**NIST IR 8351sup1**
December 2024

# History of DNA Mixture Interpretation

*Supplemental Document to*
*DNA Mixture Interpretation:*
*A NIST Scientific Foundation Review*

NIST
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

These opinions, recommendations, findings, and conclusions do not necessarily reflect the views or policies of NIST or the United States Government. This report and its supplemental documents are not intended to imply any criticism of any past, pending, or future legal proceedings involving forensic science or evidence based on forensic science.

**NIST Interagency Report**
**NIST IR 8351sup1**

# History of DNA Mixture Interpretation

*Supplemental Document to DNA Mixture Interpretation: A NIST Scientific Foundation Review*

John M. Butler

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8351sup1

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

**NIST Interagency Report**
**NIST IR 8351sup1**

# History of DNA Mixture Interpretation

*Supplemental Document to DNA Mixture Interpretation:*
*A NIST Scientific Foundation Review*

John M. Butler
*Special Programs Office*
*Laboratory Programs*
*National Institute of Standards and Technology*

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8351sup1

December 2024

U.S. Department of Commerce
*Gina M. Raimondo, Secretary*

National Institute of Standards and Technology
*Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology*

NISTIR 8351sup1
December 2024

Certain equipment, instruments, software, or materials, commercial or non-commercial, are identified in this paper in order to specify the experimental procedure adequately. Such identification does not imply recommendation or endorsement of any product or service by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

These opinions, recommendations, findings, and conclusions do not necessarily reflect the views or policies of NIST or the United States Government. This report and its supplemental documents are not intended to imply any criticism of any past, pending, or future legal proceedings involving forensic science or evidence based on forensic science.

**NIST Technical Series Policies**
Copyright, Use, and Licensing Statements
NIST Technical Series Publication Identifier Syntax

**Publication History**
Approved by the NIST Editorial Review Board on 2024-12-17

**How to Cite this NIST Technical Series Publication**
Butler, J.M. (2024) History of DNA Mixture Interpretation. *Supplemental Document to DNA Mixture Interpretation: A NIST Scientific Foundation Review.* (National Institute of Standards and Technology, Gaithersburg, MD), NIST IR 8351sup1. https://doi.org/10.6028/NIST.IR.8351sup1

**Author ORCID iDs**
John M. Butler: 0000-0001-6472-9157

**Contact Information**
john.butler@nist.gov

## Preface

When NISTIR 8351-draft was released for public comment on June 9, 2021, this supplemental document was originally Appendix 1 of the draft. While preparing the final version of NISTIR 8351 **DNA Mixture Interpretation: *A NIST Scientific Foundation Review***, the report authors decided to move this material from an appendix into this supplemental document that will be available on the same NIST website[1]. Updates and revisions have been made since the draft report. Historical information shared herein was informed and shaped by the author's experiences in the field (>30 years) and through writing several textbooks on the topic (e.g., Butler 2009, Butler 2011, Butler 2014).

[1] See https://www.nist.gov/spo/forensic-science-program/dna-mixture-interpretation-nist-scientific-foundation-review (accessed July 1, 2024)

## Acknowledgments

Valuable input has been received over the course of this project from members of the NIST team and the DNA Mixture Resource Group mentioned in Chapter 1 of the full report. In particular, Melissa Taylor was instrumental in organizing the content along with feedback from public comments provided on the draft report. These suggestions strengthened the final content of this supplemental document and are greatly appreciated. Additional edits and suggestions were also provided by Katherine Sharpless, Kelly Sauerwein, Sandra Koch, and Laura Baxter. Elements of the cover art were created by Natasha Hanacek from the NIST Web Content Development Group.

## Abstract

This report provides supplemental information to NISTIR 8351 **DNA Mixture Interpretation: *A NIST Scientific Foundation Review*** and summarizes the history of DNA mixture interpretation. Over the past 40 years, there has been an evolution of new technologies for DNA analysis (different markers, kits, instruments, and software) and new strategies for DNA mixture interpretations (manual deconvolution, binary and probabilistic models) along with a steady stream of peer-reviewed publications. Forensic DNA testing exists in this evolving environment with an increasingly complex set of technologies. Often, important changes and advances have been made across the community because of experiences in high-profile court cases or awareness of issues raised through participation in interlaboratory studies or collaborative exercises. Several of these cases and studies are highlighted here. Histories of guidance documents and training courses related to DNA mixture interpretation are also described. Recommendations on DNA mixture interpretation from a 2006 International Society for Forensic Genetics DNA Commission (see Box S1.4) serve as core foundational principles. The field would benefit from future guidance documents promoting performance-based approaches to validation studies and continued education and research in DNA mixture interpretation.

## Keywords

DNA; DNA mixture interpretation; forensic genetics; forensic science; history of DNA mixture interpretation; scientific foundation review.

## Table of Contents

**1. History of DNA Mixture Interpretation ........ 8**
1.1. Introduction and Purpose of This Document ........ 8
1.2. The Daubert Standard ........ 8
1.3. Important Issues with DNA Mixtures ........ 9
1.4. Four Decades of Change ........ 10
**2. Measurement Advances ........ 11**
2.1. Early Method Development and Assessment of DNA Mixtures ........ 11
2.1.1. The Rise and Impact of STR Methods ........ 15
2.2. Increased Sensitivity in DNA Test Methods ........ 16
2.2.1. Low Copy Number (LCN) Efforts ........ 16
2.2.2. Reliability Concerns with Increased Sensitivity ........ 17
2.2.3. Relevance Concerns with Increased Sensitivity ........ 17
2.3. Future of Measurement Advances ........ 18
**3. Interpretation Advances ........ 20**
3.1. Initial Interpretation Approaches Explored for DNA Mixtures ........ 20
3.1.1. Early U.S. Mixture Approaches – The NRC I and NRC II Reports ........ 21
3.1.2. First High-Profile Case with DNA Mixtures ........ 23
3.1.3. Development of LR Methods ........ 24
3.1.4. Mixture Deconvolution ........ 25
3.2. Probabilistic Genotyping Software (PGS) ........ 26
3.2.1. Development of PGS ........ 27
3.2.2. Movement to PGS in the United States ........ 27
3.2.3. FTCOE 2015 Landscape Study of PGS Systems ........ 28
3.3. Towards a Future Incorporating Artificial Intelligence into Interpretation ........ 29
3.3.1. Number of Contributor (NoC) Estimation ........ 29
3.3.2. “Lights-Out” Data Interpretation ........ 30
**4. Assessing Measurement and Interpretation Variation Across Laboratories ........ 32**
4.1. Initial NIST Studies ........ 32
4.2. Mixture Interpretation Studies by NIST and Others ........ 33
4.3. Proficiency Tests for DNA Mixture Interpretation ........ 34
**5. Sources of Guidance on DNA Mixture Interpretation and Validation ........ 36**
5.1. Overview ........ 36
5.2. ISFG DNA Commission and European Efforts in Mixture Interpretation ........ 37
5.3. SWGDAM and U.S. Efforts in Mixture Interpretation ........ 40

5.3.1. U.S. Validation Guidance Regarding DNA Mixture Interpretation ........ 42
5.4. Requirements and Expectations for PGS Validation ........ 49
5.4.1. Published Input from Software Developers ........ 49
5.4.2. SWGDAM 2015 PGS Validation Guidelines ........ 50
5.4.3. ISFG 2016 DNA Commission on Software Validation ........ 52
5.4.4. ENFSI DNA Working Group 2017 Best Practice Manual ........ 53
5.4.5. UK Forensic Science Regulator 2018 Guidance ........ 54
5.4.6. ANSI/ASB 018 PGS Validation Standard ........ 55
5.5. Concerns Raised by PCAST 2016 on DNA Mixture Interpretation ........ 56
**6. History of DNA Mixture Interpretation Training ........ 58**
6.1. Initial U.S. Training Workshop on Mixtures ........ 58
6.2. Training on Principles ........ 59
6.3. Training on Probabilistic Genotyping Software ........ 59
**7. Looking to the Future: Opportunities and Needs ........ 61**
**References ........ 62**

## List of Figures

**Figure S1.1.** DNA typing methodologies and markers ..........11

## List of Tables

**Table S1.1.** Measurement advances timeline summary..........19

**Table S1.2.** Interpretation advances timeline summary..........31

**Table S1.3.** Documents that govern DNA operations in accredited laboratories..........36

**Table S1.4.** Validation requirements that relate to DNA mixture interpretation..........43-46

**Table S1.5.** Webinar series held in 2019 on probabilistic genotyping..........60

## List of Boxes

**Box S1.1.** Differential extraction..........14

**Box S1.2.** Statistical approaches used for DNA mixture interpretation (SWGDAM 2017)..........21

**Box S1.3.** Impact of O.J. Simpson Trial on U.S. Approach to DNA Mixtures..........24

**Box S1.4**. ISFG 2006 recommendations on DNA mixture interpretation (Gill et al. 2006)..........38

# 1. History of DNA Mixture Interpretation

## 1.1. Introduction and Purpose of This Document

This supplemental document was originally part of the draft of **DNA Mixture Interpretation: *A NIST Scientific Foundation Review*** (NISTIR 8351) as Appendix 1 when it was released for public comment on June 9, 2021. While preparing the final version of the report, the authors decided to move this material from an appendix of the main report into this supplemental document to enable a better focus on key issues. The information contained here has also been reorganized to reflect measurement and interpretation advances, followed by a summary of guidance documents and training materials.

Understanding the historical background of a method or practice can assist modern users to appreciate the method or practice's scientific foundations and limitations. With forensic science methods and practices, this knowledge becomes particularly important in court admissibility hearings because when scientific evidence is presented in courts of law, it should be carefully communicated and meet certain criteria.

## 1.2. The Daubert Standard

The 1993 U.S. Supreme Court case *Daubert v. Merrell Dow Pharmaceuticals Inc.* established the "Daubert Standard" as a systematic framework for a trial court judge to act as a gatekeeper of scientific evidence, such that the reliability and relevance of expert witness testimony can be assessed before it is presented to a jury. To defend the use of a forensic science method or practice, forensic practitioners may undergo a Daubert admissibility hearing that scrutinizes the methodology and the underlying scientific principles.

The following factors are typically considered under the Daubert Standard[2]:

1. Whether the technique or theory in question can be, and has been tested;
2. Whether it has been subjected to publication and peer review;
3. Its known or potential error rate;
4. The existence and maintenance of standards controlling its operation; and
5. Whether it has attracted widespread acceptance within a relevant scientific community.

In preparing this history of the field supplement to the NIST DNA mixture interpretation report (NISTIR 8351), the five Daubert Standard factors have been considered and examined through the lens of several questions:

(a) When were the earliest documented uses of this forensic method and practice?

[2] See https://www.law.cornell.edu/wex/daubert_standard. The Daubert Standard supplanted the Frye Standard, which focused primarily on the general acceptance of scientific evidence within a particular field (see *Frye v. United States, 293 F. 1013 (D.C. Cir. 1923)*). While some U.S. state courts still adhere to the Frye Standard, the Daubert Standard is used in all federal courts.

(b) How has this forensic method and practice progressed over time?

(c) Are there key pivot points or cases that influenced the field?

(d) What are some of the key published research studies?

(e) Where have capabilities and limitations been documented?

(f) Have there been continuing challenges or issues?

(g) What guidance documents have been and/or are currently available for measurement and interpretation approaches used in the field?

(h) What documented training materials have been available over time and thus potentially impactful to practitioners?

### 1.3. **Important Issues with DNA Mixtures**

There are several things to keep in mind when interpreting DNA mixtures.

First, DNA mixtures arise from the contributions of multiple individuals. Samples containing DNA mixtures are characteristic of many forensic casework scenarios, such as swabs collected from sexual assault victims where cellular material from the victim and the perpetrator(s) may be recovered together. In some situations, a physical separation of the victim's epithelial cells and the perpetrator(s)'s sperm cells, known as differential extraction, may be possible to reduce the complexity of the resulting DNA profiles. A process for doing differential extraction was first described in 1985 (Gill et al. 1985) and has not significantly changed over the subsequent decades (e.g., Cotton & Fisher 2015).

Second, multiple approaches have been used for DNA mixture interpretation since the early 1990s, including likelihood ratios (LRs) (Evett et al. 1991, DAB 2000) and combined probability of inclusion (CPI) (NRC 1992, DAB 2000).

Third, trace levels of DNA are detectable with high sensitivity methods employed over the past several decades. Using these high sensitivity methods to analyze low quantities of DNA has directly led to greater complexity with the resulting DNA profiles and an expansion of approaches to performing mixture interpretation (e.g., Gill et al. 2015).

Fourth, probabilistic genotyping software (PGS) systems have been steadily growing in use by forensic DNA laboratories over the past decade or more (Coble & Bright 2019) to assist with the complex DNA profiles being obtained from the high sensitivity methods in use.

As discussed in the accompanying report (NISTIR 8351, see Figure 2.1 therein), the overall DNA testing process can be subdivided into measurement and interpretation aspects. Measurement is defined in this context as a series of steps to generate a DNA profile. This profile will depend on the genetic markers and technologies utilized. Separate steps and even different skill sets are involved with interpretation of the DNA profile and its potential meaning in the context of a forensic case. Thus, interpretation advances are discussed separately.

## 1.4. Four Decades of Change

Forensic DNA analysis has undergone numerous changes in the four decades since DNA methods were first applied to criminal investigations (Gill et al. 1985). An examination of the history of DNA mixture interpretation reveals an evolving set of technologies, DNA tests, and statistical approaches (e.g., Gill et al. 2015, Coble & Bright 2019). Many forensic laboratories have begun moving from "binary" approaches (i.e., is the genotype of interest present or not in the observed mixture?) to "probabilistic genotyping" methods (i.e., could the genotype of interest be present and, if so, how strongly does the data support this possibility?).

This shift has occurred as techniques and approaches to mixture interpretation have evolved over time with detection methods becoming more sensitive and the questions from end-users changing.

Samples analyzed by laboratories have changed from large visible stains to small invisible samples due to highly sensitive detection methods. Questions asked by the legal community have expanded from asking "to *whom* does the DNA belong?" (source-level) to also asking "*how* did it get there?" (activity-level) (see EWG 2024, pp. 172-182). The technology and marker sets have evolved from restriction fragment length polymorphism (RFLP) to simple polymerase chain reaction (PCR) assays to multiplex short tandem repeat (STR) markers with different commercial kits. Statistical approaches have changed in many jurisdictions e.g., from combined probability of inclusion (CPI) to likelihood ratios (LRs). However, as the core principles underlying relevant and reliable DNA mixture interpretation remain the same (see Chapter 2 in the accompanying report, NISTIR 8351), "interpretation of evidence continues to be the most difficult challenge that faces scientists, lawyers, and judges" as noted by forensic DNA pioneer Peter Gill (Gill 2019).

A study of the early literature on DNA mixture interpretation is influenced by several authors, some of whom are still active in the field. These authors include Charles Brenner (DNA View, Berkeley, California), John Buckleton (Institute of Environmental Science and Research, New Zealand), Bruce Budowle (formerly at the FBI Laboratory and now retired from the University of North Texas Health Science Center), James Curran (University of Auckland, New Zealand), Ian Evett (formerly at the Forensic Science Service and now Principal Forensic Services, United Kingdom), Peter Gill (formerly at the Forensic Science Service, United Kingdom and now University of Oslo, Norway), and Bruce Weir (formerly at North Carolina State University and now retired from the University of Washington).

## 2. **Measurement Advances**

Over the past 40 years, there has been an evolution of new technologies to enable DNA measurements including different genetic markers and commercial kits, instruments, and software (Butler 2009, pp. 43-78). Forensic DNA testing exists in this evolving environment with an increasingly complex set of technologies that have typically been developed initially for other fields, such as genomics or biotechnology. Measurement advances have come in the form of both methodologies and different genetic markers (**Figure S1.1**). The advantages and deficiencies of these methodologies and markers are described in the following sections.

1985 through 1990s | 1990s | ~1995 to present | ~2015 to present

Restriction Fragment Length Polymorphisms (RFLP) Single-Locus Probes | Reverse Dot Blot DNA Hybridization | Fluorescently-Labeled, Short Tandem Repeats (STRs) with Capillary Electrophoresis (CE) | Next-Generation Sequencing (NGS); Massively Parallel Sequencing (MPS)

Variable Number of Tandem Repeats (VNTRs) | Single Nucleotide Polymorphisms (SNPs) | Short Tandem Repeats (STRs) | STRs, SNPs, Microhaplotypes

*Poor sensitivity (no PCR), many alleles, few loci* | *Improved sensitivity (PCR), few alleles, few loci* | *High sensitivity (PCR), many alleles, many loci* | *High sensitivity (PCR), more alleles, more loci*

**Figure S1.1.** DNA typing methodologies and markers.

### 2.1. **Early Method Development and Assessment of DNA Mixtures**

Forensic DNA analysis began with *restriction fragment length polymorphism* (RFLP) techniques (Wyman & White 1980) and *variable number of tandem repeat* (VNTR) minisatellite probes (Jeffreys et al. 1985a) that typically required hundreds of nanograms of DNA to obtain results. Sizable blood or semen stains were the most commonly examined evidence in initial forensic cases.

Initial RFLP methods utilized multi-locus VNTR probes (e.g., Gill et al. 1985). The barcode-like patterns from these multi-locus probes were complex enough for samples originating from single contributors much less mixtures (Aronson 2007). Therefore, single-locus VNTR probes (Wong et al. 1987) quickly overtook the original multi-locus probes so that DNA mixtures could be more easily discerned (Kirby 1990, p. 140).

In the United States, private laboratories (e.g., Lifecodes and Cellmark) began performing casework in 1986 so the initial U.S. court cases on DNA involved these private companies (Aronson 2007). Public forensic laboratories began using DNA in some states (e.g., Virginia and California). The FBI Laboratory went “on-line” with DNA testing using RFLP methods in

December 1988 (Aronson 2007). Many state and local forensic laboratories followed over the next few years.

An early publication from Alec Jeffrey's laboratory at the University of Leicester in the UK claimed that autoradiograms of single-locus VNTR probes produced a signal "from 60 ng or less of human genomic DNA" and "depending on the genotypes of the individuals tested, … detect an admixture of 2% or less of one individual's DNA with another" (Wong et al. 1987). This same article notes: "locus-specific probes, unlike [multi-locus] DNA fingerprint probes, can be used to estimate the number of individuals represented in a mixed DNA sample" (Wong et al. 1987). In the late 1980s, in parallel with these developments in RFLP testing and its application to forensic analysis, a new technique helped improve DNA sensitivity. This involved generating millions of copies of targeted portions of each DNA molecule in a process known as the *polymerase chain reaction* (PCR).

PCR was originally developed in the mid-1980s (Saiki et al. 1985) and quickly became a valuable tool in molecular biology for examining small amounts of DNA. By the late 1980s and early 1990s, the first PCR methods were being implemented for forensic DNA testing purposes (Saiki et al. 1989, Blake et al. 1992). These initial methods were sensitive (i.e., detecting only a few cells' worth of DNA), but did not use highly polymorphic genetic markers (i.e., differentiating only a limited number of possible alleles and genotypes). And these early tests were not extremely effective in distinguishing individual components of DNA mixtures. Many of these first PCR assays utilized *single nucleotide polymorphisms* (SNPs) that typically possess only two alleles (often generically designated "A" and "B") and thus, three genotypes ("AA," "AB," or "BB").

The first forensic PCR test involved the single-locus human leukocyte antigen (HLA) DQα with 6 possible alleles and 21 possible genotypes when examined with the AmpliType HLA DQα typing kit (Cetus Corporation, Emeryville, CA) using dot blot and reverse dot blot techniques (Walsh et al. 1991). A few years later, the AmpliType PM PCR Amplification and Typing Kit, which was developed by Roche Molecular Systems (Alameda, CA) and marketed by the PerkinElmer Corporation (Norwalk, CT), added five additional loci to the HLA DQα locus (Fildes & Reynolds 1995). These kits used either a "C" (control) dot or an "S" (sensitivity) dot "designed to be the lightest dot on the nylon strip and intended to act as a threshold for evaluating stochastic effects" (Budowle et al. 1995). According to the manufacturer, "the 'S' and the 'C' dots are designed not to be visible if the amount of template DNA is less than approximately 0.3 ng to 0.5 ng" (Fildes & Reynolds 1995).

In some of the earliest reported DNA mixture experiments, the FBI Laboratory performed validation experiments with the AmpliType DQα typing kit that involved two-person DNA mixtures, with combinations of non-overlapping heterozygous genotypes spanning ratios of 1:1 to 1,000:1 with DNA quantities in the 200 ng to 200 pg (0.2 ng) range (Comey & Budowle 1991). The authors of that study note several limitations in the method used including (a) that mixture ratios appeared to matter more than the overall quantity of DNA in terms of dot intensity and (b) that shared alleles between contributor genotypes could prevent mixture detection with a single-locus system exhibiting a limited number of possible alleles.

When the first multiplex PCR kit became available to examine more than one DNA marker simultaneously, a publication containing FBI validation studies of the AmpliType PM (PolyMarker) kit discussed the ability to detect mixed body-fluid samples created by combining saliva and semen:

> "the presence of two or more contributors to a sample generally is inferred by the presence of unbalanced dots and/or … extra dots in [two of the loci which were tri-allelic SNPs]" (Budowle et al. 1995).

However, the authors also note:

> "the exact percentage of samples that exhibit unbalanced allele dot intensities is difficult to determine, because the determination of unbalanced intensity is somewhat subjective." This study found that "the minor contributor of a 1:20 mixture of two samples was barely detectable, and the allele dot for the minor component was less intense than the S dot" (Budowle et al. 1995).

This study concludes:

> "Because of the potential for unbalanced allele dot intensities and the limitations for detecting some mixed samples containing equivalent amounts of DNA, caution should be exercised when interpreting evidentiary samples that potentially may be from more than one donor" (Budowle et al. 1995).

Thus, the FBI alerted specialists of the challenges posed by multi-donor samples.

A study involving seven laboratories, organized by the manufacturer of the AmpliType PM PCR Amplification and Typing Kit, was published about the same time as the FBI study detailed above. The publication described the kit's ability with mixture detection a little differently than the FBI researchers. Authors of the study wrote:

> "The balance of dots within a locus of the PM DNA probe strip proved to be a valuable asset of the system for the analysis of mixtures. This feature is an important benefit of the PM system since a high percentage of forensic casework involves the analysis of sexual assault samples" (Fildes & Reynolds 1995).

The differences in perspectives highlighted here illustrate that sometimes a disconnect can exist between researchers and commercial suppliers in the types of studies performed and the language used in sharing their results. For example, compare

> "because of the potential for unbalanced allele dot intensities, … caution should be exercised" (Budowle et al. 1995)

versus

> "the balance of dots … proved to be a valuable asset of the system for the analysis of mixtures" (Fildes & Reynolds 1995).

This observation exemplifies the reasoning of the President's Council of Advisors on Science and Technology (PCAST), who wrote in their 2016 report:

> "While it is completely appropriate for method developers to evaluate their own methods, establishing scientific validity also requires scientific evaluation by other scientific groups that did not develop the method" (PCAST 2016, p. 80).

In the early to mid-1990s when the AmpliType PM kit was used, most DNA mixtures seen in forensic laboratories derived from "incomplete separation of the sperm and female epithelial cell fractions from postcoital swab extractions" (Fildes & Reynolds 1995; see Box S1.1). After reviewing the field trial results from seven forensic laboratories, the authors of this study commented:

> "The potential for sample mixtures in forensic casework analysis has always required careful and thoughtful interpretation. Individual laboratories will need to develop their own policies for the interpretation of mixtures based on their experience and case history information" (Fildes & Reynolds 1995).

**Box S1.1. Differential Extraction**

Many forensic samples, particularly those coming from vaginal swabs collected from a sexual assault victim, typically contain DNA from both the victim and the perpetrator. In the 1985 *Nature* article that launched forensic DNA analysis, authors Peter Gill and David Werrett from the UK Forensic Science Service and Alec Jeffreys from the University of Leicester introduced differential extraction as a method to separate the perpetrator's sperm cells from the victim's epithelial cells based on the chemical composition of the sperm head (Gill et al. 1985). When DNA mixtures cannot be resolved into single-source components through techniques such as differential extraction, mixture interpretation is required.

The developers of these early PCR test kits encouraged users to avoid interpreting low levels of DNA (i.e., attempting to interpret results below their "C" or "S" dots) to avoid problems with unbalanced allele detection. In a 1992 article, they note:

> "Preferential amplification due to stochastic fluctuation can occur when amplifying very low amounts of target DNA molecules; the possibility of an unequal sampling of the two alleles of a heterozygote is increased when only a few DNA molecules are used to initiate PCR. This problem can be avoided by adjusting the cycle number such that approximately 20 or more copies of target DNA [i.e., >120 pg genomic DNA assuming 6 pg per diploid copy of the genome] are required to give a typing result for that PCR system" (Walsh et al. 1992).

For the first decade of DNA testing (circa 1985 to 1995), where many nanograms of DNA were required to obtain a result, most of the samples examined involved visible bloodstains or sexual assault evidence. This meant that only a limited number of mixtures were observed in casework during the 1990s. For example, a review of DNA casework in a Spanish laboratory from 1997 through 2000 reported observing less than 7% mixture profiles (Torres et al. 2003). If mixtures were observed, they were often treated as "uninterpretable" (e.g., Fildes & Reynolds 1995).

