Forensic Science International: Genetics Supplement Series xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/fsigss

GenoProof Mixture 3—New software and process to resolve complex DNA mixtures

Frank M. Götz[a,*], Holger Schönborn[a], Viktoria Borsdorf[a], Anne-Marie Pflugbeil[b], Dirk Labudde[b]

[a] qualitype GmbH, Moritzburger Weg 67, Germany
[b] University of Applied Sciences, Mittweida, Germany

ARTICLE INFO

Keywords:
DNA interpretation
Mixture interpretation
Continuous models
Probabilistic genotyping

ABSTRACT

GenoProof Mixture 3 is an expert system for interpretation of complex DNA mixtures. The implementation of a fully continuous model enables the usage of all available profile information and increases the analysis capabilities of the software strongly. First validation data show the advantages of the fully continuous approach compared with other models. This paper provides an overview of the software and the validation results.

1. Introduction

Improvements of the sensitivity of DNA analysis techniques allow a growing number of analyzable DNA samples to be analyzed and encourage the usage of previously unresolvable traces like low template DNA samples. As a result observations of mixtures increase [1]. Resolving these complex DNA mixtures creates a new challenge for DNA experts. This technological progress has been accompanied by a further development of statistical methods and recommendations for mixture interpretation [2,3]. Today different statistical methods like binary, semi-continuous and fully continuous models are used for the interpretation of DNA mixtures. The binary model only considers whether an allele is present in a mixture or not and fails to consider drop-in/drop-out events. The semi-continuous approach is allowing for drop-in/drop-out events, while the fully-continuous model additionally takes the biological parameters peak height, pre-stutter ratio, and fragment size into consideration utilizing as much available information as possible within a profile.

Accompanied by the ongoing development in mixture interpretation we developed GenoProof Mixture supporting a binary, semi-continuous and fully continuous approach for mixture interpretation in one software. The paper presents a short overview of the software and the validation of the fully continuous model.

2. Material and methods

2.1. Probabilistic models

We implemented two different probabilistic genotyping approaches for DNA mixture interpretation: First, the discrete or semi-continuous model, which uses information of present alleles considering drop-in/dropout events and second, the fully continuous model, which additionally takes biological parameters like peak heights and pre-stutter ratios into consideration. We conducted an extensive study of probabilistic methods for mixture analysis considering stutter events, drop-in/drop-out events and peak heights [4–6]. Based on this evaluation we designed and implemented algorithms for a fully-continuous model.

The basic principle of the model is the estimation of the parameters DNA degradation, amplification efficiency, replica multiplier, DNA amount via a random process. In doing so, the likelihood of a match of the expected peak heights (generated by random genotype constellations and random parameter combinations) with the observed peak heights is determined.

During each iteration step the acceptance of parameters is reviewed. For this purpose, tally charts for each genotype constellation are listed. Parameters are accepted, if the likelihood is improved compared to the previous step or is in compliance with the tolerance value.

Subsequently, if the combination is accepted, the tally chart of the genotype constellation is raised by one. If the parameters have not been accepted, the tally chart of the previous genotype constellation is raised. The algorithm creates all possible genotype constellations, whereas all not justifiable genotype constellations are filtered by heuristics. While doing so, Markov chains are executed, using the first 10.000 acceptance steps as the transient phase. After completing the transient phase, only the genotype constellations are varied in the next 40.000 acceptance steps. By normalizing the genotype constellation checklist, the weighting of genotypes is accomplished and the likelihood of the hypothesis and the LR are calculated. In order to assess the

* Corresponding author.
E-mail address: f.goetz@qualitype.de (F.M. Götz).

http://dx.doi.org/10.1016/j.fsigss.2017.09.212
Received 8 September 2017; Received in revised form 25 September 2017; Accepted 26 September 2017

