



## Completely automated interpretation of reference samples

Volker Weirich

State Office of Criminal Investigation Mecklenburg-Vorpommern, Division 5, Department 54, Retzendorfer Straße 9, 19067 Rampe, Germany



### ARTICLE INFO

**Keywords:**  
Forensic DNA  
Probabilistic genotyping  
Automated interpretation

### ABSTRACT

Inspired by fully continuous interpretation a completely automated interpretation workflow for reference samples was implemented in our lab. It's based on a quite simple model, which includes DNA amount and degradation as well as backward and forward stutters. The current implementation can handle CE results of any number of replicates, even from different autosomal or Y-chromosomal STR kits. Results, lessons learned in a first set of test series and potential future extensions are discussed.

### 1. Introduction

Both subjective effects on EPG interpretation [1] as well as increasing number of cases raised the need for implementing an automated interpretation in our lab. Several software solutions based on full-continuous models clearly show the potential of the so-called probabilistic genotyping [2–5]. While none of them perfectly fits our lab's needs, an own approach was developed as an additional feature of Statistefix [6] as the core functionality.

### 2. Material studied, methods, techniques

Basically, there are five steps, which can be automated separately:

- 1 Import of *fsa* and / or *hid* files into GeneMapper ID-X (GMIDX; Thermo Fisher Scientific) using its command line interface (CLI)
- 2 Analyses of these files in GMIDX with Analysis Methods without any filters using CLI (with versions earlier than 1.6 the creation of separate projects with specialized Analysis Methods is recommended; there is no need for this anymore with version 1.6 because of its new "Export Table With Stutter"-option)
- 3 Export of the GMIDX results as *csv* or *txt* file, again using its CLI
- 4 Automated interpretation using the above mentioned new Statistefix-feature
- 5 Automated generation of database records including QR-codes for National DNA Database

For step 5 an AddIn for Word 2013 and newer (Microsoft Corporation) was developed.

Steps 1 to 3 and step 5 address mainly technical issues. Core functionality is realized by probabilistic genotyping within Statistefix. It's based on a quite simple model that takes DNA amount and degree of

degradation as well as backward and forward stutters into account. The current version is able to use any number of replicates, even from different STR kits.

### 3. Results and discussion

A first test series with all reference samples from case work between July 2018 and June 2019 was started. For these 555 persons 17 STR loci including Amelogenin had to be analyzed. Hence, 18,870 alleles had to be reported. While 18,856 (99.926%) automated allele calls were concordant with classical interpretation by an experienced DNA expert, only 14 discordant allele calls were observed: 7 allele calls were automatically classified as "ambiguous", the remaining 7 allele calls were erroneous (0.037% each).

Most of the discordant allele calls were caused by alleles not present in the bin sets provided by the kit vendors. Therefore, an Excel macro was developed to fill up vendors' bin sets with appropriate virtual bins. Using these new bin sets to analyze the same 555 reference samples in a second test series, only four discordant allele calls (0.02%) were observed: Three of these allele calls were caused by extreme locus imbalances, probably due to primer binding site mutations, the fourth one was caused by a huge bleed through, accompanied by a failed size calling of the EPG of the complement second kit.

Given those useful results even in cases of poor-quality reference samples, an additional test series for crime scene samples was performed on lab's backlog data, without EPG inspection by an expert. Basically, the model is extended by one variable only: the number of contributors. Above mentioned parameters (DNA amount, degree of degradation, possible genotypes) are all handled individually. In a serial analysis 5834 *hid* files of 3167 crime scene samples were analyzed using Statistefix. After 30 h 901 complete profiles (17 loci) were deduced automatically, from both single stains and mixtures. These

E-mail address: [volker.weirich@lka.mv.de](mailto:volker.weirich@lka.mv.de).

<https://doi.org/10.1016/j.figss.2019.09.027>

Received 20 September 2019; Accepted 21 September 2019

Available online 23 September 2019

1875-1768/ © 2019 Elsevier B.V. All rights reserved.

profiles were submitted to the National DNA Database and searched for matches; causing a hit ratio of 28.7%. Even incomplete DNA profiles (> = 10 complete loci), generated from poor quality stains, provided valid hits in the National DNA Database.

#### 4. Conclusion

In conclusion, although the above-mentioned probabilistic interpretation is still to improve, it is a very useful tool in our lab.

#### Declaration of Competing Interest

The author of this manuscript certifies that he has no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

#### Acknowledgments

The author thanks Øyvind Bleka, Charles Brenner and the STRmix crew (no particular order) for valuable discussions.

#### References<sup>1</sup>

- [1] I.E. Dror, G. Hampikian, Subjectivity and bias in forensic DNA mixture interpretation, *Sci. Justice* 51 (2011) 204–208.
- [2] M.D. Coble, J.A. Bright, Probabilistic genotyping software: an overview, *Forensic Sci. Int. Genet.* 38 (2019) 219–224.
- [3] Ø. Bleka, P. Gill, EuroForMix: <http://www.euroformix.com/>.
- [4] C.H. Brenner, DNA-view Mixture solution: <http://dna-view.com/>.
- [5] Institute of Environmental Science and Research New Zealand, Forensic Science South Australia, STRmix: <https://www.strmix.com/>.
- [6] V. Weirich, Statistefix: <https://www.statistefix.de/>.

---

<sup>1</sup> Refs. [3–6] are Web References.