UNITED STATES DISTRICT COURT
WESTERN DISTRICT OF MICHIGAN
SOUTHERN DIVISION

UNITED STATES OF AMERICA,

      Plaintiff,

                                    Case No. 1:17-cr-130

v.

                                    HON. JANET T. NEFF

DANIEL GISSANTANER,

      Defendant.

_____/

OPINION

        The heart of this opinion for decisional purposes is found in Sections III, IV, and V. The preceding sections provide an essential scientific context for the decisional process. A Glossary of key terms is appended as Attachment 1 for reference.

The Case

Defendant Daniel Gissantaner is charged with a single offense of felon in possession of a firearm, subject to a penalty of not less than 15 years' and up to life imprisonment; not more than a $250,000 fine; and supervised release of not more than 5 years. The case against Gissantaner rests fundamentally, if not entirely, on a small amount of "touch" DNA taken from a gun in a locked cedar chest during a search of Gissantaner's house on September 25, 2015, following a dispute with his neighbors over parking in a shared driveway, and police officers' response to the neighbor's 911 call. The locked chest containing the gun belonged to Gissantaner's wife's daughter's boyfriend, Cory Patton. Patton, also a convicted felon, had the only key to the locked chest, which was located in Patton's upstairs bedroom and opened by him at police officers' request during the search. According to the various police reports, Patton gave conflicting statements about the gun, stating that he heard an argument, went outside, and took the gun away from Gissantaner, but also stating that he never saw Gissantaner with the gun, he found it on the kitchen counter after the argument, and then placed it in the chest.

The evidentiary handling of the gun is far from pristine. It appears that the gun was moved or handled by at least one police officer before it was taken into evidence. There are also some unexplained delays and unknown whereabouts of the gun between the time it was taken from Gissantaner's house and the submission to the Michigan State Police (MSP) lab for analysis.

Ultimately, a touch-DNA analysis by the MSP from a swab of the gun determined that three individuals contributed to the DNA found on the gun. The DNA analysis produced a report based on STRmix™[1] probabilistic genotyping software that Gissantaner was a 7% minor

---

[1] STRmix™ is properly referenced as a trademark, but is used herein both with and without such indication.

2

contributor of the DNA on the gun, and that it was at least 49 million times more likely that the DNA was that of Gissantaner and two unrelated, unknown individuals, than that the DNA was that of three unrelated, unknown contributors.[2]

Gissantaner filed a motion to exclude the DNA evidence, challenging the admission of the STRmix™ DNA report by the Government. The matter is before the Court for decision after the arduous task placed on the parties, counsel, the witnesses, and court-appointed experts to explain and examine the intricacies of DNA analysis generally, and the use of the STRmix™ probabilistic genotyping software specifically, in the context of the evidence in this case.

I. INTRODUCTION

This Court is not the first to grapple with the difficult question of the admissibility of probabilistic genotyping DNA evidence. While a number of courts have found the evidence admissible, under varying standards and circumstances, others have not. *See, e.g., People v. Collins*, 15 N.Y.S.3d 564, 566 (N.Y. Sup. Ct. 2015). This Court now falls into the latter category, and concludes that the probabilistic genotyping evidence in this case does not pass scrutiny under the *Daubert*[3] lens.

In the initial two-day *Daubert* hearing, May 23-24, 2018, the Court heard testimony from a number of well-informed expert and lay witnesses, including for the Government—Dr. John Buckleton, co-developer of STRmix™, ESR, Crown Research Institute, government of New Zealand; Jeffrey Nye, assistant director, Forensic Science Division, MSP; and Amber Smith, forensic scientist, MSP Lansing Laboratory Biology DNA Unit; and for the defense—Stephen

---

[2] The Court has used various expressions of the DNA Report results for ease of reference in this opinion, but the Court recognizes that the actual likelihood ratio requires a statement in precise terms, as set forth in the MSP Report.

[3] *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 589 (1993).

Lund, mathematical statistician, National Institute of Standards and Technology (NIST), U.S. Department of Commerce; and Nathan Adams, systems engineer, Forensic Bioinformatics.

Finding that the initial briefing, testimony and evidence lacked sufficient clarity and completeness to resolve the complex issues presented, the Court solicited recommendations from the parties for a court-appointed expert to examine the issues and provide an independent opinion. The Court ultimately appointed two experts, Dr. Michael Coble and Dr. Dan E. Krane, both well-recognized for their specialized expertise and contributions to the advancing field of probabilistic genotyping in forensic DNA analysis in the U.S. Following written reports from each court-appointed expert and a day of testimony from the experts in a continued *Daubert* hearing on July 8, 2019, the Court provided the parties a last opportunity for limited supplemental briefing. With the benefit of a continually supplemented record over a year-and-a-half's time, comes the disadvantage of new insights, arguments, publications, and clarifications amid a rapidly changing technology base. The Court now has the benefit of a more complete, but voluminous, record on which it proceeds to rule.

As the court in *Collins* aptly observed in addressing the admissibility of complex-mixture DNA analysis, "judges are, far and away, not the people best qualified to explain science," particularly when novel scientific techniques are at issue. *Collins*, 15 N.Y.S.3d at 566. Nor does the manner of admitting testimony and evidence in a *Daubert* hearing lend itself to a methodical explanation of DNA science. "But courts are bound to do their best." *Id*. Because a fundamental understanding of the science and technology is critical for the *Daubert* analysis in this case, the Court endeavors to set out a scientific and technological foundation in conjunction with its decision.

4

DNA Analysis—

"Deoxyribonucleic acid, or DNA, is a molecule that encodes the genetic information in all living organisms." 4 Mod. Sci. Evidence § 30:1 (2018-2019 ed.), The Law and Science of Expert Testimony, DNA Typing, *Introduction to basic principles* (footnotes omitted). Despite its universal application in crime investigation, the science of forensic DNA analysis is relatively young.

In late 1984, Geneticist Sir Alec Jeffreys developed a DNA profiling process, "DNA typing," while working in the Department of Genetics at the University of Leicester in the United Kingdom. 36 AM. JUR. PROOF OF FACTS 3D *Proof of Criminal Identity or Paternity Through Polymerase Chain Reaction (PCR) Testing* § 5, n.57 (Sept. 2019 update) (citing Jeffreys et al., *Individual Specific "Fingerprints" of Human DNA*, 316 NATURE 76 (1985)); 97 AM. JUR. PROOF OF FACTS 3D *Identification of Seminal Fluids* § 10 (Sept. 2019 update). The first widely noted use of DNA analysis in a criminal case occurred in Britain two years later in 1986, demonstrating the "unique power of DNA typing to exonerate, as well as incriminate," when Scotland Yard called upon Dr. Jeffreys to assist in the investigation of two brutal rape and strangulation cases:

> The murders occurred in two neighboring villages in Narborough, England. Police soon focused on a suspect, Richard Buckland, who provided a graphic confession after several hours of interrogation. In it, he described details of the crime that police proclaimed were only known to the killer.
>
> In order to solidify the case against Buckland, police submitted semen samples from both crimes to Jeffreys, who had developed a process he called "DNA fingerprinting," for analysis and comparison against Buckland's blood sample. Jeffrey's conclusion, which stunned the police and the community, was that Buckland was not the perpetrator. The DNA tests confirmed that both girls had been raped by the same perpetrator, but Buckland was not that man. Buckland became the first person in the world to be cleared through the use of DNA tests. When their prime suspect was excluded from consideration, police embarked upon a campaign of "voluntary" blood testing, obtaining samples from over 5000 men in the environs of the crime. The results of this first-reported DNA dragnet did not identify the rapist. However, it did lead the police to Colin Pitchfork. A coworker revealed that Pitchfork had persuaded him to provide a sample in his stead. The

ruse was eventually uncovered and Pitchfork was arrested in 1987. After his arrest, Pitchfork confessed to the crimes and subsequent DNA tests linked him to the crimes.

Robert Aronson & Jacqueline McMurtrie, *The Use and Misuse of High-Tech Evidence by Prosecutors: Ethical and Evidentiary Issues*, 76 FORDHAM L. REV. 1453, 1473 n.123 (2007) (citing HENRY C. LEE & FRANK TIRNADY, BLOOD EVIDENCE: HOW DNA IS REVOLUTIONIZING THE WAY WE SOLVE CRIMES, 1-2 (2003) and JOSEPH WAMBAUGH, THE BLOODING (1989)).

Why DNA Analysis Works—

Although 99.9% of the DNA sequences in human cells are the same between any two individuals, enough of the DNA is different that it is possible to distinguish one individual from another (other than identical twins). David H. Kaye & George Sansabaugh, *Reference Guide on DNA Identification Evidence, Reference Manual on Scientific Evidence*, 136-137 (Federal Judicial Center, 3d ed. 2011). This remaining 0.1% variation makes each person genetically unique. *Id.* at 137.

Modern DNA profiling uses repetitive sequences that are highly variable, called short tandem repeats (STRs), at specific locations, "loci," of the human genome. In general terms, by identifying the STRs at agreed upon "loci," forensic scientists are able to develop DNA profiles for forensic comparison. The DNA profile is comprised of the particular sequences that an individual has at each locus, each of which is called an "allele." STRs are useful features for comparison because while every person has STRs at the loci, there is variation in the number of repeats in a given STR for each person (that is, different people can have different alleles), and the range of variation is known by population studies (Govt Br., ECF No. 52 at PageID.1742; citation omitted).

By analyzing a sufficiently large number of loci, a unique DNA profile is determined, such that it is highly improbable that any two people who are not identical twins would have the exact

same DNA profile (Govt Br., ECF No. 52 at PageID.1742). Each forensic lab determines the

number of loci targeted for analysis. The FBI at one point used thirteen core loci, and in 2017

increased that number to twenty (*id*., citations omitted). The MSP laboratory generally analyzes

24 loci (*id*., citation omitted).

DNA profiling is perhaps the greatest advancement of the 20th century in the criminal

justice system. It has freed the innocent, corralled the guilty, and given closure to decades long

cold cases. But—it is not a surefire solution to all crime investigation.

Why DNA Analysis Doesn't Work—

"In the thirty years since its debut, DNA has assumed an ever-increasing role in criminal

cases, and forensic DNA databases now flourish." 4 Mod. Sci. Evidence § 30:1. "No other

scientific technique has gained such widespread acceptance so quickly. No other technique is as

complex or so subject to rapid change." PAUL C. GIANNELLI, The DNA Story: An Alternative

View, 88 J. Crim. L. & Criminology 380, 381 (1997) (footnote omitted).

