Forensic Science International: Genetics 29 (2017) 126–144

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles

Tamyra R. Moretti[a,*], Rebecca S. Just[a], Susannah C. Kehl[b], Leah E. Willis[a], John S. Buckleton[c,d], Jo-Anne Bright[c], Duncan A. Taylor[e,f], Anthony J. Onorato[a]

[a] DNA Support Unit, Federal Bureau of Investigation Laboratory, 2501 Investigation Parkway, Quantico, VA 22135, USA
[b] Biometrics Analysis Section, Federal Bureau of Investigation Laboratory, 2501 Investigation Parkway, Quantico, VA 22135, USA
[c] Institute of Environmental Science and Research, Private Bag 92021, Auckland 1025, New Zealand
[d] National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA
[e] Forensic Science South Australia, 21 Divett Place, Adelaide, SA 5000, Australia
[f] School of Biological Sciences, Flinders University, GPO Box 2100 Adelaide, SA, 5001 Australia

ARTICLE INFO

Article history:
Received 28 July 2016
Received in revised form 15 March 2017
Accepted 3 April 2017
Available online 5 April 2017

Keywords:
STRs
DNA Mixtures
Probabilistic Genotyping
Likelihood Ratios

ABSTRACT

The interpretation of DNA evidence can entail analysis of challenging STR typing results. Genotypes inferred from low quality or quantity specimens, or mixed DNA samples originating from multiple contributors, can result in weak or inconclusive match probabilities when a binary interpretation method and necessary thresholds (such as a stochastic threshold) are employed. Probabilistic genotyping approaches, such as fully continuous methods that incorporate empirically determined biological parameter models, enable usage of more of the profile information and reduce subjectivity in interpretation. As a result, software-based probabilistic analyses tend to produce more consistent and more informative results regarding potential contributors to DNA evidence. Studies to assess and internally validate the probabilistic genotyping software STRmix™ for casework usage at the Federal Bureau of Investigation Laboratory were conducted using lab-specific parameters and more than 300 single-source and mixed contributor profiles. Simulated forensic specimens, including constructed mixtures that included DNA from two to five donors across a broad range of template amounts and contributor proportions, were used to examine the sensitivity and specificity of the system via more than 60,000 tests comparing hundreds of known contributors and non-contributors to the specimens. Conditioned analyses, concurrent interpretation of amplification replicates, and application of an incorrect contributor number were also performed to further investigate software performance and probe the limitations of the system. In addition, the results from manual and probabilistic interpretation of both prepared and evidentiary mixtures were compared.

The findings support that STRmix™ is sufficiently robust for implementation in forensic laboratories, offering numerous advantages over historical methods of DNA profile analysis and greater statistical power for the estimation of evidentiary weight, and can be used reliably in human identification testing. With few exceptions, likelihood ratio results reflected intuitively correct estimates of the weight of the genotype possibilities and known contributor genotypes. This comprehensive evaluation provides a model in accordance with SWGDAM recommendations for internal validation of a probabilistic genotyping system for DNA evidence interpretation

1. Introduction

As the sensitivity of forensic DNA typing procedures has improved with the development of better DNA extraction and amplification chemistries and detection instrumentation, more DNA profiles originating from the DNA of two or more individuals are being encountered in forensic casework. The complexity of profile interpretation increases with each additional contributor to

* Corresponding author.
E-mail addresses: Tamyra.Moretti@ic.fbi.gov, trmoretti828@gmail.com (T.R. Moretti).

http://dx.doi.org/10.1016/j.fsigen.2017.04.004
1872-4973/

a mixture, particularly if the DNA contribution is low and therefore subject to stochastic effects (e.g., allele dropout and greater heterozygous peak height variance). Binary decision making can be applied to the interpretation of mixed profiles and has historically been used in many aspects of the analysis of DNA for human identification purposes. This approach has provided an easily applied means of addressing biological phenomena exhibited in PCR-based typing results at short tandem repeat (STR) loci [1–3]. Two primary outcomes are considered in a binary interpretation method. For example, (a) a peak observed in an electropherogram at an expected stutter position is interpreted as either stutter or an allelic peak based on relative height, (b) two allelic peaks are interpreted as having originated from the same or different individuals depending on whether they fall within height variance expectations for heterozygous alleles, and (c) an allele is either used or not used to estimate evidential weight based on whether its height meets an empirically determined stochastic threshold [4].

Such "either-or" determinations, however, can be difficult to make given the characteristics of STR mixture results. The primary criterion used in STR interpretation is peak amplitude, relative to the size and position of the peak in the electropherogram. Yet, the sharing of an allele with that of another contributor and/or with a stutter product renders peak height information less meaningful. Furthermore, locus-specific amplification efficiencies and DNA degradation, which can vary in degree among contributors in a mixture, impact relative peak heights. Also, an allelic component of peaks that qualify as stutter cannot be ruled out when alleles from a minor contributor(s) are in the same general height range as stutter peaks [2]. Together with the possibility of allele dropout, the intricacies of mixture analysis create scientific uncertainty in the determination of possible contributor genotypes and can complicate manual interpretation of mixed DNA profiles.

The use of safeguards (such as a stochastic threshold) was recommended by the Scientific Working Group on DNA Analysis Methods (SWGDAM) to mitigate the uncertainty inherent to binary interpretation of single source, mixed-source and low-level typing results [5]. These safeguards, if applied correctly, tend to limit the usage of profile information and thereby typically lead to more common profile probability estimates, as well as more inconclusive conclusions.

Statistical software programs that incorporate probabilistic interpretation models overcome these limitations and fully utilize the available DNA typing information [6–9]. Probabilistic genotyping refers to the use of software and computer algorithms to apply biological modeling, statistical theory, and probability distributions to infer the probability of the profile from single source and mixed DNA typing results given different contributor genotypes [10]. The software weighs potential genotypic solutions for a mixture by utilizing more DNA typing information (e.g., peak height, allelic designation and molecular weight) and accounting for uncertainty in random variables within the model, such as peak heights (e.g., via peak height variance parameters and probabilities of allelic dropout and drop-in, rather than a stochastic or dropout threshold). Likelihood ratios (*LR*s) are generated to express the weight of the DNA evidence given two user-defined propositions. Probabilistic genotyping software has been demonstrated to reduce subjectivity in the interpretation of DNA typing results and, compared to binary interpretation methods, is a more powerful tool supporting the inclusion of contributors to a DNA sample and the exclusion of non-contributors [11]. Despite the effectual incorporation of higher level interpretation features, though, probabilistic software programs are not Expert Systems as defined under the National DNA Index System (NDIS) Procedures [12]. The DNA typing data and probabilistic genotyping results require human interpretation and review in accordance with the *Quality Assurance Standards for Forensic DNA Testing Laboratories* [13].

The fundamental onus on the forensic laboratory with regard to the analysis of DNA mixtures is to seek to remain current with technological developments and relevant issues and to ensure the reliability of its procedures and usage in casework by properly and thoroughly validating any new method prior to use. The interpretation of complex mixtures in particular requires that the laboratory design and execute thorough, targeted experimental studies as part of its internal validation, recognize limitations revealed through the results, and use the results of validation studies to develop detailed, reliable procedures that can be applied uniformly and consistently among analysts. SWGDAM provides guidelines and the *Quality Assurance Standards for Forensic DNA Testing Laboratories* provide quality assurance requirements for validation [13,14].

We outline here the internal validation of STRmix™ [6,15] at the FBI Laboratory in accordance with *SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems* [10]. STRmix™ is software that employs a continuous model for DNA profile interpretation and genotype determination based on a Markov Chain Monte Carlo (MCMC) sampling method. Using weights assigned to the resultant genotypes or genotype sets, STRmix™ calculates *LR*s, which are the probability of the DNA evidence under two opposing hypotheses referred to as H_1 and H_2. The terms H_1 and H_2 are used in lieu of "Prosecution hypothesis" (H_p) and "Defense hypothesis" (H_d), respectively, given that they are assigned by the scientist, usually without consultation with legal representatives.

A *LR* greater than 1 provides support for a specified person of interest as a contributor to the DNA evidence (H_1), whereas an *LR* less than 1 provides support that the person of interest is not a contributor (H_2). An *LR* of 1 provides no greater support for either proposition. We describe suitable experiments using single source samples and a breadth of mixed DNA samples to meet the recommendations and requirements for internal validation and detail additional testing conducted at the FBI Laboratory to aid in procedural and policy development.

2. Methods

All single source and mixed DNA profiles were generated in-house using DNA samples (collected and typed with informed consent) that were amplified for 27 cycles using the Applied Biosystems AmpFlSTR® Identifiler® Plus PCR Amplification Kit (Thermo Fisher Scientific, Waltham, MA), followed by detection on a 3130xl Genetic Analyzer (Thermo Fisher Scientific). 3130xl data were subsequently analyzed using Applied Biosystems GeneMapper® ID-X version 1.3 (Thermo Fisher Scientific). Protocols and analysis settings (including an analytical threshold of 50 relative fluorescent units, or rfu) previously validated by the FBI Laboratory for casework usage were used to prepare DNA samples, generate DNA typing results and perform preliminary interpretations prior to STRmix™ analysis. Given that STRmix™ version 2.3 models back stutter of one repeat unit, such peaks were retained in all input files for questioned profiles. All other artifact peak labels were removed following FBI Laboratory guidelines, including forward stutter, which is not modeled by the software versions tested.

Laboratory-specific STRmix™ parameters were established using more than 1400 single source Identifiler® Plus profiles of variable quantity and quality (Table S1). Some DNA extracts used for parameter setting were artificially degraded during 90 or 180 seconds of UV irradiation using a Spectrolinker™ XL-1000 UV Crosslinker (Spectronics Corporation, Westbury, NY), with the samples placed, caps open, 2.5 inches from the ultraviolet light

source. Degradation was confirmed following amplification of the irradiated extracts (serially diluted for amplification of 1 ng–0.03 ng template DNA) and demonstration of complete locus dropout, particularly of higher molecular weight alleles, at one or more loci. Per allele stutter ratio (SR) expectations were determined by regressing SR against allele designation and SR against the longest uninterrupted stretch (LUS) of repeats within an allele for some compound or complex repeats (e.g., TH01 9.3 allele has a LUS of 6 repeat units [16]). The maximum allowable stutter ratio was set arbitrarily high at 0.3, or 30% (this parameter is used in the initial assessment of potential alleles, not in stutter modeling, only for run time pick up). For saturated data (peaks generated from high template amounts and/or over-amplification that saturate the camera within the genetic analyzer), an alternate model is invoked within STRmix™ since the relationship of DNA template to peak height is no longer linear. Specifically, the height of a given stutter peak is not determined from the observed (saturated) allele but from an expected allele height based on the proposed model. Based on empirical 3030xl data, a peak height upper limit of 7000 rfu was established for saturation.

The ModelMaker function within STRmix™ uses a MCMC system to analyze a set of laboratory data of known origin in order to determine the distribution of peak height variability specific to the laboratory [17]. This process is used to establish a distribution of expected values for allele, stutter and locus specific amplification variances that are used by STRmix™ in the analysis of data [15,17,18]. By assessing over 700 single source profiles of varying quantity and quality (Table S1) using ModelMaker, peak height variance constant prior distributions for allelic peaks [c^2, Γ(4.2818,1.0671) with a mode of 4.219] and for stutter peaks [k^2, Γ(9.1442,1.1239) with a mode of 7.528] and a mean locus-specific amplification efficiency variance (0.0113) were determined.

The peak height variance constant is explained using the allelic example (the other distributions have similar forms): the distribution for allelic peaks has a mode at 4.219, and 95% of values for allele variance fall between 1.3 and 9.9. These values depict the way that variance changes with peak height (high at low template and low at high template). There is a relationship between peak variance and heterozygote balance (*Hb*) [19,20]. To show this, $\log_{10}(Hb)$ was plotted against average peak height (APH, based on rfu values of alleles at heterozygote loci), and the expected 95% bounds were calculated at $\pm\sqrt{2} \times 1.96 \times \sqrt{\frac{c^2}{APH}}$, where $c^2 = 4.219$. Such a graph (Fig. 1) allows for assessment of the parameters: with 97.6% of the *Hb* data falling within the 95% bounds, allelic variance was demonstrated to be sufficiently optimized.

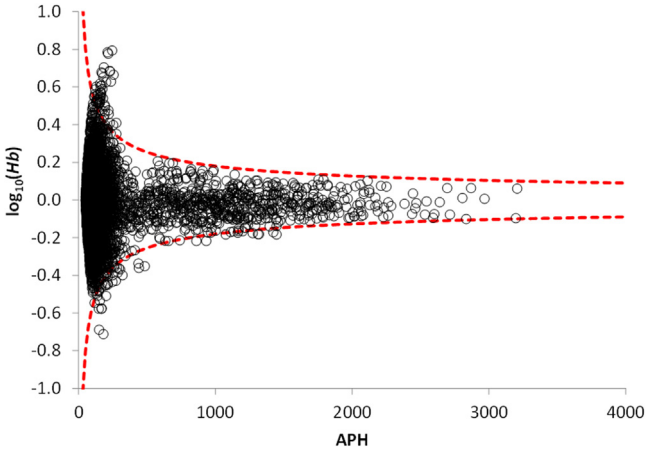

Fig. 1. Plot of $\log_{10}(Hb)$ versus APH for 4125 heterozygote loci from 709 single source ModelMaker profiles. The dashed lines represent the expected 95% bounds and encapsulate 97.6% of all data points.

The drop-in rate was set to zero since no allelic peaks ≥50 rfu were detected at the 16 Identifiler® Plus loci following 27 cycles of amplification of 500 reagent blanks extracted using the EZ1® Advanced XL (QIAGEN Sciences, Inc., Gaithersburg, MD), nor in any amplified DNA sample throughout the study.