### 2.1.1. **The Rise and Impact of STR Methods**

By the mid-1990s, the field began to move towards multi-allelic *short tandem repeat* (STR) markers where multiple STR loci could be co-amplified and labeled using multiplex PCR (Caskey et al. 1989, Edwards et al. 1991, Frégeau & Fourney 1993, Kimpton et al. 1993). STR markers assist mixture interpretation due to the existence of many alleles per marker (sometimes a dozen or more) compared to two and sometimes three alleles present in SNP loci (Butler et al. 2007), such as those used in the AmpliType PM kit. In the 1990s, the UK Home Office's Forensic Science Service (FSS) led the forensic community in advancing knowledge of STR markers and their application to forensic science including DNA mixture interpretation (Gill et al. 1995, Gill et al. 1997, Clayton et al. 1998, Gill et al. 1998, Evett et al. 1998).

Efforts were also made to extend interpretation of STR typing results to DNA quantities originating from less than approximately 15 to 20 cells (≈100 pg to 120 pg) (Gill et al. 2000) – a limit that had previously been recommended to avoid stochastic effects seen with earlier PCR marker systems (Walsh et al. 1992).

For the United States, a core set of 13 STR markers were selected in 1997 (Budowle et al. 1998) and two decades later expanded to 20 STR markers (Hares 2015)

Commercial STR kits, either from Promega Corporation (Madison, WI) or Applied Biosystems[3] (previously Foster City and now South San Francisco, CA), have been widely used since the late 1990s to enable forensic DNA testing. Later, Qiagen (Hilden, Germany) began offering STR typing kits.

While STRs have the benefit of many possible alleles per tested marker, making it more likely to detect contributor genotypes when they exist in mixtures, artifacts known as stutter products arise during the PCR amplification process that complicates the interpretation process (Butler 2014, pp. 70-79). The presence of these stutter products increases interpretation uncertainty when attempting to consider potential minor contributors where the stutter products of alleles from the major contributor(s) are present in similar amounts as the alleles of minor contributor(s). When the potential minor contributor alleles become indistinguishable from the stutter products of the major contributor alleles, then this ambiguity needs to be accounted for in the statistical treatment of the strength of evidence (Gill et al. 2006). Methods for doing so are discussed in Section 3.2 involving probabilistic genotyping.

For more than 25 years, STR typing has been performed using fluorescently-labeled PCR products followed by capillary electrophoresis (CE) separation and laser-induced fluorescence (LIF) detection (Butler et al. 2004, Butler 2011, pp. 141-165). With the use of allelic ladders and internal size standards, each STR allele and genotype may be reliably measured using the overall PCR product length and dye color (Butler 2014, pp. 47-127). Using this approach, national DNA databases containing millions of STR profiles have been constructed from single-source DNA reference samples – and a vast amount of forensic casework involving DNA mixtures has been performed over the years.

[3] Applied Biosystems has undergone multiple name changes over the years and in 2019 is known as Thermo Fisher Scientific (for ten names spanning 1981 to 2014, see Butler 2014, p. 26).

Yet the quest continues for additional markers and methods that might improve DNA mixture interpretation.

## 2.2. **Increased Sensitivity in DNA Test Methods**

The PCR method can be tuned to amplify and recover low quantities of DNA through increasing the number of amplification cycles or amount of PCR product injected into a genetic analyzer (see Butler 2011, pp. 311-346). As early as 1997, researchers demonstrated that with such tuning STR typing results could be obtained from DNA found in a single cell (Findlay et al. 1997). This capability encouraged attempts to recover DNA profiles from invisible samples left by touching or speaking over a surface, for example. Some laboratories, rather than using a specific enhanced detection method such as an increased number of PCR cycles, pushed the limits of their existing protocols by expanding their sampling approaches to include smaller and smaller quantities of biological material.

While information from invisible samples (sometimes called "touch evidence[4]" or "trace DNA") can be helpful in an investigation, increasing the sensitivity of the PCR method to obtain results from invisible samples can impact reliability and relevance. From a historical perspective, this increase in DNA test method sensitivity and willingness to attempt examination of smaller quantities of DNA have resulted in an increase in samples and sample types submitted to forensic laboratories. This has led to more mixtures being observed, and to development of modern interpretation techniques involving probabilistic genotyping software (see Section 3.2).

### 2.2.1. **Low Copy Number (LCN) Efforts**

As leaders in developing and implementing forensic DNA methods during the 1990s and early 2000s, the UK's Forensic Science Service ventured into increased sensitivity (Findlay et al. 1997, Gill et al. 2000) and new approaches for the interpretation of evidence (Gill et al. 2007). The FSS method was initially referred to as *low copy number* (LCN) DNA testing and later as low template DNA (LT-DNA). The original FSS LCN method involved an in-house 6-plex STR assay or a commercial STR kit 10-plex amplified with 34 cycles (Gill et al. 2000, Whitaker et al. 2001) rather than the traditional 28 cycles widely used at the time (e.g., Wallin et al. 1998).

In the foundational article "An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA," the authors report:

> "By increasing the PCR amplification regime to 34 cycles, we have demonstrated that it is possible routinely to analyze <100 pg DNA [representing around 15 diploid cells]…Compared to amplification of 1 ng DNA at 28 cycles, it was shown that increased imbalance of heterozygotes occurred, along with an increase in the size (peak area) of stutters. *The analysis of mixtures by peak area [or height] measurement becomes increasingly difficult*…Laboratory-based contamination cannot be completely avoided,

[4] Touch evidence or touch DNA is a misleading term as it infers an activity that may not be relevant to how the DNA got there (see EWG 2024, p. 172).

> even when analysis is carried out under stringent conditions of cleanliness…" (Gill et al. 2000, emphasis added).

Attempts to recover information from low amounts of DNA present in evidentiary samples using LCN methods inevitably led to increased imbalance in heterozygotes, higher levels of stutter products, allele drop-out, and allele drop-in (contamination). These phenomena are artifacts of stochastic, or random sampling, effects that occur in the early cycles of PCR amplification when there are a limited number of target molecules to amplify (Butler & Hill 2010).

To alleviate stochastic effects, LCN protocols typically involve the formation of consensus profiles using replicate amplifications from aliquots of a DNA extract. Alleles that are observed in replicate amplifications are deemed "reliable" and form a consensus profile (e.g., Benschop et al. 2011). While attempting to replicate alleles from single-source samples is straightforward, the replicate tests are unlikely to maintain relative peak height ratios and mixture ratios needed for traditional DNA mixture deconvolution and interpretation. It was recognized in the seminal LCN article that "these guidelines [for replicate testing and building consensus profiles] will be superseded by expert systems utilizing the Bayesian principles described in this paper" (Gill et al. 2000). However, such expert systems would not be available for almost another decade.

### 2.2.2. **Reliability Concerns with Increased Sensitivity**

A judge's ruling in the Omagh (Northern Ireland) terrorist bombing trial in 2007 raised concerns about the reliability of the FSS LCN method used in the case. This ruling, in turn, led to a formal investigation of LCN, some intense debates in several scientific meetings (see Butler 2011, pp. 313-319), and the creation of the UK Forensic Science Regulator, which monitored quality assurance. Concerns regarding the use of LCN in criminal casework had been raised previously (e.g., Budowle et al. 2001), but this case revived scrutiny.

In addition, there was increased recognition of the challenges that higher-sensitivity DNA results brought to DNA mixture interpretation. For example, a December 2007 article states:

> "With the improved sensitivity of modern DNA methods coupled with the increased use of forensic genetics in crime case investigations, the number of DNA mixtures composed of full or partial DNA profiles from two or more contributors has increased dramatically" (Morling et al. 2007).

The authors continue:

> "The biostatistical interpretation of mixed DNA profiles is a challenge – especially if DNA profiles are incomplete" (Morling et al. 2007).

### 2.2.3. **Relevance Concerns with Increased Sensitivity**

Along with the aforementioned sensitivity efforts, it was recognized early on that low levels of DNA template on items or surfaces might not be associated with the crime, but rather left innocently before or after the crime occurred (Gill 2001). Secondary or tertiary transfer of DNA due to casual

contact, such as hand shaking, has been shown to vary. This variance is based on what has been termed the "shedder" status of the individuals involved (Lowe et al. 2002). Even as far back as 1997 in a landmark study in the journal *Nature,* "DNA fingerprints from fingerprints" (van Oorschot & Jones 1997) discussed the possibility of DNA transfer.

A 2013 review article "DNA transfer: review and implications for casework" increased awareness of relevance concerns with "trace DNA," which the authors termed "DNA that cannot be attributed to an identifiable body fluid" (Meakin & Jamieson 2013). When DNA cannot be attributed to an identifiable body fluid, it can no longer address source level questions on the hierarchy of propositions (offense, activity, and source levels), which were outlined by the FSS in 1998 (Cook et al. 1998b). Thus, sub-source (Gill 2001) and even sub-sub-source levels (Taylor et al. 2014, Taylor et al. 2018) became part of DNA mixture interpretation considerations[5].

As discussed in Chapter 5 of the accompanying report (NISTIR 8351), there is still a lot to learn in this area and many gaps remain to be filled (e.g., Burrill et al. 2019, van Oorschot et al. 2019, van Oorschot et al. 2021). Assigning likelihood ratios with activity-level propositions have been suggested as the most appropriate approach to dealing with small quantities of DNA detected due to increased sensitivity of DNA tests. In some cases, there has been a shift in focus by the court from questions about the source of the DNA to the mechanism by which it was deposited (Taylor et al. 2018, Gill et al. 2020). This topic is discussed in more depth in Chapter 7 of the DNA Human Factors report (EWG 2024, pp. 172-182).

### 2.3. **Future of Measurement Advances**

Measurement advances to date (Table S1.1) have both pushed the field forward and created new challenges, for example, the need to deconvolute complex mixtures after a DNA profile has been generated. Current and future efforts are looking to address the issues that increased sensitivity has brought about. With single-cell analysis, individual cells from multiple contributors can be separated prior to DNA extraction and PCR amplification and thereby reduce the mixture sample complexity when the individual DNA profiles are generated. Several groups are working on refining these single-cell analysis techniques (Duffy et al. 2023, Grgicak et al. 2024, Huffman & Ballantyne 2023a, Huffman & Ballantyne 2023b, Huffman et al. 2023, Kulhankova et al. 2023, Kulhankova et al. 2024, Schulte et al. 2023, Schulte et al. 2024). Individual-specific mixture deconvolution has been demonstrated with mixtures of up to 11 individuals, balanced mixtures containing as few as 20 cells (10 per each individual), and imbalanced mixtures with a ratio as low as 1:80 (Kulhankova et al. 2024).

Another factor that adds to the interpretation complexity is that contributors to a DNA mixture may share length-based alleles (e.g., both have 13 repeats) that could potentially be resolved from one another if the sequences of these alleles differ (e.g., one has 12 GATA repeats plus a CATA repeat and the other has 13 GATA repeats). *Massively parallel sequencing* (MPS) methods enable resolution of STR alleles by sequence, which can improve the ability of estimating the

[5] Sub-source level of the hierarchy of propositions refers to examining a DNA profile, without necessarily knowing the cellular source of the sample while sub-sub-source refers to examining a portion of a DNA mixture profile, such as the major contributor.

number of contributors in a mixture (Young et al. 2019). With MPS, more STR loci can be examined simultaneously than with CE. In addition, other genetic markers known as *microhaplotypes*, which do not exhibit the stutter products seen with STRs, have shown potential for mixture interpretation when contributors are present in extremely unbalanced ratios (e.g., Bennett et al. 2019).

Time will tell whether a new set of microhaplotype markers analyzed by MPS will replace or supplement traditional STR analysis with CE for DNA mixture interpretation (see Butler 2015a). For more information, see Chapter 6 in the accompanying report (NISTIR 8351).

**Table S1.1**. Timeline summary for some key DNA measurement advances.

| Year | Event |
|---|---|
| **1980** | Restriction fragment length polymorphism (RFLP) marker demonstrated to vary in human DNA from 56 individuals examined (Wyman & White 1980) |
| **1985** | RFLP variable number of tandem repeat (VNTR) multi-locus system (Jeffreys et al. 1985a) used with immigration/relationship testing (Jeffreys et al. 1985b) and forensic DNA applications (Gill et al. 1985); differential extraction method described as a means to physically separate sperm and epithelial cells to reduce DNA mixtures in sexual assault cases (Gill et al. 1985); PCR as a means to copy specific DNA target regions (Saiki et al. 1985) |
| **1986** | DNA testing goes public with Cellmark and Lifecodes in the United States; Pitchfork case in England involves first person exonerated by DNA testing (Gill & Werrett 1987) |
| **1990** | PCR methods with DQα begin but are not very effective for mixtures (Blake et al. 1992) |
| **1991** | Fluorescently labeled short tandem repeat (STR) markers introduced (Edwards et al. 1991) |
| **1995** | UK National DNA Database launched (Werrett 1997); capillary electrophoresis (CE) became available with multi-color fluorescence detection (Butler 2011, p. 142) |
| **1997** | Demonstration of single cell STR analysis (Findlay et al. 1997) and DNA recovery from touched surfaces (van Oorschot & Jones 1997); selection of the FBI Combined DNA Index System (CODIS) 13 core STR loci (Budowle et al. 1998) |
| **2000** | Multiple STR kits become available for analysis of 13 CODIS STR loci |
| **2001** | Low copy number (LCN) method in the UK uses 34 cycle PCR (Whitaker et al. 2001) |
| **2007** | Beginning with MiniFiler (Mulero et al. 2008), a new generation of STR kits become more sensitive to improve performance for low quantity and poor quality DNA samples |
| **2009** | European Standard Set (ESS) expands from 7 to 12 STR loci (Schneider 2009) |
| **2017** | United States expands the CODIS core set to 20 STR loci including ESS loci (Hares 2015) |
| **2018** | Measurement capabilities for DNA mixtures demonstrated with STR allele sequencing (e.g., Gettings et al. 2018) and microhaplotype markers (e.g., Bennett et al. 2019) |
| **2024** | Single-cell analysis used to reduce mixture complexity (e.g., Grgicak et al. 2024) |

## 3. **Interpretation Advances**

The presence of a mixture can be identified by the observation of more than two alleles at an STR locus. Also, the occurrence of more than two alleles will typically be seen at two or more loci in the DNA profile for almost all mixtures. Exceptions exist for any rule though. Occasionally tri-allelic patterns have been reported at one STR locus in a single-source DNA profile (e.g., Clayton et al. 2004). Artifacts, such as stutter products created due to strand slippage during PCR amplification of STR markers (see chapter 3 in Butler 2014), can give rise to additional DNA peaks and increase the complexity and challenge of mixture interpretation. For this reason, guidelines have been developed and refined over the past several decades to assist in designating STR alleles versus artifacts and interpreting DNA profiles (Gill et al. 1997, SWGDAM 2000, SWGDAM 2010, SWGDAM 2017).

Advances in approaches for DNA interpretation have followed advances in DNA measurement methods described in the previous section. For example, improved sensitivity resulting from PCR amplification meant that control and stochastic dots were needed with reverse dot blot hybridization assays to help decipher whether allele dropout was a possibility (Budowle et al. 1995, Fildes & Reynolds 1995).

However, not everyone in the field moves forward to new approaches at the same time, so variation exists and can be seen when assessing DNA mixture interpretation from the same samples across multiple laboratories (see Section 4). Drivers for change in approaches may come from observing variation compared to others following participation in interlaboratory studies or when commercial solutions become available and are implemented by a specific laboratory.

### 3.1. **Initial Interpretation Approaches Explored for DNA Mixtures**

In some of the first articles describing mixture interpretation with STR markers, Peter Gill and his FSS colleagues noted the need to understand heterozygote peak balance within each locus to conduct mixture analysis (Gill et al. 1995, Gill et al. 1997). They point out that "interpretation of mixtures also needs to take account of the possible confusion between a true mixture and the presence of stutter bands" (Gill et al. 1995), which was described in more detail as part of the International Society for Forensic Genetics (ISFG) DNA Commission recommendations about a decade later (Gill et al. 2006). Based on their observations with a six-locus STR multiplex in use at the time, these FSS researchers share:

> "If the mixture [has components in the ratio of] 1:5 then reliable identification of the components of a [two-person] mixture is normally possible" (Gill et al. 1995).

They continue:

> "When mixtures are observed, and the components cannot be separated, there will inevitably be occasions when it will be more appropriate to present all the possible alternatives using statistical methods described by Evett et al. [Evett et al. 1991]" (Gill et al. 1995).

DNA mixture interpretation considers possible genotype combinations that could create the observed data. Different statistical approaches have been used to describe mixture results (Box S1.2). Some forensic laboratories may use one of these approaches for a specific type of sample result and another approach with a different category of mixture as permitted by an early recommendation of the DNA Advisory Board (DAB 2000).

**Box S1.2. Statistical Approaches Used for DNA Mixture Interpretation**
(as defined by SWGDAM 2017)

**RMP (random match probability)**: the probability of randomly selecting from the population an unrelated individual who could be a potential contributor to an evidentiary profile

**CPI (combined probability of inclusion)**: produced by multiplying the probabilities of inclusion from each locus; probability of inclusion is the percentage of the population that can be included as potential contributors to a DNA mixture at a given locus; also known as Random Man Not Excluded (RMNE)

**LR (likelihood ratio)**: the ratio of two probabilities of the same event under different and mutually exclusive hypotheses; typically, the numerator contains the prosecution's hypothesis and the denominator the defense's hypothesis

In 1991, Ian Evett of the FSS and several colleagues introduced a likelihood ratio (LR) approach (Evett et al. 1991). In this initial mixture interpretation article, which uses examples from RFLP single-locus probes available at the time, the authors note:

> "This paper has been *restricted to fairly simple case situations*; as the number of bands increases the evaluation is liable to become quite complicated. Also, it is important for caseworkers to recognize that *the evidential strength falls rapidly with increasing number of bands*…" (Evett et al. 1991, emphasis added).

The authors also observe:

> "In some cases, where there are unequal band intensities, it may be possible to determine which bands are paired. Thus, two very strong bands might be clearly distinguished from two weak ones. However, this would have to be considered carefully because there can be differences in intensities between the two bands from one individual…" (Evett et al. 1991).

Thus, this initial article using an LR approach recognizes the challenge of accounting for an increasing number of alleles as the number of contributors goes up. Furthermore, the article emphasizes that reliable allele pairing into contributor genotypes may be difficult and needs "to be considered carefully" because of the natural variation in heterozygote allele balance, which increases with lower amounts of starting DNA template.

### 3.1.1. **Early U.S. Mixture Approaches – The NRC I and NRC II Reports**

While LR approaches for mixtures were under development in the UK in the early 1990s, the National Research Council (NRC) in the United States completed a report in 1992 that mentions

the combined probability of inclusion (CPI) as an appropriate method for mixture interpretation:

> "If the samples are mixtures from more than one person, one should see additional bands for all or most polymorphic probes, but not for a single-copy monomorphic probe. Mixed samples can be very difficult to interpret, because the components can be present in different quantities and states of degradation. It is important to examine the results of multiple RFLPs, as a consistency check. *Typically, it will be impossible to distinguish the individual genotypes of each contributor*. If a suspect's pattern is found within the mixed pattern, the appropriate frequency to assign such a 'match' is the sum of the frequencies of all genotypes that are contained within (i.e., that are a subset of) the mixed pattern" (NRC 1992, p. 59, emphasis added).

Thus, this early report recognizes some of the difficulties in mixture interpretation including distinguishing contributor genotypes when components vary in quantity and quality.

The NRC 1992 report emphasizes the following five principles: (1) that polymorphic loci containing many possible alleles enable mixtures to be more easily detected, (2) mixtures are complicated by the ratio of contributors and their possible states of degradation, (3) checking the consistency of the mixture across multiple loci aids quality assurance, (4) distinguishing the individual genotypes of each contributor is not always possible, and (5) when individual contributor genotypes cannot be distinguished, the CPI statistic should be used, which involves summing the frequencies of all genotypes that are contained with the mixed pattern.

It is important to note that at the time the first NRC report was written, high-quantities of DNA were needed to obtain an RFLP result and therefore the possibility of allele drop-out was not considered an issue. As emphasized in a 2016 publication (Bieber et al. 2016), the CPI statistic is only fit-for-purpose at a tested locus if all alleles of all contributors present are detected in the DNA mixture. In other words, the CPI statistic cannot be applied to DNA mixture profiles with potential allele drop-out because it would not fully account for all possible genotypes. Therefore, the CPI statistic is not suitable for use with DNA mixture profiles containing low levels of DNA.

A second NRC report published in 1996 (known as the NRC II, NRC 1996), was intended to replace the 1992 report. The NRC II report observes:

> "In many cases, one of the contributors – for example, the victim – is known, and the genetic profile of the unknown is readily inferred. In some cases, it might be possible to distinguish the genetic profiles of the contributors to a mixture from differences in intensities of bands in an RFLP pattern or dots in a dot-blot typing; in either case, the analysis is similar to the unmixed case. However, when the contributors to a mixture are not known or cannot otherwise be distinguished, a likelihood-ratio approach offers a clear advantage and is particularly suitable" (NRC 1996, pp. 129-130).

The report references a simple RFLP case example in which there are four distinguishable alleles coming from two individuals – and the CPI calculation is performed as recommended from the 1992 NRC report, page 59.

The NRC II report continues:

> "That [CPI] calculation is hard to justify because it does not make use of some of the information available, namely, the genotype of the suspect. The correct procedure, we believe, was described by Evett et al. (1991)" (NRC 1996, p. 130).

After working through this example, the NRC II report notes:

> "We *have considered only simple cases*. With VNTRs, it is possible, though very unlikely, that the four bands were contributed by more than two persons, who either were homozygous or shared rare alleles. *With multiple loci, it will usually be evident if the sample was contributed by more than two persons*. Calculations taking those possibilities into account could be made if there were reason to believe that more than two persons contributed to the sample. Mixed samples are often difficult to analyze in systems where several loci are analyzed at once…. The problem is complex, and some forensic experts follow the practice of making several reasonable assumptions and then using the calculation that is most conservative. For a fuller treatment of mixed samples, see [Weir et al. (1997)]" (NRC 1996, p. 130, emphasis added).

This report discusses the benefits of an LR approach with consideration of the suspect's genotype in the context of the case and notes that multiple assumptions and calculations may be needed particularly when going beyond simple cases.

About a decade after the NRC II report was released, an article was written discussing the merits of CPI and LR approaches (Buckleton & Curran 2008). The authors noted that LR results must assume a number of contributors and are more difficult to present in court. On the other hand, CPI (RMNE) statistics waste information and cannot be interpreted directly in the context of potential propositions of interest in a court case.

### 3.1.2. **First High-Profile Case with DNA Mixtures**

In June 1994, U.S. football star O.J. Simpson was accused of murdering his ex-wife Nicole Brown and her friend Ronald Goldman. The trial was televised and became a worldwide event with DNA evidence playing a prominent role in the trial (Weir 1995). Decisions during the O.J. Simpson case may well have impacted the early trajectory of mixture interpretation in U.S. courts and forensic laboratories (Box S1.3).

Shortly after he appeared as a prosecution witness in the Simpson trial, Professor Bruce Weir, then at North Carolina State University, wrote:

> "Reliance on the [1992] NRC report prevented an adequate treatment of mixtures and population structure in the Simpson trial" (Weir 1995).

He continues:

> "It is incumbent on both prosecution and defense to explain the meaning of a conditional probability of a DNA profile… Simple frequencies do not address the issue of mixtures. When there are several contributors to a bloodstain (a mixed stain), the probability calculations can become quite complex…" (Weir 1995).

**Box S1.3. Impact of O.J. Simpson Trial on U.S. Approach to DNA Mixtures**

Experiences from the O.J. Simpson "trial-of-the century" in 1995 brought "the nature and strength of DNA evidence to wide public notice" (Weir 1995) and aided adoption of quality control measures in forensic DNA laboratories (see Butler 2009, pp. 84-85). Due to concerns raised during the trial, procedures for biological evidence collection and storage in many forensic laboratories were improved going forward. Within a few years, the FBI's DNA Advisory Board created the initial Quality Assurance Standards (QAS), which have been widely used and revised several times since then (e.g., QAS 1998, QAS 2020).

Another impact of the O.J. Simpson trial was that prior to this case in 1995, "no U.S. court had ever heard statistical testimony concerning mixtures" (Weir 2000). Towards the end of the trial, when reviewing statistics for DNA mixtures involved, Judge Ito denied the admissibility of likelihood ratio (LR) calculations performed (Kaye 2010, pp. 200-209). Reliance on the NRC I 1992 report apparently influenced this decision (Weir 1995), which may very well have delayed wider adoption of the LR approach in the United States for many years (Weir 2000) even though the NRC II 1996 report would be supportive of LRs (NRC 1996, p. 130). The trial experience did have the benefit of renewing the interest of some members of the community to work on improving interpretation of mixtures (e.g., Weir et al. 1997, Buckleton et al. 1998) and led to the first book on the topic (Evett & Weir 1998).

To improve mixture interpretation and remove some of the misconceptions that arose during the O.J. Simpson trial, Professor Weir and his collaborators began research that enabled the field to move forward in significant ways with DNA mixture interpretation.

### 3.1.3. **Development of LR Methods**

In March 1997, Professor Bruce Weir and colleagues from New Zealand published an article titled "Interpreting DNA Mixtures" in the *Journal of Forensic Sciences* that described LR calculations with two-person mixtures based on assuming independence of alleles within and between loci (Weir et al. 1997). An example was even worked from an RFLP mixture result in the O.J. Simpson case using the "2p" rule. This rule had been introduced in the NRC II report for single-banded VNTR loci used in RFLP but declared inappropriate for PCR-based systems (see NRC 1996, p. 5). However, the authors note:

> "The '2p' rule is not always conservative, and we suggest caution in its use" (Weir et al. 1997).