During this time there have been improvements, refinements and ongoing review of the

analytical processes involved in criminal DNA analysis. While this has undoubtedly led to

important insights and advancements, as with any scientific process, new methods have revealed

the shortcomings of the old. And objections remain with respect to current DNA analysis:

> Although the basic scientific principles behind PCR amplification and capillary electrophoreses are well established, there remain aspects of interpretation that might still be objectionable. Many lay persons believe that DNA testing methods produce unambiguous, or mathematically precise, results, but the truth—especially with regard to often problematic or perplexing crime scene samples—can be far from the case. In general, examiners employ rules of thumb to help resolve these ambiguities, but admissibility issues might nonetheless arise in one of two ways.

> First, if one of these rules of thumb proved scientifically unsound, then exclusion would be appropriate on Rule 702 grounds. …

* * *

7

> Secondly, even if interpretative rules meet the Rule 702 standard, it might still in some cases be appropriate to exclude evidence in a case in which bona fide disagreements as to the proper interpretation of ambiguous results might arise. …

4 Mod. Sci. Evidence § 30:11, Current objections to DNA admissibility—Science—General subjectivity (2018-2019 ed.) (footnote omitted).

II. DNA ANALYTICAL PROCESS

For the uninitiated, forensic DNA analysis is a complex scientific process. Couple that with the complexity of the mathematical theories, algorithms and likelihood statistics used in probabilistic genotyping software, and a full explanation would far exceed the bounds of this legal opinion. The Court will strive for a happy medium.

"The usual objective of forensic DNA analysis is to detect variations in the genetic material that differentiate one individual from another. But 'forensic DNA typing' is not a single scientific process. The term encompasses different kinds of testing methods, at times using different sources of bodily material, and may also refer to differing statistical means of assessing the significance of a match." 4 Mod. Sci. Evidence § 30:1 (footnote omitted).

DNA testing in the United States is generally done using commercial test kits that examine specific loci on the human genome where there are "Short Tandem Repeats, or STRs, which are genetic markers that contain short repeated sequences of DNA base pairs" (Def. Br., ECF No. 41 at PageID.749; Michigan State Police Biology Procedures and Training Manuals (MSP PMBIO) § 2.11.1). *See also* William C. Thompson, Laurence D. Mueller, & Dan E. Krane, *Forensic DNA Statistics: Still Controversial in Some Cases*, The Champion, Dec. 2012, p. 13. As noted, the region at which a particular STR is found is called a "locus" (Govt Br., ECF No. 52 at PageID.1742). The number of times that a particular sequence repeats itself at a locus varies from person to person, such that STRs "represent a good source to differentiate individuals" (ECF No.

8

41 at PageID.749-750, citing MSP PMBIO § 2.11.1). The specific number of repeats is called an "allele" (ECF No. 52 at PageID.1742). A person will have two "alleles" at each STR, one inherited from each parent. *See* William C. Thompson, et al., *Forensic DNA Statistics: Still Controversial in Some Cases*, p. 13.

"If two DNA samples are from the same person, they will have the same alleles at each locus examined; if two samples are from different people, they will almost always have different alleles at some of the loci. The goal of forensic DNA testing is to detect the alleles present at the tested loci in evidentiary samples so that they can be compared with the alleles detected at the same loci in reference samples from possible contributors." *Id*. (footnote omitted).

A. PCR and Electrophoresis

The process for obtaining a DNA profile begins with taking a swab, which then is subjected to a chemical process that extracts and amplifies or replicates the DNA at specific loci, and ultimately outputs a chart of the alleles present in the form of a graph for interpretation by the DNA analyst.

STRs can be detected through a process using polymerase chain reaction ("PCR") and an analytical technique called capillary electrophoresis (*see* MSP PMBIO § 2.11.1).   Briefly described, during PCR, the DNA sample is copied, or "amplified," utilizing commercially-produced fluorescent primers.   The genetic material then is passed through a capillary electrophoresis instrument, which "separates the DNA fragments by size," allowing for allele identification.   The number of times that a particular sequence repeats at a particular site corresponds to an "allele."   This process, "electrophoresis," generates a chart called an electropherogram (EPG or e-gram) of a series of peaks, wherein the peaks are proportionate to the amount of DNA present or the length of the STR.

9

"Either an analyst, or the [DNA testing] software, can 'call' peaks to differentiate signal from noise. Noise, by way of example, can come from artifacts of the PCR process that result in small peaks not indicative of actual alleles" (*see* Govt Br., ECF No. 52 at PageID.1743 n.3).

In the case of a single DNA profile, an analyst would expect to see either one or two signal peaks at each locus. Where, however, three or more identified peaks appear at a locus, the analyst knows that the profile usually contains a mixture of DNA. When the DNA profile is found to contain a mixture of more than one contributor, probabilistic genotyping software is used to further analyze the DNA sample.

B. Probabilistic Genotyping

Probabilistic genotyping software (PGS) is the most recent purported advancement in forensic DNA analysis. Probabilistic genotyping refers to "the use of biological modeling, statistical theory, computer algorithms, and probability distributions to calculate likelihood ratios (LRs) and/or infer genotypes for the DNA typing results of forensic samples ('forensic DNA typing results')." SWGDAM, Guidelines for the Validation of Probabilistic Genotyping Systems, https://docs.wixstatic.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf   (last   visited 10/14/19).

A probabilistic approach to DNA interpretation was first publicized around 2000, with the first software program created around 2006 (ECF No. 152 at PageID.4087). At about the same time, in 2006–2008, a shift occurred in the kinds of cases that were being submitted to laboratories (*id*. at PageID.4088). In 2006–2007, mixtures were predominantly two-person mixtures, high quality, high quantity blood or sexual assault evidence (*id*.). Beginning in 2004–2006, there was a recognition that DNA could be obtained from materials that had been touched, such as steering wheels, cell phones and guns (*id.* at PageID.4088-4089). The labs began receiving more and more

complex mixtures of two, three, four or more contributors (*id.* at PageID.4089). These converging

circumstances, in a little more than a decade, have led to PGS becoming the anointed "crown

jewel" of DNA analysis.

> As the sensitivity of forensic DNA typing procedures has improved with the development of better DNA extraction and amplification chemistries and detection instrumentation, more DNA profiles originating from the DNA of two or more individuals are being encountered in forensic casework. The complexity of profile interpretation increases with each additional contributor to a mixture, particularly if the DNA contribution is low and therefore subject to stochastic effects (e.g., allele dropout and greater heterozygous peak height variance). …

> * * *

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

(ECF No. 52-4 at PageID.1831-1832; footnotes omitted).[4]

---

[4] *See* Tamyra R. Moretti, Rebecca S. Just, Susannah C. Kehl, Leah E. Willis, John S. Buckleton, Jo-Anne Bright, Duncan A. Taylor, Anthony J. Onorato, Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles (accepted manuscript, 3/4/17), Forensic Science International: Genetics.

In order to analyze complex mixtures of DNA, with three or more contributors, software is necessary, since it performs hundreds of thousands of calculations, which could not conceivably be performed by hand.

C. STRmix™ Software

STRmix™ is one of the leading probabilistic genotyping software programs in use in the United States. The software uses the electropherogram results, along with inputs from the lab analyst, to generate a statistical estimate in the form of a likelihood ratio (LR) to communicate the laboratory's assessment of how strongly forensic evidence can be tied to a suspect (*see* ECF No. 41 at PageID.751, citing MSP STRMix Validation Summary, ECF No. 41-14 at PageID.1018 and ECF No. 41-15, Article, National Institute of Standards and Technology, "NIST Experts Urge Caution in Use of Courtroom Evidence Presentation Method," October 12, 2017).

"The *LR* considers the probability of obtaining the evidence profile(s) given two competing propositions, usually aligned with the prosecution case and defence case." Jo-Anne Bright, Duncan Taylor, Catherine McGovern, Stuart Cooper, Laura Russell, Damien Abarno, John Buckleton, *Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles*, Forensic Science International: Genetics 23 (2016) 226-239 (FSI 23), ECF No. 41-16 at PageID.1068). In essence, an LR represents the likelihood of whether a particular person's DNA is present in a mixture, as compared to a random person's DNA, based on standard population reference databases (ECF No. 41 at PageID.752). In short:

> STRmix™ is software that employs a continuous model for DNA profile interpretation and genotype determination based on a Markov Chain Monte Carlo (MCMC) sampling method. Using weights assigned to the resultant genotypes or genotype sets, STRmix™ calculates *LRs*, which are the probability of the DNA evidence under two opposing hypotheses referred to as *H1* and *H2*. The terms *H1* and *H2* are used in lieu of "Prosecution hypothesis" (*Hp*) and "Defense hypothesis" (*Hd*), respectively, given that they are assigned by the scientist, usually without consultation with legal representatives.

> A LR greater than 1 provides support for a specified person of interest [POI] as a contributor to the DNA evidence (*H1*), whereas an *LR* less than 1 provides support that the person of interest is not a contributor (*H2*). An *LR* of 1 provides no greater support for either proposition. …

(ECF No. 52-4 at PageID.1832). However, due to the methodologies used in STRMix, "[t]he results of no two analyses will be completely the same" (FSI 23, ECF No. 41-16 at PageID.1075).

STRmix™ is a proprietary software, created by the Institute of Environmental Science and Research (ESR), located in New Zealand. It is offered for commercial sale in the United States. The software was co-developed by Dr. Buckleton, who testified at the *Daubert* hearing. The genesis of probabilistic genotyping dates back to a mathematical solution Dr. Buckleton developed in 1999, making him the father of probabilistic genotyping (5/23/19 Hrg. Tr., ECF No. 77 at PageID.2522-2523). According to Dr. Buckleton, STRmix software was begun in May 2011, although likelihood ratios have been used in forensic science since the 1940s, even back to the 1910s (*id.* at PageID.2513-2514). Dr. Buckleton stated that the mathematics and the majority of modern probabilistic genotyping software either comes directly from him or dates back to work he has done (*id.* at PageID.2523).