To confirm proper calculation of the *LR* by STRmix™, the *LR* for two single source profiles and two two-person mixtures (57 individual loci), where weights determined by STRmix™ equaled both one and less than one for the known contributor profile, were calculated "manually" within Excel. Loci included both heterozygous and homozygous examples, and calculations were undertaken with $\theta = 0.01$ and $\theta = 0$. Setting θ to zero returns the product rule, where *LR* equals:

$2p_ip_j$ for heterozygous loci ($i \neq j$)

p_i^2 for homozygous loci.

When θ >0, the Balding and Nichols formulae [21] (or equations 4.10 from NRC II [22]) are applied. For single source profiles:

$$\frac{2[\theta + (1-\theta)p_i]\left[\theta + (1-\theta)p_j\right]}{(1+\theta)(1+2\theta)} \quad \text{for heterozygous loci} \tag{1}$$

$$\frac{[3\theta + (1-\theta)p_i][2\theta + (1-\theta)p_i]}{(1+\theta)(1+2\theta)} \quad \text{for homozygous loci} \tag{2}$$

where p_i is the allele frequency for allele i, p_j the allele frequency for allele j and θ is the F_{ST} value. The allele frequencies used within equations 1 and 2 are posterior mean frequencies. These are calculated using the following equation:

$$\frac{x_i + \frac{1}{k}}{N_a + 1} \tag{3}$$

where x_i is the number of observations of allele i in a database, N_a is the number of alleles at that locus in the database and k is the number of allele designations with non-zero observations in the database at the locus which the allele, whose probability is being calculated, resides.

Data used for internal validation studies included DNA typing results from serially diluted single source samples (0.031–1 ng) amplified in duplicate. Additionally, a total of 290 two, three, four and five-person mixture profiles, prepared using DNA from thirteen contributors with varying individual template amounts (0.006–3.2 ng) and total template amounts (0.019–4 ng), were created in a range of contributor ratios (Table S2). Two DNA extracts used for mixture preparation were artificially degraded by UV irradiation as described. Some contributor samples were selected based on alleles shared with other contributors or unique to a single contributor (obligate). The resulting profiles variously exhibited inter- and intra-locus peak height variation, complete profile recovery, allele and locus dropout, DNA degradation, additive effects of allele or stutter peak sharing and peak saturation (off-scale peaks). APH for a given contributor was calculated as the average of the heights of all obligate alleles, with undetected obligate alleles assigned a peak height of either 0 or 25 rfu, as noted.

Adjudicated case studies entailed 30 evidentiary mixtures that had been previously developed using Identifiler® Plus and reported as originating from two to five individuals. STRmix™ analyses included the reference DNA profiles of one to four subjects of investigation.

DNA typing results from all single source, mixed and forensic specimens were exported from GeneMapper® ID-X and imported

into STRmix™, where they were interpreted using the established laboratory-specific parameters in STRmix™ version 2.3.06 (http://strmix.esr.cri.nz/). A subset of five-person mixtures was examined using STRmix™ version 2.3.06. Subsequently, the complete set of five-person mixtures was interpreted using STRmix™ version 2.4.02. STRmix™ analyses of single source profiles were conducted according to the propositions:
H_1: The DNA originated from the person of interest
H_2: The DNA originated from an unknown, unrelated individual

For mixtures of N contributors, STRmix™ analyses were conducted according to the propositions:
H_1: The DNA originated from the person of interest and N-1 unknown, unrelated individual(s)
H_2: The DNA originated from N unknown, unrelated individuals

Obligate alleles, template amount and APH were evaluated relative to the STRmix™ results.
Conditional STRmix™ analyses of some mixtures were performed assuming the presence of DNA from a known individual, according to the propositions:
H_1: The DNA originated from the assumed individual, the person of interest and N-2 unknown individual(s)
H_2: The DNA originated from the assumed individual and N-1 unknown individual(s)

Also, for some mixtures, multiple contributors were assessed concurrently in STRmix™, according to the proposition:
H_1: The DNA originated from person of interest 1, person of interest 2 and N-2 unknown individual(s)
H_2: The DNA originated from N unknown individuals.

Where indicated for H_2-true tests, two-hundred non-contributor profiles, which were artificially constructed in Excel™ by randomly sampling alleles from the FBI's U.S. Caucasian allele frequency database [23–26] based on their observed frequency, were analyzed in STRmix™ as persons of interest.

For one, two and three-contributor profiles, STRmix™ analyses were also performed assuming $N+1$ contributors. To assess $N-1$ contributors in a manner that would not result in an exclusion outright (i.e., as would five alleles per locus under the assumption of two contributors), some two-person mixture results were modified to simulate a third contributor with no new alleles. This was done by manipulating the input files directly in Excel™, as follows: in order to avoid creating a 5th allele, a 'child' of the two contributors was constructed by adding 50 rfu to the peaks selected to be shared by parent and child. Two additional mixtures were thus created by increasing peak heights by 100 rfu and 200 rfu. With this approach, a virtual child, present as a trace or minor contributor, represented a third contributor in a mixture that could be interpreted as a two-person mixture.

The default MCMC number of accepts (100,000 burn-in and 400,000 post burn-in) were used to assign weights, which were used in the calculation of *LR*s in STRmix™ using the population genetic model described in Balding and Nichols [21], referencing equations 4.10 in NRC II [22] to correct for population substructure. Statistical calculations were based on allele frequencies from the FBI U.S. Caucasian database following STRmix™ execution in either (a) the standard analysis mode with a θ point estimate of 0.01 or (b) the Database Search mode.

The Database Search function, which produces a total "investigative" *LR* that does not include θ in the calculation, was used to facilitate rapid consideration of a large number of non-contributor propositions (i.e., specificity testing), as well as known contributor propositions (i.e., sensitivity testing) for the mixtures summarized in Table 1 [27]. Some mixtures were interpreted or reinterpreted using the standard mixture analysis mode to develop a lower-bound highest posterior density (HPD) *LR*, in addition to a total "evaluative" *LR*, with the HPD interval set to 99.0% and the N! calculation [28] enabled. For sensitivity and specificity assessment, equations derived from a best fit regression analysis of the investigative *LR* and evaluative HPD *LR* results of the same mixtures (Fig. S1) were applied to the investigative *LR* values to derive HPD *LR* estimates. Based on these estimates, tests that generated a HPD $LR < 1$ for H_1-true propositions and a HPD $LR > 1$ for H_2-true propositions were re-analyzed in STRmix™ for verification purposes. Given that the equation derived for four-person mixtures was suitable for application to the five-person mixtures for purposes of establishing HPD *LR* estimates, re-analyses in STRmix™ to develop HPD *LR*s were not performed for five-person mixtures due to computational constraints.

Precision was evaluated through five repeated interpretations of one, two, three and four-person typing results, with both the minor and major contributor considered the person of interest in H_1. The effect on repeatability by increasing the total number of MCMC accepts from 500,000 to 600,000 and 700,000 and the Random Walk Standard Deviation (RWSD) from 0.005 to 0.01 and 0.02 was evaluated.

STRmix™ analyses of the adjudicated case mixtures were performed in standard analysis mode using the U.S. Caucasian database ($\theta = 0.01$) or, where match probabilities had been reported for Native Americans, a Navajo database [25,26] ($\theta = 0.03$) to generate total *LR*s and HPD *LR*s. For some assessments, the reciprocal of HPD *LR*s between 0 and 1 was calculated.

For both the evidentiary mixtures and a subset of the prepared mixtures, the results of STRmix™ and manual analyses of the same data were compared to evaluate general consistency of the results. Manual profile interpretations and calculation of random match probabilities (RMPs) and combined probabilities of inclusion (CPIs) were performed in accordance with FBI Laboratory standard operating procedures, including usage of a stochastic threshold (200 rfu).

3. Results and Discussion

3.1. Verification of model performance, accuracy and precision

For a small subset of profiles, the *LR* is evident without calculation or can be estimated easily as described in Bright *et al.* [29]. These include single source profiles where the genotype at

Table 1
Summary of mixtures and propositions tested in STRmix™.

Number of contributors	Contributor template range	Total mixture template range	Contributor ratio range	Number of mixtures interpreted	Number of H_1-true propositions tested	Number of H_2-true propositions tested
2	0.006 to 0.9 ng	0.019 to 1 ng	10:1 to 1:1	105	202	22,504
3	0.021 to 1 ng	0.38 to 3 ng	16:1:1 to 1:1:1	64	192	13,620
4	0.05 to 3.2 ng	1 to 4 ng	16:1:1:1 to 1:1:1:1	84	336	17,808
5	0.016 to 1.25 ng	0.25 to 2 ng	10:1:1:2:2 to 1:1:1:1:1	24	120	5,256

each locus is unambiguous, and hence the weight for the correct genotype combination is expected to be 1. As an initial verification of software performance, manually calculated *LR*s for individual loci in a single source sample were identical to the corresponding *LR*s produced by STRmix™, and STRmix™ reported correct genotypes for the known contributor.

As an additional verification of model performance, using a serially-diluted single source sample, *LR*s based on STRmix™ analyses were demonstrated to decrease with template amount (1–0.03 ng) (Fig. S2). As expected, the *LR* progressed from the maximum value for the full profile (attained at ≥0.25 ng) towards log(*LR*) = 0 due to allele dropout as DNA template decreased [11]. For profiles exhibiting higher levels of dropout, simultaneous analysis of amplification replicates in STRmix™ resulted in a higher *LR*. The mass parameter *t* (template, or DNA amount) in the STRmix™ output declined steadily with decreasing peak heights, as expected (Table 2).

Weights generated by STRmix™ were assessed as a measure of the deconvolution process and model performance. Any contributor genotypes deconvoluted manually by a skilled analyst should exhibit intuitively correct, high weights and thereby indicate proper modeling. The most weight is expected to be assigned to genotype sets that correspond to the genotypes of the DNA donors. The result of such an assignment of weight is high levels of support for the inclusion of DNA contributors and exclusion of non-contributors when assessed using *LR*s (depending on profile quality) [11]. Counterintuitive weights are an indication of poor biological modeling or incorrect tuning of the models. Inspection of the STRmix™ output, including weights over the range of single source and mixed DNA profiles, showed the anticipated response to relative template amounts, with lower weights for genotypes that exhibited allelic dropout (Fig. S2).

Mixture proportions were assessed as a final check of model performance, using two-person mixtures constructed in the ratios 1:10, 1:5, 1:3 and 1:1. Mixture proportions obtained from the STRmix™ output (1:10.1, 1:4.0, 1:1.9, 1:1.0) were similar to the targeted proportions. The minor differences from the expected values may be attributable to variability of quantitative PCR results and/or pipetting. A plot of log(*LR*) for each mixture type considering both the major and minor contributors is provided in Fig. S3. The maximum potential log(*LR*s) based on a single source, full profile for each contributor are plotted as horizontal dashed lines. As expected, the *LR* calculated for the major contributor trended from the maximum potential *LR* at 10:1 downward, with the lowest *LR* produced for the 1:1 mixture. The decrease in *LR* occurs where peak heights of major and minor alleles begin to fall within heterozygous peak height variance expectations (here, less than 1:3). The *LR* for the minor contributor did not reach the maximum potential *LR* for a single source profile. At 1:10, some alleles from the minor contributor may be dropped, masked by major contributor alleles, or obscured by stutter peaks from the major contributor. At 1:5, an increase in *LR* for the minor contributor was observed, perhaps as the distinction between minor contributor and major contributor stutter peaks is greater and allele sharing is more evident from assessment of peak heights. At 1:3 and 1:1, the *LR* decreased as the minor and major contributor alleles were less distinct. These data demonstrate that as an analyst's ability to manually deconvolute a mixture decreases, the weights assigned to genotype sets also decrease and are reflected in lower *LR*s.

Table 2
Summary of single source dilution series interpreted in STRmix™ demonstrating the decrease in *LR* with APH and template quantity, *t*.

Input DNA, ng	log(*LR*)	log(HPD)	APH, rfu	*t*
1	18.38	17.71	1366	1975
	18.38	18.11	1241	1935
0.5	18.38	18.12	612	929
	18.38	18.01	618	861
0.25	18.38	18.06	325	484
	18.38	18.01	226	302
0.125	16.99	16.48	138	190
	14.58	14.15	151	194
0.063	9.02	8.82	107	115
	11.70	11.37	101	99
0.031	6.12	5.89	82	61
	4.66	4.35	68	42

For the mixed typing results, *LR*s for both the major and minor contributors varied within one order of magnitude (comparing the minimum and maximum *LR*) across five repeated interpretations, as expected due to MCMC sampling [30]. For mixtures with similar donor contributions (1:2 and 1:1 ratios), greater *LR* variability was occasionally encountered but typically still fell within two orders of magnitude. While increasing the number of MCMC iterations and RSWD might be expected to improve repeatability, no consistent benefit was observed in such trials relative to the minor contributors tested. However, given the conservatism inherent in the use of NRC II equations 4.10 and θ point estimates in STRmix™, along with a HPD interval set to 99.0%, the observed variability in *LR* is within acceptable levels [31–34].

3.2. Sensitivity and specificity studies

Sensitivity of a probabilistic genotyping system refers to the ability of the software to reliably support the presence of a contributor's DNA within the DNA typing results. Sensitivity studies demonstrate the propensity of the system to return support for H_2 for a H_1-true test (i.e., the presence of a true contributor's DNA in the profile is not supported) [10]. It should be noted, however, that failure to detect alleles and/or return support for the presence of a low-level known contributor do not necessarily constitute an error in the analytical process or probabilistic genotyping system. In general, the *LR* for a true contributor should be high but trend to 1 as less typing information to aid interpretation is available and as the number of contributors increases. Information that aids interpretation includes the detection of more alleles from a given contributor, a conditioning profile (e.g., from the donor of an intimate sample) and replicate amplification results from a DNA sample.

