Int J Legal Med
DOI 10.1007/s00414-008-0244-4

REVIEW ARTICLE

The German Stain Commission: recommendations for the interpretation of mixed stains

P. M. Schneider · R. Fimmers · W. Keil · G. Molsberger · D. Patzelt · W. Pflug · T. Rothämel · H. Schmitter · H. Schneider · B. Brinkmann

Received: 8 April 2008 / Accepted: 16 April 2008

Abstract In the course of forensic DNA analysis, the interpretation of DNA profiles of mixed stains, i.e. cell material from more than a single donor, has become increasingly more important. The German Stain Commission, a joint commission of Institutes of Forensic Science and Legal Medicine, has therefore developed guidelines aiming to harmonize the evaluation of mixed stains in German criminal cases.

Keywords Short tandem repeat typing · Biostatistical analysis · Likelihood ratio · Probability of exclusion · Mixtures

Preface

Since the beginning of forensic stain analysis, mixed stains have been observed [1, 2]. Over the past few years, they have

Schneider, Keil, Molsberger, Patzelt, Pflug, Rothämel, Schmitter, Schneider, and Brinkmann are members of the Stain Commission, the joint commission of Institutes of Legal Medicine and Forensic Science in Germany.

P. M. Schneider
Institut für Rechtsmedizin, Universitätsklinikum Köln,
Melatengürtel 60-62,
50823 Cologne, Germany

W. Keil
Institut für Rechtsmedizin, Universität München,
Nußbaumstraße 26,
80336 Munich, Germany

H. Schmitter
Bundeskriminalamt,
Äppelallee 45,
65203 Wiesbaden, Germany

R. Fimmers
Institut für Medizinische Biometrie,
Informatik und Epidemiologie, Universität Bonn,
Sigmund-Freud-Str. 25,
53105 Bonn, Germany

B. Brinkmann (✉)
Forensische Genetik,
Röntgenstr. 23,
48149 Münster, Germany
e-mail: brinkma@uni-muenster.de

G. Molsberger
Landeskriminalamt Nordrhein-Westfalen,
Völklingerstr. 49,
40221 Düsseldorf, Germany

D. Patzelt
Institut für Rechtsmedizin,
Julius-Maximilians-Universität,
Versbacher Str. 3,
97078 Würzburg, Germany

W. Pflug
Landeskriminalamt Baden-Württemberg,
Taubenheimstr. 85,
70372 Stuttgart, Germany

T. Rothämel
Institut für Rechtsmedizin, Medizinische Hochschule,
Carl-Neuberg-Str. 1,
30625 Hannover, Germany

H. Schneider
Hessisches Landeskriminalamt,
Hölderlinstrasse 5,
65187 Wiesbaden, Germany

gained importance as evidence due to improved analytical methods and the enormous increase in the numbers of investigated stains [3, 4]. While the interpretation of single source stains usually does not cause problems [5], the evaluation and interpretation of mixed DNA stains requires particular attention [6–8]. Our recommendations – first published in German [9] – are intended to build a framework for an adequate means of treating typical cases. However, it is beyond the scope of these basic recommendations to address all possible constellations.

Definitions

A stain exhibiting more than two alleles in a single DNA system[1] shall be considered a mixed stain except in the case of genetic irregularities (e.g., trisomy, somatic mosaicism, or duplication). If more than two alleles are observed in at least two DNA systems, the presence of a mixed stain shall be assumed.

The number of possible contributors to a mixed stain shall be derived, if possible:

- In general, the presence of not more than four alleles in a given system allows the assumption of at least two independent stain donors.
- In general, the presence of not more than six alleles in a given system allows the assumption of at least three independent stain donors.
- In general, if more than six alleles are observed in a given system, the exact number of stain donors cannot be reliably determined.

Classification of mixed stains

Type A has no obvious major contributor with no evidence of stochastic effects.[2] Type B has clearly distinguishable major and minor DNA components; consistent peak height ratios of approximately 4:1 (major to minor component) across all heterozygous systems, and no evidence of stochastic effects. Type C has mixtures with no major component(s) and evidence of stochastic effects.

[1] A DNA system is a genetic locus exhibiting a short tandem repeat polymorphism amplified with a pair of defined primers using the polymerase chain reaction (PCR).

[2] DNA profiles obtained from the amplification of samples with low DNA content and/or poor DNA quality, where the occurrence of allelic drop out and/or locus drop out has to be assumed.

