The Magazine for Digital Forensics Practitioners

Issue 57 - 2026

DIGITAL FORENSICS MAGAZINE

COURT ADMISSIBLE GAIT RECOGNITION

How a single stride became decisive biometric evidence

PLUS

- Eliminating the "Agent Obstacle" in Hyperscale Environments
- AI & LLMs in DFIR
- Modern Forensic DNA Profile Analysis & Interpretation
- Scheduled Ransomware Attacks

9 772042 060007

MODERN FORENSIC DNA PROFILE ANALYSIS

Distinguishing True Signals from False Signals

By Meng-Han Lin & Judi Morawitz

Modern forensic DNA profile analysis and interpretation processes involve multi-stage workflows. These can be summarised in three main steps: analysis; technical review; and export/further interpretation.

These stages operate through a DNA typing workflow within a laboratory. These detailed steps and requirements are the focus of the design, feature set and usability of DNA analysis software.

/Fundamental Analysis
Raw DNA profile data is non human-readable and requires the use of analysis software to produce a graphical representation of the DNA data, as well as to determine relevant metrics such as the allele and genotype, peak size, and peak heights. DNA analysis software is used to analyse raw data generated from amplified DNA, separated by capillary electrophoresis. Several core analysis algorithms in FaSTR™ DNA are adapted from those published [1] and implemented in the open-source OSIRIS analysis software. The developmental validation of FaSTR™ DNA software has been published [2]

/Analysis Rules/Thresholds
Beyond initial peak calling, the aim of DNA profile analysis is to distinguish true signals (e.g., allelic peaks) from false signals. This is complicated by the presence of undesirable artifactual peaks, such as stutter, pull-up, and dye blobs. DNA analysis software enables a trained analyst to assess the peaks that have been called in raw DNA profile data and determine whether they should be retained (likely genuine) or removed (likely artifactual). This often-manual process can be automated in some DNA analysis software, where artefact positions are automatically detected and no longer require the analyst to undertake a range of calculations relevant to different artefact peak considerations such as stutter ratio, pull-up ratio, and peak morphology.

/Stutter Model and Filter
With widespread adoption of probabilistic genotyping, it has become increasingly important to more accurately model and filter stutter peaks to help determine the number of contributors (NoC) to a profile. The ability to detect and filter stutter within DNA analysis software is helpful to analysts during analysis and brings substantial ease of use to initial manual interpretation/ contributor deductions of the profile. This facilitates downstream processes such as the estimation of the NoC to a mixed DNA sample.

For high template DNA profiles from a single contributor, distinguishing between allelic peaks and stutter artefacts is relatively straightforward. This becomes more complex when interpreting mixed DNA profiles where allelic peaks from a minor contributor can be of similar height to stutter artefacts from a major contributor since morphologically they look identical.

The appearance of stutter peaks can be characterised using a model. An example is given in Figure 1, where observed back stutter ratios at the D22S1045 locus have been plotted against allelic designation. A linear trend can be observed between allele length and stutter ratio: higher stutter ratios were obtained for larger alleles.

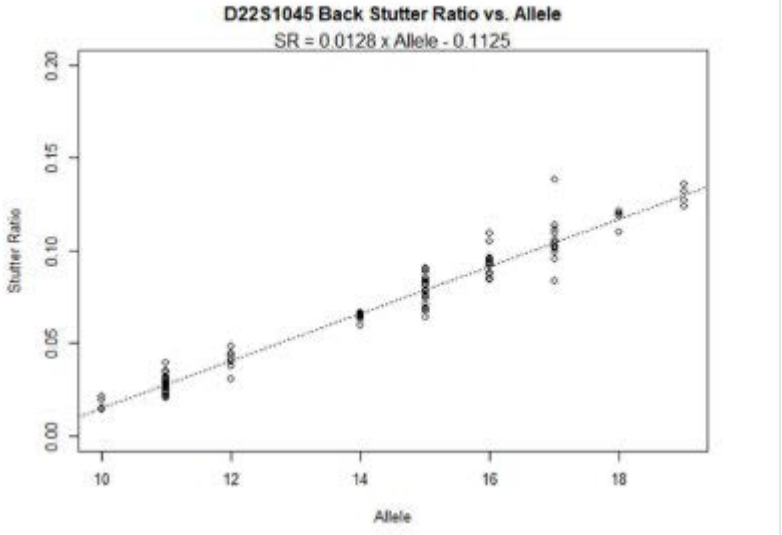


Figure 1: Allele-specific stutter regression model for back stutter at locus D22S1045.