Commenting on the value of LR calculations compared to the CPI approach, Professor Weir and colleagues state:

> "Interpretations based simply on the frequencies with which random members of a population would not be excluded from a mixed-stain profile [i.e., CPI] do not make use of all the information, and may overstate the strength of the evidence against included people," and they emphasize "only by comparing the probabilities of the evidentiary

> profile under alternative explanations [i.e., using LRs] is it possible to arrive at a complete analysis of mixtures" (Weir et al. 1997).

Thus, from the very beginning of mixture interpretation efforts, LR methods were emphasized as being superior to CPI calculations.

An important aspect of LR methods involves the number of potential contributors. Weir wrote:

> "the [LR] results given so far depend on the number of contributors to the mixed sample" (Weir et al. 1997).

Referring to an article from Charles Brenner, Rolf Fimmers, and Max Baur (the latter two of whom are German mathematicians) (Brenner et al. 1996), Professor Weir and colleagues note:

> "Whenever there is doubt as to the number of contributors, there can be considerable variation in the likelihood ratio." (Weir et al. 1997).

Using the formulas outlined in this initial article (Weir et al. 1997) and a follow-up one allowing for population sub-structure (Curran et al. 1999), a software program named *DNAMIX* was developed (Storey & Weir 1998). It is important to keep in mind that *DNAMIX* was built at a time when fairly high levels of DNA were being tested and was not designed to account for the possibility of allele drop-out.

#### 3.1.4. **Mixture Deconvolution**

The UK Forensic Science Service, using in-house developed STR assays, published several landmark articles on mixture interpretation in 1998. This included approaches to using peak heights/areas to enable mixture deconvolution with simple two-person mixtures (Clayton et al. 1998, Evett et al. 1998, Gill et al. 1998). These articles are foundational and a valuable addition to training programs in DNA mixture interpretation.

In their article in which allele peak areas were used to interpret simple STR mixtures, FSS researchers examined 39 different mixtures prepared from five different individuals with mixture ratios ranging from 1:10 to 10:1 (Gill et al. 1998). This article also notes that with lower quality data, there was a poor fit to the model and so the correct genotype did not rank as well, and they suggest "caution should be exercised with low peaks" (Gill et al. 1998). In addition, this is the first attempt to define a "complex mixture" as a profile containing "more than four alleles at any locus" – and the authors note that their method does not apply to these complex mixtures (Gill et al. 1998).

Three important points and principles suggested by the authors (Gill et al. 1998):

(1) The lower the peak heights, the higher the variability in relative peak heights due to stochastic variation in PCR amplification of the mixture components. The report noted:

> "if the peak areas [or heights] are low, then the relative peak areas [or heights] become less predictable for a given mixture."

In other words, the variability and uncertainty in relative peak heights increase as overall peak heights decrease.

(2) The reproducibility of mixture results and relative peak heights of mixture components should be verified through repeated testing if there is sufficient DNA available. The authors of the article state:

> "it is important to repeat the experiment – possibly at a higher concentration of DNA."

(3) Reducing the number of loci, simultaneously amplified, improves the relative peak balance.

The authors write:

> "Singleplex analysis (where just a single locus is amplified) is another option, to improve the signal strength" and "also improve the relative peak balance, so that peak areas better reflect the actual DNA concentration" (Gill et al. 1998).

The first commercial STR kits were becoming available at the same time that the FSS was sharing their mixture interpretation results. In the mid-1990s, Promega Corporation released triplex and quadruplex kits abbreviated CTT and CTTV using the first initials of the STR loci CSF1PO, TPOX, TH01, and vWA. Applied Biosystems followed Technical Working Group on DNA Analysis Methods (TWGDAM) guidance when validating their first STR kit "AmpFlSTR Blue," which was a triplex amplifying DNA markers D3S1358, vWA, and FGA (developmental validation published in Wallin et al. 1998). From these studies with two-person mixtures, which explored ratios of 1:1 to 1:50 at 1 ng or 5 ng total DNA template, they concluded:

> "The limit of detection for mixtures in which a total of 1 ng was amplified was 1:10" while "the limit of detection of mixtures in which a total of 5 ng was amplified was 1:30" (Wallin et al. 1998).

Therefore, the limit of detection for the minor contributor was in the range of 100 pg to 160 pg. These authors summarized:

> "Taken together, these mixture studies indicate that it is possible to detect a mixture and sometimes resolve the genotypes of each contributor, depending on the genomic DNA ratios, number of contributors, and particular combination of alleles present" (Wallin et al. 1998).

Experience gained from these early studies, as well as the increasing sensitivity of DNA tests (Gill et al. 2000) that quickly followed, would lead to the first international recommendations on DNA mixture interpretation (Gill et al. 2006). Software programs were also developed to assist with mixture deconvolution (e.g., Bill et al. 2005, Wang et al. 2006).

### 3.2. **Probabilistic Genotyping Software (PGS)**

As techniques for obtaining DNA results from low amounts of DNA template were implemented around the beginning of the 21st century (e.g., Gill et al. 2000) and laboratories began expanding the sample types they were willing to attempt to analyze, dealing with the possibility of allele drop-out and missing information from DNA mixture profiles became important. This led to thinking probabilistically about DNA data (e.g., Balding & Buckleton 2009, Kelly et al. 2014) and the development of probabilistic genotyping software (PGS) systems.

### 3.2.1. **Development of PGS**

In the late 1990s, the UK Forensic Science Service proposed the use of computer programs to assist in DNA mixture interpretation (Evett et al. 1998) and developed the initial theory for probabilistic genotyping. This theory incorporated the probability of drop-out when examining low quantities of DNA (Gill et al. 2000). At this same time, Cybergenetics (Pittsburgh, PA) was developing computer software to aid DNA mixture interpretation – first with automated methods to cope with stutter products (Perlin et al. 1995) and then with mathematics to assist in deconvolution of mixture components (Perlin & Szabady 2001). U.S. patents on using a computer to calculate a likelihood ratio from a DNA mixture, which were filed as early as 2001, have been awarded to Cybergenetics and its TrueAllele software (e.g., Perlin 2017).

The FSS simulated and modeled each of the steps in the DNA analysis and interpretation process (Gill et al. 2005) and created the LoComatioN software (Gill et al. 2007) to assist with allele drop-out, which regularly occurs when examining low amounts of DNA template (Balding & Buckleton 2009). In addition, non-contributor assessments to explore the performance of probabilistic models were advocated (e.g., Gill & Haned 2013). As explained in further detail in Chapter 2 of NISTIR 8351, PGS systems are either (1) *discrete* (also called semi-continuous) if only alleles are considered or (2) *continuous* (also called fully-continuous) if peak height information is utilized (see Kelly et al. 2014).

Aspects of this FSS work were implemented in the LiRa system by former members of the FSS (Puch-Solis & Clayton 2014). David Balding also developed likeLTD (Balding 2013) which forms the basis of Lab Retriever (Inman et al. 2015). With European Union funding and an open-source software initiative, LRmix (Haned et al. 2012, Prieto et al. 2014) and EuroForMix (Bleka et al. 2016) were developed and tested.

STRmix was developed by Duncan Taylor in South Australia and John Buckleton and Jo-Anne Bright in New Zealand (Taylor et al. 2013). STRmix was implemented in forensic laboratories across Australia and New Zealand in late 2012 and international sales began in early 2014. Developmental validation, which followed the SWGDAM 2015 guidelines (SWGDAM 2015), was published two years later (Bright et al. 2016).

### 3.2.2. **Movement to PGS in the United States**

An increased awareness of the benefits of PGS for interpreting complex mixtures came at the NIST-FBI DNA Technical Leaders' Summit held in Norman, Oklahoma in November 2013 (see Table 6.5 in Butler 2014), where more than 95% of public U.S. forensic laboratories were represented. The following June, a weeklong PGS workshop in St. Louis, Missouri sponsored by the Midwestern Association of Forensic Scientists (MAFS) informed attendees regarding the various software programs and their capabilities. At MAFS, vendors were provided an opportunity to demonstrate their PGS systems and answer questions. Concurrently, the Scientific Working Group on DNA Analysis Methods (SWGDAM) was drafting Guidelines for the Validation of Probabilistic Genotyping Systems, published the following year (SWGDAM 2015).

The first PGS publications in the U.S. came from Mark Perlin of Cybergenetics, demonstrating his fully-continuous TrueAllele Casework software; several of these articles were written in

collaboration with scientists from the New York State Police (NYSP) or the Virginia Department of Forensic Science (VDFS) (Perlin et al. 2009, Perlin & Sinelnikov 2009, Perlin et al. 2011, Perlin et al. 2013, Perlin et al. 2014). The performance of TrueAllele Casework with two-, three-, and four-person mixtures was also explored by VDFS (Greenspoon et al. 2015) and results with five-person mixtures were described in another study involving the Kern County (California) Regional Crime Laboratory (Perlin et al. 2015).

A discrete PGS system known as the Forensic Statistical Tool (FST) was developed in-house by the New York City Office of Chief Medical Examiner (OCME) to account for the possibility of allele drop-out and drop-in when testing low amounts of single-source and mixed DNA samples (Mitchell et al. 2011, Mitchell et al. 2012). OCME began using FST in forensic casework in April 2011 (Mitchell et al. 2011).

While FST was being developed and implemented in New York City, another discrete PGS system named Lab Retriever was created in California. Instead of a proprietary, in-house program like FST, Lab Retriever is an open-source, freely available program to calculate likelihood ratios for complex DNA profiles (Inman et al. 2015). This program is based on David Balding's likeLTD discrete PGS system (Balding 2013, Lohmueller & Rudin 2013).

STRmix has been used in the U.S. since 2014. According to information on the website of one of the STRmix developers[6], early U.S. adopters of STRmix included the United States Army Criminal Investigation Laboratory (USACIL) in November 2014, Erie County (NY) in July 2015, San Diego Police Department in October 2015, and the FBI Laboratory in December 2015. The FBI and the STRmix developers co-published a summary of the FBI internal validation studies of STRmix in a peer-reviewed journal (Moretti et al. 2017). Data from many early adopters of STRmix were also compiled and published, representing results for 2825 mixtures from 31 laboratories (Bright et al. 2018).

STRmix developers and colleagues have published their perspectives on the utility and validity of their PGS system. They conclude: "The efforts to bring [probabilistic genotyping] to fruition, including the initial theoretical development for human identification applications based on STR typing, span almost two decades, and thus its use today should not be misconstrued as some sudden novel technology" (Buckleton et al. 2019).

### 3.2.3. **FTCOE 2015 Landscape Study of PGS Systems**

Given the growing interest in PGS systems among U.S. forensic laboratories, the National Institute of Justice (NIJ) funded a study to examine them. In July 2015, the NIJ Forensic Technology Center of Excellence (FTCOE) published a 45-page "Landscape Study of DNA Mixture Interpretation Software" (FTCOE 2015). The FTCOE report explored the stated capabilities and limitations of 13 DNA mixture interpretation software tools available at the time: two with binary interpretation models (ArmedXpert and GeneMarker HID), six using discrete models (FST, GenoProof Mixture, Lab Retriever, LikeLTD, LiRa, and LRmix Studio), and five incorporating continuous models (DNAmixtures, DNA View Mixture Solution, LiRaHT, STRmix, and TrueAllele).

[6] https://johnbuckleton.wordpress.com/strmix/ (accessed July 2, 2024)

For each of these 13 software tools, the FTCOE assessment examined availability (commercial, proprietary, or open-source); the developer; statistical approaches utilized (RMP, CPI, LR); input data required (.fsa or .hid file, .csv or .txt file); maximum number of unknown contributors that could be evaluated; whether training resources (yes/no), technical support (none, basic, extensive), or testimony support (yes/no) were available; whether CODIS output was possible (yes/no); whether a database could be queried (yes/no); whether Markov chain Monte Carlo (MCMC) simulations were performed (yes/no); whether the software could account for possible relatedness (yes/no); and frequency of system updates.

Since the 2015 study, there have been a few updates and additions to the PGS marketplace. PGS systems known to exist as of July 2019 are listed in a 2020 review article (Butler & Willis 2020, see also Coble & Bright 2019). A few studies have been published involving direct comparisons of PGS systems, as discussed in Chapter 4 of the accompanying report (see NISTIR 8351, Table 4.2).

## 3.3. **Towards a Future Incorporating Artificial Intelligence into Interpretation**

Human review and interpretation of DNA mixture electropherograms (EPGs) to generate allele calls is time-consuming and subject to variation among analysts. Forensic DNA laboratory protocols commonly use an analytical threshold to sort detected peaks from background noise followed by an analyst “reading” the EPG and making decisions regarding whether each peak is an allele or an artifact (Butler 2014, pp. 47-86). Typically, a second DNA analyst performs a technical review by independently reading the DNA profiles, comparing results, and resolving any differences.

In addition, a key component of most DNA mixture interpretation approaches is the estimation of the number of contributors (NoC) needed to inform the propositions used to assign likelihood ratios with PGS. As both NoC estimation and data interpretation decisions can involve human subjectivity, options are being explored to incorporate artificial intelligence (AI) solutions. The availability of a large-scale dataset of known single-source and mixed-source STR profiles has aided machine learning studies (Alfonse et al. 2018).

### 3.3.1. **Number of Contributor (NoC) Estimation**

Inferring the NoC is one of the first steps in DNA mixture interpretation (Clayton et al. 1998, Gill et al. 2006). Whether interpretation is performed manually or with the aid of a PGS system, the NOC estimate is an important input to interpretation that constrains the number of possible genotype combinations that are considered. However, there can be difficulties with inferring the NOC that result in uncertainties in that estimate (see Butler 2014, pp. 169-170). Some confounding factors include allele overlap of potential contributors (Paoletti et al. 2005, Buckleton et al. 2007) or allele drop-out (Perez et al. 2011) with low-level DNA (Alfonse et al. 2017, Norsworthy et al. 2018). Simulations have shown that uncertainty in estimating the number of contributors can decrease with the use of additional STR loci (Coble et al. 2015; see also Butler 2014, p. 170; Dembinski et al. 2018) or using sequence information with additional STR allele resolution (Young et al. 2019).

Multiple approaches have been used to estimate NOC (Haned et al. 2011) with the maximum allele count (MAC) per locus as the most common method cited in a recent assessment of approaches used by 81 of 83 laboratories that manually assess NOC (Brinkac et al. 2023). Recognizing that there can be an impact on the assigned likelihood ratio if NOC is incorrect (Benschop et al. 2015), some PGS systems permit weighting results across multiple NOC possibilities (McGovern et al. 2020).

Probabilistic approaches for estimating NOC using machine learning AI approaches have been reported by researchers at Rutgers University (Swaminathan et al. 2015, Grgicak et al. 2020, Grgicak et al. 2021), Syracuse University (Marciano & Adelman 2017, Marciano & Adelman 2019), the Netherlands Forensic Institute (Benschop et al. 2019, Veldhuis et al. 2022), and the Institute of Environmental Science and Research (Kruijver et al. 2021). A series of tests have been described for validation of NOC systems (Valtl et al. 2021).

### 3.3.2. **"Lights-Out" Data Interpretation**

Researchers are working towards the ability to fully interpret a DNA mixture in a fully automated ("lights-out") mode where the EPG could potentially be processed entirely by computer program(s) to replace one or both human reviewers (Taylor & Powers 2016, Taylor et al. 2017, Marciano et al. 2018, Karkar et al. 2019, Taylor et al. 2019, Taylor & Abarno 2023). Using artificial neural networks as a form of AI opens up the possibilities of processing EPG data in a more objective fashion as well as removing the need for DNA analysts to read the profile and could potentially even eliminate the need for analytical thresholds (Adelman et al. 2019, Taylor 2022, Taylor & Buckleton 2023).

FaSTR DNA is a software program developed by the Institute of Environmental Science and Research in New Zealand to assign NOC and rapidly analyze DNA profiles using configurable rules. One validation study with the FaSTR DNA algorithm found a 99.7% accuracy level with detecting allele peaks in 485 reference profiles (Volgin et al. 2021) while another study observed 99.95% concordant peak designations from 3403 samples and over 232,000 peaks (Lin et al. 2021). While capabilities for AI-assisted DNA mixture interpretation are improving, forensic DNA laboratories need to decide based on their own internal validation studies how to best implement these capabilities into their laboratory workflow.

We conclude this section with a summary of some key DNA interpretation advances over the years (Table S1.2).

**Table S1.2.** Timeline summary for some key DNA interpretation advances.

| Year | Event |
|---|---|
| **1991** | Likelihood ratio approach introduced for DNA mixtures by the UK Forensic Science Service (FSS; Evett et al. 1991) and would be developed further initially for two person mixtures (Weir et al. 1997) |
| **1992** | NRC I report (NRC 1992, p. 59) describes a combined probability of inclusion (CPI) approach |
| **1996** | NRC II report (NRC 1996, p. 130) supports the LR approach; replicate PCR amplifications from aliquots of the same low-level DNA sample can obtain reliable results (Taberlet et al. 1996) |
| **1998** | FSS describes mixture deconvolution for two-person mixtures (Clayton et al. 1998) and these “Clayton rules” would later be cited by ISFG (Gill et al. 2006); use of peak areas (or heights) to interpret simple mixtures (Gill et al. 1998, Evett et al. 1998); FSS describes hierarchy of propositions (Cook et al. 1998b) and case assessment and interpretation approaches (Cook et al. 1998a); *Interpreting DNA Evidence* published (Evett & Weir 1998) |
| **2000** | US DNA Advisory Board (DAB) endorses CPI and LR (DAB 2000); FSS introduces low copy number (LCN) with <100 pg and probabilistic genotyping concept (Gill et al. 2000); SWGDAM interpretation guidelines (4 pages) (SWGDAM 2000) |
| **2005** | NIST conducts the MIX05 interlaboratory study (Butler et al. 2018) |
| **2006** | ISFG DNA Commission publishes nine recommendations on DNA mixture interpretation (Gill et al. 2006); German Stain Commission proposes three mixture categories (Schneider et al. 2006b, 2009); least-squares deconvolution was published (Wang et al. 2006) |
| **2009** | Allele drop-out models are developed (Balding & Buckleton 2009); FBI mixture group publishes some CPI mixture principles (Budowle et al. 2009) |
| **2010** | SWGDAM interpretation guidelines are expanded (28 pages) (SWGDAM 2010) |
| **2012** | ISFG DNA Commission on allele drop-out (Gill et al. 2012) |
| **2013** | NIST conducts the MIX13 interlaboratory study (Butler et al. 2018) |
| **2014** | SWGDAM enhanced detection methods (SWGDAM 2014) |
| **2015** | SWGDAM probabilistic genotyping software guidelines (SWGDAM 2015); guidance for CPI (Bieber et al. 2016) |
| **2016** | PCAST report encourages probabilistic genotyping (PCAST 2016) |
| **2017** | SWGDAM interpretation guidelines are updated (90 pages) (SWGDAM 2017) |
| **2018** | PROVEDIt data set was published (Alfonse et al. 2018); ISFG on propositions (Gill et al. 2018); SWGDAM LR verbal equivalents (SWGDAM 2018) |
| **2020** | ISFG DNA Commission on activity level propositions (Gill et al. 2020) |
| **2024** | NIST/NIJ DNA Human Factors Report released (EWG 2024) |

## 4. Assessing Measurement and Interpretation Variation Across Laboratories

Interlaboratory comparison studies, sometimes referred to as collaborative exercises or round-robin studies, provide a community-focused approach to assessing whether multiple laboratories can generate comparable measurements and interpretations when provided with the same samples or DNA profiles.

### 4.1. Initial NIST Studies

During the early 1990s, NIST and FBI Laboratory researchers collaborated to conduct interlaboratory comparisons of DNA profiling RFLP measurements (Mudd et al. 1994, Duewer et al. 1995, Stolorow et al. 1996, Duewer et al. 1997, Duewer et al. 1998). The initial studies involved around 20 forensic laboratories, most of which were participants in the FBI's Technical Working Group on DNA Analysis Methods (TWGDAM). These interlaboratory comparisons assessed measurement and interpretation capabilities needed to set up an eventual national DNA database in the United States. In addition, the reproducibility and measurement uncertainties around RFLP band sizes were examined across 51 proficiency tests from 1991 to 1997 (Duewer et al. 2000b), and a certified reference material was developed to enable reliable DNA size comparisons across laboratories (Duewer et al. 2000a). Graphical tools were also developed by NIST researchers to assist with assessing RFLP DNA profiling measurements (Duewer et al. 1999, Gary et al. 1999). RFLP methods were phased out of use in most forensic DNA laboratories by the early 2000s especially with the selection of core STR markers for the U.S. national DNA database (Budowle et al. 1998).

As STR markers and commercial kits became available in the mid-1990s, NIST again conducted interlaboratory studies. Measurement consistency was examined across 34 laboratories using a wide variety of analytical systems testing the STR markers CSF1PO, TPOX, and TH01 (Kline et al. 1997). This study demonstrated that to maintain reliable and consistent STR allele measurements across laboratories information would need to be reported as an allelic name (produced through comparison to an allelic ladder available to all laboratories) rather than raw sizing results (Kline et al. 1997). NIST also developed a certified reference material (SRM 2391)[7] to enable measurement calibration for commonly used STR markers and kits. The FBI Quality Assurance Standards require that newly DNA methods are checked against an appropriate and available certified reference material to enable traceability and measurement calibration of DNA test results (QAS 1998, standard 9.5; QAS 2020, standard 8.4).

Interlaboratory comparisons of DNA mixture interpretation began with two mixed stain studies conducted in 1997 and 1999 involving blood stains, semen stains, and extracted DNA examined by 45 laboratories (Duewer et al. 2001). Based on lessons learned from these studies, NIST conducted a third mixed stain study in 2001 that examined DNA quantitation accuracy (Kline et al. 2003) and signal intensity balance in STR kits across 74 participating laboratories (Duewer et al. 2004).

[7] For details on the current version of this reference material, see https://tsapps.nist.gov/srmext/certificates/2391d.pdf (accessed July 2, 2024).

In early 2004, NIST collected 287 independent data sets with DNA quantitation information from eight DNA samples tested with 19 different quantitation methodologies and 80 participating laboratories (Kline et al. 2005). Lessons learned from this study guided the development of a human DNA quantitation reference material (SRM 2372).[8]

## 4.2. **Mixture Interpretation Studies by NIST and Others**

There have been at least 16 interlaboratory studies involving DNA mixture interpretation with STR markers beyond the initial three NIST mixed stain studies. Some studies provided samples to explore both measurement and interpretation aspects of the process and other studies provided only DNA profile EPGs to assess interpretation variability across participants. While these studies enable some insights into method differences and performance variation of voluntary participants, they would not be appropriate for estimating some kind of global error rate for DNA mixture interpretation, especially as the field has been changing in terms of markers, methods, protocols, and technology over the past two decades that the studies span.

Study organizers have included NIST (Butler et al. 2018), Noblis (Brinkac et al. 2023, Hicklin et al. 2023, Hicklin et al. 2024), the Defense Forensic Science Center (Rogers et al. 2022), the Spanish-Portuguese Working Group of the International Society for Forensic Genetics (Crespillo et al. 2014, Toscanini et al. 2016, Barrio et al. 2018), the European Forensic Genetics Network of Excellence (Prieto et al. 2014), the Netherlands Forensic Institute (Benschop et al. 2017), developers of the PGS system STRmix (Cooper et al. 2015, Bright et al. 2019a), the UK Forensic Science Regulator (Barber et al. 2015), and the UK Association of Forensic Science Providers (Thomson 2018, Mallinder et al. 2022). For additional information, see Section 2.5 in the other supplemental document prepared with this foundation study (NISTIR 8351sup2).

In the 2005 NIST MIX05 study, 69 laboratories interpreted data in the form of electropherograms of two-person DNA mixtures representing four different mock sexual assault cases with different contributor ratios. In the 2013 NIST MIX13 study, 108 laboratories interpreted electropherogram data for five different case scenarios involving two, three, or four contributors, with some of the contributors potentially related (Butler et al. 2018). Limitations in the use of CPI for complex mixtures were highlighted in several of the MIX13 cases.

Findings from the MIX05 study influenced development of the Scientific Working Group on DNA Analysis Methods (SWGDAM) "SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Laboratories" released in 2010 (SWGDAM 2010) and updated in 2017 (SWGDAM 2017). Findings from the MIX13 study were initially shared at a DNA Technical Leader's Summit in November 2013 (Butler 2014, p. 153) and influenced the U.S. forensic community to move towards probabilistic genotyping approaches for complex DNA mixtures (see Bright et al. 2017).

The authors of the MIX13 study noted:

> "Despite improvements in protocols and interpretation guidelines across the United States and Canada since the SWGDAM interpretation guidelines were released in 2010, results of mixture

[8] For details on the current version of this reference material, see https://tsapps.nist.gov/srmext/certificates/2372a.pdf (accessed July 2, 2024).