According to Dr. Buckleton, although newer versions of STRmix have been continuously released, the engine of STRmix was in place very early, the core Metropolis-Hastings algorithm and the Monte Carlo Markov Chain, and has changed very little during that time (ECF No. 77 at PageID.2529). The Monte Carlo and Markov Chain processes were synergized in the '50s to '70s to become Monte Carlo Markov Chain (also called the Markov Chain Monte Carlo) (*id.* at PageID.2529-2530). It has become a dominant and mainstream methodology for solving this type of complex problem from the '70s onwards, and is applied throughout many fields such as physics, engineering, geoscience, medicine, and a great many others (*id.* at PageID.2530).

As described by Dr. Buckleton, probabilistic genotyping works roughly like a game of hot and cold, where as children, you may have taken a step and your parents said whether you were hotter or colder, and using that feedback you would try and find your way to whatever your goal was. And a Markov Chain is quite similar to this (*id.*). This process is used to tease apart the mixed DNA data of two or more contributors and essentially make a list of plausible single source genotypes that may have contributed to that mixture (*id.*).

STRmix is in two completely separate parts: the first part does what is described above, which is termed "deconvoluting" a mixture; the second part does the relatively trivial action of assembling the likelihood ratio (*id.*). The inputs into STRmix are a set of data that derives directly from the analysis of the electropherogram, often with GeneMaker or GeneMarker, which are two commercial softwares for processing e-grams (*id.* at PageID.2530-2531). The data for STRmix consists of the set of allele names and peak heights (*id.* at PageID.2531).

Using the quantitative information from the electropherogram, STRmix™ "calculate[s] the probability of the profile given all possible genotype combinations" (FSI 23, ECF No. 41-16 at PageID.1069). The program "assigns a relative weight to the probability of the [electropherogram] given each possible genotype combination at a locus," and the weights across all combinations at that locus are normalized so that they sum to one (*id.*).

As Defendant points out, a number of factors entered into the STRMix program are under the control of the operator/analyst or the individual laboratory, and thus are variable, and affect errors (ECF No. 41 at PageID.752; FSI 23, ECF No. 41-16 at PageID.1073). For instance, although "[t]he true number of contributors to a profile is always unknown" (*id.* at PageID.1075), the individual analyst determines the number of contributors to a DNA profile for purposes of the

14

STRmix™ analysis[5] (MSP PMBIO § 2.11.8, ECF No. 41-12 at PageID.989-990; MSP STRMix

Validation Summary, ECF No. 41-14 at PageID.1018).   MSP policies characterize the

determination of the number of contributors as an estimate, based on "[t]he overall quality of the

electropherogram, the locus with the greatest number of interpretable alleles ($\geq$ 250 RFUs), the

peak height of alleles within a locus and the presence of possible alleles below the Analytical

Threshold …" (MSP PMBIO § 2.10.9.2, ECF No. 41-12 at PageID.970).   The manual cautions

that "[s]tudies indicate it is difficult to determine with certainty the actual number of donors to any

given mixture, especially as the number of donors increases" (*id*.).

STRMix also relies upon terms set by the individual laboratory, such as analytical

thresholds, stutter ratios, drop-in rates, and saturation levels[6] (Def. Br., ECF No. 41 at PageID.753,

citing MSP PMBIO, § 2.11.8, ECF No. 41-12 at PageID.989; FSI 23, ECF No. 41-16 at

PageID.1073; MSP STRMix Validation Summary, ECF No. 41-14 at PageID.1032).   Furthermore,

the input amount of DNA "can have a dramatic impact on the quantity and quality of the STR

results obtained," and "the significance of the likelihood ratios are negatively impacted as the input

DNA amount decreases and the extent of allelic and locus drop-out increases" (MSP STRMix

Validation Summary, ECF No. 41-14 at PageID.1052-1053).   Accordingly, MSP policies provide

that "[p]rofiles exhibiting significant levels of allelic and/or locus drop-out of one or more of the

contributors may not be suitable for analysis using STRmix™" (MSP PMBIO 2.10.9, ECF No.

41-12 at PageID.970; *see also* 2.11.7, ECF No. 41-12 at PageID.989).   Because STRMix requires

---

[5] The evidence in this case is focused on STRmix™ version 2.3.07; subsequent versions may have advanced technology that alters certain features.

[6] The analytical threshold is a limit selected by the laboratory to distinguish "baseline or noise" from a true allelic peak; stutter is a phenomenon that is caused by miscopying in the PCR process; "Drop-in is a non-reproducible, unexplained peak within a profile" (Def. Br., ECF No. 41 at PageID. 753 nn.8-10, citing MSP STRMix Validation Summary, ECF No. 41-14 at PageID.1019, 1025).

15

use of a numeric value for the allele, if an allele cannot be accurately assigned, then MSP policies

instruct that "the entire locus shall be removed from the STRmix™ table and analysis" (MSP

PMBIO 2.11.8, ECF No. 41-12 at PageID.989). However, if an allele is removed, the MSP manual

requires that a statement to that effect be included in the laboratory report (*id*. at PageID.995).

D. Guidelines, Standards and Government Review

The defense has raised, as a key argument, STRmix™ developers' alleged failure to adhere

to existing general standards for software development as a basis for exclusion of the DNA

evidence under *Daubert*. To date, there are no standards in the United States for the development

or use of probabilistic genotyping software in forensic DNA analysis. Guidelines have been issued

for the validation of probabilistic genotyping software by the Scientific Working Group on DNA

Analysis Methods ("SWGDAM");[7] however, they are merely guidelines. There are no current

standards that a lab can be audited against in the forensic community, either in the United States

or internationally (ECF No. 152 at PageID.4101). Consequently, a significant amount of testimony

and evidence addressed, in a round-about fashion, the applicability of existing software standards

and protocols, and related critical review of STRmix™.

Nathan Adams testified that there are specific industry standards and practices used in the

field of software development and testing for validation of software programs, set by a number of

standards-setting bodies (ECF No. 78 at PageID.2868). The Institute of Electrical and Electronic

Engineers (IEEE) is known for developing many standards specific to the development,

maintenance, testing, and inspection of software (*id*.). Many of these standards have been adopted

by international organizations like the ISO organization, which is the international standard setting

---

[7] SWGDAM, Guidelines for the Validation of Probabilistic Genotyping Systems (June 15, 2015), https://docs.wixstatic.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf (last visited 10/14/19).

body, and have also been adopted by countries formally at a federal level or at a federal departmental level (*id*.). At the request of the defense, Adams conducted a review of STRmix™ version 2.3.07 used in this case, including the source code, and provided a report addressing whether the development of STRmix™ has been demonstrated to be in accordance with software engineering standards and principles (ECF No. 78 at PageID.2892-2895). It was Adams' opinion that given departures from basic software engineering practices, STRmix™ version 2.3.07 should not be considered as verified and validated against objective criteria (*id*. at PageID.2916; 5/24/18 Hrg., Def. Ex. D)

Dr. Buckleton acknowledged that "IEEE is the most commonly used standard setting body in computer science" (ECF No. 77 at PageID.2564). However, it is not without its faults (*id*.). For instance, adherence to IEEE actually caused a bug in Excel (*id*.). Compliance with IEEE standards is not mandatory, and has not been suggested by any guidance bodies for probabilistic genotyping (*id*. at PageID.2565). The International Society of Forensic Geneticists (ISFG) does make some suggestions for including core computer science principles, and the Forensic Science Regulator suggests that probabilistic genotyping should be developed within a quality system; Dr. Buckleton's view is that STRmix™ was developed within that quality system (*id*.).

While no formal standards exist governing probabilistic genotyping software, the need for guidance and uniformity has been recognized in governmental review. The defense emphasizes a 2016 Report of the President's Council of Advisors on Science and Technology (PCAST), which addressed modern forensic DNA analysis (*see* ECF No. 41 at PageID.760-761). PCAST was tasked with determining whether additional steps should be taken, "beyond those already taken by the Administration in the aftermath of a highly critical 2009 National Research Council report on the state of the forensic sciences, that could help ensure the validity of forensic evidence used in

17

the Nation's legal system" (*id*., footnote omitted). The PCAST Report offers recommendations to

NIST, the White House, the FBI, the Attorney General, and the judiciary on "actions that could be

taken to strengthen forensic science and promote its rigorous use in the courtroom" (*id*. at

PageID.761).

The 2016 PCAST Report acknowledged that "[t]he vast majority of DNA analysis

currently involves samples from a single individual or from a simple mixture of two individuals

(such as from a rape kit). DNA analysis in such cases is an objective method in which the

laboratory protocols are precisely defined and the interpretation involves little or no human

judgment." (*Id*.; PCAST Report, ECF No. 41-17 at PageID.1102, 1165-1166). However, "[a]s

DNA testing kits have become more sensitive, there has been growing interest in 'touch DNA'—

for example, tiny quantities of DNA left by multiple individuals on the steering wheel of a car"

(ECF No. 41-17 at PageID.1170). The Report explained that "the fundamental difference"

between analysis of complex-mixture samples (defined as mixtures with two or more contributors),

versus single-source and simple mixtures, "lies not in the laboratory processing, but in the

interpretation of the resulting DNA profile" (*id*. at PageID.1102, 1170).

The Report states:

Interpreting a mixed profile is different for multiple reasons: each individual may contribute two, one or zero alleles at each locus; the alleles may overlap with one another; the peak heights may differ considerably, owing to differences in the amount and state of preservation of the DNA from each source; and the "stutter peaks" that surround alleles (common artifacts of the DNA amplification process) can obscure alleles that are present or suggest alleles that are not present. It is often impossible to tell with certainty which alleles are present in the mixture or how many separate individuals contributed to the mixture, let alone accurately to infer the DNA profile of each individual.

* * *

Because many different DNA profiles may fit within some mixture profiles, the probability that a suspect "cannot be excluded" as a possible contributor to complex mixture may be *much higher* (in some cases, millions of times higher) than the

18

probabilities encountered for matches to single-source DNA profiles. As a result, proper calculation of the statistical weight is critical for presenting accurate information in court.

(*id*. at PageID.1170-1171; emphasis in original; footnotes omitted).

The PCAST Report specifically addressed probabilistic genotyping computer programs, including STRMix (*id*. at PageID.1173-1174). PCAST cautioned that while these programs "clearly represent a major improvement over purely subjective interpretation," they still require careful scrutiny to determine "whether the methods are scientifically valid" and "whether the software correctly implements the methods" (*id*. at PageID.1174). "This is particularly important because the programs employ different mathematical algorithms and can yield different results for the same mixture profile" (*id*. at PageID.1103, 1174; footnote omitted).

