethnic group and from −6 to 2 for the African and Hispanic ethnic groups. These ranges represent the maximum range which is obtained for 15 loci. The graphs show that the majority of the $\log(\hat{s})$ values are smaller than 0 (i.e., the profile is rarer using the non-cognate database than using the cognate database). This observation holds from 4 loci up to 15 loci, yet the proportion of assigned non-cognate profile probabilities, $\hat{q}_{\overline{c}}$, that represent a rarer profile than the corresponding assigned cognate profile probabilities, $\hat{q}_c$, increases as the number of loci increases (Supplementary Table 2, sheet 1). The proportion of $\hat{q}_{\overline{c}}$within a factor of 10 of $\hat{q}_c$ decreases from over 90% for 4 loci to about one half for 15 loci (Supplementary Table 2, sheet 2). Most of this discrepancy is in favor of $\hat{q}_{\overline{c}}$ indicating a rarer profile than $\hat{q}_c$ (Supplementary Table 2, sheet 3), and occasionally by many orders of magnitude. On the other hand, the discrepancy in favor of $\hat{q}_{\overline{c}}$ indicating a more common profile than $\hat{q}_c$ is occasionally greater than a factor of 10 and not often greater than a factor of 100.

3.1.2. Experiment 1b: FBI and NIST databases

Fig. 2 presents a selection of the results of Experiment 1b. Additional graphs for the NIST databases and for the FBI combined Hispanic dataset are presented in Supplementary Fig. 1 (mmc7). Each graph is specific to one ethnic cluster and presents the results for either an FBI non-cognate database or a NIST non-cognate database representative of that ethnic group. Compared to Experiment 1a, similar but less extreme results were obtained. The $\log(\hat{s})$ values range from −6, −5 and −4 to 2. This indicates that using the FBI/NIST database for 15 loci can make a profile look up to 10^6, 10^5, and 10^4 times rarer than what it would be using the cognate database, and up to 10^2 times more common than what it would be using the cognate database. Again, the majority of the $\log(\hat{s})$ values are smaller than 0 (Supplementary Table 3, sheet 1), which means that the majority of the assigned FBI/NIST profile probabilities$\hat{q}_{\overline{c}}$ indicate rarer profiles than the corresponding assigned cognate profile probabilities, $\hat{q}_c$. Compared with the results of Experiment 1a, a greater proportion (i.e., between 60% and 80%) of $\hat{q}_{\overline{c}}$ fall within a factor of 10 of $\hat{q}_c$ for 15 loci (Supplementary Table 3, sheet 2). Again, most of the discrepancy is towards$\hat{q}_{\overline{c}}$ indicating a rarer profile than $\hat{q}_c$by a factor greater than 10 (Supplementary Table 3, sheet 3).

An additional study comparing the FBI databases with the NIST databases showed that more than 96% of the profile probabilities for 15 loci are within a factor of 10 of each other. This percentage increases for fewer loci. Figures and tables of these additional results are available upon request from the authors.

3.1.3. Experiment 1c: FBI hispanic databases

Fig. 3 presents the results of Experiment 1c as a set of four graphs. In two of these, the cognate database is SW Hispanics, and the non-cognate databases are SE Hispanics and combined Hispanics, respectively. In the other two, the cognate database is SE Hispanics, and the non-cognate databases are SW Hispanics and combined Hispanics, respectively. The use of the SW Hispanic and SE Hispanic databases as the non-cognate database produced a similar fraction of profile probabilities that differed by more than an order of magnitude from the cognate profile probabilities as in Experiment 1b, though the discrepancies were less extreme: the $\log(\hat{s})$ values range from −3 to 2 for 15 loci. In these situations, the proportion of $\hat{q}_{\overline{c}}$ indicating a rarer profile than $\hat{q}_c$reaches about 80% for 15 loci (Supplementary Table 4, sheet 1). The proportion of $\hat{q}_{\overline{c}}$ within a factor of 10 of $\hat{q}_c$ decreases to about 60% for 15 loci (Supplementary Table 4, sheet 2). The combined Hispanic database, however, assigned profile probabilities that were mostly (i.e., over 95% for 15 loci) within one order of magnitude of the assigned cognate profile probability. In all of the results, most or all of the discrepancy is towards $\hat{q}_{\overline{c}}$ indicating a profile rarer than $\hat{q}_c$ by a factor greater than 10 (Supplementary Table 4, sheet 3).

