DEVELOPMENTAL VALIDATION

GenoProof Mixture 3.0.4

qualitype GmbH | Moritzburger Weg 67 | 01109 Dresden | Germany

Tel +49 351 8838 2800 | Fax +49 351 8838 2809 | sales@qualitype.de | www.qualitype.de

IT works.

Editorial Work

A.-M. Pflugbeil

Edition 1, Dresden, 2018

Version 2018-04-17

Production and Sales

qualitype GmbH
Moritzburger Weg 67
01109 Dresden, Germany
Tel.: +49 351 8838 2800

Legal Information and Trademarks

qualitype GmbH continually develops and enhances its products. Hence the information in this document may be changed without prior notice due to product improvements, standardizations or technical reasons.

For further information please read the Terms and Conditions and the Software License Terms of qualitype GmbH.

Content

1 Introduction to probabilistic genotyping

Improvements of the sensitivity of DNA analysis techniques allow a growing number of DNA samples to be analysed and encourage the use of previously unresolvable traces like low template DNA (LTDNA) samples [3]. As a result, observations of mixtures increase. These complex traces still present an ongoing challenge to molecular forensics interpretation and biostatistical assessment. In addition to the presence of several contributors, specific phenomena, like allele drop-in or drop-out effects can occur, due to low quality and quantity of the extracted DNA. As a result, the evidential weight decreases and no biostatistically assured statement is possible by applying conventional or deterministic models. Therefore, current research in the field of DNA interpretation methods moves towards continuous interpretation strategies.

Introducing semi- and fully continuous (probabilistic) methods potentially increases the efficiency of forensic laboratories, and improves the consistency and transparency of the reported results. So-called probabilistic methods or probabilistic genotyping uses biological modelling, statistical theory, algorithms, and probability distributions to calculate likelihood ratios (LRs) and/or infer genotypes from the DNA typing results of forensic samples. The software primarily supports the expert-based evaluation process, but does not replace the expert. A probabilistic genotyping system is comprised of software, or software and hardware, with analytical and statistical functions that include complex formulae and algorithms. Particularly useful for low-level DNA samples (i.e., due to DNA quantity stochastic effects may be observed for individuals) and complex mixtures (i.e., multi-contributor samples, exhibiting allele sharing and/or stochastic effects), probabilistic genotyping approaches can reduce subjectivity in the analysis of DNA typing results. Deterministic methods of mixture interpretation consider all interpreted genotype combinations to be equally probable, whereas probabilistic approaches provide a statistical weighting to the different genotype combinations. By using more genotyping information when performing statistical calculations and evaluating potential DNA contributors, probabilistic genotyping enhances the ability to distinguish between true contributors and non-contributors. A higher likelihood ratio (LR) is typically obtained when evaluating a person of interest (POI) who is a true contributor to the evidence profile, whereas a LR lower than one indicates that the POI is not a true contributor. While the absence of an allele or the presence of additional allele(s) compared to a reference sample may support exclusion (binary model), probabilistic genotyping approaches consider inclusion and exclusion hypotheses by calculating the LR incorporating allele drop-outs and drop-ins [1].

The use of the likelihood ratio as a reporting statistic for probabilistic genotyping differs substantially from binary statistic values such as the combined probability of exclusion. Developmental validation

of a probabilistic genotyping system requires test data to verify the functionality of the system, the accuracy of statistical calculations and other results, the appropriateness of analytical and statistical parameters, and determines its limitations.

The developmental validation of GenoProof Mixture 3 is based on the "SWGDAM Guidelines for Validation of Probabilistic Genotyping Systems" (FINAL APPROVED 06/15/2015) and "DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications" [1, 2]. The validation of GenoProof Mixture 3 intends to analyze the sensitivity, specificity and precision of the implemented probabilistic model (fully continuous approach). The following chapters discuss the design and the achieved results of the validation study.

1.1 Scientific model of GenoProof Mixture 3

Common probabilistic genotyping approaches, such as fully continuous approaches, incorporate empirically determined biological models and enable the use of more detailed profile information, thereby reducing subjectivity in the interpretation of results. In contrast to the binary and discrete model, peak height information of drop-in and drop-out events or other artefacts are considered in the calculation with the fully continuous model. Thus, the informative value of the LR increases.

To conquer the ongoing challenges in mixture interpretation, we developed GenoProof Mixture 3 [4]. The software provides probabilistic (semi- and fully continuous model) and non-probabilistic models (probability of identity, RMP, RMNE) for mixture interpretation. The evaluation of mixed traces with the fully continuous approach is based on the statistical algorithm described by Taylor et al. The fully continuous model is based on the modelling of peak heights of genotype constellations including various features like DNA quantity, DNA degradation, allele peak height variance, amplification efficiency, homozygosity and heterozygosity, which depend on contributors, shedder status, DNA markers, test kits and analysis devices.

Based on the total allelic product (T_{anr}^{l}) (Equation 1) and the pre-stutter ratio of an allele, the expected peak height and the expected pre-stutter peak height are calculated [5]:

$$T_{anr}^{l} = A_{r}^{l} \times t_{n} \times e^{-d_{n} \times m_{a}^{l}} \times X_{an}^{l}$$

Equation 1: Total allelic product (biological model) [5].

T_{anr}^{l} Total allelic product of allele a, person n, marker l

A_{r}^{l} Amplification efficiency of marker l and replicate r

t_{n} DNA amount of person n

d_n	DNA degradation of person n
m_a^l	Size of allele a in marker l
X_{an}^l	2, if allele a of person n in marker l is homozygous, otherwise 1

Both, person-dependent (DNA degradation and DNA template) and replicate-dependent parameters, are simulated for the calculation of the total allelic product. A previously performed training of each parameter that is considered in the biological model is not necessary in GenoProof Mixture 3. This eliminates a time-consuming revalidation when changes in the lab process occur. The pre-stutter quotient corresponds to half of the pre-stutter limit of the test kits defined by the user. The allele variance constant varies within the given limits (uniform distribution), whereas the default value of variance, referring to the amplification efficiency, was determined empirically.