Most importantly, the Council stated that "[t]he two most widely-used methods (STRMix and TrueAllele) appear to be reliable within a certain range based on available evidence and the inherent difficulty of the problem": "Specifically, these methods appear to be *reliable for three-person mixtures in which the minor contributor 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*"[8] (*id*. at PageID.1175; emphasis added; footnotes omitted). The Council noted that "appropriate evaluation" of the proposed methods should include studies by multiple groups "*not associated with the software developers* that investigate the performance and define the limitations of programs by testing them on a wide range of mixtures with different properties" (*id*. at PageID.1174).

After the PCAST Report was issued, on October 12, 2017, NIST, a non-regulatory agency of the U.S. Department of Commerce, released a study concluding that using the LR in courtrooms

---

[8] Here, Gissantaner was determined to be a minor contributor of only 7% of the DNA analyzed.

is not consistently supported by scientific reasoning (ECF No. 41 at PageID.763, citing NIST Article, "NIST Experts Urge Caution in Use of Courtroom Evidence Presentation Method," ECF No. 41-15, and LUND, ET. AL, "Likelihood Ratio as Weight of Forensic Evidence: A Closer Look," Journal of Research of National Institute of Standards and Technology, Vol. 122, Art. No. 27 (2017) ("NIST Study"), ECF No. 41-15). The authors of the study, NIST statisticians Steve Lund, who testified at the May 2018 *Daubert* hearing, and Hari Iyer, caution that the justification for using LR in courtrooms is flawed, arguing that "it risks allowing personal preference to creep into expert testimony and potentially distorts evidence for a jury" (NIST Article, ECF No. 41-15 at PageID.1064).

In their study, Lund and Iyer explain that proponents of the LR approach appear to justify its use with Bayesian reasoning, "a paradigm often viewed as normative (*i.e.*, the *right* way; what someone *should* use) for making decisions when uncertainty exists" (NIST Study, ECF No. 41-18 at PageID.1258; emphasis in original). Bayesian decision theory has long been used as a reasoning approach "by the scientific community to create logic-based statements of probability" (NIST Article, ECF No. 41-15 at PageID.1064). "Bayesian reasoning is a structured way of evaluating and reevaluating a situation as new evidence comes up" (*id*.).

Essentially, under Bayes' rule, "individuals multiply their previous (or prior) odds by their respective likelihood ratios to obtain their updated (or posterior) odds, reflecting their revised degrees of belief regarding the claim in question" (NIST Study, ECF No. 41-18 at PageID.1258). Applying this approach "allows an expert to come up with a logic-based numerical LR that makes sense to the expert *as an individual*" (NIST Article, ECF No. 41-15 at PageID.1064-1065; emphasis added). However, Lund and Iyer argue that Bayesian reasoning "breaks down in situations where information must be conveyed from one person to another such as in courtroom

20

testimony" *(id.* at PageID.1064). They state that "[t]he trouble arises when other people—such as jurors—are instructed to incorporate the expert's LR into their own decision-making," because an expert's judgment often involves complicated statistical techniques that can generate different LRs, depending on which expert is making the judgment *(id.* at PageID.1065). Lund and Iyer observe that "[c]omputing an *LR* for anything but the simplest of problems will involve approximations" (NIST Study, ECF No. 41-18 at PageID.1264), explaining that:

> reporting a single LR value after an examination of available forensic evidence *fails to correctly communicate* to the [decision maker] the information actually contained in the data. Personal choices strongly permeate every model.

(id. at PageID.1271; emphasis added).

In their summary, Lund and Iyer warn that "Bayesian decision theory neither mandates nor proves appropriate the acceptance of a subjective interpretation of another, regardless of training, expertise, or common practice" (NIST Study, ECF No. 41-18 at PageID.1279; footnote omitted). Furthermore, although validation can demonstrate that a particular interpretation may be reasonable, "this should not be misunderstood to mean the model is accurate or authoritatively appropriate" *(id.*; footnote omitted).

While a decision maker "only needs to be personally satisfied regarding the suitability of using any given *LR* in Bayes' formula, guiding the probabilistic interpretation of others [such as jurors] requires greater care" *(id.* at PageID.1264). The authors recommend using an LR only in cases where a probability-based model is warranted, "such as the evaluation of high-quality samples of DNA from a single source" (NIST Article, ECF No. 41-15 at PageID.1064). Lund and Iyer maintain that for a technique to be broadly applicable, it must be based on measurements that can be replicated *(id*. at PageID.1065).

III. MSP DNA ANALYSIS IN THIS CASE

The analysis in this case extends beyond the standard DNA analysis in order to decipher a low copy number/low-template (quantity) mixture of more than one contributor's "touch" DNA using probabilistic genotyping software, STRmix™. Although the MSP uses STRmix™ software to analyze complex DNA mixtures, not all the MSP analysts are trained in STRmix™ (*id.* at PageID.2704-2705). That is, all analysts are trained in DNA analysis and mixture interpretation, but they are not trained in producing statistics on their mixtures (*id.* at PageID.2705).[9] According to Amber Smith, who did the STRmix™ analysis here, at the time the evidence sample was submitted to the MSP lab, only two analysts in the Lansing unit were performing STRmix™ analysis (*id.* at PageID.2705).

If the analyst "running" a case develops single-source samples, then the analyst will produce a report that has a random match statistic[10] (*id.*). However, if the statistic becomes too complex for mixture samples, then STRmix™ software must be used to produce a statistical report (*id.*). If the analyst runs the sample and sees that it is a mixture (more than one contributor/donor), the analyst then forwards the sample to a qualified STRmix analyst to determine if the sample fits the guidelines for STRmix analysis (*id.*). This occurred in this case. The initial analyst who received the evidence sample, Katie Urka, developed her profiles and then forwarded them to the two STRmix™ trained analysts to evaluate for potential STRmix analysis (which still occurs until all MSP analysts are STRmix™ qualified) (*id.* at PageID.2705-2706).

The MSP laboratory attempts amplification of STRs at 24 loci (ECF No. 52 at PageID.1742, citing ECF No. 41-12 at PageID.957). Two steps of the DNA process are first, PCR,

---

[9] The way the MSP laboratory works is that analysts perform their own lab work on their own cases and then generate reports.

[10] Random match probability is a very common statistical method used for a single source DNA profile (ECF No. 77 at PageID.2658).

and then second, capillary electrophoresis (ECF No. 77 at PageID.2660). PCR is the multiplication

of the DNA that is present so it can be viewed (*id*.). Capillary electrophoresis is the sorting of the

different fragments of DNA (*id*.). Jeffrey Nye summarized the process:

> during the amplification process each DNA fragment is tagged with a fluorescent label, and then as they are sorted in the capillary, they pass a detector, a camera, from which they could detect the fluorescence of each fragment as it goes by, then that detection is represented visually with an electropherogram or an e-gram ….

(*Id*.). The MSP lab uses GeneMapper software to generate the electropherograms; then STRmix

is the statistical tool used to generate the statistics (*id.* at PageID.2707).

The actual DNA analysis process in this case involved two forensic scientists and two

separate DNA analyses and reports at the MSP lab. Urka, the initial analyst, did her lab work, and

generated the electropherograms from the evidence sample (ECF No. 41-13). Urka issued a report,

but she could not issue the statistic on the report because the sample needed a mixture statistic

(ECF No. 77 at PageID.2706). Urka's report merely stated that the DNA results obtained "indicate

it is a mixture of multiple contributors. Further analysis of this DNA profile will be the subject of

a subsequent report" (*id.*; ECF No. 41-8 at PageID.826).

Following Urka's generation of the electropherograms, Smith utilized those same

electropherograms to proceed with her STRmix™ analysis to help determine the number of

contributors. However, the analysis under STRmix™ is subject to different MSP protocol.

Because of the different protocols, if Smith did not generate the data, she could only assess how

the data is once she received it (ECF No.77 at PageID.2713).

As part of the STRmix™ protocol, Smith is required to document her number of

contributors as well as how she arrived at that number (*id.* at PageID.2713-2714). Smith explained

that the electropherograms generated by Urka have the GeneMapper software stutter filters turned

on, based on certain levels set for the lab to filter stutter, which is an artifact generated during the

DNA analytical process (*id.* at PageID.2707). However, given that different protocols governed Smith's analysis as opposed to Urka's, Smith had to analyze and insert the electropherogram into STRmix with stutter filters off—meaning that all of the artifacts are present on the electropherograms—so that STRmix can analyze the sample and determine the probability that this peak is a true artifact or if it could possibly be a potential type (*id.* at PageID.2707-2708).

Smith issued the final DNA report in this case. Her report first states how she interpreted the sample, which says that by looking at the evidence, she ran the sample assuming there were three individuals (*id.* at PageID. 2716). Based on MSP reporting formats, she must also state her hypotheses, which is generally the person of interest and two unrelated, unknown contributors, or did this sample generate from three unrelated, unknown contributors, not the person of interest (*id.*).

The MSP analysis provided to the defense indicated that the DNA contained in the swab of the firearm consisted of three contributors (ECF No. 41 at PageID.750, citing MSP Report 2, ECF No. 41-8). The majority of the DNA was female (*id.*, citing MSP Worksheet, ECF No. 41-5). The analysis focused on 23 loci, 17 of which contained alleles that matched those of Gissantaner's sample (*id.*, citing MSP EPG, May 22, 2016 (DNA extract from swabs of gun), ECF No. 41-13). No information was provided with respect to the identity of the other contributors of the DNA on the weapon (ECF No. 41 at PageID.750).

Smith's report concluded:

Based on the DNA typing results obtained, it is at least 49 Million times more likely if the observed profile from the swabs of the textured areas of GUN-001 originated from Daniel Gissantaner and two unrelated, unknown contributors than if the data originated from three unrelated, unknown individuals.

(MSP Report 3, ECF No. 41-10 at PageID.919). The report also stated a verbal equivalent of this conclusion in lay terms, which corresponded with the 10,000 and greater part of the MSP table for

24

a verbal equivalent, which was "very strong support" that Daniel Gissantaner is a contributor to

the DNA profile developed from the swabs of the textured area of GUN-001[11] (*id.* at PageID.920;

ECF No. 77 at PageID.2717).