In Figure 1, a line of best fit is drawn through the data points (using linear regression). The equation of this line can be used to calculate the expected back stutter ratio for any allele at D22S1045. Within this equation, "Allele" represents the allelic designation of the parent allele (e.g. 10, 11, 12, etc.).

The simplest approach to stutter filtering is to apply a "general" stutter filter. For example, a value of 0.15 could be used across all loci. This is easy to implement but ignores the fact that stutter ratios have been shown to be locus- and allele-specific. An updated plot of SR versus allele for D22S1045, with a "general" stutter filter applied shows that this approach correctly filters all stutter observations within the dataset (Figure 2). The smaller alleles, however, are substantially over-filtered (there is a lot of space between the SR observations and the stutter filter ratio of 0.15 for these alleles). There is also the risk that larger alleles might produce stutters that exceed the filter ratio. A more informed approach would be to use allele-specific stutter filter and flag ratios.

To filter stutter peaks more accurately, DNA analysis software like FaSTR™ DNA can use an allele-specific stutter model and filter, as described by Kalafut et al. (2018) [3]. In this method, the stutter filter adjusts dynamically according to the empirically modelled expected stutter ratio.

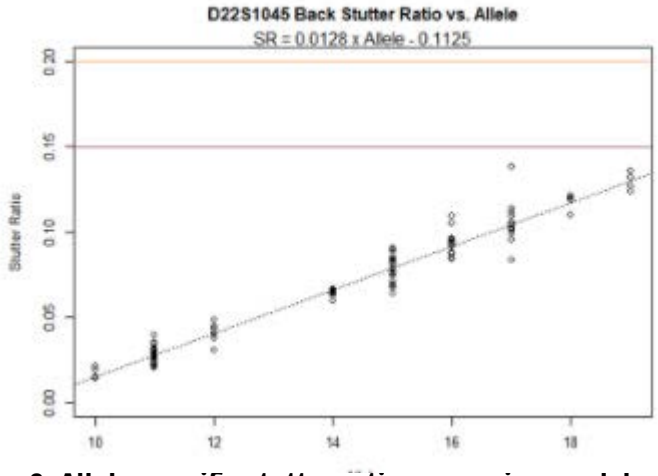


Figure 2: Allele-specific stutter ratio regression model with general filter applied at 0.15 and general flag at 0.2.

A plot of SR versus allele designation is reproduced in Figure 3, where allele-specific filter and flag ratios have now been applied. These are indicated by the green and blue dashed lines, respectively. Potential stutters below the "filter" value are automatically marked as stutter to assist with the determination of NoC.