The estimation of the parameters, included in the biological model, and the weights of all possible genotype constellations that may explain a given DNA mixture are calculated using the so-called Markov Chain Monte Carlo (MCMC) method. This method considers peak heights, stutter quotients and allele drop-in/ drop-outs to make full use of available DNA profile information. 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. The algorithm runs through all possible genotype constellations, whereas all unjustifiable genotype constellations are filtered by heuristics. In each iteration step, the acceptance of parameters is reviewed. GenoProof Mixture 3 uses the Metropolis-Hastings algorithm to reject or accept genotype constellations via tally chart listings [6, 7]. Parameters are accepted, if the likelihood is improved compared to the previous constellation or complies with the limit of tolerance. Subsequently, if the combination has been 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. While doing so, Markov chains are executed, using the first 20.000 acceptance steps (default) as the transient phase. After completing the transient phase, only the genotype constellations are varied in the next 100.000 acceptance steps (default). Where a reference profile from a person of interest (POI) is available, the LR can be calculated, including the weighting of possible genotypes. According to the definition, the LR describes the probability of obtaining the evidence profile(s) under the assumption of two competing hypotheses; the prosecution hypothesis (Hp) and defence hypothesis (Hd).

$$LR = \frac{\Pr(E|H_p)}{\Pr(E|H_d)}$$

Equation 2: Likelihood Ratio (LR). The equation describes the evidential support for the identification hypothesis that a suspect contributed their DNA to the biological evidence, whereas Hp corresponds to the hypothesis of the prosecution and Hd corresponds to the hypothesis of the defence.

GenoProof Mixture 3 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, genotype deconvolution and statistic calculations of likelihood ratios. Users can directly import both, raw data and already analysed profiles (fsa/hid files), for data interpretation with the software. The raw data analysis has been part of the product portfolio for the evaluation of DNA profiles for many years and has already been sufficiently tested.

The software also enables the reanalysis of profiles. That means, PCR replicates created either with the same STR kit or different STR kits can be used for LR calculations. Upon completion of the calculation, the convergence diagnostics will be displayed. It serves as a quality control of the LR calculation and has two views: Running Mean Plot and Trace Plot. The Running Mean Plot visualises the running mean of the parameters. It shows if the convergence of the parameters was successful. If the mean is not constant, but slightly rises, convergence has not been reached. In this case, the calculation should be repeated with an increased number of iterations (Fig. 1). The Trace Plot visualises the same parameters as the running mean plot, but shows the actual values for the different parameters of each iteration (Fig. 2). The graphs should not display any irregularities as these indicate problems during the parameterisation of the model and poor mixing of the parameters. In this case, the settings of the parameters should be examined.

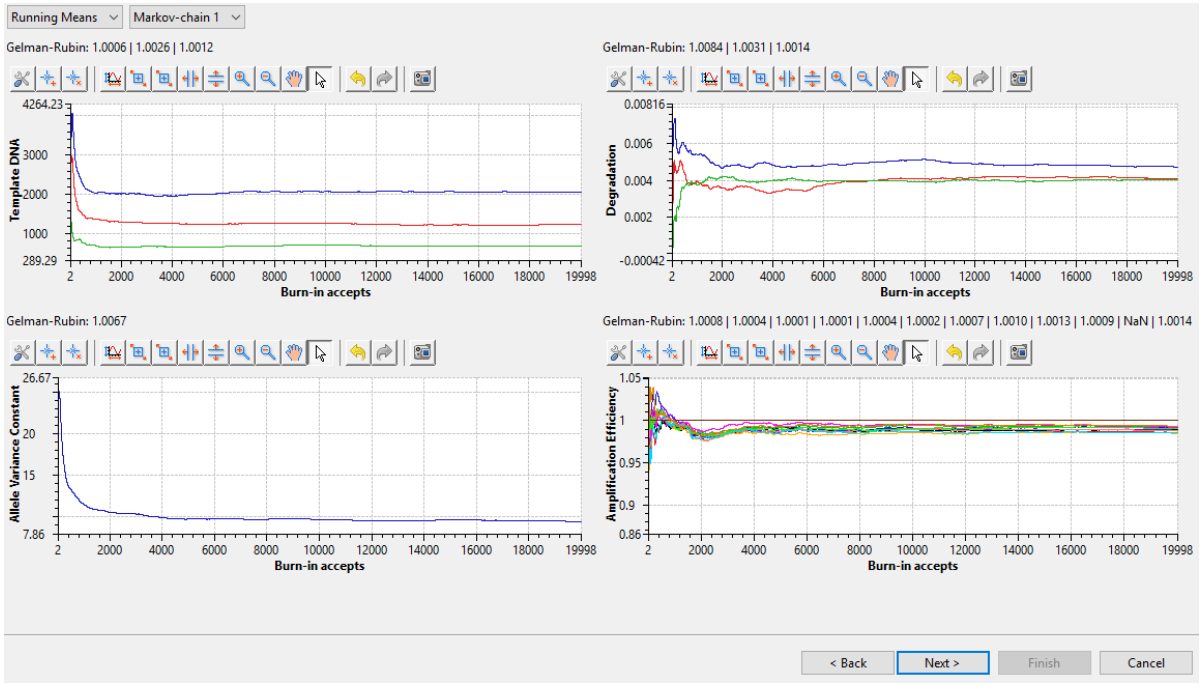

Figure 1: Running Mean plot. The figure illustrates the diagnosis of convergence regarding a three person mixture. Based on the graphs the convergence for each parameter during burn-in can be verified (DNA template, degradation, allele variance constant, amplification efficiency). The blue, red and green line in the plot "Template DNA" and "Degradation" refers to the previously defined number of possible contributors.