> interpretation were still highly variable several years later when the MIX13 study was conducted. Some of this variation was *a consequence of inappropriately using CPI to interpret complex mixtures*. As demonstrated in MIX13 Case 5, there is a risk of including a non-contributor when blindly applying CPI without interpretation of the DNA mixture itself. We recognize that many laboratories are implementing probabilistic genotyping software systems to assist in the deconvolution and statistical evaluation of complex mixtures. Future interlaboratory studies will be helpful in assessing how effective these software approaches are at improving performance across the community. These results, as with previous collaborative exercises, can be tools to draw attention to issues that can lead to improvements in the field." (Butler et al. 2018, emphasis in the original)

Findings from MIX05 and MIX13 (Butler et al. 2018) and the more recent DNAmix 2021 study (Hicklin et al. 2023, Hicklin et al. 2024) have brought awareness of differences in approaches to DNA mixture interpretation and have highlighted the need for improved training (to aid analyst understanding of methods used), more detailed and standardized protocols (to benefit consistent results within and across laboratories), and perhaps more extensive validation studies (to improve assessment of method capabilities and limitations). Hopefully lessons learned from these types of interlaboratory comparisons can lead to improved protocols over time.

### 4.3. **Proficiency Tests for DNA Mixture Interpretation**

Over the years, a variety of DNA mixture proficiency tests (PTs) have been provided to participating forensic DNA analysts. In the United States, PT providers offering DNA mixture tests include Collaborative Testing Services (Sterling, VA), Bode Technology (Lorton, VA), and Forensic Assurance (Northville, MI). In addition, the German DNA Profiling Group (GEDNAP) provides DNA proficiency tests for many laboratories in Europe (see NISTIR 8351, Section 3.1.3).

Summaries of CTS PT results are publicly available[9] on their website and information from these PTs have been consolidated and discussed in an accompanying supplemental document (NISTIR 8351sup2, Section 2.4). While test data from biological samples has been publicly available for more than 10 years, DNA interpretation test results, where participants start with an electropherogram, only became available in 2017 and probabilistic genotyping test summaries, where participants may provide an LR value, since 2022. CTS data from biological samples, in their forensic biology, DNA semen, and DNA mixture PTs, typically come in the form of single-source or two-person mixtures from mixing blood and/or semen samples in roughly equal amounts. Data from a few three-person and even one four-person PT are available in the more recent DNA interpretation and probabilistic genotyping test results (see NISTIR 8351sup2, Section 2.4). Thus, it is only in the past few years that more complex mixtures have been provided to participants to explore their performance.

As of 2024, Forensic Assurance provides[10] forensic biology with probabilistic genotyping software and forensic DNA with probabilistic genotyping software proficiency tests four times per year. However, summaries of these PT results are not publicly available. These PTs, which

[9] https://cts-forensics.com/program-1.php (accessed July 2, 2024)
[10] See https://forensicassurance.com/products/ (accessed July 2, 2024)

have been available for at least five years, can be either standard test design (where each participant receives the same test samples as others in their laboratory) or declared double blind test design (where each participant will receive a test composed of randomly selected unknown and/or known proficiency testing items).

# 5. Sources of Guidance on DNA Mixture Interpretation and Validation

## 5.1. Overview

Accredited laboratories follow written protocols and are regularly audited to assess their conformance to these protocols and compliance with applicable standards. Multiple advisory groups have provided recommendations on quality assurance measures and helpful validation studies to assess the capabilities and limitations of DNA mixture interpretation approaches (Butler 2013).

Numerous documents exist that provide guidance on DNA analysis in general and in some cases, mixture interpretation. A growing number are becoming available from various organizations around the world (see Table S1.3). A 2019 review noted that 34 guidance documents related to forensic DNA analysis and interpretation were published in the previous three years (Butler & Willis 2020). A follow-on review found 70 guidance documents related to forensic DNA released from 2019 to 2022 (Butler 2023). While many of these documents are designed to be specific for certain regions, there is value in knowing what others are doing and learning from them, as science knows no boundaries. Understanding the authority under which various documents are created, who is involved in creating them, and who uses or enforces the requirements or recommendations can be helpful.

**Table S1.3.** Documents that govern and influence DNA operations in accredited forensic laboratories. The order of the information does not imply preference. Abbreviations: AAFS = American Academy of Forensic Sciences, ANSI = American National Standards Institute, ANAB = ANSI National Accreditation Board, ASB = AAFS Standards Board, ASCLD/LAB = American Society of Crime Laboratory Directors/Laboratory Accreditation Board, ASTM = American Society for Testing and Materials, DAB = DNA Advisory Board, ENFSI = European Network of Forensic Science Institutes, FBI = Federal Bureau of Investigation, IEC = International Electrotechnical Commission, ILAC = International Laboratory Accreditation Cooperation, ISFG = International Society for Forensic Genetics, ISO = International Organization for Standardization, NDIS = National DNA Index System, OSAC = Organization of Scientific Area Committees for Forensic Science, QAS = Quality Assurance Standards, SDO = standards developing organization, SWGDAM = Scientific Working Group on DNA Analysis Methods, UK = United Kingdom, WG = Working Group.

| Document | Authority | Who Creates | Who Uses or Enforces |
|---|---|---|---|
| **FBI QAS** (1998/1999 updated in 2009, 2011, 2020) | Law passed by Congress in 1994; issued by FBI Director | Originally DAB (1995-2000), now SWGDAM | FBI and ANAB auditors to assess U.S. forensic laboratories |
| **ILAC G19** (2014) and **ISO/IEC 17025** (2017) | Standards community | ISO committee | Accrediting bodies (ANAB and formerly ASCLD/LAB) |
| **Guidelines & Best Practices** | Forensic practitioner community | SWGDAM, ENFSI DNA WG, ISFG DNA Commission | Forensic laboratories and practitioners (not required) |
| **UK Forensic Science Code of Practice** | UK Forensic Science Regulator | UK Forensic Science Regulator working group | UK forensic laboratories and practitioners |
| **ASB/ASTM Standards** (and OSAC Registry) | SDOs with forensic practitioner community input | SDOs (ASB, ASTM) and OSAC | Accrediting bodies as they are adopted |

Groups that have commented on or proposed recommendations for DNA mixture interpretation include the ISFG DNA Commission (Gill et al. 2006, Gill et al. 2012, Coble et al. 2016, Gill et al. 2018, Gill et al. 2020), the German Stain Commission (Schneider et al. 2006b, Schneider et al. 2009), the European Network of Forensic Science Institutes DNA Working Group (Morling et al. 2007, ENFSI 2017, ENFSI 2022), the Technical UK DNA Working Group on Mixture Interpretation (Gill et al. 2008), the Biology Specialist Advisory Group (BSAG) of the Australian and New Zealand forensic science community (Stringer et al. 2009), an FBI mixture committee (Budowle et al. 2009), the UK Forensic Science Regulator (UKFSR 2018a, UKFSR 2018b), AAFS Standards Board (ANSI/ASB 020, ANSI/ASB 040, Press 2020, ANSI/ASB 018), and SWGDAM (SWGDAM 2010, SWGDAM 2015, SWGDAM 2017). These efforts are briefly described below.

## 5.2. **ISFG DNA Commission and European Efforts in Mixture Interpretation**

The International Society for Forensic Genetics (ISFG) has a DNA Commission that periodically addresses important topics in the field and makes recommendations. DNA mixture interpretation has been a part of five ISFG DNA Commissions (Gill et al. 2006, Gill et al. 2012, Coble et al. 2016, Gill et al. 2018, Gill et al. 2020).

In July 2006, the DNA Commission of the International Society for Forensic Genetics (ISFG) published nine recommendations (Box S1.4) covering multiple mixture interpretation principles (Gill et al. 2006). In one of these recommendations, the ISFG DNA Commission endorsed the mixture deconvolution steps published in 1998 by the Forensic Science Service (Clayton et al. 1998). Since several forensic statisticians were part of this Commission, these recommendations favor approaches involving likelihood ratios that had previously been published (Evett et al. 1991, Weir et al. 1997).

An editorial accompanied the 2006 ISFG DNA Commission recommendations (Schneider et al. 2006a). The authors describe the purposes behind these initial DNA mixture interpretation recommendations:

> "…DNA evidence alone could be decisive for obtaining a conviction of an accused suspect. Thus, the interpretation of the observed DNA profile of a given stain in the context of the case needs to include a reasonable biostatistical evaluation of the weight of the evidence. At the same time, *the molecular biological tools available to the forensic geneticist have become more and more sensitive to the point where the genomic DNA from a few dozen cells may be sufficient to obtain a full STR profile from an unknown offender*. *As a result, the number of DNA mixtures composed from full or partial profiles from two or more contributors* (who could be offenders, victims, or individuals not associated with the crime event) *has increased significantly*. The biostatistical interpretation of such mixed DNA profiles is a very challenging task that sometimes leads to controversial views about correct mathematical approaches for estimating the weight of the evidence. Indeed, *diverse practices have already arisen between laboratories*, hence there is an urgent need to formulate recommendations… These recommendations have been written to serve two purposes: to define a generally acceptable mathematical approach for typical mixture scenarios and to address open

questions where practical and generally accepted solutions do not yet exist…This paper is a 'high level' treatise on the mathematical principles to analyse complex mixtures. We realise that it will not be possible for most laboratories to immediately implement the methods described. *Our intention is primarily to specify a consensus approach to act as the foundation stone*. *Hopefully we will encourage the development of expert systems to take care of the onerous calculations*." (Schneider et al. 2006a, emphasis added).

**Box S1.4. ISFG 2006 Recommendations on DNA Mixture Interpretation**

**Recommendation 1**: The likelihood ratio is the preferred approach to mixture interpretation. The RMNE [Random Man Not Excluded; also known as the Combined Probability of Inclusion, CPI] approach is restricted to DNA profiles where the profiles are unambiguous. If the DNA crime stain profile is low level and some minor alleles are the same size as stutters of major alleles, and/or if drop-out is possible, then the RMNE method may not be conservative.

**Recommendation 2**: Even if the legal system does not implicitly appear to support the use of the likelihood ratio, it is recommended that the scientist is trained in the methodology and routinely uses it in case notes, advising the court in the preferred method before reporting the evidence in line with the court requirements. The scientific community has a responsibility to support improvement of standards of scientific reasoning in the court-room.

**Recommendation 3**: The methods to calculate likelihood ratios of mixtures (not considering peak area) described by Evett et al. (Evett et al. 1991) and Weir et al. (Weir et al. 1997) are recommended.

**Recommendation 4**: If peak height or area information is used to eliminate various genotypes from the unrestricted combinatorial method, this can be carried out by following a sequence of guidelines based on Clayton et al. (Clayton et al. 1998).

**Recommendation 5**: The probability of the evidence under $H_p$ is the province of the prosecution and the probability of the evidence under $H_d$ is the province of the defense. The prosecution and defense both seek to maximize their respective probabilities of the evidence profile. To do this both $H_p$ and $H_d$ require propositions. There is no reason why multiple pairs of propositions may not be evaluated.

**Recommendation 6**: If the crime profile is a major/minor mixture, where minor alleles are the same size (height or area) as stutters of major alleles, then stutters and minor alleles are indistinguishable. Under these circumstances alleles in stutter positions that do not support $H_p$ should be included in the assessment.

**Recommendation 7**: If drop-out of an allele is required to explain the evidence under $H_p$: (S = ab; E = a), then the allele should be small enough (height/area) to justify this. Conversely, if a full crime stain profile is obtained where alleles are well above the background level, and the probability of drop-out approaches $Pr(D) \approx 0$, then $H_p$ is not supported.

**Recommendation 8**: If the alleles of certain loci in the DNA profile are at a level that is dominated by background noise, then a biostatistical interpretation for these alleles should not be attempted.

**Recommendation 9**: In relation to low copy number, stochastic effects limit the usefulness of heterozygous balance and mixture proportion estimates. In addition, allelic drop-out and allelic drop-in (contamination) should be taken into consideration of any assessment.

Following the 2006 ISFG DNA Commission publication, a Technical UK DNA Working Group was formed to provide a detailed response that considered their national needs and court experiences with DNA mixture interpretation (Gill et al. 2008). An FBI Laboratory working group (Budowle et al. 2009) and SWGDAM (SWGDAM 2010) also built upon the 2006 ISFG DNA Commission foundational principles.

The December 2007 issue of *Forensic Science International: Genetics* contained a letter to the editor entitled "Interpretation of DNA mixtures – European consensus on principles" that was co-authored by chairs of the European DNA Profiling Group (EDNAP), the DNA Working Group of the European Network of Forensic Science Institutes (ENFSI), the German Stain Commission, and the Technical UK DNA Working Group (Morling et al. 2007). These groups expressed their support for the 2006 ISFG recommendations on mixture interpretation (Gill et al. 2006). This letter to the editor emphasized "laboratories must invest in continuous education of the staff in the interpretation of DNA mixtures" (Morling et al. 2007).

The ISFG 2006 recommendations and principles were also supported by an Australian and New Zealand Biology Specialist Advisory Group (BSAG) (Stringer et al. 2009). The BSAG provided some additional commentary:

> "The likelihood ratio is a common approach to mixture interpretation in Australia and New Zealand. RMNE [random man not excluded] is considered an acceptable alternative approach to DNA interpretation. If the crime stain DNA profile is low level and some minor alleles are the same size as stutters of major alleles, and/or if drop-out is possible, then extra consideration needs to be given to the method of statistical interpretation… It is recommended that the scientist is trained in the primary methodology routinely used in their laboratory and has an understanding of other statistical approaches for DNA interpretation. The scientific community has a responsibility to support improvement of standards of scientific reasoning in the Justice system" (Stringer et al. 2009).

The German Stain Commission, a group of scientists from Germany's Institutes of Legal Medicine, introduced a three-part classification scheme for DNA mixtures: Type A (no major contributor), Type B (major and minor contributors distinguishable), and Type C (low-level DNA with stochastic effects). Their recommendations were first provided in German (Schneider et al. 2006b) and then republished in English (Schneider et al. 2009) to increase accessibility.

Under this classification scheme, Type A mixtures require a biostatistical analysis that can be performed with an LR or RMNE (CPI). Type B mixtures can be deconvoluted into the major and minor components, usually if they are present with consistent peak-to-height ratios of approximately 4:1. The major component following deconvolution can be treated as a single-source profile and a random match probability calculated. For Type C mixtures, where all alleles may not be seen due to allele dropout, a biostatistical interpretation is not appropriate, and a clear decision about whether to include or exclude a suspect may be difficult to reach. It is important to keep in mind that these German Stain Commission categories were originally developed when two-person mixtures were most commonly seen in forensic laboratories (see Butler 2014, p. 133) – and were not intended to address the complex mixtures processed today with PGS systems.

Later iterations of the ISFG DNA Commission provided recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods (Gill et al. 2012), the validation of software programs performing biostatistical calculations for forensic genetics applications (Coble et al. 2016), guidelines on formulating propositions for investigative and court-going purposes (Gill et al. 2018), and advice on evaluating low-level DNA results considering activity level propositions (Gill et al. 2020).

In 2017, the ENFSI DNA Working Group[11], which has members from more than 50 organizations across 35 European countries, published a best-practice manual, which outlined experiments for performing internal validation of probabilistic genotyping software used in DNA mixture interpretation (ENFSI 2017). This guidance builds upon the ISFG DNA Commission recommendations (Coble et al. 2016).

In 2018, the UK Forensic Science Regulator offered guidance on DNA mixture interpretation (UKFSR 2018a) and software validation for DNA mixture interpretation (UKFSR 2018b). For example, the software validation document promotes use of a validation library with supporting information covering software specifications, risk assessments, technical reports or scientific publications, a validation plan including the user acceptance criteria, information on the statistical models used, a statistical specifications report including underlying data on which any conclusions are based, the validation report with data summaries and assessment against the acceptance criteria, and a record of validation approval (UKFSR 2018b). A 2018 annual report[12] from the Regulator states:

> "There will always be limits to the complexity of DNA mixtures that can safely be interpreted, but the guidance published in FSR-G-222 [(UKFSR 2018a)] and FSR-G-223 [(UKFSR 2018b)] should ensure that interpretation does not stray beyond what is scientifically robust" (March 15, 2019, p. 47).

## 5.3. SWGDAM and U.S. Efforts in Mixture Interpretation

In the United States, the FBI Laboratory has sponsored the Technical Working Group on DNA Analysis Methods (TWGDAM) from 1988 to 1998 and the Scientific Working Group on DNA Analysis Methods (SWGDAM)[13] from 1998 to the present. An important purpose of TWGDAM and SWGDAM continues to be a semi-annual gathering of forensic DNA scientists to share protocols and ideas and to write guidelines where appropriate. From 1995 to 2000, the FBI also had a Federal Advisory Committee known as the DNA Advisory Board (DAB) that crafted the original Quality Assurance Standards (QAS), which were first issued in October 1998. Since 2000, when the DAB's charter expired, revisions to the QAS have been performed by SWGDAM.

Historically, DNA mixture interpretation has been minimally addressed in the QAS, with more detailed information included in SWGDAM guidance documents (see Table S1.4). For example, the 2011 version of the QAS contains one requirement regarding mixture interpretation in Standard 9.6.4:

---

[11] See http://enfsi.eu/about-enfsi/structure/working-groups/dna/

[12] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/786137/FSRAnnual_Report_2018_v1.0.pdf (p. 47)

[13] See https://www.swgdam.org/about-us

> "Laboratories analyzing forensic samples shall have and follow a documented procedure for mixture interpretation that addresses major and minor contributors, inclusions and exclusions, and policies for the reporting of results and statistics" (QAS 2011).

Contemporaneous SWGDAM guidance documents then provided more detailed suggestions (SWGDAM 2010, SWGDAM 2012).

In February 2000, the FBI's DNA Advisory Board endorsed the use of CPI and LR methods for providing statistical support of an inclusion following mixture interpretation (DAB 2000). In their first publication regarding implementation of STRs in forensic casework, the FBI Laboratory discussed the importance of a stochastic threshold when performing mixture interpretation and using the CPI statistic (Moretti et al. 2001a, Moretti et al. 2001b). An FBI Mixture Committee provided further guidance on using stochastic thresholds with CPI a few years later (Budowle et al. 2009).

An interlaboratory study conducted by NIST in 2005, designated MIX05, demonstrated variation across the community in approaches being taken at the time with two-person mixtures (Butler et al. 2018). Recognizing a need to address variability observed in approaches being taken with mixture interpretation, SWGDAM started a Mixture Committee in January 2007. The committee discussed topics surrounding mixture interpretation and drafted what was eventually published three years later as a 28-page document "SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories" (SWGDAM 2010). The SWGDAM 2010 guidelines built upon many of the 2006 ISFG DNA Commission recommendations (Gill et al. 2006), particularly in relationship to interpretation of peaks in the stutter position (see Butler 2014, pp. 148-149).

Updates were made to the 2010 guidelines by the SWGDAM Autosomal STR Committee, and a 90-page document was released in 2017, providing a variety of examples of handling binary methods of DNA mixture interpretation (SWGDAM 2017). Further revisions of the SWGDAM interpretation guidelines are under development to assist with guidance on probabilistic genotyping approaches. It is helpful to keep in mind that guidelines and standards take time to develop and are not always available when technology or interpretation approaches are initially implemented. Other documents from SWGDAM related to DNA mixture interpretation include verbal equivalents for likelihood ratios (SWGDAM 2018) and validation guidelines (see next section).

In September 2018, the U.S. Department of Justice issued a Uniform Language for Testimony and Reports (ULTR)[14] for forensic autosomal DNA examinations using probabilistic genotyping systems. This ULTR supports the LR verbal scale defined earlier by SWGDAM with qualitative equivalent categories of *uninformative* (LR=1), *limited support* (LR = 2 to <100), *moderate support* (LR = 100 to <10,000), *strong support* (LR = 10,000 to < 1 million), and *very strong support* (LR > 1 million) (SWGDAM 2018).

The Organization of Scientific Area Committees for Forensic Science (OSAC)[15] was created in 2014 as a joint venture between NIST and the Department of Justice (Butler 2015b). OSAC's

[14] https://www.justice.gov/olp/uniform-language-testimony-and-reports (accessed July 1, 2024)
[15] https://www.nist.gov/organization-scientific-area-committees-forensic-science (accessed July 1, 2024)

goal is to facilitate the development of technically sound, science-based standards through a formal standards developing organization (SDO) process and placement of approved standards and guidelines on an OSAC Registry. In May 2020, the first two DNA standards were placed on the OSAC Registry (Press 2020). Both standards relate to DNA mixture interpretation: "Standard for Validation Studies of DNA Mixtures, and Development and Verification of a Laboratory's Mixture Interpretation Protocol" (ANSI/ASB 020) and "Standard for Forensic DNA Interpretation and Comparison Protocols" (ANSI/ASB 040). These two documents were originally drafted by OSAC in 2015 and 2016 and then further developed and published by the AAFS Standards Board in 2018 and 2019 before being reviewed by OSAC for placement on the registry in 2020.

These new standards, which are meant to complement the FBI QAS and build upon SWGDAM guidelines, require laboratories to demonstrate that their protocols produce consistent and reliable conclusions with DNA samples different from the ones used in the initial validation studies. These standards also require that laboratories do not attempt to interpret DNA mixtures beyond the scope that they have validated and verified. For example, if a lab has tested its protocol for up to three-person DNA mixtures, it should not interpret casework that contains DNA from four or more people (Press 2020).

Additional standards to assist in DNA mixture interpretation in the future are in the OSAC pipeline and being finalized through the AAFS Standards Board DNA Consensus Body[16] with the SDO process.

### 5.3.1. U.S. Validation Guidance Regarding DNA Mixture Interpretation

Validation studies assist in understanding the degree of reliability of scientific methods. This section briefly reviews FBI QAS validation requirements and SWGDAM guidance related to DNA mixture interpretation. For the forensic DNA community, levels of validation have been divided into developmental validation, often performed under the auspices of the vendor, and internal validation, performed within each user laboratory or laboratory system. The purpose of these studies is to explore the capabilities and limitations of the methods being used in the laboratory.

Often publications in the forensic DNA literature state, when describing the developmental validation of, for example, a new DNA test kit or methodology, that "SWGDAM validation guidelines were followed." In making such statements, authors of these publications may be trying to convey that because suggested mixture studies were performed, the method should be accepted as robust, reliable, and reproducible.

An observation made in conducting this scientific foundation review is that, historically, FBI QAS validation requirements and SWGDAM validation guidelines have become *task-driven* rather than *performance-based*. In other words, the requirements and guidelines may be treated by some as a checklist of studies that need to be completed to satisfy requirements rather than a demonstrated performance of the accuracy or reliability of results obtained using the method.

[16] https://www.aafs.org/academy-standards-board/consensus-bodies (accessed July 1, 2024)

Laboratory decision makers and users of their validation data must understand statements such as SWGDAM validation guidelines to assess whether they were followed. These guidelines relate to DNA mixture interpretation and have changed over the years, and the general nature of current validation requirements or guidelines can create variability in the ways these studies are conducted.

Over the past several decades, SWGDAM has regularly updated its validation guidelines as well as validation requirements in the FBI Quality Assurance Standards (QAS) (Table S1.4). Validation guidelines were initially issued for RFLP techniques in 1989 (TWGDAM 1989) and for PCR techniques beginning in 1991 (TWGDAM 1991). PCR-based validation guidelines have been refined and updated in 1995, 2004, 2012, and 2016. In addition, validation guidelines for probabilistic genotyping software (PGS) systems were issued by SWGDAM in 2015 (SWGDAM 2015). Validation requirements contained in the FBI QAS were published in 1998, 1999, 2009, 2011, and 2020. Content related to DNA mixture interpretation in each of these documents is summarized in Table S1.4 with the exception of the SWGDAM PGS validation guidelines, which are covered separately (see Section 5.4.2).