Evaluation criteria

Peak analysis

The morphology of a peak shall be typical and fully consistent with an allele of a given short tandem repeat system. Generally, reproducible peaks with heights >50 relative fluorescence units (RFU) can be considered regular peaks if the noise of the baseline is low and the number of PCR cycles recommended by the manufacturer was used.

The presence of peaks exhibiting a low signal strength (i.e., typically below 100 RFU) and/or peaks exhibiting clearly variable intensities shall be annotated in the table of observed alleles. Tables in the final report shall be accompanied by a legend explaining the designations of peak characteristics.

Stutter peaks

Both $n-1$ and $n+1$ stutter peaks may occur. Their heights depend on the DNA systems and the amplification conditions. A stutter peak may, in certain cases, exhibit up to 15% of the height of the corresponding main peak. Furthermore, the following shall be considered for the evaluation of a stutter peak:

- The relative stutter intensities of the alleles of a locus, as well as those between loci of a multiplex amplification.
- The possibility that a stain allele is in the position of a stutter peak.

In case of reasonable doubt, a peak in the position of a stutter peak shall be considered a true allele and part of the DNA profile and shall be included in the biostatistical calculation.

Inclusion/exclusion criteria

Inclusion

If all alleles of a person in question are uniformly present in a mixed stain, the person shall be considered a possible contributor to the stain.

Exclusion

If alleles of a person in question are not present in a mixed stain, the person shall not be considered as a possible contributor to the stain.

Grey area between inclusion and exclusion

The following effects may occur in type C mixtures due to imbalances between the mixture components and may cause

difficulties in reaching an unambiguous decision about inclusion or exclusion across all analyzed DNA systems:

- Locus drop out and allelic drop out (e.g., caused by the sensitivity of the amplification system, as well as by stochastic effects).
- Allelic drop out is more likely to occur for longer than for shorter alleles, and in particular for DNA systems with long amplicon sizes.

Additional criteria

In every case, the decision about inclusion or exclusion shall be made after careful consideration of the issues described under the "Grey area between inclusion and exclusion" section. The reasons shall be explained in detail. If appropriate, it shall be stated why a clear decision about inclusion or exclusion was not possible.

Biostatistical calculations for mixed stains

Basis

The basis for all calculations is the knowledge of the allele frequencies in the relevant population.

Probability of exclusion (P_E)/probability of inclusion (P_I)

P_I represents the combined probability (relative population frequency) of all combinations of genotypes that cannot be excluded to have contributed to the DNA profile of a stain based on the criteria given in the "Inclusion" section. P_I is equivalent to the match probability in the case of a stain originating from a single person.

The calculation of P_I is independent of assumptions about the number of possible contributors to a stain, the genotypes, and the ethnic origin of persons involved in a given case. It is equivalent to the probability that a randomly selected person is a contributor to the stain [=random man not excluded (RMNE)]. The probability of exclusion $P_E = 1 - P_I$ indicates the probability of excluding a randomly selected person as a contributor to a given stain.

Likelihood ratio

The calculation of the likelihood ratio (LR) is based on the assumption of two mutually excluding hypotheses. This imperatively requires the description of a distinct scenario for a given stain case. Both hypotheses explicitly describe alternative scenarios for the origin of a stain. Each of these hypotheses shall clearly state who contributed to the stain and how many unknown contributors are assumed. Then, a calculation of the likelihood for the occurrence of the DNA profile of the stain is performed based on the assumption of the respective hypotheses: $L(\text{stain}|H)$. The LR

$$\text{LR} = \frac{L(\text{stain}|H_1)}{L(\text{stain}|H_2)}$$

allows the evidential value of a stain to be calculated with reference to a specific person involved in a case, e.g., an accused stain donor.

Given a two-person mixed stain M and that all observed alleles can be explained by the genotype of the victim, G_v, and the genotype of the suspect, G_s, the hypotheses can be formulated as follows:

Hypothesis H_p (view of the prosecution): The stain M originates from the victim V and the suspect S.

Hypothesis H_d (view of the defense): The stain originates from the victim V and from an unknown person U unrelated to the suspect.

$$\text{LR} = \frac{L(M|H_p)}{L(M|H_d)} = \frac{L(M|G_v, G_s)}{L(M|G_v, G_u)}$$

The resulting LR provides a numerical value, which indicates how many times more likely the observed DNA profile is under the assumption of the scenario described in H_p compared to the scenario described in H_d.

Procedures

Calculation for a mixed stain with an unambiguous major component from one person

The conclusion of a major DNA profile from a single contributor in a mixed stain shall only be drawn if a peak height ratio of at least 4:1 (major vs minor component) is observed across all heterozygous DNA systems (see "Definitions" section). In this case, the major DNA profile can be considered equivalent to that of a stain originating from a single person, and all calculations can be performed accordingly.