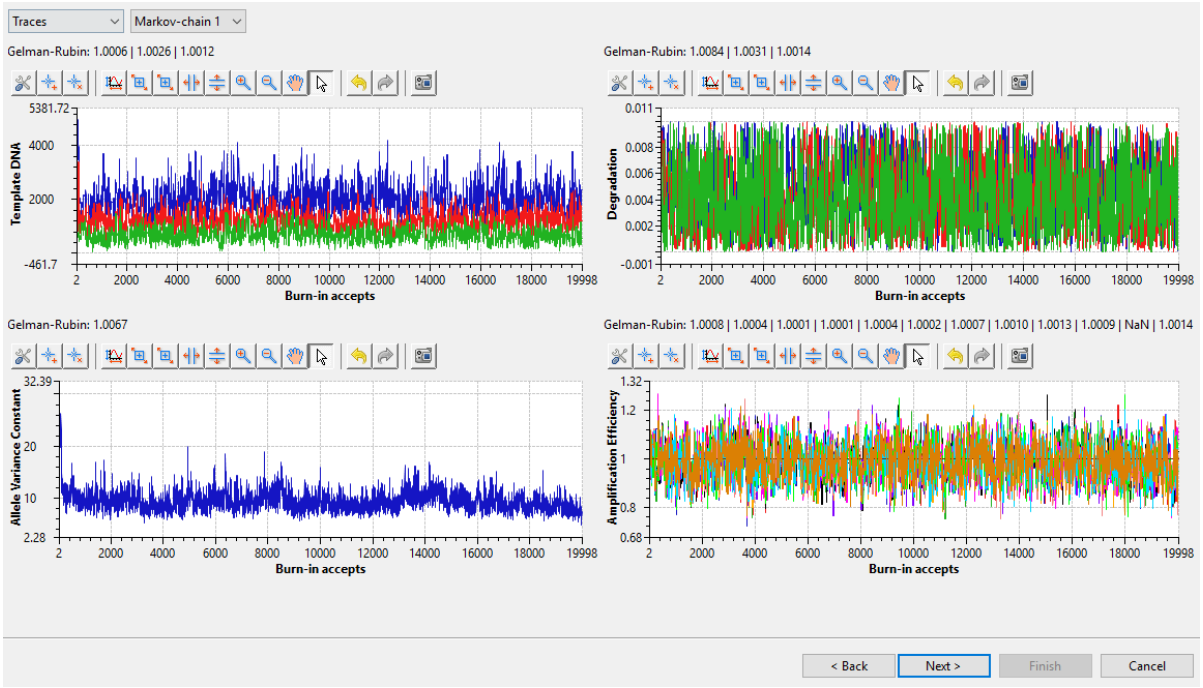

Figure 2: Trace plot. The plots show the current value for each parameter from the biological model during burn-in. It can be noted that the parameterization for the given scenario of a three person mixture was sufficient.

1.2 Design of the developmental validation study

In addition to the recommended guidelines, we have created a continuous validation procedure of the probabilistic system (Fig. 3). The hypotheses for mixture scenarios were analysed automatically by a specific test environment. Each step in software development was examined with appropriate test data of varying degrees of quality (artificially created stains, including reference samples and real cases), before the actual validation procedure started.

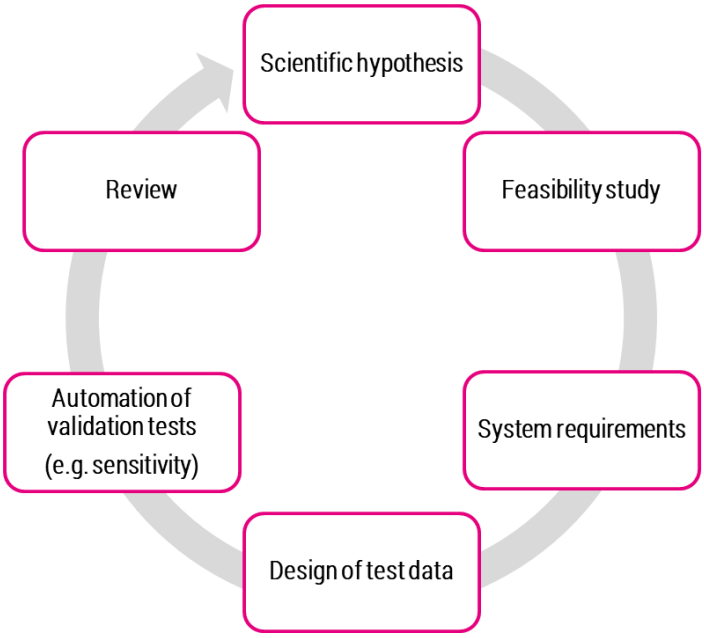

Figure 3: Process of developmental validation. The figure illustrates the process of developmental validation. The first step includes the establishment of a scientific question and the subsequent formulation of a hypothesis. Based on this hypothesis, not only feasibility and system requirements are examined, but also appropriate test data sets are chosen. Subsequently, the required tests for the validation procedure are performed, concluding with a review of the results.

1.3 Sensitivity and specificity study

With respect to interpretation methods, sensitivity studies assess the ability of the system or software to reliably determine the presence of a contributor's DNA in a mixed DNA profile over a wide range of DNA template quantity. Furthermore, different numbers of contributors and mixture proportions should be evaluated. In general, the calculated LR for known or possible contributors (Hp = true) shows high values, but tends to 0 when less information is present within the profile. Therefore, these studies should demonstrate the potential for Type I errors (e.g. incorrect rejection of a true hypothesis), in which, for example, a contributor fails to yield a LR greater than 1 and thus his/her presence in the mixture is excluded. In alliance with the samples of the sensitivity study, different non-contributor samples (min. two non-contributors) are used for the analysis of exclusion hypotheses of the specificity study.

They demonstrate the potential for Type II errors (e.g. failure to reject a false hypothesis); in which, for example, a non-contributor yield a LR greater than 1 and thus his/her presence in the mixture is supported. Sensitivity- and specificity studies also aim to demonstrate the range of LR values that can be expected for known contributors and non-contributors in a mixed DNA profile.

1.3.1 Material and methods

For sensitivity and specificity testing we created several mixed DNA profiles, which differ in DNA template and mixture ratio. Each amount of total DNA template was combined with each mixture ratio of major and minor contributor (see the list below).

Our developmental validations, including sensitivity, precision and specificity tests, consider two and three person mixtures. Artificial sample set up was characterized by known reference samples, mixture ratios and number of contributors. The mixed DNA profiles were generated in-house and amplified with the Mentype® Chimera® Kit (Biotype Diagnostic GmbH, Dresden, Germany) with standard parameters for amplification. Amplified fragments were detected by capillary electrophoresis (Applied Biosystems 3500 Genetic Analyzer).

The following variations in DNA amount and mixture ratios were analysed for the sensitivity study:

Two person mixture:

- Total DNA template: 25 pg, 50 pg, 75 pg, 100 pg, 125 pg, 150 pg, 200 pg, 250 pg and 400 pg
- Mixture proportions (major : minor component): 1:1, 2:1, 3:1, 4:1 and 5:1

Three person mixture:

- Total DNA template: 250 pg
- Mixture proportions (major : minor component): 1:1:1, 3:1:1 and 5:1:1

In total, 250 profiles (two and three person mixtures) were used to validate the implemented probabilistic model as well as the accuracy of LR calculations. For each constellation of DNA amount and mixture ratio, the LR was calculated five times with the fully continuous model and compared with the results of the binary and the semi-continuous model. All given constellations were analysed in case of inclusion and exclusion, considering the outlined hypotheses (Table 1).

In order to assess the stability of the LR calculations for each given scenario, we calculated a *coefficient of variation* (CV), also known as *relative standard deviation (SD)*. With the help of this statistical measurement, the dispersion of estimated LRs was characterized.

In addition to the samples created for the sensitivity study (inclusion hypotheses), two known non-contributor samples were used for validation of system specificity (exclusion hypotheses) to assess the evidential weight of the LR calculations.