**Table S1.4.** A chronological review of validation guidelines or requirements prepared by SWGDAM or its predecessors that relate to DNA mixture interpretation.

| Year | Document | **Information related to DNA mixture interpretation**<br>(bold font used to add emphasis on language related to mixtures) |
|---|---|---|
| **1989** | TWGDAM Quality Assurance | *(no mention of mixtures)* |
| **1991** | TWGDAM Quality Assurance | 4.1.5.5 Mixed Specimen Studies - **Investigate the ability of the system to detect the components of mixed specimens and define the limitations of the system.**<br><br>4.4.1.6 Where more than one locus is amplified in one sample mixture, **the effects of such amplification on each system (alleles) must be addressed and documented**. |
| **1995** | TWGDAM Quality Assurance | *(mixture information in TWGDAM 1995 is the same as TWGDAM 1991)* |
| **1998 and 1999** | DAB QAS Forensic (QAS 1998) and Database (QAS 1999) | 8.1.2 Novel forensic methodologies shall undergo developmental validation to ensure the accuracy, precision and reproducibility of the procedure. The developmental validation shall include the following: …<br><br>8.1.2.2 Species specificity, sensitivity, stability and **mixture studies are conducted**.<br><br>*(no mention of mixtures under 8.1.3 internal validation requirements)*<br><br>9.1.3 The laboratory shall have **a procedure for differential extraction** of stains that potentially contain semen.<br><br>9.6 The laboratory shall **have and follow written general guidelines for the interpretation of data.** *(no mention of mixtures)* |

| Year | Document | Information related to DNA mixture interpretation<br>(bold font used to add emphasis on language related to mixtures) |
|---|---|---|
| **2004** | SWGDAM Validation Guidelines | 2.8 Mixture studies: **The ability to obtain reliable results from mixed source samples should be determined.**<br><br>3.5 Mixture studies: When appropriate, forensic casework laboratories must **define and mimic the range of detectable mixture ratios**, including detection of major and minor components. Studies should be conducted **using samples that mimic those typically encountered in casework** (e.g., postcoital vaginal swabs). |
| **2009** | FBI QAS | 8.2.1 Developmental validation studies shall include, where applicable, characterization of the genetic marker, species specificity, sensitivity studies, stability studies, reproducibility, case-type samples, population studies, **mixture studies**, precision and accuracy studies, and PCR-based studies. PCR-based studies include reaction conditions, assessment of differential and preferential amplification, effects of multiplexing, assessment of appropriate controls, and product detection studies. All validation studies **shall be documented**.<br><br>8.3.1 Internal validation studies conducted after the date of this revision shall include as applicable: known and non-probative evidence samples or mock evidence samples, reproducibility and precision, sensitivity and stochastic studies, **mixture studies**, and contamination assessment. Internal validation studies **shall be documented and summarized**. The technical leader shall approve the internal validation studies.<br><br>8.3.2 Internal validation shall define quality assurance parameters and interpretation guidelines, including as applicable, **guidelines for mixture interpretation**.<br><br>9.6.4 Laboratories analyzing forensic samples **shall have and follow a documented procedure for mixture interpretation** that addresses major and minor contributors, inclusions and exclusions, and policies for the reporting of results and statistics. |
| **2011** | FBI QAS | *(mixture information is the same as QAS 2009)* |

| Year | Document | **Information related to DNA mixture interpretation**<br>(bold font used to add emphasis on language related to mixtures) |
|---|---|---|
| **2012** | SWGDAM Validation Guidelines | 2.2.2.2 Quality assurance parameters and interpretation guidelines shall be derived from internal validation studies. For example, lower template DNA may cause extreme heterozygote imbalance; as such, **empirical heterozygote peak-height ratio data could be used to formulate mixture interpretation guidelines and determine the appropriate ratio by which two peaks are determined to be heterozygotes**. In addition to establishing an analytical threshold, results from sensitivity studies could be used to determine the extent and parameters of quality control tests that reagents require prior to their being used in actual casework.<br><br>3.8 Mixture studies: The ability to obtain reliable results from mixed-source samples should be determined. These **studies will assist the laboratory to establish guidelines for mixture interpretation**, which may include determination of the number of contributors to the mixture, determination of the major and minor contributor profiles, and contributor ratios or proportions.<br><br>4.4 Mixture studies: **Mixed DNA samples that are representative of those typically encountered by the testing laboratory should be evaluated**. These studies will assist a casework laboratory to establish guidelines for mixture interpretation, which may include determination of the number of contributors to the mixture, determination of the major and minor contributor profiles, and contributor ratios or proportions. **A simplified mixture study may also assist a databasing laboratory to recognize mixtures and/or contamination**.<br><br>Table 1 ***Mixture studies will be required if the assay is intended to distinguish different contributors (male/female, major/minor, etc.)**. |
| **2016** | SWGDAM Validation Guidelines | *(mixture information is the same as SWGDAM 2012)* |

| Year | Document | **Information related to DNA mixture interpretation** (bold font used to add emphasis on language related to mixtures) |
| --- | --- | --- |
| **2020** | FBI QAS | 2. Definitions: Interpretation Software is a tool to assist the analyst in assessing the analyzed data by applying quality assurance rules, **performing mixture deconvolution**, and/or evaluating comparisons. Interpretation software may include probabilistic genotyping software or expert systems.<br><br>2. Definitions: Sensitivity studies (for the purposes of Standard 8.8) are used to assess the ability of the system to reliably determine the presence of a contributor's DNA over a broad variety of evidentiary typing results (to include **mixtures** and low-level DNA quantities).<br><br>2. Definitions: Specificity studies (for the purposes of Standard 8.8) are used to evaluate the ability of the system to provide reliable results over a broad variety of evidentiary typing results (to include **mixtures** and low-level DNA quantities).<br><br>8.2.1 Developmental validation studies shall include, where applicable, characterization of the genetic marker, species specificity, sensitivity studies, stability studies, case-type samples, population studies, **mixture studies**, precision and accuracy studies, and PCR-based studies. PCR-based studies include reaction conditions, assessment of differential and preferential amplification, effects of multiplexing, assessment of appropriate controls, and product detection studies. All validation studies shall be documented.<br><br>8.3.1 Internal validation studies shall include as applicable: known and nonprobative evidence samples or mock evidence samples, precision and accuracy studies, sensitivity and stochastic studies, **mixture studies**, and contamination assessment studies.<br><br>8.3.2 Internal validation shall define quality assurance parameters and interpretation guidelines, including, as applicable, **guidelines for mixture interpretation** and the application of appropriate statistical calculations.<br><br>8.3.2.1 **Mixture interpretation validation studies shall include samples with a range of the number of contributors, template amounts, and mixture ratios expected to be interpreted in casework.**<br><br>9.6.6 **Have and follow procedures for mixture interpretation** that address the following: 9.6.6.1 The assessment of the number of contributors. 9.6.6.2 The separation of contributors (e.g., major versus minor). 9.6.6.3 The criteria for deducing potential contributors.<br><br>9.10.5 The approaches to performing statistical calculations. 9.10.5.1 For autosomal STR typing, the procedure shall address homozygous and heterozygous typing results, multiple locus profiles, **mixtures**, minimum allele frequencies, and where appropriate, biological relationships. |

As can be seen in Table S1.4, the amount of information regarding mixture interpretation has increased over the years in newer versions of the SWGDAM validation guidelines and the FBI QAS requirements. A more detailed comparison of topics covered in the various versions for developmental and internal validation requirements is available elsewhere (see Tables 7.2 and 7.3 on pages 179-181 in Butler 2011).

Historically, limited information was provided regarding the suggested and/or required studies to inform mixture interpretation protocols. Rather, the early emphasis was to:

> "investigate the ability of the system [DNA testing method] to detect components of mixed specimens and define the limitations of the system" (TWGDAM 1991, section 4.1.5.5)

or to determine

> "the ability to obtain reliable results from mixed source samples" (SWGDAM 2004, section 2.8)

and to

> "define and mimic the range of detectable mixture ratios" in validation experiments (SWGDAM 2004, section 3.5).

The 2012 SWGDAM validation guidelines first emphasized performing validation studies that reflect the complexity of samples being examined in casework:

> "Mixed DNA samples that are representative of those typically encountered by the testing laboratory should be evaluated" (SWGDAM 2012, Guideline 4.4).

The 2012 guidelines do not specifically address the need to define the limitations of the system; rather, they suggest studies to help establish laboratory guidelines for mixture interpretation (SWGDAM 2012, Guideline 3.8). This text was maintained in the 2016 version of the document (SWGDAM 2016).

The 2020 update to the FBI QAS now requires that

> "mixture interpretation validation studies shall include samples with a range of the number of contributors, template amounts, and mixture ratios expected to be interpreted in casework" (QAS 2020, Standard 8.3.2.1).

The 2009 version included a more open requirement, stating:

> "Laboratories analyzing forensic samples shall have and follow a documented procedure for mixture interpretation" (QAS 2009, Standard 9.6.4).

Recommended studies include, for example, known and nonprobative evidence samples, sensitivity and stochastic studies, precision and accuracy assessments, mixture studies, and contamination assessments. Under mixture studies, the guidelines state:

> "mixed DNA samples that are representative of those typically encountered by the testing laboratory should be evaluated" (SWGDAM 2016, Section 4.4).

Ideally, developmental validation studies are conducted by vendors to meet specific performance measures, and internal validation experiments demonstrate similar performance under individual laboratory conditions.

Performance-based approaches are preferable over checklists of validation studies conducted because they can provide information on the limitations of the method. As noted in the previous section, a new documentary standard was published in 2018: "Standard for Validation Studies of DNA Mixtures, and Development and Verification of a Laboratory's Mixture Interpretation Protocol" (ANSI/ASB 020). This document discusses performance in more detail than previous guidance documents, but since it is new, feedback is not yet available regarding routine implementation by forensic DNA laboratories. For example, the standard requires:

> "The laboratory shall verify and document that the mixture interpretation protocols developed from the validation studies generate reliable and consistent interpretation and conclusions for the types of mixed DNA samples typically encountered by the laboratory" (ANSI/ASB 020, standard 4.4)

and explains further that

> "DNA mixture data from different sets of contributors than used in the initial validation studies shall be used to verify the protocol" (ANSI/ASB 020, p. 6).

Forensic laboratories are accredited to international standard ISO/IEC 17025:2017 "General Requirements for the Competence of Testing and Calibration Laboratories," which describes the types of information that can be used for method validation: (1) calibration or evaluation of bias and precision using reference materials, (2) systematic assessment of the factors influencing the result, (3) testing method robustness through variation of controlled parameters, (4) comparison of results achieved with other validated methods, (5) interlaboratory comparisons, and (6) evaluation of measurement uncertainty of the results based on the theoretical principles of the method and practical experience of the performance of the sampling or test method (ISO/IEC 17025:2017, Standard 7.2.2.1 note 2).

The ANAB accreditation requirements, under which most U.S. forensic laboratories are assessed, state:

> "Method validation shall: (a) be conducted according to a validation plan; (b) include the associated data analysis and *interpretation*; (c) establish the data and acceptance criteria required to report a result, opinion, *interpretation*, or statement of conformity; and (d) *identify limitations* of the method" (ANAB 2023, Section 7.2.2.1.1, emphasis added).

Historically, forensic DNA laboratories have conducted mixture studies during their internal validation experiments with emphasis on *robustness* (does the test produce a result?) and *detectability* (can minor alleles in a two-person mixture with multiple mixture ratios be detected?) rather than *reliability* (was interpretation of the mixture data accurate and consistent if repeated?). Publicly accessible performance-based validation data covering the desired factor space to achieve confidence in interpreting complex mixtures involving more than two contributors have been limited (see NISTIR 8351sup2).

## 5.4. **Requirements and Expectations for PGS Validation**

The ISFG DNA Commission from 2012 concluded:

> "The introduction of software solutions to interpret DNA profiles must be accompanied by a validation process ensuring conformity with existing standard laboratory procedures. … Software tools used for casework implementation must be evaluated with known samples and each laboratory will have to establish reporting guidelines and testimony training to properly present the results to courts" (Gill et al. 2012).

Several organizations and individual researchers have provided guidance on PGS validation. A brief history and overview of this guidance are provided here.

### 5.4.1. **Published Input from Software Developers**

In 2006, the TrueAllele PGS developer, Mark Perlin, described his thoughts on scientific validation of mixture interpretation methods in a *Proceedings of the International Symposium for Human Identification* submission with a focus on precision, accuracy, and reproducibility (Perlin 2006).

In 2014, the STRmix developers, John Buckleton, Jo-Anne Bright, Duncan Taylor, and two colleagues, Ian Evett and James Curran, provided their thoughts on some recommended tests when validating PGS systems (Bright et al. 2015). Four experiments were suggested: (1) comparison of the expected LR with no drop-out or drop-in, (2) the effect of drop-out, (3) the effect of drop-in, and (4) reproducibility. Some examples were run with single-source profiles and simple two-person mixtures using STRmix, LRmix, and Lab Retriever. They conclude:

> "*An understanding of the models within each of the program[s] and their limitations is required* in order to validate interpretation software" (Bright et al. 2015, emphasis added).

They continue:

> "Gaining an understanding of the behavior of the software under certain conditions is central to the developmental validation process prior to use in casework... [It is] an important part of the internal validation and training process prior to implementation of software. *This includes calibration based on ground-truth cases where the contributors are known* and case hardening to test how a program performs in the real world" (Bright et al. 2015, emphasis added).

Developers of the discrete PGS systems LRmix and Lab Retriever write that "model and software validation are inherently entangled" and provide an example of examining over 1,000 LR calculations for their LRmix validation (Haned et al. 2016). They describe four principle steps for software validation: (1) define the statistical specifications of the software (i.e., document the theory behind the model); (2) carry out analytical verification, which involves manually calculating LR values for simple cases and comparing results to the software output (while keeping in mind that as the model becomes more complex, analytical verification may not be possible); (3) compare results to data obtained from alternative software, which may rely on a

similar or a different probabilistic model; and (4) verify the code itself through visual inspection and recoding, which they note is most easily achievable through open-source software (Haned et al. 2016).

These authors also note:

> "The more complex the model, the greater the number of assumptions that are required. Increasing the number of variables incorporated into such a model also increases the chance of creating dependencies. Such models require a validation protocol that specifically addresses the additional interactions, and care must be taken to clearly define the variables. *We caution that complex models may at some point begin to produce unrealistic results, and hence become counter-productive*. More generally, the validation criteria should be explicit to the end users, and a determination made as to whether these criteria are fit for purpose" (Haned et al. 2016, emphasis added).

This last quote notes the challenge of defining what is "fit for purpose" with LR values assigned by PGS systems. As described in the next section, validation criteria in terms of studies that should be conducted have been outlined by SWGDAM (SWGDAM 2015). However, guidance on how to identify limitations with these PGS systems is sparse or non-existent. Knowing that "complex models may at some point begin to produce unrealistic results" (Haned et al. 2016), the identification of these points of caution or concern need to be defined by assessing performance from whether same-source and different-source sample profiles are appropriately "included" or "excluded" in situations with types of samples similar to the case in question.

### 5.4.2. **SWGDAM 2015 PGS Validation Guidelines**

The SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems were approved and posted on the SWGDAM website on June 15, 2015 (SWGDAM 2015). They begin:

> "These guidelines are not intended to be applied retroactively. It is anticipated that they will evolve with future developments in probabilistic genotyping systems."

Topics covered include documentation required, computer system control measures, developmental validation studies recommended to be performed by developers, internal validation studies to be performed by forensic laboratories, and performance checks with any software modifications (SWGDAM 2015). Suggested readings include three published references available at the time (Gill et al. 2012, Kelly et al. 2014, Steele & Balding 2014). The introduction states:

> "Prior to validating a probabilistic genotyping system, the laboratory should ensure that it possesses the appropriate foundational knowledge in the calculation and interpretation of likelihood ratios. Laboratories should also be aware of the features and limitations of various probabilistic genotyping programs and the impact that those items will have on the validation process."

The 2015 SWGDAM PGS validation guidelines state that the system shall be validated "prior to usage for forensic applications" (1.1), that "the laboratory shall document all validation studies

in accordance with the FBI Quality Assurance Standards" (1.2), and the laboratory should "have access to documentation that explains how the software performs its operations and activities" in order "to identify aspects of the system that should be evaluated through validation studies" (1.3). In addition, the laboratory is reminded to "verify that the software is installed on computers suited to run the software, that the system has been properly installed, and that the configurations are correct" (2.1) and that the following system control measures are in place: "every software release should have a unique version number" (2.2.1), "appropriate security protection [should exist] to ensure only authorized users can access the software and data" (2.2.2), that "audit trails to track changes to system data and/or verification of system settings [are] in place each time a calculation is run" (2.2.3), and that "user-level security [exists] to ensure that system users only perform authorized actions" (2.2.4).

The developmental validation section of these guidelines stresses the importance of demonstrating "any known or potential limitations of the system" and emphasizes that "the underlying scientific principle(s) of the probabilistic genotyping methods and characteristics of the software should be published in a peer-reviewed scientific journal" and that these principles may include "modeling of stutter, allelic drop-in and drop-out, Bayesian prior assumptions such as allelic probabilities, and statistical formulae used in the calculation and algorithms" (3.1).

According to the 2015 SWGDAM guidelines, studies that should be performed for developmental validation include sensitivity (3.2.1), specificity (3.2.2), precision (3.2.3), case-type samples (3.2.4), control samples (3.2.5), and accuracy (3.2.6). Studies should "assess the ability of the system to reliably determine the presence of a contributor's(s') DNA over a broad variety of evidentiary typing results (to include mixtures and low-level DNA quantities)" with "various sample types (e.g., different numbers of contributors, mixture proportions, and template quantities)" (SWGDAM 2015). The 2015 SWGDAM guidelines emphasize the need to understand the sensitivity and specificity of performance over a variety of conditions.

Under section 4 on internal validation, these guidelines state:

> "Data should be selected to test the system's capabilities and to identify its limitations. In particular, complex mixtures and low-level contributors should be evaluated thoroughly during internal validation, as the data from such samples generally help to define the software's limitations…" (SWGDAM 2015).

Internal validation should address samples with known contributors (4.1.1), hypothesis testing with contributors and non-contributors (4.1.2), variable DNA typing conditions (4.1.3), allelic peak height including off-scale data (4.1.4), single-source samples (4.1.5), DNA mixtures with various contributor ratios (4.1.6.1), various total DNA template quantities (4.1.6.2), various number of contributors (4.1.6.3), over- and under-estimating the number of contributors (4.1.6.4), allele sharing among contributors (4.1.6.5), partial profiles (4.1.7), allele drop-in (4.1.8), forward and reverse stutter (4.1.9), intra-locus peak height variation (4.1.10), inter-locus peak height variation (4.1.11), use of a different data set to establish software parameters and perform validation studies (4.1.12), sensitivity, specificity and precision studies (4.1.13), and additional challenge testing, such as the inclusion of non-allelic peaks from bleed-through or voltage spikes (4.1.14) (SWGDAM 2015).

### 5.4.3. ISFG 2016 DNA Commission on Software Validation

In November 2016, the ISFG DNA Commission provided 16 recommendations on validation of software programs used in forensic genetics to perform biostatistical calculations (Coble et al. 2016). These recommendations are summarized as follows:

(1) software should be supported by a user manual and scientific publications describing the data model(s) used to permit independent recalculation to verify reproducibility of all computations;

(2) validation should be according to specified requirements and intended use with publicly available or disclosed data sets;

(3) each software version should be distinguishable and independently validated;

(4) software developers should provide instructions to users on how to validate and configure their software;

(5) a user manual should accompany software to enable trained users to understand and explain results;

(6) laboratories are responsible to provide sufficient training resources and support for users;

(7) software source code should be placed in a secure repository and algorithms described in sufficient details to ensure continued availability of software in the future;

(8) software bugs and their fixes need to be disclosed and users notified about updates and any quality assurance issues;

(9) software using random permutation algorithms, such as MCMC, needs to have a feature to set this function to a stable mode for repeatability testing purposes;

(10) laboratories should develop a documented validation plan prior to initiating software validation and have supporting publications describing the models, propositions, and parameters used by the software;

(11) laboratories should test the software on representative data generated in-house with reagents, instruments, analysis software, and conditions used routinely for casework;

(12) laboratories should test true donors ($H_1$ true) and non-donors ($H_2$ true) as well as related and unrelated individuals across a range of situations that span or exceed the complexity of cases likely to be encountered in casework;

(13) laboratories should determine whether software results are consistent with previous interpretation procedures if the data and/or methods exist;

(14) laboratories should develop standard operating procedures based on their internal validation data and outline the types of cases and data to which the software can be applied;

(15) laboratories should develop and follow a policy or procedure for training software end users in the laboratory; and

(16) the forensic community is encouraged to establish a public repository of typing results, including results from different challenging scenarios like low-level mixtures and related contributors, in a universal, standardized file format and to have this repository governed by a neutral organization to permit equal access to all interested international parties.

### 5.4.4. **ENFSI DNA Working Group 2017 Best Practice Manual**

In May 2017, the ENFSI DNA Working Group issued a "Best Practice Manual for the internal validation of probabilistic software to undertake DNA mixture interpretation" that was intended to build upon the ISFG 2016 recommendations (see previous section and ENFSI 2022). This document focuses on internal validation performed within a forensic laboratory. Regarding previous developmental (termed "external") validation, this best practice manual notes:

> "It will be a decision for the laboratory to be satisfied that the external validation is 'fit-for-purpose' within the scope of its intended use" (ENFSI 2017).

Section 4.1 in this document states:

> "…a person(s) should be nominated to be responsible to act as the 'local expert' with the broadest knowledge about the software."

Section 4.2 recommends:

> "The software developer should create instructions on how to validate and configure software within the laboratory…and supply a user manual…for end users" (ENFSI 2017).

A documented validation plan should be developed to take into account the types of samples the laboratory plans to analyze (Section 6.1). Mock casework samples that span the kinds of samples routinely tested by that laboratory, where ground truth is known, should be used (Section 6.2) and, where possible, results produced by the software should be compared for consistency with previous interpretation procedures used by the laboratory (Section 6.3). The laboratory should "establish a series of criteria that define the limitations of testing," such as "if the profile of interest is predominantly below some defined level or a specified number of alleles have dropped-out (under the prosecution hypothesis)" (Section 6.4). This document emphasizes:

> "It is important that users have a clear understanding on the limitations. To facilitate this, users must be presented with examples considered unsuitable for testing" (ENFSI 2017).

This ENFSI guidance document also discusses the probability of drop-in (Section 6.5), proficiency testing (Section 7), training (Section 12.1), and presentation of evidence (Section 13) and contains an appendix on terminology for probabilistic mixture models (Section 16.1).

### 5.4.5. **UK Forensic Science Regulator 2018 Guidance**

In July 2018, the UK Forensic Science Regulator issued a 53-page guidance document on software validation for DNA mixture interpretation (UKFSR 2018b). A few points are highlighted here.

Section 6.1 discusses validation considerations specific to likelihood ratio calculations given that there is no "true" value for an LR. Section 6.2 reviews desired performance parameters (e.g., the software should be capable of analyzing three-person mixtures at a minimum), principles that should be incorporated into a DNA mixture interpretation model (e.g., limitations of all approaches should be made apparent to the customer), and routine operating quality checks required and data input considerations (e.g., an assessment of the evidence profile in the context of case circumstances, where possible, should always be undertaken before the use of software).

Section 7 reviews the process of validation defined in the UK Forensic Science Regulator's *Codes of Practice and Conduct* available at the time (UKFSR 2017) and a 2014 guidance document on validation (UKFSR 2014). Three additional stages are included with DNA mixture interpretation: (1) validation of the statistical model, (2) software development and testing, and (3) user acceptance testing.

Under Section 7.5 covering conceptual and operational validation of the statistical model, this guidance document states:

> "…ideally *the underlying data on which conclusions are based should also be made available*, for example, as supplementary material within the journal or access provided online to downloadable material *including all data and a full statistical description*. This enables other scientists in the field to inspect it independently and verify the results obtained in order to enable general acceptance of the model concept within the scientific community. Such transparency is essential for any software used within the [criminal justice system], for which there can be no 'secret science'" (UKFSR 2018b, p. 25, emphasis added).

The guidance continues:

> "…[software] testing should utilize a variety of ground-truth cases for which the composition is known, and are of varying degrees of quality and complexity that represent the full spectrum of data that may typically be encountered in casework" (UKFSR 2018b, p. 25).

Assessment of reproducibility is needed including the magnitude of the variation when a statistical model

> "does not return precisely the same number on replicate analyses of identical data" (UKFSR 2018b, p. 26).

Also encouraged are boundary testing to experimentally determine the impact of increasing the number of contributors and benchmarking exercises comparing results with other software

models or manual calculations that may be feasible with less complex data assessments (UKFSR 2018b, p. 26).

In addition, Section 7.10 of the UK guidance encourages creation of a validation library to maintain documentation from validation studies conducted and associated supporting materials including published articles and technical reports. Sections 8.1.4 and 8.1.5 state:

> "...the existing evidence that has been produced by a third party, and on which reliance is placed, must be relevant, available and adequate" and "the details of the analysis undertaken are both transparent and accessible to third parties" (UKFSR 2018b, p. 35).

### 5.4.6. **ANSI/ASB 018 PGS Validation Standard**

In July 2020, the AAFS Standards Board published the first standard on PGS validation (ANSI/ASB 018). The foreword states:

> "Validations of [PGS] systems provide the study results and conclusions necessary for customers or forensic science service providers to have confidence in the evidence provided."

This document continues:

> "…each laboratory will need to perform internal studies to demonstrate the reliability of the software and any potential limitations."

The bibliography cited in Annex B of the document includes 16 references.

Under this new standard, developmental validation (4.1.2) and internal validation (4.1.3) require accuracy, sensitivity, specificity, and precision studies with:

> "case-type profiles of known composition that represent (in terms of number of contributors, mixture ratios, and total DNA template quantities) the range of scenarios that would likely be encountered in casework. Studies shall not be limited to pristine DNA but shall also include compromised DNA samples (e.g., low template, degraded, and inhibited samples)" (ANSI/ASB 018).

This standard also states:

> "The internal validation shall not exceed the scope of the conditions tested in the developmental validation" (4.1.3), "All validation and performance check studies conducted by the laboratory shall be documented and retained by the laboratory" (4.5), and "Prior to implementation, the laboratory shall verify the functionality of its defined software settings and parameters utilizing different data sets than what were originally used to establish those settings and parameters" (4.7) (ANSI/ASB 018).

Annex A with supporting information states:

> "Repeated testing and data analysis are critical to the understanding of variability. While specific requirements for the minimum number of studies and sample sets used for validation studies are not detailed in this standard, the laboratory shall *perform sufficient studies* to address the variability inherent to the various aspects of DNA

> testing, data generation, analysis and interpretation of data and user input parameters" (4.1.3) (ANSI/ASB 018, emphasis added).

It continues:

> "All internal validation and performance check studies shall be documented and retained by the laboratory. Any validation and performance check studies may take a significant amount of time and are likely to result in a considerable amount of documentation output material. It is incumbent upon any laboratory performing these studies to *retain these results for the examination and evaluation by third parties*. The *results should be documented in such a way that the performance checks and validations can be reproduced* and decisions made on the basis of these studies documented…" (ANSI/ASB 018, emphasis added).

As emphasized in previous guidance documents, internal validation studies of PGS software need to be sufficient to assess variability across the types of DNA mixtures expected to be seen in a laboratory, and results from these studies should be available for third-party review.