3.1.4. Experiment 1d: extent of bias due to the methodology

Supplementary Table 5 (mmc5) presents the results obtained in the bias experiment conducted on NIST's US Caucasian dataset for 4, 8, 12 and 15 loci. Databases of size 97 produce results that are very close to what is observed for databases of size 100 (Supplementary Table 5, columns 1 and 2). If the databases are larger, the results (e.g., $N_d = 250$ in Supplementary Table 5, column 3) confirm that the bias is smaller (i.e., a much higher proportion of $\hat{q}_{\overline{c}}$ are within a factor of 10 of$\hat{q}_c$, and a smaller proportion of $\hat{q}_{\overline{c}}$ indicate a rarer profile than$\hat{q}_c$). So the results presented in Supplementary Table 5 for $N_d = 100$ ($N_d = 97$ for the Asian ethnic cluster) represent the greatest effect the bias can have on the results of Experiments 1a, 1b and 1c.

As expected from the bias [1], this experiment produced results where the profile is the most common in the cognate database (Supplementary Table 5, sheet 1). To see whether the bias can explain the results obtained in Experiments 1a, 1b and 1c, the results for 15 loci for $N_d = 100$ (Supplementary Table 5, column 2) are compared with the results obtained for 15 loci in Experiments 1a, 1b, and 1c. For the results obtained in Experiment 1c, this