To reproduce the typing result, PCR replicates of the stain are often analysed in the lab. GenoProof Mixture 3 enables the user to calculate LRs from replicated DNA profiles that were obtained from the same or different STR kits. In this case, we analysed the informative value of calculated LRs and the duration of each calculation.

Nr. of contributors to the stain	Hypotheses	
	Inclusion Pr (E | Hp)	Exclusion Pr (E | Hd)
two 2 person mixture	Hp: DNA originated from the major contributor and n-1 other individuals Hd: DNA originated from two unknown individuals	Hp: DNA originated from a known non-contributing individual (person 3) and n-1 other individuals Hd: DNA originated from two unknown individuals
	Hp: DNA originated from the minor contributor and n-1 other individuals Hd: DNA originated from two unknown individuals	Hp: DNA originated from a second known non-contributing individual (person 4) and n-1 other individual Hd: DNA originated from two unknown individuals
	Hp: DNA originated from the major and minor contributor Hd: DNA originated from two unknown individuals	2 PCR replicates (equal kit) Hp: DNA originated from a second known non-contributing individual (person 4) and n-1 other individual Hd: DNA originated from two unknown individuals

	2 PCR replicates (equal Kit) Hp: DNA originated from two known individuals Hd: DNA originated from two unknown individuals	
three person mixture	Hp: DNA originated from the major contributor and n - 2 other individuals Hd: DNA originated from three unknown individuals	Hp: DNA originated from a known non-contributing individual (person 4) and n-2 other individuals Hd: DNA originated from three unknown individuals
	Hp: DNA originated from the first minor component and n - 2 other individuals Hd: DNA originated from three unknown individuals	Hp: DNA originated from known non-contributing individual (person 5) and n-2 other individuals Hd: DNA originated from three unknown individuals
	Hp: DNA originated from the second minor component and n - 2 other individuals Hd: DNA originated from three unknown individuals	Hp: DNA originated from known non-contributing Individual (person 6) and n-2 other individuals Hd: DNA originated from three unknown individuals
	Hp: DNA originated from the major contributor; the first minor component and n - 1 other individuals Hd: DNA originated from three unknown individuals	Hp: DNA originated from known non-contributing Individuals (person 4 + 5) and n-1 other individual Hd: DNA originated from three unknown individuals
	Hp: DNA originated from the major	Hp: DNA originated from known non-

	contributor; the second minor component and n - 1 other individuals Hd: DNA originated from three unknown individuals	contributing Individuals (person 4 + 6) and n-1 other individual Hd: DNA originated from three unknown individuals
	Hp: DNA originated from both minor components and n - 1 other individuals Hd: DNA originated from three unknown individuals	Hp: DNA originated from known non-contributingd Individuals (person 5 +6) and n-1 other individual Hd: DNA originated from three unknown individuals
	Hp: DNA originated from the major component and both minor components Hd: DNA originated from three unknown individuals	Hp: DNA originated from known non-contributing Individuals (person 4 + 5 + 6) Hd: DNA originated from three unknown individuals
	2 PCR replicates (equal kit): Hp: DNA originated from the major component and both minor components Hd: DNA originated from three unknown individuals	2 PCR replicates (equal kit): Hp: DNA originated from known non-contributing Individuals (person 4 + 5 + 6) Hd: DNA originated from three unknown individuals

Table 1: Analysed Hypotheses. The table is comprised of the complete analysed set of hypotheses for two and three person mixtures. The formulation of the hypotheses was based on known contributors and known non-contributors from the experimental setup.

1.3.2 Results and discussion

Sensitivity and specificity test with GenoProof Mixture 3 should demonstrate the range of LR values that can be expected for known contributors and known non-contributors with respect to DNA dilution series combined with different mixture ratios.

For both, two and three person mixtures, the calculated LR (fully continuous model) decreased with lower DNA template amounts. For each considered set of hypotheses (Table 1), samples with a total DNA amount ≥ 100 pg showed stable and high LRs > 3.0xE+10 (Fig. 2, green area). For samples with DNA amounts < 100 pg, we achieved significant LRs up to 1.0E+10 (Fig. 4, yellow area). The variation between the five calculations of the LR for each tested hypotheses set correlates with the quality of the STR profile (Fig. 4, represent by CV). Although, even for low-template DNA (LTDNA < 100 pg) meaningful results could be achieved. All hypothetical inclusions (Hp = true) for two and three person mixtures were identified reliably. Assuming that only the second person (minor contributor) contributes to the stain, the LR decreases with increasing complexity of the mixture ratio.

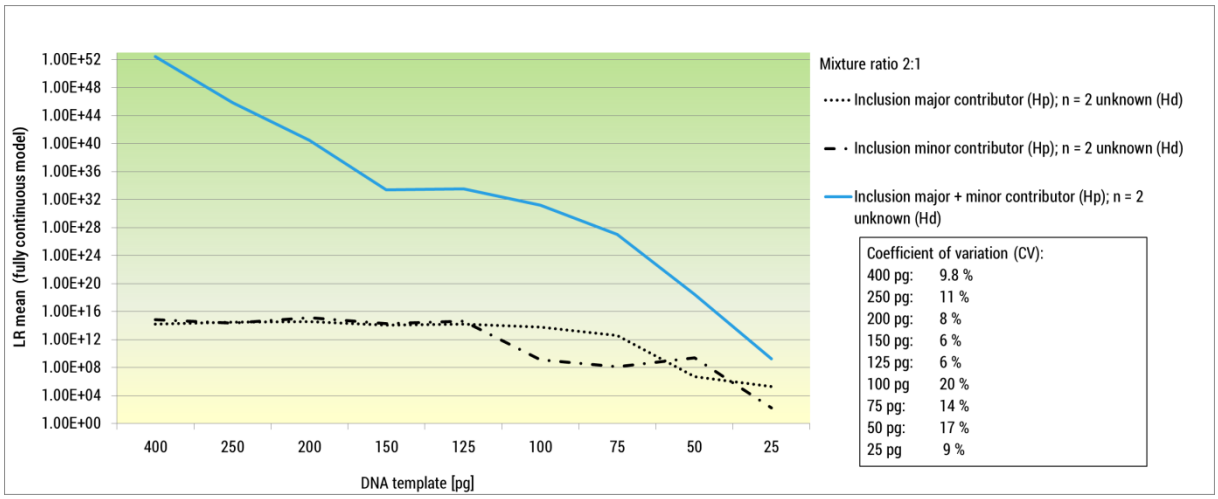

Figure 4: Impact of DNA template amount on LR (fully continuous model) for two person mixtures. The LR decreases with decreasing amount of DNA in all sets of hypotheses. The LR increases when all contributors were included in the formulation of the hypothesis (Inclusion of person 1+2). Samples with a LR > 3.0E+10 are depicted in the green area and samples with a LR between 1 and 1.0E+10 are found in the yellow area. The range for the calculated coefficient of variation (CV) for redundant LR calculations was between 6 % and 20 % (box).