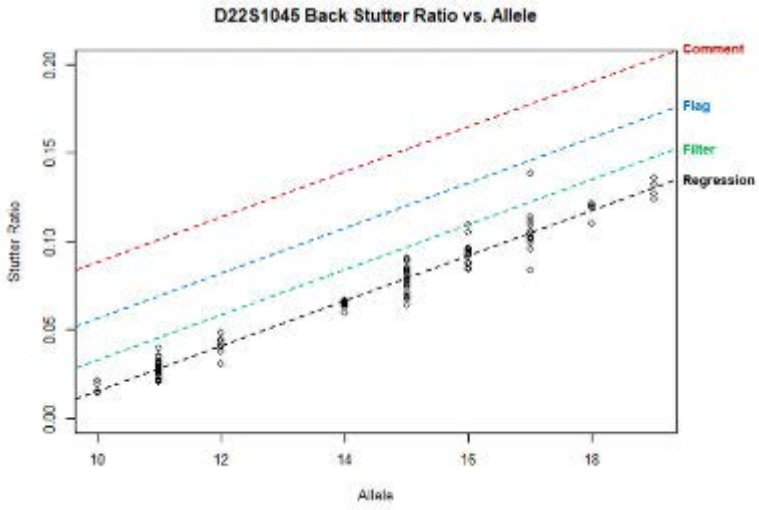
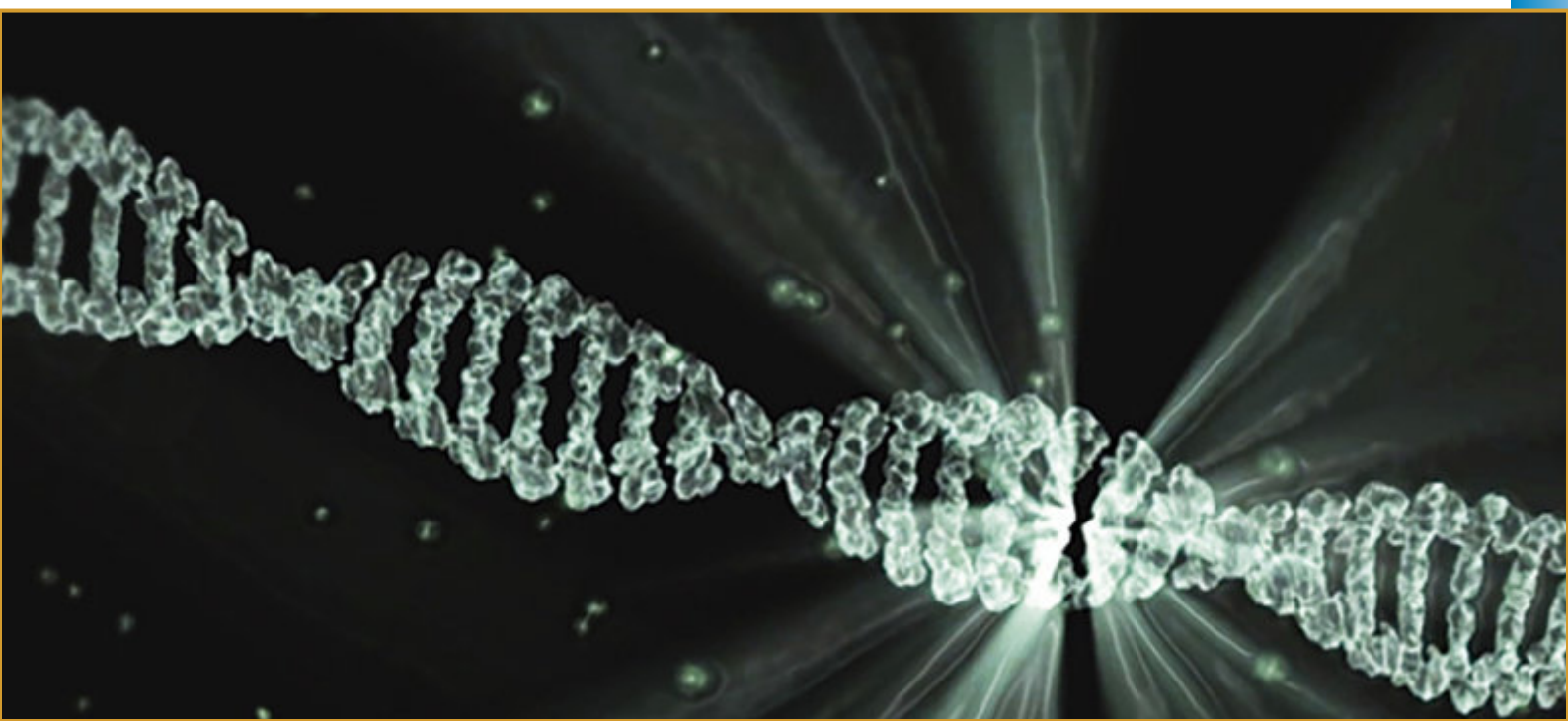


Figure 3: Allele-specific stutter ratio regression model with allele-specific stutter filter and allele-specific stutter flag levels applied.



/Composite Stutter Positions
Alleles in a profile may be positioned such that multiple stutter types can coincide to appear in a single peak position.

Expected stutter ratio considerations need to be updated for such composite peaks which consist of multiple stutter variants, as well as stutter peaks that may have been affected by spectral pull-up. FaSTR™ DNA includes options to detect and adjust stutter filter and flag ratios

for an unrestricted number of stutter types within composite stutter positions.

In the example shown in Figure 4, the 13 peak is identified as a composite peak consisting of forward stutter from the 12 allele and double back stutter from the 15 allele. The filter and flag ratios for each stutter variant have been adjusted to account for contribution to the peak height from the other stutter type. Adjusted values are indicated within FaSTR™ DNA using an asterisk (*). These calculations give substantial efficiency gains during analysis as they mitigate the need for manual calculations of stutter ratios, as well as removing the need for subsequent technical review/ replication of these calculations.

/Estimation of the Number of Contributors (NoC) to a DNA Profile
The estimation of the NoC to a forensic DNA profile is an important step in the interpretation process. Assessment of the NoC by an analyst is a time-consuming task and can be subjective. The simplest method for estimating NoC is to count the maximum number of alleles at an autosomal locus, divide by two and round up (the maximum allele count or MAC method). As well as implementing the MAC method, some DNA analysis software can also implement a relatively simple method for estimating NoC using decision trees [4]. Decision trees are fast to run and straightforward to explain.