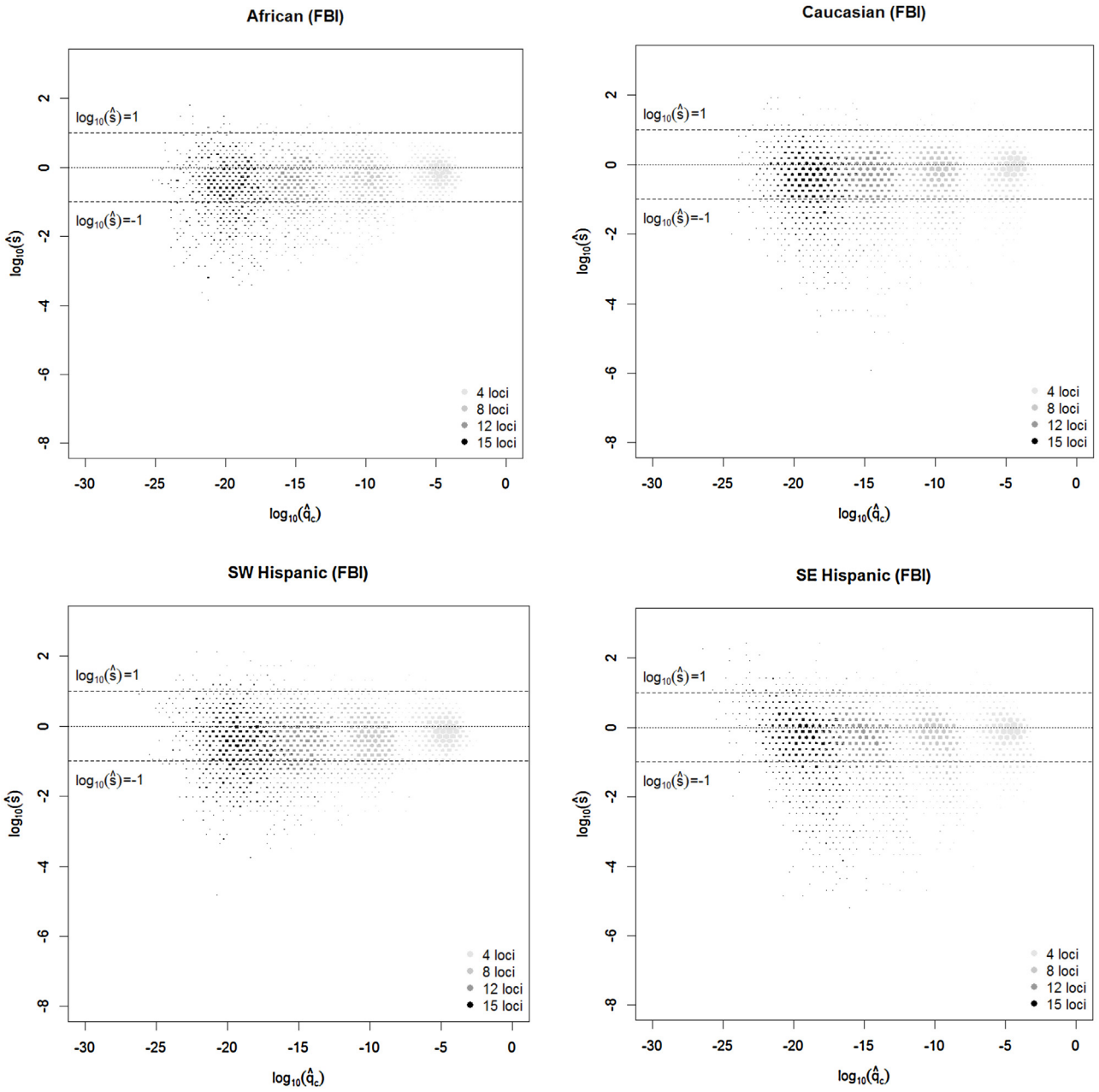

Fig. 2. A selection of results of Experiment 1b. Each database within a cluster, with the exception of the FBI and NIST databases, serves as the cognate database to simulate 400 profiles. Ratio $\hat{s}$ is obtained for the cluster's FBI and NIST databases, respectively. The graphs show the plots of the log of $\hat{s}$ in function of the log of $\hat{q}_c$ for the African, Asian, Caucasian and Hispanic clusters (note that the FBI has two Hispanic databases: one for Southwest (SW) Hispanics and one for Southeast (SE) Hispanics). Additional results are presented in Supplementary Fig. 1 (mmc7).

comparison is made only with the results for the SE Hispanic and SW Hispanic non-cognate databases. The results of the bias experiment show a smaller proportion where $\hat{q}_{\overline{c}}$ indicates a rarer profile than $\hat{q}_c$ (i.e., about 73% compared to about 80% in Experiments 1a, 1b and 1c). The proportion where $\hat{q}_{\overline{c}}$ is within a factor of 10 of $\hat{q}_c$ is larger (i.e., about 89% compared to values ranging from about 50% to 80% in Experiments 1a, 1b and 1c).

3.2. Experiment 2: the effect of using NRC II recommendation 4.2 with the FBI and NIST allele proportions

The results for experiment 2 are presented as a series of graphs (Fig. 4 and Supplementary Fig. 2 (mmc8)) of ratio $\hat{r}$ vs $\hat{m}_c$on a logarithmic scale. In each graph, the cognate database is each database within that ethnic cluster and the non-cognate database is the nominated database (i.e., FBI or NIST database for that ethnic cluster). These graphs show the log of the ratio of the 4.2 assignment in the non-cognate database to the assignment using the product rule in each cognate database. When the log of this ratio is equal to 0 (i.e., $\log(\hat{r}) = 0$), this means that the two match probability assignments give the same value. A $\log(\hat{r}) > 0$ means that the 4.2 assignment with the nominated database produces a match probability representing a more common match than the product rule assignment in the cognate database. A $\log(\hat{r}) < 0$ means that the 4.2 assignment with the nominated database produces a match probability representing a rarer match than the product rule assignment in the cognate database.