Calculation based on the LR

If the basis for clearly defined and mutually exclusive hypotheses is given, i.e.,

- The number of contributors to the stain can be determined
- Unambiguous DNA profiles across all loci are observed [type A mixtures, or type B, if the person considered as "unknown" contributor, e.g., the suspect, is part of the minor component of the mixture (see "Definitions" section)]

then the calculation of a LR is appropriate.

Calculation based on probability of exclusion/inclusion

If a major DNA profile cannot be identified based on unambiguous DNA profiles, or if the number of contributors cannot be determined, calculations of the probability of exclusion P_E or the probability of inclusion P_I, respectively, for randomly selected persons is appropriate. Also, the calculation of P_E and P_I is always possible for type A and type B mixtures.

Supplementary recommendations

Further calculations that may result in erroneous interpretations of the evidence shall not be performed (e.g. reporting the genotype frequency of a non-excluded suspect, if the mixed stain does not allow a meaningful biostatistical interpretation).

Validated computer programmes for the calculation of complex mixed stains are available.

Acknowledgements We thank K. Anslinger (Munich), P. Berschick (Düsseldorf), M. Eckert (Wiesbaden), C. Hohoff (Münster), S. Jung (Würzburg), and J. Schnee-Griese (Stuttgart) for helpful collaboration. Furthermore, we thank N. Morling (Copenhagen) and S. Rand (Antwerp) for a critical revision of the English text.

Appendix

Examples of the calculations of P_I and P_E

The probability of inclusion P_I is calculated from the sum of all genotypes of possible stain contributors. In a stain case, where a, b, and c denote the alleles of a DNA system detected in the mixture, the sum of all relevant genotypes can be calculated as follows (assuming that allele frequency data conform to Hardy–Weinberg equilibrium):

$$P_I = a^2 + b^2 + c^2 + 2ab + 2bc + 2ac$$

This term can be simplified using the formula for the binominal distribution:

$$a^2 + b^2 + c^2 + 2ab + 2bc + 2ac = (a + b + c)^2$$

Assuming a frequency of 0.1 for alleles a, b, and c, the following result is obtained:

$$P_I = 0.3^2 = 0.09$$

Thus, it is expected that 9% of a group of randomly selected persons will not be excluded as stain contributors. This is equivalent to one out of 11 randomly selected persons (=RMNE). The probability of exclusion is calculated from the difference

$$P_E = 1 - P_I = 1 - 0.09 = 0.91$$

Thus, it is expected that 91% of a group of randomly selected persons will be excluded as stain contributors. For several DNA systems, S_1, S_2,…, S_n, which are genetically unlinked (i.e., in linkage equilibrium), the general expression of $P_E(S_1, S_2,\ldots, S_n)$ can be derived from the product of the individual inclusion probabilities $P(S_j)$ as follows:

$$P_E(S_1, S_2, \ldots, S_n) = 1 - [P_I(S_1) \cdot P_I(S_2) \cdot \ldots \cdot P_I(S_n)]$$

Examples for the calculation of the LR

Simple scenario

Consider a case with a mixed stain M with three alleles, a, b, and c, composed from a victim and a perpetrator. The victim V has the genotype AB, and the suspect S has the genotype BC. The hypotheses can be given as follows:

> H_p: The stain M originates from the victim V and the suspect S.
> H_d: The stain M originates from the victim V and from an unknown person unrelated to the suspect.

Let us first derive the numerator of the LR. The prosecution claims that the stain can be explained by a combination of the genotypes of the victim and the suspect, as there are no unaccounted alleles. Hence, the numerator results as

$$L(M|H_p) = L(M|G_v, G_s) = 1$$

The defense, however, claims that the suspect has not contributed to the stain. The genotype of the suspect is not relevant since the presence of allele c in the mixture must be explained by the contribution of an unknown person. As allele c may have been contributed either by a person homozygous for allele c or from a person heterozygous for c in combination with allele a or b, the denominator is as follows:

$$L(M|H_d) = L(M|G_v, G_u) = 2ac + 2bc + c^2$$

And, thus, the entire expression is given as

$$\mathrm{LR} = \frac{1}{2ac + 2ab + c^2}$$

Assuming a frequency of 0.1 for alleles a, b, and c, the following result is obtained:

$$\mathrm{LR} = \frac{1}{0.02 + 0.02 + 0.01} = \frac{1}{0.05} = 20$$

The result can be described by the following statement: It is 20 times more likely to observe the DNA profile if the mixed stain originated from the victim and the suspect than if it originated from the victim and an unknown person (who is unrelated to the suspect[3]).

