Article:

Coble, M. D., & Bright, J.-A. (2019). **Probabilistic genotyping software: An overview**. *Forensic Science International: Genetics*, *38*, 219–224.

INSTITUTE OF ENVIRONMENTAL SCIENCE AND RESEARCH LIMITED
Kenepuru Science Centre: 34 Kenepuru Drive, Kenepuru, Porirua 5022 | PO Box 50348, Porirua 5240, New Zealand
T: +64 4 914 0700 **F:** +64 4 914 0770
www.esr.cri.nz

Probabilistic genotyping software: An overview

Michael D. Coble[1*], Jo-Anne Bright[2]

[1] *Center for Human Identification, Department of Microbiology, Immunology, and Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107, USA*

[2] *Institute of Environmental Science and Research Limited, Private Bag 92021, Auckland 1142, New Zealand*

* Corresponding author. E-mail address*:* michael.coble@unthsc.edu (M.D. Coble).

The interpretation of mixed profiles from DNA evidentiary material is one of the more challenging duties of the forensic scientist. Traditionally, analysts have used a "binary" approach to interpretation where inferred genotypes are either included or excluded from the mixture using a stochastic threshold and other biological parameters such as heterozygote balance, mixture ratio, and stutter ratios. As the sensitivity of STR multiplexes and capillary electrophoresis instrumentation improved over the past 25 years, coupled with the change in the type of evidence being submitted for analysis (from high quality and quantity (often single-source) stains to low quality and quantity (often mixed) "touch" samples), the complexity of DNA profile interpretation has equally increased. This review provides a historical perspective on the movement from binary methods of interpretation to probabilistic methods of interpretation. We describe the two approaches to probabilistic genotyping (semi-continuous and fully continuous) and address issues such as validation and court acceptance. Areas of future needs for probabilistic software are discussed.

Keywords: Forensic DNA, DNA mixture, probabilistic genotyping, mixture software, validation, interpretation

1. Introduction

DNA profile interpretation is a process undertaken to help address the question "who is the source of the DNA in this profile?" If the stain has originated from one individual (termed single-source) assigning a weight to the evidence is relatively easy. DNA profile interpretation is complicated when there is more than one contributor to a profile (termed a mixture) and by dropout and allelic drop-in. Dropout is a consequence of low template, degraded, or inhibited DNA and results in partial DNA profiles, where the DNA from one or more contributors is not present at all loci. Drop-in is the presence of low amounts of DNA within a profile that are not inherent to the DNA extract. Over recent years DNA profiling techniques have become more sensitive with improvements to both chemistry and detection technology. This has resulted in more sensitive methods leading to the generation of more mixtures and profiles exhibiting dropout and drop-in.

2. A Historical Statistical Perspective

To understand where we are today with the widespread acceptance and use of probabilistic genotyping methods, it is useful to first examine our past. With single-source samples, the statistical analysis of a match between the evidence and the person of interest (POI) can be expressed in the form of a Random Match Probability (RMP) [1] or the Likelihood Ratio (LR) [2,3]. The RMP estimates the probability of a matching DNA profile (not the person of interest's) within a chosen population. The LR is not a probability but rather a ratio of two probabilities that evaluates the evidence given two or more mutually exclusive propositions:

$$LR = \frac{\Pr(E \mid H_1)}{\Pr(E \mid H_2)}$$

Where E is the evidence (DNA profile(s)), H_1 is the proposition that the POI is the contributor of DNA to the evidence (sometimes called the prosecution proposition or hypothesis), and H_2 is the proposition that some other randomly selected individual from the population unrelated to the POI is the contributor of DNA to the evidence profile (sometimes called the defense proposition). For an unambiguous single-source profile, the probability of the evidence given H_1 (the numerator) is equal to 1 since the reference profile from the POI would match exactly at all loci. For the same profile, the probability of the evidence given H_2 (that the DNA originated from someone other than the POI) is equal to the RMP. Hence, in its simplest form, the LR for an unambiguous single-source profile is $1/\mathrm{RMP}$.

The statistical evaluation of mixed DNA profiles has traditionally been undertaken using the LR, RMP, and a third method, the Combined Probability of Inclusion (CPI) (also known as Random Man Not Excluded, RMNE). The CPI is the simplest of the three methods. It is not an interpretation method per se in that it does not resolve the profile or attempt to determine the genotypes of the individual contributor(s). It is instead used to estimate the proportion of unrelated individuals in the population that could be included as possible contributors to the profile [4].

In 1998, Clayton et al. (1998) proposed a method to interpret mixed STR profiles [5]. The focus of this work was the interpretation of two-person mixtures since that was their experience "for the overwhelming majority of mixtures encountered during casework" [5]. A survey of DNA casework by Torres et al. (2003) [6] over a four-year period around the same time as the publication of the Clayton et al. paper made similar observations. Over this

4-year period 6.7% of casework samples (163/2424) were mixed. Of these mixtures, 95.1% (155/163) were two-person mixtures with the remaining described as high order mixtures. Torres et al. noted that a majority of the mixed samples in their survey (over 98%) were body fluids involving blood or semen from sexual assaults. Forensic casework in the early 2000s investigated mostly single source-samples and occasionally low order mixtures, predominantly two-person mixtures.