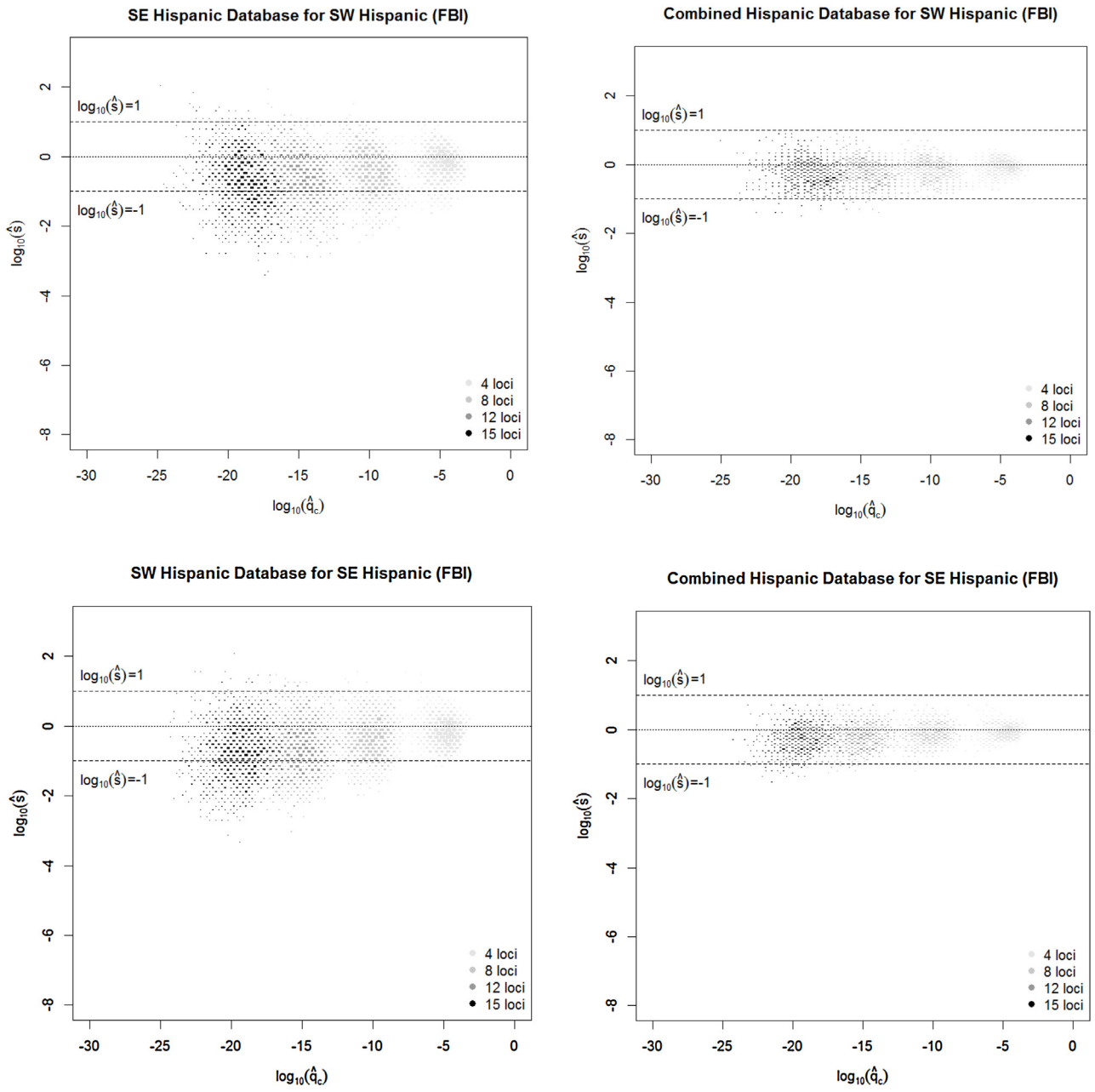

Fig. 3. Results of Experiment 1c. In the top row, the FBI SW Hispanic database serves as the cognate database to simulate 8000 profiles, and ratio $\hat{s}$ is obtained for the FBI SE Hispanic database (left) and the FBI SW and SE combined Hispanic database (right). In the bottom row, the FBI SE Hispanic database serves as the cognate database to simulate 8000 profiles, and ratio $\hat{s}$ is obtained for the FBI SW Hispanic database (left) and the FBI SW and SE combined Hispanic database (right).

Like for Experiment 1, each graph presents data for 4, 8, 12 and 15 loci. Again this produces four clusters of data: from left to right, these clusters represent the data for 15 loci, 12 loci, 8 loci and 4 loci, respectively. As the match probability increases, the mean of $\hat{r}$ regresses towards 1 from above and the range of $\hat{r}$contracts.