Specificity of a probabilistic genotyping system refers to the ability to reliably support the absence of a non-contributor's DNA within the DNA typing results. Specificity studies demonstrate the propensity of the system to return support for H_1 for a H_2-true test [10].

For any mixture study, the proportion of analyses that provide support for a true or false hypothesis (H_1 or H_2) is dependent on the design and quantities of the mixtures tested and should not be used as an indication of error expectations. In this study, the specificity and sensitivity of STRmix™ were tested using a total of 277 two, three, four and five-person mixtures (Table 1). These mixtures exhibited the range of features typically encountered in forensic casework and included many challenging specimens (e.g., with degraded DNA and/or multiple contributors of similar low-level template quantities), specifically to identify potential limitations of the software.

The Database Search mode of STRmix™ was initially used to efficiently execute more than 60,000 comparisons to a 'database' of the true mixture contributor profiles and 200 non-contributor profiles, for a total of 855 H_1-true tests and 59,188 H_2-true tests. The known number of contributors to these mixtures was used for these analyses. Generally, at high template levels STRmix™

returned high *LR*s for true contributors and *LR*s of zero for non-contributors (Figs. S4 through S7).

As contributor template amount decreased and/or contributor number increased, *LR*s trended to 1. The *LR*s for 255 mixtures were converted to HPD *LR* estimates according to the equations in Fig. S1. The HPD probability interval provides a one-sided, lower bound point value based on a *LR* distribution that reflects both population sampling effects and MCMC variability. The HPD *LR* can thus be interpreted and reported in a manner similar to a confidence interval. All propositions generating an HPD *LR* estimate that provided support for the false proposition were re-analyzed for assessment of a STRmix™-generated HPD *LR*. The estimated and

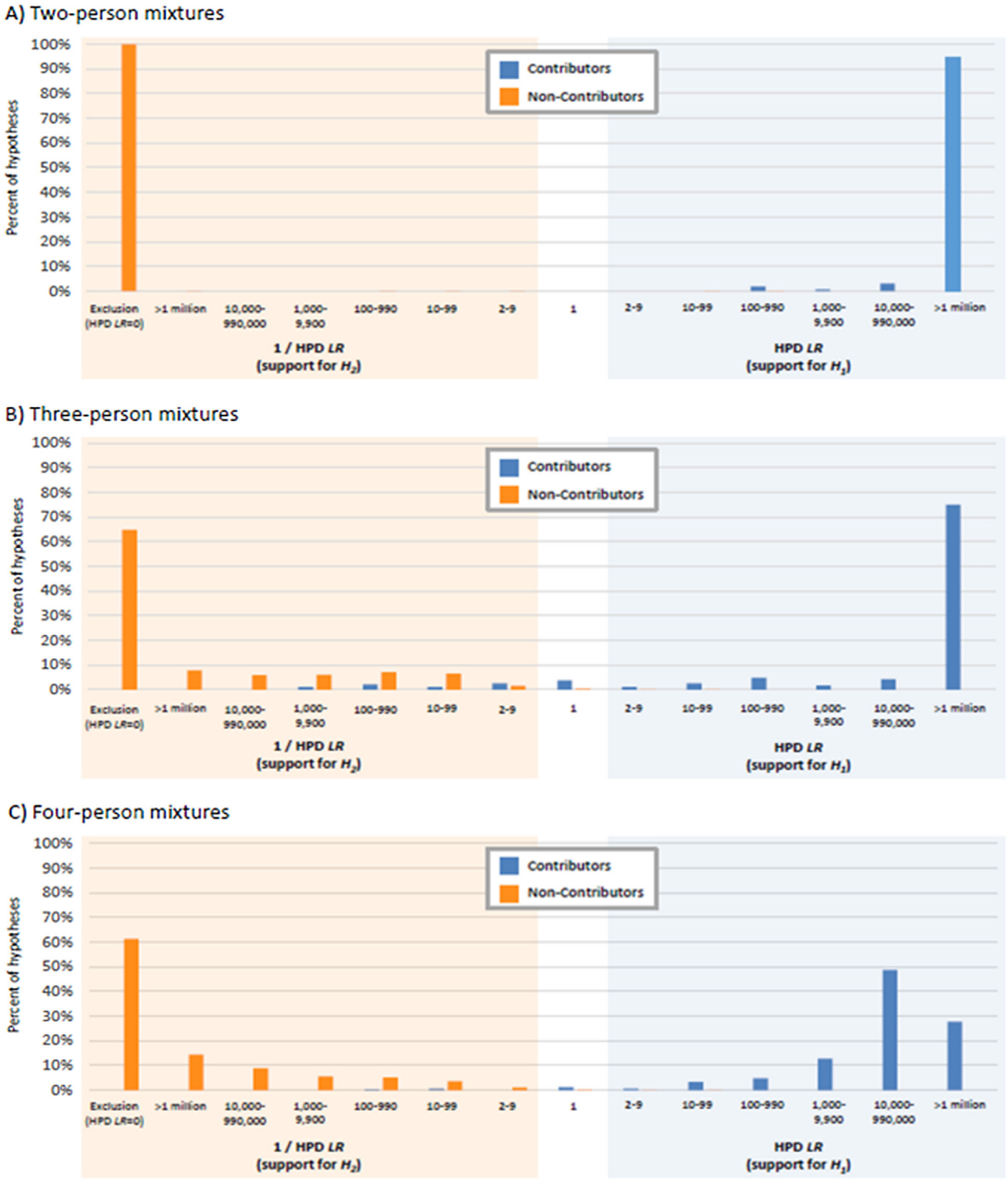

Fig. 2. Sensitivity and specificity of STRmix™ interpretation of two, three, four and five person mixtures. The plots display estimated 99.0% one-sided lower-bound HPD *LR*s for true contributors and known non-contributors proposed as contributors to (A) two-person, (B) three-person, (C) four-person, and (D) five-person mixed DNA profiles. The number of mixtures and propositions tested is provided in Table 1; results for tests using poor quality profiles, as discussed, are not plotted. HPD *LR*s greater than 1 indicate support for the H_1 (contributor) hypothesis, whereas HPD *LR*s less than 1 (here, converted from decimals to positive integers by taking the reciprocal) indicate support for the H_2 (non-contributor) hypothesis. Here, sensitivity is indicated by the percentage of H_1-true propositions (blue bars) within the blue-shaded area (HPD $LR > 1$) versus the percentage in the orange-shaded area (HPD $LR < 1$). Similarly, specificity is indicated by the percentage of H_2-true propositions (orange bars) within the orange-shaded area compared to the blue-shaded area. In total, the histograms in panels A–D represent more than 800 known contributor (H_1-true) propositions, and more than 50,000 non-contributor (H_2-true) propositions. Panel E shows the improvement in specificity for five-person mixtures when interpretations are conditioned on a known contributor.

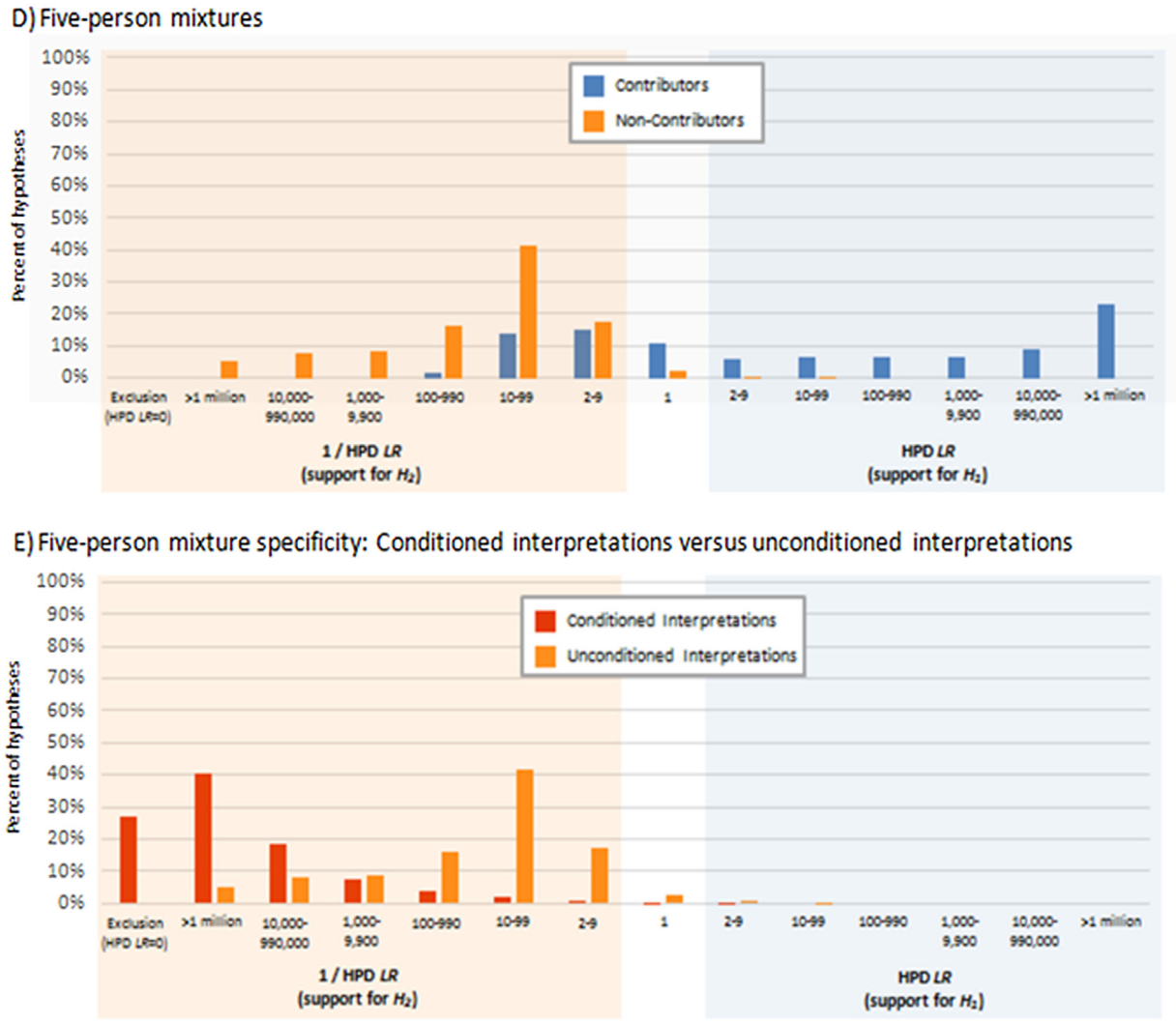

Fig. 2. (Continued)

calculated HPD *LR*s were found to be different by only 0.16 orders of magnitude on average. Overall, the HPD *LR*s indicated very high sensitivity and specificity of STRmix™ analysis (Fig. 2).

For two-person mixtures (Figs. 2 and 3) spanning ratios of 1:1–1:20 and donor contributions of 0.006 ng to 0.90 ng, all H_1-true propositions ($N = 166$) for both contributors resulted in HPD *LR*s > 100. The majority of HPD *LR*s (95%) exceeded 1 million. Nearly all H_2-true propositions for the two-person mixtures produced exclusions (HPD $LR = 0$). All non-zero *LR*s for H_2-true propositions correctly provided support for H_2, with the exception of two that incorrectly provided support for H_1 (discussed below).

For three-person mixtures (Figs. 2 and 3) spanning ratios of 1:1:1–1:1:16 and donor contributions of 0.02 ng to 1 ng, 87% of H_1-true propositions ($N = 192$ total tests) resulted in HPD *LR*s > 1. As with two-person mixtures, the majority of HPD *LR*s (73%) exceeded 1 million. A small portion of H_1-true propositions for the three-person mixtures resulted in HPD *LR*s < 1. These findings occurred when little or no indication of the known contributor was observed in the STR profile due to DNA levels typically <50 pg, allele sharing and dropout; in fact, only two H_1-true propositions produced *LR*s < 1 when the APH and number of obligate alleles for the minor contributors were >0. Five false exclusions were returned for true contributor tests due to poor profile quality (unresolved alleles differing in size by 1 bp; discussed below). As compared to two-person mixtures, fewer of the non-contributor propositions (65%) ($N = 13{,}620$ total tests) resulted in exclusions, though most HPD *LR*s (>99%) were less than 1 and correctly indicated support for H_2. Incorrect H_1-support resulted from 26 H_2-true propositions for the three-person mixtures (discussed below). Given that these analyses were conducted using the known number of contributors to the mixtures, and because several of the incorrect support results occurred with an undetected known minor contributor (APH and number of obligate alleles for the minor contributor equaling 0), the mixtures producing false H_1-support were recalculated in STRmix™ using the observed number of contributors. Based on this re-analysis, 7 of 13,620 non-contributor propositions generated *LR*s > 1 (2–27) (Table 3).