In general, GenoProof Mixture 3 enables the analysis of PCR replicates for the LR calculation. Our experiments have shown that the evidential weight of the LR increases, when two PCR replicates are included (Fig. 5, magenta line). However, this also means higher computational costs. In general, it could be observed that the calculated deviation regarding run time depends on both the quality of the STR profiles and the number of assumed persons in the stain.

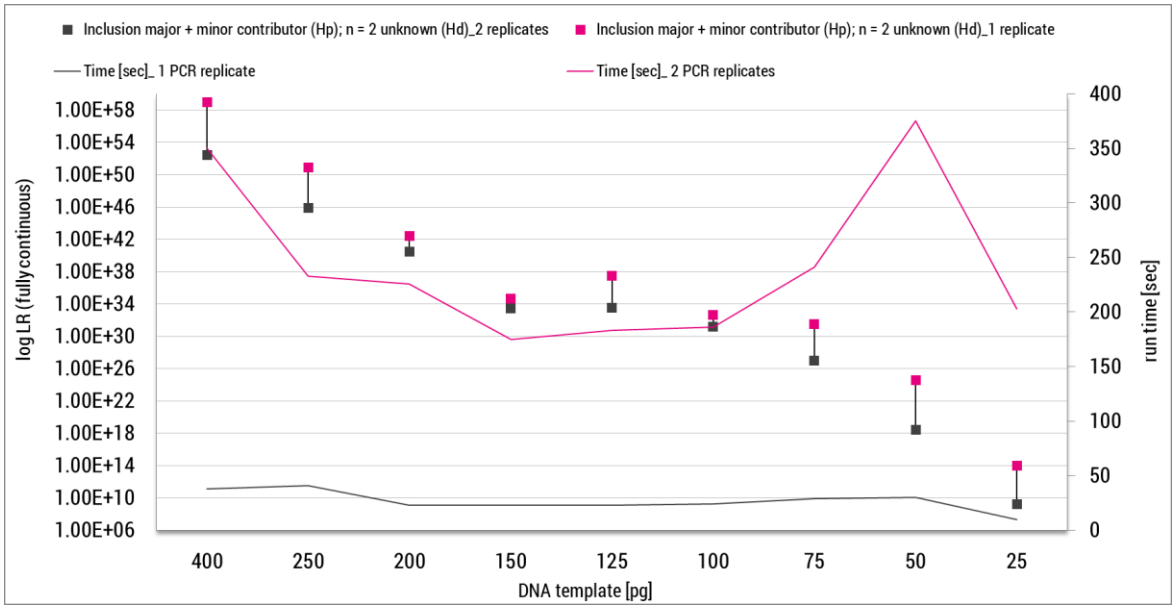

Figure 5: Impact of PCR replicates on LR (fully continuous model) for two person mixtures. The LR increases when two PCR replicates are included in the calculation (magenta squares). Average run time for LR calculation with one PCR replicate (grey line) was 26 sec and with two PCR replicates (magenta line) 241 sec. An increase of the run time was observed at 50 pg (LTDNA) and two PCR replicates. This phenomenon can be explained by quality differences regarding the two PCR replicates. Moreover, a lack of allele information in one of the replicates occurred.

Validation of three person mixtures showed a strong correlation of investigated mixture ratios with the LR, depending on the tested hypotheses set (Fig. 6). As the major component in the mixture increased, the LR also increased. Even with reduced consideration of both minor components, a continuous reduction of the LR could be observed. In the case of the hypothesis that the DNA sample originated from the major component and both minor components, no correlation between LR and mixture ratio was found in the simultaneous consideration of the major and minor component.

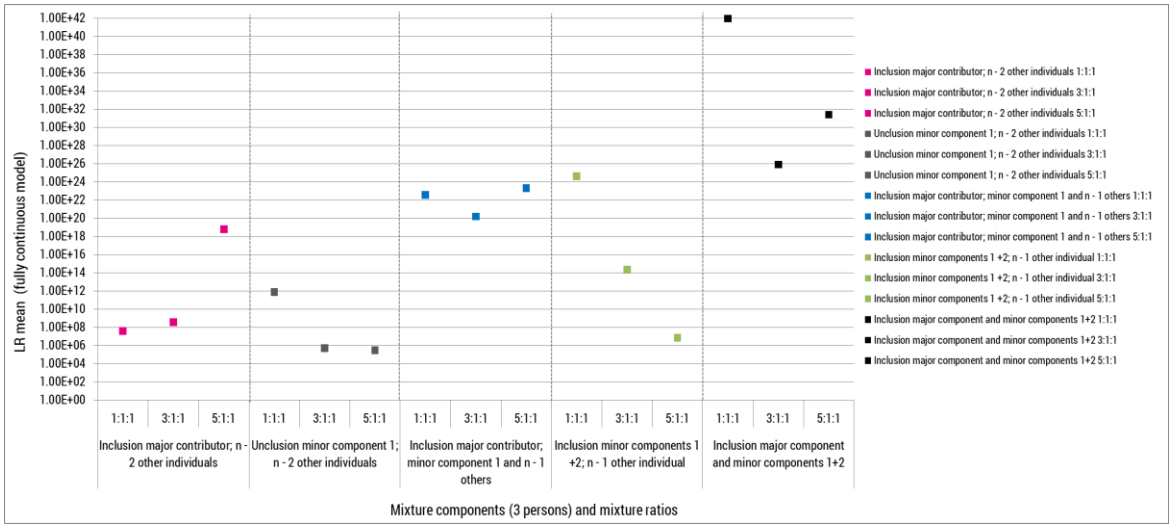

Figure 6: Impact of mixture ratio on LR (three person mixture). The influence of major and minor contributor of a three person mixture is illustrated. There is a clear trend regarding the LR (mean value) in the case of deconvolution of the single components (major and both minor components) in the mixture, depending on the mixture ratios. Taking only the major contributor into account, the LR increases with an increased major component in the mixture ratio. Even with reduced consideration of both minor components, a continuous reduction of the LR could be observed.

Assuming that two persons contributed to a two person mixture, it was possible to achieve LRs > 3.0 E+10 [8] with the fully continuous model in 88.9 % of inclusion cases (Fig. 7). By comparison, only 26 % of the calculated LRs with the semi-continuous model could be classified in the range > 3.0E+10 (Fig. 7). In some incidents, we saw cases of false negative exclusion of the analysed non-contributors. The values in these scenarios were close to 1, so there is no evidence for inclusion of this person. This occurred when a known non-contributed person had a similar genotype and a decreased profile quality.