By default, FaSTR™ DNA is provided with one trained decision tree for NoC estimation

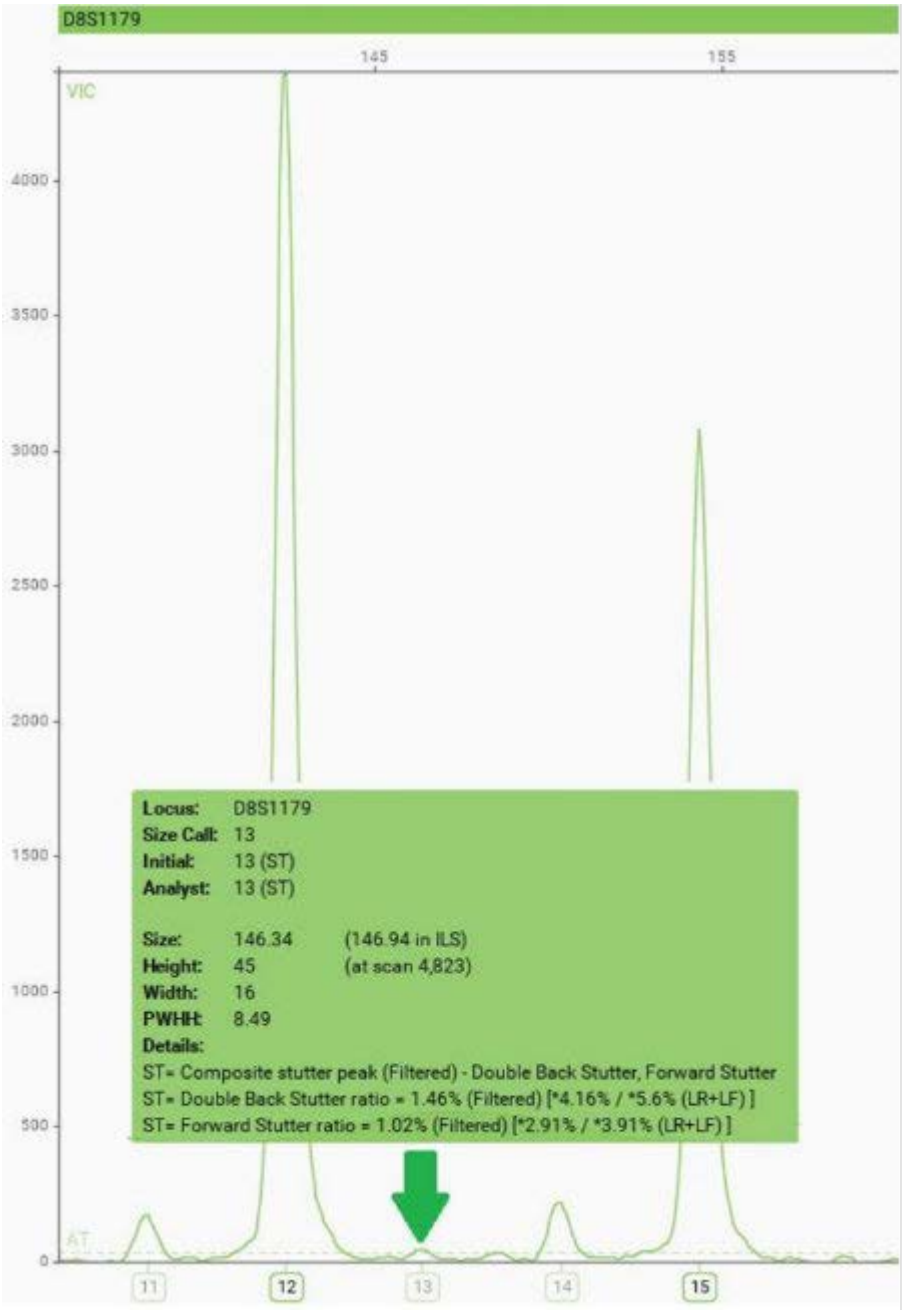


Figure 4: Composite stutter peak details for peak 13 which consists of a double back stutter and forward stutter component

of profiles generated using the GlobalFiler™ profiling kit and another for the PowerPlex® Fusion 6C profiling kit. The performance of each decision tree is summarised below.