In the U.S., a study funded by the National Institute of Justice focused on the benefits of DNA testing for "minor" property crimes such as burglary and car theft [7]. In 2006, the study concluded that DNA testing could be very useful for identifying individuals committing such crimes and therefore lead to benefits in public safety. Soon after the publication of the study, many laboratories around the U.S. started to see a shift in the types of cases being submitted for DNA testing. Rather than the "high quality, high quantity" body fluid stain evidence such as blood or semen, laboratories were now testing more "trace" evidence such as swabs from car steering wheels and handguns. Often, these type of touch samples contain low quality and/or low quantities of DNA from cellular and non-cellular material. It is not uncommon that many of these touch evidence samples produce complex mixtures with low-level contributors where allelic dropout may be an issue.

With the shift from mostly high level "simple" two-person sexual assault evidence (where the complainant can often be assumed to be in the mixture) to more low-level complex mixture evidence, mixture interpretation becomes more challenging. In 2005, Dr. Peter Gill presenting for a forensic webinar discussing DNA mixture interpretation remarked "If you show ten colleagues a mixture, you will probably end up with ten different answers." The National Institute of Standards and Technology observed similar findings from an inter-laboratory study on mixture interpretation conducted in 2005 where a wide range of inter and intra-laboratory variation was reported [8].

In 2005, the International Society for Forensic Genetics (ISFG) empaneled a DNA commission to provide guidance and recommendations for interpreting DNA mixtures with low-level contributors [9]. Published in 2006, they recommended the LR approach as the preferred method for profile interpretation over the CPI or RMNE method partly due to their wastefulness of information. The stated advantages of LR approaches were that they were the only method that could assess stutter and dropout probabilistically. They gave guidance for how to interpret profiles and also recommended the use of a stochastic threshold when interpreting low-level DNA mixtures.

After publication of the 2006 ISFG recommendations, other forensic governing bodies published letters in support of the guidelines [10–13]. In the U.S., Budowle et al. [14] and the Scientific Working Group on DNA Analysis Methods (SWGDAM) also recommended the use of a stochastic threshold in autosomal STR interpretation [15,16]. The idea of a stochastic threshold was not a novel concept for STR DNA interpretation. In the late 1980s and early 1990s before the routine use of multiplex STR kits, assays such as D1S80, HLA-DQα, and PolyMarker used PCR to amplify these alleles for evidential material too small for conventional RFLP testing. These assays would amplify the target using PCR and then the products would bind to a membrane for colorimetric detection (e.g. a blue dot would appear on the membrane to correspond with the alleles amplified. The "C" dot (for control) for the reverse dot blot PCR amplification assay HLA-DQα and the "S" dot (for sensitivity) in the PolyMarker (PM) testing system were quality assurance controls to assure sufficient amplification of DNA [17,18]. A weak or missing control was an indication that stochastic effects from amplifying low levels of DNA may be present and lead to missing or incomplete data in the assay.

LR methods optionally use peak height information, mixture ratios, and an assigned number

of contributors to interpret the mixture under the two or more propositions considered. Interpretation (or resolution) of a mixture is the attempt to determine the genotype combinations of the different contributors to the profile. The strict application of the RMP does not take into account peak heights however many laboratories have adopted a modified approach that can interpret a mixed profile after an assumption of the number of contributors has been made [1]. The use of peak areas or heights was shown to be very useful in the early days of STR profiling to assist with the interpretation of mixed DNA profiles [19–21]. The differences in peak heights between different contributors are leveraged to resolve the DNA profiles of the different contributors. These models made a number of assumptions including that peak heights are proportional to the quantity of template DNA, and that the height of 'shared' peaks between individuals is the sum of the heights of peaks from the contributing individuals. To our knowledge, studies formally examining these assumptions have not been carried out, however a vast body of empirical data produced by laboratories worldwide certainly supports their validity.

3. Binary methods of interpretation

Early methods of DNA profile interpretation were described as binary as the probability of the evidence given a proposed genotype was assigned as zero (genotype excluded) or one (genotype included). In a binary method of interpretation all included genotype combinations are considered equally likely [22]. There are two broad approaches to binary interpretation methods; quantitative (taking into account peak height) or qualitative (not using peak heights). Both approaches involve the application of rules or thresholds in order to determine genotype combinations of the contributors for which the profile is possible. Rules include the management of stutter, application of analytical and stochastic thresholds, peak height ratio thresholds, and mixture proportion thresholds [14,23,24]. The purpose of a stochastic threshold is to help identify loci where allelic dropout may be possible. Stochastic thresholds (ST) are based on empirical data and values are typically set high so that single peaks observed at a locus can confidently be labelled homozygous.

In a CPI calculation the ST was used to determine if loci should be included (all alleles observable above the ST) or excluded (one or more alleles observed below the ST) within the calculation of the match statistic [4,15]. The ST in isolation was an insufficient criterion and other loci needed to be considered when making the include/exclude decision mentioned above [4]. The CPI approach tends to waste information that should be used [25]. Rather than exclude loci below the ST from the statistic, the '2p rule' was recommended for use in the RMP and the LR calculation to account for zygosity ambiguity. The 2p rule assigns the probability $2p_i$ or $2p_i(1-p_i)+p_i^2$ to the occurrence of a single allele, *i*, whose partner may have dropped out. This approach was recommended by the National Research Council II (NRC II, [26]) for the interpretation of RFLP data. The 2p rule was used for many years to account for potential allelic dropout within a profile. It was thought to be conservative but was later shown to be non-conservative under certain conditions [27].