For each graph in Fig. 4 and Supplementary Fig. 2 (mmc8) and for each number of loci, Supplementary Table 6 (mmc6) presents:

1 the proportion of values where the non-cognate match probability, $\hat{m}_{\overline{c}}$, indicates a rarer match than the cognate match probability, $\hat{m}_c$, that is, the proportion where $\log(\hat{r}) < 0$
2 the proportion of values where the non-cognate match probability,$\hat{m}_{\overline{c}}$, differs from the cognate match probability,$\hat{m}_c$, by a factor of 10 or less
3 the proportion of values where the non-cognate match probability, $\hat{m}_{\overline{c}}$, indicates a match rarer than the cognate match probability, $\hat{m}_c$, by a factor greater than 10, that is, the proportion where $\log(\hat{r}) < -1$

As in the results for Experiment 1, the range of values obtained for $\log(\hat{r})$ increases as $\hat{m}_c$ decreases, so that this range increases as the number of loci increases. As a consequence, the proportion of ratios $\hat{r}$ within a range from 0.1 to 10 decreases as the number of loci used to assign the match probability increases (Supplementary Table 6, sheet 2). However, unlike the results of Experiment 1, the proportion of $\hat{m}_{\overline{c}}$ indicating a rarer match than $\hat{m}_c$ decreases as the number of loci increases and $\hat{m}_c$ decreases (Supplementary Table 6, sheet 1). The