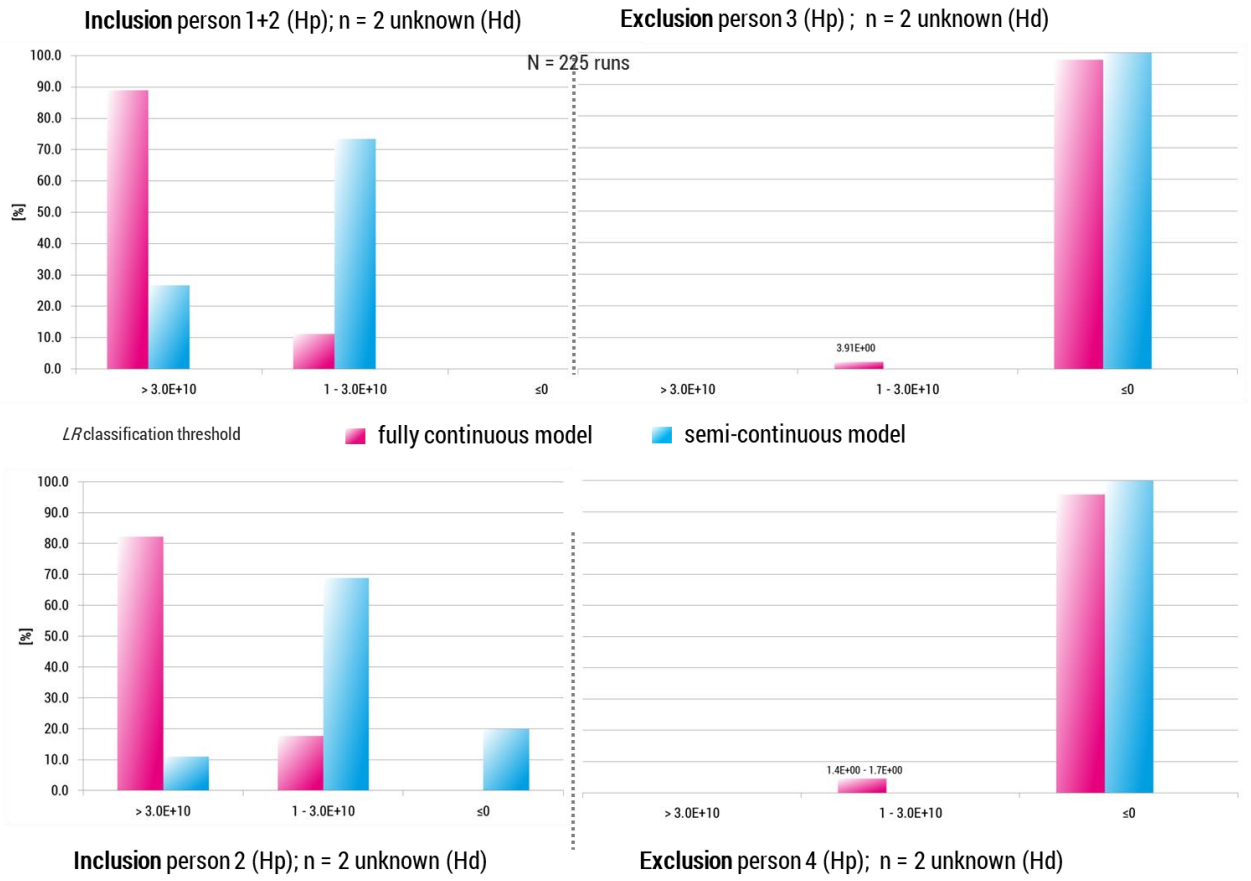

Figure 7: Resolving power by comparison of the semi- and fully continuous model (two person mixture). The chart visualizes the power of the fully continuous model (magenta boxes) in comparison with the semi-continuous model (blue boxes) in case of inclusion (left charts) and exclusion (right charts) hypotheses of two contributors.

When considering three person mixtures, the hypothesis of inclusion of all three known contributors (Hp) in opposition to three unknown individuals (Hd) could be supported with values > 3.0 E + 10 in 100 % of the respective cases. The same applies to the exclusion of known non-contributors (Fig. 8). If both minor components were considered, the robustness of the fully continuous model was confirmed with LRs of 3.0E + 10 up to 66 % of the analysed cases (Fig. 8, bottom chart, on the left).

