



Improving the Interpretation of Complex DNA Mixtures with Probabilistic Genotyping – A Guide to STRmix™ for Clients –

- *STRmix™ is software that helps forensic scientists reliably interpret complex mixtures of DNA*
- *The software employs the same principles that have always been used for interpreting mixed DNA profiles at the CFS*
- *Forensic scientists operate and control the software, and apply their expertise in critically evaluating results throughout the STRmix™ process*
- *Implementing the software expands the number of mixed DNA profiles that will be considered suitable for comparison*

1. Introduction

Until recently the interpretation of DNA profiles generated from crime scene evidence has relied exclusively on the training and expertise of forensic scientists, performed in accordance with established, validated procedures. Although this approach is highly effective with the vast majority of samples, some DNA mixtures are too complex to interpret in this manner.

Following validation in accordance with applicable international standards and guidelines, the Centre of Forensic Sciences (CFS) has implemented an improved method for resolving mixed DNA profiles. The method, known as probabilistic genotyping, employs a specialized software application called STRmix™.

This guide is intended to provide a general overview of STRmix™ and its application to forensic casework at the CFS.

2. Considerations in the Interpretation of DNA Mixtures

First, it would be of value to remind readers of the key principles involved in the

interpretation of DNA profiles, which apply whether or not the software-assisted approach is used.

Mixture interpretation is a multi-step process. The scientist evaluates the DNA profile observed following the test and makes various assessments based on his/her knowledge of factors known to influence test results. In addition to the principles governing the structure of DNA and its organization/packaging within human cells, these factors include:

- 2.1 How DNA reacts, on an item of evidence, to various environmental conditions from the time it is deposited to the time it is collected and preserved for eventual testing
- 2.2 How DNA reacts during the extraction and PCR¹ steps performed in the laboratory, including how very small quantities of DNA react relative to larger quantities

In consideration of these factors, and in relation to the observed DNA profile, the scientist assesses the following:

- 2.3 *Which contributions to the DNA profiling result are potential by-products of the testing process itself, also known as artefacts?* The PCR method generates known and predictable by-products which may appear as alleles (true copies of DNA fragments) to the untrained eye.
- 2.4 *How many different alleles are detected within the profile at each test site?* This indicates the minimum number of DNA contributors to the mixture.
- 2.5 *What are the quantities of each allele observed in the mixture?* Relatively speaking, alleles from a person contributing more DNA would be detected in larger quantities than alleles from a person contributing less DNA.
- 2.6 *What is the overall quality of the DNA in the sample tested?* Relative quantities of alleles may indicate the presence of substances in the sample which inhibit the PCR process or may indicate that the DNA in the sample has been degraded between the time it was deposited and the time it was collected for testing.

Through these assessments, the forensic scientist can, for relatively simple mixtures, determine the genetic makeup (i.e. genotype) or DNA profile of one or more contributors. However, in some cases, the mixtures are too complex for analysis without the assistance of a probabilistic genotyping system. For some samples, especially those with larger numbers of contributors, and especially where there are very low levels of DNA and/or degraded DNA present, the complexity of the assessments outlined above is such that a scientist is unable to draw reliable conclusions as to the makeup of some or any of the contributing DNA profiles. Clients are likely familiar with the term 'not suitable for comparison', which has been reported in these and other instances.

3. What Does STRmix™ Do?

STRmix™ is a software tool that serves as a computerized extension of the interpretation process that scientists at the CFS have always employed. It helps the scientist in assessing the factors outlined above (see 2.3 to 2.6). Unlike the scientist, however, it is able to manage the increased interpretative complexities created by combinations of larger numbers of contributors, low-level and degraded DNA. In doing so, various combinations of individual DNA profiles which could account for an observed mixture can be reliably identified and ranked by probability. This, in turn, permits a comparison to be undertaken against a known reference profile, a conclusion with respect to whether the donor of the reference profile is excluded or not, and in the latter case, an associated weight to the finding informed by both the probability that the profile in question is a constituent in the mixture and, as ever, by the rarity of the profile in the general population.

4. How Does STRmix™ Work?

Probabilistic genotyping can be thought of as a form of reverse engineering. As with all forms of reverse engineering, it involves taking a finished product and figuring out how it was produced. In this instance, the finished product is the results of an STR analysis of a complex mixture. The programme derives the various combinations of genotypes that could produce that particular mixture having regard to everything that is known of the factors which affect forensic DNA analysis. The programme also assesses the relative probabilities of the different combinations (all of which could possibly produce the observed complex mixture).

More specifically, for any given mixture, STRmix™ works by simulating DNA test results for virtual combinations of DNA profiles, comparing those simulated results against the actual test results generated in the laboratory, and assessing how well the simulated results fit the observed data². It does so hundreds of thousands of times for any given mixture (something a scientist on his or her own does not have the capacity to do), each time assessing whether the fit is better than the previous simulation. In this way, the software is able to narrow the field down to the best possible explanations (i.e. combinations of DNA profiles) for the observed test results.

In simulating test results for combinations of DNA profiles, STRmix™ relies on both DNA profile data generated by the CFS during validation studies and on scientifically accepted biological models to develop numerous permutations³ regarding how test results might manifest for specific profile combinations. By doing so, the software is effectively able to account for all of the factors that may influence test results (see 2.1 and 2.2 above).