GlobalFiler™ Decision Tree
The decision tree was trained using a dataset of 2,663 profiles comprising 1-5 contributors. The overall accuracy of the decision tree was determined to be 90%, as tested on a further 666 profiles comprising 1-5 contributors.

PowerPlex® Fusion 6C Decision Tree
The decision tree was trained using a dataset of 422 profiles comprising 1-5 contributors.

A decision tree is a flowchart where at every split one of two branches is taken depending on a test.

The overall accuracy of the decision tree was determined to be 91%, as tested on a further 280 profiles comprising 1-5 contributors.

Peak type classification	Specificity	ANN Call
Allele	All loci, all dye channels	Allele
Stutter (one repeat back stutter)	All loci, excludes gender loci and ILS dye channel	Stutter
FStutter (one repeat forward stutter)	All loci, excludes gender loci and ILS dye channel	Stutter
Half stutter (half repeat back stutter)	D1S1656 and SE33 only	Stutter
Pull-up	All loci, all dye channels	Remove
Baseline*	All loci, all dye channels	Remove

Table 1: ANN peak type classification specificity for default GlobalFiler™ ANN. (*the baseline classification is applied where it is not any other classification type)

/Artificial Neural Networks for the Classification of Peaks

Some DNA analysis software can optionally utilise kit-specific artificial neural networks (ANN) to provide peak type classifications for every peak detected and labelled.

The ANNs applied in FaSTR™ DNA, for example, are based on the implementation described within a series of publications by Taylor et al. [5-8]. ANN calls are an additional aid to the analyst where there is uncertainty over peak designation. ANN assigns a peak type classification for each peak detected by the software's analysis algorithms. The ANN is not used for peak detection and does not assign allele calls, size, or peak heights. As an example, the peak type classifications available in the trained ANN are displayed below alongside their respective default specificity (Table 1).

For each peak detected, the ANN assigns a probability to each possible classification type. The sum of the probabilities across all possible classification types for a peak is 1 or 100% (assuming no rounding). The ANN call assigned is the one with the highest combined probabilities for label (Allele and Stutter) or remove (pull-up and baseline) action.

/Workflows

Review Module for Technical Review

The technical review process for DNA data analysis often requires multiple post-analysis steps involving manual data manipulation and may require the use of different auxiliary commercial software or other software tools. These elaborate processes are restricted predominantly to comparing genotype tables and, therefore during technical review, the context of the profile and any considerations (analysis rules, artefact issues) and the profile epg itself are not readily available.

Ideally a project review module within DNA analysis software would fully integrate the technical review process and allow the analyst to access all analysis information (artefact information, peak details, sample details, analysis notes) including the epg for consideration during review.

The review module within FaSTR™ DNA enables the comparison of two projects containing the same set of samples and analysed using the same method to provide a side-by-side comparison of any differences in analysis (peak label, peak removal/retention, peak classification (allele or stutter), or peak size), with the epg profiles of both projects simultaneously in view. An example of peak differences highlighted on the epg profile is shown in Figure 5.

Differences at the sample level (NoC assignment and sample status) are also detected. The differences or conflicts can be reviewed and resolved in the review module and a final reviewed project can be created. Reports and exports can be created from within the review module and NoC can be assigned as usual. The reviewed project may also be exported directly to the PG software (STRmix™) for interpretation of the profiles.

The Project Review module (Figure 5) compares and displays the same profile as analysed by two analysts (Reader A and B) and highlights the difference in peak labelling. In this example Reader A has marked the highlighted peak as stutter, while Reader B has removed the peak.

/Comparison Module

A comparison module in DNA analysis software allows sample-to-sample comparison within a project, as well as comparison of samples within the project to a saved database. This feature can be used to investigate sample-to-sample correspondences, as well as correspondence to, for example, profiles within a staff elimination database.