For four-person mixtures (Figs. 2 and 3) spanning ratios of 1:1:1:1 to 1:1:1:16 and donor contributions of 0.05 ng to 3.2 ng, 96% of H_1-true propositions ($N = 336$) resulted in HPD *LR*s > 1, and 27% exceeded 1 million. For H_1-true propositions, one mixture and its replicate amplification produced false-H_2 support for a known minor contributor that was present in both mixtures at 50 pg. Five false exclusions were returned for true contributor tests due to poor profile quality ("saturated" allelic peaks derived from 4 ng contributor template inputs; discussed below). All H_2-true

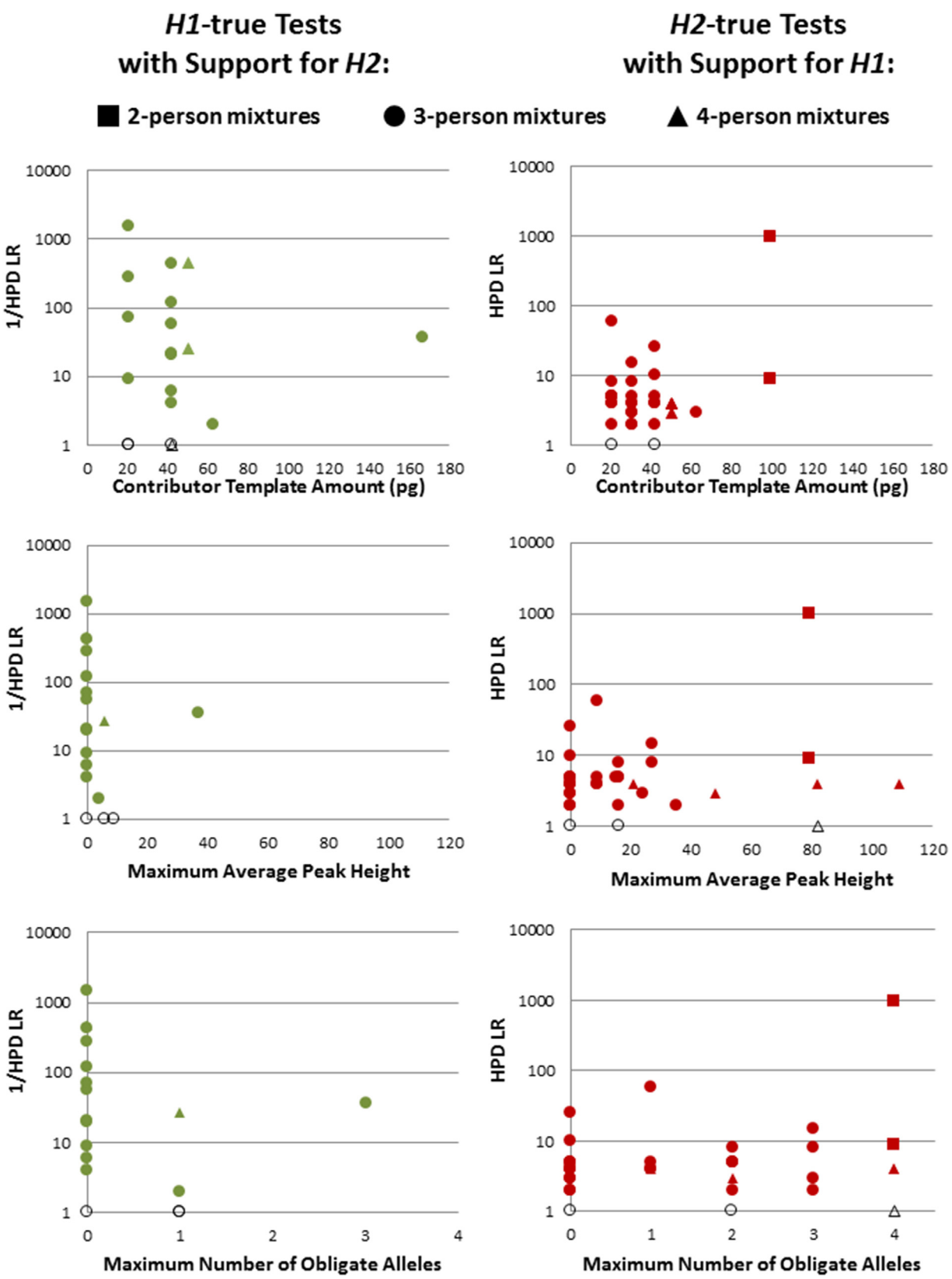

Fig. 3. HPD *LR*s that provided incorrect support for H_1 or H_2. For the contributor and non-contributor tests for which estimated HPD *LR*s indicated incorrect support for H_1 or H_2, respectively, actual HPD *LR*s were calculated using the standard mixture analysis mode and using the known number of contributors to the mixture. The number of mixtures and propositions tested is provided in Table 1. In the H_1-true plots, average peak heights (APH) are based on the rfu values for all obligate (unshared) alleles for the contributor tested, with dropout of an obligate allele captured as rfu = 0. As the known number of contributors was used for the interpretations, rather than the number of contributors that would be inferred via visual review of the electropherogram, in some instances no obligate alleles were detected for a minor contributor, and thus APH also equals zero. All x-axis values for the non-contributor tests are based on the highest value for any minor contributor to the mixture. The open markers in the plots indicate HPD *LR* = 1 (inconclusive) results.

Table 3
HPD *LR* results that failed to support the correct hypothesis.

(A) H_1-True Hypotheses

# of Contributors	Contributor Ratio	Mixture Identifier	Person of Interest	Contributor Template (pg)	Obligate Alleles	Contributor APH	1/HPD *LR*	Inconclusive (HPD *LR* = 1)
3	16:1:1	C.1	Known: J	21	1	9		1
		C.2	Known: J	21	0	0		1
3	8:1:1	D.1	Known: F	42	0	0	20	
		D.1	Known: J	42	0	0	120	
		D.2	Known: F	42	0	0	6	
		D.2	Known: J	42	0	0	21	
3	16:1:1	E.1	Known: J	21	0	0	9	
		E.2	Known: J	21	0	0	280	
3	2:1:1	F.1	Known: B	167	3	37	36	
3	8:1:1	G.1	Known: B	63	1	4	2	
3	8:1:1	H.2	Known: B	42	0	0	57	
		H.2	Known: F	42	0	0	430	
3	16:1:1	J.1	Known: B	21	0	0	71	
		J.2	Known: F	21	0	0	1500	
3	8:1:1	K.1	Known: B	42	1	6		1
		K.2	Known: B	42	0	0	4	
4	16:1:1	O.1	Known: A	50	1	6	26	
		O.2	Known: A	50	0	0	460	

(B) H_2-True Hypotheses

# of Contributors	Contributor Ratio	Mixture Identifier[a]	Non-Contributor Identifier[b]	Contributor Template (pg)	Obligate Minor Alleles	Maximum Minor Contributor APH	HPD LR	Inconclusive (HPD *LR* = 1)
2	10:01	A.1	Rand: 148	100	4	79	980	
		A.1	Rand: 179	100	4	79	9	
3	16:01:01	C.1	Rand: 19	21	1	9	27	
		C.1	Rand: 157	21	1	9	20	
		C.1	Rand: 62	21	1	9		1
3	16:01:01	I.1	Rand: 12	31	0	0	10	
		I.1	Rand: 129	31	0	0	7	
		I.1	Known: Q	31	0	0	2	
		I.1	Known: T	31	0	0	2	
		I.2	Rand: 199	31	0	0	8	
		I.2	Rand: 154	31	0	0		1

STRmix results are shown for all (A) H_1-true hypotheses that returned support for H_2 or HPD *LR* = 1, followed by (B) H_2-true hypotheses that returned support for H_2 or HPD *LR* = 1.

[a] Each mixture is designated with a letter (e.g., C), with replicate amplifications of the mixed DNA extract designated as .1 and .2.

[b] The mixtures were analyzed in STRmix™ with the specified POI.

propositions with the exception of 5 produced *LR*s < 1 (discussed below). Two of the incorrect H_1-support results occurred with the APH and number of obligate alleles for the minor contributor equaling 0. After re-analysis using the observed rather than the known number of contributors, none of the 17,808 non-contributor propositions generated *LR*s > 1 (Table 3).

For five-person mixtures (Figs. 2 and 3) spanning ratios of 1:1:1:1:1–1:1:2:2:10 and donor contributions of 0.02 ng to 1 ng, 58% of H_1-true propositions ($N = 120$) resulted in HPD *LR*s > 1, and 23% exceeded 1 million. This lowered sensitivity is expected given the high number of contributors per sample and very low DNA inputs for some contributors. Yet results from the H_2-true tests indicated that specificity is still high with five-person mixtures: 97% of HPD *LR*s were <1 ($N = 5256$ total tests). Specificity was further improved when the STRmix™ interpretations were conditioned on one known contributor. For these analyses, >99% of H_1-true propositions resulted in HPD *LR*s < 1 ($N = 3200$), and only four resulted in incorrect H_1 support.

Overall across 60,000 STRmix™ analyses, the majority of HPD *LR*s indicating false H_2-support for known contributor propositions ranged from 2 to 2,200 and occurred with 3 or more contributor mixtures that generally exhibited 0 or 1 obligate minor alleles. False H_1-support for known non-contributors was typically demonstrated by *LR*s < 10 (though the highest *LR* was 980) for mixtures of three or more contributors exhibiting 0 to 4 obligate minor alleles. To further explore the results demonstrated with low level DNA contributions, STRmix™ Database Search results for known contributor testing (H_1-true) and non-contributor testing (H_2-true) for the three and four-person mixtures are provided in Fig. 4, a plot of log(*LR*) versus average peak height of the individual contributors for profiles where individual contributions were less than 100 pg. While the HPD *LR*s calculated for these same tests (discussed and tallied above as, for example, incorrect H_1 support for three and four-person mixtures) are lower for true contributors than the *LR*s shown in Fig. 4 (and in some instances returned correct support whereas the Database Search returned incorrect support), the results indicate that STRmix™ was able to correctly separate out true from false contributors even at low levels of DNA. Some low level false positive results were observed, as expected given the complexity of mixtures (i.e., number of contributors and

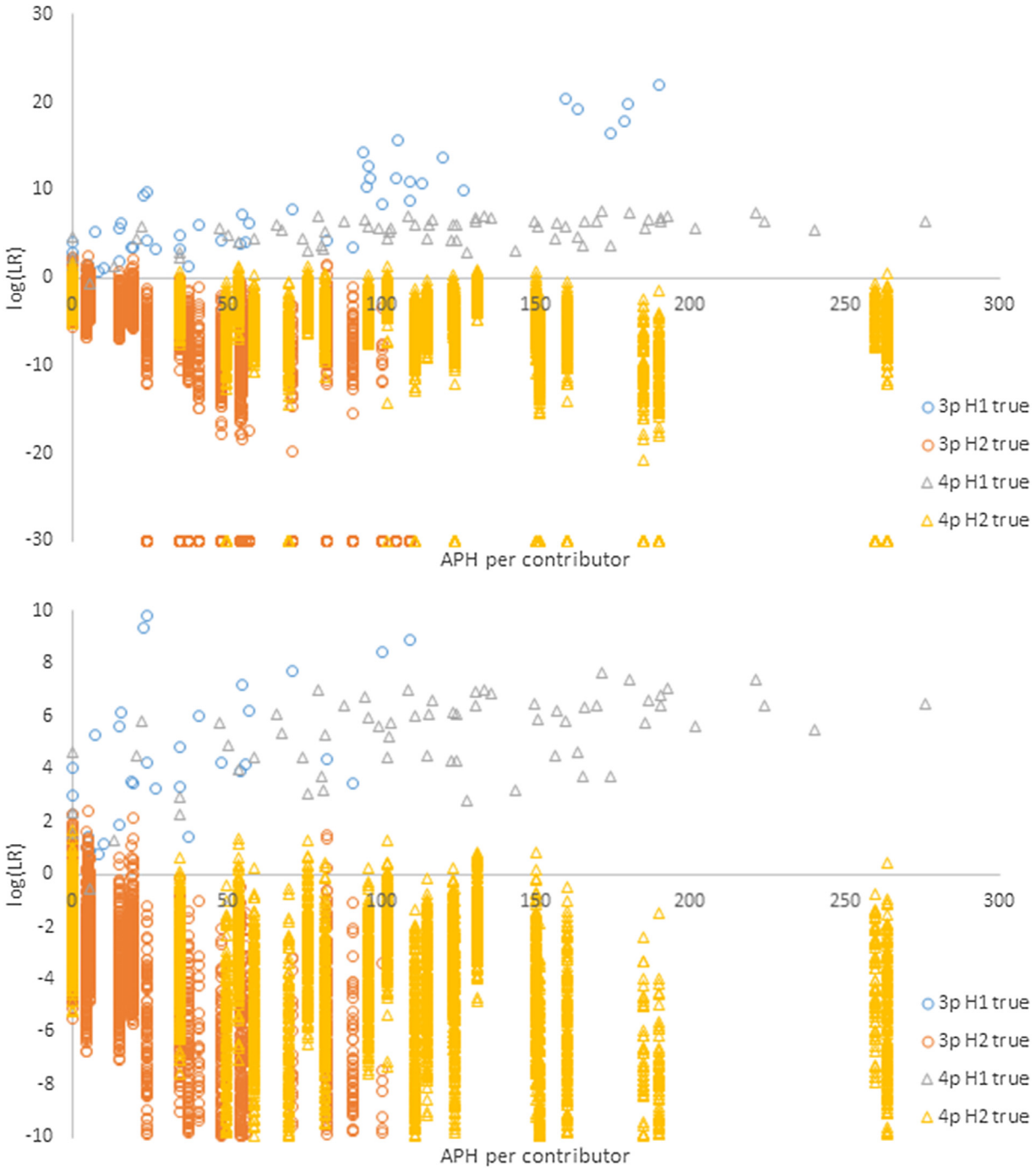

Fig. 4. Plot of log(*LR*) versus average peak height (APH) per contributor for three and four person mixtures where individual contributions are < 100 pg DNA. In the second pane the y-axis has been truncated at +/− log(LR) = 10 in order to better see the results.