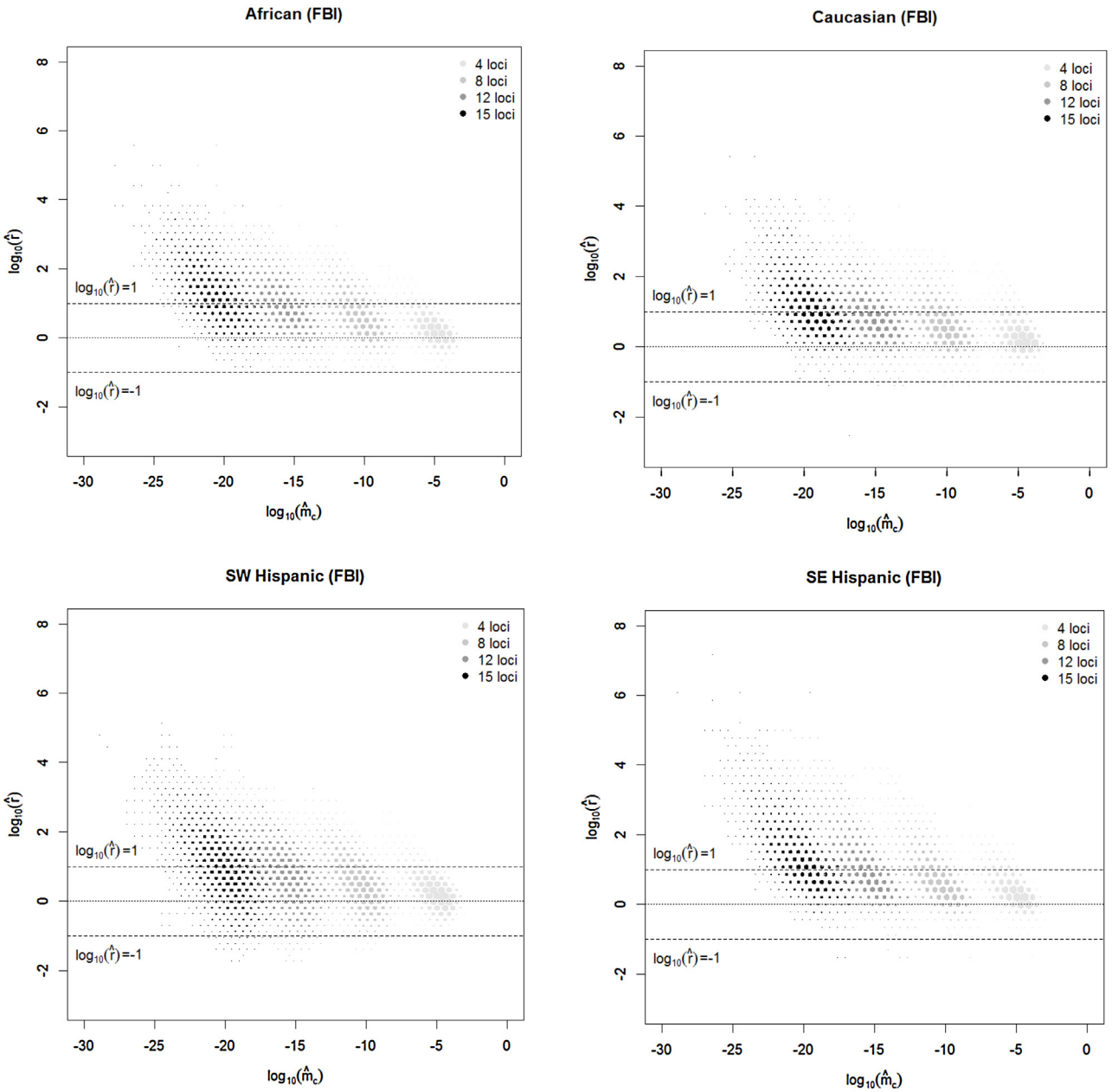

Fig. 4. A selection of results of Experiment 2. The cognate databases are all the databases within a cluster, with the exception of that ethnic group's FBI and NIST databases. Their match probabilities, $\hat{m}_c$, are obtained using the product rule. The non-cognate database is the cluster's FBI or NIST database. The match probability, $\hat{m}_{\overline{c}}$, is obtained using NRC II recommendation 4.2. The graphs show the plots of the log of $\hat{r}$ in function of the log of $\hat{m}_c$ for the African, Asian, Caucasian and Hispanic clusters (note that the FBI has two Hispanic databases: one for Southwest (SW) Hispanics and one for Southeast (SE) Hispanics). Additional results are presented in Supplementary Fig. 2 (mmc8).

discrepancy is now in favor of $\hat{m}_{\overline{c}}$ representing a more common match than $\hat{m}_c$.

The $\log(\hat{r})$ values range from -2 to 8. This means that using the FBI/NIST database with the 4.2 assignment can make the match look up to 10^2 times rarer and up to 10^8 times more common than what it would be if the cognate subpopulation database were used with the product rule. With the exception of NIST's Asian database and the FBI's SW Hispanic database, the FBI/NIST database with the 4.2 assignment makes a match of 15 loci look more common than the cognate subpopulation database with the product rule more than 95% of the time (Supplementary Table 6, sheet 1). NIST's Asian database with the 4.2 assignment makes a match of 15 loci look more common a little over 93% of the time, and the FBI's SW Hispanic database with the 4.2 assignment makes a match of 15 loci look more common a little over 85% of the time. The proportions of non-cognate match probabilities, $\hat{m}_{\overline{c}}$, that are within a factor of 10 of the cognate match probabilities, $\hat{m}_c$, for 15 loci range from 37.7% for NIST's African American data to 50.2% for the FBI's SW Hispanic data (Supplementary Table 6, sheet 2). The proportions of $\hat{m}_{\overline{c}}$ that represent a match rarer than $\hat{m}_c$ by a factor greater than 10 are a fraction of a percentage for all the databases except the SW Hispanic database where this proportion reaches values slightly over 1% (Supplementary Table 6, sheet 3).

4. Discussion and conclusions

The results of Experiment 1 show a majority of $\log(\hat{s})$ values smaller than 0. This observation occurs when the profile

probability assignment in the cognate database is greater (i.e., the profile is more common) than in the non-cognate database. There are several possible explanations for this observation. On the one hand, this observation is a manifestation that a profile is usually the most common in its own database (i.e., the cognate database) confounded with the bias described by Budowle et al. [1]. If one samples profiles according to the distribution of the cognate database, one will generally obtain profiles with high profile probabilities. It is therefore not too surprising that a lower profile probability will be assigned using a non-cognate database. On the other hand, the presence of subpopulations within an ethnic cluster would lead the profile to seem more common in the cognate database than in the non-cognate database. The results of the bias experiment indicate that the bias described by Budowle et al. [1] alone cannot produce results as extreme as the ones observed in Experiments 1a, 1b and 1c. This suggests that the differences between the values presented in Supplementary Table 5 and the results of Experiments 1a, 1b and 1c (Supplementary Tables 2 through 4) are due to the subpopulations within each ethnic cluster. This discrepancy, which can be greater than a factor of 10, tends towards the profile probability representing a rarer profile in the non-cognate database than in the cognate database. From a forensic science perspective, this is non-conservative because it could make a person of interest's profile appear rarer.