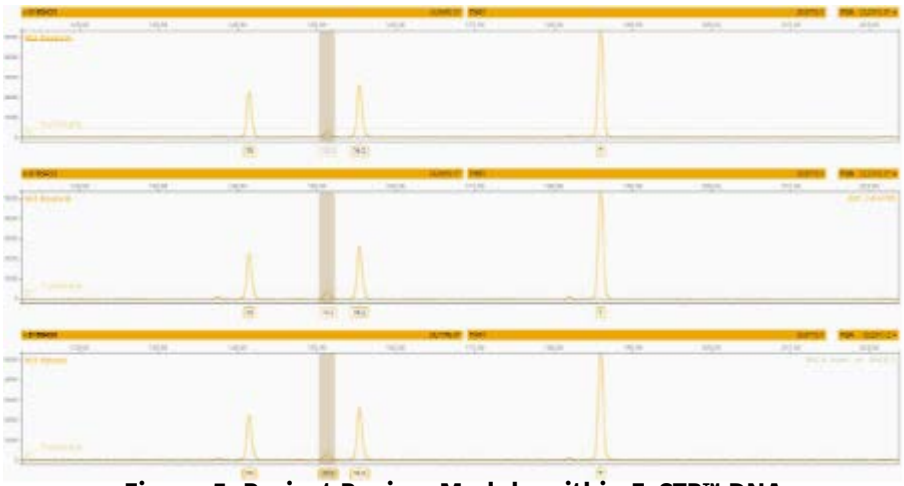


Figure 5: Project Review Module within FaSTR™ DNA

/Sample-to-Sample Comparison

Samples are compared pair-wise within a project and matches are generated where the number of alleles in common between two

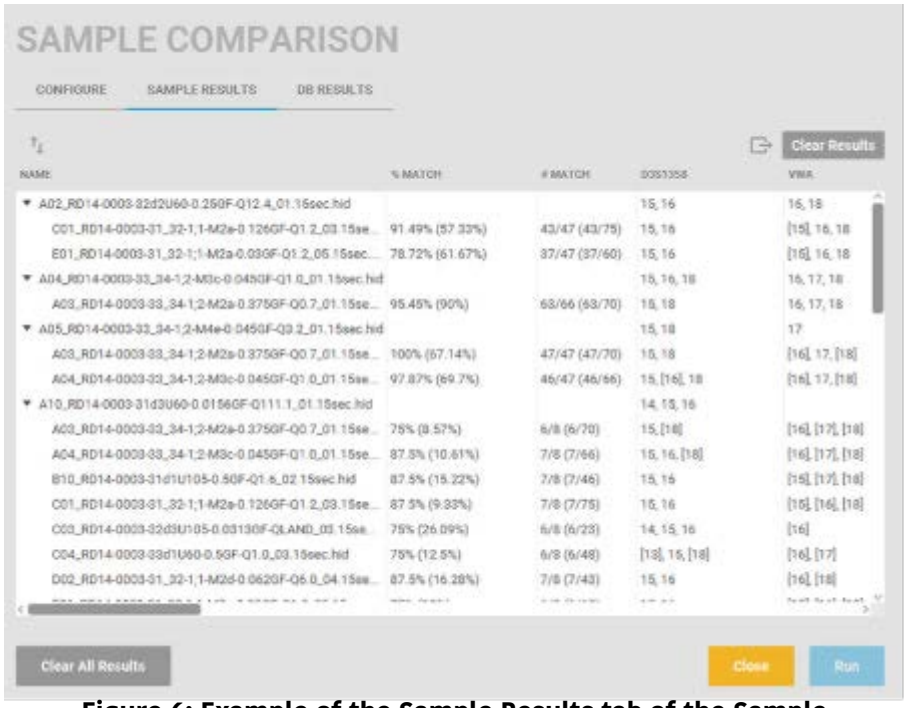


Figure 6: Example of the Sample Results tab of the Sample Comparison window

samples exceeds the match threshold configured. The matches are based on counts of peaks matching. An example of the results from a comparison module run in FaSTR™ DNA is shown in Figure 6.

An additional way to view the matches between samples is through the inspect match visualisation, where the matched peaks are highlighted in a side-by-side sample comparison view in the epg (Figure 7).

/Database Comparison

In a database comparison, each sample within a project is compared against individuals from a database, e.g., a staff elimination database or known offenders database. The allele matching mechanism can be the same as the sample-to-sample matching, but in this feature the database entries (individuals) are the subject of the comparison match. Any project samples that match (above the match threshold) to a database individual are listed as matches against that individual. An example of sample to database comparison results from FaSTR™ DNA is shown in Figure 8.

/Analysis to Mixture Interpretation

The interpretation of DNA profile data using probabilistic genotyping software is now commonplace for many forensic laboratories. A software-based solution that integrates this link between DNA profile analysis and interpretation would streamline the DNA typing and interpretation workflow by reducing setup requirements (for interpretation) and minimising manual handling by the analyst.