low level DNA contributions). To further assess the utility of STRmix™ for discerning minor contributors, the range of HPD *LR* estimates for all minor contributors to the two and three-person mixtures was considered with respect to differing contributor ratios, DNA inputs and, for the two-person mixtures, DNA degradation (Fig. 5). As would be expected due to profile quality, comparisons to two-person mixtures in which one or both DNA extracts were degraded resulted in both lower and more variable HPD *LR* values than when no degradation was present. However, aside from the lowest DNA input tested for the degraded extracts set (0.006 ng), all other minor contributor HPD *LR* estimates exceeded 100,000. A few general trends were also apparent from the three-person mixture plots. As expected, the HPD *LR*s were overall more variable than was observed with the two-person mixtures, and the values generally decreased with decreasing DNA template amounts. But, within three broad template categories representing all minor contributor DNA amounts greater than 0.05 ng (colored in red, blue and purple in Fig. 5), the minor contributor HPD *LR*s increased as the gap between the major to minor contributor input increase. At template amounts below

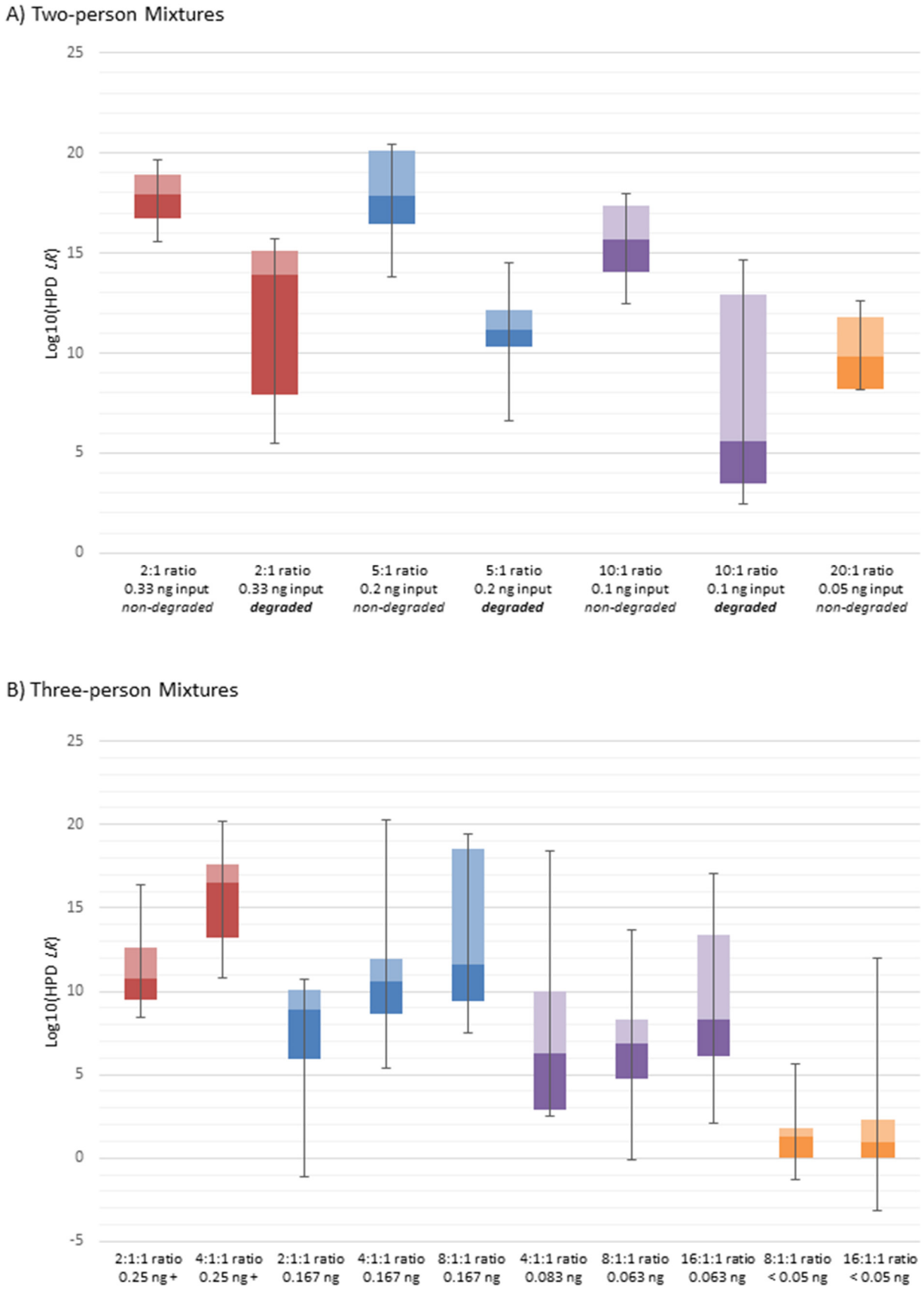

Fig. 5. Range of minor contributor HPD *LR*s. Box and whisker representation of the HPD *LR* estimates for known minor contributors to two-person (panel A) and three-person (panel B) mixtures considering DNA input of the minor contributor, major:minor contributor ratio, and (in the case of the two-person mixtures) DNA degradation of one or both contributor extracts by ultraviolet irradiation.

0.05 ng, most minor contributor comparisons resulted in HPD *LR* estimates indicating some support for H_1, but approximately 25% produced values less than 1.

There were two notable exceptions to the general conclusions from the sensitivity and specificity testing of STRmix™:

(1) Two non-contributors provided HPD *LR*s of 9 and 980 for a single 1:10 mixture (Fig. S8, panel A) developed from two DNA extracts that were each artificially degraded by ultraviolet irradiation. The mixture had a minor contribution of 100 pg. At several loci, the mixture displayed complete dropout of all alleles or all minor contributor alleles. Comparison of the known minor contributor to the mixture resulted in a low HPD *LR* estimate of 320, correctly in support H_1. For the minor contributor (APH = 79 rfu), STRmix™ deconvolution assigned only 5 alleles with >99% probability. Both non-contributor profiles included all 5 of these alleles, as well as several undetectable alleles shared with the major contributor. The absence of all other undetected alleles could be reasonably attributed to dropout. As for alleles that could not be accounted for by the non-contributors, for one of the comparisons (HPD *LR* = 980), a single additional peak was modeled as stutter (9%); for the other comparison (HPD *LR* = 9), two additional peaks were modeled as stutter (9% and 15%). Accordingly, the probabilistic interpretation relative to these non-contributors makes intuitive sense given the mixture results. Under a binary interpretation approach, there is no basis for exclusion of these non-contributors; however, because all minor contributor alleles are below a stochastic threshold (200 rfu), the mixture is not suitable for statistical analysis (i.e., CPI calculation) and would be reported as inconclusive.

A replicate amplification of the same 1:10 mixture was considered with respect to the non-contributor comparison that had produced the HPD *LR* = 980 result. Deconvolution of the duplicate amplification (Fig. S8, panel B) assigned to the minor contributor with >99% probability an allele (not detected in the first amplification) that was inconsistent with the non-contributor, resulting in a HPD *LR* = 0. When the PCR replicates were analyzed simultaneously in STRmix™, the non-contributor was also unambiguously excluded as a potential contributor to the mixture (HPD *LR* = 0). Moreover, the concurrent consideration of both mixed profiles with the true minor contributor as the hypothesized person of interest resulted in a HPD *LR* higher by 2.5 orders of magnitude than when either STR profile was interpreted alone. As a point of reference, repeated STRmix™ analyses of the same mixed profile are generally expected to produce *LR*s with a maximum of a 10-fold (one order of magnitude) difference between the highest and lowest values [30]. Thus, *LR* differences

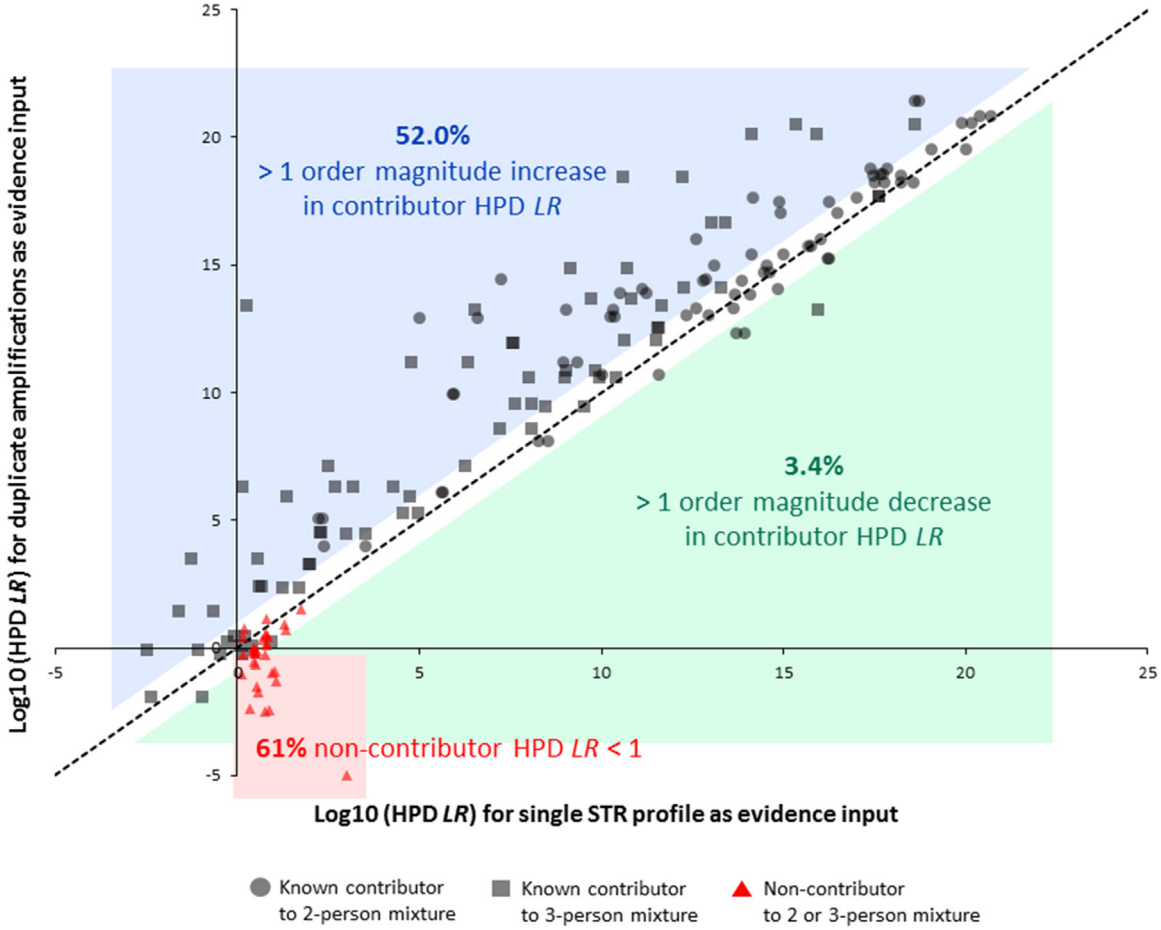

Fig. 6. HPD *LR*s resulting from interpretation of single STR profiles versus PCR replicates. Two-person and three-person mixtures were interpreted with respect to known minor contributors and non-contributors to assess the impact of considering PCR replicates. The data are plotted on a log10 scale, with HPD *LR* = 0 plotted as −5. The dashed line represents where the HPD *LR*s for the single profile (x-axis) and PCR replicates (y-axis) interpretations are equal, and the blue and green-shaded areas represent an increase or decrease (respectively) of one or more orders of magnitude. Only non-contributor propositions that produced any degree of support for inclusion (HPD *LR* > 1) for the single mixed profile interpretations were examined, thus all non-contributor data points (red triangles) are found to the right of the y-axis. The pink-shaded area highlights the non-contributor PCR replicates analyses that produced HPD *LR*s < 1 (support for exclusion).

exceeding one order of magnitude are greater than what would be expected due to variations in the statistical sampling process (MCMC) alone.

(2) One known minor contributor to a 2:1:1 mixture, with template input of 167 pg, resulted in a reciprocal HPD *LR* of 36 in support of H_2. The STRmix™ deconvolution indicated a 1:1:1 contributor ratio, which may in part have been due to incomplete resolution of a shared TH01 allele 9.3 and the minor contributor's TH01 allele 10. The inferred contributor ratio of 1:1:1 resulted in low weights for genotype sets with allelic dropout. However, only half of the minor contributor's obligate alleles were detected, and the APH for the contributor was 37 rfu. As a result, loci with allelic dropout produced *LR*s < 1, resulting in an overall HPD *LR* that incorrectly supported H_2.

The *LR*s from sensitivity and specificity testing in STRmix™ tend towards one as the information in the profile declines, usually correlating with lower template amounts (Figs. 2 and 4 and S4 through S7). Where profiles exhibit stochastic effects and allele dropout, particularly at very low template where few or no obligate alleles for a given contributor are detected, the *LR* for false contributors (as well as true contributors) tends to spread slightly above and below one. Given probabilistic modeling within the stochastic range, *LR*s >1 are expected for some non-contributors. In validation testing, failure to demonstrate false support would indicate that the system is either not functioning properly or has not been queried with sufficiently challenging specimens. In fact, a high *LR* for a simulated non-contributor may even result from a high template single source profile, since simulation of a large number of non-contributor genotypes will eventually produce one that matches the profile. In general, however, the results demonstrate the accuracy of support (i.e., inclusionary or exclusionary), with a much greater probability of excluding a true contributor (6.1% of H_1-true tests) than of including a non-contributor (0.1% of H_2-true tests).