The results for the Asian cluster show the greatest variation between the cognate profile probability and the non-cognate profile probability. The results of the bias experiment showed that these extreme values cannot be explained by the database size and the bias described by Budowle et al. [1] alone. This observation suggests that the greater variation is due to more divergent subpopulations within the Asian cluster.

According to the results of Experiment 1c, assigning profile probabilities using the FBI's combined Hispanic database produces values that are almost all within a factor of 10 of the cognate profile probabilities for SE Hispanics and SW Hispanics. This observation suggests that the use of the FBI's combined Hispanic database is acceptable for assigning profile probabilities when assigned to any Hispanic profile in this study.

Experiment 2 is an attempt to assess the discrepancy induced by the subpopulation effect when using a generalised database and the 4.2 assignment. This experiment simulates the situation where the person of interest and the true offender are from the same subgroup and the match probability is assigned using the 4.2 assignment and an FBI or NIST database.

In Fig. 4, data above $\log(\hat{r}) = 0$ occur when the 4.2 assignment is conservative. The overwhelming majority of data suggests that the 4.2 assignment (with $\theta = 0.01$) is conservative (see Supplementary Table 6). These results are in agreement with previous research results [25,26].

This experiment also suggests that a combined Hispanic database and the 4.2 assignment would operate reasonably well at assigning match probabilities for Hispanic subpopulations.

The results of both experiments illustrate how the range of values of the ratio ($\hat{s}$ in Experiment 1 and $\hat{r}$ in Experiment 2) increases as the number of loci used increases. This result is reasonable because one can expect an amount of variation due to the different allele probabilities for each locus for different subpopulations. Hence, using more loci creates a multiplicative effect of these differences, which can lead to greater variation between the cognate probability and the non-cognate probability.

For the kits used today, even using as few as 4 loci produces a variation between the cognate probability and the non-cognate probability that exceeds a factor of 10, although not often. The use of 15 loci increases the proportion of results that exceed a factor of 10 to over 35% when any dataset is used as the non-cognate database, and to over 20% when the FBI or NIST dataset is used as the non-cognate database. Using the 4.2 assignment, however, provides a conservative solution.

Acknowledgements

The authors sincerely thank Johanna Veth, Michael Coble, Adam Pintar and two anonymous reviewers for their valuable comments. This work was supported in part by grant number 2011-DN-BX-K541 from the US National Institute of Justice. The views expressed in this document are those of the authors and do not necessarily reflect the official position or policies of the U.S. Department of Justice, the National Institute of Standards and Technology (NIST), the Federal Bureau of Investigation (FBI) or the U.S. Government. Certain commercial equipment, instruments, or materials (or suppliers, or software) are identified in this paper to foster understanding. Names of commercial manufacturers are provided for identification purposes only, and inclusion does not imply recommendation or endorsement of the manufacturer, or its products or services, by the FBI or NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2017.02.007.