One can see in the discussion of future needs in many of these early papers on mixture interpretation the recognition that the models of incorporating peak heights and number of contributors were somewhat limited beyond very simple examples. Evett and colleagues [19] give a prescient need for probabilistic software:

> "The complexity of the actual analysis may best be resolved by computer programs of sufficient flexibility and we see the development of such facilities as an important area for the future. There is also a need to combine the requirement for such a system with the capability of dealing impartially with other factors, in particular, peak artefacts, such as pull-up and stutter, and we believe that expert systems

technology represents the most profitable path."

4. Probabilistic methods of interpretation

Binary methods of forensic DNA interpretation are restricted as they are unable to deal completely with complex low level or mixed DNA profiles. These types of data have become more prevalent as DNA typing technologies and STR multiplex chemistries become more sensitive. In addition, binary methods make partial but not full use of the information available from peak heights. These shortcomings led to the development of improved models that factor in the probability of dropout.

The concept of probabilistic genotyping is not a recent discovery or innovation. In the classic paper by Gill and colleagues (Gill et al. 2000) that introduced the forensic DNA world to "Low Copy Number" or "Low Template DNA" testing, the first part of the paper discusses the methodology for increasing sensitivity and the need for replicate analyses to develop a consensus profile. The second half of the paper provides a probabilistic approach to incorporate the probabilities of dropout, drop-in, and stutter into the LR statistic. One of the first probabilistic genotyping software systems (LoComatioN) was developed by the Forensic Science Service and published in 2007 [28] using the underlying mathematical approach in [29] and extended to include population substructure [30]. Additional publications with variations in the mathematical approaches but interpreting mixtures in a probabilistic framework followed. The DNA Commission of the ISFG has also provided guidance for moving to methods to account for dropout and drop-in in low-level DNA profiles [31].

In general, probabilistic genotyping software can be classified into two categories: semi-continuous and fully continuous. The *semi-continuous method* (also known as the drop model or discrete model) can optionally incorporate a probability for dropout, Pr(D), and/or a probability for drop-in, Pr(C). Semi-continuous methods do not use peak heights when generating possible genotype sets and do not model artefacts such as stutter. One limitation of the semi-continuous methods is that peak height information is not fully utilized by the software. Peak heights are important for the end-user of the software to determine parameters such as the probability of dropout or the number of contributors in the mixture.

The dropout probability Pr(D) can be modelled from validation data using logistic regression, by using empirical data directly, or by extending models based on Hb [28,32–36]. The probability of drop-in Pr(C) is generally modelled from empirical data. The calculations are sufficiently complex that software is needed and there are a number currently in use by forensic laboratories.

A number of probabilistic methods using different distributions have been described to model allelic and/or stutter peaks within a DNA profile [19,37–41]. These models take the quantitative information from the DNA profile and calculate the probability of the peak heights given all the possible genotype combinations for the individual contributors. These *continuous methods* make assumptions about the underlying behavior of peak heights to evaluate the probability of a set of peak heights. They use more information from within the profile and reduce the requirement for the subjective manual assignment of peaks as allelic within evidence profiles. These methods require the use of specialized software, and again, there are a number in use (Table S1). Software implementing full and semi-continuous models are generally called probabilistic genotyping methods. Rather than fixing the probability of the profile given a genotype to 1 or 0 as done in the binary approach, the probability may fall between these two values. Probabilistic genotyping refers to the use of

biological modeling and statistical theory to calculate LRs and/or infer genotypes for the DNA typing results of forensic samples [42]. They allow for the interpretation of more complex low level and higher order mixtures that were previously considered too complicated. This has been described as a paradigm shift in DNA interpretation [43].

There has been significant adoption of probabilistic genotyping systems by forensic laboratories for the interpretation of mixed DNA profiles in recent years. In 2014 Prieto et al. reported a European Forensic Genetics Network of Excellence collaborative study involving 18 laboratories using LRmix [44]. In 2017, Bright et al. reported the collated internal validation data for the continuous probabilistic method STRmix™ from 31 laboratories [45]. In 2018, Barrio et al. reported on the results of the Spanish and Portuguese-Speaking Group of the International Society for Forensic Genetics (GHEP-ISFG) collaborative exercise from 25 laboratories mostly using the semi-continuous software LRmix Studio [46].

5. Validation and standards

SWGDAM, ISFG, and the UK Forensic Regulator have all published recommendations for the developmental and internal validation of probabilistic genotyping software [42,47,48]. Generally, developmental validation is undertaken by the software developers and involves the demonstration of the verification of the functionality of the system, the accuracy of statistical calculations and other results, the appropriateness of analytical and statistical parameters, and the determination of the limits of the software [49]. Internal or end user validation is undertaken by the laboratory. It is the accumulation of test data within the laboratory to demonstrate that the software is performing as expected. Developmental validation studies have been published in peer reviewed journals for a number of different probabilistic genotyping software [38,50–52]. In addition, a number of laboratories have published results from their internal validation [45,53,54]. Internal validation studies are less likely to be published as the findings are no longer considered novel.

In September 2016 the U.S. President's Council of Advisors on Science and Technology (PCAST) published a report discussing the scientific validity of feature-comparison methods including DNA profile interpretation [55,56]. Generally, PCAST were very supportive of the use of probabilistic genotyping for the interpretation of mixtures, however they called for the expansion of empirical studies and testing for the validity and reliability of methods across a broader range of profile types. PCAST emphasized that evaluation of software should be undertaken by more than the developers, claiming that to establish scientific validity, scientific evaluation is required by other groups not involved in developing the method. Further, PCAST urged sharing within the forensic community, through publication, of validation studies that properly establish the range of reliability of methods for the analysis of complex DNA mixtures. They also proposed that researchers investigate under what circumstances and why different methods produce different results.