It is these conclusions that the Government seeks to admit as evidence, the defense seeks

to exclude, and the Court must examine under *Daubert*.

IV. *DAUBERT* STANDARD

The Federal Rules of Evidence require a trial court judge to ensure that scientific testimony

or evidence is both reliable and relevant before it may be admitted. *Daubert v. Merrell Dow*

Pharms., Inc., 509 U.S. 579, 589 (1993). The court's two-prong obligation derives specifically

from Federal Rule of Evidence 702, "which clearly contemplates some degree of regulation of the

subjects and theories about which an expert may testify." *Id.*

Rule 702. Testimony by Expert Witnesses

A witness who is qualified as an expert by knowledge, skill, experience, training,
or education may testify in the form of an opinion or otherwise if:

(a) the expert's scientific, technical, or other specialized knowledge will help the trier
of fact to understand the evidence or to determine a fact in issue;

(b) the testimony is based on sufficient facts or data;

(c) the testimony is the product of reliable principles and methods; and

(d) the expert has reliably applied the principles and methods to the facts of the case.

FED. R. EVID. 702.

---

[11] Nye explained that a likelihood ratio statistic is in a numerical form, and because the MSP has
90 DNA scientists, who, along with other experts might review that material or testify to those
findings, all of whom may interpret the number to mean something different, he helped develop a
table of verbal equivalents to the LR numerical value to assist the jury and assist MSP scientists to
normalize or standardize how they present the data and the significance, i.e., how strong or weak
the evidence is, in testimony (ECF No. 77 at PageID.2687-2688).

Under Rule 702, the court has discretionary authority to determine reliability and relevancy, given the particular facts and circumstances. *Kumho Tire Co., Ltd. v. Carmichael*, 526 U.S. 137, 158 (1999). This "gatekeeping" duty of the district court applies to all specialized knowledge, including, but not limited to technical and scientific knowledge. *Id*. at 147-48; *see also Daubert*, 509 U.S. at 589. The proponent must show by a preponderance of the evidence that scientific evidence is admissible. *Daubert*, 509 U.S. at 592 n.10.

In *Daubert*, the Court set forth four non-exclusive factors for a court to evaluate in determining the admissibility of scientific or expert testimony:

(1) whether the theory or technique can be, and has been, tested;

(2) whether the theory or technique has been subjected to peer review and publication;

(3) the known or potential rate of error of the particular scientific technique or theory and the existence and maintenance of standards controlling the technique's operation; and

(4) whether the theory or technique has general acceptance in the relevant scientific community.

Id. at 593-94. However, no single factor alone is necessarily dispositive, and other factors may be relevant. *See id.* at 593; *see also Kumho Tire, supra*, 526 U.S. at 149. The inquiry envisioned by Rule 702 is a flexible one. *Daubert*, 509 U.S. at 594. "[S]ubmission to the scrutiny of the scientific community is a component of 'good science,' in part because it increases the likelihood that substantive flaws in methodology will be detected." *Id.* at 593.

Finally, even relevant evidence must be excluded "if its probative value is substantially outweighed" by a danger of unfair prejudice, confusion of the issues, misleading of the jury, undue delay, wasting time, or the needless presentation of cumulative evidence. FED. R. EVID. 403.

26

V. ANALYSIS

Defendant asserts that this case implicates important questions about the standards for the collection, analysis, and interpretation of DNA material—questions appropriately evaluated under *Daubert*. He argues that the MSP DNA analysis, reporting a likelihood ratio generated by STRMix probabilistic genotyping software, is an unreliable statistical estimate because the program's code relies upon information that is subjective and can vary to an impermissible degree depending on the individual analyst and laboratory. Furthermore, the evidence does not meet accepted standards for use of this program. Lastly, submission of these results to a jury would be unfairly prejudicial because the LR created by STRMix in this case is unreliable regarding the possibility that some of the DNA material found on the firearm belongs to Gissantaner, and not another individual.

As noted, the Court conducted a *Daubert* hearing on May 23 and 24, 2018, which included expert and lay witnesses on both sides of the issue. The *Daubert* hearing was continued over the course of more than a year to identify and appoint independent court experts and secure their opinions on the DNA analysis and use of probabilistic genotyping software in this case. The Court heard an additional day of *Daubert* testimony from the court-appointed experts on July 8, 2019, followed by supplemental briefing from the parties. The Court now undertakes its *Daubert* gatekeeping responsibility to examine the evidence and determine whether the DNA testimony and other evidence is both reliable and relevant, before it may be admitted as evidence of Defendant's guilt.

As noted, the *Daubert* hearing culminated in a day of testimony from two court-appointed experts following their written reports, which is where the Court begins its analysis. Both court-appointed experts are well-recognized in the field of forensic DNA analysis and for their study of probabilistic genotyping software analysis.

27

Michael Coble, PhD, is the Associate Director of the Center for Human Identification and Associate Professor in the Department of Microbiology, Immunology, and Genetics at the University of North Texas Health Science Center in Fort Worth Texas. Prior to his work in academia, Coble was employed as a federal employee at NIST and as a federal contractor at the Armed Forces DNA Identification Laboratory. Dr. Coble has extensive experience in all aspects of forensic DNA analysis, including analyses of low level samples, DNA mixture interpretation, population genetics, statistical interpretations, autosomal and Y-chromosomal STR markers, SNP markers, mitochondrial DNA, and probabilistic genotyping.

Dan E. Krane, PhD, is a Professor of Biological Sciences at Wright State University in Dayton, Ohio, where he has been a faculty member since 1993. His "research interests are primarily in the areas of molecular evolution and the way that gene frequencies change over the course of time in populations of organisms."[12] Dr. Krane is also involved with the use and development of computer-based tools to evaluate DNA evidence associated with criminal investigations. He is President/CEO of Forensic Bioinformatics, a consulting company founded in 2002, which reviews DNA testing results from hundreds of court cases around the world each year. The large amounts of data associated with those reviews lends itself to meta-analyses that allows development of tools and approaches that make forensic DNA profiling more reliable and objective.

Each expert prepared a written report pursuant to the Court's order of appointment (ECF Nos. 139, 140). The experts' reports addressed considerations specific to *Daubert* regarding STRmix software as well as the application of STRmix to the evidence in this case. And the

---

[12] https://people.wright.edu/dan.krane (last visited 10/3/19); *see also* ECF No. 146-1 at PageID.3424).

experts had the benefit of reviewing the transcripts and all of the evidentiary exhibits from the initial two-day *Daubert* hearing, i.e., a bird's eye review of the pertinent evidence from the perspective of the Government and the defense.

It is patently clear from the experts' written opinions and testimony that much divergence remains on the reliability of probabilistic genotyping software under the circumstances presented in this case—the likelihood ratio generated from the analysis of a complex mixture of low-template touch DNA consisting of at least three contributors in which the person of interest is determined to be a minor contributor of only 7%.

With respect to the admission of the DNA analysis in this case, it is the Court's gatekeeping responsibility to first determine if it is reliable scientific evidence or testimony, and it is the Government's obligation to make that showing. In its concluding arguments, the Government emphasizes that the *Daubert* standard is reliability, not infallibility; that Dr. Coble recommends admitting the evidence; and that most of the *Daubert* factors are not at issue. Further, Dr. Coble and Dr. Krane agree that probabilistic genotyping is the new paradigm for DNA mixture analysis, and *Daubert* is not a high bar. The Government asserts that the defense contrasts STRmix with perfection, but to the extent the defense challenges to STRmix are even proper under *Daubert*, they are merely fodder for cross-examination—not exclusion of the evidence.

The court-appointed experts addressed each *Daubert* factor with regard to the STRmix DNA analysis in this case, both in their reports and in extensive direct and cross-examination. The Court has considered the experts' reports and testimony, and other supporting or opposing evidence on the *Daubert* factors. While the Court agrees with the Government that the factors are not of equal concern given the *Daubert* evidence, the Court gives them each due consideration.

29

That consideration leads to but one conclusion: the Government has failed to show sufficient support for admission of the evidence under *Daubert*.

(1)   Whether the theory or technique can be, and has been, tested

The first question posed by the Court to the experts addressed testing: whether the use of STRmix has been adequately tested and validated, independently of the testing by the developer.

In response, Dr. Coble noted that any software program used in a forensic DNA laboratory must go through a rigorous validation study. The developers of the software program first perform a "developmental validation" "to provide the community an introduction of the program" (Coble Report, ECF No. 139 at PageID.3349). Often, this study is published in a peer-reviewed journal and provides the forensic community information on the software's performance and features. The adoption of a new software program by a forensic laboratory requires a separate validation study, called an "internal validation." "The software program is tested independently from the developer within the laboratory using data generated in the laboratory using **their STR kit**, analyzed on **their capillary electrophoresis instrument**, under **their established laboratory conditions**" (*id.* at PageID.3349-3350; emphasis in original). Dr. Coble's report states, "[t]here are 45 laboratories in the United States alone that have tested, validated and implemented STRmix for case work in the   US   as   of   5/22/19"   (*id.*   at   PageID.3350,   citing   Buckleton   website, https://johnbuckleton.files.wordpress.com/2019/04/labs-live-1.pdf). He concludes that he "would personally consider this as evidence that STRmix has been adequately tested and validated independently of the developers" (*id.*).

In Dr. Krane's view, two different kinds of internal testing/validation should be of interest to the Court: (a) that performed by other laboratories using STRmix in their casework, and (b) that performed by the Michigan State Police laboratory prior to its use of STRmix in its casework (Krane Report, ECF No. 140 at PageID.3400). Dr. Krane observed that more than 30 laboratories

30

worldwide had performed internal validation studies of STRmix prior to this Court's May 2018

Daubert hearing, many of which had published their findings in publications/presentations, some

of which were independent of the developer. However, he noted there was relatively little

testimony about the MSP laboratory's testing/validation of STRmix prior to its use in casework

beginning in early 2016. Upon review of the limited evidence, Dr. Krane concluded that the MSP

Validation Summary for STRmix (Govt Ex. 10) did "not appear to have identified any limitations

such as at what level alleles are dominated by background noise (at least in terms of quantity of

template DNA or mixture ratio) for 2-, 3-, and 4-person low-level DNA mixtures" (ECF No. 140

at PageID.3402). While the testing identified in the MSP Validation Summary variously involved

three- and four-person mixtures, Dr. Krane pointed to a number of shortcomings in the available

report to establish validation under the circumstances presented in this case.