Complex scenario

Let us consider a case with a mixed stain M with four alleles a, b, c, and d found on the victim's clothes. The victim's genotype is EF and, hence, the corresponding alleles e and f are not observed in the stain. Suspect S has genotype AB, but there is no known second person who may have contributed the alleles c and d. The hypotheses can be given as follows:

H_p: Stain M originates from suspect S and an unknown person U.
H_d: Stain M originates from two unknown persons $U1$ and $U2$.

The prosecution claims that the stain can be explained by a combination of the suspect's genotype and a second person with the genotype CD. Hence, the numerator results as

$$L(M|H_p) = L(M|G_s, G_u) = 2cd$$

The defense claims that no genotypes of the contributors are known. Thus, the sum of all possible genotype combinations from two persons $U1$ and $U2$ must be considered for the denominator:

Genotypes $U1$	$U2$	Combined frequency $U2$
AB	CD	$2ab \times 2cd = 4abcd$
AC	BD	$4abcd$
AD	BC	$4abcd$
BC	AD	$4abcd$
BD	AC	$4abcd$
CD	AB	$4abcd$
$L(M\|H_d) = L(M\|G_{U1}, G_{U2}) =$		$24abcd$

After reducing the term and by assuming a frequency of 0.1 for alleles a, b, c, and d, the following result is obtained:

$$\mathrm{LR} = \frac{2cd}{24abcd} = \frac{1}{12ab} = \frac{1}{0.12} = 8.3$$

[3] A familial relationship between S and the unknown stain contributor can be considered for calculating LR. However, the exact degree of relationship must be known.

Thus, it is eight times more likely to observe the DNA profile if the mixed stain originated from the suspect and an unknown person than if it originated from two unknown persons. If two suspects $S1$ and $S2$ with the genotypes AB and CD are considered for the same mixed stain scenario, the hypotheses and, hence, the LR change, as no unknown person remains for H_p:

H_p: Stain M originates from the suspects $S1$ and $S2$.
H_d: Stain M originates from two unknown persons $U1$ and $U2$.

Thus, the numerator of the LR is, again, 1. The term cannot be reduced further and the resulting LR is as follows:

$$\mathrm{LR} = \frac{1}{24abcd} = \frac{1}{0.0024} = 416.7$$

Thus, it is 416 times more likely to observe the DNA profile if the mixed stain originated from suspects $S1$ and $S2$ than if it originated from two unknown persons.

We give the following caveat: Additional hypotheses, which are not discussed here, can be formulated. Depending on the precise scenario, such additional hypotheses may be highly relevant in a given case, such as (a) H_p: the stain originates from $S1$ and $S2$; H_d: the stain originates from $B1$ and U, or (b) H_p: the stain originates from $S1$ and $S2$; H_d: the stain originates from $S2$ and U. Depending on the genotype frequencies of $S1$ and $S2$, the resulting LRs may differ significantly.

References

1. Wiegand P, Rolf B (2003) Analyse biologischer Spuren, Teil 1. Rechtsmedizin 13:103–113
2. Hohoff C, Brinkmann B (2003) Trends in der forensischen Molekulargenetik. Rechtsmedizin 13:183–189
3. Pflug W, Nguyen TMH, Merkel J (1997) Zuordnung von Schußwaffen mittels DNA-Analyse. Kriminalistik 12:799–802
4. Schöneberg A, Gerl L, Oesterreich W, Bastisch I, Gerhard M, Kärgel HJ, Fesefeldt A, Pflug W (2003) DNA-Analyse von Hautabriebspuren. Kriminalistik 8–9:497–499
5. National Research Council (1996) The evaluation of forensic DNA evidence. National Academy Press, Washington, D.C.
6. Evett IW, Weir BS (1998) Interpreting DNA evidence: statistical genetics for forensic science. Sinauer, Sunderland
7. Buckleton J, Triggs CM, Walsh SJ (2005) Forensic DNA evidence interpretation. CRC, London
8. Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, Morling N, Prinz M, Schneider PM, Weir BS (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): recommendations on the interpretation of mixtures. Forensic Sci Int 160:90–101
9. Schneider PM, Fimmers R, Keil W, Molsberger G, Patzelt D, Pflug W, Rothämel T, Schmitter H, Schneider H, Brinkmann B (2006) Allgemeine Empfehlungen der Spurenkommission zur Bewertung von DNA-Mischspuren. Rechtsmedizin 16:401–404