A number of analyst decisions are made during the analysis and interpretation of forensic DNA profiles. These subjective decisions can lead to differences in the reported match statistics. In 2018, Butler et al. reported the results of a 2013 inter-laboratory exercise undertaken by 108 laboratories across the U.S. and Canada, called MIX13 [8]. Results for the five mixed DNA profiles varied significantly between analysts from different laboratories. These differences were due in part to use of different interpretation methods, ST and AT values used, allele frequency databases, and differences in population genetic models. Differences were also observed between analysts from within the same laboratory. Butler et al. reported that this variation "may reflect lack of training or variation in understanding mixture interpretation principles. Observed variation in reported results may also reflect a

lack of protocol specificity or sufficiency." [8]

Probabilistic genotyping methods reduce the number of subjective decisions made by an analyst and hence may improve consistency [57]. There are two groups of 'subjective' decisions that are made during interpretation. The first group relates to the decisions on what models to use. The second group of subjective decisions exist because sufficiency modelling is not part of the interpretation scheme. It is this second group that can be removed as technology and modelling improves, for example such decisions as 'is this a stutter?'. Subjectivity will always exist to some extent.

Differences between results have also been reported for analysts using the probabilistic genotyping software LRmix [44,46] and STRmix™ [58]. The larger differences were mostly due to subjective decisions made prior to the use of the software, such as assignment of the number of contributors (NOC) and choice of proposition set. Minor differences in the STRmix™ study were also observed due to the inherent variability of the Markov chain Monte Carlo (MCMC) algorithms used within the software. Probabilistic genotyping software utilizing MCMC methods will not produce an identical answer after repeat interpretations of the same profile because of the Monte Carlo aspect. The magnitude of this variability within the software STRmix™ has previously been explored and compared with other sources of variability in forensic DNA profiling including PCR, capillary electrophoresis load and injection, and the makeup of allele frequency databases [59]. The MCMC variability was shown to be the smallest under the conditions tested. A recent interpretation of the MIX13 mixtures using different probabilistic genotyping showed some variation between the different software, particularly those employing different methods of interpretation (semi versus fully continuous) [60]. Other studies have been published comparing the results from different models for different profiles [61–64]. As different probabilistic genotyping software use different models, differences in LRs are expected.

6. Court acceptance

There have been a number of admissibility hearings for evidence interpreted using different probabilistic genotyping software in the U.S., and a few in Australia. Different jurisdictions have different requirements that allow for the presentation of expert evidence from new scientific techniques. Commonly, this is a test of general acceptance within the community. In the U.S. it is generally based on the Frye standard [65], and more recently the Daubert standard [66].

The Frye standard, established in 1923, set the bar for determining if evidence has a valid scientific basis. Determining if a new scientific procedure has met "general acceptance" is a two-step process: (1) identify the area and scientific community into which the scientific principle or discovery falls, and (2) to determine if the scientific community has accepted the new technology, principle, or discovery [65]. The Daubert standard was established in the late 1980s and early 1990s. The US Supreme Court established five criteria for accepting scientific evidence: has the theory or technique (1) been tested? (2) been subjected to peer review and publication? (3) has a known or potential rate of error? (4) are there standards for the technique's operation? and (5) is there acceptance in the relevant scientific community?

The number of different probabilistic genotyping methods available (Table S1) and the significant uptake by laboratories adopting probabilistic genotyping software for use in forensic casework is proof of the general acceptance of the methods by the forensic community.

A non-exhaustive list of admissibility hearings in the U.S. and Australia is given in Table S2. In addition to questioning general acceptance, one common theme is the request for access to the software source code. Some probabilistic genotyping software is open source, which means the source code is freely available under license for use or modification. For commercial software such as TrueAllele® and STRmix™, the code is not freely available but the developers will make it available to defense experts on request.

7. Future

As mentioned earlier, there has been an increase in the recovery of low-level and/or complex mixed DNA profiles due to increases in the sensitivity of modern STR multiplex kits and capillary electrophoresis instrumentation, as well as a shift by investigators towards submitting increasing numbers of trace DNA samples. Probabilistic methods of interpretation have allowed laboratories the ability to analyze cases that were too complex for binary interpretation. Therefore, probabilistic software will be an integral component of the DNA testing procedure for years to come.

One area of potential enhancement is in the estimation of the number of contributors in the mixture. Presently, for nearly all probabilistic software programs, the number of contributors is assigned by the analyst and this information is input when setting the hypotheses to be tested by the program. Most software programs "fix" the number of contributors to be equal in both the numerator and denominator. Assigning the number of contributors can be challenging and is often a crucial step in the decision process to analyze a mixture or not – if the number of contributors cannot be assigned with a reasonable degree of confidence then many laboratories will not progress an interpretation. This may be the case where there are multiple low-level contributors with indications of allelic dropout or where dealing with higher order mixtures that appear to originate from four or more contributors. One possible approach is to carry out several interpretations under different NOC and then report all results. Of course care may need to be taken in explaining the results if, for example, a person of interest is excluded under one NOC but included under another. A growing number of software programs can maximize the number of contributors in the numerator and denominator to allow variation in the number of contributors in each proposition [67]. Slooten and Caliebe [68] have published an approach that considers the number of contributors a "nuisance" parameter and they provide examples using this method. It should be noted that in general, overestimation of the number of contributors has very little effect on the LR of the true contributors in the mixture and tends to give an uninformative (i.e. approaching 1) LR for a true non-contributor [69].