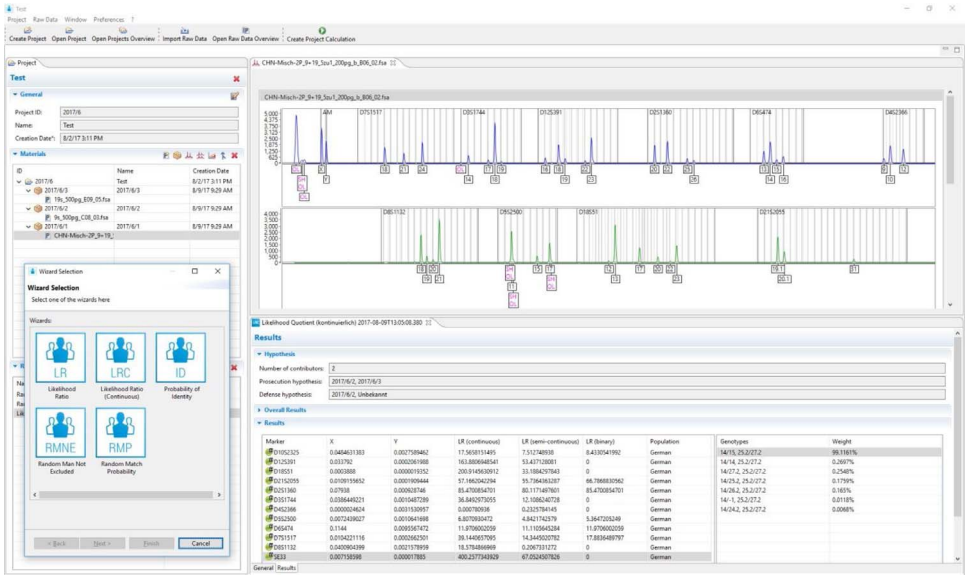

Fig. 1. GenoProof Mixture 3 User Interface.

convergence of parameter and the quality of a calculation, we included a graphical monitoring tool for convergence diagnostics.

2.2. Software implementation

GenoProof Mixture (Fig. 1) is designed as a complete solution for analysis of forensic samples and complex DNA mixtures. It covers the entire evaluation process from raw data analysis and statistic calculations of likelihood ratios.

Users can directly import the raw data (fsa/hid files) of their samples and start the data interpretation. Profiles from replicate analyses with different test kits can be combined in one interpretation. GenoProof Mixture provides probabilistic and non-probabilistic models (probability of identity, RMP, RMNE, binary LR) for mixture interpretation. For LR calculation users can define their hypotheses for the statistical analysis. The major improvement of GenoProof Mixture 3 is the implementation of a fully continuous model for mixture interpretation. It includes the weighting of contributor genotype constellations as a central part of our evaluation. The weights of all possible genotype constellations that may explain a given DNA mixture are calculated using a Markov chain Monte Carlo (MCMC) method. This method considers peak heights, stutter quotients and allele drop-in/drop-out to make full use of available information within a DNA profile. Algorithms were optimized to deliver an excellent runtime behavior also on ordinary desktop computers. Providing quality and transparency of your results, we included graphical representations to assess result quality as the base for a clear and correct explanation in court.

2.3. Validation study

We conducted a developmental validation of our probabilistic genotyping system based on test data to verify the functionality of the system and to determine limitations. We performed our validation according to SWGDAM guidelines for validation of probabilistic genotyping systems [7].

Several single source DNA samples were analyzed using Mentype Chimera multiplex kit (Biotype Diagnostic GmbH, Dresden, Germany). The samples were saliva stains on swabs (4N6FLOQSwabs Genetics, Copan, Brescia, Italy) and DNA was isolated with QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). Amplified products were separated on an Applied Biosystems PRISM 3500 Genetic Analyser (Thermo Fisher, Waltham, MA) and data was analyzed using GenoProof Mixture 3 (Qualitype GmbH, Dresden, Germany). Utilizing this single source DNA we created several 2, 3 and 4-person DNA mixtures in different ratios and DNA concentrations (25, 50, 75,100, 125, 150, 200, 250, 400 pg) and evaluated them with GenoProof Mixture 3.

3. Results

All mixtures were analyzed three times. In total 225 samples were used to evaluate the software sensitivity, specificity, precision and accuracy (Table 1).

As shown in Table 1 all results were compared against the binary model and semi-continuous model and resulted in increased LR values utilizing the fully continuous approach. We performed the evaluation as an examination of mixed samples of known contributors testing true contributors (that produced a large LR in all cases) and non-contributors (that could be excluded in all cases).

We analyzed the samples 10 times to demonstrate reproducibility. The produced LR for a two person mixture (200 pg, 3:1 ratio, Table 1 italic) varied between $1.42 \cdot 10^{15}$ and $3.57 \cdot 10^{15}$ (mean LR $2.04 \cdot 10^{15}$). The average simulation time was 77 s. Modification of model parameters such as increasing the number of iterations of the MCMC method can further improve reproducibility. However, it is accompanied by an increase of time needed for simulation.

In addition to our validation study a prototype version of our software has been tested on real case samples by several forensic labs.

4. Conclusion

GenoProof Mixture 3 has been developed as mixture interpretation software providing all established statistical methods for mixture interpretation. We were able to increase the mixture interpretation capabilities of the software with implementation of a fully continuous model. The first validation study has shown strongly improved LRs for all analyzed mixtures and exclusion for all tested non-contributors. Additional validation studies and case work studies in several labs are in progress.

The GenoProof Mixture website (https://www.qualitype.de/en/solutions/products/evaluation-software/genoproof-mixture/genoproof-mixture-3/) offers detailed information about the software.

Conflict of interest statement

Frank Götz declares no conflicts.

Frank Götz is Managing Director & Partner of qualitype GmbH