3.3. Concurrent STRmix™ analysis of replicate amplification results

Analyses of duplicate amplifications of two and three-person mixtures ($N = 56$) yielded improvements in the *LR* results. Considering known minor contributors across a broad range of DNA template quantities and contributor ratios, 52% of concurrent analyses of PCR replicates produced a HPD *LR* at least one order of magnitude higher than when an amplification result was analyzed singly, and approximately one third of the time the HPD *LR* increased by two or more orders of magnitude (Fig. 6). Notably, over a range of a HPD *LR*s from10 to 100 million for a single amplification, with the addition of amplification replicate results, the HPD *LR* was never reduced. In fact, 77% of HPD *LR*s increased by

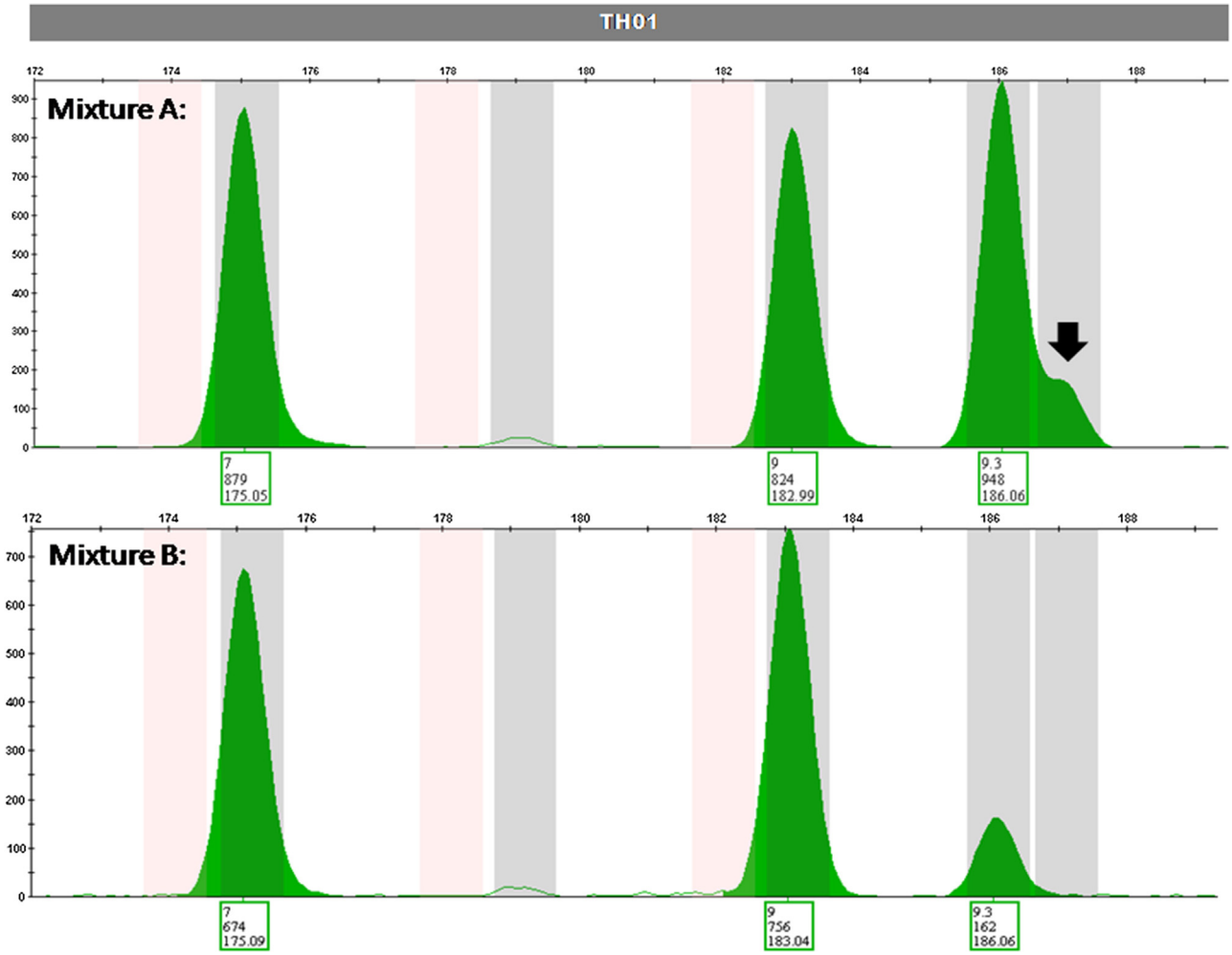

Fig. 7. Typing results at the TH01 locus generating false exclusion results (*LR* = 0) by STRmix™. An unresolved allele 10 (arrow) is apparent for the minor contributor to Mixture A (the multi-locus electropherogram is provided as Fig. S9). Mixture B (Fig. S10) exhibits all alleles at TH01 for its two contributors. STRmix™ unambiguously determined the major contributor genotype to be 7,9, but did not consider 9.3,9.3 as a possible minor contributor type. A *LR* = 0 was returned for the known minor contributor, 9.3,9.3.

more than one order of magnitude (on average by 2.8 orders of magnitude). In addition, 33 non-contributor propositions that had produced some degree of support for inclusion (HPD *LR* > 1) were reanalyzed with the addition of a duplicate amplification. In 61% of these analyses, concurrent interpretation of the PCR replicates resulted in a HPD *LR* less than 1 (Fig. 6). Overall, these data demonstrate that the inclusion of more information in the form of replicate amplifications tends to improve both sensitivity and specificity. However, in practice, replicate amplifications (e.g., with the same template amount) are not routinely performed.

3.3.1. False exclusions

In limited instances, STRmix™ returned readily recognizable false exclusions (*LR* = 0) for known contributors. These had two distinct causes: profile quality and mixture deconvolution issues. False exclusions due to poor profile quality included (a) saturated peaks and (b) electrophoretic failure to resolve major and minor alleles differing in size by 1 bp.

In the present study, false exclusion occasionally occurred in saturated three and four-person mixtures for which the major contribution was 2 ng or more. Additionally, some H_2-true propositions for saturated profiles resulted in inconclusive results (HPD *LR*s = 1). While some saturated peaks may have a nominal effect on *LR*s and weights in some STRmix™ analyses, it is advisable to reprocess the sample (e.g., inject for less time for capillary electrophoresis), given that no useful quantitative information is associated with such peaks and there is a greater potential for elevated stutter, electrophoretic artifacts resulting from amplification of high template amounts and false exclusion with saturated data.

False exclusions also occurred in a three-person mixture presenting with a 3:1:1 ratio. STRmix™ unambiguously determined the major contributor genotype at all loci except for TH01, which exhibited three alleles (Fig. 7, Mixture A; the electropherogram is provided as Fig. S9). At this locus the genotypes of the three known contributors are 7,9 (major contributor), 9.3,9.3 and 9.3,10. On inspection of the electropherogram (Fig. 7), the 10 allele corresponding to the third contributor appears to have been unresolved from the 9.3 during capillary electrophoresis (the 10 allele was thus not 'called' by GeneMapper ID-X and therefore not subsequently analyzed by STRmix™). Based on the modeling of template and degradation for the third contributor, STRmix™ did not consider dropout at TH01 (of, say, the unresolved allele 10), and therefore a *LR* = 0 was returned for this locus only (Table 4, Mixture A). As a general indicator of a potential problem with the data/analysis, review of the STRmix™ output data showed intuitively correct results for all loci except for TH01. Five repeat reinterpretations of the profile with the unresolved allele returned LR = 0 or a very low LR providing incorrect H_2 support (approximately 10^{-4}), as expected given the flawed input data. A replicate amplification of the sample resolved the minor allele 10 and provided a *LR* of 2×10^9. If the unresolved/uncalled peak were not evident in the mixture, given the height of the peaks at this locus, manual interpretation of the profile would also have resulted in an exclusion. However, given the apparent presence of the unresolved allele and the evident interpretation issue, re-injection or potentially re-amplification of the sample is warranted to improve the input data for STRmix™ analysis. Should such repeat processing not provide conclusive typing results, setting the software to ignore a problematic locus is appropriate and, in this instance, correctly produced a non-zero total *LR* as expected based on truth data. With respect to these false exclusions occurring with saturated data and unresolved alleles, STRmix™ performed as expected given the profile quality.

Four instances of a mixture deconvolution problem also presented as an exclusion of a known contributor due to a *LR* = 0 at a single locus (TH01), with the remaining loci producing non-zero *LR*s (Table 4, Mixture B). All four instances occurred with comparison of the minor contributor to mixtures constructed from the same two individuals exhibiting the types 7,9 (major contributor) and 9.3,9.3 (minor contributor) (Fig. 7, Mixture B; the electropherogram is provided as Fig. S10). STRmix™ unambiguously assigned the major contributor type at TH01 (weighted 1.000). However, a weight was not assigned to a genotype combination that, based on the electropherogram, one would reasonably consider a possible minor contributor type: 9.3,9.3. Review of STRmix™ results file in each of these instances indicated that the only genotypes considered were 9,9.3 (weighted 0.919), 7,9.3 (weighted 0.080) and Q,9.3 (weighted 0.001), where Q represents an undetected allele. *LR* = 0 was therefore returned for the known minor contributor. This phenomenon is an infrequent result of the statistical sampling process (MCMC) and occurs when the probability space that includes the true genotype is not sampled. To investigate the *LR* = 0 result, the four mixtures were each deconvoluted ten or more times in STRmix™, and most of the time the repeated analysis (which proceeded from a different

Table 4
LR results for two different mixtures, A and B, that incorrectly returned exclusionary results (underlined) at a single locus.

	Mixture A	Mixture A omit TH01	Mixture A replicate amplification	Mixture B	Mixture B replicate STRmix™
	*LR*s for individual loci				
D8S1179	2.90	2.94	2.11	9.93	9.88
D21S11	25.12	21.76	17.65	79.87	77.69
D7S820	3.41	3.51	3.01	32.88	31.41
CSF1PO	2.91	3.04	3.75	4.71	4.52
D3S1358	1.96	1.99	6.22	9.63	9.63
TH01	0.00	–	11.66	0.00	1.69
D13S317	4.45	4.52	3.93	9.45	9.42
D16S539	2.81	2.48	1.57	2.77	2.87
D2S1338	7.96	8.93	7.55	5.31	4.84
D19S433	1.79	1.67	1.52	10.79	9.93
vWA	2.51	2.41	2.36	2.62	2.65
TPOX	2.16	1.99	1.85	0.83	0.81
D18S51	30.25	27.83	19.92	2.92	3.01
D5S818	2.14	2.16	2.31	20.88	20.59
FGA	6.25	5.41	4.69	11.65	11.17
	*LR*s for the multi-locus profile				
LR total	0	3.55E + 08	2.03E + 09	0	3.44E + 12
Factor of N! *LR*	0	1.85E + 08	9.68E + 08	0	1.72E + 12
HPD *LR*	0	3.72E + 07	3.20E + 08	0	8.10E + 11

random starting seed for the MCMC) produced a non-zero *LR* for the affected locus. In fact, the *LR* = 0 result could sometimes only be replicated by setting the MCMC starting seed to the value that produced the initial false exclusion. Repeated deconvolutions of the profiles were also performed with an increased number of MCMC accepts (up to 5 million) and with a larger random walk standard deviation (RWSD; up to 0.02), but at least one *LR* = 0 result was observed even with these changes to the STRmix™ parameters (data not shown). The false exclusion is likely due to a larger than expected variability in peak heights at this locus that was atypical for the dataset used in establishment of variance parameters. In casework one should attempt to remedy any issue stemming from amplification or electrophoretic phenomena rather than change the interpretation parameters in STRmix™. However, these four particular instances of erroneous STRmix™ results stemmed from the software, not the typing results. Such inconsistency of the electropherogram and STRmix™ results indicates a need for repeating the STRmix™ analysis.

In the examples presented, a weighting of >0.99 was returned for nearly all but the problematic locus. All other loci returning *LR*s > 1 while a single locus has a *LR* = 0 result is a clear indicator that careful review of the typing results and the genotype weights for all loci is merited. Consideration by a skilled analyst of the DNA typing results and all STRmix™ results data is critical in identifying the source of error in the analysis. In all such instances in our study, both the presence of a potential false exclusion and its cause were readily identified by examination of the profile and the STRmix™ results output, with results appearing inconsistent with scientific expectations based on manual review and comparison of the profiles and, in these instances from validation studies, truth data.

3.4. Alternative propositions

Although reference samples for more than one individual may be provided to a laboratory for comparison to evidentiary profiles, the relevant question in the typical legal context relates to a single individual (e.g., complainant, defendant 1 or defendant 2). Nonetheless, STRmix™ analyses were conducted to assess the effects of testing known contributors concurrently, as well as contributors and non-contributors. When two known contributors were assessed concurrently (i.e., for a three-person mixture, H_1 = contributor 1 + contributor 2 + unknown contributor; H_2 = three unknown contributors), the *LR* was additive, approximating the combined *LR*s of testing the contributors individually. However, when one of the two contributors thus assessed was a non-contributor (i.e., for a three-person mixture, H_1 = contributor + non-contributor + unknown contributor; H_2 = three unknown contributors), different outcomes were observed. When a non-contributor tested individually returned a *LR* of zero, the *LR* for concurrent testing with a known contributor was zero. This result is a correct assessment for both individuals considered together but does not appropriately represent the presence of the known contributor in the mixture. When the non-contributor tested singly was not excluded but returned support for H_2 ($0 < LR < 1$), the results of concurrent testing with a known contributor varied: (a) *LR* of zero, or (b) an additive *LR* > 1. The latter indicates incorrect support for H_1 given that one of the individuals tested concurrently is a non-contributor.