As the forensic DNA community explores Massively Parallel Sequencing (MPS) for autosomal DNA analyses, the promises of improved mixture interpretation have been demonstrated [70,71]. The presence of "isoalleles" (alleles indistinguishable using size-based resolution methods such as CE but resolvable at the sequence level) or SNPs in the amplicon from individual contributors [72] can readily be identified in MPS data. Bleka et al. [73] have shown that mixtures using SNPs from MPS data can be analyzed effectively using a fully continuous program. Autosomal STR mixtures analyzed with MPS will also likely require probabilistic methods of interpretation. Models to account for variation in the "read" rates generated from library preparation, rather than peak heights in CE systems, will be required. Stutter will also likely require probabilistic methods of interpretation for detecting artifacts that may not be an issue in CE-based interpretation, but may be present in MPS data [74]. In addition, complex sequence motifs in the current STR loci can lead to variation in stutter products which will require a probabilistic interpretation.

Although the focus of this review is on autosomal DNA mixture interpretation with

probabilistic approaches, there is a strong need to apply probabilistic methods to interpreting mixtures with haploid marker systems. Progress has been made in separating mitochondrial DNA mixtures using MPS data [75–79]. There is a need for the same advancement in Y-STR mixture analysis, particularly given that modern autosomal STR multiplexes routinely include Y-STR markers. Taylor et al. [80] have provided a roadmap for a continuous approach to interpreting Y-STR mixtures, but recognize that such systems are still in the early stages of development. More recently, Andersen and Balding have described a method for the interpretation of Y-STR mixtures [81].

8. Conclusions

The concept of probabilistic methods to interpret DNA mixtures is neither a new nor a recent innovation to forensic testing. Based on the number of publications, there has been a recent "explosion" in the number of software programs available for the laboratory (Table S1). These include a wide variety of solutions that range from freely available, to programs with an open source code, to programs that are commercially supported. At least in the United States, there are at least 48 laboratory systems, representing around 100 individual laboratories, that have implemented and are using a probabilistic software solution in casework. This represents a "tipping point" in that nearly half of the United States has transitioned from a binary approach to interpretation to probabilistic methods of interpretation. Approximately 55 additional laboratory systems in the United States are at some point in the process of procuring or validating a software program.

With the transition from binary to probabilistic methods of interpretation, we agree with Gill et al. [82] that probabilistic genotyping software cannot and should not replace trained, qualified, and experienced analysts. The analyst should first assess the DNA profile prior to software analysis and then critically evaluate the results produced. It should be remembered that probabilistic software is only a tool for the forensic scientist to assist with interpretation.

This crucial evaluation of the input (mixture) and output (LR) from the software therefore requires that the analyst have a strong foundation in LRs and understand how their particular software works. Many laboratories bringing on a probabilistic genotyping system (particularly in the US) may have spent most of their time reporting DNA mixture statistics with CPI and/or the RMP. Transitioning to the LR and understanding the nuances of building relevant propositions based upon case scenarios can be challenging to users accustomed to a frequentist view of probability.

It is also important, resonating the advice of the DNA Commission of the ISFG and other scientists [31], that software should not be treated as a "black box" where something magical happens to generate the statistic. The LR generated from the program is based upon modeling parameters, population genetic theory, and other assumptions being used by the software. It is imperative the end user understand the underlying mathematics (at least to a conceptual level), assumptions, models and limitations of the software program to convey how the program works to the trier of fact. To quote Buckleton et al. [83], "By understanding the strengths and limitations of any PG software, users and stakeholders will better understand the system and hopefully use it in a thoughtful manner for the public good."

The forensic DNA community is presently in the midst of a paradigm shift in the approach to interpreting complex mixtures. Probabilistic Genotyping software provides substantial advantages for interpreting low-level, complex DNA mixtures over existing binary approaches. This is especially true for any profile where peak heights contain information that assists in resolving the genotypes of contributors, and has been found to be true in most

circumstances [84]. The software makes better use of the data by accounting for dropout rather than excluding the locus or using the "2p" rule. As the field evolves to implementing new technologies (MPS) or marker systems (SNPs) compared to the traditional STRs on CE, probabilistic methods of interpretation will be a necessary part of the forensic DNA toolkit.

Acknowledgements

The authors would like to thank Dr. James Curran (University of Auckland) for helpful historical discussions and guidance. The authors would also like to acknowledge Dr. John Buckleton (ESR), Zane Kerr (ESR), and two anonymous reviewers for their helpful comments and suggestions that improved this manuscript. This work was supported in part by grant NIJ 2017-DN-BX-0136 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of their organizations.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fsigen.2018.11.009.

References

[1] Bille T, Bright J-A, Buckleton J. Application of Random Match Probability Calculations to Mixed STR Profiles. Journal of Forensic Sciences. 2013;58:474-85.

[2] Evett IW, Buffery C, Willott G, Stoney D. A guide to interpreting single locus profiles of DNA mixtures in forensic cases. Journal of the Forensic Science Society. 1991;31:41-7.

[3] Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KAJ, Buckleton JS. Interpreting DNA Mixtures. Journal of Forensic Sciences. 1997;42:213-22.