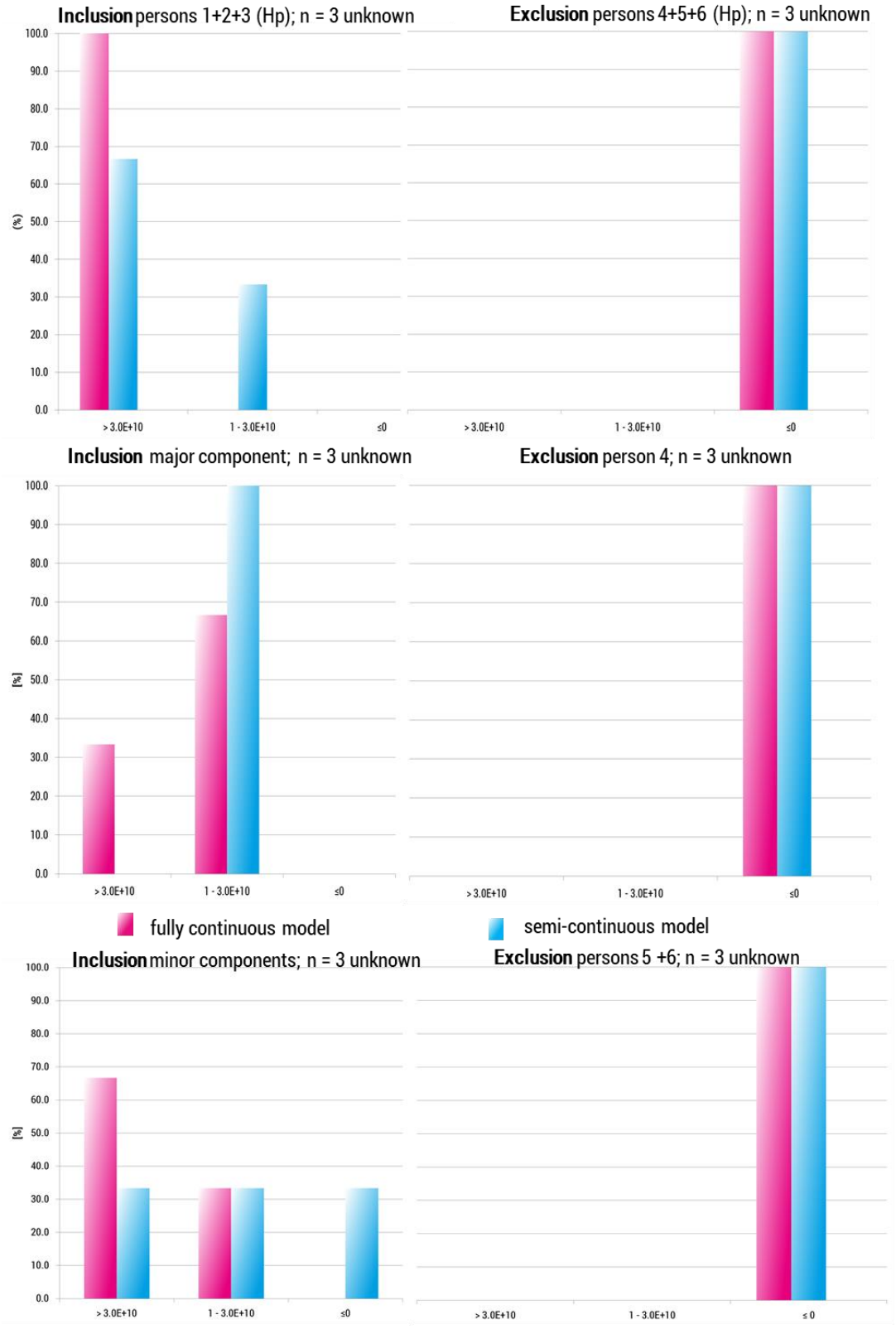

Figure 8: Resolving power by comparison of semi- and fully continuous model (three person mixture). The chart illustrates the power of the fully continuous model (magenta boxes) in comparison with the semi-continuous model (blue boxes) in case of inclusion (left charts) and exclusion (right charts) hypotheses and with three contributors.

The variance between calculated LRs for one analysed hypothesis, represented by the coefficient of variation, is due to the characteristic of the MCMC method. Since the method is based on simulations, every calculation result is different but in a similar range without a decrease of significance of the LR.

With an increasing number of PCR replicates, possible constellations of mass parameters and genotypes increase, resulting in higher calculation time. In this context, individual profile characteristics (e.g. peak height, degradation index, allele variance, number of contributors, DNA template) influence simulation as well.

Overall false inclusions and false exclusions may occur as a result of a combination of specific software, multiplex and operator factors. High-order mixtures, e.g. three or more contributors, increase the chance of false inclusion. A false exclusion can occur when the PCR run was not sufficient and the peak or the stutter heights give misleading information or the operator didn't include all peak information of the respective STR profile in the calculation with the fully continuous model. If that is the case, a lack of possible genotype constellations would occur and the genotype of the POI cannot be aligned sufficiently to the stain or other reference samples.

1.4 Precision study

Precision studies evaluate the variation in LRs calculated from repeated software analyses of the same input data. A phenomenon that is relatively new for the forensic community is the disparity between two analyses using stochastic systems like MCMC. Variation in redundant LR calculations produced from GenoProof Mixture 3 depends on both the sample (number of contributors, DNA profile quality, and STR information in the profile) and the run parameters (e.g. number of accepted iterations).

Where applicable, these studies therefore demonstrate the range of LR values that can be expected from multiple analyses of the same data and are the basis for establishing an acceptable amount of variation in LRs.

Any parameter settings (e.g. iterations of the MCMC) that can reduce variability should be evaluated. For example, for some complex mixtures (e.g. partial profiles with more than three contributors), increasing the number of MCMC iterations can reduce variation in the likelihood ratio. For non-probabilistic methods the results must be consistent every time.

1.4.1 Material and methods

For validation, samples of the sensitivity study were used. The variation of the calculated LR values, due to sample factors and run-specific parameters in GenoProof Mixture 3, has been explored for two person and three person mixtures with DNA amounts close to the respective sensitivity threshold of the used STR-Kit.

Only one parameter was changed per analysis run:

- Burn-in accepts: 10000 to 100000, step size 10000
- Iteration accepts: 10000 to 100000, step size 10000
- Maximum degradation: 0 to 0.03, step size 0.005
- Maximum pre-stutter ratio: 0.2 to 0.4, step size 0.05 and 0
- Drop-in value alpha: 0 to 0.1, step size 0.01
- Drop-in value beta: 0 to 0.1, step size 0.01
- Drop-in frequency: 0.0005 to 0.002, step size 0.0005
- Minimum allele variance: 0 to 10, step size 1
- Maximum allele variance: 10 to 50, step size 5
- Amplification efficiency variance: 0 to 0.1, step size 0.01

Each STR profile was analysed five times for each parameter setting in the given ranges, by calculating the LR with the fully continuous, semi-continuous and binary model. The data was analysed to determine the parameter configuration resulting in stable and plausible LRs, also considering the calculation time. The sample set up was the following:

Two person mixture:

- Total DNA template: 100 pg, 125 pg, 150 pg, 200 pg, 250 pg and 400 pg
- Mixture proportion (major : minor component): 3:1

Three person mixture:

- Total DNA template: 250 pg
- Mixture proportion (major : minor component 1 : minor component 2): 3:1:1

1.4.2 Results and discussion

Variation of the LRs in the tested parameter ranges, CV and runtime for two- and three person mixtures for the given default parameter settings were assessed (e.g. Burn-In, accepted-iterations, maximum degradation, maximum pre-stutter value). In summary, all hypothetical inclusions and exclusions were identified reliably. Nevertheless, specific changes in person- and PCR- replicate

dependent parameters have major effects on the overall result (LR) for hypotheses including or excluding one or more persons.

A relative variation of 8 % to 12 % between calculated LRs, but no improvement or deterioration could be determined for the parameters "burn-in" and "iteration accepts". However, a decrease of the iteration step < 50.000 is not recommended, thereby ensuring a meaningful solution finding in the sequence of simulation and iteration steps [5].