References

[1] B. Budowle, K.L. Monson, A.M. Guisti, A reassessment of frequency estimates of PVUII-generated VNTR profiles in a Finnish, an Italian, and a general Caucasian database—no evidence of ethnic subgroups affecting forensic estimates, Am. J. Hum. Genet. 55 (1994) 533–539.
[2] I.W. Evett, B.S. Weir, Interpreting {DNA} Evidence, Sinauer Associates, Sunderland MA, 1998.
[3] D. Zaykin, L.A. Zhivotovsky, B.S. Weir, Exact tests for association between alleles at arbitrary numbers of loci, Genetica 96 (1995) 169–178.
[4] D. Zaykin, L.A. Zhivotovsky, P.H. Westfall, B.S. Weir, Truncated product method for combining p-values, Genet. Epidemiol. 22 (2002) 170–185.
[5] I.W. Evett, J.S. Buckleton, Statistical analysis of STR data, in: A. Carracedo, B. Brinkmann, W. Bär (Eds.), Advances in Forensic Haemogenetics, Springer Verlag, Berlin, 1996, pp. 79–86.
[6] B. Law, B.S. Weir, C.M. Triggs, J.S. Buckleton, Effects of population structure and admixture on exact tests for association between loci, Genetics 164 (2003) 361–387.
[7] C.M. Triggs, J.S. Buckleton, Logical implications of applying the principles of population genetics to the interpretation of DNA profiling evidence, Forensic Sci. Int. 128 (2002) 108–114.
[8] NRC I, National Research Council, Committee on DNA Technology in Forensic Science, Board on Biology, Commission on Life Sciences, DNA Technology in Forensic Science, National Academy Press, Washington, D.C, 1992.
[9] NRC II, National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence, National Academy Press, Washington, D. C, 1996.
[10] G.W. Beecham, B.S. Weir, Confidence interval of the likelihood ratio associated with mixed stain DNA evidence, J. Forensic Sci. (2011) 2011.
[11] B. Budowle, K.L. Monson, A.M. Giusti, B.L. Brown, The assessment of frequency estimates of Hae III-generated VNTR profiles in various reference databases, J. Forensic Sci. 39 (1994) 319–352.
[12] B. Budowle, K.L. Monson, A.M. Giusti, B.L. Brown, The assessment of frequency estimates of Hinf I-generated VNTR profiles in various ethnic databases, J. Forensic Sci. 39 (1994) 988–1008.
[13] J.M. Hartmann, B.T. Houlihan, R.S. Keister, E.L. Buse, The effect of ethnic and racial population substructuring on the estimation of multi-locus fixed-bin VNTR RFLP genotype probabilities, J. Forensic Sci. 42 (1997) 232–240.
[14] S. Sawyer, A. Podleski, D. Krane, D. Hartl, Fingerprinting Loci do show population differences: comments on Budowle et al, Am. J. Hum. Genet. 59 (1996) 272–274.
[15] P.D. Gill, L.A. Foreman, J.S. Buckleton, C.M. Triggs, H. Allen, A comparison of adjustment methods to test the robustness of an STR DNA database comprised of 24 European populations, Forensic Sci. Int. 131 (2003) 184–196.

[16] D.J. Balding, R.A. Nichols, {DNA} profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, Forensic Sci. Int. 64 (1994) 125–140.
[17] D. Taylor, J.-A. Bright, J. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Sci. Int. Genet. 7 (2013) 516–528.
[18] M.W. Perlin, M.M. Legler, C.E. Spencer, J.L. Smith, W.P. Allan, J.L. Belrose, et al., Validating TrueAllele® DNA mixture interpretation, J. Forensic Sci. 56 (2011) 1430–1447.
[19] R. Puch-Solis, T. Clayton, Evidential evaluation of DNA profiles using a discrete statistical model implemented in the DNA LiRa software, Forensic Sci. Int. Genet. 11 (2014) 220–228.
[20] H. Haned, Forensim: an open-source initiative for the evaluation of statistical methods in forensic genetics, Forensic Sci. Int. Genet. 5 (2011) 265–268.
[21] K. Lohmueller, N. Rudin, Calculating the weight of evidence in low-template forensic DNA casework, J. Forensic Sci. 58 (s1) (2013) s234–59.
[22] D.J. Balding, J. Buckleton, Interpreting low template DNA profiles, Forensic Sci. Int. Genet. 4 (2009) 1–10.
[23] C.R. Hill, D.L. Duewer, M.C. Kline, M.D. Coble, J.M. Butler, U.S. population data for 29 autosomal STR loci, Forensic Sci. Int. Genet. 7 (2013) e82–e83.
[24] T.R. Moretti, L.I. Moreno, J.B. Smerick, M.L. Pignone, R. Hizon, J.S. Buckleton, et al., Population data on the expanded CODIS core STR locifor eleven populations of significance for forensic DNA analyses in the United States, Forensic Sci. Int. Genet. 25 (2016) 175–181.
[25] J. Buckleton, J. Curran, S. Walsh, How reliable is the sub-population model in DNA testimony? Forensic Sci. Int. 157 (2006) 144–148.
[26] J.M. Curran, J.S. Buckleton, C.M. Triggs, What is the magnitude of the subpopulation effect, Forensic Sci. Int. 135 (2003) 1–8.