[4] Bieber FR, Buckleton JS, Budowle B, Butler JM, Coble MD. Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion. BMC Genetics. 2016;17:125.

[5] Clayton T, Whitaker JP, Sparkes RL, Gill P. Analysis and interpretation of mixed forensic stains using DNA STR profiling. Forensic Science International. 1998;91:55 - 70.

[6] Torres Y, Flores I, Prieto V, Lopez-Soto M, Farfan MJ, Carracedo A, et al. DNA mixtures in forensic casework: a 4-year retrospective study. Forensic Science International. 2003;134:180-6.

[7] Zedlewski E, Murphy MB. DNA Analysis for "Minor" Crimes: A Major Benefit for Law Enforcement. NIJ Journal. 2006:https://www.nij.gov/journals/253/pages/dna_analysis.aspx.

[8] Butler JM, Kline MC, Coble MD. NIST Interlaboratory Studies Involving DNA Mixtures (MIX05 and MIX13): Variation Observed and Lessons Learned. Forensic Science International: Genetics. 2018;37:81-94.

[9] Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, et al. DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. Forensic Science International. 2006;157:187-97.

[10] Gill P, Brown RM, Fairley M, Lee L, Smyth M, Simpson N, et al. National recommendations of the Technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes. Forensic Science International: Genetics. 2008;2:76-82.

[11] Schneider PM, Fimmers R, Keil W, Molsberger G, Patzelt D, Pflug W, et al. The German Stain Commission: recommendations for the interpretation of mixed stains. International Journal of Legal Medicine. 2009;123:1-5.

[12] Morling N, Bastisch I, Gill P, Schneider PM. Interpretation of DNA mixtures--European consensus on principles. Forensic Science International: Genetics. 2007;1:291-2.

[13] Stringer P, Scheffer JW, Scott P, Lee J, Goetz R, Ientile V, et al. Interpretation of DNA Mixtures--Australian and New Zealand consensus on principles. Forensic Science International: Genetics. 2009;3:144-5.

[14] Budowle B, Onorato AJ, Callaghan TF, Manna AD, Gross AM, Guerreri RA, et al. Mixture Interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. Journal of Forensic Sciences. 2009;54:810-21.

[15] Scientific Working Group on DNA Analysis Methods (SWGDAM). SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. 2010. http://www.forensicdna.com/assets/swgdam_2010.pdf Accessed 1 June 2018.

[16] Scientific Working Group on DNA Analysis Methods (SWGDAM). Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. 2017. https://media.wix.com/ugd/4344b0_2a08f65be531488caa8037ed55baf23d.pdf Accessed 17 March 2017.

[17] Budowle B, Lindsey JA, DeCou JA, Koons BW, Giusti AM, Comey CT. Validation and population studies of the loci LDLR, GYPA, HBGG, D7S8, and Gc (PM loci) and HLA-DQ-alpha using a multiplex amplification and typing procedure. Journal of Forensic Sciences. 1995;40:45-54.

[18] Moretti T, Baumstark AL, Defenbaugh BS, Keys KM, Smerick JB, Budowle B. 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. 2001;46:647 - 60.

[19] Evett IW, Gill PD, Lambert JA. Taking account of peak areas when interpreting mixed DNA profiles. Journal of Forensic Sciences. 1998;43:62-9.

[20] Gill P, Sparkes RL, Pinchin R, Clayton T, Whitaker JP, Buckleton JS. Interpreting simple STR mixtures using allelic peak areas. Forensic Science International. 1998;91:41 - 53.

[21] Perlin MW, Szabady B. Linear mixture analysis: A mathematical approach to resolving mixed DNA samples. Journal of Forensic Sciences 2001;46:1372-7.

[22] Kelly H, Bright J-A, Buckleton JS, Curran JM. A comparison of statistical models for the analysis of complex forensic DNA profiles. Science and Justice. 2014;54:66-70.

[23] Bright J-A, Turkington J, Buckleton J. Examination of the variability in mixed DNA profile parameters for the Identifiler multiplex. Forensic Science International: Genetics. 2009;4:111-4.

[24] Bright J-A, McManus K, Harbison S, Gill P, Buckleton J. A comparison of stochastic variation in mixed and unmixed casework and synthetic samples. Forensic Science International: Genetics. 2012;6:180-4.

[25] Buckleton J, Curran J. A discussion of the merits of random man not excluded and likelihood ratios. Forensic Science International: Genetics. 2008;2:343-8.

[26] National Research Council II. National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence. Washington, D.C.: National Academy Press; 1996.

[27] Buckleton J, Triggs CM. Is the 2p rule always conservative? Forensic Science International. 2006;159:206-9.

[28] Gill P, Kirkham A, Curran J. LoComatioN: A software tool for the analysis of low copy number DNA profiles. Forensic Science International. 2007;166:128-38.

[29] Gill P, Whitaker JP, Flaxman C, Brown N, Buckleton JS. An investigation of the rigor of interpretation rules for STR's derived from less than 100 pg of DNA. Forensic Science International. 2000;112:17-40.

[30] Bill M, Gill P, Curran J, Clayton T, Pinchin R, Healy M, et al. PENDULUM - A guideline based approach to the interpretation of STR mixtures. Forensic Science International. 2005;148:181-9.

[31] Gill P, Gusmão L, Haned H, Mayr WR, Morling N, Parson W, et al. DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. Forensic Science International: Genetics. 2012;6:679-88.