Dr. Krane cited both Dr. Lund's testimony in the May 2018 *Daubert* hearing and a 2016

publication by Dr. Coble stressing the implications of validation sample limitations in the

application of probabilistic software. Dr. Coble's 2016 publication explains:

> The goal of an internal validation study is to explore the limitations of the software
> and test the reliability, robustness, and reproducibility of the system. Samples that
> mimic the types of cases encountered should be tested. … Determination of the
> limits of the software is important to establish the types of profiles that are suitable
> for handling by the laboratory. … Probabilistic software, especially for low-level
> DNA mixtures, may allow a laboratory to widen the scope of their casework in
> terms of the type of evidence handled. However, there may also be a temptation to
> submit all complex mixtures to particularly versatile software. Therefore, the
> community is reminded of a previous recommendation of the DNA Commission
> (2) that is still valid: "Gill et al., 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."

(*id*., citing Govt Ex. 23). Dr. Krane noted that Amber Smith, the MSP analyst who performed the

STRmix analysis in this case, testified that the evidence sample here contained DNA from three

contributors but was open to the possibility that it arose from four contributors. She believed that

31

"around 3 microliters" of a 0.2344 nanogram (ng) per microliter solution was used for testing—corresponding to approximately 0.7 ng of total DNA. STRmix estimated that if the sample did arise from three contributors that the lowest level contributor (the one who might be the person of interest) was responsible for approximately 7% of the total amount of DNA—corresponding to just less than 50 picograms of DNA (ECF No. 140 at PageID.3401, n.1).

Although the Government filed a letter and additional evidence from the MSP DNA technical leader (ECF No. 146-14) following the expert reports to address the testing shortcomings identified by Dr. Krane, it is the Court's conclusion that the evidence on the record does not establish adequate testing and validation of the STRmix software under the conditions of the DNA evidence in this case—a complex mixture of low-copy number/low-template DNA, approximately 0.7 nanograms, with at least three contributors and possibly four, where the person of interest is a minor contributor of 7% of the DNA analyzed, approximately 49 picograms—approximately 8-9 human cells (*see* 7/8/19 Tr., ECF No. 152 at PageID.4183).

Additional evidence filed by the Government indicates five adjudicated case mixtures analyzed in the MSP lab that set outer bounds encompassing the conditions in this case: a 7% minor contributor with 49 picograms of DNA (*id*. at PageID.4140-4144; table, ECF No. 146-14 at PageID.3983). However, this testing was not included in the MSP Validation Summary, which Dr. Krane found concerning, since it would be helpful to a review of the MSP Validation Summary and would have answered some reasonable and appropriate questions (ECF No. 152 at PageID.4164-4166). In any event, Dr. Krane pointed to concerns that remain with validation and that, in the Court's view, preclude a conclusion that the STRmix software used by the MSP in this case has been sufficiently tested and validated to rely on this likelihood ratio evidence in Defendant's criminal prosecution.

Dr. Krane testified that some aspects of the table of adjudicated case mixtures required additional scrutiny. He stated that it would be extremely important to know the likelihood ratios that were generated, particularly in the instances where the lowest level contributor contributes two or three picograms of DNA (ECF No. 152 at PageID.4144). The mixture study information also did not indicate what the false-positive and the false-negative rates were, which would be especially helpful with respect to the very marginal samples (*id.*). Dr. Krane stated that if STRmix was able to generate significant likelihood ratios for samples of only two or three picograms, that would be significantly at odds with the published literature about STRmix's behavior for such low level samples (*id.* at PageID.4145, 4165-4166). In any event, in adjudicated case mixture studies, the true nature of the sample is unknown, and the purpose is essentially to see if the results are generally consistent with the results obtained during the original investigation, before the tool was available (*id.* at PageID.4145).

A second table of lab-created mixture study results (ECF No. 146-14 at PageID.3984) submitted by the Government to address validation questions raised in Dr. Krane's report also left concerns. In this instance, the experiments were controlled, and the contributors were known. Dr. Krane indicated that it would be extremely helpful to see the actual test results, the electropherograms, but at the very least in the context of validation, it would be very important to see another column that gave the likelihood ratios that STRmix reported, and if possible, something about the false-positive and the false-negative rates associated with those types of samples (*id.* at PageID.4146).

In response to Government counsel's suggestion that it was safe to say that Dr. Krane always would like to see a little more data, no matter the experiment, Dr. Krane elaborated on the importance of not "glossing over" what is a fundamental concern about missing information in the

33

table: "Because in the absence of that information, all we have to work with is a superficial suggestion in the validation summary that the likelihood ratios were significant" (ECF No. 152 at PageID.4146-4147). He reiterated the basis for concern: "the crux here is I want to know, and I think the Court needs to know what are the limits beyond which we should be suspect of STRmix results?" (*id.* at PageID.4147). Ultimately, the analysts for the Michigan State Police need to know where it is that they should proceed with caution, and they get that kind of information from the validation study (*id.*). In Dr. Krane's view:

> [T]he validation summary and the validation study itself ultimately are there to provide guidance to analysts in the use of a new methodology, STRmix. If this information is not within the validation summary it can't make its way into the protocols, the interpretation guidelines for the laboratory, and it can't provide guidance to analysts, and the analysts then are not getting guidance on what I would suggest is perhaps one of the most important questions that they will encounter, which is at what point they should *not* be relying upon STRmix to help them in their work.

(*id.* at PageID.4165; emphasis added).

Dr. Krane noted that there are essentially two categories of objectives for internal validation. One is to demonstrate that a particular methodology or approach or tool generates results that are consistent with what the proponents of the tool or methodology suggest could be obtained—essentially to confirm that the developmental validation is consistent with what it is that the testing laboratory obtains in their own hands (ECF No. 152 at PageID.4167).

The other category is that an internal validation study should establish limits that give guidance to analysts about how and when to use this methodology and, perhaps more importantly when not to use a methodology. Dr. Krane testified that there is "very little guidance" in the MSP standard operating procedures for analysts regarding when and how to use STRmix (*id.*). That is so despite published recommendations that advise using "**extreme caution**" with fully continuous software on low-template DNA samples (*id.* at PageID.4169; emphasis added).

In view of the open questions about the testing and validation of the MSP's use of STRmix in circumstances equivalent to the low-template, low level complex DNA mixture in this case, this *Daubert* factor weighs strongly against a finding that the DNA evidence at issue is reliable.

(2)   Whether the theory or technique has been subjected to peer review and publication

The testimony and evidence on this factor is equivocal, depending on the definitional context ascribed to "peer review." Dr. Coble states in his report (ECF No. 139) that as of June 1, 2019, there are 53 articles in peer-reviewed journals that focus on STRmix theory and application, or relate to some aspect of STRmix (e.g., the use of Likelihood Ratios or software validation). He observes that this is over four times as many publications of the closest competitor to STRmix (TureAllele). While most of the articles written have the developers of STRmix (Drs. Bright, Taylor and Buckleton) as co-authors, Dr. Coble personally does not consider that an issue for three reasons. First, most crime laboratories are unable to publish their internal validation studies in peer-reviewed journals since the results would no longer be considered "novel" once the first paper was published. Second, many laboratories performing casework are generally too busy with the number of cases to analyze, having little time for the laboratory to conduct independent research for peer-reviewed publication. Dr. Coble does point to a publication, "Bright et al. (2018) 'Internal Validation of STRmix: A multi-laboratory response to PCAST.'" *Forensic Science International: Genetics* 34:11-24 (Govt Ex. 4), as an excellent example of independent peer-review, since the data is from 31 independent labs.

Finally, Dr. Coble notes that many research laboratories will need to purchase and receive training in STRmix before they can produce independent peer-reviewed publications in the literature. It is much easier for academic and research laboratories to use open-source and freely available tools in the forensic domain. Thus, it may take some time before a stream of independent

researchers is publishing (although there are a couple of papers in the last year that are independent of the developers).

Dr. Krane acknowledges in his report (ECF No. 140) that, as was noted at the May 2018 hearing, Dr. Buckleton and his colleagues are prolific publishers of peer-reviewed papers. However, he states a fairly common criticism of STRmix, as well as its largest competitors, is that there have been virtually no publications by individuals who either are not directly affiliated with the development and sale of STRmix and/or who have not made a financial commitment to utilize STRmix.   He concludes, persuasively, that there is unquestionably a value associated with extensive independent peer review of any powerful analytical tool or approach.[13]   And it is unlikely that that value will be realized if courts or government procurement guidelines do not insist on truly independent review of both the performance and the coding/implementation of programs like STRmix.

In his testimony, Dr. Coble acknowledged that it would help to have more peer-reviewed materials examining STRmix, particularly because it is still fairly early in this process (ECF No. 152 at PageID.4016).   The first lab to start using STRmix was the U.S. Army crime lab in late 2013 or maybe 2014. He qualified his statements in his report concerning the more than 50 articles in peer-reviewed journals, noting that not all of those articles were specifically about STRmix itself, but may relate to aspects of STRmix, like how stutter is modeled, and how peak height variability is modeled (*id.* at PageID.4016-4017).

The Court concludes that, at the very least, the testimony and evidence concerning the publication of peer review articles addressing complex mixtures is a mixed review. While the peer

---

[13] The Court observes that this is particularly true where the use of the tool is likely to have serious consequences, such as here, where Defendant's conviction would result in a minimum of 15 years of imprisonment.

review factor carries some weight in favor of the use of STRmix generally, in this Court's view, the evidence adds nothing of significance to the reliability determination specific to this case, involving complex mixtures of low-template, low level DNA. Further, the Court takes note of the studies and articles, such as the NIST Study by Lund and Iyer, which have determined that review of probabilistic genotyping software, independent of that of the developers, is critical for an assessment of its reliability with respect to use in the courts.

(3)   Known or potential rate of error of the particular scientific technique or theory and the existence and maintenance of standards controlling the technique's operation

Based on the May 2018 *Daubert* hearing and the evidence to date, the Court asked that the experts specifically address: (1) evidence of the rate of error in applying STRmix and the significance to the DNA testing in this case; (2) the maintenance of standards, certifications, and extent of validation of STRmix by the MSP laboratory; and (3) whether the MSP validation was reviewed by any external auditors and whether it is accepted protocol that self/internal validation of DNA analysis software by a laboratory is sufficient or whether scientific protocol requires external and independent review of the validation.