5. What is the Scientist's Role?

STRmix™ is operated under the control of the same qualified forensic scientists who routinely interpret DNA mixtures based on the very same factors underpinning the software. Each scientist has also received specialized supplemental training with

respect to STRmix™.

Critical evaluations of DNA profile data are necessary for successful interpretations and are undertaken by scientists before and after the STRmix™ process. The knowledge, skill and judgement applied by the scientist when using the software ensures: 1) that it is appropriately used within the range of mixtures determined through validation to generate reliable results and 2) that the outcome of the mixture interpretation aligns with scientific expectations.

6. Reporting Results – The Likelihood Ratio (LR)

Readers may be familiar with the random match probability (RMP), which is used to describe the statistical significance of an association between a crime scene DNA profile and a reference DNA profile from a specific individual. The RMP, however, is not suitable for use when describing combinations of individual DNA profiles in mixtures from more than one person. In such instances, another well-established measure, the likelihood ratio (LR), must be employed.

The use of likelihood ratios in forensic DNA analysis is widespread throughout the world. They are already used by the CFS to express the significance of familial DNA analyses as well as male-specific Y-STR associations. A LR is defined as the probability of the observed DNA test results given one proposition (proposition A) divided by the probability of the same observed DNA test results given a different proposition (proposition B). A LR equal to 1 indicates that the test results are equally likely under both propositions. A LR greater than 1 provides support for proposition A and the greater the LR, the greater the support. A LR of less than 1, on the other hand, provides support for proposition B.

In the context of a two-person mixture of DNA, the propositions to be assessed may be, for example:

- A: The DNA mixture is comprised of DNA from the suspect and one unknown individual
- B: The DNA mixture is comprised of DNA from two unknown individuals

Using this approach, it is possible to weigh various propositions. This provides a more versatile and, depending on the case context, a more meaningful way of assessing the significance of the DNA evidence.

7. Validation

STRmix™ has undergone both developmental validation⁴, performed by forensic DNA experts who developed the software, as well as an in-house validation, performed by the CFS. Additionally, the software has been implemented for forensic casework in numerous accredited forensic labs throughout North America and in other parts of the world.

The Centre's in-house validation, performed in accordance with the SWGDAM guidelines⁵, entailed an assessment of hundreds of mixed samples of varying quantity, quality and numbers of contributors in order to assess how the software performed in response to these variables. Assessments were also performed, where possible, to demonstrate that results produced using the software were in accordance with interpretations undertaken by scientists without computer assistance.

Validation studies performed at the CFS and elsewhere demonstrate that the software is fit for its intended purpose as an interpretation aid for a wide range of DNA profiles, including complex mixtures of DNA. Not only is it an excellent tool to assess the statistical weight of results when an individual is not excluded from a complex mixture, the validations, which include thousands of simulations, demonstrate that STRmix™ can reliably exclude individuals who could not be contributors to these same samples.

8. Applicability / Limitations

Advances in the sensitivity of DNA testing over the past decade or so have led to the increased testing of samples containing low levels of DNA, including samples commonly touched, handled, or worn over time by multiple individuals. Not surprisingly, mixtures of DNA are also detected more frequently. There are often inherent limitations in associating such profiles, even when they can be attributed to specific people, to specific activities or time frames related to a crime.

CFS scientists assess the scientific merits of pursuing testing strategies, including the use of STRmix™, in the context of the circumstances of the case and may decide that a test will not be performed if its associated limitations outweigh its probative value.

In addition to the above stated limitations, CFS scientists continue to apply their expertise in assessing whether the complexity of a mixture prevents a valid assessment of the number of potential contributors. In these instances, especially when results suggest more than four contributors of DNA in a sample, the software may be of limited to no value as an interpretation aid.

9. Further Information

CFS scientists can provide more information about this technology and its potential applicability to investigations. Should you have questions regarding a specific case, please contact the scientist assigned. If you would like more general information, please contact the STRmix™ program manager, Linda Parker (linda.parker@ontario.ca; 647-329-1547).

¹ The Polymerase Chain Reaction (PCR) is a key step in any forensic DNA testing. It involves copying specific target areas of interest on the DNA, while also tagging these copies to facilitate subsequent detection processes.

² This process is driven by what is known as the Metropolis-Hastings algorithm. Not unique to STRmix™, it is commonly used in similar testing to assess whether a simulated result is a 'good' or 'bad' fit relative to the actual result.

³ These simulations are based on what is known as a Markov Chain Monte Carlo (MCMC) process, a widely used approach that has been used in other fields such as code breaking and engineering for many years.

⁴ Bright, J-A, D Taylor, C McGovern, S Cooper, L Russell, D Abarno and J Buckleton. 2016. *Developmental validation of STRmix™*, expert software for the interpretation of forensic DNA profiles. *Forensic Science International: Genetics* 23: 226-239.

⁵ Guidelines for Validation of Probabilistic Genotyping Systems. Approved Jun 15, 2015 by the Scientific Working Group on DNA Analysis Methods (SWGDM)