Within FaSTR™ DNA, once samples are analysed to completion and

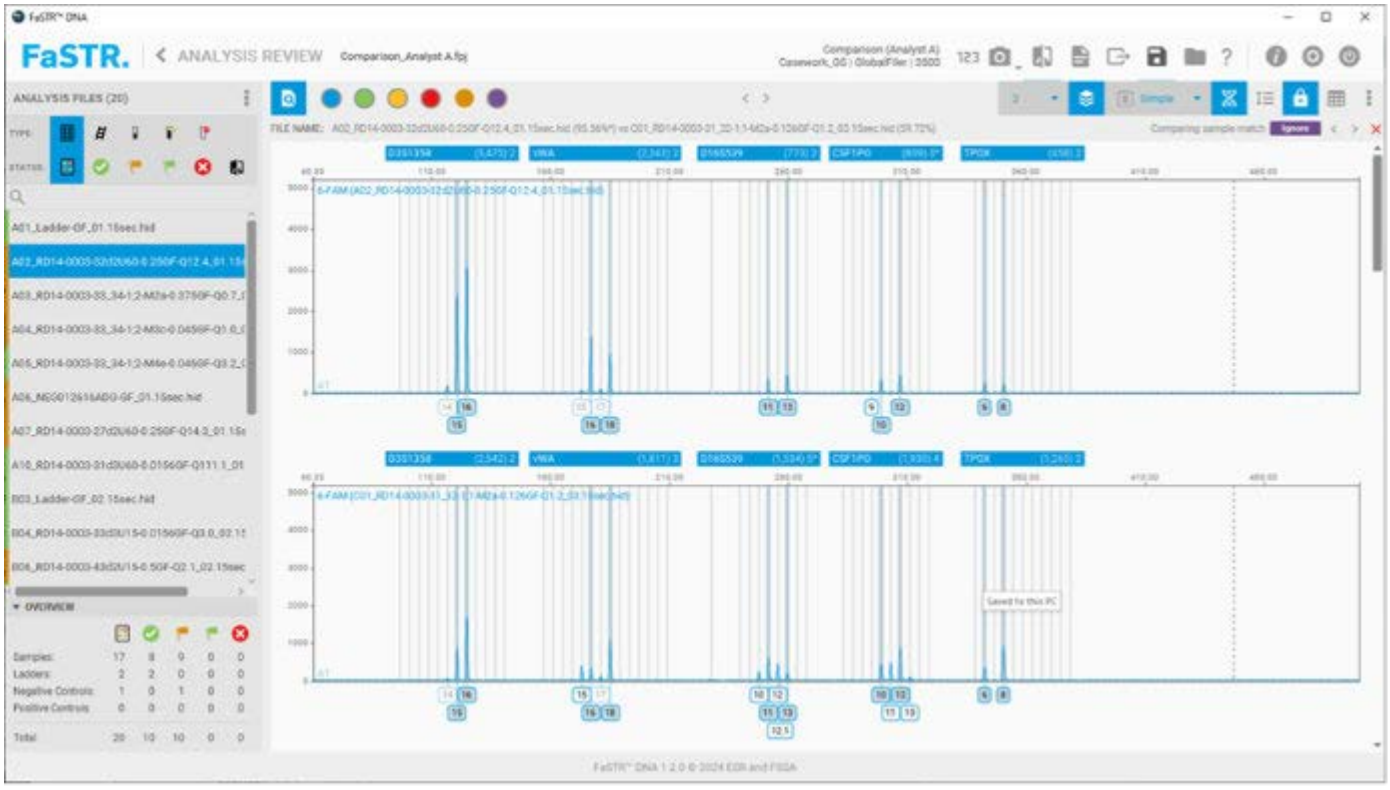


Figure 7: The sample to sample epg comparison view (Inspect Match)