Where appropriate, mixture interpretation can be conditioned on the assumption that the DNA of a given individual (e.g., donor of a vaginal swab) is present in the sample. The use of such information has been shown to increase the *LR* for H_1-true propositions and reduce the *LR* for H_2-true propositions [11]. A total of 94 two, three, four and five-person mixtures that had been analyzed in STRmix™ with no contributor assumed were reinterpreted in STRmix™ to test the effect of conditioning the analysis on a known contributor to the mixture (Fig. S11). For all analyses, a known contributor to the mixture was tested as a person of interest (H_1-true). In general, conditioning the analysis improved the *LR* when the DNA input amounts for the assumed contributor and the person of interest were similar [e.g., the minor contribution to the mixture is at least 50% (red data points) as compared to at most 20% (green data points). As an exception, conditioning the analysis of 1:1 mixtures produced no substantial difference in *LR* when the DNA of one of two contributors was degraded, since the differences in peak heights due to degradation enabled resolution of the two genotypes. In addition, *LR*s increased when both the assumed contributor and person of interest were minor contributors, with the magnitude of the effect decreasing as the number of individuals in the mixture increased. For example, for three-person mixtures with a minor person of interest, the *LR* increased by 4.2 orders of magnitude upon conditioning on another minor contributor, but for four-person mixtures, the benefit was reduced to 2.6 orders of magnitude. By contrast, if the person of interest was a minor contributor, the *LR* rarely improved when the analysis was conditioned on a definitive major contributor, and vice versa (green data points). This result is intuitive: conditioning on a clear major contributor, for example, does not typically improve resolution of the minor component(s).

For the five-person mixtures, the same general trends in the data were most apparent when “trace” contributors (in this instance, defined as individuals with DNA inputs of 0.18 ng or less who also represented less than 10% of the total DNA load for the mixture) were distinguished from non-trace contributors. In these mixtures, *LR*s increased to the greatest degree and most consistently when the assumed contributor and person of interest were both (a) trace contributors, or (b) non-trace contributors. Conditioned analyses of 1:1:1:1:1 mixtures, as well as analyses conditioned on a trace contributor in which a non-trace conditioner was the person of interest (or vice versa), increased the *LR* by less than one order of magnitude on average.

3.5. Incorrect number of contributors

Within STRmix™ the number of contributors to a DNA profile must be assigned prior to analysis. The true number of contributors to an evidence profile, however, is unknown. Uncertainty in determination of the number of contributors may increase due to artifacts, stutter percentages that exceed expectation, allele dropout and, particularly with higher contributor numbers, allele sharing. Given a contributor number of N and the assumption of an additional contributor ($N+1$), STRmix™ adds the additional (unseen) contributor at trace levels which, when considered with the true trace contributor, diffuses the genotype weights and can either return a false exclusion or lower the *LR* of a true contributor [27,35]. The *LR* of the major contributor is not appreciably different when $N+1$ contributors are assigned.

In the present study, the effect on the *LR* of assuming an incorrect number of contributors was tested by both increasing ($N+1$) and decreasing (N-1) the number. For the $N+1$ tests, 27 total one, two and three-person profiles were interpreted as originating from two, three and four individuals, respectively. The *LR* was calculated using the Database Search function for both true contributors and 200 non-contributors, which were converted to HPD *LR* estimates as previously described.

For true contributors (H_1-true), the majority of HPD *LR*s under the assumptions of N and $N+1$ contributors were similar (within one order of magnitude); for 13% of the analyses, the HPD *LR*s decreased by more than one order of magnitude (Fig. S12). With regard to non-contributors (H_2-true), 89.6% were excluded (HPD *LR* = 0) under assumption of the correct number of contributors,

and the remainder (excepting the false H_1-support instances noted above) returned HPD *LR*s < 1. Under the incorrect assumption of an additional contributor, only 5.3% of non-contributors were excluded outright, though overall 94.3% returned HPD *LR*s < 1.; 4.0% of such analyses were inconclusive (HPD *LR* = 1), and only 1.7% of H_2-true tests analyzed with $N+1$ contributors returned incorrect support for H_1. The vast majority (91.9%) of results with incorrect H_2 were HPD *LR*s ≤ 10; only one result out of 5,716 $N+1$ analyses was >100 (HPD *LR* = 126).

STRmix™ generates a *LR* = 0 if contributor number is underestimated since any "extra" allele cannot be accounted for under the assigned number of contributors. Therefore, as a means of examining the impact of assuming too few contributors without returning an exclusion outright, three mixtures were artificially created from a two-person mixture (1:5 contributor ratio) by adding an additional contributor without adding any new (exclusionary) alleles. The "third" contributor was constructed as if a child of the two true contributors, sharing alleles at all loci, by increasing the rfu values of the alleles by approximately 50 rfu, 100 rfu or 200 rfu per mixture. Each artificially constructed three-person profile was analyzed as a two-person mixture and compared with the true contributors and 200 non-contributors. The resulting *LR*s for the major or minor contributor were not affected by the addition of a third contributor at any of the three average peak heights. All non-contributors resulted in exclusions (*LR* = 0)

3.6. Manual and probabilistic interpretation of the same mixed profiles

To assess general consistency between manual interpretation and STRmix™ analysis, a set of mixtures prepared over a range of contributor ratios and DNA template amounts was analyzed using both methods. Where a person of interest was not excluded as a possible contributor to a mixture based on manual interpretation, STRmix™ analysis demonstrated support for H_1, with HPD *LR*s ranging from 8.7×10^8 to 1.8×10^{19} (Table S3). Where loci or entire profiles were manually disqualified for CPI calculation following application of a stochastic threshold of 200 rfu, STRmix™ results varied: HPD *LR*s for true contributors ranged from 2,200 to 250 billion in support of H_1, and for non-contributors were 0 (exclusion), 850 trillion in support of H_2 and 2 in support of H_1. The latter result occurred in a three-person mixture noted above as providing incorrect support for H_1.

A comparative examination of mixtures from 30 authentic forensic specimens was also performed (Fig. 8). For reference samples included by manual interpretation as potential major

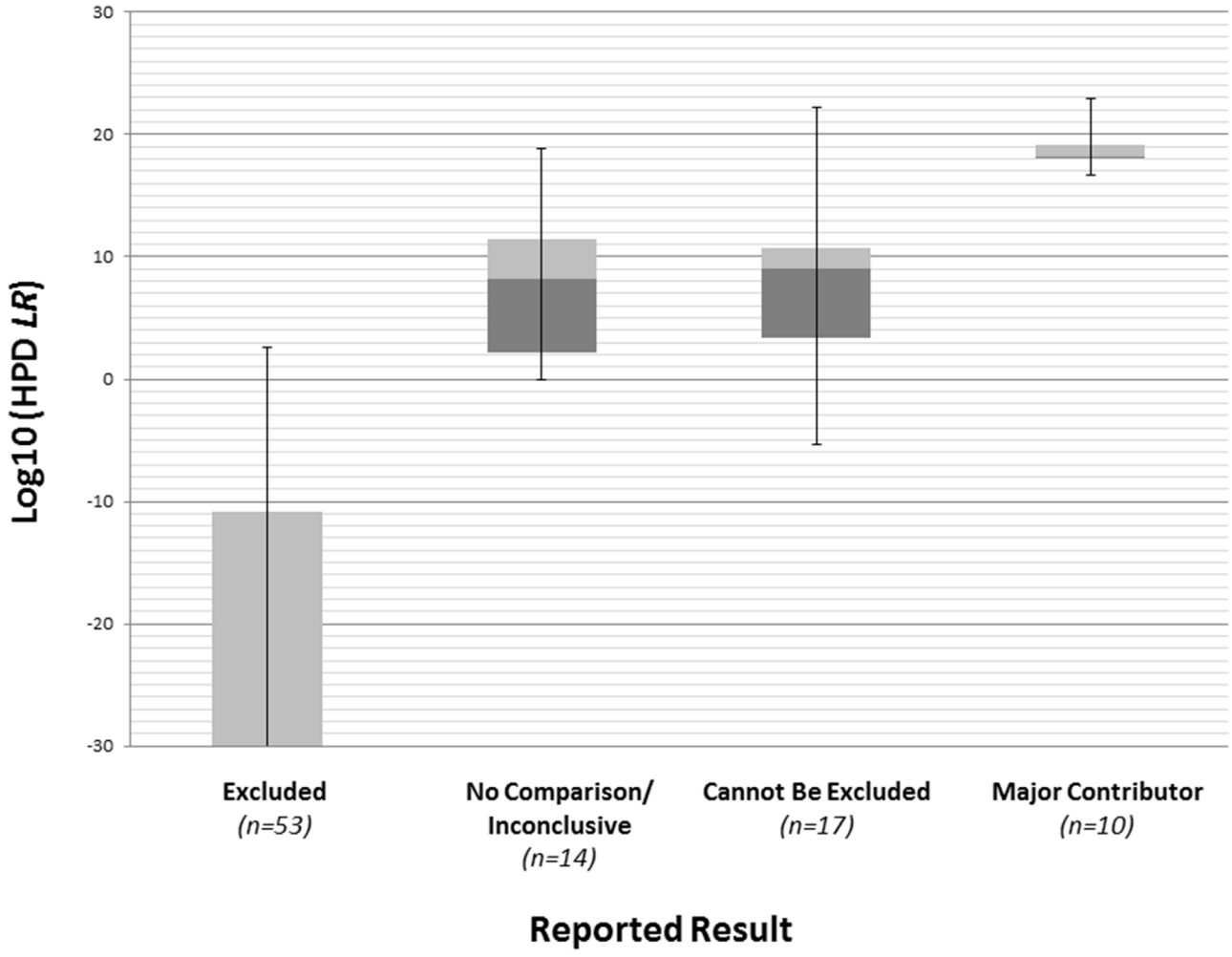

Fig. 8. Manual versus probabilistic interpretation of mixed profiles from evidentiary specimens. Box and whisker representation of the HPD *LR*s calculated for 94 total reference genotype comparisons to Identifiler® Plus profiles developed from 30 authentic forensic specimens from adjudicated FBI cases. The manual interpretation results (x-axis) were categorized based on how the person of interest comparison was reported: excluded as a contributor to the DNA evidence, inconclusive (did not meet standards for match comparison), cannot be excluded (potential contributor to the mixed DNA profile reported with a CPI statistic), or major contributor (deduced single-source profile reported with a RMP statistic). 99.0% lower-bound HPD *LR*s (y-axis) are plotted using a log scale, with HPD *LR* = 0 plotted as −30. All 30 evidentiary profiles indicated mixtures of DNA from two or more individuals.

contributors and previously reported with a RMP statistic, the STRmix™ results produced HPD *LR*s in excess of 1×10^{16}. Considering 53 manual exclusions: in 51 instances the probabilistic interpretation also supported exclusion, in one instance the HPD *LR* was 1 (denoting an inconclusive result), and in the last instance the HPD *LR* was 410 in support of inclusion. In this last instance, the probabilistic interpretation was conducted assuming four contributors, and STRmix™ indicated the reference as one of three trace contributors, comprising just 5% of the DNA load. Examination of the evidence electropherogram and reference genotype (Fig. S13) revealed a peak disqualified as stutter in the manual interpretation that was modeled by STRmix™ as an allele for a trace contributor that exhibited dropout at multiple loci.

Among the 94 total hypotheses tested for the evidentiary specimens, reference samples were manually designated as potential contributors ("cannot be excluded") in 17 instances (involving ten distinct mixtures; Table 5). These comparisons were previously reported with CPI statistics, which ranged from a low of 1 in 1 to a high of 1 in 30,000. Where the statistical estimate was 1 in 1, the percentage of the population that would be included as possible contributors (75%–98%) was also reported (Table 5). For STRmix™ analyses of these ten mixtures, nine were analyzed as originating from four persons and one was analyzed as originating from three persons. For all potential contributors previously reported with a CPI of 1 in 2 or greater, the STRmix™ results supported inclusion, with HPD *LR*s ranging from 2,600 to 16 sextillion. Considering the six reference comparisons for which the reported CPI was 1 in 1, the probabilistic interpretations produced HPD *LR*s less than 1 (denoting support for H_2) in four instances. This is not a surprising result given the complexity of the mixtures and the weak statistical support calculated for the manual inclusions. For two of these four instances, both involving the same mixture (C1Q10, compared to reference samples C1K10 and C1K18), the total and HPD *LR* values from the STRmix™ output indicated a far larger HPD interval than is typical (Table 5). Given the approximate 4:4:1:1 contributor ratio for this mixture, it seems likely that the similarities in DNA loads may have resulted in MCMC uncertainty, creating the wide HPD interval, which in turn accounts for the strength of the support for H_2. The remaining two CPI results of 1 in 1 provided support for H_1.

For the evidentiary mixtures that were deemed inconclusive by manual interpretation, STRmix™ produced wide-ranging HPD *LR*s, as with the prepared mixtures, from 1 to greater than 1×10^{18} in support of H_1 (Fig. 8). These results indicate that a fully continuous probabilistic method enabling usage of more profile information and modeling features in STRmix™ yields more refined conclusions for some mixed DNA profiles as compared to a binary interpretation method.

4. Conclusions

The internal validation studies described herein involved the examination of more than 300 autosomal STR profiles, derived from one to five contributors and representing a wide range of contributor ratios and DNA template amounts. The probabilistic interpretations using laboratory-specific parameters totaled more than 800 known contributor propositions, nearly 60,000 non-contributor tests, and nearly 100 reference sample comparisons to mixed profiles developed from authentic forensic specimens. Overall, the study results demonstrate that STRmix™ software

Table 5
Manual interpretation statistics compared to STRmix™ results for casework comparisons reported as "cannot be excluded".