For both the selected two-person and three-person mixtures, the greatest observed impact on the overall outcome for the hypotheses sets was found for the parameters “degradation”, “drop-in alpha and beta value” and “minimal allele variance “. In the setup for the parameter degradation (range: 0 - 0.03) an increase of the LR from 1.36E+32 up to 1.58E+33 could be observed (two person mixture). Starting from the default settings a higher degradation value increases the runtime by a factor of at least 10 (min value - max value). With the determination of a conservative, kit-independent parameter selection, it is possible to map the influence of person-dependent degradation to amplification efficiency. Assessing the degree of degradation is within the responsibility of the user. If, in the opinion of the user, a high degree of degradation should be taken into account for the calculation, the default value can be increased up to 0.2. However, a significant increase of runtime must be expected. The pre-stutter value can be adapted according to the kit specifications. Ignoring the parameter with a setting (0) can lead to false positive results. Essentially, the extension of the potential solution space for each parameter leads to an extension of the runtime. The same was observed with the variation of the drop-in alpha values (0 - 0.1). Here, we recommend the values in the literature. The height-dependence of an observed drop-in event is defined by the beta value. The higher the value, the greater the influence of the peak height of a drop-in event. For this parameter, a 10-fold increase in run-time was observed in the tested limits, but had no influence on the LR.

If the alpha value equals the beta value, high or low drop-ins are equally probable. Depending on the case study, the expert can decide on the analysis limits. An increase of the minimal limit for “allele variance” (up to 10) led to a decrease of the LR from 1.37E+32 up to 8.18E+29 (two person mixture). Changes to the upper limit of the allele variance parameter did not significantly change the predictive power of the LR. The parameter “amplification efficiency” should be evaluated depending on the profile quality, together with the degradation. The following table presents the accepted range for each parameter setting (Table 2). In general, the user should take profile quality and number of contributors into account. Adaption of parameters should always be carried out with respect to quality control visualised by the respective Gelman-Rubin value, running-mean and trace plot. Non-convergence can be diagnosed in the software using the Gelman-Rubin statistic.

A Gelman rubin value > 1.01 suggests the extension of the iteration acceptance steps.

It can be concluded that the evaluation of the specificity study and the observed effects on the LR depends on the test data. The choice of parameters therefore also depends on the number of contributors to the stain and the quality of the profile.

Nr. of contributed persons	Parameter	Recommend range (min – max)
two and three person mixture	Burn-In	20000
	Iteration accepts	100000
	Max degradation	0 – 0.02
	Max pre-stutter	0.3 (see STR kit specification)
	Drop-in value alpha	0.05
	Drop-in value beta	0 – 0.5
	Drop-in frequency	0.001
	Min allele variance	0
	Max allele variance	50 (increased number of PCR cycles) - 30
	Amplification efficacy	0.001

Table 2: Recommended parameter settings in GenoProof Mixture 3.0.4. This table presents recommended settings for each parameter. The settings can vary from stain to stain, depending on the influencing factors (e.g. quality of the STR-profile, number of the contributors, number of replicates etc.)

1.5 Number of contributors

GenoProof Mixture 3 offers the user the possibility to determine the number of possible contributors via two methods (consensus and composite method). Thus, the subjective impression of the stain can be minimized. Nevertheless, in exceptional cases an overestimation of the number of possible contributors cannot always be determined unambiguously due to STR profile artefacts (e.g. high stutter or drop-ins). In general, the determination of effects of incorrectly assigned number of contributors is not a required part of developmental validation. However, it is an interesting fact that we wanted to explore. In many cases it may be assigned with some confidence with information from the profile itself, and with case and sample information. In the presence of ambiguous peaks it may be tempting to increase the number of assumed contributors, resulting in the over- or underestimation of the LR, depending on the defined hypothesis.

The probability of a given number of contributors is influenced by how likely this number is, given the case circumstances and how well this number of contributors explains the profile [9]. Therefore, we can monitor the behaviour of the LR under certain situations, for example when the Hp or Hd is supported.

1.5.1 Material and methods

In order to investigate the influence of the number of possible contributors, a known two person mixture from the previous sensitivity and specificity study was selected. The ratio of major and minor contributor was 3:1 and the total amount of DNA used, was 100 pg.

The tested number of hypothetical contributors of the stain varied from two to four. The influence of an increasing number of persons on the LR was examined under three different inclusion scenarios and the exclusion of two persons, which were not included in the stain. In the result, the LR was calculated and evaluated regarding the supported hypothesis (Hd or Hp).

1.5.2 Results and discussion

When using the continuous approach, the assumption of an incorrect number of contributors of a mixed DNA profile does not affect the weight of evidence assigned to a distinct major contributor. It could be observed, that an overestimated number of contributors decrease the LR assigned to known minor contributors and in assumption of the hypothesis of the prosecution, that both known persons are included in the stain. In case of the exclusion of the two known persons, the LR increases from 0 up to 1.58E+4 when three or four persons were assumed hypothetically. A direct exclusion of known non-contributors took place only with the assumption of the real number of contributors to the stain.

References:

[1] S. W. G. o. D. A. Methods, “SWGDAM Guidelinesfor the Validation of Probabilistic Genotyping Systems,” 2015. [Online]. Available: https://www.swgdam.org/publications. [Accessed 2017].

[2] M. Coble, J. Buckelton, J. Butler and T. Egeland, “DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications,” *Forensic Science International: Genetics,* 2016.

[3] D. a. B. J. Balding, “ Interpreting low template DNA profiles,” *Forensic Science International: Genetics,* pp. 1-10, 2009.

[4] F. Götz, H. Schönborn, V. Bordorf, A.-M. Pflugbeil and D. Labudde, “GenoProof Mixture 3 - New software and process to resolve complex DNA mixtures,” *Forensic Science International: Genetics Supplement Series,* 2017.

[5] D. Taylor, J. Bright and J. Buckelton, “The interpretation of single source and mixed DNA profiles,” *Forensic Science International: Genetics,* 2013.

[6] W. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika,* pp. 97-109, 1970.

[7] N. Metropolis, M. Rosenbluth, A. Teller and E. Teller, “Equations of state calculations by fast computing machines,” *J. Chem. Phys.,* pp. 1087-1091, 1953.

[8] W. Ulbrich, K. Anslinger, G. Bäßler und M. Eckert, „Gemeinsame Empfehlungen der Projektgruppe "Biostatistische DNA-Berechnungen" und der Spurenkommission zur biostatistischen Bewertung von DNA-analytischen Befunden,“ *Rechtsmedizin,* 2016.

[9] D. K. T. D. M. R. D.R. Paoletti, “Inferring the number of contributors to mixed DNA profiles,” *Comput. Biol. Bioinformatics,* pp. 113-122, 2012.

[10] J.-A. Bright, D. Taylor, Curran James M. and J. S. Buckelton, “Developing allelic and stutter peak height models for a continuous method of DNA,” *Forensic Science International: Genetics,* pp. 296 - 304, 2013.

[11] D. Tayler, J. Bright and B. J., “The interpretation of single source and mixed DNA profiles,” *Forensic Science International: Genetics,* 2013.