[32] Tvedebrink T, Eriksen PS, Mogensen HS, Morling N. Estimating the probability of allelic drop-out of STR alleles in forensic genetics. Forensic Science International: Genetics. 2009;3:222-6.

[33] Haned H, Egeland T, Pontier D, Pène L, Gill P. Estimating drop-out probabilities in forensic DNA samples: A simulation approach to evaluate different models. Forensic Science International: Genetics. 2011;5:525-31.

[34] Tvedebrink T, Eriksen PS, Asplund M, Mogensen HS, Morling N. Allelic drop-out probabilities estimated by logistic regression—Further considerations and practical implementation. Forensic Science International: Genetics. 2012;6:263-7.

[35] Buckleton JS, Kelly H, Bright J-A, Taylor D, Tvedebrink T, Curran JM. Utilising allelic dropout probabilities estimated by logistic regression in casework. Forensic Science International: Genetics. 2014;9:9-11.

[36] Lohmueller KE, Rudin N, Inman K. Analysis of allelic drop-out using the Identifiler® and PowerPlex® 16 forensic STR typing systems. Forensic Science International: Genetics. 2014;12:1-11.

[37] Cowell RG, Lauritzen SL, Mortera J. A gamma model for DNA mixture analyses. Bayesian Analysis. 2007;2:333-48.

[38] Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, et al. Validating TrueAllele® DNA mixture interpretation. Journal of Forensic Sciences. 2011;56:1430-47.

[39] Taylor D, Bright J-A, Buckleton J. The interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics. 2013;7:516-28.

[40] Puch-Solis R, Rodgers L, Mazumder A, Pope S, Evett I, Curran J, et al. Evaluating forensic DNA profiles using peak heights, allowing for multiple donors, allelic dropout and stutters. Forensic Science International: Genetics. 2013;7:555-63.

[41] Bright J-A, Taylor D, Curran JM, Buckleton JS. Developing allelic and stutter peak height models for a continuous method of DNA interpretation. Forensic Science International: Genetics. 2013;7:296-304.

[42] Scientific Working Group on DNA Analysis Methods (SWGDAM). Guidelines for the Validation of Probabilistic Genotyping Systems. 2015. http://media.wix.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf Accessed 3 October 2016.

[43] Bright J-A, Taylor D, Gittelson S, Buckleton J. The paradigm shift in DNA profile interpretation. Forensic Science International: Genetics. 2017;31:e24-e32.

[44] Prieto L, Haned H, Mosquera A, Crespillo M, Alemañ M, Aler M, et al. Euroforgen-NoE collaborative exercise on LRmix to demonstrate standardization of the interpretation of complex DNA profiles. Forensic Science International: Genetics. 2014;9:47-54.

[45] Bright J-A, Richards R, Kruijver M, Kelly H, McGovern C, Magee A, et al. Internal validation of STRmix™ – A multi laboratory response to PCAST. Forensic Science International: Genetics. 2018;34:11-24.

[46] Barrio PA, Crespillo M, Luque JA, Aler M, Baeza-Richer C, Baldassarri L, et al. GHEP-ISFG collaborative exercise on mixture profiles (GHEP-MIX06). Reporting conclusions: Results and evaluation. Forensic Science International: Genetics. 2018;35:156-63.

[47] Coble MD, Buckleton J, Butler JM, Egeland T, Fimmers R, Gill P, et al. DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications. Forensic Science International: Genetics. 2016;25:191-7.

[48] Forensic Science Regulator. Software Validation for DNA Mixture Interpretation, FSR-G-223 Issue 1. 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/730994/G223_Mixture_software_validation_Issue1.pdf Accessed 11 September 2018.

[49] Haned H, Gill P, Lohmueller K, Inman K, Rudin N. Validation of probabilistic genotyping software for use in forensic DNA casework: Definitions and illustrations. Science and Justice. 2016;56:104-8.

[50] Bright J-A, Taylor D, McGovern CE, Cooper S, Russell L, Abarno D, et al. Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles. Forensic Science International: Genetics. 2016;23:226-39.

[51] Götz FM, Schönborn H, Borsdorf V, Pflugbeil A-M, Labudde D. GenoProof Mixture 3—New software and process to resolve complex DNA mixtures. Forensic Science International: Genetics Supplement Series. 2017;6:e549-e51.

[52] Mitchell AA, Tamariz J, O'Connell K, Ducasse N, Budimlija Z, Prinz M, et al. Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in. Forensic Science International: Genetics. 2012;6:749-61.

[53] Moretti TR, Just RS, Kehl SC, Willis LE, Buckleton JS, Bright J-A, et al. Internal validation of STRmix for the interpretation of single source and mixed DNA profiles. Forensic Science International: Genetics. 2017;29:126-44.

[54] Greenspoon SA, Schiermeier-Wood L, Jenkins BC. Establishing the Limits of TrueAllele® Casework: A Validation Study. Journal of Forensic Sciences. 2015;60:1263-76.

[55] President's Council of Advisors on Science and Technology. Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. 2016. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf Accessed 22 April 2017.

[56] President's Council of Advisors on Science and Technology. An addendum to the PCAST report on forensic science in criminal courts. 2016. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf Accessed 20 July 2017.

[57] Jeanguenat AM, Budowle B, Dror IE. Strengthening forensic DNA decision making through a better understanding of the influence of cognitive bias. Science and Justice. 2017;57:415-20.