In his report, Dr. Coble details circumstances in which STRmix could result in "false exclusions" and "false inclusions," especially the former, in a very low level sample with a minor contributor representing a relatively "small" quantity of DNA (a "trace" contributor) in a mixture (ECF No. 139). Dr. Coble also acknowledged this in his testimony: "obviously, there is the potential for having false inclusions and false exclusions …" (ECF No. 152 at PageID.4101). In his testimony, he explained that when you have someone who is not in the mixture that gives you a likelihood ratio of greater than one, that is considered to be a false-positive; alternatively, a false-negative would be someone who is actually in the mixture but they give you a likelihood ratio that is less than one (*id*. at 4025).

However, Dr. Coble's opinion is that these are not "errors" of the software: "The software has modeled the contributors to the mixture and we have by chance created a reference that would closely match the person of interest" (ECF No. 139 at PageID.3351). He noted that statistician Alan Turing, who worked on code-breaking during World War II, stated that the error of a likelihood ratio is "about one over the likelihood ratio that you would expect" (ECF No. 152 at PageID.4026-4027). So, when he thinks about errors, it is about catching errors in running samples, not catching errors by looking at software code (*id.* at PageID.4027).

Dr. Coble expressed his opinion that any "error" rate would be expected to be very low (ECF No. 139 at PageID.3351). His reasoning for this latter opinion is that laboratories (including the MSP laboratory, based on the testimony of analyst Smith along with a review of their protocols) are not treating STRmix as a "black box" and are performing diagnostic interpretations after the analysis of the software, which add an additional level of scrutiny of the results and not just "copying and pasting the LR ratio into a report" (*id.*).

Importantly, Dr. Coble states that at the moment, there are no standards for probabilistic genotyping software (*id.* at PageID.3352). The American Academy of Forensic Sciences Standards Board (ASB) is a Standards Development Organization that is currently creating standards for the forensic community. Dr. Coble was on the Organization of Scientific Area Committees (OSAC) task group that published a set of proposed standards for the validation of probabilistic genotyping software, which is currently working its way through the ASB process (*id.*). The proposed standards are based on the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Software. And it is Dr. Coble's opinion that laboratories following the SWGDAM guidelines during STRmix validation will have no issues meeting the standards and requirements of the ASB once they are published (*id.* at PageID.3353). However, Dr. Coble

38

certainly thinks there is a need for more standards when it comes to mixture interpretation than

what exists currently, which are in the process (ECF No. 152 at PageID.4102). The FBI is updating

the quality assurance standards and the OSAC has also submitted several standards (*id*.).

Dr. Coble reviewed the MSP STRmix validation summary (Govt Ex. 10) and found that it

followed the SWGDAM Guidelines for the Validation of Probabilistic Genotyping Software (Govt

Ex. 19). His report states that the results of the validation summary are consistent with other

internal validations and publications of probabilistic genotyping software (ECF No. 139 at

PageID.3353).

In response to the Court's more specific inquiry, Dr. Coble stated that he saw no

information in the record of any external review of the MSP validation study by auditors.

However, in the forensic DNA community, it is generally accepted that self/internal validation of

DNA analysis software is sufficient for the laboratory to start using the software (*id.*). In his

testimony, he acknowledged the Government's additional evidence that the MSP lab did have an

external audit, which included a review of STRmix and there were no findings or nonconformances

(ECF No. 152 at 4045-4046).

With regard to the rate of error in applying STRmix and the significance to the DNA testing

in this case, Dr. Krane noted testimony in the May 2018 *Daubert* hearing that the likelihood ratio

reported for Defendant might be 49 million or as little as 5 million (ECF No. 140 at PageID.3403).

He also noted reasonable concerns about the ease with which errors could be introduced through

things such as Prosecutor's Fallacy.[14] He stated that these concerns could be mitigated by a careful

---

[14] As explained by Steven Lund, "prosecutor's fallacy is a misunderstanding that when somebody
speaks to the value of or the probability of the evidence under competing hypotheses or
propositions, that they misinterpret it as the probabilistic characterization of the hypotheses
themselves given the evidence. So they're being told the probability of A assuming B is correct
but they interpret it as the probability of B assuming A is correct" (5/24/18 Hrg. Tr., ECF No. 78

explanation of the values in both the numerator and denominator of likelihood ratios. He explained the underlying reasoning for this given STRmix software modeling, noting: "Different models might yield different numerators and/or denominators and all models are just approximations" (ECF No. 140 at PageID.3403).

Dr. Krane stated there was relatively little testimony about a false positive or false negative error rate per se for STRmix at the 2018 *Daubert* hearing. But publications submitted by the Government suggest that fewer than 1 in 1000 known non-contributors to a sample would be associated with a likelihood ratio of 1000 or more. "By extension, many millions of known non-contributors might need to be evaluated before one was found to have a likelihood ratio greater than the one reported for the defendant in this case" (ECF No. 140 at PageID.3403).

With regard to standards, certification, and validation, Dr. Krane referenced his earlier discussion concerning testing, and stated that "[t]he evidence sample in this case appears to fall outside of (below) the ranges of %-contribution and quantity-of-template-contributed for which the MSP Laboratory has validated STRmix" (*id*. at PageID.3403).

As in Dr. Coble's report, Dr. Krane found no affirmative indication in the record of an external audit of the MSP laboratory, but stated that one would be expected given protocols. Likewise, Dr. Krane stated it is common practice for U.S. testing laboratories to use new approaches/tools after an internal validation, and often in dozens or even hundreds of cases before an external and independent review of the validation (*id*. at PageID.3404).

---

at PageID.2785). Lund testified that in his interactions with other employees at NIST, other scientists, "or in the courses that we have taught, we found that this is a very common tendency. That people want to think you're providing a characterization [] about the truth of the hypothesis as opposed to the plausibility of the evidence under the hypothesis, or the frequency of occurrence of the evidence under the hypothesis" (*id*. at PageID.2785-2786).

It is the Court's view that the difference of expert opinion on this *Daubert* factor, both in the reports and testimony, does not support a finding of reliability under the circumstances in this case. **There are no controlling standards governing the application of STRmix™ generally, and more troublesome is that this case involves the outer limits of a complex mixture of low-template, low level DNA, which has very limited comparable validation, if any, by the MSP lab.**

(4)   Whether the theory or technique has general acceptance in the relevant scientific community

It is Dr. Coble's opinion that the relevant scientific community is the "forensic DNA community," and given the 45 laboratory systems already online and using STRmix and another 68 at some point in their validation of the software, some with multiple laboratories, his best estimate is that nearly 90 laboratories are using STRmix in casework (ECF No. 139 at PageID.3353).   Considering additional labs using other probabilistic genotyping software (TrueAllele and LabRetriever), probably 100 laboratories—or nearly one-half of the U.S.—are using a probabilistic genotyping system. Moreover, a yearly meeting of STRmix users had about 100 attendees. In Dr. Coble's opinion, this is sufficient to show general acceptance of STRmix in the forensic DNA community. Dr. Coble additionally pointed to the large number of cases that have already presented STRmix results in the courts, and a non-exhaustive list of admissibility hearings in the U.S. involving STRmix.

Dr. Krane, on the other hand, states that there is a range of opinion on the definition of the relevant scientific community. On one extreme, it has been argued that only a very small number of individuals who have been intimately involved with the development of probabilistic genotyping should constitute the relevant scientific community. He finds this a paradoxical suggestion because most, if not all, of those individuals require that an evaluator sign a non-

disclosure agreement that would preclude any involvement in the development of probabilistic genotyping software (ECF No. 140 at PageID.3404). Quoting a 2018 letter to the editor of the Journal of Forensic Sciences (Def. Ex. E), cosigned by Dr. Krane and Nathan Adams, his report states:

> software-based probabilistic genotyping approaches (like STRmix) "are necessarily rooted in collaboration between experts in the area of molecular biology, population genetics, statistics, forensic science, computer science, and software engineering. While it is important to consider the perspective of all of these disciplines on the validation issue, we think that the perspective of software engineers [is] particularly important. Decades of experience with software failures have led to established practices for what is commonly known as verification and validation (V&V) of software. We urge that those practices be followed when evaluating PG systems."

(ECF No. 140 at PageID.3404). As such, Dr. Krane's opinion is that experts in all those disciplines (molecular biology, population genetics, statistics, forensic science, computer science, and software engineering) be considered part of the relevant scientific community that determines the general acceptance of a computer program like STRmix (*id*. at PageID.3404-3405).

Despite his advocacy for a broader-based definition of the relevant scientific community with regard to probabilistic genotyping software, Dr. Krane states that he agrees with Dr. Buckleton's testimony that, of all the probabilistic genotyping systems being used by crime laboratories in the U.S. for casework at this time, STRmix comes the closest to following the IEEE V&V standards that are embraced by software engineering professionals. But, he states:

> no one is suggesting that STRmix has adhered to IEEE V&V standards. It would be difficult to find a software engineer who maintained that a program whose output has such potential to lead to loss of liberty or life be deemed acceptable simply because alternatives were less rigorously developed. While there are credible alternatives to IEEE's V&V standards, adherence to software development/testing practices that are generally accepted [by] software-engineering professionals should not be considered a matter of stylistic preference.

(*id*. at PageID.3405).

In his testimony, Dr. Coble expanded on his position that the forensic DNA community should determine general acceptance, stating that if the issue is developmental validation of the software, "it's nice to include statisticians, coders, whatever …," but internal validation of the software "is best conducted by the people who are going to be using the software" (ECF No. 152 at PageID.4110).

Given the testimony and evidence, it is the Court's conclusion that STRmix™ does have some general acceptance in the scientific community, particularly with respect to simple mixture or "mainstream" higher quality and quantity DNA. However, the application of probabilistic genotyping software, including STRmix™, to the interpretation of complex mixture low-template, low level DNA in the manner used in this case to present a likelihood ratio in a criminal prosecution, remains controversial. This factor does not add weight for a finding that the STRmix™ DNA analysis is reliable.

Additional considerations:

At the Court's request, in light of their responses/opinions addressing the *Daubert* factors, as a concluding matter in their reports, the experts each specifically addressed the implications for the use of STRmix probabilistic genotyping software in this case. Their opinions reflect their divergent perspectives concerning the performance and outer parameters of STRmix in analyzing the complex DNA mixture from the gun swab in this case.