## 5.5. **Concerns Raised by PCAST 2016 on DNA Mixture Interpretation**

Of the 2100 references[17] compiled in conjunction with the September 2016 report by the President's Council of Advisors on Science and Technology (PCAST), there were 294 publications listed in the DNA section. In the PCAST discussion of complex mixtures (PCAST 2016, pp. 75-83), the authors cited eight articles on PGS (Bille et al. 2014, Bright et al. 2014, Taylor 2014, Greenspoon et al. 2015, Perlin et al. 2015, Taylor et al. 2015, Taylor & Buckleton 2015, Bright et al. 2016). After examining these PGS references, the PCAST authors share their judgments (but not their specific criteria for reliability):

> "…*current studies have adequately explored only a limited range of mixture types* (with respect to number of contributors, ratio of minor contributors, and total amount of DNA). The two most widely used methods (STRmix and TrueAllele) appear to be reliable within a certain range, based on the available evidence and the inherent difficulty of the problem. Specifically, these methods appear to be reliable for three-person mixtures in which the minor[18] constitutes at least 20 percent of the intact DNA in the mixture and in which the DNA amount exceeds the minimum level required for the method.[19] *For more complex mixtures (e.g., more contributors or lower proportions), there is relatively little published evidence*… When further studies are published, it will likely be possible to extend the range in which scientific validity has been established to include more challenging samples" (PCAST 2016, pp. 80-81, emphasis added).

Since specific judgment criteria used by PCAST are not stated in their report, it is unclear on what basis PCAST claims that PGS "methods appear to be reliable." Publicly available data from validation studies, whether or not this information has been published in a peer-reviewed

[17] https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_references.pdf (accessed July 1, 2024)

[18] Changed to "person of interest" in a January 2017 Addendum to the PCAST Report (see p.8 of https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf)

[19] We note that this PCAST statement does not say anything about the quantity of DNA from the minor contributor(s).

journal, can be utilized by a user (e.g., the DNA analyst when the provider is the PGS developer or the court when the analyst is providing their results) to scrutinize the underlying data and supporting details for what is currently possible in research settings (what PCAST terms "scientific or foundational validity") and ascertain whether what is actually happening in casework settings (what PCAST calls "validity as applied") reflects the claims of the validation data.

A follow-on Addendum to the PCAST Report published four months later states:

> "PCAST found that empirical testing of [PGS] had largely been limited to a narrow range of parameters (number and ratio of contributors)… The path forward is straightforward. The validity of specific [probabilistic genotyping] software should be validated by testing a diverse collection of samples within well-defined ranges." (PCAST 2017, pp. 8-9).

In a footnote to their statement "there has been little empirical validation," the PCAST Addendum concludes:

> "The few studies that have explored 4- or 5-person mixtures often involve mixtures that are derived from only a few sets of people (in some cases, only one). Because *the nature of overlap among alleles is a key issue*, it is critical to examine mixtures from various different sets of people. In addition, the studies involve few mixtures in which a sample is present at an extremely low ratio. *By expanding these empirical studies, it should be possible to test validity and reliability across a broader range*" (PCAST 2017, footnote #11, emphasis added).

Thus, the PCAST Report (PCAST 2016) and its Addendum (PCAST 2017) emphasize a need to have casework with factor space coverage represented in the empirical studies that are performed and shared for independent review as the field adopts PGS methods to assist with DNA mixture interpretation. PCAST specifically mentions the benefits of testing "mixtures from various different sets of people" to explore PGS performance in terms of the degree of allele overlap from contributors and "mixtures in which a sample is present at an extremely low ratio."

As noted above, when their analysis was performed in 2016, PCAST provided their opinion that "current studies *have adequately explored only a limited range of mixture types* (with respect to number of contributors, ratio of minor contributors, and total amount of DNA)" (PCAST 2016, emphasis added). Since 2016, there have been a number of additional studies exploring aspects of various PGS systems for assigning likelihood ratios involving sub-source propositions (see chapter 4 in NISTIR 8351 and NISTIR 8351sup2). For example, in response to PCAST's call for more validation data, a group of 31 laboratories compiled 2825 DNA mixtures from eight different STR kits consisting of 3-, 4-, 5-, and 6-contributors. These profiles were then interpreted using the PGS system STRmix by staff members of the software developer (Bright et al. 2018).

## 6. History of DNA Mixture Interpretation Training

The 2007 article "Interpretation of DNA mixtures – European consensus on principles" emphasizes that:

> "laboratories must invest in continuous education of the staff in the interpretation of DNA mixtures" (Morling et al. 2007).

This point had been made previously by the ISFG DNA Commission:

> "Our discussions have highlighted a significant need for continuing education and research in this area [DNA mixture interpretation]" (Gill et al. 2006).

A brief history of training workshops on this topic is included below. Further thoughts on needs and potential solutions for education, training, and professional credentialing may be found in a May 2024 NIST report (EWG 2024, pp. 241-274).

### 6.1. Initial U.S. Training Workshop on Mixtures

The first DNA mixture training course in the United States was held as part of a scientific conference in Annapolis, Maryland, sponsored by International Business Communications on July 31, 1998 (IBC 1998). This workshop, titled "Resolution and Interpretation of Mixtures," included presentations by Peter Gill of the UK Forensic Science Service ("Distinguishing between Alleles, Artifacts and Genetic Anomalies in Mixture Interpretation"); James Curran, then working with Bruce Weir in the statistics department of North Carolina State University ("Calculating the Evidentiary Strength of Mixed DNA Profiles"); and Charles Brenner, a consultant in forensic mathematics ("Some Considerations of Race, Number and Accuracy").

Peter Gill began his July 1998 workshop presentation with the admonition: "Don't do mixture interpretation unless you have to!" He explained that forensic cases often have multiple stains and that a selection should be made, where possible, of samples that do not contain mixtures. He also emphasized that it was important to ensure that any mixtures obtained were consistent with casework circumstances (IBC 1998).

At this workshop, James Curran taught

> "if numbers are to be provided, they must be calculated with the same attention to appropriate methods as is given to the generation of the profiles in the first place" and "the key issue is to decide upon possible explanations for the mixed stains."

He worked through some examples in calculating likelihood ratios and the underlying assumptions (IBC 1998). Both James Curran and Peter Gill acknowledged John Buckleton's contribution to their work. Most of the individuals who participated in this first DNA mixture workshop almost 25 years ago are still active in the field, and the primary issues discussed have not changed.

## 6.2. **Training on Principles**

To assist forensic DNA analysts in understanding issues and principles underpinning DNA mixture interpretation, more than 50 training workshops and presentations were organized or given by researchers from the National Institute of Standards and Technology (NIST) and collaborators (see below) between 2005 and 2014 (see Butler 2014, Table 6.5). Slides for many of these workshops (e.g., AAFS 2008, AAFS 2011, ISHI 2010, ISHI 2011, ISHI 2012) are available on the NIST STRBase website[20].

Researchers from Boston University (BU) received a training grant from the National Institute of Justice (NIJ) that funded DNA mixture interpretation training workshops in 2010, 2011, and 2012. In addition, BU created a training website[21] with 12 lessons that examine the various steps of mixture interpretation. In addition, the BU website contains more than 2,700 .fsa files with single-source, two-person, three-person, and four-person mixtures at different mixture ratios and DNA amounts that can be downloaded and used in training programs. An even more extensive set of DNA mixture profiles, known as PROVEDIt (Alfonse et al. 2018), is available[22] from Professor Catherine Grgicak now at Rutgers University.

The ISFG also maintains educational workshop materials shared at biennial conferences for its members[23] on a variety of topics including DNA mixture interpretation.

## 6.3. **Training on Probabilistic Genotyping Software**

With the development and implementation of PGS systems, software-specific training courses have been created. In 2012, the European Forensic Genetics Network of Excellence (EuroForGen-NoE) created an online training academy[24] with webinars discussing DNA mixture interpretation using an open-source PGS system LRmix. The EuroForGen-NoE group demonstrated that training and use of a common PGS system could lead to uniformity of results obtained with DNA mixtures (Prieto et al. 2014).

Vendors providing PGS programs conduct training courses to support their appropriate use. For example, the STRmix team has provided almost 100 training courses between 2014 and 2018 with durations ranging from one to five days[25].

A 2019 webinar series organized by the FBI Laboratory introduced hundreds of DNA analysts to PGS theory, methods, and software (Table S1.5).

---

[20] See https://strbase-archive.nist.gov/ (accessed July 1, 2024)
[21] https://www.bu.edu/dnamixtures/ (accessed July 1, 2024)
[22] https://lftdi.camden.rutgers.edu/provedit/files/ (accessed July 1, 2024)
[23] https://www.isfg.org/Members+Area/Education (accessed July 1, 2024)
[24] https://www.euroforgen.eu/training/online-training-academy/ (no longer available)
[25] See https://johnbuckleton.wordpress.com/wp-content/uploads/2018/08/training.pdf (accessed July 1, 2024)

**Table S1.5.** Webinar series on DNA mixture interpretation and probabilistic genotyping organized by FBI Laboratory and NIJ's Forensic Technology Center of Excellence. Original webinars were held from May 1, 2019 to July 17, 2019 and are available in archived format at https://forensiccoe.org/webinar/online-workshop-series-probabilistic-genotyping-of-evidentiary-dna-typing-results/ (accessed July 1, 2024). Abbreviations: DOJ = Department of Justice, ESR = Institute of Environmental Science and Research, FBI = Federal Bureau of Investigation, FSSA = Forensic Science South Australia, NFI = Netherlands Forensic Institute, UNTHSC = University of North Texas Health Science Center.

| Lesson | Topics | Presenters |
|---|---|---|
| Module 1 | Mixture interpretation and introduction to probabilistic genotyping software (PGS) | **Tamyra Moretti** (FBI Laboratory, USA)<br>**Peter Gill** (University of Oslo, Norway)<br>**Lynn Garcia** (Texas Forensic Science Commission, USA) |
| Module 2 | Statistical aspects of PGS | **David Balding** (University of Melbourne, Australia)<br>**Mike Coble** (UNTHSC, USA)<br>**Steven Myers** (California DOJ, USA)<br>**John Buckleton** (ESR, New Zealand) |
| Module 3 | PGS software and output: instructive overviews | **John Buckleton** (ESR, New Zealand)<br>**Mike Coble** (UNTHSC, USA)<br>**Peter Gill** (University of Oslo, Norway)<br>**Mark Perlin** (Cybergenetics, USA) |
| Module 4 | Validation of PGS | **Tamyra Moretti** (FBI Laboratory, USA)<br>**Sarah Noël** (Montreal, Canada)<br>**Duncan Taylor** (FSSA, Australia) |
| Module 5 | Representation of statistical weight to stakeholders and the court | **David Kaye** (Penn State Law School, USA)<br>**Tamyra Moretti** (FBI Laboratory, USA)<br>**Steven Myers** (California DOJ, USA) |
| Module 6 | PGS in U.S. courts | **John Buckleton** (ESR, New Zealand)<br>**Jerrilyn Conway** (FBI Laboratory, USA)<br>**Dawn Herkenham** (Leidos, USA)<br>**Mark Perlin** (Cybergenetics, USA) |
| Module 7 | Uncertainty and limitations of PGS | **Amke Caliebe** (University of Kiel, Germany)<br>**Zane Kerr** (ESR, New Zealand)<br>**Klaas Slooten** (NFI, The Netherlands)<br>**Bianka Szkuta** (Victoria Police, Australia) |
| Module 8 | PGS summation and special topics | **Jo-Anne Bright** (ESR, New Zealand)<br>**Ted Hunt** (USDOJ, USA)<br>**Klaas Slooten** (NFI, The Netherlands) |

It is anticipated that more workshops and online training webinars will continue to be needed and valued as use of PGS grows.

## 7. **Looking to the Future: Opportunities and Needs**

Since initially described in the mid-1980s (Gill et al. 1985), forensic DNA methods have evolved and become more sensitive. This change has necessitated new approaches to DNA mixture interpretation (e.g., Gill et al. 2000). Often important changes and advances have been made across the community because of experiences in high-profile court cases or awareness of issues raised through participation in interlaboratory studies or collaborative exercises. Several of these cases and studies are highlighted in this brief history of the field.

Guidance documents and training efforts have played an important role in the history of DNA mixture interpretation, in particular the 2006 ISFG DNA Commission (e.g., Gill et al. 2006). These recommendations emphasize the value of using likelihood ratios with mixture deconvolution and review difficulties when interpreting minor components in the presence of artifacts like STR stutter products and stochastic variation present with low amounts of DNA.

Effective training and continuing education of forensic practitioners are crucial to keep up with the evolving forensic DNA technologies and applications. Given these ongoing changes, “laboratories must invest in continuous education of the staff in the interpretation of DNA mixtures” (Morling et al. 2007). Stakeholders in the criminal justice system (e.g., law enforcement personnel, lawyers, and judges) using DNA results also benefit from regular training and continuing education to understand capabilities and limitations.

A culture of critical thinking and clear communication regarding DNA mixture interpretation is crucial as probabilistic genotyping software programs are implemented and results from low-level, complex mixtures are shared in written reports and court testimony. Defining what analysts *need to know* rather than simply what they *need to do* will increase confidence and enhance practice.

Future needs include the promotion of performance-based approaches to validation studies (see NISTIR 8351, chapter 4) and continuing education and research in DNA mixture interpretation. It would be helpful to have training workshops and seminars on validation to assist the forensic DNA community and stakeholders in strengthening DNA mixture interpretation. A report published in May 2024 on improving practice with forensic DNA interpretation has an entire chapter devoted to education, training, and professional credentialing (EWG 2024, pp. 241-274).

## References

AAFS 2008 DNA Mixture Workshop: https://strbase-archive.nist.gov/training/AAFS2008_MixtureWorkshop.htm (accessed July 1, 2024).

AAFS 2011 DNA Mixture Workshop: https://strbase-archive.nist.gov/training/AAFS2011_MixtureWorkshop.htm (accessed July 1, 2024).

Adelman JD, Zhao A, Eberst DS, Marciano MA (2019) Automated detection and removal of capillary electrophoresis artifacts due to spectral overlap. *Electrophoresis* 40(14):1753-1761. https://doi.org/10.1002/elps.201900060

Alfonse LE, Tejada G, Swaminathan H, Lun DS, Grgicak CM (2017) Inferring the number of contributors to complex DNA mixtures using three methods: Exploring the limits of low-template DNA interpretation. *Journal of Forensic Sciences* 62(2):308-316. https://doi.org/10.1111/1556-4029.13284

Alfonse LE, Garrett AD, Lun DS, Duffy KR, Grgicak CM (2018) A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt. *Forensic Science International: Genetics* 32:62-70. https://doi.org/10.1016/j.fsigen.2017.10.006

ANAB ANSI National Accreditation Board (2023) Accreditation Requirements for Forensic Testing and Calibration – (AR 3125 2023/02/01). Available at https://anab.qualtraxcloud.com/ShowDocument.aspx?ID=12371 (accessed July 1, 2024).

ANSI/ASB Standard 020 (2018) Standard for Validation Studies of DNA Mixtures, and Development and Verification of a Laboratory's Mixture Interpretation Protocol. Available at https://www.aafs.org/asb-standard/standard-validation-studies-dna-mixtures-and-development-and-verification-laboratorys (accessed July 1, 2024).

ANSI/ASB Standard 040 (2019) Standard for Forensic DNA Interpretation and Comparison Protocols. Available at https://www.aafs.org/asb-standard/standard-forensic-dna-interpretation-and-comparison-protocols (accessed July 1, 2024).

ANSI/ASB Standard 018 (2020) Standard for Validation of Probabilistic Genotyping Systems. Available at https://www.aafs.org/asb-standard/standard-validation-probabilistic-genotyping-systems (accessed July 1, 2024).

Aronson JD (2007) *Genetic Witness: Science, Law, and Controversy in the Making of DNA Profiling*. Rutgers University Press: New Brunswick, NJ.

Balding DJ & Buckleton J (2009) Interpreting low template DNA profiles. *Forensic Science International: Genetics* 4(1):1-10. https://doi.org/10.1016/j.fsigen.2009.03.003

Balding DJ (2013) Evaluation of mixed-source, low-template DNA profiles in forensic science. *Proceedings of the National Academy of Sciences of the United States of America* 110(30):12241-12246. https://doi.org/10.1073/pnas.1219739110

Barber M, Evett I, Pope S, Sullivan K, Whitaker J (2015) HOS/13/037 Forensic Science Regulator Codes of Practice and Conduct – Proficiency testing guidance for DNA mixtures. Final report. January 2015. Principal Forensic Services Ltd.

Barrio PA, Crespillo M, Luque JA, Aler M, Baeza-Richer C, Baldassarri L, Carnevali E, Coufalova P, Flores I, García O, García MA, González R, Hernández A, Inglés V, Luque GM, Mosquera-Miguel A, Pedrosa S, Pontes ML, Porto MJ, Posada Y, Ramella MI, Ribeiro T, Riego E, Sala A, Saragoni VG, Serrano A, Vannelli S (2018) GHEP-ISFG collaborative exercise on mixture profiles (GHEP-MIX06). Reporting conclusions: results and evaluation. *Forensic Science International: Genetics* 35: 156-163. https://doi.org/10.1016/j.fsigen.2018.05.005

Bennett L, Oldoni F, Long K, Cisana S, Maddela K, Wootton S, Chang J, Hasegawa R, Lagace R, Kidd KK, Podini D (2019) Mixture deconvolution by massively parallel sequencing of microhaplotypes. *International Journal of Legal Medicine* 133(3):719-729. https://doi.org/10.1007/s00414-019-02010-7

Benschop CC, van der Beek CP, Meiland HC, van Gorp AG, Westen AA, and Sijen, T. (2011) Low template STR typing: effect of replicate number and consensus method on genotyping reliability and DNA database search results. *Forensic Science International: Genetics* 5(4): 316-328. https://doi.org/10.1016/j.fsigen.2010.06.006

Benschop CCG, Haned H, Jeurissen L, Gill PD, Sijen T (2015) The effect of varying the number of contributors on likelihood ratios for complex DNA mixtures. *Forensic Science International: Genetics* 19:92-99. https://doi.org/10.1016/j.fsigen.2015.07.003

Benschop CC, Connolly E, Ansell R, Kokshoorn B (2017) Results of an inter and intra laboratory exercise on the assessment of complex autosomal DNA profiles. *Science & Justice* 57(1): 21-27. https://doi.org/10.1016/j.scijus.2016.10.001

Benschop CCG, van der Linden J, Hoogenboom J, Ypma R, Haned H (2019) Automated estimation of the number of contributors in autosomal short tandem repeat profiles using a machine learning approach. *Forensic Science International: Genetics* 43:102150. https://doi.org/10.1016/j.fsigen.2019.102150

Bieber FR, Buckleton JS, Budowle B, Butler JM, Coble MD (2016) Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion. *BMC Genetics* 17(1):125. https://doi.org/10.1186/s12863-016-0429-7

Bill M, Gill P, Curran J, Clayton T, Pinchin R, Healy M, Buckleton J (2005) PENDULUM – a guideline-based approach to the interpretation of STR mixtures. *Forensic Science International* 148:181-189. https://doi.org/10.1016/j.forsciint.2004.06.037

Bille TW, Weitz SM, Coble MD, Buckleton J, Bright JA (2014) Comparison of the performance of different models for the interpretation of low level mixed DNA profiles. *Electrophoresis* 35:3125-3133. https://doi.org/10.1002/elps.201400110

Blake E, Mihalovich J, Higuchi R, Walsh PS, Erlich H (1992) Polymerase chain reaction (PCR) amplification and human leukocyte antigen (HLA)-DQ alpha oligonucleotide typing on biological evidence samples: casework experience. *Journal of Forensic Sciences* 37(3):700-726.

Bleka Ø, Storvik G, Gill P (2016) EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. *Forensic Science International: Genetics* 21:35-44. https://doi.org/10.1016/j.fsigen.2015.11.008

Brenner CH, Fimmers R, Baur MP (1996) Likelihood ratios for mixed stains when the number of donors cannot be agreed. *International Journal of Legal Medicine* 109(4):218-219. https://doi.org/10.1007/BF01225523

Bright JA, Taylor D, Curran J, Buckleton J (2014) Searching mixed DNA profiles directly against profile databases. *Forensic Science International: Genetics* 9:102-110. https://doi.org/10.1016/j.fsigen.2013.12.001

Bright JA, Evett IW, Taylor D, Curran JM, Buckleton J (2015) A series of recommended tests when validating probabilistic DNA profile interpretation software. *Forensic Science International: Genetics* 14:125-131. https://doi.org/10.1016/j.fsigen.2014.09.019

Bright JA, Taylor D, McGovern C, Cooper S, Russell L, Abarno D, Buckleton J (2016) Developmental validation of STRmix, expert software for the interpretation of forensic DNA profiles. *Forensic Science International: Genetics* 23:226-239. https://doi.org/10.1016/j.fsigen.2016.05.007

Bright JA, Taylor D, Gittelson S, Buckleton J (2017) The paradigm shift in DNA profile interpretation. *Forensic Science International: Genetics* 31:e24-e32. https://doi.org/10.1016/j.fsigen.2017.08.005

Bright JA, Richards R, Kruijver M, Kelly H, McGovern C, Magee A, McWhorter A, Ciecko A, Peck B, Baumgartner C, Buettner C, McWilliams S, McKenna C, Gallacher C, Mallinder B, Wright D, Johnson D, Catella D, Lien E, O'Connor C, Duncan G, Bundy J, Echard J, Lowe J, Stewart J, Corrado K, Gentile S, Kaplan M, Hassler M, McDonald N, Hulme P, Oefelein RH, Montpetit S, Strong M, Noel S, Malsom S, Myers S, Welti S, Moretti T, McMahon T, Grill T, Kalafut T, Greer-Ritzheimer MM, Beamer V, Taylor DA & Buckleton JS (2018) Internal validation of STRmix™– a multi laboratory response to PCAST. *Forensic Science International: Genetics* 34:11-24. https://doi.org/10.1016/j.fsigen.2018.01.003

Bright JA, Cheng K, Kerr Z, McGovern C, Kelly H, Moretti TR, Smith MA, Bieber FR, Budowle B, Coble MD, Alghafri R, Allen PS, Barber A, Beamer V, Buettner C, Russell M, Gehrig C, Hicks T, Charak J, Cheong-Wing K, Ciecko A, Davis CT, Donley M, Pedersen N, Gartside B, Granger D, Greer-Ritzheimer M, Reisinger E, Kennedy J, Grammer E, Kaplan M, Hansen D, Larsen HJ, Laureano A, Li C, Lien E, Lindberg E, Kelly C, Mallinder B, Malsom S, Yacovone-Margetts A, McWhorter A, Prajapati SM, Powell T, Shutler G, Stevenson K, Stonehouse AR, Smith L, Murakami J, Halsing E, Wright D, Clark L, Taylor DA, Buckleton J (2019) STRmix™ collaborative exercise on DNA mixture interpretation. *Forensic Science International: Genetics* 40:1-8. https://doi.org/10.1016/j.fsigen.2019.01.006

Brinkac LM, Richetelli N, Davoren JM, Bever RA, Hicklin RA (2023) DNAmix 2021: Laboratory policies, procedures, and casework scenarios summary and dataset. *Data in Brief* 48:109150. https://doi.org/10.1016/j.dib.2023.109150

Buckleton JS, Evett IW, Weir BS (1998) Setting bounds for the likelihood ratio when multiple hypotheses are postulated. *Science & Justice* 38(1):23-26. https://doi.org/10.1016/S1355-0306(98)72069-6

Buckleton JS, Curran JM, Gill P (2007) Towards understanding the effect of uncertainty in the number of contributors to DNA stains. *Forensic Science International: Genetics* 1(1):20-28. https://doi.org/10.1016/j.fsigen.2006.09.002

Buckleton J & Curran J (2008) A discussion of the merits of random man not excluded and likelihood ratios. *Forensic Science International: Genetics* 2(4):343-348. https://doi.org/10.1016/j.fsigen.2008.05.005

Buckleton JS, Bright J-A, Gittelson S, Moretti TR, Onorato AJ, Bieber FR, Budowle B, Taylor DA (2019) The probabilistic genotyping software STRmix: utility and evidence for its validity. *Journal of Forensic Sciences* 64(2):393-405. https://doi.org/10.1111/1556-4029.13898

Budowle B, Lindsey JA, DeCou JA, Koons BW, Giusti AM, Comey CT (1995) 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* 40(1):45-54.

Budowle B, Moretti TR, Niezgoda SJ, Brown BL (1998). CODIS and PCR-based short tandem repeat loci: law enforcement tools. *Proceedings of the Second European Symposium on Human Identification* (pp. 73–88). Madison, Wisconsin: Promega Corporation. Available at https://www.promega.com/-/media/files/resources/conference-proceedings/ishi-02/oral-presentations/17.pdf?la=en (accessed July 1, 2024).

Budowle B, Hobson DL, Smerick JB, Smith JAL (2001) Low copy number – consideration and caution. *Proceedings of the Twelfth International Symposium on Human Identification*. Madison, WI: Promega Corporation. Available at https://www.promega.com/~/media/files/resources/conference%20proceedings/ishi%2012/oral%20presentations/budowle.pdf?la=en (accessed July 1, 2024).