		Manual Interpretation		STRmix Interpretation		
Forensic Sample Identifier	Reference Sample Identifier	Reported CPI	Percentage of Population Included	Assigned Number of Contributors	HPD *LR* or 1/HPD *LR**	*Orders of magnitude difference between total LR and HPD LR*
C1Q10	C1K10	1 in 1	75%	4	210,000*	*5.1*
	C1K14	1 in 1	75%	4	85 million	*2.8*
	C1K18	1 in 1	75%	4	2,000*	*8.2*
C1Q14	C1K6	1 in 4	n/a	4	8.1 billion	*0.7*
C1Q15	C1K6	1 in 18	n/a	4	27 million	*0.6*
	C1K10	1 in 18	n/a	4	51 billion	*0.8*
	C1K14	1 in 18	n/a	4	4.4 billion	*0.5*
C1Q19	C1K14	1 in 33	n/a	3	1.1 billion	*0.4*
C1Q22	C1K14	1 in 1	98%	4	2*	*0.7*
C1Q23	C1K14	1 in 1	98%	4	370 trillion	*0.8*
	C1K18	1 in 1	98%	4	2*	*0.6*
C1Q26	C1K6	1 in 9	n/a	4	140 million	*0.7*
	C1K10	1 in 9	n/a	4	52 billion	*0.6*
C1Q39	C1K14	1 in 2	46%	4	1.6 trillion	*1.0*
C1Q40	C1K14	1 in 2	64%	4	16 sextillion	*0.8*
C3Q2	C3K10	1 in 30,000	n/a	4	2,600	*0.3*
	C3K11	1 in 30,000	n/a	4	230 trillion	*0.4*

Combined probabilities of inclusion (CPIs), population percentages and *LR*s are based on U.S. Caucasian or Navajo population sample allele frequencies and theta values of 0.01 or 0.03 (respectively), as appropriate based on the case details. Population percentages were included in the FBI Laboratory Report of Examination for all reported populations if the CPI statistic was 1 in 1 for any of the reported populations.

* Asterisks denote conversion of HPD *LR*s that were less than 1 to a positive integer (1/HPD *LR*) to convey the degree of support for the H_2 hypothesis on the same scale as HPD *LR*s >1 results. All HPD *LR* and 1/HPD *LR* values were truncated to two significant figures.

n/a = not applicable (percentage of population included is only provided for "1 in 1" CPIs)

performed as expected. With very few exceptions, genotype weights were intuitively correct, and the statistical results were consistent with scientific expectations. Across multiple studies, the data showed that as the informative content of a profile increased such as with higher DNA template amounts, greater disparity in contributor ratios, and simultaneous consideration of PCR replicates, *LR*s increased for true contributors and decreased for known non-contributors.

When a 99.0% one-sided lower-bound HPD *LR* value was used to assess the STRmix™ results for the two, three, four and five-person prepared mixtures, the software proved to be appropriately sensitive and specific. Across more than 60,000 tests, 93.4% of true contributors produced HPD *LR*s supporting inclusion, and greater than 99.9% of non-contributors resulted in HPD *LR*s supporting exclusion. Specificity with five-person mixtures was further increased by conditioning the interpretation on a known contributor. In all cases where non-contributor comparisons generated HPD *LR*s > 1, the results were consistent with scientific expectations given the mixture quality and complexity (e.g., degradation, allelic dropout) and the number of contributors. A few exclusionary results (*LR* = 0) for known contributors occurred due to poor profile quality (e.g., inadequate capillary electrophoresis resolution) and when MCMC sampling failed to identify the true genotype combination for a single locus. In all instances, the cause of the unexpected results could be deduced upon review of the STRmix™ output files in relation to the mixture electropherogram and resolved by a repeat STRmix™ analysis with or without modifications (according to error type). Taken together, the results of the various studies (a) aptly demonstrate the reliability of the STRmix™ software in terms of sensitivity and specificity when laboratory-specific parameters are employed for analyses and (b) underscore the importance of analyst review of both the DNA typing and probabilistic genotyping results.

These studies establish that STRmix™ version 2.3.06 is fit for purpose for the interpretation and statistical assessment of single source profiles and mixtures originating from two, three, four and five individuals. To convey the statistical weight and aid comprehension of the reported statistical results, which in this study ranged from *LR*s of 0 to approximately 10^{27}, the *LR* may be accompanied by a qualitative description of the degree of support for the H_1 or H_2 hypothesis [10]. The FBI Laboratory reports the HPD *LR* for Identifiler® Plus typing results with a verbal expression of evidential strength as recommended by the European Network of Forensic Science Institutes (ENFSI) [36] and founded by the Association of Forensic Science Providers [37]; HPD *LR*s of 0 are reported as exclusions, and HPD *LR*s of 1 are reported as uninformative.

The implementation of a fully continuous probabilistic genotyping system on December 1, 2015 represents a major step forward in the interpretation of autosomal STR data at the FBI Laboratory. As evidenced by the comparative examinations of prepared mixtures and evidentiary profiles from prior FBI cases, the conclusions derived from the results of probabilistic genotyping can be expected to align with properly applied historical methods. The probabilistic approach used by STRmix™ greatly increases the information that can be used to deconvolute mixtures and estimate evidentiary weight, showing distinct advantages with mixtures with three or more individuals and low-level contributors. Our analysis of findings supports that STRmix™ reliably applies suitable biological modeling and statistical methods, is sufficiently robust for usage with forensic-type specimens and, as a probabilistic genotyping system, represents a vital advancement in the field of human identification testing.

Acknowledgements

The authors would like to thank the many individuals who provided scientific or technical support for this work, including: Jerrilyn Conway, Jade Gray, Jeremy Fletcher and Baxter Cohen of the DNA Casework Unit, FBI Laboratory; Jill Smerick and Jodi Irwin of the DNA Support Unit, FBI Laboratory; Laura Russell, Catherine McGovern and Stuart Cooper of the Institute of Environmental Science and Research; Jeffrey Monaghan of Robotech Science, Inc.; and Luigi Armogida of NicheVision. This work was supported in part by Award No. 2014-DN-BX-K028, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. This work was also supported in part through the FBI's Visiting Scientist Program, an educational opportunity administered by the Oak Ridge Institute for Science and Education (ORISE). The opinions and assertions presented herein are those of the authors and should not be construed as official or as reflecting the views of the U.S. Department of Justice, the U.S. Department of Commerce, the U.S. Department of Energy or the U.S. Government. Certain commercial equipment, instruments, materials, suppliers and software are identified to specify experimental procedures and foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, the Federal Bureau of Investigation or any branch of the U.S. Government, 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.2017.04.004.

References

[1] T. Moretti, A.L. Baumstark, B.S. Defenbaugh, K.M. Keys, J.B. Smerick, B. Budowle, Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples, Journal of Forensic Sciences 46 (647) (2001) 60.

[2] B. Budowle, A.J. Onorato, T.F. Callagham, A.D. Manna, A.M. Gross, R.A. Guerreri, et al., Mixture Interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework, Journal of Forensic Sciences 54 (2009) 810–821.

[3] P. Gill, J. Buckleton, Commentary on: Budowle B, Onorato AJ, Callaghan TF, della Manna A, Gross AM, Guerrieri RA, Luttman JC, McClure DL. Mixture interpretation: Defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. J Forensic Sci 2009;54 (4):810-21, Journal of Forensic Sciences 55 (2010) 265–268.

[4] F.R. Bieber, J.S. Buckleton, B. Budowle, J.M. Butler, M.D. Coble, Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion, BMC Genetics 17 (2016) 125.

[5] Scientific Working Group on DNA Analysis Methods (SWGDAM), SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories, (2010) . [Accessed] 23 November 2014 http://www.fbi.gov/hq/lab/html/codis_swgdam.pdf.

[6] D. Taylor, J.-A. Bright, J. Buckleton, The interpretation of single source and mixed DNA profiles, Forensic Science International: Genetics. 7 (2013) 516–528.

[7] 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, Journal of Forensic Sciences. 56 (2011) 1430–1447.

[8] D.J. Balding, J. Buckleton, Interpreting low template DNA profiles, Forensic Science International: Genetics. 4 (2009) 1–10.

[9] P. Gill, H. Haned, A new methodological framework to interpret complex DNA profiles using likelihood ratios, Forensic Science International: Genetics. 7 (2013) 251–263.

[10] Scientific Working Group on DNA Analysis Methods (SWGDAM), Guidelines for the validation of probabilistic genotyping systems, (2015) . [Accessed] 3 October 2016. http://media.wix.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515. pdf.

[11] D. Taylor, Using continuous DNA interpretation methods to revisit likelihood ratio behaviour, Forensic Science International: Genetics 11 (2014) 144–153.
[12] FBI Laboratory, National DNA Index System (NDIS) Operational Procedures Manual, Version Effective January 1, 2015, (2015) . [Accessed] April 4, 2016 https://www.fbi.gov/about-us/lab/biometric-analysis/codis/ndis-procedures-manual/view.
[13] FBI Quality Assurance Standards for Forensic DNA Testing Laboratories, (2011) . [Accessed] 19 November 2014 http://www.fbi.gov/about-us/lab/biometric-analysis/codis/qas-standards-for-forensic-dna-testing-laboratories-effective-9-1-2011.
[14] Scientific Working Group on DNA Analysis Methods (SWGDAM), Guidelines for validation of DNA analysis methods, (2016) . [Accessed] 1 February 2017 https://media.wix.com/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf.
[15] J.-A. Bright, D. Taylor, J.M. Curran, J.S. Buckleton, Developing allelic and stutter peak height models for a continuous method of DNA interpretation, Forensic Science International: Genetics 7 (2013) 296–304.
[16] C. Brookes, J.-A. Bright, S. Harbison, J. Buckleton, Characterising stutter in forensic STR multiplexes, Forensic Science International: Genetics 6 (2012) 58–63.
[17] D. Taylor, J. Buckleton, Bright J-A: Factors affecting peak height variability for short tandem repeat data, Forensic Science International: Genetics 21 (2016) 126–133.
[18] D. Taylor, J.-A. Bright, C. McGoven, C. Hefford, T. Kalafut, J. Buckleton, Validating multiplexes for use in conjunction with modern interpretation strategies, Forensic Science International: Genetics 20 (2016) 6–19.
[19] J.-A. Bright, E. Huizing, L. Melia, J. Buckleton, Determination of the variables affecting mixed MiniFiler DNA profiles, Forensic Science International: Genetics 5 (2011) 381–385.
[20] J.-A. Bright, J. Turkington, J. Buckleton, Examination of the variability in mixed DNA profile parameters for the Identifiler multiplex, Forensic Science International: Genetics 4 (2009) 111–114.
[21] D.J. Balding, R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, Forensic Science International 64 (1994) 125–140.
[22] National Research Council II, National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence, National Academy Press, Washington, D.C, 1996.
[23] B. Budowle, T.R. Moretti, A.L. Baumstark, D.A. Defenbaugh, K.M. Keys, Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US, Caucasians, Hispanics, Bahamanians, Jamaicans and Trinidadians, Journal of Forensic Sciences 44 (1999) 1277–1286.
[24] B. Budowle, P. Collins, P. Dimsoski, C. Ganong, L. Hennessy, C. Leibelt, et al., Population data on the STR loci D2S1338 and D19S433, Forensic Sci Communications (2001) 3.
[25] B. Budowle, B. Shea, S.J. Niezgoda, R. Chakraborty, CODIS STR. loci data from 41 sample populations, Journal of Forensic Sciences 46 (453) (2001) 89.
[26] T.R. Moretti, B. Budowle, J.S. Buckleton, Notice of Amendment of the FBI's STR Population Data Published in 1999 and 2001, Journal of Forensic Sciences 60 (2015) 1114–1116.
[27] J.-A. Bright, D. Taylor, J. Curran, J. Buckleton, Searching mixed DNA profiles directly against profile databases, Forensic Science International: Genetics 9 (2014) 102–110.
[28] D. Taylor, J.-A. Bright, J. Buckleton, The 'factor of two' issue in mixed DNA profiles, Journal of Theoretical Biology 363 (2014) 300–306.
[29] J.-A. Bright, I.W. Evett, D. Taylor, J.M. Curran, J. Buckleton, A series of recommended tests when validating probabilistic DNA profile interpretation software, Forensic Science International: Genetics 14 (2015) 125–131.
[30] J.-A. Bright, D. Taylor, C.E. McGovern, S. Cooper, L. Russell, D. Abarno, et al., Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles, Forensic Science International: Genetics 23 (2016) 226–239.
[31] J.M. Curran, J.S. Buckleton, C.M. Triggs, What is the magnitude of the subpopulation effect, Forensic Science International 135 (2003) 1–8.
[32] J. Buckleton, J. Curran, J. Goudet, D. Taylor, A. Thiery, B.S. Weir, Population-specific F_{ST} values for forensic STR markers: A worldwide survey, Forensic Science International: Genetics 23 (2017) 91–100.
[33] C.M. Triggs, J.M. Curran, The sensitivity of the Bayesian HPD method to the choice of prior, Science & Justice 46 (2006) 169–178.
[34] D. Taylor, J.-A. Bright, J. Buckleton, J. Curran, An illustration of the effect of various sources of uncertainty on DNA likelihood ratio calculations, Forensic Science International: Genetics 11 (2014) 56–63.
[35] J.-A. Bright, J.M. Curran, J.S. Buckleton, The effect of the uncertainty in the number of contributors to mixed DNA profiles on profile interpretation, Forensic Science International: Genetics 12 (2014) 208–214.
[36] European Network of Forensic Science Institutes, ENFSI Guideline for Evaluative Reporting in Forensic Science, (2015) . [Accessed] 3 March 2017 http://enfsi.eu/news/enfsi-guideline-evaluative-reporting-forensic-science/.
[37] Association of Forensic Science Providers. Standards for the formulation of evaluative forensic science expert opinion, Science & Justice 49 (2009) 161–164.