Dr. Coble:

I have reviewed the electropherograms, the STRmix outputs, the MSP report and the testimony from Ms. Amber Smith with regard to this case. Given my responses and opinions from points 1-4 above, I believe that STRmix was properly applied in this case. I agree with Ms. Smith's analysis that this is a 3-person mixture. I concur with her conclusion to omit the D8S1179 locus due to oversaturation. The diagnostics of the STRmix report are intuitive and there is nothing here that would give me pause to [accept] the conclusions of the MSP laboratory. Given the fact that the MSP is reporting the HPD LR, this is a conservative LR that is more favorable to the POI in this case that just reporting the "point" LR which is one

43

order of magnitude larger. It is my opinion that STRmix was properly applied in this case and the results are valid and reliable.

(ECF No. 139 at PageID.3354).

Dr. Krane:

In section 3.b. of this report, I say "The evidence sample in this case seems to fall outside of (below) the ranges of %-contribution and quantity-of-template-contributed for which the MSP Laboratory has validated STRmix." From the materials that have been provided for me to review it appears that the evidence sample (LS15-377, a gun swab) is a mixture of DNA from at least three individuals where the individual who contributed the least material is responsible for only 7% of the total DNA that was used for testing (approximately 49 pg). These values are well below the levels at which the 2016 PCAST report felt that some probabilistic genotyping systems had been foundationally validated. 49 pg corresponds to about what would be obtained from eight or nine human cells and is much less than what the test kit used by the MSP Laboratory recommends for optimum results. It would be inappropriate to assume that an approach or tool worked reliably outside of the range of values upon which it had been tested.

(ECF No. 140 at PageID.3405).

This divergence in the expert opinions on the STRmix DNA analysis in this case is underscored by the conflicting testimony and other evidence, and cannot support a conclusion of reliability under *Daubert*. In its Guidelines for Validation of Probabilistic Genotyping Systems, SWGDAM observed that an element of human interpretation is inherent in forensic DNA typing:

Human interpretation and review is required for the interpretation of forensic DNA typing results in accordance with the FBI Director's Quality Assurance Standards for Forensic DNA Testing Laboratories. Probabilistic genotyping is a tool to assist the DNA analyst in the interpretation of forensic DNA typing results. Probabilistic genotyping is not intended to replace the human evaluation of the forensic DNA typing results or the human review of the output prior to reporting.[15]

The open questions concerning STRmix™ validation and interpretation with regard to evidence in this case only further confirm the reliability determination reached.

---

[15] SWGDAM, Guidelines for the Validation of Probabilistic Genotyping Systems (June 15, 2015), https://docs.wixstatic.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf (last visited 10/14/19) (footnote omitted).

The concluding lesson from the extensive testimony and complex documentary evidence presented in this case is that the specific care required for low-template, low level DNA testing has largely faded into the background as the shortcomings of the technology and need for stringent controls on its use have been glossed over in the rush to embrace the technological advancements. With low-copy number typing, problems that can arise at every step of the sampling and testing process are amplified. As a result, a low-copy number typing profile is apt to contain greater instances of drop-in and drop-out, extreme peak imbalances and significant stutter imbalances. The high sensitivity testing that occurs with probabilistic genotyping software such as STRmix, should be undertaken with extreme care.

VI. CONCLUSION

Both court-appointed experts in this case are clearly eminently qualified to speak to the complex and difficult issues facing the Court in evaluating the MSP STRmix DNA analysis. Their insights and considerations have been invaluable to the Court in deciphering not only the highly technical scientific evidence and concepts presented in this case, but also the more abstract ramifications of probabilistic genotyping software as the next technological horizon in forensic DNA analysis. The Court has heard testimony from some of the world's foremost experts on the topics presented, including Dr. Buckleton, the father of probabilistic genotyping and co-developer of STRmix from New Zealand. Both Government counsel and defense counsel have commendably assumed the burden of educating the bench while ably presenting their cases and advocating their positions.

Based on the entire record, the Court determines that the STRmix DNA report at issue does not meet the *Daubert* reliability standard for admissibility as evidence. This decision is not an indictment of probabilistic genotyping, and certainly not of STRmix software in particular. The

Court does not question the usefulness of probabilistic genotyping software as a sophisticated tool in forensic DNA analysis. Rather, this decision is a conclusion based on the testimony and other voluminous evidence, presented over a year-and-a-half of hearings, briefing and examination by counsel, experts and the Court.

Probabilistic genotyping software is a burgeoning technology with respect to its use in forensic DNA laboratories, and more specifically, in analyzing low-copy number/low template complex mixtures of touch DNA. Forensic DNA analysis has been in use since the late 1980s. STRmix software had been in use a mere two years before its application to this case. And it had been in use by the MSP merely *three months* at the time of the DNA report and the conclusions reached by the MSP analysts—in simple terms, that the 7% contributor was 49 million times more likely to be Defendant than some unknown contributor.[16] At that time, and even currently, there are no standards in place governing the development of PGS and its use in forensic DNA analysis.

Nathan Adams testified that the motivation for further looking into software behaviors of the now-outdated 2.3.07 version of STRmix at issue here, effectively concludes with this case. But software development must adhere to accepted scientific guidelines for proper and adequate critique and analysis by those charged with this responsibility, and most especially, the courts, with their limited expertise and resources but the monumental responsibility of achieving justice for all concerned.

---

[16] Dr. Buckleton testified that STRmix first became commercially available in the United States in January 2014 (ECF No. 77 at PageID.2526).   The SWGDAM Guidelines for Validating Probabilistic Genotyping Systems were first published in June 2015 (*id*. at PageID.2644-2645). Thus, ESR completed the developmental validation or engaged in a significant amount of it and sold the product in the United States but in the absence of formal standards under SWGDAM (*id.* at PageID.2620). The DNA in this case was received by the MSP lab January 27, 2016 (ECF No. 41-7). The MSP lab first started using STRmix February 22, 2016 (*id.* at PageID.2554, 2667, 2670). The STRmix analysis was run on June 2, 2016, and the MSP Report 3 was completed June 10, 2016 (ECF Nos. 41-9, 41-10).

The Court has no doubt that this evolving technology will advance the forefront of crime investigation and other forensic analysis in the United States and around the world. But such advancements are accompanied with unique concerns when life, liberty and justice are at stake. As Government counsel stated in previewing the *Daubert* hearing testimony: "STRmix is a big step forward in that it imposes uniformity on how probabilistic genotyping is done, but [] inevitably it is a human exercise and so there is, there are human judgment calls that go into operating the software" (5/23/18 Hrg. Tr., ECF No. 77 at PageID.2505). While some or perhaps many aspects of DNA analysis have been proven reliable at this juncture of technology evolution, other aspects are merely on the cutting edge in laboratory applications and are much less tested.

The DNA evidence sought to be admitted in this case—in essence, that it is 49 million times more likely if Daniel Gissantaner is a contributor to the DNA on the gun than if he is not— is not really evidence at all. It is a combination of forensic DNA techniques, mathematical theory, statistical methods (including Monte Carlo-Markov Chain modeling, as in the Monte Carlo gambling venue), decisional theory, computer algorithms, interpretation, and subjective opinions that cannot in the circumstances of this case be said to be a reliable sum of its parts. Our system of justice requires more.

It is the Court's hope that this decision brings to light the shortcomings or, at the very least, points of inquiry necessary in evaluating this advancing technology as a tool in forensic DNA analysis. There must be a dialogue among key players in the general interest of the development and refinement of the technology, the software and its application by the individuals charged with its use in the field, rather than post-hoc testing of its reliability in the context of a criminal prosecution where the ultimate question is the freedom and guilt or innocence of the person of interest.

Here, because the sum of the parts simply does not add up to a reliable whole, the DNA analysis/likelihood ratio resulting from the use of the STRmix probabilistic genotyping software must be excluded.   Defendant's motion to exclude evidence is granted.   An Order will enter consistent with this Opinion.

Dated: October 16, 2019   /s/ Janet T. Neff
                                                             JANET T. NEFF
                                                             United States District Judge

<div align="right">Attachment 1</div>

<div align="center">**Glossary**</div>

Allele (peak): One of two or more alternative forms of a gene, a peak appears on an electropherogram for each allele that is detected.

Bayes' Rule: A mathematical equation that describes how subjective estimates of probability should be revised in light of new evidence.

Degradation: The chemical or physical breaking down of DNA.

Drop in: Detection of an allele that is not from a contributor to an evidence sample, usually due to low levels of contamination.

Drop out: Failure to detect an allele that is actually present in a sample, usually due to small amounts of starting material.

Electropherogram: The output of a genetic analyzer, typically displayed as a graph where individual peaks correspond to the presence of alleles detected in a tested sample.

Likelihood ratio: A statistic reflecting the relative probability of a particular finding under alternative theories about its origin.

Locus (pl. loci): The physical location of a gene on a chromosome.

Low-copy number (LCN)/low-template (LT) DNA: DNA test results at or below the stochastic threshold.

Monte Carlo-Markov Chain (MCMC) modeling: A computer-intensive statistical method that proposes millions of possible scenarios that might have produced the observed results, computes the probability of the observed results under each scenario, and uses the resulting distributions (and Bayes' Rule) to determine which scenarios best explain the observed results.

Random match probability (RMP): The probability that a randomly chosen unrelated individual would have a DNA profile that cannot be distinguished from that observed in an evidence sample.

Scientific Working Group on DNA Analysis Methods (SWG-DAM): A group of forensic scientists from Canada and the U.S. crime laboratories appointed by the director of the FBI to provide guidance on crime laboratory policies and practices.

Stochastic effects: Random fluctuations in testing results that can adversely influence DNA profile interpretation (e.g., exaggerated peak height imbalance, exaggerated stutter, allelic drop-out, and allelic drop-in).

STR (short tandem repeat) testing: A locus where alleles differ in the number of times that a string of four nucleotides are tandemly repeated.

Stutter: A spurious peak that is typically one repeat unit less (or more) in size than a true allele. Stutter arises during DNA amplification because of strand slippage.

Source: William C. Thompson, Laurence D. Mueller, & Dan E. Krane, *Forensic DNA Statistics: Still Controversial in Some Cases*, The Champion, Dec. 2012, pp. 17-18.