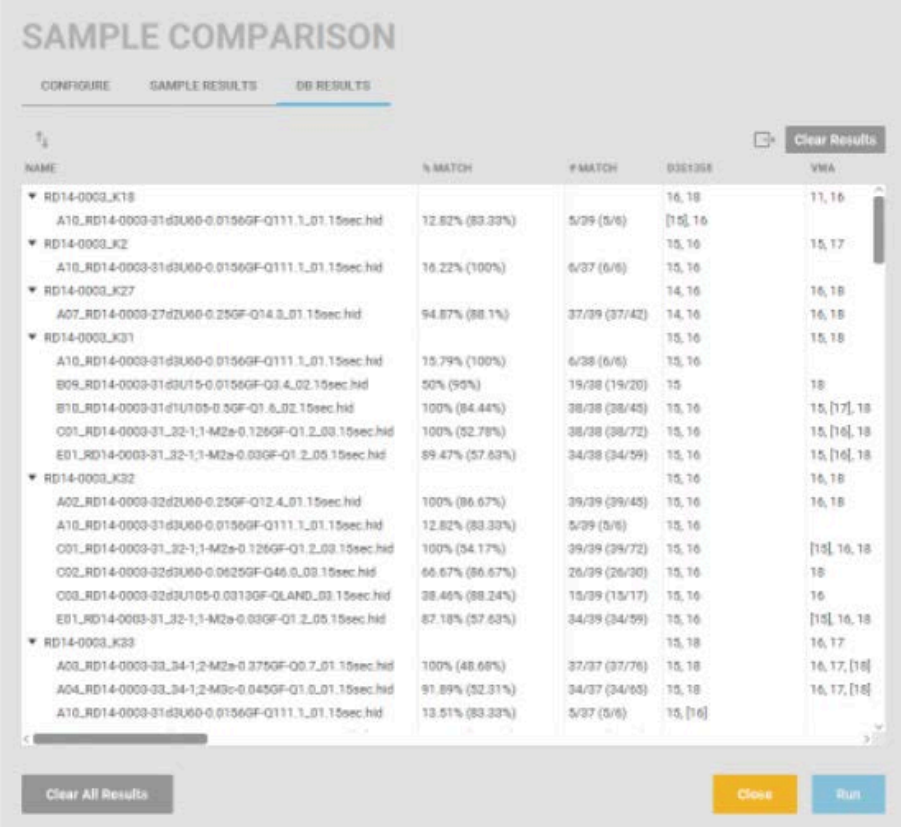


Figure 8: Example of the DB (database) Results tab of the Sample Comparison window

the NoC assigned, the project can then be exported to STRmix™ for either an automated deconvolution batch, or for further setup of an interpretation batch with references for an LR calculation.

/Conclusion

The advances in DNA analysis and interpretation have led to ever-increasing applications for forensic DNA profile data. DNA analysis and interpretation technology, software, and workflows are critical to facilitating the use of such data to its full potential. Although the specific process may vary among individual laboratories, the feature set, usability, and advanced analyses available within FaSTR™ DNA analysis software have been designed to address these requirements.

/References:

1. R.M. Goor, L. Forman Neall, D. Hoffman, S.T. Sherry, A Mathematical Approach to the Analysis of Multiplex DNA Profiles, Bulletin of Mathematical Biology 73(8) (2011) 1909-1931.
2. M.-H. Lin, S.-I. Lee, X. Zhang, L. Russell, H. Kelly, K. Cheng, S. Cooper, R. Wivell, Z. Kerr, J. Morawitz, Developmental validation of FaSTR™ DNA: Software for the analysis of forensic DNA profiles, Forensic Science International: Reports 3 (2021) 100217.
3. T. Kalafut, C. Schuerman, J. Sutton, T. Faris, L. Armogida, J.-A. Bright, J. Buckleton, D. Taylor, Implementation and validation of an improved allele specific stutter filtering method for electropherogram interpretation, Forensic Science International: Genetics 35 (2018) 50-56.
4. M. Kruijver, H. Kelly, K. Cheng, M.-H. Lin, J. Morawitz, L. Russell, J. Buckleton, J.-A. Bright, Estimating the number of contributors to a DNA profile using decision trees, Forensic Science International: Genetics 50 (2021) 102407.
5. D. Taylor, Using a multi-head, convolutional neural network with data augmentation to improve electropherogram classification performance, Forensic Science International: Genetics 56 (2022) 102605.
6. D. Taylor, A. Harrison, D. Powers, An artificial neural network system to identify alleles in reference electropherograms, Forensic Science International: Genetics 30 (2017) 114-126.
7. D. Taylor, M. Kitselaar, D. Powers, The generalisability of artificial neural networks used to classify electrophoretic data produced under different conditions, Forensic Science International: Genetics 38 (2019) 181-184.
8. D. Taylor, D. Powers, Teaching artificial intelligence to read electropherograms, Forensic Science International: Genetics 25 (2016) 10-18.

Meng-Han Lin is a Senior Scientist at The New Zealand Institute for Public Health and Forensic Science Limited (PHF Science) and has been employed in the STRmix™ team since 2016. A molecular biologist by background, his role is primarily responsible for the development, validation, and support of FaSTR™ DNA analysis software.

Judi Morawitz was previously a scientist in The New Zealand Institute for Public Health and Forensic Science Limited (PHF Science). On her return she has taken on the role of integration lead of software within the STRmix team. She has been actively involved in the design and development of the FaSTR™ DNA software.