[58] Cooper S, McGovern C, Bright JA, Taylor D, Buckleton J. Investigating a common approach to DNA profile interpretation using probabilistic software. Forensic Science International: Genetics. 2015;16:121-31.

[59] Bright J-A, Stevenson KE, Curran JM, Buckleton JS. The variability in likelihood ratios due to different mechanisms. Forensic Science International: Genetics. 2015;14:187-90.

[60] Buckleton JS, Bright J-A, Cheng K, Budowle B, Coble MD. NIST Interlaboratory Studies Involving DNA Mixtures (MIX13): A modern analysis. Forensic Science International: Genetics. 2018;37:172-9.

[61] Bright J-A, Evett IW, Taylor D, Curran JM, Buckleton J. A series of recommended tests when validating probabilistic DNA profile interpretation software. Forensic Science International: Genetics. 2015;14:125-31.

[62] Bille TW, Weitz SM, Coble MD, Buckleton JS, Bright J-A. Comparison of the performance of different models for the interpretation of low level mixed DNA profiles. Electrophoresis. 2014;35:3125-33.

[63] Alladio E, Omedei M, Cisana S, D'Amico G, Caneparo D, Vincenti M, et al. DNA mixtures interpretation – A proof-of-concept multi-software comparison highlighting different probabilistic methods' performances on challenging samples. Forensic Science International: Genetics. 2018;37:143-50.

[64] Bleka Ø, Benschop CCG, Storvik G, Gill P. A comparative study of qualitative and quantitative models used to interpret complex STR DNA profiles. Forensic Science International: Genetics. 2016;25:85-96.

[65] Frye v The United States of America, 54 AppDC 46, 293Fed 1013 (1923). 1923.

[66] Daubert et al. v Merrell Dow Pharmaceuticals Inc., 509 US 579 (1993). 1993.

[67] Taylor D, Bright J-A, Buckleton J. Interpreting forensic DNA profiling evidence without specifying the number of contributors. Forensic Science International: Genetics. 2014;13:269-80.

[68] Slooten K, Caliebe A. Contributors are a nuisance (parameter) for DNA mixture evidence evaluation. Forensic Science International: Genetics. 2018;37.

[69] Taylor D. Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. Forensic Science International: Genetics. 2014;11:144-53.

[70] Churchill JD, Schmedes SE, King JL, Budowle B. Evaluation of the Illumina® Beta Version ForenSeq™ DNA Signature Prep Kit for use in genetic profiling. Forensic Science International: Genetics. 2016;20:20-9.

[71] Van Neste C, Vandewoestyne M, Van Criekinge W, Deforce D, Van Nieuwerburgh F. My-Forensic-Loci-queries (MyFLq) framework for analysis of forensic STR data generated by massive parallel sequencing. Forensic Science International: Genetics. 2014;9:1-8.

[72] Gettings KB, Aponte RA, Vallone PM, Butler JM. STR allele sequence variation: Current knowledge and future issues. Forensic Science International: Genetics. 2015;18:118-30.

[73] Bleka Ø, Eduardoff M, Santos C, Phillips C, Parson W, Gill P. Open source software EuroForMix can be used to analyse complex SNP mixtures. Forensic Science International: Genetics. 2017;31:105-10.

[74] Aponte RA, Gettings KB, Duewer DL, Coble MD, Vallone PM. Sequence-based analysis of stutter at STR loci: Characterization and utility. Forensic Science International: Genetics Supplement Series. 2015;5:e456-e8.

[75] Holland MM, McQuillan MR, O'Hanlon KA. Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy. Croatian Medical Journal. 2011;52:299-313.

[76] Vohr SH, Gordon R, Eizenga JM, Erlich HA, Calloway CD, Green RE. A phylogenetic approach for haplotype analysis of sequence data from complex mitochondrial mixtures. Forensic Science International: Genetics. 2017;30:93-105.

[77] Churchill JD, Stoljarova M, King JL, Budowle B. Massively parallel sequencing-enabled mixture analysis of mitochondrial DNA samples. International Journal of Legal Medicine. 2018;132:1263-72.

[78] Egeland T, Salas A. A statistical framework for the interpretation of mtDNA mixtures: forensic and medical applications. PloS One. 2011;6:e26723.

[79] Pereira V, Longobardi A, Børsting C. Sequencing of mitochondrial genomes using the Precision ID mtDNA Whole Genome Panel. Electrophoresis. 2018;doi: 10.1002/elps.201800088.

[80] Taylor D, Bright J-A, Buckleton J. Using probabilistic theory to develop interpretation guidelines for Y-STR profiles. Forensic Science International: Genetics.21:22-34.

[81] Andersen MM, Balding DJ. Y-profile evidence: Close paternal relatives and mixtures. Forensic Science International: Genetics. 2019;38:48-53.

[82] Gill P, Haned H, Bleka O, Hansson O, Dorum G, Egeland T. Genotyping and interpretation of STR-DNA: Low-template, mixtures and database matches - twenty years of research anddevelopment. Forensic Science International: Genetics. 2015;18:100-17.

[83] Buckleton JS, Bright J-A, Gittelson S, Moretti TR, Onorato AJ, Bieber FR, et al. The Probabilistic Genotyping Software STRmix: Utility and Evidence for its Validity. Journal of Forensic Sciences. 2018;doi: 10.1111/1556-4029.13898.

[84] Taylor D, Buckleton J. Do low template DNA profiles have useful quantitative data? Forensic Science International: Genetics. 2015;16:13-6.