Budowle B, Onorato AJ, Callaghan TF, Della Manna A, Gross AM, Guerrieri RA, Luttman JC, McClure DL (2009) Mixture interpretation: defining the relevant features for guidelines for the assessment of mixed DNA profiles in forensic casework. *Journal of Forensic Sciences* 54(4):810-821. https://doi.org/10.1111/j.1556-4029.2009.01046.x

Burrill J, Daniel B, Frascione N (2019) A review of trace "touch DNA" deposits: Variability factors and an exploration of cellular composition. *Forensic Science International: Genetics* 39:8-18. https://doi.org/10.1016/j.fsigen.2018.11.019

Butler JM, Buel E, Crivellente F, McCord BR (2004). Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis* 25:1397-1412. https://doi.org/10.1002/elps.200305822

Butler JM, Coble MD, Vallone PM (2007) STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Science, Medicine and Pathology* 3(3):200-205. https://doi.org/10.1007/s12024-007-0018-1

Butler JM (2009) *Fundamentals of Forensic DNA Typing*. Elsevier Academic Press: San Diego. [see Chapter 3 "Historical Methods", pp. 43-78]

Butler JM & Hill CR (2010) Scientific issues with analysis of low amounts of DNA. *Profiles in DNA,* 13(1). Available at https://www.promega.com/resources/profiles-in-dna/2010/scientific-issues-with-analysis-of-low-amounts-of-dna/ (accessed July 1, 2024).

Butler JM (2011) *Advanced Topics in Forensic DNA Typing: Methodology*. Elsevier Academic Press: San Diego.

Butler JM (2013) Forensic DNA advisory groups: DAB, SWGDAM, ENFSI, and BSAG. In *Encyclopedia of Forensic Sciences, Second Edition*, pp. 339-343. Elsevier: San Diego.

Butler JM (2014) *Advanced Topics in Forensic DNA Typing: Interpretation*. Elsevier Academic Press: San Diego.

Butler JM (2015a) The future of forensic DNA analysis. *Philosophical Transactions of the Royal Society B* 370:20140252. https://doi.org/10.1098/rstb.2014.0252

Butler JM (2015b) U.S. initiatives to strengthen forensic science & international standards in forensic DNA. *Forensic Science International: Genetics* 18:4-20. https://doi.org/10.1016/j.fsigen.2015.06.008

Butler JM, Kline MC, Coble MD (2018) NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): variation observed and lessons learned. *Forensic Science International: Genetics* 37: 81-94. https://doi.org/10.1016/j.fsigen.2018.07.024

Butler JM & Willis S (2020) INTERPOL review of forensic biology and forensic DNA typing 2016-2019. *Forensic Science International: Synergy* 2:352-367. https://doi.org/10.1016/j.fsisyn.2019.12.002. Presentation available at https://strbase-archive.nist.gov/pub_pres/Butler-INTERPOL-DNAreview-Oct2019.pdf (accessed July 1, 2024).

Butler JM (2023) Recent advances in forensic biology and forensic DNA typing: INTERPOL review 2019-2022. *Forensic Science International: Synergy* 6:100311. https://doi.org/10.1016/j.fsisyn.2022.100311

Caskey CT, Edwards A, Hammond HA (1989) DNA: the history and future use in forensic analysis. *Proceedings of the International Symposium on the Forensic Aspects of DNA Analysis*. U.S. Government Printing Office: The Laboratory Division of the Federal Bureau of Investigation (June 19-23, 1989), pp. 3-9.

Clayton TM, Whitaker JP, Sparkes R, Gill P (1998) Analysis and interpretation of mixed forensic stains using DNA STR profiling. *Forensic Science International* 91(1):55-70. https://doi.org/10.1016/s0379-0738(97)00175-8

Clayton TM, Guest JL, Urquhart AJ, Gill PD (2004) A genetic basis for anomalous band patterns encountered during DNA STR profiling. *Journal of Forensic Sciences* 49(6):1207-1214.

Coble MD, Bright JA, Buckleton JS, Curran JM (2015) Uncertainty in the number of contributors in the proposed new CODIS set. *Forensic Science International: Genetics* 19:207-211. https://doi.org/10.1016/j.fsigen.2015.07.005

Coble MD, Buckleton J, Butler JM, Egeland T, Fimmers R, Gill P, Gusmao L, Guttman B, Krawczak M, Morling N, Parson W, Pinto N, Schneider PM, Sherry ST, Willuweit S, Prinz M (2016) DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications. *Forensic Science International: Genetics* 25:191-197. https://doi.org/10.1016/j.fsigen.2016.09.002

Coble MD & Bright J-A (2019) Probabilistic genotyping software: an overview. *Forensic Science International: Genetics* 38:219-224. https://doi.org/10.1016/j.fsigen.2018.11.009

Comey CT & Budowle B (1991) Validation studies on the analysis of the HLA DQ alpha locus using the polymerase chain reaction. *Journal of Forensic Sciences* 36(6):1633-1648.

Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998a) A model for case assessment and interpretation. *Science & Justice* 38(3):151-156. https://doi.org/10.1016/S1355-0306(98)72099-4

Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998b) A hierarchy of propositions: Deciding which level to address in casework. *Science & Justice* 38(4):231-239. https://doi.org/10.1016/S1355-0306(98)72117-3

Cooper S, McGovern C, Bright JA, Taylor D, Buckleton J (2015) Investigating a common approach to DNA profile interpretation using probabilistic software. *Forensic Science International: Genetics* 16:121-131. https://doi.org/10.1016/j.fsigen.2014.12.009

Cotton RW & Fisher MB (2015) Review: Properties of sperm and seminal fluid, informed by research on reproduction and contraception. *Forensic Science International: Genetics* 18:66-77. https://doi.org/10.1016/j.fsigen.2015.03.009

Crespillo M, Barrio PA, Luque JA, Alves C, Aler M, Alessandrini F, Andrade L, Barretto RM, Bofarull A, Costa S, García MA, García O, Gaviria A, Gladys A, Gorostiza A, Hernández A, Herrera M, Hombreiro L, Ibarra AA, Jiménez MJ, Luque GM, Madero P, Martínez-Jarreta B, Masciovecchio MV, Modesti NM, Moreno F, Pagano S, Pedrosa S, Plaza G, Prat E, Puente J, Rendo F, Ribeiro T, Sala A, Santamaría E, Saragoni VG, Whittle MR (2014) GHEP-ISFG collaborative exercise on mixture profiles of autosomal STRs (GHEP-MIX01, GHEP-MIX02 and GHEP-MIX03): results and evaluation. *Forensic Science International: Genetics* 10:64-72. https://doi.org/10.1016/j.fsigen.2014.01.009

Curran JM, Triggs CM, Buckleton J, Weir BS (1999) Interpreting DNA mixtures in structured populations. *Journal of Forensic Sciences* 44(5):987-995.

[DAB 2000] DNA Advisory Board (2000) Statistical and population genetics issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database(s). *Forensic Sci. Comm.* 2(3). Available at https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2000/dnastat.htm (accessed July 1, 2024).

Dembinski GM, Sobieralski C, Picard CJ (2018) Estimation of the number of contributors of theoretical mixture profiles based on allele counting: Does increasing the number of loci increase success rate of estimates?. *Forensic Science International: Genetics* 33:24-32. https://doi.org/10.1016/j.fsigen.2017.11.007

Duewer DL, Currie LA, Reeder DJ, Leigh SD, Liu H-K, Mudd JL (1995) Interlaboratory comparison of autoradiographic DNA profiling measurements. 2. Measurement uncertainty and its propagation. *Analytical Chemistry* 67:1220-1231. https://doi.org/10.1021/ac00103a013

Duewer DL, Currie LA, Reeder DJ, Leigh SD, Filiben JJ, Liu H-K, Mudd JL. (1997) Interlaboratory comparison of autoradiographic DNA profiling measurements. 4. Protocol effects. *Analytical Chemistry* 69:1882-1892. https://doi.org/10.1021/ac961070k

Duewer DL, Lalonde SA, Aubin RA, Fourney RM, Reeder DJ. (1998) Interlaboratory comparison of autoradiographic DNA profiling measurements: precision and concordance. *Journal of Forensic Sciences* 43:465-471.

Duewer DL, Liu H-K, Reeder DJ (1999) Graphical tools for RFLP DNA profiling. Single-locus charts. *Journal of Forensic Sciences* 44:969-977.

Duewer DL, Richie KL, Reeder DJ. (2000a) RFLP band size standards: NIST Standard Reference Material® 2390. *Journal of Forensic Sciences* 45:1093-1105.

Duewer DL, Gary KT, Reeder DJ. (2000b) RFLP band size standards: Cell line K562 values from 1991 – 1997 proficiency studies. *Journal of Forensic Sciences* 45:1106-1118.

Duewer DL, Kline MC, Redman JW, Newall PJ, Reeder DJ (2001) NIST Mixed Stain Studies #1 and #2: interlaboratory comparison of DNA quantification practice and short tandem repeat multiplex performance with multiple-source samples. *Journal of Forensic Sciences* 46:1199-1210.

Duewer DL, Kline MC, Redman JW, Butler JM (2004) NIST Mixed Stain Study #3: signal intensity balance in commercial short tandem repeat multiplexes. *Analytical Chemistry* 76:6928-6934. https://doi.org/10.1021/ac049178k

Duffy KR, Lun DS, Mulcahy MM, O'Donnell L, Sheth N, Grgicak CM (2023) Evidentiary evaluation of single cells renders highly informative forensic comparisons across multifarious admixtures. *Forensic Science International: Genetics* 64:102852. https://doi.org/10.1016/j.fsigen.2023.102852

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics* 49:746-756.

[ENFSI 2017] ENFSI DNA Working Group (2017) Best practice manual for the internal validation of probabilistic software to undertake DNA mixture interpretation. Available at https://enfsi.eu/wp-content/uploads/2024/04/Best-Practice-Manual-for-the-internal-validation-of-probabilistic-software-to-undertake-DNA-mixture-interpretation-v1.docx-1-1.pdf (accessed July 1, 2024).

[ENFSI 2022] ENFSI Best Practice Manual for Human Forensic Biology and DNA Profiling. Available at https://enfsi.eu/wp-content/uploads/2022/12/ENFSI-DNA-BPM-03.pdf (accessed August 8, 2024).

Evett IW, Buffery C, Willott G, Stoney D (1991) A guide to interpreting single locus profiles of DNA mixtures in forensic cases. *Journal of Forensic Science Society* 31(1):41-47. https://doi.org/10.1016/s0015-7368(91)73116-2

Evett IW & Weir BS (1998) *Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists*. Sinauer Associates: Sunderland, Massachusetts.

Evett IW, Gill PD, Lambert JA (1998) Taking account of peak areas when interpreting mixed DNA profiles. *Journal of Forensic Sciences* 43(1):62-69.

[EWG 2024] Expert Working Group on Human Factors in Forensic DNA Interpretation (2024) *Forensic DNA Interpretation and Human Factors: Improving Practice Through a Systems Approach.* National Institute of Standards and Technology, Gaithersburg, MD. NISTIR 8305. https://doi.org/10.6028/NIST.IR.8503

Fildes N & Reynolds R (1995) Consistency and reproducibility of AmpliType PM results between seven laboratories: field trial results. *Journal of Forensic Sciences* 40(2):279-286.

Findlay I, Taylor A, Quirke P, Frazier R, Urquhart A (1997) DNA fingerprinting from single cells. *Nature* 389:555-556. https://doi.org/10.1038/39225

[FTCOE 2015] Forensic Technology Center of Excellence (2015) *Landscape Study of DNA Mixture Interpretation Software*. Available at https://forensiccoe.org/report-dna-mixture-interpretation-software/ (accessed July 1, 2024).

Frégeau CJ & Fourney RM (1993) DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. *BioTechniques* 15(1):100-119.

Gary KT, Duewer DL, Reeder DJ. (1999) Graphical tools for RFLP DNA profiling. Laboratory Performance charts. *Journal of Forensic Sciences* 44:978-982.

Gettings KB, Borsuk LA, Steffen CR, Kiesler KM, Vallone PM (2018) Sequence-based U.S. population data for 27 autosomal STR loci. *Forensic Science International: Genetics* 37:106-115. https://doi.org/10.1016/j.fsigen.2018.07.013

Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic application of DNA 'fingerprints.' *Nature* 318:577-579. https://doi.org/10.1038/318577a0

Gill P & Werrett DJ (1987) Exclusion of a man charged with murder by DNA fingerprinting. *Forensic Science International* 35(2-3): 145-148. https://doi.org/10.1016/0379-0738(87)90050-8

Gill P, Kimpton CP, Urquhart A, Oldroyd N, Millican ES, Watson SK, Downes TJ (1995) Automated short tandem repeat (STR) analysis in forensic casework – a strategy for the future. *Electrophoresis* 16:1543-1552. https://doi.org/10.1002/elps.11501601257

Gill P, Sparkes R, Kimpton C (1997) Development of guidelines to designate alleles using an STR multiplex system. *Forensic Science International* 89(3):185-197. https://doi.org/10.1016/s0379-0738(97)00131-x

Gill P, Sparkes R, Pinchin R, Clayton T, Whitaker J, Buckleton J (1998) Interpreting simple STR mixtures using allele peak areas. *Forensic Science International* 91(1):41-53. https://doi.org/10.1016/s0379-0738(97)00174-6

Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Science International* 112(1):17-40. https://doi.org/10.1016/s0379-0738(00)00158-4

Gill P (2001) Application of low copy number DNA profiling. *Croatian Medical Journal* 42(3):229-232.

Gill P, Curran J, Elliot K (2005) A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. *Nucleic Acids Research* 33(2):632-643. https://doi.org/10.1093/nar/gki205

Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, Morling N, Prinz M, Schneider PM, Weir BS (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International* 160:90-101. https://doi.org/10.1016/j.forsciint.2006.04.009

Gill P, Kirkham A, Curran J (2007) *LoComatioN*: a software tool for the analysis of low copy number DNA profiles. *Forensic Science International* 166:128-138. https://doi.org/10.1016/j.forsciint.2006.04.016

Gill P, Brown RM, Fairley M, Lee L, Smyth M, Simpson N, Irwin B, Dunlop J, Greenhalgh M, Way K, Westacott EJ, Ferguson SJ, Ford LV, Clayton T, Guiness J (2008) National recommendations of the Technical UK DNA working group on mixture interpretation for the NDNAD and for court going purposes. *Forensic Science International: Genetics* 2(1):76-82. https://doi.org/10.1016/j.fsigen.2007.08.008

Gill P, Gusmao L, Haned H, Mayr WR, Morling N, Parson W, Prieto L, Prinz M, Schneider H, Schneider PM, Weir BS (2012) 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* 6(6):679-688. https://doi.org/10.1016/j.fsigen.2012.06.002

Gill P & Haned H (2013) A new methodological framework to interpret complex DNA profiles using likelihood ratios. *Forensic Science International: Genetics* 7(2):251-263. https://doi.org/10.1016/j.fsigen.2012.11.002

Gill P, Haned H, Bleka Ø, Hansson O, Dørum G, Egeland T (2015) Genotyping and interpretation of STR-DNA: Low-template, mixtures and database matches—Twenty years of research and development. *Forensic Science International: Genetics* 18:100-117. https://doi.org/10.1016/j.fsigen.2015.03.014

Gill P, Hicks T, Butler JM, Connolly E, Gusmão L, Kokshoorn B, Morling N, van Oorschot RAH, Parson W, Prinz M, Schneider PM, Sijen T, Taylor D (2018) DNA Commission of the International Society for

Forensic Genetics: Assessing the value of forensic biological evidence—Guidelines highlighting the importance of propositions. Part I: Evaluations of DNA profiling comparisons given (sub-) source propositions. *Forensic Science International: Genetics* 36:189-202. https://doi.org/10.1016/j.fsigen.2018.07.003

Gill, P. (2019) Interpretation continues to be the main weakness in criminal justice systems: developing roles of the expert witness and court. *WIREs Forensic Sci.* 1:e1321. https://doi.org/10.1002/wfs2.1321

Gill P, Hicks T, Butler JM, Connolly E, Gusmão L, Kokshoorn B, Morling N, van Oorschot RAH, Parson W, Prinz M, Schneider PM, Sijen T, Taylor D (2020) DNA Commission of the International Society for Forensic Genetics: Assessing the value of forensic biological evidence - guidelines highlighting the importance of propositions. Part II: Evaluation of biological traces considering activity level propositions. *Forensic Science International: Genetics* 44:102186. https://doi.org/10.1016/j.fsigen.2019.102186

Greenspoon SA, Schiermeier-Wood L, Jenkins BC (2015) Establishing the limits of TrueAllele® Casework: a validation study. *Journal of Forensic Sciences* 60(5):1263-1276. https://doi.org/10.1111/1556-4029.12810

Grgicak CM, Karkar S, Yearwood-Garcia X, Alfonse LE, Duffy KR, Lun DS (2020) A large-scale validation of NOCIt's a posteriori probability of the number of contributors and its integration into forensic interpretation pipelines. *Forensic Science International: Genetics* 47:102296. https://doi.org/10.1016/j.fsigen.2020.102296

Grgicak CM, Duffy KR, Lun DS (2021) The a posteriori probability of the number of contributors when conditioned on an assumed contributor. *Forensic Science International: Genetics* 54:102563. https://doi.org/10.1016/j.fsigen.2021.102563

Grgicak CM, Bhembe Q, Slooten K, Sheth NC, Duffy KR, Lun DS (2024) Single-cell investigative genetics: Single-cell data produces genotype distributions concentrated at the true genotype across all mixture complexities. *Forensic Science International: Genetics* 69:103000. https://doi.org/10.1016/j.fsigen.2023.103000

Haned H, Pène L, Lobry JR, Dufour AB, Pontier D (2011) Estimating the number of contributors to forensic DNA mixtures: does maximum likelihood perform better than maximum allele count? *Journal of Forensic Sciences* 56(1):23-28. https://doi.org/10.1111/j.1556-4029.2010.01550.x

Haned H, Slooten K, Gill P (2012) Exploratory data analysis for the interpretation of low template DNA mixtures. *Forensic Science International: Genetics* 6(6):762-774. https://doi.org/10.1016/j.fsigen.2012.08.008

Haned H, Gill P, Lohmueller K, Inman K, Rudin N (2016) Validation of probabilistic genotyping software for use in forensic DNA casework: Definitions and illustrations. *Science & Justice* 56(2):104-108. https://doi.org/10.1016/j.scijus.2015.11.007

Hares DR (2015) Selection and implementation of expanded CODIS core loci in the United States. *Forensic Science International: Genetics* 17:33-34. https://doi.org/10.1016/j.fsigen.2015.03.006

Hicklin RA, Richetelli N, Emerick BL, Bever RA, Davoren JM (2023) Variation in assessments of suitability and number of contributors for DNA mixtures. *Forensic Science International: Genetics* 65:102892. https://doi.org/10.1016/j.fsigen.2023.102892

Hicklin RA, Richetelli N, Bever RA, Davoren JM (2024) Inter-laboratory variation in interpretations, comparisons, and statistical analyses of DNA mixtures. (submitted)

Huffman K & Ballantyne J (2023a) Validation of probabilistic genotyping software for single cell STR analysis. *Genes* 14(3): 674. https://doi.org/10.3390/genes14030674

Huffman K & Ballantyne J (2023b) Single cell genomics applications in forensic science: Current state and future directions. *iScience* 26(11): 107961. https://doi.org/10.1016/j.isci.2023.107961

Huffman K, Kruijver M, Ballantyne J, Taylor D (2023) Carrying out common DNA donor analysis using DBLR™ on two or five-cell mini-mixture subsamples for improved discrimination power in complex DNA mixtures. *Forensic Science International: Genetics* 66:102908. https://doi.org/10.1016/j.fsigen.2023.102908

Inman K, Rudin N, Cheng K, Robinson C, Kirschner A, Inman-Semerau L, Lohmueller KE (2015) Lab Retriever: a software tool for calculating likelihood ratios incorporating a probability of drop-out for forensic DNA profiles. *BMC Bioinformatics* 16:298. https://doi.org/10.1186/s12859-015-0740-8

[IBC 1998] International Business Communications (1998). Resolution and Interpretation of Mixtures. July 31, 1998. Post-conference workshop taught by Peter Gill, James Curran, and Charles Brenner. Copy of workshop materials held by attendee John Butler.

ISHI 2010 workshop: https://strbase-archive.nist.gov/training/ISHI2010_MixtureWorkshop.htm (accessed July 1, 2024).

ISHI 2011 workshop: https://strbase-archive.nist.gov/training/MixtureWorkshop-ISHI2011-no-lit.pdf (accessed July 1, 2024).

ISHI 2012 workshop: https://strbase-archive.nist.gov/training/ISHI2012-MixtureWorkshop-Outline.pdf (accessed July 1, 2024). **M**ixtures **U**sing *SOUND* **S**tatistics, **I**nterpretation & **C**onclusions (ISHI 2012 Workshop: Butler, Coble, Cotton, Grgicak, Word) [Introduction] [Overview] [Profile1] [Statistics] [Profile Enhancement] [Different Assumptions] [BU Website] [Complex Mixtures] [Probabilistic Genotyping] [Complexity Threshold] [Court] [Future]

ISO/IEC 17025:2017 (2017) General requirements for the competence of testing and calibration laboratories. Available at https://www.iso.org/ISO-IEC-17025-testing-and-calibration-laboratories.html (accessed July 1, 2024). *Section 7.2.2.1(e) mentions interlaboratory comparisons as a technique used for method validation and Section 7.7.2(b) encourages laboratories to monitor their performance by participation in interlaboratory comparisons.*

Jeffreys AJ, Wilson V, Thein SL (1985a) Hypervariable 'minisatellite' regions in human DNA. *Nature* 314:67-73. https://doi.org/10.1038/314067a0

Jeffreys AJ, Brookfield JF, Semeonoff, R (1985b) Positive identification of an immigration test-case using human DNA fingerprints. *Nature* 317:818-819. https://doi.org/10.1038/317818a0

Karkar S, Alfonse LE, Grgicak CM, Lun DS (2019) Statistical modeling of STR capillary electrophoresis signal. *BMC Bioinformatics* 20(Sup 16):584. https://doi.org/10.1186/s12859-019-3074-0

Kaye DH (2010) *The Double Helix and the Law of Evidence*. Harvard University Press: Cambridge, MA. *See pp. 200-209, where Bruce Weir was requested to produce statistical analyses of multiple mixtures from the O.J. Simpson case, but Judge Ito would not permit LRs to be used (Bruce Weir's testimony took place on June 22-23, 1995).*

Kelly H, Bright JA, Buckleton JS, Curran JM (2014) A comparison of statistical models for the analysis of complex forensic DNA profiles. *Science & Justice* 54(1):66-70. https://doi.org/10.1016/j.scijus.2013.07.003

Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods & Applications* 3:13-22. https://doi.org/10.1101/gr.3.1.13

Kirby LT (1990) *DNA Fingerprinting: An Introduction.* Stockton Press: New York.

Kline MC, Duewer DL, Newall P, Redman JW, Reeder DJ, Richard M (1997) Interlaboratory evaluation of STR triplex CTT. *Journal of Forensic Sciences* 42:897-906.

Kline MC, Duewer DL, Redman JW, Butler JM (2003) NIST mixed stain study 3: DNA quantitation accuracy and its influence on short tandem repeat multiplex signal intensity. *Analytical Chemistry* 75:2463-2469. https://doi.org/10.1021/ac026410i

Kline MC, Duewer DL, Redman JW, Butler JM (2005) Results from the NIST 2004 DNA Quantitation Study. *Journal of Forensic Sciences* 50(3):571-578.

Kruijver M, Kelly H, Cheng K, Lin MH, Morawitz J, Russell L, Buckleton J, Bright JA (2021) Estimating the number of contributors to a DNA profile using decision trees. *Forensic Science International: Genetics* 50:102407. https://doi.org/10.1016/j.fsigen.2020.102407

Kulhankova L, Montiel González D, Bindels E, Kling D, Kayser M, Mulugeta E (2023) Single-cell transcriptome sequencing allows genetic separation, characterization and identification of individuals in multi-person biological mixtures. *Communications Biology* 6(1):201. https://doi.org/10.1038/s42003-023-04557-z

Kulhankova L, Bindels E, Kayser M, Mulugeta E (2024) Deconvoluting multi-person biological mixtures and accurate characterization and identification of separated contributors using non-targeted single-cell DNA sequencing. *Forensic Science International: Genetics* 71:103030. https://doi.org/10.1016/j.fsigen.2024.103030

Lin MH, Lee SI, Zhang X, Russell L, Kelly H, Cheng K, Cooper S, Wivell R, Kerr Z, Morawitz J, Bright JA (2021) Developmental validation of FaSTR™ DNA: software for the analysis of forensic DNA profiles. *Forensic Science International: Reports* 3:100217. https://doi.org/10.1016/j.fsir.2021.100217

Lohmueller KE & Rudin N (2013) Calculating the weight of evidence in low-template forensic DNA casework. *Journal of Forensic Sciences* 58 Suppl 1:S243-S249. https://doi.org/10.1111/1556-4029.12017

Lowe A, Murray C, Whitaker J, Tully G, Gill P (2002) The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces. *Forensic Science International* 129:25-34. https://doi.org/10.1016/s0379-0738(02)00207-4

Mallinder B, Pope S, Thomson J, Beck LA, McDonald A, Ramsbottom D, Court DS, Vanhinsbergh D, Barber M, Evett I, Sullivan K, Whitaker J (2022) Interpretation and reporting of mixed DNA profiles by seven forensic laboratories in the UK and Ireland. *Forensic Science International: Genetics* 58: 102674. https://doi.org/10.1016/j.fsigen.2022.102674

Marciano MA & Adelman JD (2017) PACE: Probabilistic Assessment for Contributor Estimation- A machine learning-based assessment of the number of contributors in DNA mixtures. *Forensic Science International: Genetics* 27:82-91. https://doi.org/10.1016/j.fsigen.2016.11.006

Marciano MA, Williamson VR, Adelman JD (2018) A hybrid approach to increase the informedness of CE-based data using locus-specific thresholding and machine learning. *Forensic Science International: Genetics* 35:26-37. https://doi.org/10.1016/j.fsigen.2018.03.017

Marciano MA & Adelman JD (2019) Developmental validation of PACE™: Automated artifact identification and contributor estimation for use with GlobalFiler™ and PowerPlex® Fusion 6c generated data. *Forensic Science International: Genetics* 43:102140. https://doi.org/10.1016/j.fsigen.2019.102140

McGovern C, Cheng K, Kelly H, Ciecko A, Taylor D, Buckleton JS, Bright JA (2020) Performance of a method for weighting a range in the number of contributors in probabilistic genotyping. *Forensic Science International: Genetics* 48:102352. https://doi.org/10.1016/j.fsigen.2020.102352

Meakin G & Jamieson A (2013) DNA transfer: Review and implications for casework. *Forensic Science International: Genetics* 7:434-443. https://doi.org/10.1016/j.fsigen.2013.03.013

Mitchell AA, Tamariz J, O'Connell K, Ducasse N, Prinz M, Caragine T (2011) Likelihood ratio statistics for DNA mixtures allowing for drop-out and drop-in. *Forensic Science International: Genetics Supplement Series* 3:e240-e241. https://doi.org/10.1016/j.fsigss.2011.08.119

Mitchell AA, Tamariz J, O'Connell K, Ducasse N, Budimlija Z, Prinz M, Caragine T (2012) Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in. *Forensic Science International: Genetics* 6(6):749-761. https://doi.org/10.1016/j.fsigen.2012.08.007

Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B (2001a) 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(3):647-660.

Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Brown AL, Budowle B (2001b) Validation of STR typing by capillary electrophoresis. *Journal of Forensic Sciences* 46(3):661-676.

Moretti TR, Just RS, Kehl SC, Willis LE, Buckleton JS, Bright JA, Taylor DA, Onorato AJ (2017) Internal validation of STRmix for the interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics* 29:126-144. https://doi.org/10.1016/j.fsigen.2017.04.004

Morling N, Bastisch I, Gill P, Schneider PM (2007) Interpretation of DNA mixtures--European consensus on principles. *Forensic Science International: Genetics* 1(3-4):291-292. https://doi.org/10.1016/j.fsigen.2007.06.007

Mudd JL, Baechtel FS, Duewer DL, Currie LA, Reeder DJ, Leigh SD, Liu H-K (1994) Interlaboratory comparison of autoradiographic DNA profiling measurements. 1. Data and summary statistics. *Analytical Chemistry* 66:3303-3317. https://doi.org/10.1021/ac00092a005

Mulero JJ, Chang CW, Lagacé RE, Wang DY, Bas JL, McMahon TP, Hennessy LK (2008) Development and validation of the AmpFlSTR MiniFiler PCR Amplification Kit: a MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA. *Journal of Forensic Sciences* 53(4): 838-852. https://doi.org/10.1111/j.1556-4029.2008.00760.x

[NRC 1992] National Research Council (1992) *DNA Technology in Forensic Science.* The National Academy Press: Washington DC. (sometimes referred to as “NRC I”)

[NRC 1996] National Research Council (1996) *The Evaluation of Forensic DNA Evidence*. The National Academy Press: Washington DC. (sometimes referred to as “NRC II”)

[NISTIR 8351] Butler JM, Iyer H, Press R, Taylor MK, Vallone PM, Willis S (2024) DNA Mixture Interpretation: A NIST Scientific Foundation Review. (National Institute of Standards and Technology, Gaithersburg, MD), NISTIR 8351. https://doi.org/10.6028/NIST.IR.8351

[NISTIR 8351sup2] Butler, J.M. (2024) Summarized Information from Publicly Accessible Validation Data and Proficiency Testing Results Covering DNA Mixture Interpretation. *Supplemental Document to DNA Mixture Interpretation: A NIST Scientific Foundation Review.* (National Institute of Standards and Technology, Gaithersburg, MD), NIST IR 8351sup2. https://doi.org/10.6028/NIST.IR.8351sup2

Norsworthy S, Lun DS, Grgicak CM (2018) Determining the number of contributors to DNA mixtures in the low-template regime: Exploring the impacts of sampling and detection effects. *Legal Medicine* 32:1-8. https://doi.org/10.1016/j.legalmed.2018.02.001

Paoletti DR, Doom TE, Krane CM, Raymer ML, Krane DE (2005) Empirical analysis of the STR profiles resulting from conceptual mixtures. *Journal of Forensic Sciences* 50(6):1361-1366.

Perez J, Mitchell AA, Ducasse N, Tamariz J, Caragine T (2011) Estimating the number of contributors to two-, three-, and four-person mixtures containing DNA in high template and low template amounts. *Croatian Medical Journal* 52(3):314-326. https://doi.org/10.3325/cmj.2011.52.314

Perlin MW, Lancia G, Ng SK (1995) Toward fully automated genotyping: genotyping microsatellite markers by deconvolution. *American Journal of Human Genetics* 57(5):1199-1210.

Perlin MW & Szabady B (2001) Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. *Journal of Forensic Sciences* 46(6):1372-1378.

Perlin MW (2006) Scientific validation of mixture interpretation methods. *Proceedings of the Seventeenth International Symposium on Human Identification*. Available at https://www.promega.com/products/pm/genetic-identity/ishi-conference-proceedings/17th-ishi-oral-presentations/ (accessed July 1, 2024).

Perlin MW, Kadane JB, Cotton RW (2009) Match likelihood ratio for uncertain genotypes. *Law, Probability and Risk* 8(3):289-302. https://doi.org/10.1093/lpr/mgp024

Perlin MW & Sinelnikov A (2009) An information gap in DNA evidence interpretation. *PLoS ONE* 4(12):e8327. https://doi.org/10.1371/journal.pone.0008327

Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW (2011) Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences* 56(6):1430-1447. https://doi.org/10.1111/j.1556-4029.2011.01859.x

Perlin MW, Belrose JL, Duceman BW (2013) New York State TrueAllele® casework validation study. *Journal of Forensic Sciences* 58(6):1458-1466. https://doi.org/10.1111/1556-4029.12223

Perlin MW, Dormer K, Hornyak J, Schiermeier-Wood L, Greenspoon S (2014) TrueAllele casework on Virginia DNA mixture evidence: computer and manual interpretation in 72 reported criminal cases. *PLoS ONE* 9(3):e92837. https://doi.org/10.1371/journal.pone.0092837

Perlin MW, Hornyak JM, Sugimoto G, Miller KW (2015) TrueAllele® genotype identification on DNA mixtures containing up to five unknown contributors. *Journal of Forensic Sciences* 60(4):857-868. https://doi.org/10.1111/1556-4029.12788

Perlin MW (2017) Method and system for DNA mixture analysis. United States Patent 9,708,642 B2 (July 18, 2017).

[PCAST 2016] President's Council of Advisors on Science and Technology. Report to the President - Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. September 2016. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf (accessed July 1, 2024).

[PCAST 2017] President's Council of Advisors on Science and Technology. *An Addendum to the PCAST Report on Forensic Science in Criminal Courts*. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_addendum_finalv2.pdf (accessed July 1, 2024).

Press R (2020) Two New Forensic DNA Standards Added to the OSAC Registry: National forensic science organization approves standards for interpreting DNA mixtures. NIST press release. Available at https://www.nist.gov/news-events/news/2020/05/two-new-forensic-dna-standards-added-osac-registry (accessed July 1, 2024).

Prieto L, Haned H, Mosquera A, Crespillo M, Aleman M, Aler M, Alvarez F, Baeza-Richer C, Dominguez A, Doutremepuich C, Farfán MJ, Fenger-Grøn M, García-Ganivet JM, González-Moya E, Hombreiro L, Lareu MV, Martínez-Jarreta B, Merigioli S, Milans Del Bosch P, Morling N, Muñoz-Nieto M, Ortega-González E, Pedrosa S, Pérez R, Solís C, Yurrebaso I, Gill P (2014) Euroforgen-NoE collaborative exercise on LRmix to demonstrate standardization of the interpretation of complex DNA profiles. *Forensic Science International: Genetics* 9:47-54. https://doi.org/10.1016/j.fsigen.2013.10.011

Puch-Solis R & Clayton T (2014) Evidential evaluation of DNA profiles using a discrete statistical model implemented in the DNA LiRa software. *Forensic Science International: Genetics* 11:220-228. https://doi.org/10.1016/j.fsigen.2014.04.005

[QAS 1998] Federal Bureau of Investigation (2000). Quality assurance standards for forensic DNA testing laboratories. *Forensic Science Communications*, *2*(3). Available at https://www2.fbi.gov/hq/lab/fsc/backissu/july2000/codispre.htm or https://www2.fbi.gov/hq/lab/fsc/backissu/july2000/codis2a.htm (accessed July 1, 2024).

[QAS 1999] Federal Bureau of Investigation (2000). Quality assurance standards for convicted offender DNA databasing laboratories. *Forensic Science Communications*, *2*(3). Available at https://www2.fbi.gov/hq/lab/fsc/backissu/july2000/codis1a.htm (accessed July 1, 2024).

[QAS 2009] Federal Bureau of Investigation (2009) Quality Assurance Standards (QAS) for Forensic DNA Testing Laboratories. Available at https://strbase-archive.nist.gov/QAS/Final-FBI-Director-Forensic-Standards.pdf (accessed July 1, 2024).

[QAS 2011] Federal Bureau of Investigation (2011) Quality Assurance Standards (QAS) for Forensic DNA Testing Laboratories. Available at https://ucr.fbi.gov/lab/biometric-analysis/codis/quality-assurance-standards-for-forensic-dna-testing-laboratories (accessed July 1, 2024).

[QAS 2020] Federal Bureau of Investigation (2020) Quality Assurance Standards (QAS) for Forensic DNA Testing Laboratories. Available at https://www.swgdam.org/publications (accessed July 1, 2024).

Rogers E, Aranda R, Spencer PM, Myers DR (2022) DNA mixture study: Novel metrics to quantify the intra- and inter-laboratory variability in forensic DNA mixture interpretation. National Institute of Justice report on award 2013-DNR-5042. Available at https://www.ojp.gov/pdffiles1/nij/grants/304317.pdf (accessed July 2, 2024).

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350-1354. https://doi.org/10.1126/science.2999980

Saiki RK, Walsh PS, Levenson CH, Erlich HA (1989) Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proceedings of the National Academy of Sciences of the United States of America* 86(16):6230-6234. https://doi.org/10.1073/pnas.86.16.6230

Schneider PM, Gill P, Carracedo A (2006a) Editorial on the recommendations of the DNA commission of the ISFG on the interpretation of mixtures. *Forensic Science International* 160:89. https://doi.org/10.1016/j.forsciint.2006.05.036

Schneider PM, Fimmers R, Keil W, Molsberger G, Patzelt D, Pflug W, Rothämel T, Schmitter H, Schneider H, Brinkmann B (2006b) Allgemeine Empfehlungen der Spurenkommission zur Bewertung von DNA-Mischspuren. *Rechtsmedizin* 16: 401-404. [German version was published in 2009 in English] https://doi.org/10.1007/s00194-006-0411-1

Schneider PM, Fimmers R, Keil W, Molsberger G, Patzelt D, Pflug W, Rothämel T, Schmitter H, Schneider H, Brinkmann B (2009) The German Stain Commission: recommendations for the interpretation of mixed stains. *International Journal of Legal Medicine* 123(1):1-5. [originally published in German in 2006 – *Rechtsmedizin* 16: 401-404] https://doi.org/10.1007/s00414-008-0244-4

Schneider PM (2009) Expansion of the European Standard Set of DNA database loci – the current situation. *Profiles in DNA* 12(1):6-7. https://www.promega.com/resources/profiles-in-dna/2009/expansion-of-the-european-standard-set/ (accessed July 2, 2024)

Schulte J, Marciano MA, Scheurer E, Schulz I (2023) A systematic approach to improve downstream single-cell analysis for the DEPArray™ technology. *Journal of Forensic Sciences* 68(6):1875-1893. https://doi.org/10.1111/1556-4029.15344

Schulte J, Caliebe A, Marciano M, Neuschwander P, Seiberle I, Scheurer E, Schulz I (2024) DEPArray™ single-cell technology: A validation study for forensic applications. *Forensic Science International: Genetics* 70:103026. https://doi.org/10.1016/j.fsigen.2024.103026

Steele CD & Balding DJ (2014) Statistical evaluation of forensic DNA profile evidence. *Annual Review of Statistics and Its Application* 1:361-384. https://doi.org/10.1146/annurev-statistics-022513-115602

Stolorow AM, Duewer DL, Reeder DJ, Buel E, Herrin G Jr (1996) Interlaboratory comparison of autoradiographic DNA profiling measurements. 3. Repeatability and reproducibility of RFLP band sizing, particularly bands of molecular size > 10k base pairs. *Analytical Chemistry* 68:1941-1947. https://doi.org/10.1021/ac951138g

Storey JD & Weir BS (1998) *DNAMIX* software. Available at http://genomics.princeton.edu/storeylab/dnamix/index.html (accessed April 3, 2023).

Stringer P, Scheffer JW, Scott P, Lee J, Goetz R, Ientile V, Eckhoff C, Turbett G, Carroll D, Harbison S (2009) Interpretation of DNA mixtures—Australian and New Zealand consensus on principles. *Forensic Science International: Genetics* 3(2):144-145. https://doi.org/10.1016/j.fsigen.2008.09.003

Swaminathan H, Grgicak CM, Medard M, Lun DS (2015) NOCIt: a computational method to infer the number of contributors to DNA samples analyzed by STR genotyping. *Forensic Science International: Genetics* 16:172-180. https://doi.org/10.1016/j.fsigen.2014.11.010

SWGDAM (2000) Short Tandem Repeat (STR) Interpretation Guidelines. *Forensic Science Communications* 2(3). Available at https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2000/strig.htm (accessed July 1, 2024).

SWGDAM (2004) Revised Validation Guidelines. *Forensic Science Communications* 6(3). Available at https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2004/standards/2004_03_standards02.htm (accessed July 1, 2024).

SWGDAM (2010) SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories.

SWGDAM (2012) SWGDAM Validation Guidelines for DNA Analysis Methods.

SWGDAM (2014) SWGDAM Guidelines for STR Enhanced Detection Methods. Available at https://www.swgdam.org/publications (accessed August 8, 2024).

SWGDAM (2015) SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems. Available at https://www.swgdam.org/publications (accessed July 1, 2024).

SWGDAM (2016) SWGDAM Validation Guidelines for DNA Analysis Methods. Available at https://www.swgdam.org/publications (accessed July 1, 2024).

SWGDAM (2017) SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Available at https://www.swgdam.org/publications (accessed July 1, 2024).

SWGDAM (2018) Recommendations of the SWGDAM Ad Hoc Working Group on Genotyping Results Reported as Likelihood Ratios. Available at https://www.swgdam.org/publications (accessed July 1, 2024).

Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* 24(16):3189-3194. https://doi.org/10.1093/nar/24.16.3189

Taylor D, Bright JA, Buckleton J (2013) The interpretation of single source and mixed DNA profiles. *Forensic Science International: Genetics* 7(5):516-528. https://doi.org/10.1016/j.fsigen.2013.05.011

Taylor D (2014) Using continuous DNA interpretation methods to revisit likelihood ratio behaviour. *Forensic Science International: Genetics* 11:144-153. https://doi.org/10.1016/j.fsigen.2014.03.008

Taylor D, Bright JA, Buckleton J (2014) The 'factor of two' issue in mixed DNA profiles. *Journal of Theoretical Biology* 363:300-306. https://doi.org/10.1016/j.jtbi.2014.08.021

Taylor D & Buckleton J (2015) Do low template DNA profiles have useful quantitative data? *Forensic Science International: Genetics* 16:13-16. https://doi.org/10.1016/j.fsigen.2014.11.001

Taylor D, Buckleton J, Evett I (2015) Testing likelihood ratios produced from complex DNA profiles. *Forensic Science International: Genetics* 16:165-171. https://doi.org/10.1016/j.fsigen.2015.01.008

Taylor D & Powers D (2016) Teaching artificial intelligence to read electropherograms. *Forensic Science International: Genetics* 25:10-18. https://doi.org/10.1016/j.fsigen.2016.07.013

Taylor D, Harrison A, Powers D (2017) An artificial neural network system to identify alleles in reference electropherograms. *Forensic Science International: Genetics* 30:114-126. https://doi.org/10.1016/j.fsigen.2017.07.002

Taylor D, Kokshoorn B, Biedermann A (2018) Evaluation of forensic genetics findings given activity level propositions: A review. *Forensic Science International: Genetics* 36:34-49. https://doi.org/10.1016/j.fsigen.2018.06.001

Taylor D, Kitselaar M, Powers D (2019) The generalisability of artificial neural networks used to classify electrophoretic data produced under different conditions. *Forensic Science International: Genetics* 38:181-184. https://doi.org/10.1016/j.fsigen.2018.10.019

Taylor D (2022) Using a multi-head, convolutional neural network with data augmentation to improve electropherogram classification performance. *Forensic Science International: Genetics* 56:102605. https://doi.org/10.1016/j.fsigen.2021.102605

Taylor D & Buckleton J (2023) Combining artificial neural network classification with fully continuous probabilistic genotyping to remove the need for an analytical threshold and electropherogram reading. *Forensic Science International: Genetics* 62:102787. https://doi.org/10.1016/j.fsigen.2022.102787

Taylor D & Abarno D (2023). A lights-out forensic DNA analysis workflow for no-suspect crime. *Forensic Science International Genetics* 66:102907. https://doi.org/10.1016/j.fsigen.2023.102907

Thomson J (2018) AFSP Mixture Evaluation Collaborative Exercise. Presentation given at the English Speaking Working Group of the International Society for Forensic Genetics held September 2018 in St. Petersburg, Russia. Available at http://www.isfg.org/files/ESWG2018_Thomson_AFSP_Mixture.pdf (accessed August 8, 2024).

Torres Y, Flores I, Prieto V, López-Soto M, Farfán MJ, Carracedo A, Sanz P (2003) DNA mixtures in forensic casework: a 4-year retrospective study. *Forensic Science International* 134(2-3):180-186. https://doi.org/10.1016/s0379-0738(03)00161-0

Toscanini U, Gusmao L, Alava Narvaez MC, Alvarez JC, Baldassarri L, Barbaro A, Berardi G, Betancor HE, Camargo M, Carreras-Carbonell J, Castro J, Costa SC, Coufalova P, Domínguez V, Fagundes de Carvalho E, Ferreira STG, Furfuro S, García O, Goios A, González R, de la Vega AG, Gorostiza A, Hernández A, Jiménez Moreno S, Lareu MV, León Almagro A, Marino M, Martínez G, Miozzo MC, Modesti NM, Onofri V, Pagano S, Pardo Arias B, Pedrosa S, Penacino GA, Pontes ML, Porto MJ, Puente-Prieto J, Pérez RR, Ribeiro T, Rodríguez Cardozo B, Rodríguez Lesmes YM, Sala A, Santiago B, Saragoni VG, Serrano A, Streitenberger ER, Torres Morales MA, Vannelli Rey SA, Velázquez Miranda M, Whittle MR, Fernández K, Salas A (2016) Analysis of uni and bi-parental markers in mixture samples: Lessons from the 22nd GHEP-ISFG Intercomparison Exercise. *Forensic Science International: Genetics* 25:63-72. https://doi.org/10.1016/j.fsigen.2016.07.010

[TWGDAM 1989] Technical Working Group on DNA Analysis Methods with James J. Kearney, chair (1989) Guidelines for a quality assurance program for DNA restriction fragment length polymorphism analysis. *Crime Laboratory Digest* 16(2):40-59.

[TWGDAM 1991] Technical Working Group on DNA Analysis Methods with James J. Kearney, chair, and California Association of Criminalistics Ad Hoc Committee on DNA Quality Assurance with Carol L. Hunter, chair (1991) Guidelines for a quality assurance program for DNA analysis. *Crime Laboratory*

*Digest* 18(2):44-75. Available at https://www.ojp.gov/pdffiles1/Digitization/132230NCJRS.pdf (accessed July 1, 2024)

[TWGDAM 1995] Technical Working Group on DNA Analysis Methods with Bruce Budowle, chair (1995) Guidelines for a quality assurance program for DNA analysis. *Crime Laboratory Digest* 22(2):20-43. Available at https://www.ojp.gov/pdffiles1/Digitization/153914NCJRS.pdf (accessed July 1, 2024).

[UKFSR 2014] UK Forensic Science Regulator (2014) Validation, FSR-G-201. Available at https://www.gov.uk/government/publications/forensic-science-providers-validation (accessed July 1, 2024).

[UKFSR 2017] UK Forensic Science Regulator (2017) Codes of Practice and Conduct for Forensic Science Providers and Practitioners in the Criminal Justice System. Available at https://www.gov.uk/government/collections/forensic-science-providers-codes-of-practice-and-conduct (accessed July 1, 2024).

[UKFSR 2018a] UK Forensic Science Regulator (2018a) DNA mixture interpretation, FSR-G-222. Available at https://www.gov.uk/government/publications/dna-mixture-interpretation-fsr-g-222 (accessed July 1, 2024).

[UKFSR 2018b] UK Forensic Science Regulator (2018b) Software validation for DNA mixture interpretation, FSR-G-223. Available at https://www.gov.uk/government/publications/software-validation-for-dna-mixture-interpretation-fsr-g-223 (accessed July 1, 2024).

Valtl J, Mönich UJ, Lun DS, Kelley J, Grgicak CM (2021) A series of developmental validation tests for Number of Contributors platforms: Exemplars using NOCIt and a neural network. *Forensic Science International: Genetics* 54:102556. https://doi.org/10.1016/j.fsigen.2021.102556

van Oorschot RA & Jones MK (1997) DNA fingerprints from fingerprints. *Nature* 387:767. https://doi.org/10.1038/42838

van Oorschot RAH, Szkuta B, Meakin GE, Kokshoorn B, Goray M (2019) DNA transfer in forensic science: A review. *Forensic Science International: Genetics* 38:140–166. https://doi.org/10.1016/j.fsigen.2018.10.014

van Oorschot RAH, Meakin GE, Kokshoorn B, Goray M, Szkuta B (2021) DNA transfer in forensic science: recent progress towards meeting challenges. *Genes* 12(11): 1766. https://doi.org/10.3390/genes12111766

Veldhuis MS, Ariëns S, Ypma RJF, Abeel T, Benschop CCG (2022) Explainable artificial intelligence in forensics: Realistic explanations for number of contributor predictions of DNA profiles. *Forensic Science International: Genetics* 56:102632. https://doi.org/10.1016/j.fsigen.2021.102632

Volgin L, Taylor D, Bright JA, Lin MH (2021) Validation of a neural network approach for STR typing to replace human reading. *Forensic Science International: Genetics* 55:102591. https://doi.org/10.1016/j.fsigen.2021.102591

Wallin JM, Buoncristiani MR, Lazaruk KD, Fildes N, Holt CL, Walsh PS (1998) TWGDAM validation of the AmpFlSTR Blue PCR amplification kit for forensic casework analysis. *Journal of Forensic Sciences* 43(4):854-870.

Walsh PS, Fildes N, Louie AS, Higuchi R (1991) Report of the blind trial of the Cetus Amplitype HLA DQ alpha forensic deoxyribonucleic acid (DNA) amplification and typing kit. *Journal of Forensic Sciences* 36(5):1551-1556.

Walsh PS, Erlich HA, Higuchi R (1992) Preferential PCR amplification of alleles: mechanisms and solutions. *PCR Methods & Applications* 1(4):241-250. https://doi.org/10.1101/gr.1.4.241

Wang T, Xue N, Birdwell JD (2006) Least-square deconvolution: a framework for interpreting short tandem repeat mixtures. *Journal of Forensic Sciences* 51(6):1284-1297. https://doi.org/10.1111/j.1556-4029.2006.00268.x

Weir BS (1995) DNA statistics in the Simpson matter. *Nature Genetics* 11:365-368. https://doi.org/10.1038/ng1295-365

Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KA, Buckleton J (1997) Interpreting DNA mixtures. *Journal of Forensic Sciences* 42(2):213-222.

Weir BS (2000) Court experiences in the USA: People v. Simpson. *First International Conference on Forensic Human Identification*. London, 23-26 October 1999.

Werrett DJ (1997) The National DNA Database. *Forensic Science International* 88:33-42. https://doi.org/10.1016/S0379-0738(97)00081-9

Whitaker JP, Cotton EA, Gill P (2001) A comparison of the characteristics of profiles produced with the AMPFlSTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Science International* 123(2-3):215-223. https://doi.org/10.1016/s0379-0738(01)00557-6

Wong Z, Wilson V, Patel I, Povey S, Jeffreys AJ (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. *Annals of Human Genetics* 51(4):269-288. https://doi.org/10.1111/j.1469-1809.1987.tb01062.x

Wyman AR & White R (1980) A highly polymorphic locus in human DNA. *Proceedings of the National Academy of Sciences of the United States of America* 77(11):6754-6758. https://doi.org/10.1073/pnas.77.11.6754

Young BA, Gettings KB, McCord B, Vallone PM (2019) Estimating number of contributors in massively parallel sequencing data of STR loci. *Forensic Science International: Genetics* 38:15-22. https://doi.org/10.1016/j.fsigen.2018.09.007